UNIVERSITY OF CALIFORNIA, SAN DIEGO

Estimation and Inference of Directionally Differentiable Functions: Theory and Applications

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Economics

by

Zheng Fang

Committee in charge:

Professor Andres Santos, Chair
Professor Brendan K. Beare
Professor Bruce Driver
Professor Graham Elliott
Professor Ruth J. Williams

2015

The dissertation of Zheng Fang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

To my mother for her unconditional support;

And to those who have been constant inspirations to me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

SETS

$\mathbf{R}^+, \bar{\mathbf{R}}^+$:	$\mathbf{R}^+ \equiv \{x \in \mathbf{R} : x \geq 0\}$, $\bar{\mathbf{R}}^+ \equiv \mathbf{R}^+ \cup \{\infty\}$.
$A^\epsilon, \epsilon > 0$:	For A in a metric space (T, d), $A^\epsilon \equiv \{a \in T : d(a, A) \leq \epsilon\}$.
$A \subset_f B$:	For sets A and B, A is a finite subset of B.
$M^\intercal$:	For an $m \times n$ matrix M, $M^\intercal$ is the transpose of M.
K^m_λ :	For $\lambda > 0$, $K^m_\lambda \equiv \{x \in \mathbf{R}^m : \|x\| \leq \lambda\}$.
$[A]$:	For A in a normed space, $[A]$ is the closed linear span of A.
$A^\perp$:	For A in a Hilbert space $\mathbb{H}$, $A^\perp \equiv \{x \in \mathbb{H} : \langle x, y\rangle_{\mathbb{H}} = 0 \ , \forall y \in A\}$.

NORMS AND METRICS

$\|\cdot\|_\infty$:	For a function $f : T \to \mathbf{R}$, $\|f\|_\infty \equiv \sup_{t\in T} |f(t)|$.
$\|\cdot\|_{L^p}, p \in [1, \infty)$:	For f on a measure space $(T, \mathcal{M}, \mu)$, $\|f\|_{L^p} \equiv \{\int |f|^p \, d\mu\}^{1/p}$.
$\|\cdot\|_{L^p(W)}$:	The $\|\cdot\|_{L^p}$ norm with measure μ induced by a random variable W.
$d(a, B)$:	For B in a metric space (T, d), $d(a, B) \equiv \inf_{b\in B} d(a, b)$.
$\vec{d}_H(\cdot, \cdot)$:	For sets A, B, $\vec{d}_H(A, B) \equiv \sup_{a\in A} d(a, B)$.
$d_H(\cdot, \cdot)$:	For sets A, B, $d_H(A, B) \equiv \max\{\vec{d}_H(A, B), \vec{d}_H(B, A)\}$.

SPACES

$\mathbb{B}, \mathbb{D}, \mathbb{E}$:	Normed or Banach spaces.
$\mathrm{BL}_a(T)$:	$\mathrm{BL}_a(T) \equiv \{f : T \to \mathbf{R} \,\big|\, \|f\|_\infty \leq a \ , \ |f(t_1) - f(t_2)| \leq a \cdot d(t_1, t_2)\}$.
$C(T)$:	The space of continuous functions on T with norm $\|\cdot\|_\infty$.
$\ell^\infty(T)$:	The space of bounded functions on T with norm $\|\cdot\|_\infty$.
$L^p(T), p \in [1, \infty)$:	$L^p(T) \equiv \{f : T \to \mathbf{R} : \|f\|_{L^p} \equiv \{f : T \to \mathbf{R} \,\big|\, \|f\|_{L^p} < \infty\}$.

MISCELLANY

$a \lesssim b$:	$a \leq Mb$ for some constant M that is universal in the proof.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude as my Ph.D. life at UCSD is drawing to an end. First of all, I am deeply indebted to my advisor, Andres Santos, for his guidance and support. I got to know him from his Econ 220C class in which I learned microeconometrics. It is, however, his elective class Econ 227 on empirical process theory that led me into a fascinating world and greatly opened my eyes. Later on, as my advisor, he guided me into the BKRW realm on efficient estimation (Bickel et al., 1993) and Le Cam's elegant framework of limits of experiments. I am grateful to him for setting me on the "right" path. I have also benefited a lot from the way he thinks and writes. He is one of the constant inspirations and serves as my role model.

I thank Brendan Beare, Graham Elliott and James Hamilton for their classes and feedback on my research. I am especially thankful for their help and suggestions when I was on the job market. My thanks also go to my committee members, Bruce Driver and Ruth J. Williams, for providing me the training in mathematics which I have enjoyed and benefited very much. I thank Yixiao Sun for making the decision of admitting me into the program in the very beginning given my mediocre profile and for his continued help throughout my time in the department.

I am forever grateful to Kunpeng Li, Qi Li and Zhijie Xiao. They are the main reasons why I made up my mind to take Econometrics as my research field even before I came to UCSD and why I got admitted by the department. I also thank Tirthankar Chakravarty and Jungbin Hwang for all the econometricians' time we had together, and Qihui Chen, Marina Kutyavina, Kam Razavi, Juwon Seo, Zhen Shi, Igor Vaynman and Yang Xie for all their generous help at various stages.

I am thankful to my parents for their understanding and support, and to my girlfriend Shuang Song for being with me through hard times as well as sweet ones.

Chapter 1 and Section 3.1 together is a coauthored work with Andres Santos, titled "Inference on Directionally Differentiable Functions," while Section 3.2 is coauthored with

Brendan K. Beare under the same title. I thank them for their permissions to include the coauthored papers in this dissertation.

VITA

2008	B.A. in Economics, Tsinghua University
2010	M.A. in Economics, Tsinghua University
2013	M.S. in Statistics, University of California, San Diego
2015	Ph.D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Estimation and Inference of Directionally Differentiable Functions: Theory and Applications

by

Zheng Fang

Doctor of Philosophy in Economics

University of California, San Diego, 2015

Professor Andres Santos, Chair

This dissertation addresses a large class of irregular models in economics and statistics – settings in which the parameters of interest take the form $\phi(\theta_0)$, where ϕ is a known directionally differentiable function and θ_0 is estimated by $\hat{\theta}_n$. Chapter 1 provides a tractable framework for conducting inference, Chapter 2 focuses on optimality of estimation, and Chapter 3 applies the developed theory to construct a test whether a Hilbert space valued parameter belongs to a convex set and to derive the uniform weak convergence of the Grenander distribution function – i.e. the least concave majorant of the empirical distribution function – under minimal assumptions.

Chapter 1

Inference on Directionally Differentiable Functions

Abstract

This chapter 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.

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., 2015), 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, 2015).

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

$$r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0 \ . \tag{1.1}$$

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

[1] Interestingly, despite its importance, the origins of the Delta method remain obscure. Ver 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).

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

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0) , \tag{1.2}$$

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 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 (1.2)). 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 (2013b) 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 (1.2) 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

$$H_0 : \phi(\theta_0) \leq 0 \qquad\qquad H_1 : \phi(\theta_0) > 0 \ . \tag{1.3}$$

Special cases of (1.3) 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 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 (Zarantonello, 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 (2014), 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 1.2 formally introduces the model we study and contains a minor extension of the Delta method for directionally differentiable functions. In Section 1.3 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 1.4 concludes. All proofs are contained in the Appendix.

1.2 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.

1.2.1 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 1.2.3-1.2.6 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

$$r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0 \ , \tag{1.4}$$

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 (1.4) 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.

1.2.1.1 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 1.2.1 (Absolute Value of Mean)**.** Let $X \in \mathbf{R}$ be a scalar valued random variable, and suppose we wish to estimate the parameter

$$\phi(\theta_0) = |E[X]| \; . \tag{1.5}$$

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 1.2.2 (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

$$\phi(\theta_0) = \max\{E[X^{(1)}], E[X^{(2)}]\} \; . \tag{1.6}$$

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 (1.6) 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 (2014), and Chetverikov (2012).

Example 1.2.3 (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

$$\phi(\theta_0) = \sup_{f \in \mathcal{F}} E[Yf(Z)] \; . \tag{1.7}$$

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 1.2.4 (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

$$\nu(p, \Lambda) \equiv \sup_{\lambda \in \Lambda} \langle p, \lambda \rangle \qquad p \in \mathbb{S}^d \ . \tag{1.8}$$

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

$$\phi(\theta_0) = \sup_{p \in \mathbb{S}^d} \{\langle p, \lambda \rangle - \nu(p, \Lambda)\} \ , \tag{1.9}$$

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 1.2.5 (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

$$\phi(\theta_0) = \int_{\mathbf{R}} \max\{F^{(1)}(u) - F^{(2)}(u), 0\} w(u) du \ , \tag{1.10}$$

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 (2015).

Example 1.2.6 (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

$$\mathcal{M}f(u) = \inf\{g(u) : g \in \ell^\infty([0,1]) \text{ is concave and } f(u) \leq g(u) \text{ for all } u \in [0,1]\} \quad (1.11)$$

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 (2015) note

$$\phi(\theta_0) = \Big\{ \int_0^1 (\mathcal{M}G(u) - G(u))^2 du \Big\}^{\frac{1}{2}} \quad (1.12)$$

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 $\theta \in \ell^\infty([0,1])$. ■

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 1.2.1 and 1.2.2, in which the domain of ϕ is a finite dimensional space. In order to address Examples 1.2.3-1.2.6, 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 1.2.1. 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:

$$\lim_{n\to\infty} \|\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h)\|_{\mathbb{E}} = 0 \ , \quad (1.13)$$

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:

$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h)\|_{\mathbb{E}} = 0 \ , \tag{1.14}$$

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 (1.13) and (1.14)). However, despite this similarity, two key differences distinguish Hadamard differentiability from Hadamard directional differentiability. First, in (1.14) 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 (1.14) implies ϕ'_θ must be homogenous of degree one. It is in fact this latter property that distinguishes the two differentiability concepts.

Proposition 1.2.1. *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 1.2.1 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 (1.14) holds. Linearity, however, will play an instrumental role in determining whether the bootstrap is consistent or not.

Remark 1.2.1. 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 1.2.1. ■

Remark 1.2.2. 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 (1.14) holds for some map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ suffices for establishing the Hadamard directional differentiability of ϕ at θ. ■

1.2.2.1 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 1.2.1 (cont.) In this example, simple calculations reveal $\phi'_\theta : \mathbf{R} \to \mathbf{R}$ is

$$\phi'_\theta(h) = \begin{cases} h & \text{if } \theta > 0 \\ |h| & \text{if } \theta = 0 \\ -h & \text{if } \theta < 0 \end{cases} . \tag{1.15}$$

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

Example 1.2.2 (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

$$\phi'_\theta(h) = \begin{cases} h^{(j^*)} & \text{if } \theta^{(1)} \neq \theta^{(2)} \\ \max\{h^{(1)}, h^{(2)}\} & \text{if } \theta^{(1)} = \theta^{(2)} \end{cases} . \tag{1.16}$$

As in (1.15), 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 1.2.3 (cont.) Suppose $E[Y^2] < \infty$ and that $\mathcal{F}$ is compact when endowed with the norm $\|\cdot\|_{L^2(Z)}$. Then, $\theta_0 \in \mathcal{C}(\mathcal{F})$, and Lemma 1.6.8 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

$$\phi'_\theta(h) = \sup_{f\in\Psi_{\mathcal{F}}(\theta)} h(f) \ . \tag{1.17}$$

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 Hadamard differentiable at such θ. We note in this example, $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$. ■

Example 1.2.4 (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 (2013b) then shows that $\phi'_\theta : \mathcal{C}(\mathbb{S}^d) \to \mathbf{R}$ is given by

$$\phi'_\theta(h) = \sup_{p\in\Psi_{\mathbb{S}^d}(\theta)} -h(p) \ . \tag{1.18}$$

As in Example 1.2.3, $\phi : \mathcal{C}(\mathbb{S}^d) \to \mathbf{R}$ 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 1.2.5 (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

$$\phi'_\theta(h) = \int_{B_+(\theta)} (h^{(1)}(u) - h^{(2)}(u))w(u)du + \int_{B_0(\theta)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\}w(u)du \tag{1.19}$$

for $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ – see Lemma 1.6.9 in the Appendix. In particular, if $B_0(\theta)$ has zero Lebesgue measure, then ϕ is Hadamard differentiable at θ. ■

Example 1.2.6 (cont.) Lemma 3.2 in Beare and Moon (2015) 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

$$\phi'_\theta(h) = \Big\{\int_0^1 (\mathcal{M}'_\theta(h)(u) - h(u))^2 du\Big\}^{\frac{1}{2}} \tag{1.20}$$

where $\mathcal{M}'_\theta : \mathcal{C}([0,1]) \to \ell^\infty([0,1])$ is the Hadamard directional derivative of $\mathcal{M}$ at θ. Note in this example, $\mathbb{D}_0 = \mathcal{C}([0,1])$. ■

1.2.3 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 (2015), Kaido (2013b), and Kaido and Santos (2014) 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 1.2.1. (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 1.2.2. (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 $r_n \uparrow \infty$, $r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0$ in $\mathbb{D}$; (ii) $\mathbb{G}_0$ is tight and its support is included in $\mathbb{D}_0$.

Assumption 1.2.3. (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 1.2.1 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 1.2.2(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 1.2.1-1.2.6, but may differ in nonparametric problems. In turn, Assumption 1.2.2(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 1.2.3(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 1.2.3. We emphasize, however, that Assumption 1.2.3(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 (1.14) for $h \in \mathbb{D} \setminus \mathbb{D}_0$. For instance, in Example

1.2.3 ϕ is differentiable tangentially to $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$, but the map ϕ'_θ in (1.17) is naturally well defined and continuous on $\mathbb{D} = \ell^\infty(\mathcal{F})$. Finally, Assumption 1.2.3(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 1.2.3. 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 1.2.3(i) is automatically satisfied. ■

Assumptions 1.2.1 and 1.2.2 suffice for establishing the validity of the Delta method. The probabilistic version of the Delta method, however, additionally requires Assumption 1.2.3.

Theorem 1.2.1. *If Assumptions 1.2.1 and 1.2.2 hold, then* $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$. *If in addition Assumption 1.2.3(i) is also satisfied, then it follows that*

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} = \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) \ . \tag{1.21}$$

The intuition behind Theorem 1.2.1 is the same that motivates the traditional Delta method. Heuristically, the theorem can be obtained from the approximation

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \approx \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) \ , \tag{1.22}$$

Assumption 1.2.2(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 (1.22) holds in an appropriate sense – a condition ensured by Hadamard directional differentiability. Following this insight, Theorem 1.2.1 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 1.2.1 and 1.2.2. Strengthening the Delta method to hold in probability further requires Assumption 1.2.3(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 1.2.1. *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*

$$\hat{\theta}_n(f) \equiv \frac{1}{n}\sum_{i=1}^n f(X_i) \qquad\qquad \theta_0(f) \equiv \int f(x)dP(x) \ . \tag{1.23}$$

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

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

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 1.2.1 specializes Theorem 1.2.1 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 1.2.1 immediately delivers the relevant asymptotic distributions in Examples 1.2.1, 1.2.2, 1.2.3 and 1.2.5, but not in Examples 1.2.4 or 1.2.6. In the latter two examples $\hat{\theta}_n$ and θ_0 do not take the form in (1.23), and we therefore need to employ Theorem 1.2.1 together with the asymptotic distribution of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ as available, for example, in Kaido and Santos (2014) for support functions and Beare and Moon (2015) for Example 1.2.6.

[2] Without Assumption 1.2.3(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.

1.3 The Bootstrap

While Theorem 1.2.1 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 is often lost under Hadamard directional differentiable transformations, and propose an alternative resampling scheme which generalizes existing approaches in the literature.

1.3.1 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

$$r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \tag{1.24}$$

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 1.3.2.

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

$$\mathrm{BL}_1(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} : \sup_{a \in \mathbf{A}} |f(a)| \leq 1 \text{ and } |f(a_1) - f(a_2)| \leq \|a_1 - a_2\|_{\mathbf{A}}\} . \tag{1.25}$$

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

$$d_{\mathrm{BL}}(L_1, L_2) \equiv \sup_{f \in \mathrm{BL}_1(\mathbf{A})} |\int f(a) dL_1(a) - \int f(a) dL_2(a)| \,. \tag{1.26}$$

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]

$$\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| \,. \tag{1.27}$$

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 (1.27), converges in probability to zero. This type of consistency can in turn be exploited to 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 1.3.1.

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

Assumption 1.3.1. (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 $\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| = o_p(1)$.

Assumption 1.3.2. (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 1.3.1(i) defines $\hat{\theta}_n^*$ in accord with our discussion, while Assumption 1.3.1(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 1.3.2 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 1.2.1, where $\hat{\theta}_n$ and $\hat{\theta}_n^*$ correspond to the empirical and bootstrapped empirical processes respectively.

[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.

Remark 1.3.1. In the special case where $\mathbb{D} = \mathbf{R}^d$, Assumption 1.3.1(ii) implies that:

$$\sup_{t \in A} |P(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \leq t | \{X_i\}_{i=1}^n) - P(\mathbb{G}_0 \leq t)| = o_p(1) \tag{1.28}$$

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 (1.28) then readily yields consistency of the corresponding quantiles at points at which the cdf of $\mathbb{G}_0$ is continuous and strictly increasing. ■

Remark 1.3.2. 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

$$\sqrt{n}\{\frac{1}{n}\sum_{i=1}^n X_i^* - \bar{X}\} \ , \tag{1.29}$$

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 (1.29) becomes

$$\sqrt{n}\{\frac{1}{n}\sum_{i=1}^n W_i X_i - \bar{X}\} \ . \tag{1.30}$$

Thus, letting $\hat{\theta}_n^* = \frac{1}{n}\sum_i W_i X_i$ we may express (1.29) in the form $\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}$. ■

1.3.2 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 1.3.1(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

$$r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} \tag{1.31}$$

conditional on the data (van der Vaart and Wellner, 1996). For conciseness, we refer to the law of (1.31) 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 1.3.1. 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

$$\phi'_{\theta_0}(\mathbb{G}_0 + \mathbb{G}_1) - \phi'_{\theta_0}(\mathbb{G}_0) \text{ is independent of } \mathbb{G}_0 \ . \tag{1.32}$$

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

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \tag{1.33}$$

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 1.2.1 and 1.2.2, 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 1.3.1. *Let Assumptions 1.2.1, 1.2.2, 1.2.3, 1.3.1, and 1.3.2 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*

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) \ . \tag{1.34}$$

A powerful implication of Theorem 1.3.1 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 1.3.3. 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 (1.33), which recall does not depend on h. If in addition $0 \in \mathbb{D}$ is in the support of $\mathbb{G}_0$, then from (1.33) we can conclude the bootstrap limit is the desired one, since then

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0) \ , \tag{1.35}$$

where "$\stackrel{d}{=}$" denotes equality in distribution.[4] Relationship (1.35) 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 (1.35) is violated whenever ϕ is not Hadamard differentiable, and hence that the standard bootstrap is inconsistent.

Remark 1.3.3. In Examples 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.5, and 1.2.6 the map $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbf{R}$ satisfies

$$\phi'_{\theta_0}(h_1 + h_2) \le \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2) \tag{1.36}$$

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

$$P\Big(\phi'_{\theta_0}(\mathbb{G}_0 + h^\star) < \phi'_{\theta_0}(\mathbb{G}_0) + \phi'_{\theta_0}(h^\star)\Big) > 0 \ . \tag{1.37}$$

Results (1.36) and (1.37) 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 (1.35), ϕ'_{θ_0} is not $\mathbb{G}_0$-translation invariant, and from Theorem 1.3.1 we conclude the bootstrap fails. ■

[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 1.6.3 in the Appendix for a formal derivation of (1.35).

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

As Theorem 1.3.1 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 1.3.2. *If Assumptions 1.2.1, 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 (1.32)). The converse, however, is a far subtler result which we establish by relying on insights in van der Vaart (1991c) and Hirano and Porter (2012); see Remark 1.3.4. While perhaps not of independent interest, Theorem 1.3.2 has important implications when combined with our previous results. First, in conjunction with Theorem 1.3.1, Theorem 1.3.2 implies that establishing bootstrap consistency reduces to simply verifying the linearity of ϕ'_{θ_0}. Second, together with Proposition 1.2.1, 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 1.3.1 and 1.3.2 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 1.3.1. *Let Assumptions 1.2.1, 1.2.2, 1.2.3, 1.3.1, 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 1.3.4. 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 (1.35)). The proof of Theorem 1.3.2 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 (1991c) and Hirano and Porter (2012) who compare characteristic functions in a limit experiment to conclude regular estimability of a functional implies its differentiability. ■

1.3.3 An Alternative Approach

Theorems 1.3.1 and 1.3.2 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 1.2.1 ($\phi'_{\theta_0}(\mathbb{G}_0)$).

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 1.2.1 can therefore be constructed employing the law of

$$\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) \tag{1.38}$$

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 1.3.3.1 below.

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

Assumption 1.3.3. $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ 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:

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P\Big(\sup_{h \in K^\delta} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > \epsilon \Big) = 0 \ . \tag{1.39}$$

Unfortunately, the requirement in (1.39) is complicated by the presence of the δ-enlargement of K. Without such enlargement, requirement (1.39) 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 1.3.3. It is worth noting, however, that in many applications stronger, but simpler, conditions than (1.39) can be easily verified. For instance, under appropriate additional requirements, the δ factor in (1.39) 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 1.3.5 and 1.3.6.

Remark 1.3.5. In certain applications, it is sufficient to require $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ to satisfy

$$\sup_{h \in K} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1) \ , \tag{1.40}$$

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 (1.40) implies (1.39). 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 (1.40) may be used in place of (1.39).[5] ■

Remark 1.3.6. Assumption 1.3.3 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 $h \in \mathbb{D}_0$

$$\|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1) \tag{1.41}$$

suffices for establishing (1.39) holds; see Lemma 1.6.6 in the Appendix. This observation is particularly helpful in the analysis of Examples 1.2.3 and 1.2.4; see Section 1.3.3.1. ■

Given Assumption 1.3.3 we can establish the validity of the proposed procedure.

Theorem 1.3.3. *Under Assumptions 1.2.1, 1.2.2, 1.2.3(i), 1.3.1, 1.3.2 and 1.3.3, it follows that*

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) \ . \tag{1.42}$$

Theorem 1.3.3 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 1.2.1. In particular, when $\phi(\hat{\theta}_n)$ is a test statistic, and hence scalar valued, Theorem 1.3.3 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 1.3.2. *Let Assumptions 1.2.1, 1.2.2, 1.2.3(i), 1.3.1, 1.3.2 and 1.3.3 hold,* $\mathbb{E} = \mathbf{R}$*, and*

$$\hat{c}_{1-\alpha} \equiv \inf\{c : P(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) \leq c|\{X_i\}_{i=1}^n) \geq 1 - \alpha\} \ . \tag{1.43}$$

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 1.3.3 or Corollary 1.3.2. 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 1.3.3. Nonetheless, alternative estimators are still easily obtained as we next discuss in the context of Examples 1.2.1-1.2.6.

1.3.3.1 Examples Revisited

In order to illustrate the applicability of Theorem 1.3.3, we now return to Examples 1.2.1-1.2.6 and show existing inferential methods may be reinterpreted to fit (1.38). For conciseness, we group the analysis of examples that share a similar structure.

Examples 1.2.1 and 1.2.2 (cont.) In the context of Example 1.2.2, 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 $\kappa_n \to 0$ satisfy $\kappa_n\sqrt{n} \to \infty$, we then define

$$\hat{\phi}'_n(h) = \begin{cases} h^{(\hat{j}^*)} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| > \kappa_n \\ \max\{h^{(1)}, h^{(2)}\} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| \leq \kappa_n \end{cases} , \tag{1.44}$$

(compare to (1.16)). Under appropriate moment restrictions, it is then straightforward to verify Assumption 1.3.3 holds, since $\hat{\phi}'_n : \mathbf{R}^2 \to \mathbf{R}$ in fact satisfies

$$\liminf_{n\to\infty} P\Big(\hat{\phi}'_n(h) = \phi'_{\theta_0}(h) \text{ for all } h \in \mathbf{R}^2\Big) = 1 \ . \tag{1.45}$$

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 (1.38) 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 1.2.1 can be studied in a similar manner and we therefore omit its analysis. ■

Examples 1.2.3 and 1.2.4 (cont.) In Example 1.2.3, 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]

$$d_H(\Psi_{\mathcal{F}}(\theta_0), \hat{\Psi}_{\mathcal{F}}(\theta_0), \|\cdot\|_{L^2(Z)}) = o_p(1) \ . \tag{1.46}$$

[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 $d_H(A,B,\|\cdot\|) \equiv \max\{\vec{d}_H(A,B,\|\cdot\|), \vec{d}_H(B,A,\|\cdot\|)\}$.

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 (1.17))

$$\hat{\phi}'_n(h) = \sup_{f \in \hat{\Psi}_{\mathcal{F}}(\theta_0)} h(f) \ , \tag{1.47}$$

which can easily be shown to satisfy Assumption 1.3.3; see Lemma 1.6.10 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 (1.38) becomes

$$\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}) = \sup_{f \in \hat{\Psi}_{\mathcal{F}}(\theta_0)} \frac{1}{\sqrt{n}} \sum_{i=1}^n \{Y_i^* f(Z_i^*) - \frac{1}{n}\sum_{i=1}^n Y_i f(Z_i)\} \ , \tag{1.48}$$

which was originally proposed in Andrews and Shi (2013) for conducting inference in conditional moment inequalities models. A similar approach is pursued in Kaido (2013b) and Kaido and Santos (2014) in the context of Example 1.2.4. ■

Examples 1.2.5 and 1.2.6 (cont.) Recall that in Example 1.2.5, $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)})$ 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 $B_+(\theta)$ respectively, it is then natural for any $h \in (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ to define

$$\hat{\phi}'_n(h) = \int_{\hat{B}_+(\theta_0)} (h^{(1)}(u) - h^{(2)}(u))w(u)du + \int_{\hat{B}_0(\theta_0)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\}w(u)du \tag{1.49}$$

(compare to (1.19)). For $A \triangle B$ the symmetric set difference between sets A and B, it is then straightforward to verify Assumption 1.3.3 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 1.2.6 we refer the reader to Beare and Shi (2015). ■

1.3.4 Local Analysis

As evidenced in Examples 1.2.1-1.2.6, $\phi(\hat{\theta}_n)$ 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 1.2.1 and 1.3.3 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 1.3.4. (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 1.3.2. 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

$$\lim_{t \to 0} \int \frac{1}{t^2}\Big(\frac{d\wp_t^{\frac{1}{2}}}{d\mu}(x) - \frac{dP^{\frac{1}{2}}}{d\mu}(x) - t\wp_0'(x)\Big)^2 d\mu(x) = 0 \ . \tag{1.50}$$

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 (Le Cam, 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:

[7]For instance, in Examples 1.2.1 and 1.2.2 the known map $P \mapsto \theta(P)$ is given by $\theta(P) \equiv \int x dP(x)$.

Assumption 1.3.5. (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 1.3.5(i) demands that the distributional convergence of $\hat{\theta}_n$ be robust to local perturbations of P, while Assumption 1.3.5(ii) imposes that the parameter $P \mapsto \theta(P)$ be smooth in P. As shown in van der Vaart (1991c), these requirements are closely related, whereby Assumption 1.3.5(i) and mild regularity conditions on $\hat{\theta}_n$ and the tangent space actually imply Assumption 1.3.5(ii). Assumption 1.3.5 is immediately satisfied, for instance, when $\theta(P)$ is a (possible uncountable) collection of moments, as in Examples 1.2.1, 1.2.2, 1.2.3 and 1.2.5. 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 1.3.7.

Remark 1.3.7. 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 1.3.1. *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 1.2.1, 1.2.2, 1.2.3, 1.3.4, and 1.3.5 hold, then*

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P_n))\} \overset{L_n}{\to} \phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)) \ . \tag{1.51}$$

Lemma 1.3.1 characterizes the asymptotic distribution of $\phi(\hat{\theta}_n)$ under a sequence of local perturbations to P. As expected, the asymptotic limit in (1.51) need not equal the pointwise asymptotic distribution derived in Theorem 1.2.1. Intuitively, the asymptotic approximation in (1.51) reflects the importance of local parameters and for this reason

[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 $\bigotimes_{i=1}^n P_n$.

can be expected to provide a better approximation to finite sample distributions – a point forcefully argued in the study of moment inequality models by Andrews and Soares (2010) and Andrews and Shi (2013); see Remark 1.3.8 below.

Remark 1.3.8. In the context of Example 1.2.2, 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 1.2.1 we then obtain

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} \stackrel{L}{\to} \begin{cases} \mathbb{G}_0^{(j^*)} & \text{if } \theta^{(1)}(P) \neq \theta^{(2)}(P) \\ \max\{\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)}\} & \text{if } \theta^{(1)}(P) = \theta^{(2)}(P) \end{cases} , \tag{1.52}$$

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 (1.16)). As argued in Andrews and Soares (2010), the discontinuity of the pointwise asymptotic distribution in (1.52) 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, let 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 1.3.1 yields

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P_n))\} \stackrel{L_n}{\to} \max\{\mathbb{G}_0^{(1)} + h^{(1)}, \mathbb{G}_0^{(2)} + h^{(2)}\} - \max\{h^{(1)}, h^{(2)}\} \,. \tag{1.53}$$

Thus, by reflecting the importance of the "slackness" parameter h, result (1.53) 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 (1.51) we can conclude from Lemma 1.3.1 that $\phi(\hat{\theta}_n)$ is a regular estimator for $\phi(\theta(P))$ if and only if

$$\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0) \tag{1.54}$$

for all curves $\wp$ in $\mathbf{P}$ and all scalars $\eta \in \mathbf{R}$. Therefore, we immediately obtain from Lemma 1.3.1 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 1.3.1 implies $\phi(\hat{\theta}_n)$ will often not be regular when ϕ is directionally, but not fully, Hadamard differentiable at θ_0. Condition (1.54) in fact closely resembles the requirement that ϕ'_{θ_0} be

$\mathbb{G}_0$-translation invariant (compare to (1.35)). 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 1.3.9), $\phi(\hat{\theta}_n)$ is indeed a regular estimator if and only if ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant.

Corollary 1.3.3. *If Assumptions 1.2.1, 1.2.2, 1.2.3, 1.3.4, 1.3.5 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 1.3.3 arises from combining it with Theorem 1.3.1. 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 1.3.3 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 1.3.9. 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 is equal to $\overline{\bigcup_{\wp} \theta'(\wp)}$; 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. ■

1.3.4.1 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 1.2.2, 1.2.3, and 1.2.5 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 1.3.1 to show that whenever such property holds, the bootstrap procedure of Theorem 1.3.3 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

$$H_0 : \phi(\theta(P)) \leq 0 \qquad\qquad H_1 : \phi(\theta(P)) > 0 \ . \tag{1.55}$$

A natural test statistic for this problem is then $\sqrt{n}\phi(\hat{\theta}_n)$, while Theorem 1.2.1 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 booootstrap estimator for $c_{1-\alpha}$ (see (1.43)), Theorem 1.2.1 and Corollary 1.3.2 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, 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 1.3.1, 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

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

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 1.3.6. (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 1.3.6(i) formalizes the requirement that ϕ be scalar valued. In turn, in Assumption 1.3.6(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 1.3.6(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 1.3.6(iii) is in fact equivalent to demanding that ϕ'_{θ_0} be convex, which greatly simplifies verifying Assumption 1.3.6(ii) when $\mathbb{G}_0$ is Gaussian; see Remark 1.3.11. We further note that Assumption 1.3.6 is trivially satisfied when ϕ'_{θ_0} is linear, which by Lemma 1.3.1 also implies $\phi(\hat{\theta}_n)$ is regular. However, we emphasize that Assumption 1.3.6 can also hold at points $\theta(P)$ at which ϕ is not Hadamard differentiable, as is easily verified in Examples 1.2.1-1.2.6.

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

Theorem 1.3.4. *Let Assumptions 1.2.1, 1.2.2, 1.2.3, 1.3.1, 1.3.2, 1.3.3, 1.3.4, 1.3.5, and 1.3.6(i)-(ii) hold. It then follows that for any curve $\wp$ in $\mathbf{H}$, and every $\eta \in \mathbf{R}$ we have*

$$\liminf_{n\to\infty} \pi_n(\wp_{\eta/\sqrt{n}}) \geq P(\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) > c_{1-\alpha}) \ . \tag{1.56}$$

If in addition Assumption 1.3.6(iii) also holds, then we can conclude that for any $\eta \leq 0$

$$\limsup_{n\to\infty} \pi_n(\wp_{\eta/\sqrt{n}}) \leq \alpha \ . \tag{1.57}$$

The first claim of the Theorem derives a lower bound on the power against local alternatives, with (1.56) 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 1.3.4 establishes the ability of the test to locally control size along parametric submodels. Heuristically, the role of subadditivity can be seen from (1.56) 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 (1.57). We note a similar logic can be employed to evaluate confidence regions built using Theorems 1.2.1 and 1.3.3; see Remark 1.3.10.

Since the results of Theorem 1.3.4 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 1.3.5(i), 1.3.6(ii), and 1.3.6(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 1.3.6(ii) as mainly a technical requirement that can be dispensed with following insights in Andrews and Shi (2013); see Remark 1.3.12. 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 1.2.1, 1.2.2, 1.2.3, and 1.2.5, both these requirements are seen to be easily satisfied for a large class of possible P. However, in other instances, such as in Example 1.2.4 applied to estimator in Kaido and Santos (2014), ϕ'_{θ_0} is always subadditive, but the regularity of $\hat{\theta}_n$ can fail to hold for an important class of P.

Remark 1.3.10. As usual, we can obtain confidence regions for $\phi(\theta(P))$ by test inverting

$$H_0 : \phi(\theta(P)) = c_0 \qquad\qquad H_1 : \phi(\theta(P)) \neq c_0 \ , \tag{1.58}$$

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

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

straightforward to see (1.58) can be expressed as in (1.55) 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 1.3.11. Under Assumptions 1.2.1 and 1.3.6(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 1.3.12. In certain applications, such as in Examples 1.2.3 and 1.2.5, Assumption 1.3.6(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 1.3.6(ii) fails, we still have

$$\liminf_{n\to\infty} P(\hat{c}_{1-\alpha} + \delta \geq c_{1-\alpha}) = 1 \; . \tag{1.59}$$

Therefore, by contiguity it follows that the local size control established in (1.57) holds without Assumption 1.3.6(ii) if we employ $\hat{c}_{1-\alpha} + \delta$ instead of $\hat{c}_{1-\alpha}$. ■

1.4 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 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. 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.

1.5 Acknowledgement

Chapter 1 is part of the paper: "Inference on Directoinally Differentiable Functions," coauthored with Andres Santos.

1.6 Appendix

1.6.1 Proofs of Main Results

Proof of Proposition 1.2.1: 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:

$$\lim_{j\to\infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} = \phi'_\theta(h) \ . \tag{1.60}$$

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:

$$\lim_{j\to\infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} = - \lim_{j\to\infty} \frac{\phi(\theta + (-t_{n_{k_j}})(-h_{n_{k_j}})) - \phi(\theta)}{-t_{n_{k_j}}} = -\phi'_\theta(-h) = \phi'_\theta(h) \ , \tag{1.61}$$

where the final equality holds by the assumed linearity of ϕ'_θ. Thus, results (1.60) and (1.61) imply that every subsequence $\{t_{n_k}, h_{n_k}\}$ has a further subsequence along which

$$\lim_{j\to\infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} = \phi'_\theta(h) \ . \tag{1.62}$$

Since the subsequence $\{t_{n_k}, h_{n_k}\}$ is arbitrary, it follows that (1.62) must hold along the original sequence $\{t_n, h_n\}$ and hence ϕ is Hadamard differentiable tangentially to $\mathbb{D}_0$. ■

Proof of Theorem 1.2.1: 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, 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

$$g_n(h_n) \equiv r_n\{\phi(\theta_0 + \frac{h_n}{r_n}) - \phi(\theta_0)\} \tag{1.63}$$

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 1.2.1(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:

$$f_n(h_n, h) = (g_n(h_n), \phi'_{\theta_0}(h)) \ , \tag{1.64}$$

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:

$$\begin{bmatrix} r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\ \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) \end{bmatrix} \stackrel{L}{\to} \begin{bmatrix} \phi'_{\theta_0}(\mathbb{G}_0) \\ \phi'_{\theta_0}(\mathbb{G}_0) \end{bmatrix} \ . \tag{1.65}$$

In particular, result (1.65) and the continuous mapping theorem allow us to conclude:

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} - \phi'_{\theta_0}(r_n(\hat{\theta}_n - \theta_0)) \stackrel{L}{\to} 0 \ . \tag{1.66}$$

The second claim then follows from (1.66) and Lemma 1.10.2(iii) in van der Vaart and Wellner (1996). ■

Proof of Corollary 1.2.1: Follows immediately from Theorem 1.2.1 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 1.3.1: 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 1.6.2 and the continuous mapping theorem imply that:

$$\begin{aligned}(r_n\{\hat{\theta}_n^* - \theta_0\}, r_n\{\hat{\theta}_n - \theta_0\}) \\ &= (r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} + r_n\{\hat{\theta}_n - \theta_0\}, r_n\{\hat{\theta}_n - \theta_0\}) \xrightarrow{L} (\mathbb{G}_1 + \mathbb{G}_2, \mathbb{G}_2) \quad (1.67)\end{aligned}$$

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 1.2.1(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

$$\Phi'_{\theta_0}(h_1, h_2) = \phi'_{\theta_0}(h_1) - \phi'_{\theta_0}(h_2) \quad (1.68)$$

for any $(h_1, h_2) \in \mathbb{D}_0 \times \mathbb{D}_0$. Thus, by Assumptions 1.2.2(ii) and 1.2.3(ii), Theorem 1.2.1, result (1.67), and $r_n\{\hat{\theta}_n^* - \theta_0\} = \mathbb{G}_n^* + \mathbb{G}_n$ we can conclude that

$$\begin{aligned}r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} &= r_n\{\Phi(\hat{\theta}_n^*, \hat{\theta}_n) - \Phi(\theta_0, \theta_0)\} \\ &= \Phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n, \mathbb{G}_n) + o_p(1) = \phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n) + o_p(1)\ . \quad (1.69)\end{aligned}$$

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

$$\sup_{h\in\mathrm{BL}_1(\mathbb{E})} |E^*[h(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\}) - h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]|$$
$$\leq \epsilon + 2P^*(\|r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon|\{X_i\}_{i=1}^n) \quad (1.70)$$

Moreover, Lemma 1.2.6 in van der Vaart and Wellner (1996) and result (1.69) also yield:

$$E^*[P^*(\|r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon|\{X_i\}_{i=1}^n)]$$
$$\leq P^*(\|r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon) = o(1)\ . \quad (1.71)$$

Therefore, since $\epsilon > 0$ was arbitrary, we obtain from results (1.70) and (1.71) that:

$$\sup_{h\in\mathrm{BL}_1(\mathbb{E})} |E^*[h(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]|$$
$$= \sup_{h\in\mathrm{BL}_1(\mathbb{E})} |E^*[h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]| + o_p(1) \quad (1.72)$$

Thus, in establishing the Theorem, it suffices to study the right hand side of (1.72).

First Claim: 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 1.6.2 implies

$$(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n), \mathbb{G}_n) \xrightarrow{L} (\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2), \mathbb{G}_2) \quad (1.73)$$

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 (1.73) we then have:

$$\lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))g(\mathbb{G}_2)] \quad (1.74)$$

On the other hand, also note that if the bootstrap is consistent, then result (1.72) yields

$$\sup_{h\in\mathrm{BL}_1(\mathbb{E})} |E^*[h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1)\ . \quad (1.75)$$

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

$$\begin{aligned}
&\lim_{n\to\infty} E^*[|E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]|g(\mathbb{G}_n)] \\
&\quad \leq \lim_{n\to\infty} E^*[|E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]|] \\
&\quad \leq \lim_{n\to\infty} 2P^*(|E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| > \epsilon) + \epsilon\ .
\end{aligned} \tag{1.76}$$

Thus, result (1.75), ϵ being arbitrary in (1.76), Lemma 1.6.5(v), $g(h) \geq 0$ for all $h \in \mathbb{D}$, and $\mathbb{G}_n \xrightarrow{L} \mathbb{G}_2$ by result (1.73) allow us to conclude that:

$$\begin{aligned}
&\lim_{n\to\infty} E^*[E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]g(\mathbb{G}_n)] \\
&\qquad = \lim_{n\to\infty} E^*[E[f(\phi'_{\theta_0}(\mathbb{G}_0))]g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)]\ .
\end{aligned} \tag{1.77}$$

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

$$\begin{aligned}
&\lim_{n\to\infty} E_*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] \\
&\qquad \leq \lim_{n\to\infty} E^*[E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]g(\mathbb{G}_n)] \\
&\qquad \leq \lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)]
\end{aligned} \tag{1.78}$$

since $\mathbb{G}_n$ is a function of $\{X_i\}_{i=1}^n$ only and $g(\mathbb{G}_n) \geq 0$. However, by (1.73) 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 (1.77) and (1.78) we can conclude:

$$\lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)]\ . \tag{1.79}$$

Hence, comparing (1.74) and (1.79) with $g \in \mathrm{BL}_1(\mathbb{D})$ given by $g(a) = 1$ for all $a \in \mathbb{D}$,

$$\begin{aligned}
E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)] &= E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))]E[g(\mathbb{G}_2)] \\
&= E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))g(\mathbb{G}_2)]\ ,
\end{aligned} \tag{1.80}$$

where the second equality follows again by (1.74) and (1.79). Since (1.80) must hold for any $f \in \text{BL}_1(\mathbb{E})$ and $g \in \text{BL}_1(\mathbb{D})$ with $f(h_1) \geq 0$ and $g(h_2) \geq 0$ 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.

Second Claim: 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 1.2.2, Lemma 1.6.1, and Lemma 1.3.8 in van der Vaart and Wellner (1996), $\mathbb{G}_n$ and $\mathbb{G}_n^*$ are asymptotically tight. Therefore, there exists a compact set $K \subset \mathbb{D}$ such that for any $\delta > 0$:

$$\liminf_{n\to\infty} P_*(\mathbb{G}_n^* \in K^\delta) \geq 1 - \epsilon \qquad \liminf_{n\to\infty} P_*(\mathbb{G}_n \in K^\delta) \geq 1 - \epsilon \ , \tag{1.81}$$

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 1.6.4 and continuity of $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$, there exist scalars $\delta_0 > 0$ and $\eta_0 > 0$ such that:

$$\sup_{a,b\in(K+K)^{\delta_0}:\|a-b\|_{\mathbb{D}}<\eta_0} \|\phi'_{\theta_0}(a) - \phi'_{\theta_0}(b)\|_{\mathbb{E}} < \epsilon \ . \tag{1.82}$$

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 (1.82) that for every

$b \in K^{\delta_1}$ there exists a $1 \leq j(b) \leq J$ such that

$$\sup_{a \in K^{\delta_1}} \|\{\phi'_{\theta_0}(a+b) - \phi'_{\theta_0}(b)\} - \{\phi'_{\theta_0}(a + a_{j(b)}) - \phi'_{\theta_0}(a_{j(b)})\}\|_{\mathbb{E}}$$
$$\leq \sup_{a,b \in (K+K)^{\delta_0}: \|a-b\|_{\mathbb{D}} < \eta_0} 2\|\phi'_{\theta_0}(a) - \phi'_{\theta_0}(b)\|_{\mathbb{E}} < 2\epsilon \ . \quad (1.83)$$

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 (1.83) implies that for every realization of $\mathbb{G}_n$ there is an a_j independent of $\mathbb{G}_n^*$ such that:

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |(f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)) - f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j)))1\{\Delta_n\}| < 2\epsilon \ . \quad (1.84)$$

Letting Δ_n^c denote the complement of Δ_n, result (1.84) then allows us to conclude

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| \leq 2P^*(\Delta_n^c|\{X_i\}_{i=1}^n)$$
$$+ \max_{1 \leq j \leq J} \sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| + 2\epsilon \quad (1.85)$$

since $\|f\|_\infty \leq 1$ for all $f \in \mathrm{BL}_1(\mathbb{E})$. However, by Assumptions 1.3.1(i)-(ii) and 1.3.2(ii), and Theorem 10.8 in Kosorok (2008) it follows that for any $1 \leq j \leq J$:

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_j) - \phi'_{\theta_0}(a_j))]| = o_p(1) \ . \quad (1.86)$$

Thus, since K is a subset of the support of $\mathbb{G}_0$, Lemma 1.6.3, result (1.86), the continuous mapping theorem, and $J < \infty$ allow us to conclude that:

$$\max_{1 \leq j \leq J} \sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) \ . \quad (1.87)$$

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 (1.81) that:

$$\limsup_{n\to\infty} P^*(2P^*(\Delta_n^c | \{X_i\}_{i=1}^n) + 2\epsilon > 6\sqrt{\epsilon}) \leq \limsup_{n\to\infty} P^*(P^*(\Delta_n^c|\{X_i\}_{i=1}^n) > 2\sqrt{\epsilon})$$
$$\leq \frac{1}{2\sqrt{\epsilon}} \times \limsup_{n\to\infty}\{P^*(\mathbb{G}_n \notin K^{\delta_1}) + P^*(\mathbb{G}_n^* \notin K^{\delta_1})\} \leq \sqrt{\epsilon}\ . \quad (1.88)$$

Since $\epsilon > 0$ was arbitrary, combining (1.72), (1.85), (1.87), and (1.88) imply (1.34) holds, or equivalently, that ϕ'_{θ_0} being $\mathbb{G}_0$-translation invariant implies bootstrap consistency. ■

Proof of Theorem 1.3.2: Let P denote the distribution of $\mathbb{G}_0$ on $\mathbb{D}_0$, and note that by Assumption 1.2.2(ii) and Lemma 1.6.7 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:

$$\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_0) - \phi'_{\theta_0}(\mathbb{G}_0) = \phi'_{\theta_0}(\mathbb{G}_1)\ , \quad (1.89)$$

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

$$\mathbb{D}'_P \equiv \left\{d' : \mathbb{D} \to \mathbf{R} : d' \text{ is linear, Borel-measurable, and } \int_{\mathbb{D}} (d'(d))^2 dP(d) < \infty\right\}\ , \quad (1.90)$$

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:

$$\mathbb{H}_P \equiv \{h \in \mathbb{D} : P^{rh} \ll P \text{ for all } r \in \mathbf{R}\}\ . \quad (1.91)$$

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 $h \in \mathbb{H}_P$:

$$\frac{dP^h}{dP}(d) = \exp\left\{\langle d, Ih\rangle_{\mathbb{D}} - \frac{1}{2}\sigma^2(h)\right\} \qquad \sigma^2(h) \equiv \int_{\mathbb{D}} \langle d, Ih\rangle_{\mathbb{D}}^2 dP(d) \ . \tag{1.92}$$

Next, fix an arbitrary $e^* \in \mathbb{E}^*$ and $h \in \mathbb{H}_P$. Then note that Lemma 1.6.3 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 $t \in \mathbf{R}$:

$$\begin{aligned} E[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0)\rangle_{\mathbb{E}}\}] &= E[\exp\{it\{\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh)\rangle_{\mathbb{E}}\}\}] \\ &= \exp\{-itr\langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}}\}E[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0 + rh)\rangle_{\mathbb{E}}\}] \ , \end{aligned} \tag{1.93}$$

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 (1.93) we can then obtain by direct calculation that for all $t \in \mathbf{R}$

$$\begin{aligned} itC(t) \times \langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}} &= \lim_{r\downarrow 0} \frac{1}{r}\{E[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0 + rh)\rangle_{\mathbb{E}}\}] - C(t)\} \\ &= \lim_{r\downarrow 0} \frac{1}{r} \int_{\mathbb{D}} \left\{ \exp\left\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\right\} - C(t)\right\} dP(d) \ , \end{aligned} \tag{1.94}$$

where in the second equality we exploited result (1.92), 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

$$\begin{aligned} &\sup_{r\in(0,1]} \frac{1}{r}\left| \exp\left\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\right\} - \exp\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}}\}\right| \\ &\quad \leq \sup_{r\in(0,1]} \left| \exp\left\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\right\} \times \{\langle d, Ih\rangle_{\mathbb{D}} - r\sigma^2(h)\}\right| \\ &\quad \leq \exp\{|\langle d, Ih\rangle_{\mathbb{D}}|\} \times \{|\langle d, Ih\rangle_{\mathbb{D}}| + \sigma^2(h)\} \ , \end{aligned} \tag{1.95}$$

where the final inequality follows from $\sigma^2(h) > 0$ and $|\exp\{it\{e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}}\}\}| \leq 1$. 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:

$$\int_{\mathbb{D}} \exp\{|\langle d, Ih\rangle_{\mathbb{D}}|\} \times \{|\langle d, Ih\rangle_{\mathbb{D}}| + \sigma^2(h)\}dP(d)$$
$$= \int_{\mathbf{R}} \frac{\{|u| + \sigma^2(h)\}}{\sigma(h)\sqrt{2\pi}} \times \exp\left\{|u| - \frac{u^2}{2\sigma^2(h)}\right\}du < \infty \ . \quad (1.96)$$

Hence, results (1.95) and (1.96) justify the use of the dominated convergence theorem in (1.94). 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

$$\langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}} = -\frac{iE[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0)\rangle_{\mathbb{E}}\}\langle \mathbb{G}_0, Ih\rangle_{\mathbb{D}}]}{t_0 C(t_0)} \ . \quad (1.97)$$

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 (1.97) 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 (1991c) implies $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ must be linear and continuous, which establishes the Theorem. ■

Proof of Corollary 1.3.1: By Theorems 1.3.1 and 1.3.2 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 1.3.3: 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 1.2.2(ii) there is a compact set $K_0 \subseteq \mathbb{D}_0$ such that

$$P(\mathbb{G}_0 \notin K_0) < \frac{\epsilon\eta}{2} \ . \quad (1.98)$$

Thus, by Lemma 1.6.1 and the Portmanteau Theorem, we conclude that for any $\delta > 0$

$$\limsup_{n\to\infty} P(\mathbb{G}_n^* \notin K_0^\delta) \leq P(\mathbb{G}_0 \notin K_0^\delta) \leq P(\mathbb{G}_0 \notin K_0) < \frac{\epsilon\eta}{2} . \tag{1.99}$$

On the other hand, since K_0 is compact, Assumption 1.3.3 yields that for some $\delta_0 > 0$:

$$\limsup_{n\to\infty} P(\sup_{h\in K_0^{\delta_0}} \|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)\|_{\mathbb{E}} > \epsilon)/\epsilon < \eta/2 . \tag{1.100}$$

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:

$$\begin{aligned}
\sup_{f\in \mathrm{BL}_1(\mathbb{E})} &|E[f(\hat{\phi}_n'(\mathbb{G}_n^*))|\{X_i\}] - E[f(\phi_{\theta_0}'(\mathbb{G}_n^*))|\{X_i\}]| \\
&\leq \sup_{f\in \mathrm{BL}_1(\mathbb{E})} E[|f(\hat{\phi}_n'(\mathbb{G}_n^*)) - f(\phi_{\theta_0}'(\mathbb{G}_n^*))||\{X_i\}] \\
&\leq E[2 \times 1\{\mathbb{G}_n^* \notin K_0^{\delta_0}\} + \sup_{f\in K_0^{\delta_0}} \|\hat{\phi}_n'(f) - \phi_{\theta_0}'(f)\|_{\mathbb{E}}|\{X_i\}] \\
&\leq 2P(\mathbb{G}_n^* \notin K_0^{\delta_0}|\{X_i\}_{i=1}^n) + \sup_{f\in K_0^{\delta_0}} \|\hat{\phi}_n'(f) - \phi_{\theta_0}'(f)\|_{\mathbb{E}} ,
\end{aligned} \tag{1.101}$$

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 (1.99) yield:

$$\limsup_{n\to\infty} P(P(\mathbb{G}_n^* \notin K_0^{\delta_0}|\{X_i\}_{i=1}^n) > \epsilon) \leq \limsup_{n\to\infty} P(\mathbb{G}_n^* \notin K_0^{\delta_0}) < \eta . \tag{1.102}$$

Next, also note that Assumption 1.3.1(i) and Theorem 10.8 in Kosorok (2008) imply that:

$$\sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E[f(\phi_{\theta_0}'(\mathbb{G}_n^*))|\{X_i\}_{i=1}^n] - E[f(\phi_{\theta_0}'(\mathbb{G}_0))]| = o_p(1) . \tag{1.103}$$

Thus, by combining results (1.100), (1.101), (1.102) and (1.103) we can finally conclude:

$$\limsup_{n\to\infty} P(\sup_{f\in \mathrm{BL}_1(\mathbb{E})} |E[f(\hat{\phi}_n'(\mathbb{G}_n^*))|\{X_i\}_{i=1}^n] - E[f(\phi_{\theta_0}'(\mathbb{G}_0))]| > 3\epsilon) < 3\eta . \tag{1.104}$$

Since ϵ and η were arbitrary, the claim of the Theorem then follows from (1.104). ■

Proof of Corollary 1.3.2: Let F denote the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$, and similarly define:

$$\hat{F}_n(c) \equiv P(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) \leq c|\{X_i\}_{i=1}^n) \ . \tag{1.105}$$

Next, observe that Theorem 1.3.3 and Lemma 10.11 in Kosorok (2008) imply that:

$$\hat{F}_n(c) = F(c) + o_p(1) \ , \tag{1.106}$$

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:

$$\begin{aligned} &\limsup_{n\to\infty} P(|\hat{c}_{1-\alpha} - c_{1-\alpha}| > \epsilon) \\ &\qquad \leq \limsup_{n\to\infty}\{P(|\hat{F}_n(c_1) - F(c_1)| > \delta) + P(|\hat{F}_n(c_2) - F(c_2)| > \delta)\} = 0 \ , \end{aligned} \tag{1.107}$$

due to (1.106). Since $\epsilon > 0$ was arbitrary, the Corollary then follows. ■

Proof of Lemma 1.3.1: First note that by Assumption 1.3.5(ii) we can conclude:

$$\lim_{n\to\infty} \|\sqrt{n}\{\theta(P_n) - \theta(P)\} - \eta\theta'(\wp)\|_{\mathbb{D}} = 0 \ . \tag{1.108}$$

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 (1.108) 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 1.3.5(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 1.2.1(ii) and Definition 1.2.1 we can conclude that

$$\begin{aligned}\lim_{n\to\infty} \|\sqrt{n}\{\phi(\theta(P_n)) - \phi(\theta(P))\} - \phi'_{\theta_0}(\eta\theta'(\wp))\|_{\mathbb{E}} \\ = \lim_{n\to\infty} \|\frac{\phi(\theta(P) + t_n h_n) - \phi(\theta(P))}{t_n} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = 0 \ . \quad (1.109)\end{aligned}$$

Next, let $P^n \equiv \bigotimes_{i=1}^n P$ and $P_n^n \equiv \bigotimes_{i=1}^n P_n$. By Theorem 1.2.1 we then have that:

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} = \phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) + o_p(1) \quad (1.110)$$

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 (1.109) and (1.110) we obtain

$$\begin{aligned}\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P_n))\} &= \sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} - \sqrt{n}\{\phi(\theta(P_n)) - \phi(\theta(P))\} \\ &= \phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) - \phi'_{\theta_0}(\eta\theta'(\wp)) + o_p(1) \ . \quad (1.111)\end{aligned}$$

under P_n^n. Furthermore, by regularity of $\hat{\theta}_n$ and result (1.108) we also have that:

$$\sqrt{n}\{\hat{\theta}_n - \theta(P)\} = \sqrt{n}\{\hat{\theta}_n - \theta(P_n)\} + \sqrt{n}\{\theta(P_n) - \theta(P)\} \xrightarrow{L_n} \mathbb{G}_0 + \eta\theta'(\wp) \ . \quad (1.112)$$

Thus, we may conclude by (1.111), (1.112) and the continuous mapping theorem. ■

Proof of Corollary 1.3.3: 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 1.6.3, and $\overline{\bigcup_{\wp} \theta'(\wp)} = \mathbb{D}_L$ implies for any $\eta \in \mathbf{R}$ and curve $\wp$ in $\mathbf{P}$:

$$E[h(\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)))] = E[h(\phi'_{\theta_0}(\mathbb{G}_0))] \quad (1.113)$$

for all bounded and continuous $h : \mathbb{E} \to \mathbf{R}$. Letting "$\stackrel{d}{=}$" denote equality in distribution, we then conclude from (1.113) and Lemma 1.3.12 in van der Vaart and Wellner (1996):

$$\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0) \tag{1.114}$$

for any $\eta \in \mathbf{R}$ and curve $\wp$ in $\mathbf{P}$. Thus, result (1.114) and Lemma 1.3.1 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:

$$\Phi(h_0, h_1) \equiv \phi'_{\theta_0}(h_0 + h_1) - \phi'_{\theta_0}(h_0) \ . \tag{1.115}$$

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

$$E[f(\Phi(h_0, \mathbb{G}_1))] = \lim_{n\to\infty} E[f(\Phi(h_{0,n}, \mathbb{G}_1))] = E[f(\Phi(0, \mathbb{G}_1))] = E[f(\phi'_{\theta_0}(\mathbb{G}_1))] \ , \tag{1.116}$$

where the second equality follows from $0 \in \bigcup_{\wp} \theta'(\wp)$ together with Lemma 1.3.1 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 (1.115) and $\phi'_{\theta_0}(0) = 0$. Hence, Fubini's theorem, result (1.116) and $\mathbb{G}_0$ and $\mathbb{G}_1$ being independent allow us to conclude that:

$$\begin{aligned} &E[f(\phi'_{\theta_0}(\mathbb{G}_0 + \mathbb{G}_1) - \phi'_{\theta_0}(\mathbb{G}_0))g(\mathbb{G}_0)] \\ &\qquad = \int_{\mathbb{D}_L} E[f(\Phi(h_0, \mathbb{G}_1))]g(h_0)dP(h_0) = E[f(\phi'_{\theta_0}(\mathbb{G}_1))]E[g(\mathbb{G}_0)] \end{aligned} \tag{1.117}$$

where with some abuse of notation we let P also denote the distribution of $\mathbb{G}_0$ on $\mathbb{D}$. Since (1.117) 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 1.3.4: 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 1.3.2 $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$ under P^n, it follows that we also have:

$$\hat{c}_{1-\alpha} = c_{1-\alpha} + o_p(1) \text{ under } P_n^n \ . \tag{1.118}$$

Moreover, since $\phi(\theta(P)) = 0$, we also obtain from result (1.110) that under P_n^n we have:

$$\begin{aligned}\sqrt{n}\phi(\hat{\theta}_n) &= \sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} \\ &= \phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) + o_p(1) \xrightarrow{L_n} \phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) \ ,\end{aligned} \tag{1.119}$$

where the final result holds for L_n denoting law under P_n^n by result (1.112) and the continuous mapping theorem. Thus, (1.56) holds by (1.119) and the Portmanteau Theorem.

In order to establish (1.57) holds whenever $\eta \leq 0$, first note that (1.109) implies

$$0 \geq \lim_{n\to\infty} \sqrt{n}\{\phi(\theta(P_n)) - \phi(\theta(P))\} = \phi'_{\theta_0}(\eta\theta'(\wp)) \ , \tag{1.120}$$

where we have exploited that $\phi(\theta(P)) = 0$ and $\phi(\theta(P_n)) \leq 0$ for all $\eta \leq 0$. Therefore, result (1.118) together with the second equality in (1.119) allow us to conclude

$$\begin{aligned}&\limsup_{n\to\infty} P_n^n(\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) \\ &\quad \leq \limsup_{n\to\infty} P_n^n(\phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) \geq c_{1-\alpha}) \\ &\quad \leq \limsup_{n\to\infty} P_n^n(\phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P_n)\}) + \phi'_{\theta_0}(\sqrt{n}\{\theta(P_n) - \theta(P)\}) \geq c_{1-\alpha}) \\ &\quad \leq \limsup_{n\to\infty} P_n^n(\phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P_n)\}) \geq c_{1-\alpha}) \\ &\quad = P(\phi'_{\theta_0}(\mathbb{G}_0) \geq c_{1-\alpha}) \ ,\end{aligned} \tag{1.121}$$

where the second inequality follows from subadditivity of ϕ'_{θ_0}, the third inequality is implied by (1.120), and the final result follows from $\sqrt{n}\{\hat{\theta}_n - \theta(P_n)\} \xrightarrow{L_n} \mathbb{G}_0$ by Assumption 1.3.5(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 (1.57) follows. ■

Lemma 1.6.1. *If Assumptions 1.2.1(i), 1.2.2(ii), 1.3.1, 1.3.2(i) hold, then* $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\} \xrightarrow{L} \mathbb{G}_0$.

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 1.6.5(i) and Lemma 1.2.6 in van der Vaart and Wellner (1996) we have that:

$$\begin{aligned} E^*[f(\mathbb{G}^*_n)] - E[f(\mathbb{G}_0)] &\geq E^*[E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n]] - E[f(\mathbb{G}_0)] \\ &\geq -E^*[|E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\ &\geq -E^*[\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \ . \end{aligned} \tag{1.122}$$

Similarly, applying Lemma 1.2.6 in van der Vaart and Wellner (1996) once again together with Lemma 1.6.5(ii), and exploiting that $f \in \mathrm{BL}_1(\mathbb{D})$ we can conclude that:

$$\begin{aligned} E_*[f(\mathbb{G}^*_n)] - E[f(\mathbb{G}_0)] &\leq E_*[E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n]] - E[f(\mathbb{G}_0)] \\ &\leq E^*[|E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\ &\leq E^*[\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \ . \end{aligned} \tag{1.123}$$

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:

$$\begin{aligned} &E^*[\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\ &\qquad \leq 2P^*(\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| > \eta) + \eta \ . \end{aligned} \tag{1.124}$$

Moreover, by Assumption 1.3.2(i), $E^*[f(\mathbb{G}^*_n)] = E_*[f(\mathbb{G}^*_n)] + o(1)$. Thus, Assumption

1.3.1(ii), η being arbitrary, and results (1.122) and (1.123) together imply that:

$$\lim_{n\to\infty} E^*[f(\mathbb{G}_n^*)] = E[f(\mathbb{G}_0)] \tag{1.125}$$

for any $f \in \text{BL}_1(\mathbb{D})$. Further note that since $\mathbb{G}_0$ is tight by Assumption 1.2.2(ii) and $\mathbb{D}$ is a Banach space by Assumption 1.2.1(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 (1.125), Theorem 1.12.2 and Addendum 1.12.3 in van der Vaart and Wellner (1996). ■

Lemma 1.6.2. *Let Assumptions 1.2.1(i), 1.2.2, 1.3.1 and 1.3.2(i) 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 $\mathbb{D}\times\mathbb{D}$:*

$$(r_n\{\hat{\theta}_n - \theta_0\}, r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}) \xrightarrow{L} (\mathbb{G}_1, \mathbb{G}_2)\ . \tag{1.126}$$

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 1.2.2(i)-(ii), Lemma 1.6.1, 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:

$$(\mathbb{G}_{n_{k_j}}, \mathbb{G}_{n_{k_j}}^*) \xrightarrow{L} (\mathbb{Z}_1, \mathbb{Z}_2) \tag{1.127}$$

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 $f_1, f_2 \in \text{BL}_1(\mathbb{D})$ satisfy $f_1(h) \geq 0$ and $f_2(h) \geq 0$ for all $h \in \mathbb{D}$. Then note that by result (1.127) it follows that:

$$\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}_{n_{k_j}}^*)] = E[f_1(\mathbb{Z}_1) f_2(\mathbb{Z}_2)]\ . \tag{1.128}$$

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 1.6.5(iii) imply that:

$$\begin{aligned}
&\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})f_2(\mathbb{G}^*_{n_{k_j}})] - E^*[f_1(\mathbb{G}_{n_{k_j}})E[f_2(\mathbb{G}_0)]] \\
&\qquad \geq \lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n]] - E^*[f_1(\mathbb{G}_{n_{k_j}})E[f_2(\mathbb{G}_0)]] \\
&\qquad \geq -\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})|E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f_2(\mathbb{G}_0)]|] \\
&\qquad \geq -\lim_{j\to\infty} E^*[\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \; , \qquad (1.129)
\end{aligned}$$

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 1.6.5(iv), and $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ also imply that:

$$\begin{aligned}
&\lim_{j\to\infty} E_*[f_1(\mathbb{G}_{n_{k_j}})f_2(\mathbb{G}^*_{n_{k_j}})] - E_*[f_1(\mathbb{G}_{n_{k_j}})E[f_2(\mathbb{G}_0)]] \\
&\qquad \leq \lim_{j\to\infty} E_*[f_1(\mathbb{G}_{n_{k_j}})E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n]] - E_*[f_1(\mathbb{G}_{n_{k_j}})E[f_2(\mathbb{G}_0)]] \\
&\qquad \leq \lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})|E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f_2(\mathbb{G}_0)]|] \\
&\qquad \leq \lim_{j\to\infty} E^*[\sup_{f\in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \; . \qquad (1.130)
\end{aligned}$$

Thus, combining result (1.124) together with (1.129) and (1.130), and the fact that $(\mathbb{G}_n, \mathbb{G}^*_n)$ and $\mathbb{G}_n$ are asymptotically measurable, we can conclude that:

$$\begin{aligned}
\lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})f_2(\mathbb{G}^*_{n_{k_j}})] &= \lim_{j\to\infty} E^*[f_1(\mathbb{G}_{n_{k_j}})E[f_2(\mathbb{G}_0)]] \\
&= E[f_1(\mathbb{G}_0)]E[f_2(\mathbb{G}_0)] \; , \qquad (1.131)
\end{aligned}$$

where the final result follows from $\mathbb{G}_n \xrightarrow{L} \mathbb{G}_0$ in $\mathbb{D}$. Hence, (1.128) and (1.131) imply

$$E[f_1(\mathbb{Z}_1)f_2(\mathbb{Z}_2)] = E[f_1(\mathbb{G}_0)]E[f_2(\mathbb{G}_0)] \qquad (1.132)$$

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 1.2.1(i), and hence result (1.132) 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 (1.127), the claim of the Lemma then follows. ■

Lemma 1.6.3. *Let Assumptions 1.2.1, 1.2.2(ii), and 1.2.3 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:*

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] \ . \tag{1.133}$$

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:

$$\lim_{n\to\infty} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_n) - \phi'_{\theta_0}(a_n))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] \ . \tag{1.134}$$

Next, let $B_\epsilon(a_0) \equiv \{a \in \mathbb{D} : \|a_0 - a\|_{\mathbb{D}} < \epsilon\}$, and observe that result (1.134) implies:

$$\begin{aligned} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] &= \liminf_{\epsilon\downarrow 0} \inf_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a) - \phi'_{\theta_0}(a))] \\ &\leq \limsup_{\epsilon\downarrow 0} \sup_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a) - \phi'_{\theta_0}(a))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] \ . \end{aligned} \tag{1.135}$$

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:

$$\begin{aligned} &\inf_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + a) - \phi'_{\theta_0}(a))] P(\mathbb{G}_2 \in B_\epsilon(a_0)) \\ &\qquad \leq \int_{B_\epsilon(a_0)} \int_{\mathbb{D}_L} f(\phi'_{\theta_0}(z_1 + z_2) - \phi'_{\theta_0}(z_2)) dL(z_1) dL(z_2) \\ &\qquad \leq \sup_{a\in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + a) - \phi'_{\theta_0}(a))] P(\mathbb{G}_2 \in B_\epsilon(a_0)) \ . \end{aligned} \tag{1.136}$$

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:

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] = \lim_{\epsilon \downarrow 0} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))|\mathbb{G}_2 \in B_\epsilon(a_0)]$$
$$= \lim_{\epsilon \downarrow 0} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))|\mathbb{G}_2 \in B_\epsilon(0)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))] \ , \quad (1.137)$$

where the first equality follows from (1.135) and (1.136), 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 (1.135), (1.136), and $\phi'_{\theta_0}(0) = 0$ due to ϕ'_{θ_0} being homogenous of degree one. ■

Lemma 1.6.4. *Let Assumption 1.2.1(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:*

$$\sup_{(a,b) \in K^\delta \times K^\delta : \|a-b\|_{\mathbb{D}} < \eta} \|\psi(a) - \psi(b)\|_{\mathbb{E}} < \epsilon \ . \quad (1.138)$$

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

$$\eta \equiv \min_{1 \le j \le J} \frac{\zeta_{a_j}}{4} \qquad\qquad \delta \equiv \min_{1 \le j \le J} \frac{\zeta_{a_j}}{4} \ . \quad (1.139)$$

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

$$\|a - a_{\bar{j}}\|_{\mathbb{D}} \le \|a - \Pi a\|_{\mathbb{D}} + \|\Pi a - a_{\bar{j}}\|_{\mathbb{D}} < \delta + \frac{\zeta_{a_{\bar{j}}}}{4} \le \frac{\zeta_{a_{\bar{j}}}}{2} \ , \quad (1.140)$$

due to the choice of δ in (1.139). Moreover, if $b \in \mathbb{D}$ satisfies $\|a - b\|_{\mathbb{D}} < \eta$, then:

$$\|b - a_{\bar{j}}\|_{\mathbb{D}} \le \|a - b\|_{\mathbb{D}} + \|a - a_{\bar{j}}\|_{\mathbb{D}} < \eta + \frac{\zeta_{a_{\bar{j}}}}{2} \le \zeta_{a_{\bar{j}}} \ , \quad (1.141)$$

by the choice of η in (1.139). We conclude from (1.140), (1.141) that $a, b \in B_{\zeta_{a_{\bar{j}}}}(a_{\bar{j}})$, and

$$\|\psi(a) - \psi(b)\|_{\mathbb{E}} \leq \|\psi(a) - \psi(a_{\bar{j}})\|_{\mathbb{E}} + \|\psi(b) - \psi(a_{\bar{j}})\|_{\mathbb{E}} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \tag{1.142}$$

by our choice of $\zeta_{a_{\bar{j}}}$. Thus, the Lemma follows from result (1.142). ■

Lemma 1.6.5. *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 $U(\omega) \geq 0$ and $V(\omega) \geq 0$ for all $\omega \in \Omega$. If E^* and E_* denote outer and inner expectations respectively, then it follows that:*

(i) $E^*[U] - c \geq -E^*[|U - c|]$.

(ii) $E_*[U] - c \leq E^*[|U - c|]$.

(iii) $E^*[UV] - E^*[Uc] \geq -E^*[U|V - c|]$ *whenever* $\min\{E^*[UV], E^*[Uc]\} < \infty$.

(iv) $E_*[UV] - E_*[Uc] \leq E^*[U|V - c|]$ *whenever* $\min\{E_*[UV], E_*[Uc]\} < \infty$.

(v) $|E^*[UV] - E^*[Uc]| \leq E^*[U|V - c|]$ *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:

$$E_*[T] = -E^*[-T] \ , \tag{1.143}$$

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, (1.143) and $E_* \leq E^*$ yield:

$$\begin{aligned} E^*[U] - c = E[U^* - c] = E[(U - c)^*] = E^*[U - c] \\ \geq E^*[-|U - c|] = -E_*[|U - c|] \geq -E^*[|U - c|] \ . \end{aligned} \tag{1.144}$$

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:

$$E_*[U] - c \leq E^*[U] - c = E[U^* - c] = E[(U-c)^*] \leq E^*[|U-c|] \ . \tag{1.145}$$

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 (1.143) and $E_* \leq E^*$ that:

$$\begin{aligned} E^*[UV] - E^*[Uc] = E[(UV)^* - (Uc)^*] \geq E[-|(UV)^* - (Uc)^*|] \\ \geq E[-|UV - Uc|^*] = -E_*[U|V-c|] \geq -E^*[U|V-c|] \ . \end{aligned} \tag{1.146}$$

Similarly, for the fourth claim of the Lemma, employ (1.143), 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:

$$\begin{aligned} E_*[UV] - E_*[Uc] = E[(-Uc)^* - (-UV)^*] \leq E[|(-Uc)^* - (-UV)^*|] \\ \leq E[|(-Uc) - (-UV)|^*] = E^*[U|V-c|] \ . \end{aligned} \tag{1.147}$$

Finally, for the fifth claim of the Lemma, note the same arguments as in (1.147) yield

$$\begin{aligned} E^*[UV] - E^*[Uc] = E[(Uc)^* - (UV)^*] \leq E[|(Uc)^* - (UV)^*|] \\ \leq E[|(Uc) - (UV)|^*] = E^*[U|V-c|] \ . \end{aligned} \tag{1.148}$$

Thus, part (v) of the Lemma follows from part (iii) and (1.148). ■

Lemma 1.6.6. *Let Assumptions 1.2.1, 1.2.3(i) hold, and suppose that for some $\kappa > 0$ and $C_0 < \infty$ we have $\|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)\|_{\mathbb{E}} \leq C_0 \|h_1 - h_2\|^{\kappa}_{\mathbb{D}}$ for all $h_1, h_2 \in \mathbb{D}$ outer almost surely. Then, Assumption 1.3.3 holds provided that for all $h \in \mathbb{D}_0$ we have:*

$$\|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1) \ . \tag{1.149}$$

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 1.6.4 implies there exists a $\delta_1 > 0$ such that:

$$\sup_{h \in K_0^{\delta_1}} \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}} < \epsilon \ . \tag{1.150}$$

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

$$\sup_{h \in K_0^{\delta_2}} \|\hat{\phi}'_n(h) - \hat{\phi}'_n(\Pi h)\|_{\mathbb{E}} \leq \sup_{h \in K_0^{\delta_2}} C_0 \|h - \Pi h\|_{\mathbb{E}}^{\kappa} \leq C_0 \delta_2^{\kappa} < \epsilon \ . \tag{1.151}$$

Defining $\delta_3 \equiv \min\{\delta_1, \delta_2\}$, exploiting (1.150), (1.151), and $\Pi h \in K_0$ we then conclude:

$$\begin{aligned} & \sup_{h \in K_0^{\delta_3}} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} \\ & \qquad \leq \sup_{h \in K_0^{\delta_3}} \{\|\hat{\phi}'_n(h) - \hat{\phi}'_n(\Pi h)\|_{\mathbb{E}} + \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}} + \|\hat{\phi}'_n(\Pi h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}}\} \\ & \qquad \leq \sup_{h \in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} + 2\epsilon \end{aligned} \tag{1.152}$$

outer almost surely. Thus, since $K_0^{\delta} \subseteq K_0^{\delta_3}$ for all $\delta \leq \delta_3$ we obtain from (1.152) that:

$$\begin{aligned} & \lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(\sup_{h \in K_0^{\delta}} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 5\epsilon) \\ & \qquad \leq \limsup_{n \to \infty} P(\sup_{h \in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 3\epsilon) \ . \end{aligned} \tag{1.153}$$

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:

$$\sup_{h \in K_0} \min_{1 \leq j \leq J} \max\{C_0 \|h - h_j\|_{\mathbb{D}}^{\kappa}, \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}}\} < \epsilon \ . \tag{1.154}$$

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 (1.154) that:

$$\sup_{h \in K_0} \|\hat{\phi}'_{\theta_0}(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} \leq \max_{1 \leq j \leq J} \|\hat{\phi}'_{\theta_0}(h_j) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}} + 2\epsilon \ . \tag{1.155}$$

Thus, we can conclude from (1.155) and $\hat{\phi}'_{\theta_0}$ satisfying (1.149) for any $h \in \mathbb{D}_0$ that:

$$\begin{aligned} &\limsup_{n\to\infty} P(\sup_{h \in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 3\epsilon) \\ &\qquad \leq \limsup_{n\to\infty} P(\max_{1 \leq j \leq J} \|\hat{\phi}'_n(h_j) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}} > \epsilon) = 0 \ . \end{aligned} \tag{1.156}$$

Since ϵ and K_0 were arbitrary, the Lemma follows from (1.153) and (1.154). ■

Lemma 1.6.7. *Let Assumptions 1.2.1, 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 1.2.2(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

$$\mathbb{P} \equiv \overline{\text{span}\{\mathbb{S}(\tau)\}}^{\tau} \tag{1.157}$$

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

$$\mathbb{N}_P(\tau_w) = \mathbb{S}(\tau_w) \ . \tag{1.158}$$

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 (1.158):

$$\mathbb{S}(\tau) \subseteq \mathbb{N}_P(\tau) = \mathbb{N}_P(\tau_w) = \mathbb{S}(\tau_w) \subseteq \mathbb{S}(\tau) \ . \tag{1.159}$$

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 (1.157) and (1.158) we thus conclude $\mathbb{S}(\tau) = \mathbb{P}$ and the claim of the Lemma follows. ■

1.6.2 Results for Examples 1.2.1-1.2.6

Lemma 1.6.8. *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:

$$|\sup_{a\in\mathbf{A}}\{\theta(a) + t_n h_n(a)\} - \sup_{a\in\mathbf{A}}\{\theta(a) + t_n h(a)\}| \leq t_n\|h_n - h\|_\infty = o(t_n) \ . \tag{1.160}$$

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

$$\Psi_{\bar{\mathbf{A}}}(\theta+g)=\arg\max_{a\in\bar{\mathbf{A}}}\Gamma_\theta(g)(a) \tag{1.161}$$

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 $\|t_nh\|_\infty=o(1)$ that $\Psi_{\bar{\mathbf{A}}}(\theta+t_nh)\subseteq\Psi_{\bar{\mathbf{A}}}(\theta)^{\delta_n}$ 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

$$\begin{aligned}|\sup_{a\in\bar{\mathbf{A}}}\{\theta(a)+t_nh(a)\}-&\sup_{a\in\Psi_{\bar{\mathbf{A}}}(\theta)}\{\theta(a)+t_nh(a)\}|\\&=\sup_{a\in\Psi_{\bar{\mathbf{A}}}(\theta)^{\delta_n}}\{\theta(a)+t_nh(a)\}-\sup_{a\in\Psi_{\bar{\mathbf{A}}}(\theta)}\{\theta(a)+t_nh(a)\}\\&\leq\sup_{a_0,a_1\in\bar{\mathbf{A}}:\|a_0-a_1\|_{\mathbf{A}}\leq\delta_n}t_n|h(a_0)-h(a_1)|\\&=o(t_n)\ ,\end{aligned} \tag{1.162}$$

where the final result follows from h being uniformly continuous by compactness of $\bar{\mathbf{A}}$. Therefore, exploiting (1.160), (1.162) and θ being constant on $\Psi_{\bar{\mathbf{A}}}(\theta)$ yields

$$\begin{aligned}&|\sup_{a\in\mathbf{A}}\{\theta(a)+t_nh_n(a)\}-\sup_{a\in\mathbf{A}}\theta(a)-t_n\sup_{a\in\Psi_{\bar{\mathbf{A}}}(\theta)}h(a)|\\&\quad\leq|\sup_{a\in\Psi_{\bar{\mathbf{A}}}(\theta)}\{\theta(a)+t_nh(a)\}-\sup_{a\in\Psi_{\bar{\mathbf{A}}}(\theta)}\theta(a)-t_n\sup_{a\in\Psi_{\bar{\mathbf{A}}}(\theta)}h(a)|+o(t_n)=o(t_n)\ ,\end{aligned} \tag{1.163}$$

which verifies the claim of the Lemma. ∎

Lemma 1.6.9. *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 $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, and

$$B_-(\theta) \equiv \{u \in \mathbf{R} : \theta^{(1)}(u) < \theta^{(2)}(u)\} \ . \tag{1.164}$$

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:

$$\begin{aligned}\int_{B_-(\theta)} & \max\{(\theta^{(1)}(u) - \theta^{(2)}(u)) + t_n(h_n^{(1)}(u) - h_n^{(2)}(u)), 0\} w(u) du \\ & \lesssim t_n \int_{B_-(\theta)} 1\{t_n(h_n^{(1)}(u) - h_n^{(2)}(u)) \geq -(\theta^{(1)}(u) - \theta^{(2)}(u))\} w(u) du = o(t_n) \ . \end{aligned} \tag{1.165}$$

Thus, (1.165), $B_-(\theta)^c = B_+(\theta) \cup B_0(\theta)$ and dominated convergence theorem imply

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

which establishes the claim of the Lemma. ∎

Lemma 1.6.10. *Let Assumptions 1.2.1, 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:*

$$\phi'_{\theta_0}(h) = \sup_{a \in A_0} h(a) \ . \tag{1.166}$$

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* (1.39).

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 1.6.6 it suffices to verify that for any $h \in \mathcal{C}(\mathbf{A})$, $\hat{\phi}'_n$ satisfies

$$|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| = o_p(1) \ . \tag{1.167}$$

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

$$\sup_{\|a_1 - a_2\|_{\mathbf{A}} < \eta} |h(a_1) - h(a_2)| < \epsilon_0 \ . \tag{1.168}$$

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

$$|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| \leq \sup_{\|a_1 - a_2\|_{\mathbf{A}} \leq d_H(\hat{A}_0, A_0, \|\cdot\|_{\mathbf{A}})} |h(a_1) - h(a_2)| \ . \tag{1.169}$$

Thus, by results (1.168) and (1.169), and the Hausdorff consistency of $\hat{A}_0$, we obtain:

$$\limsup_{n\to\infty} P(|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| > \epsilon_0) \leq \limsup_{n\to\infty} P(d_H(\hat{A}_0, A_0, \|\cdot\|_{\mathbf{A}}) > \eta) = 0 \ . \tag{1.170}$$

It follows that (1.167) indeed holds, and the claim of the Lemma follows. ■

Chapter 2

Optimal Plug-in Estimators of Directionally Differentiable Functionals

Abstract

This chapter studies optimal estimation of parameters taking the form $\phi(\theta_0)$, where θ_0 is unknown but can be regularly estimated and ϕ is a known directionally differentiable function. The irregularity caused by nondifferentiability of ϕ makes traditional optimality criteria such as semiparametric efficiency and minimum variance unbiased estimation impossible to apply. We instead consider optimality in the sense of local asymptotic minimaxity – i.e. we seek estimators that locally asymptotically minimize the maximum of the risk function. We derive the lower bound of local asymptotic minimax risk within a class of plug-in estimators and develop a general procedure for constructing estimators that attain the bound. As an illustration, we apply the developed theory to the estimation of the effect of Vietnam veteran status on the quantiles of civilian earnings.

2.1 Introduction

In many econometric problems, parameters of interest embody certain irregularity that presents significant challenges for estimation and inference (Hirano and Porter, 2012; Fang and Santos, 2014). A large class of these parameters take the form $\phi(\theta_0)$ where θ_0 is a well-behaved parameter that depends on the underlying distribution of the data while ϕ is a known but potentially nondifferentiable function. Economic settings in which such irregularity arises with ease include treatment effects (Manski and Pepper, 2000; Hirano and Porter, 2012; Song, 2014, 2015), interval valued data (Manski and Tamer, 2002), incomplete auction models (Haile and Tamer, 2003), and estimation under shape restrictions (Chernozhukov et al., 2010).

The aforementioned examples share the common feature of ϕ being directionally differentiable despite a possible failure of full differentiability. In this paper, we study optimal estimation of $\phi(\theta_0)$ for such irregular ϕ. In regular settings, one usually thinks of optimality in terms of semiparametric efficiency (Bickel et al., 1993). Unfortunately, the irregularity caused by nondifferentiability of ϕ makes traditional optimality criteria including semiparametric efficiency impossible to apply – in particular, if ϕ is nondifferentiable, then any estimator for $\phi(\theta_0)$ is necessarily irregular and biased(Hirano and Porter, 2012). Hence, the first question we need to address is: what is an appropriate notion of optimality for nondifferentiable ϕ? Following the decision theoretic framework initiated by Wald (1950) and further developed by Le Cam (1955, 1964), we may compare the competing estimators under consideration by examining their expected losses. Specifically, let T_n be an estimator of $\phi(\theta_0)$ and ℓ a loss function that measures the loss of estimating $\phi(\theta_0)$ using T_n by $\ell(r_n\{T_n - \phi(\theta_0)\})$, where $r_n \uparrow \infty$ is the rate of convergence for estimation of θ_0. The resulting expected loss or *risk function* is then

$$E_P[\ell(r_n\{T_n - \phi(\theta(P))\})] \ , \tag{2.1}$$

where E_P denotes the expectation taken with respect to P that generates the data and

$\theta_0 \equiv \theta(P)$ signifies the dependence of θ_0 on P. The function (2.1) can in turn be employed to assess the performance of the estimator T_n – in particular, we would like an estimator to have the smallest possible risk at every P in the model. Unfortunately, it is well known that there exist no estimators that minimize the risk uniformly for all P (Lehmann and Casella, 1998).

As ways out of this predicament, one can either restrict the class of competing estimators, or seek an estimator that has the smallest risk in some overall sense. For the former approach, common restrictions imposed on estimators include mean unbiasedness, quantile unbiasedness and equivariance (including regularity which is also known as asymptotic equivariance in law). By Hirano and Porter (2012), however, if ϕ is only directionally differentiable, then no mean unbiased, quantile unbiased or regular estimators exist. It is noteworthy that non-existence of unbiased estimators implies that bias correction procedures cannot fully eliminate the bias of any estimator; in fact, any procedure that tries to remove the bias would push the variance to infinity (Doss and Sethuraman, 1989). As to equivariance in terms of groups of transformations, it is unclear to us what a suitable group of invariant transformations should be. Alternatively, one may translate the risk function (2.1) into a single number such as Bayesian risk that leads to average risk optimality or the maximum risk that leads to minimaxity. Since our analysis shall focus on local risk, one may not have natural priors on the space of localization parameters in order to evaluate the Bayesian risk. Moreover, when the model is semiparametric or nonparametric which our setup accommodates, Bayes estimators entail specification of priors on infinite dimensional spaces which practitioners may lack.

The approach we adopt in this paper towards optimal estimation of $\phi(\theta_0)$ is a combination of the above two: we confine our attention to the important class of plug-in estimators of the form $\phi(\hat{\theta}_n)$, where $\hat{\theta}_n$ is a regular estimator of θ_0, and seek estimators that minimize the maximum of the risk function – i.e. the risk under the worst case scenario. In addition, the optimality shall be in local sense, that is, we consider maximum risk over neighborhoods around the distribution that generates the data. This is justified by the facts that global risk is somewhat too restrictive for infinite dimensional P (Bickel et al., 1993, p.21) and that one can locate the unknown parameter with considerable precision as sample size increases (Hájek, 1972). Specifically, for $\hat{\theta}_n$ an arbitrary regular estimator of θ_0, we establish the lower bound of the following local asymptotic minimax risk:

$$\sup_{I\subset_f H} \liminf_{n\to\infty} \sup_{h\in I} E_{P_{n,h}}[\ell(r_n\{\phi(\hat{\theta}_n) - \phi(\theta(P_{n,h}))\})] \ , \tag{2.2}$$

where H is the set of localization parameters, and $I \subset_f H$ signifies that I is a finite subset of H so that the first supremum is taken over all finite subsets of H.[1] For detailed explanations on why we take the above version of local asymptotic minimaxity, which dates back to van der Vaart (1988a, 1989), we defer our discussion to Section 2.2.3. The lower bound derived relative to the local asymptotic minimax risk (2.2) is consistent with the regular case (van der Vaart and Wellner, 1996); moreover, it is also consistent with previous work by Song (2014) who studies a more restrictive class of irregular parameters.

We also present a general procedure of constructing optimal plug-in estimators. An optimal plug-in estimator is of the form $\phi(\hat{\theta}_n + \hat{u}_n/r_n)$, where $\hat{\theta}_n$ is an efficient estimator of θ_0 usually available from efficient estimation literature, and $\hat{u}_n$ is a correction term that depends on the particular loss function ℓ. It is interesting to note that optimality is preserved under simple plug-in for differentiable maps (van der Vaart, 1991b), but in general not for nondifferentiable ones due to the presence of the correction term $\hat{u}_n$ – i.e. $\hat{u}_n$ equals zero when ϕ is differentiable but may be nonzero otherwise. Heuristically, the need of the correction term $\hat{u}_n$ arises from the fact that the simple plug-in estimator $\phi(\hat{\theta}_n)$ may have undesirably high risk at θ_0 where ϕ is nondifferentiable. By adding a correction term, one is able to reduce the risk under the worst case scenario. As an illustration, we apply the construction procedure to the estimation of the effect of Vietnam veteran status on the quantiles of civilian earnings. In the application, the structural quantile functions of earnings exhibit local nonmonotonicity, especially for veterans. Nonetheless, by estimating the closest monotonically increasing functions to the population quantile processes, we are

[1] For example, if P is parametrized as $\theta \mapsto P_\theta$ where θ belongs to an open set $\Theta \subset \mathbf{R}^k$, one typically considers local parametrization $h \mapsto P_{\theta_0+h/\sqrt{n}}$ with local parameter h ranging over the whole space $\mathbf{R}^k$. We shall have a formal definition of H in Section 2.2.2.

able to resolve this issue and provide locally asymptotically minimax plug-in estimators.

There has been extensive study on optimal estimation of regular parameters (Ibragimov and Has'minskii, 1981; Bickel et al., 1993; Lehmann and Casella, 1998). The best known optimality results are characterized by the convolution theorems and the local asymptotic minimax theorems (Hájek, 1970, 1972; Le Cam, 1972; Koshevnik and Levit, 1976; Levit, 1978; Begun et al., 1983; Millar, 1983, 1985; Chamberlain, 1987; van der Vaart, 1988b; van der Vaart and Wellner, 1990; van der Vaart, 1991a). However, little work has been done on nondifferentiable parameters. Blumenthal and Cohen (1968a,b) considered minimax estimation of the maximum of two translation parameters and pointed out the link between biased estimation and nondifferentiability of the parameter. Hirano and Porter (2012) formally established the connection between differentiability of parameters and possibility of regular, mean unbiased and quantile unbiased estimation, and emphasized the need for alternative optimality criteria when the parameters of interest are nondifferentiable. Chernozhukov et al. (2013) considered estimation of intersection bounds in terms of median-bias-corrected criterion. The work by Song (2014, 2015) is mostly closely related to ours. By restricting the parameter of interest to be a composition of a real valued Lipschitz function having a finite set of nondifferentiability points and a translation-scale equivariant real-valued map, Song (2014, 2015) was able to establish local asymptotic minimax estimation within the class of arbitrary estimators. In present paper, we consider a much wider class of parameters at the expense of restricting the competing estimators to be of a plug-in form. We note also that for differentiable ϕ, the optimality of the plug-in principle has been established by van der Vaart (1991b).

The remainder of the paper is structured as follows. Section 2.2 formally introduces the setup, presents a convolution theorem for efficient estimation of θ that will be essential for later discussion, and specifies the suitable version of local asymptotic minimaxity criterion for our purposes. In Section 2.3 we derive the minimax lower bound for the class of plug-in estimators, and then present a general construction procedure. Section 2.4 applies the theory to the estimation of the effect of Vietnam veteran status on the quantiles of civilian earnings. Section 2.5 concludes. All proofs are collected in Appendices.

2.2 Setup, Convolution and Minimaxity

In this section, we formally set up the problem under consideration, present a convolution theorem for the estimation of θ, and establish the optimality criterion that will be employed to assess the statistical performance of plug-in estimators $\phi(\hat{\theta}_n)$.

2.2.1 Setup and Notation

In order to accommodate applications such as incomplete auction models and estimation under 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 2.2.3 and 2.2.4 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 an arbitrary map of the sample $\{X_i\}_{i=1}^n$ into the domain of ϕ. Thus, the distributional convergence in our context is understood to be in the Hoffman-Jørgensen sense and expectations throughout should be interpreted as outer expectations (van der Vaart and Wellner, 1996), though we obviate the distinction in the notation.

We introduce notation that is recurrent in this paper. For two sets A and B, we write $A \subset_f B$ to signify that A is a finite subset of B. For a finite set $\{g_1, \ldots, g_m\}$, we write $g^m \equiv (g_1, \ldots, g_m)^\intercal$. Lastly, we define $K_\lambda^m \equiv \{x \in \mathbf{R}^m : \|x\| \leq \lambda\}$ for $\lambda > 0$.

2.2.1.1 Examples

To illustrate the applications of our framework, we begin by presenting some examples that arise in the econometrics and statistics literature. We shall revisit these examples later on as we develop our theory. To highlight the essential ideas and for ease of exposition, we base our discussion on simplifications of well known examples. The general case can be handled analogously.

In the treatment effect literature one might be interested in estimating the maximal treatment effect. Our first example has been considered in Hirano and Porter (2012) and

Song (2014, 2015).

Example 2.2.1 (Best Treatment)**.** Let $X = (X^{(1)}, X^{(2)})^\intercal \in \mathbf{R}^2$ be a pair of potential outcomes under two treatments. Consider the problem of estimating the parameter

$$\phi(\theta_0) = \max\{E[X^{(1)}], E[X^{(2)}]\} \ . \tag{2.3}$$

One can think of $\phi(\theta_0)$ as the expected outcome under the best treatment. Here, $\theta_0 = (E[X^{(1)}], E[X^{(2)}])^\intercal$, $\mathbb{D} = \mathbf{R}^2$, $\mathbb{E} = \mathbf{R}$, and $\phi : \mathbf{R}^2 \to \mathbf{R}$ is given by $\phi(\theta) = \max(\theta^{(1)}, \theta^{(2)})$. Parameters of this type are essential in characterizing optimal decision rules in dynamic treatment regimes which, as opposed to classical treatment, incorporate heterogeneity across both individuals and time (Murphy, 2003). The functional form of (2.3) is also related to the study of bounds of treatment effects under monotone instruments (Manski and Pepper, 2000, 2009). Minimax estimation of $\phi(\theta)$ when $X^{(1)}$ and $X^{(2)}$ are independent normal random variables with equal variances has been studied in Blumenthal and Cohen (1968a,b). ■

Partial identification is an inherent feature of statistical analysis based on interval censored data. In these settings, one might still want to estimate identified features of the model under consideration. Our second example is based on Manski and Tamer (2002) who study inference on regressions with interval data on a regressor or outcome.

Example 2.2.2 (Interval Regression Model)**.** Let $Y \in \mathbf{R}$ be a random variable generated by

$$Y = \alpha + \beta W + \epsilon \ ,$$

where $W \in \{-1, 0, 1\}$ is a discrete random variable, and $E[\epsilon|W] = 0$. Suppose that Y is unobservable but there exist (Y_l, Y_u) such that $Y_l \leq Y \leq Y_u$ almost surely. Let $\vartheta = (\alpha, \beta)^\intercal$ and $Z = (1, W)^\intercal$. Then the identified region for ϑ is

$$\Theta_0 \equiv \{\vartheta \in \mathbf{R}^2 : E[Y_l|Z] \leq Z^\intercal\vartheta \leq E[Y_u|Z]\} \ .$$

Interest often centers on either the maximal value of a particular coordinate of ϑ or the maximal value of the conditional expectation $E[Y|W]$ at a specified value of the covariates, both of which can be expressed as

$$\sup_{\vartheta \in \Theta_0} \lambda^\intercal \vartheta \ , \tag{2.4}$$

for some known $\lambda \equiv (\lambda^{(1)}, \lambda^{(2)})^\intercal \in \mathbf{R}^2$. Let $\theta_0 \equiv (P(W=-1), P(W=1))^\intercal$. It is shown in Appendix 2.6.2 that the analysis of (2.4) reduces to examining terms of the form[2]

$$\phi(\theta_0) = \max\{\psi(\theta_0), 0\} \ , \tag{2.5}$$

where for each $\theta = (\theta^{(1)}, \theta^{(2)})^\intercal \in \mathbf{R}^2$, $\psi(\theta)$ is defined by

$$\psi(\theta) = \lambda^{(1)} \frac{\theta^{(1)} + \theta^{(2)}}{\theta^{(1)} + \theta^{(2)} - (\theta^{(2)} - \theta^{(1)})^2} + \lambda^{(2)} \frac{\theta^{(1)} - \theta^{(2)}}{\theta^{(1)} + \theta^{(2)} - (\theta^{(2)} - \theta^{(1)})^2} \ .$$

In this example, $\mathbb{D} = \mathbf{R}^2$, $\mathbb{E} = \mathbf{R}$ and $\phi : \mathbf{R}^2 \to \mathbf{R}$ satisfies $\phi(\theta) = \max\{\psi(\theta), 0\}$ with $\psi(\theta)$ defined as above. The functional form of ϕ here is common in a class of partially identified models (Beresteanu and Molinari, 2008; Bontemps et al., 2012; Chandrasekhar et al., 2012; Kaido and Santos, 2014; Kaido, 2013a; Kline and Santos, 2013). ■

The next example presents a nondifferentiable function which appears as an identification bound on the distribution of valuations in an incomplete model of English auctions (Haile and Tamer, 2003; Hirano and Porter, 2012).

Example 2.2.3 (Incomplete Auction Model)**.** In an English auction model with symmetric independent private values, a robust approach of interpreting bidding data proposed by Haile and Tamer (2003) is to assume only that bidders neither bid more than their valuations nor let an opponent win at a price they would be willing to beat. Consider two auctions in which bidders' valuations are i.i.d. draws from F. Let B_i and V_i be bidder i's bid and valuation respectively, and let F_1 and F_2 be the distributions of bids in two auctions. The

[2]Here we work with $\phi(\theta_0)$ for simplicity and ease of exposition.

first assumption implies $B_i \leq V_i$ for all i, which in turn imposes a upper bound of F:[3]

$$F(v) \leq \min\{F_1(v), F_2(v)\} \ .$$

Similarly, by exploiting the assumption that bidders do not let an opponent win at a price below their willingness to pay, one may obtain a lower bound on F. For simplicity, we consider only the upper bound which we write as

$$\phi(\theta_0)(v) = \min\{F_1(v), F_2(v)\} \ . \tag{2.6}$$

In this example, $\theta_0 = (F_1, F_2)$, $\mathbb{D} = \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, $\mathbb{E} = \ell^\infty(\mathbf{R})$ and $\phi : \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R}) \to \ell^\infty(\mathbf{R})$ satisfies $\phi(\theta)(v) \equiv \min\{\theta^{(1)}(v), \theta^{(2)}(v)\}$. ■

Our final example involves a map that monotonizes estimators in linear quantile regressions. Being estimated in pointwise manner, the quantile regression processes need not be monotonically increasing (Bassett and Koenker, 1982; He, 1997). This problem can be fixed by considering the closest monotonically increasing function.[4]

Example 2.2.4 (Quantile Functions without Crossing)**.** Let $Y \in \mathbf{R}$ and $Z \in \mathbf{R}^d$ be random variables. Consider the linear quantile regression model:

$$\beta(\tau) \equiv \arg\min_{\beta \in \mathbf{R}^d} E[\rho_\tau(Y - Z'\beta)] \ ,$$

where $\rho_\tau(u) \equiv u(\tau - 1\{u \geq 0\})$. Let $\mathcal{T} \equiv [\epsilon, 1-\epsilon]$ for some $\epsilon \in (0, 1/2)$ and $\theta_0 \equiv c'\beta(\cdot) : \mathcal{T} \to \mathbf{R}$ be the quantile regression process, for fixed $Z = c$. Under misspecification, θ_0 need not be monotonically increasing. In order to avoid the quantile crossing problem, we may instead consider projecting θ_0 onto the set of monotonically increasing functions – i.e. the

[3] Haile and Tamer (2003) actually exploit order statistics of bids in order to obtain tighter bounds on F.
[4] Alternatively, Chernozhukov et al. (2010) propose employing a sorting operator to monotonize possibly nonmonotone estimators.

closest monotonically increasing function to θ_0:

$$\phi(\theta_0) = \Pi_\Lambda \theta_0 \equiv \arg\min_{\lambda\in\Lambda} \|\lambda - \theta_0\|_{L^2} \ , \tag{2.7}$$

where Λ be the set of monotonically increasing functions in $L^2(\mathcal{T}, \nu)$ with ν the Lebesgue measure on $\mathcal{T}$, and Π_Λ is the metric projection onto Λ – i.e. the mapping that assigns every point in $L^2(\mathcal{T})$ with the closest point in Λ.[5] In this example, $\mathbb{D} = L^2(\mathcal{T})$, $\mathbb{E} = \Lambda$ and $\phi : L^2(\mathcal{T}) \to \Lambda$ is defined by $\phi(\theta) = \Pi_\Lambda\theta$. We note that the metric projection approach introduced here can in fact handle a larger class of estimation problems under shape restrictions; see Remark 2.2.1. ■

Remark 2.2.1. Let $\theta_0 : \mathcal{T} \to \mathbf{R}$ be a unknown real valued function where $\mathcal{T} = [a, b]$ with $-\infty < a < b < \infty$. Then one may monotonize θ_0 by considering the nearest monotonically increasing function $\phi(\theta_0) \equiv \Pi_\Lambda\theta_0$ where $\Lambda \subset L^2(\mathcal{T})$ is the set of increasing functions. More generally, one may take Λ to be a closed and convex set of functions satisfying certain shape restrictions such as convexity and homogeneity. Then the projection $\Pi_\Lambda\theta_0$ of θ_0 onto Λ is the closest function to θ_0 with desired shape restrictions. ■

2.2.2 The Convolution Theorem

In this section, before delving into the discussion of the defining ingredient ϕ, we formalize basic regularity assumptions and then present a convolution theorem for the estimation of θ_0, which in turn will be employed when deriving the asymptotic minimax lower bound for the estimation of $\phi(\theta_0)$.

Following the literature on limits of experiments (Blackwell, 1951; Le Cam, 1972; van der Vaart, 1991a), we consider a sequence of experiments $\mathcal{E}_n \equiv (\mathscr{X}_n, \mathcal{A}_n, \{P_{n,h} : h \in H\})$, where $(\mathscr{X}_n, \mathcal{A}_n)$ is a measurable space, and $P_{n,h}$ is a probability measure on $(\mathscr{X}_n, \mathcal{A}_n)$, for each $n \in \mathbf{N}$ and $h \in H$ with H a subspace of some Hilbert space equipped with inner product $\langle\cdot,\cdot\rangle_H$ and induced norm $\|\cdot\|_H$. We observe a sample $X_1, \ldots, X_n$ that is jointly distributed according to some $P_{n,h}$. This general framework allows us to consider non

[5] The set Λ is closed and convex so that the metric projection Π_Λ exists and is unique; see Appendix 2.6.2 for detailed discussion.

i.i.d. models (Ibragimov and Has'minskii, 1981; van der Vaart, 1988b; van der Vaart and Wellner, 1990) as well as common i.i.d. setup. We confine our attention to the family of probability measures $\{P_{n,h} : h \in H\}$ possessing local asymptotic normality; see Assumption 2.2.1(ii).[6] This is perhaps the most convenient class to begin with in the literature of efficient estimation, since mutual contiguity implied by local asymptotic normality allows us, by Le Cam's third lemma, to deduce weak limits along sequence $\{P_{n,h}\}_{n=1}^{\infty}$ from that under the fixed sequence $\{P_{n,0}\}_{n=1}^{\infty}$ – usually thought of as the underlying truth. Formally, we impose

Assumption 2.2.1. (i) The set H is a subspace of some separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_H$ and induced norm $\| \cdot \|_H$.

(ii) The sequence of experiments $(\mathscr{X}_n, \mathcal{A}_n, \{P_{n,h} : h \in H\})$ is asymptotically normal, i.e.

$$\log \frac{dP_{n,h}}{dP_{n,0}} = \Delta_{n,h} - \frac{1}{2}\|h\|_H^2 \ , \tag{2.8}$$

where $\{\Delta_{n,h} : h \in H\}$ is a stochastic process which converges to $\{\Delta_h : h \in H\}$ marginally under $\{P_{n,0}\}$,[7] with $\{\Delta_h : h \in H\}$ a Gaussian process having mean zero and covariance function given by $E[\Delta_{h_1}\Delta_{h_2}] = \langle h_1, h_2 \rangle_H$.[8]

Separability as in Assumption 2.2.1(i) is only a minimal requirement in practice, while linearity is standard although not entirely necessary.[9] The essence of Assumption 2.2.1(ii) is that the sequence of experiments $\mathcal{E}_n$ can be asymptotically represented by a Gaussian shift experiment. Thus, one may "pass to the limit first", "argue the case for the limiting problem" which has simpler statistical structure, and then translate the results back to the original experiments $\mathcal{E}_n$ (Le Cam, 1972).[10] In the i.i.d. case, Assumption 2.2.1(ii) is guaranteed by the so-called differentiability in quadratic mean; see Remark 2.2.2.

[6] Our results in fact extend to models having local asymptotic mixed normality; see Jeganathan (1981, 1982) and van der Vaart (1998, Section 9.6).

[7] That is, for any finite set $I \subset H$, $(\Delta_{n,h} : h \in I) \xrightarrow{L} (\Delta_h : h \in I)$ under $\{P_{n,0}\}$.

[8] Here, $dP_{n,0}$ and $dP_{n,h}$ can be understood as densities of $P_{n,0}$ and $P_{n,h}$ with respect to some σ-finite measure μ_n, respectively. Fortunately, the log ratio above is independent of the choice of μ_n; see van der Vaart (1998, p.189-91).

[9] In fact, H can be relaxed to be a convex cone; see van der Vaart and Wellner (1996) and van der Vaart (1989).

[10] From a technical level, for any finite set $I \subset H$, weak convergence of likelihoods in Assumption 2.2.1(ii) is equivalent to convergence in terms of Le Cam's deficiency distance (Le Cam, 1972, 1986).

Regularity conditions on the parameter θ and an estimator $\hat{\theta}_n$ are imposed as follows. In our setup, we recognize θ as a map $\theta : \{P_{n,h}\} \to \mathbb{D}$ and write $\theta_n(h) \equiv \theta(P_{n,h})$.

Assumption 2.2.2. The parameter $\theta : \{P_{n,h}\} \to \mathbb{D}_\phi \subset \mathbb{D}$, where $\mathbb{D}$ is a Banach space with norm $\|\cdot\|_{\mathbb{D}}$, is regular, i.e. there exists a continuous linear map $\theta_0' : H \to \mathbb{D}$ such that for every $h \in H$,

$$r_n\{\theta_n(h) - \theta_n(0)\} \to \theta_0'(h) \text{ as } n \to \infty , \tag{2.9}$$

for a sequence of $\{r_n\}$ with $r_n \to \infty$ as $n \to \infty$.

Assumption 2.2.3. $\hat{\theta}_n : \{X_i\} \to \mathbb{D}_\phi$ is regular, i.e. there is a fixed tight random variable $\mathbb{G} \in \mathbb{D}$ such that for any $h \in H$,

$$r_n\{\hat{\theta}_n - \theta_n(h)\} \overset{L_{n,h}}{\to} \mathbb{G} \text{ in } \mathbb{D} , \tag{2.10}$$

where $\overset{L_{n,h}}{\to}$ denotes weak convergence under $\{P_{n,h}\}$.

Assumption 2.2.2, which dates back to Pfanzagl and Wefelmeyer (1982), is essentially a Hadamard differentiability requirement; see Remark 2.2.3. Our optimality analysis shall extend from Hadamard differentiable parameters to a class of (Hadamard) directionally differentiable parameters. The derivative $\theta_0' : H \to \mathbb{D}$ is crucial in determining the efficiency bound for estimating θ. If $\mathbb{D} = \mathbf{R}^m$, the derivative $\theta_0' : H \to \mathbf{R}^m$ uniquely determines through the Riesz representation theorem a $m \times 1$ vector $\tilde{\theta}_0$ of elements in the completion $\overline{H}$ of H such that $\theta_0'(h) = \langle \tilde{\theta}_0, h \rangle$ for all $h \in H$. The matrix $\Sigma_0 \equiv \langle \tilde{\theta}_0, \tilde{\theta}_0^\intercal \rangle$ is called the efficiency bound for θ. For general $\mathbb{D}$, the efficiency bound is characterized through the topological dual space $\mathbb{D}^*$ of $\mathbb{D}$ (Bickel et al., 1993); see Theorem 2.2.1.

Assumption 2.2.3 means that $\{\hat{\theta}_n\}$ is asymptotically equivariant in law for estimating $\theta_n(h)$, or put it another way, the limiting distribution of $\{\hat{\theta}_n\}$ is robust to "local perturbations" $\{P_{n,h}\}$ of the "truth" $\{P_{n,0}\}$. In this way it restricts the class of plug-in estimators we consider. For instance, superefficient estimators such as Hodges' estimator and shrinkage estimators are excluded from our setup (Le Cam, 1953; Huber, 1966; Hájek, 1972;

van der Vaart, 1992). Finally, we note that while regularity of θ, as ensured by Assumption 2.2.2, is necessary for Assumption 2.2.3 to hold (Hirano and Porter, 2012), it is in general not sufficient unless the model is parametric (Bickel et al., 1993).

Assumptions 2.2.1, 2.2.2 and 2.2.3 together place strong restrictions on the structure of the asymptotic distribution of $\hat{\theta}_n$. In particular, for every $\hat{\theta}_n$ satisfying the above regularity conditions, its weak limit can be represented as the efficient Gaussian random variable plus an independent noise term, as illustrated in the following convolution theorem taken directly from van der Vaart and Wellner (1990). The derivative $\theta_0' : H \to \mathbb{D}$ as a continuous linear map has an adjoint map $\theta_0'^{*} : \mathbb{D}^* \to \overline{H}$ satisfying $d^*\theta_0'(h) = \langle \theta_0'^{*} d^*, h\rangle_H$ for all $d^* \in \mathbb{D}^*$; that is, $\theta_0'^{*}$ maps the dual space $\mathbb{D}^*$ of $\mathbb{D}$ into $\overline{H}$.

Theorem 2.2.1 (Hájek-Le Cam Convolution Theorem)**.** *Let $(\mathscr{X}_n, \mathcal{A}_n, \{P_{n,h} : h \in H\})$ be a sequence of statistical experiments, and $\hat{\theta}_n$ be an estimator for the parameter $\theta : \{P_{n,h}\} \to \mathbb{D}$. Suppose that Assumptions 2.2.1, 2.2.2 and 2.2.3 hold. It follows that*[11]

$$\mathbb{G} \overset{d}{=} \mathbb{G}_0 + \mathbb{U} \ , \tag{2.11}$$

where $\mathbb{G}_0$ is a tight Gaussian random variable in $\mathbb{D}$ satisfying $d^*\mathbb{G}_0 \sim \mathcal{N}(0, \|\theta_0'^{*} d^*\|_H^2)$ for every $d^* \in \mathbb{D}^*$, and $\mathbb{U}$ is a tight random variable in $\mathbb{D}$ that is independent of $\mathbb{G}_0$. Moreover, the support of $\mathbb{G}_0$ is $\overline{\theta_0'(H)}$ (the closure of $\{\theta_0'(h) : h \in H\}$ relative to $\|\cdot\|_{\mathbb{D}}$).[12]

One important implication of Theorem 2.2.1 is that a regular estimator sequence $\{\hat{\theta}_n\}$ is considered efficient if its limiting law is such that $\mathbb{U}$ is degenerate at 0. In addition, normality being "the best limit" is a result of optimality, rather than an *ex ante* restriction. If ϕ is Hadamard differentiable, then we may conclude immediately that $\phi(\hat{\theta}_n)$ is an efficient estimator for $\phi(\theta_0)$ if $\hat{\theta}_n$ is for θ_0 (van der Vaart, 1991b). When ϕ is Hadamard directionally differentiable only, however, we have to base our optimality analysis within the class of irregular estimators because no regular estimators exist in this context (Hirano and Porter, 2012). As a result, the convolution theorem is not available in general, which motivates the

[11] The symbol $\overset{d}{=}$ denotes equality in distribution.

[12] The support of $\mathbb{G}_0$ refers to the intersection of all closed subsets $\mathbb{D}_0 \subset \mathbb{D}$ with $P(\mathbb{G}_0 \in \mathbb{D}_0) = 1$.

optimality analysis in terms of asymptotic minimax criterion.

Remark 2.2.2. Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample with common distribution P that is known to belong to a collection $\mathcal{P}$ of Borel probability measures, and let $\{P_t : t \in (0, \epsilon)\} \subset \mathcal{P}$ with $P_0 = P$ be a submodel such that

$$\int [\frac{dP_t^{1/2} - dP^{1/2}}{t} - \frac{1}{2} h dP^{1/2}]^2 \to 0 \text{ as } t \downarrow 0 , \tag{2.12}$$

where h is called the score of this submodel. In this situation, we identify $P_{n,h}$ with $\prod_{i=1}^n P_{1/\sqrt{n},h}$ where $\{P_{1/\sqrt{n},h}\}$ is differentiable in quadratic mean with score h, and the set $\dot{\mathcal{P}}^0$ of all score functions thus obtained, which are necessarily elements of $L^2(P)$, will be the index set H, also known as the tangent set of $\mathcal{P}$. It can be shown that the sequence $\{P_{n,h}\}$ satisfies Assumption 2.2.1(ii) (van der Vaart and Wellner, 1996). ∎

Remark 2.2.3. Let $\{X_i\}_{i=1}^n$ be an i.i.d. sample generated according to some $P \in \mathcal{P}$ where $\mathcal{P}$ is dominated by a σ-finite measure μ. Since $\mathcal{P}$ can be embedded into $L^2(\mu)$ via the mapping $Q \mapsto \sqrt{dQ/d\mu}$, we can obtain a tangent set $\dot{\mathcal{S}}^0$ consisting of Fréchet derivatives of differentiable paths $\{dP_t^{1/2}\}$ in $L^2(\mu)$ (Bickel et al., 1993). Define the continuous linear operator $\dot{\theta}_0 : \dot{\mathcal{S}}^0 \to \mathbb{D}$ by $\dot{\theta}_0(g) \equiv \theta_0'(2g/dP^{1/2})$, then (2.9) can be read as

$$\lim_{t \downarrow 0} t^{-1}\{\theta(dP_t^{1/2}) - \theta(dP^{1/2})\} = \dot{\theta}_0(g) , \tag{2.13}$$

where $\{dP_t^{1/2}\}$ is a curve passing $dP^{1/2}$ with Fréchet derivative $g \equiv \frac{1}{2} h dP^{1/2}$. This is exactly Hadamard differentiability if we view θ as a map from $\{\sqrt{dQ/d\mu} : Q \in \mathcal{P}\} \subset L^2(\mu)$ to the space $\mathbb{D}$. ∎

2.2.3 Local Asymptotic Minimaxity

There are different versions of local asymptotic minimax risk. In this section we briefly review some of these and specify the one that is appropriate for our purposes. For simplicity of exposition, let us confine our attention to the i.i.d. case. Let $\mathcal{P}$ be a collection of probability measures, θ the parameter of interest and ℓ a loss function. In an asymptotic framework, a global minimax principle would imply that an asymptotically best estimator sequence $\{T_n\}$ of θ should be the one for which the quantity

$$\liminf_{n\to\infty} \sup_{P\in\mathcal{P}} E_P[\ell(r_n\{T_n - \theta(P)\})] \tag{2.14}$$

is minimized, where E_P denotes expectation under P, and $r_n \uparrow \infty$ is the rate of convergence for estimating θ. While this version is suitable when $\mathcal{P}$ is parametric, it is somewhat too restrictive for semiparametric or nonparametric models. In addition, this approach is excessively cautious since we are able to learn about P with arbitrary accuracy as sample size $n \to \infty$ and hence it would be unreasonable to require nice properties of the estimator sequence around regions too far away from the truth (Hájek, 1972; Ibragimov and Has'minskii, 1981; van der Vaart, 1992). The strategy is then to minimize the asymptotic maximum risk over (shrinking) neighborhoods of the truth.

The earliest consideration of *local* asymptotic minimaxity in the literature is perhaps Chernoff (1956), according to whom the idea actually originated from Charles Stein and Herman Rubin. Different variants have been developed since then (Hájek, 1972; Koshevnik and Levit, 1976; Levit, 1978; Chamberlain, 1987), among which an important version is of the form

$$\lim_{a\to\infty} \liminf_{n\to\infty} \sup_{P\in V_{n,a}} E_P[\ell(r_n\{T_n - \theta(P)\})] \ , \tag{2.15}$$

where $V_{n,a}$ shrinks to the truth as $n \to \infty$ for each fixed $a \in \mathbf{R}$ and spans the whole parameter space as $a \to \infty$ for each fixed $n \in \mathbf{N}$ (Ibragimov and Has'minskii, 1981; Millar, 1983). For instance, Begun et al. (1983) and van der Vaart (1988b) take $V_{n,a}$ to be:

$$V_{n,a} = \{Q \in \mathcal{P} : r_n d_H(Q,P) \le a\} \ , \quad d_H(Q,P) \equiv [\int (dQ^{1/2} - dP^{1/2})^2]^{1/2} \ .$$

However, the above neighborhood versions may invite two problems. First, the neighborhoods might be too large so that the sharp lower bounds are infinite. This is more easily seen in the Hellinger ball version. As pointed out by van der Vaart (1988b, p.32), one may pick $Q_n \in V_n(P, a)$ for each $n \in \mathbf{N}$ such that $\prod_{i=1}^n Q_n$ is not contiguous to $\prod_{i=1}^n P$ (Oosterhoff and van Zwet, 1979, Theorem 1), which in turn implies that $r_n\{T_n - \theta\}$ escapes to "infinity" under $\prod_{i=1}^n Q_n$ (Lehmann and Romano, 2005, Theorem 12.3.2). Second, when it comes to the construction of an optimal estimator, one typically has to establish uniform convergence over the neighborhoods, which may be impossible if the neighborhoods are "too big".

In this paper, we shall consider local asymptotic minimax risk over smaller neighborhoods – more precisely, neighborhoods that consist of finite number of distributions – as in van der Vaart (1988b, 1989, 1998) and van der Vaart and Wellner (1990, 1996):

$$\sup_{I \subset_f \dot{\mathcal{P}}^0} \liminf_{n\to\infty} \sup_{h \in I} E_{P_{n,h}}[\ell(r_n\{T_n - \theta(P_{n,h})\})] , \tag{2.16}$$

where the first supremum is taken over all finite subsets I in the tangent set $\dot{\mathcal{P}}^0$ as defined in Remark 2.2.2, and $\{P_{n,h}\}$ is a differentiable path with score h. This resolves the aforementioned concerns as well as two subtleties that are worth noting here. First, it is necessary to take supremum over neighborhoods of the truth (the second supremum) in order to obtain robust finite sample approximation and as a result rule out superefficient estimators, while the first supremum is needed to remove the uncertainty of the neighborhoods.[13] Second, the local nature of the risk may be translated to the global one if one replaces the second supremum with $\sup_{h\in\dot{\mathcal{P}}^0}$ and ignore the first supremum, so that we are back to the aforementioned uniformity issue. Another possibility is to consider finite dimensional submodels; see Remark 2.2.4.

Remark 2.2.4. As another approach to circumvent the contiguity and uniformity concerns aforementioned, van der Vaart (1988b) considers a version of asymptotic minimaxity based on finite dimensional submodels. Let $h_1, \ldots, h_m \in \dot{\mathcal{P}}^0$ be linearly independent and $\{P_{n,\lambda}^m\}_{n=1}^\infty$ a differentiable path with score $\sum_{j=1}^m \lambda_j h_j$ for each fixed $\lambda \in \mathbf{R}^m$. As λ ranges over $\mathbf{R}^m$, we obtain a full description of local perturbations of some parametric submodel.

[13]The role played by $\sup_{I\subset_f \dot{\mathcal{P}}^0}$ is the same as that by $\lim_{a\to\infty}$ in display (2.15).

Then one may consider the following:

$$\sup_{H_m} \lim_{a\to\infty} \liminf_{n\to\infty} \sup_{\|\lambda\|\leq a} E_{P^m_{n,\lambda}}[\ell(r_n\{T_n - \theta(P^m_{n,\lambda})\})] \ , \tag{2.17}$$

where the first supremum is taken over all finite dimensional subspaces $H_m \subset \dot{\mathcal{P}}^0$ spanned by $h_1, \ldots, h_m$. The same approach has been employed by van der Vaart (1988b, 1989) to obtain generalized convolution theorems for weakly regular estimators. We note however that this version of local asymptotic minimaxity is equivalent to (2.16) in the sense that they yield the same lower bound that is attainable and hence induce the same optimal plug-in estimators. This is essentially because for any parametric submodel $\mathcal{P}^m$ with scores $h_1, \ldots, h_m$, the expansion of the log likelihood ratio (2.8) holds uniformly over $\lambda \in K$ with K any compact set in $\mathbf{R}^m$ (Bickel et al., 1993, Proposition 2.1.2). ■

2.3 Optimal Plug-in Estimators

Building on the ingredients established for θ in previous section, we now proceed to investigate optimal plug-in estimators of $\phi(\theta)$. To begin with, we first review the notion of Hadamard directional differentiability, then establish the minimax lower bound for the class of plug-in estimators, and finally show the attainability by presenting a general procedure of constructing optimal plug-in estimators.

2.3.1 Hadamard Directional Differentiability

A common feature of the examples introduced in Section 2.2.1.1 is that 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 2.2.1 and 2.2.2, in which the domain of ϕ is a finite dimensional space. In order to address Examples 2.2.3 and 2.2.4, 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 2.3.1. Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces equipped with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$ respectively, and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$. The map ϕ is said to be *Hadamard directionally differentiable* at $\theta \in \mathbb{D}_\phi$ if there is a map $\phi'_\theta : \mathbb{D} \to \mathbb{E}$ such that:

$$\lim_{n\to\infty} \|\frac{\phi(\theta + t_n z_n) - \phi(\theta)}{t_n} - \phi'_\theta(z)\|_{\mathbb{E}} = 0 \ , \tag{2.18}$$

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

As various notions of differentiability in the literature, Hadamard directional differentiability can be understood by looking at the restrictions imposed on the approximating map (i.e. the derivative) and the way the approximation error is controlled (Averbukh and Smolyanov, 1967, 1968). Specifically, let

$$\mathrm{Rem}_\theta(z) \equiv \phi(\theta + z) - \{\phi(\theta) + \phi'_\theta(z)\} \ , \tag{2.19}$$

where $\phi(\theta) + \phi'_\theta(z)$ can be viewed as the first order approximation of $\phi(\theta + z)$. Hadamard directional differentiability of ϕ then amounts to requiring the approximation error $\mathrm{Rem}_\theta(z)$ satisfy that $\mathrm{Rem}_\theta(tz)/t$ tends to zero uniformly in $z \in K$ for any compact set K – i.e.

$$\sup_{z\in K} \|\frac{\mathrm{Rem}_\theta(tz)}{t}\|_{\mathbb{E}} \to 0 \ , \ \text{as } t \downarrow 0 \ .$$

However, unlike Hadamard differentiability that requires the approximating map ϕ'_θ be linear and continuous, linearity of the directional counterpart is often lost though the continuity is automatic (Shapiro, 1990). In fact, linearity of the derivative is the exact gap between these two notions of differentiability.

The way that Hadamard directional differentiability controls the approximation error ensures the validity of the Delta method, which we exploit in our asymptotic analysis. Moreover, the chain rule remains valid for compositions of Hadamard directional differentiable maps; see Remark 2.3.1.[14] We note also that though Definition 3.2.2 is adequate for

[14] In fact, by slight modifications of the arguments employed in Averbukh and Smolyanov (1968), one can

our purposes in this paper, there is a *tangential* version of Hadamard directional differentiability, which restricts the domain of the derivative ϕ'_{θ_0} to be a subset of $\mathbb{D}$.

Remark 2.3.1. Suppose that $\psi : \mathbb{B} \to \mathbb{D}_\phi \subset \mathbb{D}$ and $\phi : \mathbb{D}_\phi \to \mathbb{E}$ are Hadamard directionally differentiable at $\vartheta \in \mathbb{B}$ and $\theta \equiv \psi(\vartheta) \in \mathbb{D}_\phi$ respectively, then $\phi \circ \psi : \mathbb{B} \to \mathbb{E}$ is Hadamard directionally differentiable at ϑ with derivative $\phi'_\theta \circ \psi'_\vartheta : \mathbb{B} \to \mathbb{E}$ (Shapiro, 1990, Proposition 3.6). Thus, if $\theta : \{P_{n,h}\} \to \mathbb{D}_\phi$ is not regular but $\theta(P_{n,h}) = \psi(\vartheta(P_{n,h}))$ for some parameter $\vartheta : \{P_{n,h}\} \to \mathbb{B}$ admitting a regular estimator $\hat{\vartheta}_n$ and a Hadamard directionally differentiable map ψ, then the results in this paper may be applied with $\tilde{\phi} \equiv \phi \circ \psi$, $\tilde{\theta}(P_{n,h}) \equiv \vartheta(P_{n,h})$, and $\hat{\vartheta}_n$ in place of ϕ, $\theta(P_{n,h})$ and $\hat{\theta}_n$ respectively. ■

2.3.1.1 Examples Revisited

We next verify Hadamard directional differentiability of the maps in the examples introduced in Section 2.2.1.1, and hence show that they indeed fall into our setup. The first example is straightforward.

Example 2.2.1 (Continued). Let $j^* = \arg\max_{j \in \{1,2\}} \theta^{(j)}$. For any $z = (z^{(1)}, z^{(2)})' \in \mathbf{R}^2$, simple calculations reveal that $\phi'_\theta : \mathbf{R}^2 \to \mathbf{R}$ is given by

$$\phi'_\theta(z) = \begin{cases} z^{(j^*)} & \text{if } \theta^{(1)} \neq \theta^{(2)} \\ \max\{z^{(1)}, z^{(2)}\} & \text{if } \theta^{(1)} = \theta^{(2)} \end{cases} . \tag{2.20}$$

Note that ϕ'_θ is nonlinear when Hadamard differentiability is not satisfied. ■

Example 2.2.2 (Continued). In this example, by the chain rule (see Remark 2.3.1) it is easy to verify that

$$\phi'_\theta(z) = \psi'_\theta(z) 1\{\psi(\theta) > 0\} + \max\{\psi'_\theta(z), 0\} 1\{\psi(\theta) = 0\} , \tag{2.21}$$

show that Hadamard directional differentiability is the weakest directional differentiability that satisfies the chain rule, just as Hadamard differentiability is the weakest differentiability that does the same job.

where $1\{\cdot\}$ denotes the indicator function, and

$$\begin{aligned}\psi'_\theta(z) =& \frac{[\lambda^{(1)}(z^{(1)}+z^{(2)})+\lambda^{(2)}(z^{(1)}-z^{(2)})][\theta^{(1)}+\theta^{(2)}-(\theta^{(2)}-\theta^{(1)})^2]}{[\theta^{(1)}+\theta^{(2)}-(\theta^{(2)}-\theta^{(1)})^2]^2}\\ &-\frac{[\lambda^{(1)}(\theta^{(1)}+\theta^{(2)})+\lambda^{(2)}(\theta^{(1)}-\theta^{(2)})][z^{(1)}+z^{(2)}-2(\theta^{(2)}-\theta^{(1)})(z^{(2)}-z^{(1)})]}{[\theta^{(1)}+\theta^{(2)}-(\theta^{(2)}-\theta^{(1)})^2]^2}\ .\end{aligned}$$

Clearly, the directional derivative ϕ'_θ is nonlinear at θ with $\psi(\theta)=0$. ■

Example 2.2.3 and 2.2.4 are more involved in that the domain and range of ϕ are both infinite dimensional.

Example 2.2.3 (Continued). Let $B_1 = 1\{x : \theta^{(1)}(x) > \theta^{(2)}(x)\}$, $B_2 = 1\{x : \theta^{(2)}(x) > \theta^{(1)}(x)\}$ and $B_0 = 1\{x : \theta^{(1)}(x) = \theta^{(2)}(x)\}$. Then it is not hard to show that ϕ is Hadamard directionally differentiable at any $\theta \in \ell^\infty(\mathbf{R})\times\ell^\infty(\mathbf{R})$ satisfying for any $z \in \ell^\infty(\mathbf{R})\times\ell^\infty(\mathbf{R})$,

$$\phi'_\theta(z) = z^{(1)}1_{B_1} + z^{(2)}1_{B_2} + \max\{z^{(1)}, z^{(2)}\}1_{B_0}\ . \tag{2.22}$$

Here, nonlinearity occurs when the set of points at which $\theta^{(1)}$ and $\theta^{(2)}$ are equal is not empty, implying Hadamard directional differentiability. ■

Example 2.2.4 (Continued). For a set $A \subset L^2(\mathcal{T})$, denote the closed linear span of A by $[A]$, and define the complement $A^\perp$ of A by $A^\perp \equiv \{z \in L^2(\mathcal{T}) : \langle z,\lambda\rangle = 0 \text{ for all } \lambda \in A\}$. Lemma 2.6.11 shows that Π_Λ is Hadamard directionally differentiable at every $\theta \in L^2(\mathcal{T})$ and the resulting derivative satisfies for all $z \in L^2(\mathcal{T})$

$$\phi'_\theta(z) = \Pi_{C_\theta}(z)\ , \tag{2.23}$$

where

$$C_\theta = T_{\bar\theta} \cap [\theta - \bar\theta]^\perp\ ,\ T_{\bar\theta} = \overline{\bigcup_{\alpha\ge 0} \alpha\{\Lambda - \bar\theta\}}\ , \tag{2.24}$$

with $\bar\theta = \Pi_\Lambda\theta$. Note that C_θ is a closed convex cone, which can be thought of as a local approximation to Λ at θ along the direction perpendicular to the projection residual

$\theta - \Pi_\Lambda \theta$. Unlike Fang and Santos (2014), the consideration of nonboundary points $\theta \notin \Lambda$ here is necessitated by the possible misspecification of conditional quantile functions. ■

2.3.2 The Lower Bounds

As the first step towards establishing the minimax lower bound, we would like to leverage the Delta method for Hadamard directionally differentiable maps (Shapiro, 1991; Dümbgen, 1993) to derive the weak limits of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\}$ under $\{P_{n,h}\}$. This is not a problem in i.i.d. settings since we may write

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\} = r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(0))\} - r_n\{\phi(\theta_n(h)) - \phi(\theta_n(0))\} \ ,$$

and then Delta method can be employed right away in view of the fact that $\theta_n(0)$ is typically a constant. In general, however, we would hope the directional differentiability of ϕ is strong enough to possess uniformity to certain extent.

There are two ways to obtain uniform differentiability in general. One natural way, of course, is to incorporate uniformity into the definition of differentiability (van der Vaart and Wellner, 1996, Theorem 3.9.5). For differentiable maps, continuous differentiability suffices for uniform differentiability; for directionally differentiable ones, unfortunately, continuous differentiability is a rare phenomenon. In fact, one can show by way of example that it is unwise to include uniformity in the definition of Hadamard directional differentiability. The other general principle of obtaining uniformity is to require $\theta_n(0)$ converge sufficiently fast. Following Dümbgen (1993), we take this latter approach and require $\theta_n(0)$ converge in the following manner:

Assumption 2.3.1. There are fixed $\theta_0 \in \mathbb{D}_\phi$ and $\Delta \in \theta_0'(H)$ such that as $n \to \infty$,

$$r_n\{\theta_n(0) - \theta_0\} \to \Delta \ . \tag{2.25}$$

Assumption 2.3.2. The map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$, where $\mathbb{E}$ is a Banach space with norm $\|\cdot\|_{\mathbb{E}}$, is Hadamard directionally differentiable at θ_0.

In the i.i.d. setup, Assumption 2.3.1 is automatically satisfied with $\theta_n(0) = \theta_0 \equiv \theta(P)$, $\Delta = 0$, and $\{r_n\}$ any sequence. Assumption 2.3.2 simply formalizes the appropriate notion of directional differentiability of ϕ. It is worth noting that directional differentiability is only assumed at θ_0. This Hadamard directional differentiability condition, together with Assumptions 2.2.2, 2.2.3, and 2.3.1, allows us to deduce weak limits of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\}$ under $\{P_{n,h}\}$.

Next, minimaxity analysis necessitates the specification of a loss function or a family of loss functions. As recommended by Strasser (1982), we shall consider a collection of loss functions and they are specified as follows:

Assumption 2.3.3. The loss function $\ell : \mathbb{E} \to \mathbf{R}^+$ is such that $\ell_M \equiv \ell \wedge M$ is Lipschitz continuous, i.e. for each $M > 0$, there is some constant $C_{\ell,M} > 0$ such that:

$$|\ell_M(x) - \ell_M(y)| \leq C_{\ell,M}\|x - y\|_{\mathbb{E}} \text{ for all } x, y \in \mathbb{E}\ . \tag{2.26}$$

Assumption 2.3.3 includes common loss functions such as quadratic loss, absolute loss, and quantile loss but excludes the zero-one loss. We emphasize that the symmetry of ℓ is not needed here. From a technical level, this is because we no longer need Anderson's lemma to derive the lower bound of minimax risk. Moreover, we note that Assumption 2.3.3 clearly implies continuity of ℓ and Lipschitz continuity if ℓ is bounded.

Given the ability to derive weak limits of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\}$, asymptotic normality of $\{P_{n,h}\}$, and a loss function ℓ, we are able to obtain the lower bound of local asymptotic minimax risk as the first main result of this paper.

Theorem 2.3.1. *Let $(\mathscr{X}_n, \mathcal{A}_n, \{P_{n,h} : h \in H\})$ be a sequence of statistical experiments, and $\hat{\theta}_n$ a map from the data $\{X_i\}_{i=1}^n$ into a set $\mathbb{D}_\phi$. Suppose that Assumptions 2.2.1, 2.2.2, 2.2.3, 2.3.1, 2.3.2 and 2.3.3 hold. Then it follows that*

$$\sup_{I \subset_f H} \liminf_{n\to\infty} \sup_{h\in I} E_{n,h}[\ell(r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\})]$$
$$\geq \inf_{u\in\mathbb{D}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))]\ , \tag{2.27}$$

where $E_{n,h}$ denotes the expectation evaluated under $P_{n,h}$.

The lower bound takes a minimax form which in fact is consistent with regular cases – i.e. when ϕ is Hadamard differentiable or equivalently ϕ'_{θ_0} is linear, in which the lower bound is given by $E[\ell(\phi'_{\theta_0}(\mathbb{G}_0))]$ provided that ℓ is subconvex (van der Vaart and Wellner, 1996). To see this, note that if ϕ'_{θ_0} is linear, then

$$\inf_{u\in\mathbb{D}}\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]$$
$$=\inf_{u\in\mathbb{D}} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0)+\phi'_{\theta_0}(u))]=E[\ell(\phi'_{\theta_0}(\mathbb{G}_0))]\ ,$$

where the last step is by Anderson's lemma since ℓ is subconvex and $\phi'_{\theta_0}(\mathbb{G}_0)$ is Gaussian in view of ϕ'_{θ_0} being continuous and linear. Thus, the minimax form in (2.27) is caused entirely by the nonlinearity of ϕ'_{θ_0}. We note also that the lower bound in Theorem 2.3.1 is consistent with that in Song (2014) for the special class of parameters studied there.

If the lower bound in (2.27) is infinite, then any estimator is "optimal". One should then change the loss function or work with an alternative optimality criteria so that the problem becomes nontrivial. Given a particular loss function, finiteness of the lower bound hinges on the nature of both the model and the parameter being estimated. For the sake of finiteness of the lower bound, we thus require the derivative ϕ'_{θ_0} satisfy:

Assumption 2.3.4. The derivative ϕ'_{θ_0} is Lipschitz continuous, i.e. there exists some constant $C_{\phi'}>0$ possibly depending on θ_0 such that

$$\|\phi'_{\theta_0}(z_1)-\phi'_{\theta_0}(z_2)\|_{\mathbb{E}}\le C_{\phi'}\|z_1-z_2\|_{\mathbb{D}} \text{ for all } z_1,z_2\in\mathbb{D}_\phi\ . \tag{2.28}$$

Assumption 2.3.4 in fact is satisfied in all of our examples; see Section 2.3.2.1. The following Lemma shows that Assumption 2.3.4 ensures finiteness of the lower bound in (2.27) for a class of popular loss functions.

Lemma 2.3.1. *Let $\ell(\cdot)=\rho(\|\cdot\|_{\mathbb{E}})$ for some nondecreasing lower semicontinuous function*

$\rho : \mathbf{R}^+ \to \mathbf{R}^+$. *If Assumption 2.3.4 holds and* $E[\rho(C_{\phi'}\|\mathbb{G}_0\|_{\mathbb{D}})] < \infty$, *then*

$$\inf_{u\in\mathbb{D}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))] < \infty \ .$$

The moment condition in Lemma 2.3.1 is easy to verify in practice when combined with Lipschitz property of ρ (Bogachev, 1998, Theorem 4.5.7) or tail behavior of the CDF of $\|\mathbb{G}_0\|_{\mathbb{E}}$ (Davydov et al., 1998, Proposition 11.6) but by no means necessary. If the lower bound is finite, this would not be a concern in the first place. As another example, if $\mathbb{D}$ is Euclidean, then it suffices that there is some $\delta > 0$ such that

$$\sup_{c\in\mathbf{R}^m} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + c) - \phi'_{\theta_0}(c))^{1+\delta}] < \infty \ .$$

In cases when θ is Euclidean valued – i.e. $\mathbb{D} = \mathbf{R}^m$ for some $m \in \mathbf{N}$, we have a simpler form of the lower bound in (2.27). This includes semiparametric and nonparametric models as well as parametric ones; see Examples 2.2.1 and 2.2.2.

Corollary 2.3.1. *Let* $(\mathscr{X}_n, \mathcal{A}_n, \{P_{n,h} : h \in H\})$ *be a sequence of statistical experiments, and* $\hat{\theta}_n$ *an estimator for the parameter* $\theta : \{P_{n,h}\} \to \mathbb{D}_\phi \subset \mathbb{D}$ *with* $\mathbb{D} = \mathbf{R}^m$ *for some* $m \in \mathbf{N}$. *Suppose that Assumptions 2.2.1, 2.2.2, 2.2.3, 2.3.1, 2.3.2 and 2.3.3 hold. If the efficiency bound* $\Sigma_0 \equiv \langle\tilde{\theta}_0, \tilde{\theta}_0^\intercal\rangle$ *is nonsingular, then it follows that*

$$\sup_{I\subset_f H} \liminf_{n\to\infty} \sup_{h\in I} E_{n,h}[\ell(r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\})] \geq \inf_{u\in\mathbf{R}^m} \sup_{c\in\mathbf{R}^m} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + c) - \phi'_{\theta_0}(c))] \ . \tag{2.29}$$

The lower bound in (2.29) is a minimax optimization problem over $\mathbf{R}^m$; in particular, the supremum is taken over $\mathbf{R}^m$ instead of the tangent set. This simply follows from the facts that the support of $\mathbb{G}_0$ is $\overline{\theta'_0(H)}$ by Theorem 2.2.1 and that a nondegenerate Gaussian random variable in $\mathbf{R}^m$ has support $\mathbf{R}^m$. As a result, the construction of optimal plug-in estimators in Section 2.3.3 becomes much easier when θ is Euclidean valued.

2.3.2.1 Examples Revisited

In this section we explicitly derive the lower bound for each example introduced in Section 2.2.1.1. For simplicity of illustration, we confine our attention to the simplest i.i.d. setup. That is, we assume that the sample $X_1, \ldots, X_n$ is i.i.d. and distributed according to $P \in \mathcal{P}$, and we are interested in estimating $\phi(\theta)$.

Example 2.2.1 (Continued). Simple algebra reveals that ϕ'_θ is Lipschitz continuous. In order to compare with previous literature, consider the case when X is bivariate normal with covariance matrix $\sigma^2 I_2$, and take the squared loss function. As shown in Appendix 2.6.2, the lower bounds is given by

$$\begin{aligned} &\inf_{u \in \mathbf{R}^2} \sup_{c \in \mathbf{R}^2} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + c) - \phi'_{\theta_0}(c))] \\ &= \inf_{u \in \mathbf{R}^2} \sup_{c \in \mathbf{R}^2} E[(\max\{\mathbb{G}_0^{(1)} + u^{(1)} + c^{(1)}, \mathbb{G}_0^{(2)} + u^{(2)} + c^{(2)}\} - \max\{c^{(1)}, c^{(2)}\})^2] = \sigma^2 \ , \end{aligned}$$

where $\mathbb{G}_0 \equiv (\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)}) \sim N(0, \sigma^2 I_2)$, and the infimum is achieved when $u = (-\infty, 0)$ and $c = (-\infty, c^{(2)})$ with $c^{(2)} \in \mathbf{R}$ arbitrary. In fact, the lower bound can be also achieved at $u = 0$ and $c = 0$. We note that this lower bound is consistent with Song (2014) and Blumenthal and Cohen (1968b). ■

Example 2.2.2 (Continued). In this case, it is also easy to see that ϕ'_θ is Lipschitz continuous. For the squared loss function, the lower bound at the point θ_0 with $\psi(\theta_0) = 0$ becomes

$$\inf_{u \in \mathbf{R}^2} \sup_{c \in \mathbf{R}^2} E[(\max\{\psi'_{\theta_0}(\mathbb{G}_0 + u + c), 0\} - \max\{\psi'_{\theta_0}(c), 0\})^2] \ ,$$

where $\mathbb{G}_0$ is the efficient Gaussian limit for estimating θ_0. ■

Example 2.2.3 (Continued). In this example, it can be shown that ϕ'_θ is Lipschitz continuous. For the loss function $\ell(z) = \|z\|_\infty$, the lower bound becomes

$$\inf_{u^{(1)},u^{(2)}\in\ell^\infty(\mathbf{R})} \sup_{h^{(1)},h^{(2)}\in H} \{E[\|(\mathbb{G}_0^{(1)} + u^{(1)})1_{B_1} + (\mathbb{G}_0^{(2)} + u^{(2)})1_{B_2} + \max\{\mathbb{G}_0^{(1)} + u^{(1)} + h^{(1)}, \mathbb{G}_0^{(2)} + u^{(2)} + h^{(2)}\}1_{B_0} - \max\{h^{(1)}, h^{(2)}\}1_{B_0}\|_\infty]\} ,$$

where H consists of all bounded measurable real valued functions on $\mathbf{R}$ with $\int_{\mathbf{R}} h\,dP = 0$, and $(\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)})$ is the efficient Gaussian limit in $\ell^\infty(\mathbf{R})\times\ell^\infty(\mathbf{R})$ for estimating $\theta_0 \equiv (F_1, F_2)$.

■

Example 2.2.4 (Continued). Since C_{θ_0} is closed and convex, ϕ'_{θ_0} or equivalently $\Pi_{C_{\theta_0}}$ is Lipschitz continuous (Zarantonello, 1971, p.241). If the loss function $\ell(\cdot) : L^2(\mathcal{T}) \to \mathbf{R}$ is $\ell(z) = \|z\|^2_{L^2}$, then the lower bound is finite and given by

$$\inf_{u\in L^2(\mathcal{T})} \sup_{h\in H} E[\|\Pi_{C_{\theta_0}}(\mathbb{G}_0 + u + \theta'_0(h)) - \Pi_{C_{\theta_0}}(\theta'_0(h))\|^2_{L^2}] ,$$

where $H \equiv \{(h_1, h_2) : h_1 \in H_1, h_2 \in H_2\}$ with[15]

$$H_1 \equiv \{h_1 : \mathscr{Z} \to \mathbf{R} : E[h_1(Z)] = 0\} ,$$
$$H_2 \equiv \{h_2 : \mathscr{Y} \times \mathscr{Z} \to \mathbf{R} : E[h_2(Y, z)] = 0 \text{ for a.s. } z \in \mathscr{Z}\} ,$$

$\mathbb{G}_0$ is a zero mean Gaussian process in $L^2(\mathcal{T})$ with covariance function $\mathrm{Cov}(\tau_1, \tau_2) \equiv J(\tau_1)^{-1}\Gamma(\tau_1, \tau_2)J(\tau_2)^{-1}$ in which for $f_Y(y|Z)$ the density of Y conditional on Z,

$$J(\tau) \equiv c'E[f_Y(Z'\beta(\tau)|Z)ZZ'] , \forall\, \tau \in \mathcal{T} ,$$
$$\Gamma(\tau_1, \tau_2) \equiv E[(\tau_1 - 1\{Y \le Z'\beta(\tau_1)\})(\tau_2 - 1\{Y \le Z'\beta(\tau_2)\})ZZ'] , \forall\, \tau_1, \tau_2 \in \mathcal{T} ,$$

[15] see Severini and Tripathi (2001). Technically, H here is not the tangent set; however, every element in the tangent set can be written as a unique decomposition involving some pair in H. This shouldn't bother us since tangent set *per se* is not of our interest.

and,

$$\theta_0'(h)(\tau) \equiv -J(\tau)^{-1}\int c'z1\{y \le z'\beta(\tau)\}h_1(y,z)P(dy,dz)$$
$$- J(\tau)^{-1}\int c'z(1\{y \le z'\beta(\tau)\} - \tau)h_2(z)P(dy,dz)\ .$$

For a detailed discussion on the efficient estimation of θ, see Lee (2009, Theorem 3.1). ■

2.3.3 Attainability via Construction

Having established the lower bounds as in Theorem 2.3.1 and Corollary 2.3.1, we now proceed to show the attainability of the bounds by developing a general procedure of constructing optimal plug-in estimators. The lower bounds in (2.27) and (2.29) suggest that an optimal plug-in estimator is of the form $\phi(\hat{\theta}_n + \hat{u}_n/r_n)$ where $\hat{u}_n$ is an estimator of the optimal noise term in Theorem 2.2.1 – i.e. $\hat{u}_n$ should be an estimator of the minimizer(s) in the lower bounds. We deal with infinite dimensional $\mathbb{D}$ first in order to accommodate Examples 2.2.3 and 2.2.4, and then specialize to Euclidean $\mathbb{D}$.

Recall from Theorem 2.3.1 that the lower bound for the local asymptotic minimax risk is given by

$$\inf_{u\in\mathbb{D}}\sup_{h\in H} E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u + \theta_0'(h)) - \phi_{\theta_0}'(\theta_0'(h)))]\ . \tag{2.30}$$

If the objective function in (2.30) were known, we would pick the optimal correction term by solving a minimax optimization problem. However, this is not the case since there are four unknown objects here: the law of the efficient Gaussian component $\mathbb{G}_0$, the derivatives ϕ_{θ_0}' and θ_0', and the space H. We thus work with the sample analog of (2.30) by replacing $\mathbb{G}_0$, ϕ_{θ_0}', θ_0', and H with their sample counterparts.

We shall assume that the law of $\mathbb{G}_0$ can be estimated by bootstrap or simulation. Specifically, let $\hat{\theta}_n$ be an efficient estimator of θ, and $\hat{\theta}_n^*$ a bootstrapped version of it – i.e. $\hat{\theta}_n^*$ is a function mapping the data $\{X_i\}_{i=1}^n$ and random weights $\{W_i\}$ that are independent of $\{X_i\}$ into the domain $\mathbb{D}_\phi$ of ϕ. This abstract definition suffices for encompassing the nonparametric, Bayesian, block, score, and weighted bootstrap as special cases. The hope is then that the limiting law of $r_n\{\hat{\theta}_n - \theta_0\}$ can be consistently estimated by the (finite sample) law of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$, which necessitates a metric that measures distances between probability measures. Since the law $\mathbb{G}_0$ is tight and hence separable, we may employ the bounded Lipschitz metric d_{BL} introduced by Dudley (1966, 1968): for two Borel probability measures L_1 and L_2 on $\mathbb{D}$, define

$$d_{\mathrm{BL}}(L_1, L_2) \equiv \sup_{f \in \mathrm{BL}_1(\mathbb{D})} |\int f\, dL_1 - \int f\, dL_2| \ ,$$

where recall that $\mathrm{BL}_1(\mathbb{D})$ is the set of bounded and Lipschitz continuous functions as defined in (**??**). We may now measure the distance between the law of $\hat{\mathbb{G}}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ conditional on $\{X_i\}$ and the limiting law $\mathbb{G}_0$ of $r_n\{\hat{\theta}_n - \theta_0\}$ by

$$d_{\mathrm{BL}}(\hat{\mathbb{G}}_n^*, \mathbb{G}_0) = \sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}] - E[f(\mathbb{G}_0)]| \ . \tag{2.31}$$

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 (2.31), converges in probability to zero.

The estimation of θ_0' can be done by analogy principle since the derivative θ_0' typically takes the form $\theta_0' \equiv \theta_0'(P)$, that is, we may estimate θ_0' by $\hat{\theta}_n' = \theta_0'(\mathbb{P}_n)$ with $\mathbb{P}_n$ the empirical measure. Estimation of the derivative ϕ_{θ_0}' is trickier. In this regard, we impose sufficient conditions so as to meet Assumption 3.3 in Fang and Santos (2014). The following assumption formalizes our discussion so far.

Assumption 2.3.5. (i) $\hat{\mathbb{G}}_n^* : \{X_i, W_i\}_{i=1}^n \to \mathbb{D}_\phi$ with $\{W_i\}$ independent of $\{X_i\}$ satisfies $\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(\hat{\mathbb{G}}_n^*)|\{X_i\}] - E[f(\mathbb{G}_0)]| = o_p(1)$ under $\{P_{n,0}\}$.

(ii) $\hat{\theta}_n' : H \to \mathbb{D}$ depends on $\{X_i\}$ and satisfies $\|\hat{\theta}_n'(\hat{h}_n) - \theta_0'(h)\|_{\mathbb{D}} \xrightarrow{p} 0$ under $\{P_{n,0}\}$ whenever $\|\hat{h}_n - h\|_H \xrightarrow{p} 0$ under $\{P_{n,0}\}$ with $\hat{h}_n : \{X_i\} \to H$.

(iii) $\hat{\phi}_n' : \mathbb{D} \to \mathbb{E}$ depends on $\{X_i\}$ satisfying (a) for any $z \in \mathbb{D}$, $\hat{\phi}_n'(z)$ is consistent for $\phi_{\theta_0}'(z)$ – i.e. $\|\hat{\phi}_n'(z) - \phi_{\theta_0}'(z)\|_{\mathbb{E}} \xrightarrow{p} 0$ under $\{P_{n,0}\}$; and (b) there is some deterministic constant $C_{\hat{\phi}'}$ such that $\|\hat{\phi}'_n(z_1) - \hat{\phi}'_n(z_2)\|_{\mathbb{E}} \leq C_{\hat{\phi}'}\|z_1 - z_2\|_{\mathbb{D}}$ outer almost surely for all $z_1, z_2 \in \mathbb{D}$.

Assumption 2.3.5(i) is simply a bootstrap consistency condition on $\hat{\mathbb{G}}_n^*$ for the target law of $\mathbb{G}_0$, including Song (2014)'s simulation method as a special case. Assumption 2.3.5(ii) imposes a weak consistency condition on the estimator $\hat{\theta}_n$. One might require $\hat{\theta}'_n$ be consistent in the sense that $\|\hat{\theta}'_n - \theta'_0\|_{op} \xrightarrow{p} 0$ where $\|\cdot\|_{op}$ is the operator norm. However, such an assumption is too restrictive for a Glivenko-Cantelli argument to hold since the operator norm is a supremum taken over all $h \in H$ with $\|h\|_H \leq 1$. The pointwise consistency condition on $\hat{\phi}'_n$ in Assumption 2.3.5(iii)-(a) is a minimal requirement, while Assumption 2.3.5(iii)-(b) imposes Lipschitz continuity on $\hat{\phi}'_n$, a condition inherited from ϕ'_{θ_0} as in Assumption 2.3.4. Assumptions 2.3.5(iii)-(a) and -(b) together imply that $\hat{\phi}'_n$ converges in probability to ϕ_{θ_0} uniformly over all δ-enlargement of compact sets in $\mathbb{D}$, a condition that has been employed in Fang and Santos (2014) to construct a valid inference procedure for the parameter $\phi(\theta)$.

We next deal with approximating the spaces H and $\mathbb{D}$ as needed to construct an analog to the bound (2.30). To understand the unknown nature of H, consider the i.i.d. setup in which case $H \equiv \dot{\mathcal{P}}^0$ where $\dot{\mathcal{P}}^0$ is the tangent set as defined in Remark 2.2.2. In these settings, it is common that $\dot{\mathcal{P}}^0$ is equal to the largest possible tangent set $L_0^2(P) \equiv \{h \in L^2(P) : \int h dP = 0\}$, which depends on the unknown probability measure P. It is worth noting that $L_0^2(P)$ can be viewed as the projection of $L^2(P)$ onto the complement of the subspace of constant functions. In fact, this projection nature of $\dot{\mathcal{P}}^0$ is prevalent in efficient estimation (Bickel et al., 1993), an insight helpful to the estimation of H.

Since both H and $\mathbb{D}$ are infinite dimensional, we need to approximate H and $\mathbb{D}$ by sequences of sieve spaces, which typically consist of compact subsets or finite dimensional subspaces that grow dense in H and $\mathbb{D}$. Consider the space H first. If we have a "basis" $\{g_m\}$ for $\dot{\mathcal{P}}^0$, then we may approximate H by finite dimensional subspaces constructed from $\{g_m\}$. For example, the space $C_c^0(\mathbf{R}^{d_x})$ of mean zero continuous functions on $\mathbf{R}^{d_x}$ with compact support is dense in $L_0^2(P)$; by the Stone-Weierstrass theorem, the set of polynomial functions are in turn dense in $C_c^0(\mathbf{R}^{d_x})$. Thus, following Chamberlain (1987) who approximates the efficiency bound in models defined by conditional moment restrictions based on polynomials, we may take the polynomials, properly projected or truncated, as a complete sequence in $\overline{H}$. As for the space $\mathbb{D}$ over which the infimum is taken, we may employ linear sieves as approximation. These being said, we assume the following:

Assumption 2.3.6. (i) $\{g_m\}_{m=1}^{\infty} \subset H$ is complete in the sense that for each $h \in H$ and $\epsilon > 0$, there exists $\alpha_1, \ldots, \alpha_m$ such that $\|h - \sum_{j=1}^{m} \alpha_j g_j\|_H < \epsilon$; (ii) for each $m \in \mathbf{N}$, $\hat{g}_m : \{X_i\} \to H$ satisfies $\|\hat{g}_m - g_m\|_H \xrightarrow{p} 0$ under $\{P_{n,0}\}$; (iii) $\{\psi_k\}_{k=1}^{\infty} \subset \mathbb{D}$ is complete.

Assumption 2.3.6(i) formalizes the approximation property of $\{g_m\}$, in a way like the Schauder basis except that the representation coefficients α_j might not be unique, while Assumption 2.3.6(iii) is a similar approximation condition imposed on $\{\psi_k\}$. Assumption 2.3.6(ii) requires that $\{g_m\}$ be estimated by a sequence $\{\hat{g}_m\}$ of random variables to accommodate the unknown nature of H.

Given the availability of complete sequences $\{g_m\}$ and $\{\psi_k\}$ in H and $\mathbb{D}$ respectively, we may approximate the lower bound (2.30) by

$$\min_{v \in K_{\tau_k}^k} \max_{c \in K_{\lambda_m}^m} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta'_0(g^m)^\intercal c) - \phi'_{\theta_0}(\theta'_0(g^m)^\intercal c))] \,, \tag{2.32}$$

where $K_{\tau_k}^k$ and $K_{\lambda_m}^m$ are balls in $\mathbf{R}^k$ and $\mathbf{R}^m$ respectively as defined in the beginning of Section 2.2.1, $\{\lambda_m\}$ and $\{\tau_k\}$ are sequences that diverge to infinity as $m, k \to \infty$ respectively, and $\theta'_0(g^m) \equiv (\theta'_0(g_1), \ldots, \theta'_0(g_m))^\intercal$. Heuristically, (2.32) is the bound for the parametric submodel whose tangent set is $\{c^\intercal g^m : c \in K_{\lambda_m}^m\}$ and noise term v is restricted to be bounded in norm by τ_k. As the approximation indices m, k increase to infinity, (2.32) converges to the lower bound (2.30). With g^m, $\mathbb{G}_0$, θ'_0 and ϕ'_{θ_0} in (2.32) replaced by the corresponding estimates $\{\hat{g}_m\}$, $\hat{\mathbb{G}}_n^*$, $\hat{\theta}'_n$ and $\hat{\phi}'_n$, the bound (2.32) can in turn be estimated by

$$\min_{v \in K_{\tau_k}^k} \max_{c \in K_{\lambda_m}^m} E[\ell(\hat{\phi}'_n(\hat{\mathbb{G}}_n^* + (\psi^k)^\intercal v + \hat{\theta}'_n(\hat{g}^m)^\intercal c) - \hat{\phi}'_n(\hat{\theta}'_n(\hat{g}^m)^\intercal c))|\{X_i\}] \,, \tag{2.33}$$

where $\hat{\theta}'_n(\hat{g}^m) \equiv (\hat{\theta}'_n(\hat{g}_1), \ldots, \hat{\theta}'_n(\hat{g}_m))^\intercal$, and the expectation is evaluated with respect to the bootstrap weights $\{W_i\}_{i=1}^n$ holding $\{X_i\}_{i=1}^n$ fixed. For notational simplicity, define

$$\hat{B}_m(v) \equiv \max_{c \in K_{\lambda_m}^m} E[\ell(\hat{\phi}_n'(\hat{\mathbb{G}}_n^* + (\psi^k)^\intercal v + \hat{\theta}_n'(\hat{g}^m)^\intercal c) - \hat{\phi}_n'(\hat{\theta}_n'(\hat{g}^m)^\intercal c))|\{X_i\}] ,$$

$$\hat{\Psi}_{k,m} \equiv \{v \in K_{\tau_k}^k : \hat{B}_m(v) \leq \min_{v' \in K_{\tau_k}^k} \hat{B}_m(v') + \epsilon_n\} ,$$

where $\epsilon_n = o_p(1)$ as $n \to \infty$. Here, $\hat{\Psi}_{k,m}$ is the set of minimizers for the sample analog approximating problem (2.33), allowing negligible computational error ϵ_n that tends to zero in probability.

We are now ready to construct the optimal plug-in estimators. For any $\hat{v}_{n,k,m} \in \hat{\Psi}_{k,m}$, we consider estimating $\phi(\theta_n(h))$ by

$$\phi(\hat{\theta}_n + \frac{\hat{u}_{n,k,m}}{r_n}) , \quad \hat{u}_{n,k,m} \equiv (\psi^k)^\intercal \hat{v}_{n,k,m} , \tag{2.34}$$

where $\hat{\theta}_n$ is an efficient estimator of θ – i.e. it satisfies

Assumption 2.3.7. $\{\hat{\theta}_n\}$ is an efficient estimator of θ – i.e. for each $h \in H$,

$$r_n\{\hat{\theta}_n - \theta_n(h)\} \overset{L_{n,h}}{\to} \mathbb{G}_0 \text{ in } \mathbb{D} ,$$

where $\mathbb{G}_0$ is the efficient Gaussian random variable as in Theorem 2.2.1.

Our first construction result shows that the plug-in estimator (2.34) attains the local asymptotic minimax lower bound (2.30).

Theorem 2.3.2. *Suppose that Assumptions 2.2.1, 2.2.2, 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5, 2.3.6, and 2.3.7 hold. Let $\{\lambda_m\}$ and $\{\tau_k\}$ be sequences that diverge to infinity as $m, k \to \infty$ respectively. If $\hat{v}_{n,k,m} \in \hat{\Psi}_{k,m}$, then*

$$\limsup_{k\to\infty} \limsup_{m\to\infty} \sup_{I \subset_f H} \limsup_{n\to\infty} \sup_{h \in I} E_{n,h}\big[\ell\big(r_n\big(\phi(\hat{\theta}_n + \frac{\hat{u}_{n,k,m}}{r_n}) - \phi(\theta_n(h))\big)\big)\big]$$
$$\leq \inf_{u \in \mathbb{D}} \sup_{h \in H} E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u + \theta_0'(h)) - \phi_{\theta_0}'(\theta_0'(h)))] , \tag{2.35}$$

where $\hat{u}_{n,k,m} \equiv (\psi^k)^\intercal \hat{v}_{n,k,m}$.

We note that, though unpleasant, the first two lim sup's over k and m are necessary in general and more importantly are taken after letting $n \to \infty$. The reason is that minimizers in $\hat{\Psi}_{k,m}$ would possibly diverge to "infinity" as the search ranges $K^k_{\tau_k}$ and $K^m_{\lambda_m}$ grow to the whole (noncompact) spaces, rendering the Delta method inapplicable under just Hadamard directional differentiability. Nonetheless, by restricting u to be in a compact set $\mathbb{D}_u \subset \mathbb{D}$, for example a class of smooth functions, we are able to remove the first lim sup; see Section 2.3.3.1.

The general construction of optimal plug-in estimators for infinite dimensional $\mathbb{D}$ is intrinsically complicated. When $\mathbb{D}$ is Euclidean – i.e. $\mathbb{D} = \mathbf{R}^m$ for some $m \in \mathbf{N}$, the computation greatly simplifies. Recall that by Corollary 2.3.1, the lower bound in this case is given by

$$\inf_{u \in \mathbf{R}^m} \sup_{c \in \mathbf{R}^m} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + c) - \phi'_{\theta_0}(c))] \ . \tag{2.36}$$

Comparing (2.36) with (2.30), it is clear that we can dispense with the computation burden of estimating H and θ'_0. Instead we now only have to estimate the directional derivative ϕ'_{θ_0} and the law of $\mathbb{G}_0$. Following the same idea as before, we therefore define

$$\hat{B}_\lambda(u) \equiv \max_{c \in K^m_\lambda} E[\ell(\hat{\phi}'_n(\hat{\mathbb{G}}^*_n + u + c) - \hat{\phi}'_n(c))|\{X_i\}] \ ,$$

$$\hat{\Psi}_{\tau,\lambda} \equiv \{u \in K^m_\tau : \hat{B}_\lambda(u) \le \min_{u' \in K^m_\tau} \hat{B}_\lambda(u') + \epsilon_n\} \ ,$$

where $\epsilon_n = o_p(1)$ as $n \to \infty$. As expected, if we pick $\hat{u}_{n,\tau,\lambda} \in \hat{\Psi}_{\tau,\lambda}$, then

$$\phi(\hat{\theta}_n + \frac{\hat{u}_{n,\tau,\lambda}}{r_n}) \tag{2.37}$$

will be an optimal plug-in estimator, as confirmed by the following theorem.

Theorem 2.3.3. *Let $\mathbb{D} = \mathbf{R}^m$ for some $m \in \mathbf{N}$ and $\Sigma_0 \equiv \langle \tilde{\theta}_0, \tilde{\theta}_0^\intercal \rangle$ be nonsingular. Suppose* *that Assumptions 2.2.1, 2.2.2, 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5(i)(iii), and 2.3.7 hold. Then*

$$\limsup_{\tau\to\infty}\limsup_{\lambda\to\infty}\sup_{I\subset_f H}\limsup_{n\to\infty}\sup_{h\in I} E_{n,h}\big[\ell\big(r_n\big(\phi(\hat{\theta}_n+\frac{\hat{u}_{n,\tau,\lambda}}{r_n})-\phi(\theta_n(h))\big)\big)\big]$$
$$\leq \inf_{u\in\mathbf{R}^m}\sup_{c\in\mathbf{R}^m} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+c)-\phi'_{\theta_0}(c))]\ . \qquad (2.38)$$

It is worth noting that the optimal plug-in estimators (2.34) and (2.37) depend, through the correction terms $\hat{u}_{n,m,k}$ and $\hat{u}_{n,\tau,\lambda}$ respectively, on the choice of the loss function ℓ, which in turn hinges on the nature of the problem at hand and practitioners' risk preference.

2.3.3.1 Smoothed Optimal Plug-in Estimators

By letting $k,m\to\infty$ and $\tau,\lambda\to\infty$ after n tends to infinity in the lower bounds, one essentially confines the minimizers $\hat{u}_{n,m,k}$ and $\hat{u}_{n,\tau,\lambda}$ to compact subsets. We may alternatively start with compact (possibly infinite dimensional) spaces and base our analysis therein.

In the literature of nonparametric and semi-(non)parametric methods, compactness can be obtained by attaching an appropriate norm different from the one that defines the space under consideration (Gallant and Nychka, 1987). For detailed discussions we refer the readers to Gallant and Nychka (1987), Newey and Powell (2003) and Santos (2012). We instead impose the following high level conditions.

Assumption 2.3.8. (i) $\mathbb{D}_u\subset\mathbb{D}$ is compact; (ii) $\{\mathbb{D}_k\}_{k=1}^{\infty}$ with $\mathbb{D}_k\subset\mathbb{D}_u$ for each $k\in\mathbf{N}$ is a sequence of compact sieves satisfying for any $u\in\mathbb{D}_u$, there exists $u_k\in\mathbb{D}_k$ such that $\|u_k-u\|_{\mathbb{D}}\to 0$ as $k\to\infty$.

Suppose that we are interested in the following restricted version of lower bound:

$$\min_{u\in\mathbb{D}_u}\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]\ , \qquad (2.39)$$

which is equal to the bound (2.30) if the infimum in the latter is attained in $\mathbb{D}_u$. In turn, (2.39) can be approximated by

$$\min_{u\in\mathbb{D}_u}\max_{c\in K^m_{\lambda_m}} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(g^m)^\intercal c)-\phi'_{\theta_0}(\theta'_0(g^m)^\intercal c))] \;, \tag{2.40}$$

where $\lambda_m\to\infty$ as $m\to\infty$. Replacing g^m, $\mathbb{G}_0$, θ'_0 and ϕ'_{θ_0} in (2.40) by their corresponding estimates $\{\hat{g}_m\}$, $\hat{\mathbb{G}}^*_n$, $\hat{\theta}'_n$ and $\hat{\phi}'_n$, and approximating $\mathbb{D}_u$ by the sequence of compact sieves $\{\mathbb{D}_k\}$, we may in turn estimate the bound (2.40) by considering

$$\hat{B}_m(u)\equiv\max_{c\in K^m_{\lambda_m}} E[\ell(\hat{\phi}'_n(\hat{\mathbb{G}}^*_n+u+\hat{\theta}'_n(\hat{g}^m)^\intercal c)-\hat{\phi}'_n(\hat{\theta}'_n(\hat{g}^m)^\intercal c))|\{X_i\}] \;,$$

$$\hat{\Psi}_m\equiv\{u\in\mathbb{D}_{k_n}:\hat{B}_m(u)\le\min_{u'\in\mathbb{D}_{k_n}}\hat{B}_m(u')+\epsilon_n\} \;,$$

where $\epsilon_n=o_p(1)$ as $n\to\infty$. Notice that the set $\hat{\Psi}_m$ of minimizers of $\hat{B}_m(u)$ is obtained on the approximating space $\mathbb{D}_{k_n}$, though we have suppressed the dependence of $\hat{\Psi}_m$ on n for notational simplicity.

Now take arbitrary $\hat{u}_{n,m}\in\hat{\Psi}_m$ and define the plug-in estimator

$$\phi(\hat{\theta}_n+\frac{\hat{u}_{n,m}}{r_n}) \;. \tag{2.41}$$

Optimality of (2.41) in the sense of local asymptotic minimaxity is confirmed as follows.

Theorem 2.3.4. *Suppose that Assumptions 2.2.1, 2.2.2, 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5, 2.3.6(i)(ii), 2.3.7, and 2.3.8 hold. Let $\hat{u}_{n,m}\in\hat{\Psi}_m$. If $\lambda_m,k_n\to\infty$ as $m,n\to\infty$ respectively, then*

$$\begin{aligned}\limsup_{m\to\infty}\sup_{I\subset_f H}\limsup_{n\to\infty}\sup_{h\in I} E_{n,h}\big[\ell\big(r_n\big(\phi(\hat{\theta}_n+\frac{\hat{u}_{n,m}}{r_n})-\phi(\theta_n(h))\big)\big)\big]\\ \le\inf_{u\in\mathbb{D}_u}\sup_{h\in H}E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))] \;.\end{aligned} \tag{2.42}$$

We note that similar as the sieve approximation for $\mathbb{D}$, one may also consider construction based on a general sequence of compact sieves of H. While one might have different tastes on the choice of compact sieves for $\mathbb{D}$ – for instance, one might choose different degrees of smoothness which in turn directly affects the smoothness of the correction term $\hat{u}_n$, approximation for H is purely for computational purposes and has more indirect effect on $\hat{u}_n$. We thus skip the general approximation for H here.

2.3.3.2 Examples Revisited

We now turn to Examples 2.2.1-2.2.4. For the sake of brevity, we omit the bootstrap procedure, and instead focus on verifying Assumptions 2.3.5(ii)(iii), 2.3.6, and 2.3.7. For Examples 2.2.1 and 2.2.2, there is no need to estimate H and θ_0'; see Corollary 2.3.1.

Example 2.2.1 (Continued). The sample mean $\overline{X}_n$ serves as an efficient estimator of θ. Denote $\hat{j}^* = \arg\max_{j\in\{1,2\}} \bar{X}^{(j)}$ and pick $t_n \uparrow \infty$ satisfying $t_n/\sqrt{n} \downarrow 0$. Define

$$\hat{\phi}_n'(z) = \begin{cases} z^{(\hat{j}^*)} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| > t_n \\ \max\{z^{(1)}, z^{(2)}\} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| \le t_n \end{cases} . \tag{2.43}$$

Then it is straightforward to verify that $\hat{\phi}_n'$ is Lipschitz continuous and pointwise consistent. ■

Example 2.2.2 (Continued). The efficient estimation of θ_0 in this example can be conducted in the conditional moment restriction framework (Newey, 1993). Then we may estimate ϕ_{θ_0}' by

$$\hat{\phi}_n'(z) = \psi_{\hat{\theta}_n}'(z)1\{\psi(\hat{\theta}_n) > t_n\} + \max\{\psi_{\hat{\theta}_n}'(z), 0\}1\{|\psi(\hat{\theta}_n)| \le t_n\} ,$$

where $\hat{\theta}_n$ is an efficient estimator of θ_0, and t_n is a sequence specified as in Example 2.2.1.■

Example 2.2.3 (Continued). Let $\hat{F}_1$ and $\hat{F}_2$ be the empirical CDFs of F_1 and F_2 respectively. It is known that empirical CDFs $\hat{F}_1$ and $\hat{F}_2$ are efficient in estimating F_1 and F_2 respectively (van der Vaart and Wellner, 1996), and hence $(\hat{F}_1, \hat{F}_2)$ is efficient in estimating (F_1, F_2) in the product space $\ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ (van der Vaart, 1991b). The form of the derivative ϕ_{θ_0}' as in (2.22) suggests a natural estimator for it. Define $\hat{B}_1 \equiv \{x \in \mathbf{R} : \hat{F}_{1n}(x) - \hat{F}_{2n}(x) > t_n\}$, $\hat{B}_2 \equiv \{x \in \mathbf{R} : \hat{F}_{2n}(x) - \hat{F}_{1n}(x) > t_n\}$, and $\hat{B}_0 \equiv \{x \in \mathbf{R} : |\hat{F}_{1n}(x) -$

$\hat{F}_{2n}(x)| \leq t_n\}$ where t_n is again as in Example 2.2.1. Let $\hat{\phi}_n' : \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R}) \to \ell^\infty(\mathbf{R})$ be defined by

$$\hat{\phi}_n'(z) = z^{(1)} 1_{\hat{B}_1} + z^{(2)} 1_{\hat{B}_2} + \max\{z^{(1)}, z^{(2)}\} 1_{\hat{B}_0} .$$

Then $\hat{\phi}_n'$ is Lipschitz continuous and pointwise consistent.

In this example, we have to estimate θ_0' and a basis $\{g_j\}$ of $L_0^2(P)$, as well as the derivative ϕ_{θ_0}' and the law of $\mathbb{G}_0$. By Section 3.11.1 in van der Vaart and Wellner (1996), for each $h \equiv (h_1, h_2) \in H \times H$ with H being the set of bounded measurable functions on $\mathbf{R}$,

$$\theta_0'(h)(v) = \Big(\int_{-\infty}^{v} h_1(t)\, P_1(dt), \int_{-\infty}^{v} h_2(t)\, P_2(dt) \Big) .$$

Thus, we may take the following estimator of θ_0':

$$\hat{\theta}_n'(h)(v) = \Big(\int_{-\infty}^{v} h_1(t)\, \mathbb{P}_{1n}(dt), \int_{-\infty}^{v} h_2(t)\, \mathbb{P}_{2n}(dt) \Big) .$$

As to Assumption 2.3.6(ii), if $\{g_m^{(i)}\}$ is complete in $L^2(P_i)$ with $i = 1, 2$, then we may take

$$g_1^{(1)}(v) - \frac{1}{n}\sum_{i=1}^{n} g_1^{(1)}(B_{1i}) \ , \ g_2^{(1)}(v) - \frac{1}{n}\sum_{i=1}^{n} g_2^{(1)}(B_{1i}) \ , \ldots \ ,$$
$$g_1^{(2)}(v) - \frac{1}{n}\sum_{i=1}^{n} g_1^{(2)}(B_{2i}) \ , \ g_2^{(2)}(v) - \frac{1}{n}\sum_{i=1}^{n} g_2^{(2)}(B_{2i}) \ , \ldots \ ,$$

where $\{B_{1i}\}_{i=1}^n$ and $\{B_{2i}\}_{i=1}^n$ are bids from auctions 1 and 2 respectively; see Lemma 2.6.9.[16] In this example, since functions in $\ell^\infty(\mathbf{R})$ can be rather irregular, one might want to follow the compact version of construction, for instance, let $\mathbb{D}_u$ be a class of smooth $\mathbf{R}^2$-valued functions. For concrete constructions, see Gallant and Nychka (1987), Newey and Powell (2003), and Santos (2012). ■

Example 2.2.4 (Continued). Since $\beta(\cdot) : \mathcal{T} \to \mathbf{R}$ can be efficiently estimated by the quantile regression process $\hat{\beta}_n(\cdot)$, we thus conclude that $\hat{\theta}_n \equiv c'\hat{\beta}_n(\cdot)$ is efficient in estimating θ_0 (van der Vaart, 1991b). As to estimation of the derivative ϕ_{θ_0}', we follow the approach

[16] For example, if P_i satisfies $\int e^{M|v|}\, P_i(dv) < \infty$ for all $M \in (0, \infty)$, then $\{1, v, v^2, \ldots\}$ is complete in $L^2(P_i)$.

pursued by Hong and Li (2014) and propose the following estimator:

$$\hat{\phi}'_n(z) = t_n^{-1}\{\Pi_\Lambda(\hat{\theta}_n + t_n z) - \Pi_\Lambda(\hat{\theta}_n)\} ,$$

where t_n satisfies $t_n \to 0$ and $t_n\sqrt{n} \to \infty$ as $n \to \infty$.[17] The derivative θ'_0 can be estimated as follows:

$$\begin{aligned}\hat{\theta}'_n(h) \equiv -\hat{J}(\tau)^{-1}\int c'z1\{y \le z'\hat{\beta}(\tau)\}h_1(y,z)\mathbb{P}_n(dy,dz)\\ - \hat{J}(\tau)^{-1}\int c'z(1\{y \le z'\hat{\beta}(\tau)\} - \tau)h_2(z)\mathbb{P}_n(dy,dz) ,\end{aligned}$$

where $\hat{J}(\tau)$ is constructed as in Angrist et al. (2006):

$$\hat{J}(\tau) \equiv \frac{1}{2n\kappa_n}\sum_{i=1}^{n} 1\{|Y_i - Z'_i\hat{\beta}(\tau)| \le \kappa_n\}Z_iZ'_i ,$$

where κ_n satisfies $\kappa_n \to 0$ and $\kappa_n^2 n \to \infty$. A complete sequence in H_1 can be estimated similarly as in Example 2.2.3. As to H_2, if $\{g_j(y,z)\}$ is complete in $L^2(\mathscr{Y} \times \mathscr{Z})$, then we may take

$$g_1(y,z) - \frac{1}{n}\sum_{i=1}^{n} g_1(Y_i,z) \ , \ g_2(y,z) - \frac{1}{n}\sum_{i=1}^{n} g_2(Y_i,z) \ , \ldots \ .$$

A complete sequence $\{\psi_k\}$ in $L^2(\mathcal{T})$ can be a sequence of polynomials, while the compact space $\mathbb{D}_u$ can be chosen to be a class of smooth functions in $L^2(\mathcal{T})$ as in Example 2.2.3. ■

2.4 Empirical Application

In this section, we apply the theory developed in previous sections to the estimation of the effect of Vietnam veteran status on the quantiles of civilian earnings (Angrist, 1990). Since certain types of men are more likely to service in the military, making the veteran status endogenous, a conventional quantile regression method is inappropriate to recover

[17] Song (2014, 2015) essentially took the same approach.

the casual relationship. Following Angrist (1990), we employ the Vietnam draft lottery eligibility indicator as an instrument for veteran status. In particular, we apply the instrumental quantile regression framework developed by Chernozhukov and Hansen (2005, 2006) to the Current Population Survey data set as in Chernozhukov et al. (2010), which consists of four variables: annual labor real earnings, weakly real wage, veteran status indicator with value 1 for veterans, and Vietnam draft lottery eligibility indicator as an instrument with value 1 for eligible men. As in Chernozhukov et al. (2010), we focus on the annual labor earnings throughout.

Let Y denote the annual labor real earnings, D the veteran status, and Z the Vietnam draft lottery eligibility. Under instrument independence and rank similarity, Chernozhukov and Hansen (2005) showed that the quantile regression coefficients $\beta(\tau)$ for veterans can be identified by the following conditional moment restriction:

$$E[(\tau - 1\{Y \leq \beta(\tau)D\})|Z] = 0 \text{ a.s.}, \forall \tau \in (0,1) , \tag{2.44}$$

much like the counterpart in mean regression models. Chernozhukov and Hansen (2006) developed the instrumental variable quantile regression based on restriction (2.44), which can be viewed as a quantile regression analog of two stage least squares.

Unfortunately, since $\beta(\tau)$ is estimated pointwise, there is in general no guarantee that the quantile function $\hat{\beta}(\cdot)$ is monotonically increasing. To circumvent the nonmonotonicity when estimating the structural quantile functions of earnings, we therefore employ the metric projection operator introduced in Example 2.2.4. In estimating the correction terms, we take polynomials as basis functions for $\mathbb{D} \equiv L^2(\mathcal{T})$ and H, and set $m = 4, k = 3$. The quantile index set $\mathcal{T}$ is taken to be the grid on $[0.25, 0.75]$ with increment 0.001, while the number of bootstrap repetitions is set to be two hundred. As for the estimation of the Hadamard directional derivative, we follow the same approach as in Example 2.2.4 and set $t_n = n^{-1/3}$. The correction terms are estimated relative to the L^2 loss function.

In Figures 2.1 and 2.2 we show the structural quantile functions of earnings for vet-

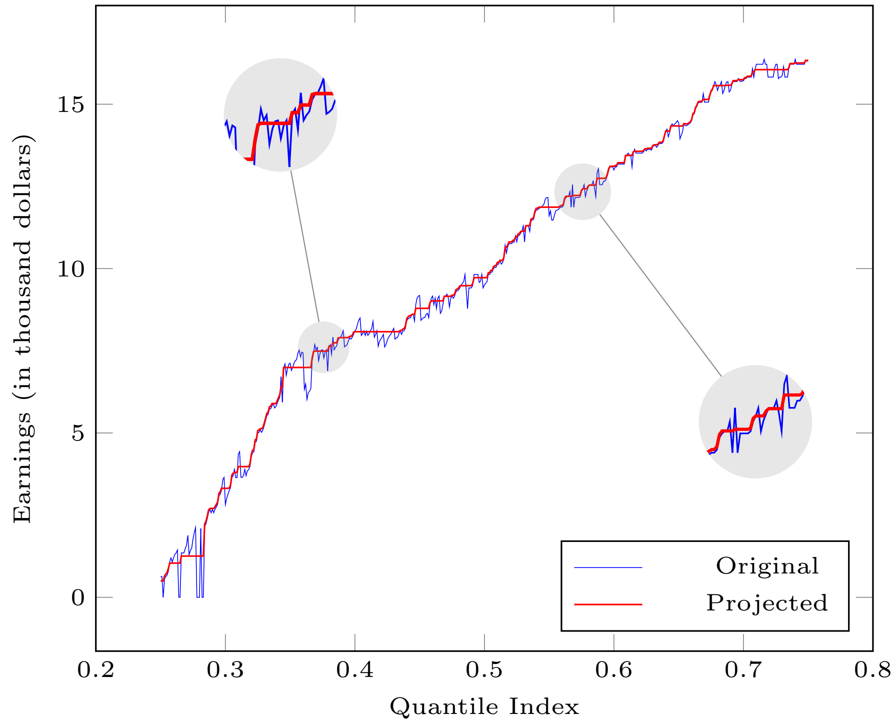

Figure 2.1: Structural Quantile Functions of Earnings for Veterans

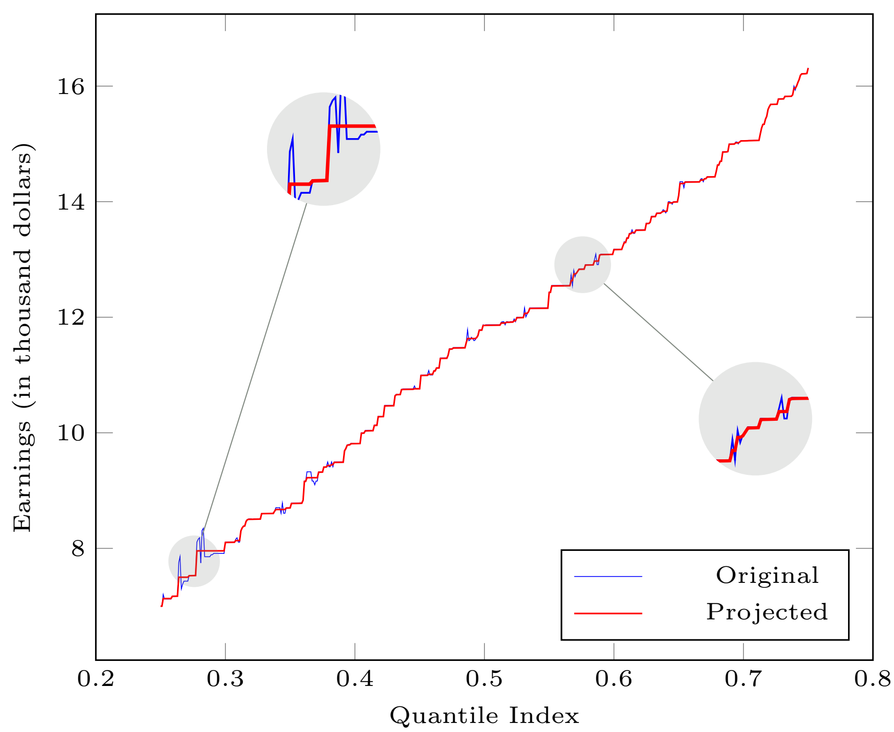

Figure 2.2: Structural Quantile Functions of Earnings for Non-Veterans

erans and non-veterans respectively, as well as their optimal projected counterparts with correction terms. In both figures, the original quantile functions exhibit obvious nonmonotonicity at certain regions, especially for veterans. The projected counterparts are by construction monotone and optimal in terms of local asymptotic minimaxity. We note significant differences between original quantile curve and the optimal projected one for veterans. For example, the median of the annual earnings for veterans is 9,819 dollars according to the original estimate and 9,929 dollars according to the projected estimate. The maximal difference of 1,767 dollars occurs at the 0.725 quantile. In contrast, we find less difference between the original structural quantile function and the optimal projected counterpart for the non-veterans, with the maximal gap being 403 dollars at the 0.725 quantile.

2.5 Conclusion

In this paper, we have derived the local asymptotic minimax lower bound for a class of plug-in estimators of directionally differentiable parameters, which arise in a large class of econometric problems. The employment of minimaxity criterion, although perhaps not fully necessary, seems to the most suitable one for our purposes. The derived lower bound is intrinsically complicated. Nonetheless, we have been able to present a general construction procedure to show attainability of the lower bound.

2.6 Appendix

2.6.1 Proofs of Main Results

Proof of Theorem 2.3.1: For each finite subset $I \subset H$, we have

$$\begin{aligned}\liminf_{n\to\infty} \sup_{h\in I} & E_{n,h}[\ell(r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\})] \\ & \geq \sup_{h\in I} \liminf_{n\to\infty} E_{n,h}[\ell(r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\})] \,. \qquad (2.45)\end{aligned}$$

By Assumption 2.3.3, ℓ is continuous and positive. In turn, Lemma 2.6.1 allows us to invoke the portmanteau theorem to conclude that

$$\liminf_{n\to\infty} E_{n,h}[\ell(r_n\{\phi(\hat{\theta}_n)-\phi(\theta_n(h))\})] \\ \geq E[\ell(\phi'_{\theta_0}(\mathbb{G}+\theta'_0(h)+\Delta)-\phi'_{\theta_0}(\theta'_0(h)+\Delta))] \ . \tag{2.46}$$

Combining results (2.45) and (2.46) we thus have

$$\liminf_{n\to\infty}\sup_{h\in I} E_{n,h}[\ell(r_n\{\phi(\hat{\theta}_n)-\phi(\theta_n(h))\})] \\ \geq \sup_{h\in I} E[\ell(\phi'_{\theta_0}(\mathbb{G}+\theta'_0(h)+\Delta)-\phi'_{\theta_0}(\theta'_0(h)+\Delta))] \ . \tag{2.47}$$

Taking supremum on both sides in (2.47) over all finite $I \subset H$ yields that

$$\begin{aligned}\sup_{I\subset_f H}\liminf_{n\to\infty}\sup_{h\in I} & E_{n,h}[\ell(r_n\{\phi(\hat{\theta}_n)-\phi(\theta_n(h))\})] \\ & \geq \sup_{I\subset_f H}\sup_{h\in I} E[\ell(\phi'_{\theta_0}(\mathbb{G}+\theta'_0(h)+\Delta)-\phi'_{\theta_0}(\theta'_0(h)+\Delta))] \\ & = \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}+\theta'_0(h)+\Delta)-\phi'_{\theta_0}(\theta'_0(h)+\Delta))] \\ & = \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))] \ ,\end{aligned} \tag{2.48}$$

where the last equality is due to the fact that $\Delta \in \theta'_0(H)$ by Assumption 2.3.1 and the fact that H is linear by Assumption 2.2.1(i).

In view of (2.48) and the desired lower bound in (2.27), it suffices to show that,

$$\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))] \\ \geq \inf_{u\in\mathbb{D}}\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))] \ . \tag{2.49}$$

Towards this end, we follow the idea of Song (2014) but, instead of employing the purification theorem initially developed by Dvoretzky et al. (1950, 1951), we appeal to a more generalized version in Feinberg and Piunovskiy (2006) and hence are able to simplify the proof that would be otherwise involved.

Since H is separable by Assumption 2.2.1(i), we may pick a sequence $\{h_j\}_{j=1}^{\infty}$ that is dense in H. By positivity and continuity of ℓ implied by Assumption 2.3.3 and continuity of θ_0' and ϕ_{θ_0}' implied by Assumptions 2.2.2 and 2.3.4, we may conclude by Fatou's lemma that $E[\ell(\phi_{\theta_0}'(\mathbb{G} + \theta_0'(h)) - \phi_{\theta_0}'(\theta_0'(h)))]$ is lower semicontinuous in h. It follows by Lemma 2.6.5 that

$$\sup_{h \in H} E[\ell(\phi_{\theta_0}'(\mathbb{G} + \theta_0'(h)) - \phi_{\theta_0}'(\theta_0'(h)))] = \sup_{j \in \mathbf{N}} E[\ell(\phi_{\theta_0}'(\mathbb{G} + \theta_0'(h_j)) - \phi_{\theta_0}'(\theta_0'(h_j)))] \; . \quad (2.50)$$

Fix $J \in \mathbf{N}$. For $j = 1, \ldots, J$, write $\rho(z, u) = (\rho_1(z, u), \ldots, \rho_J(z, u))^\intercal$ where

$$\rho_j(z, u) = E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u + \theta_0'(h_j)) - \phi_{\theta_0}'(\theta_0'(h_j)))] z \; .$$

By Assumptions 2.2.1, 2.2.2 and 2.2.3, Theorem 2.2.1 applies so that we may write $\mathbb{G} \overset{d}{=} \mathbb{G}_0 + \mathbb{U}$, where $\mathbb{G}_0$ is the efficient Gaussian component and $\mathbb{U}$ is the noise term independent of $\mathbb{G}_0$. Denote the distribution of $\mathbb{U}$ by Q. For fixed $\lambda > 1$, let Z follow the uniform distribution ν_λ supported on $[1, \lambda]$. By Theorem 1 in Feinberg and Piunovskiy (2006), there is a measurable map $u^* : [1, \lambda] \to \mathbb{D}$ such that

$$\int_{\mathbf{R}} \int_{\mathbb{D}} \rho(z, u) \, Q(du) \nu_\lambda(dz) = \int_{\mathbf{R}} \rho(z, u^*(z)) \, \nu_\lambda(dz) \; ,$$

which in turn implies that, for all $j = 1, \ldots, J$,

$$\begin{aligned} \frac{1+\lambda}{2} \int_{\mathbb{D}} & E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u + \theta_0'(h_j)) - \phi_{\theta_0}'(\theta_0'(h_j)))] \, Q(du) \\ & = \int_{\mathbf{R}} E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u^*(z) + \theta_0'(h_j)) - \phi_{\theta_0}'(\theta_0'(h_j)))] z \, \nu_\lambda(dz) \\ & \geq \int_1^\lambda E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u^*(z) + \theta_0'(h_j)) - \phi_{\theta_0}'(\theta_0'(h_j)))] \, \nu_\lambda(dz) \; , \end{aligned} \quad (2.51)$$

where the inequality exploits the facts that $z \geq 1$ almost everywhere and that $\ell \geq 0$. By change of variable applied to the right hand side of (2.51), we have for all $j = 1, \ldots, J$,

$$\frac{1+\lambda}{2}\int_{\mathbb{D}} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]\, Q(du)$$
$$\geq \int_0^1 E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u^*((\lambda-1)y+1) + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]\, dy\ . \tag{2.52}$$

It follows that

$$\frac{1+\lambda}{2}\max_{j=1,\ldots,J} E[\ell(\phi'_{\theta_0}(\mathbb{G} + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]$$
$$\geq \inf_{\lambda>1}\max_{j=1,\ldots,J}\int_0^1 E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u^*((\lambda-1)y+1) + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]\, dy$$
$$\geq \inf_{u\in\mathcal{R}(u^*)}\max_{j=1,\ldots,J}\int_0^1 E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]\, dy$$
$$\geq \inf_{u\in\mathbb{D}}\max_{j=1,\ldots,J} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]\ , \tag{2.53}$$

where $\mathcal{R}(u^*)$ denotes the range of u^*.

Letting $\lambda \downarrow 1$ and then $J \to \infty$ in (2.53) yields

$$\sup_{j\in\mathbf{N}} E[\ell(\phi'_{\theta_0}(\mathbb{G} + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]$$
$$\geq \inf_{u\in\mathbb{D}}\sup_{j\in\mathbf{N}} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h_j)) - \phi'_{\theta_0}(\theta'_0(h_j)))]\ . \tag{2.54}$$

Combining (2.50), (2.54), and the fact that the expectation on the right hand side is also lower semicontinuous in h by Assumptions 2.2.2, 2.3.2 and 2.3.3, we thus conclude that

$$\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G} + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))]$$
$$\geq \inf_{u\in\mathbb{D}}\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))]\ , \tag{2.55}$$

proving (2.49) and hence the Theorem. ■

Lemma 2.6.1. *Let* $(\mathscr{X}_n, \mathcal{A}_n, \{P_{n,h} : h \in H\})$ *be a sequence of statistical experiments, and* $\hat{\theta}_n$ *be an estimator for the parameter* $\theta : \{P_{n,h}\} \to \mathbb{D}$*. If Assumptions 2.2.2, 2.2.3, 2.3.1* *and 2.3.2 hold, then*

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\} \overset{L_{n,h}}{\to} \phi'_{\theta_0}(\mathbb{G} + \theta'_0(h) + \Delta) - \phi'_{\theta_0}(\theta'_0(h) + \Delta) \tag{2.56}$$

for every $h \in H$.

PROOF: Rewrite

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_n(h))\} = r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} - r_n\{\phi(\theta_n(h)) - \phi(\theta_0)\} \ . \tag{2.57}$$

By Assumptions 2.2.3, 2.2.2, and 2.3.1, we have

$$\begin{aligned} r_n(\hat{\theta}_n - \theta_0) &= r_n\{\hat{\theta}_n - \theta_n(h)\} + r_n\{\theta_n(h) - \theta_n(0)\} + r_n\{\theta_n(0) - \theta_0\} \\ &\overset{L_{n,h}}{\to} \mathbb{G} + \theta'_0(h) + \Delta \ , \end{aligned}$$

for every $h \in H$. By Assumption 2.3.2, ϕ is Hadamard directionally differentiable at θ_0 tangentially to $\mathbb{D}$, and hence by the Delta method (Fang and Santos, 2014, Theorem 2.1) we may conclude that

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \overset{L_{n,h}}{\to} \phi'_{\theta_0}(\mathbb{G} + \theta'_0(h) + \Delta) \ . \tag{2.58}$$

On the other hand, Assumptions 2.2.2, and 2.3.1 imply that for all $h \in H$,

$$r_n\{\theta_n(h) - \theta_0\} = r_n\{\theta_n(h) - \theta_n(0)\} + r_n\{\theta_n(0) - \theta_0\} \to \theta'_0(h) + \Delta \ ,$$

whence by Assumption 2.3.2,

$$r_n\{\phi(\theta_n(h)) - \phi(\theta_0)\} \to \phi'_{\theta_0}(\theta'_0(h) + \Delta) \ . \tag{2.59}$$

The Lemma then follows from displays (2.57), (2.58) and (2.59). ■

Proof of Lemma 2.3.1: By Assumption 2.3.4, we have

$$\|\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h))\|_{\mathbb{E}} \leq C_{\phi'}\|\mathbb{G}_0 + u\|_{\mathbb{D}} ,$$

and hence by ρ being nondecreasing

$$\begin{aligned}
&\inf_{u\in\mathbb{D}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))] \\
&\qquad\qquad = \inf_{u\in\mathbb{D}} \sup_{h\in H} E[\rho(\|\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h))\|_{\mathbb{E}})] \\
&\qquad\qquad \leq \inf_{u\in\mathbb{D}} E[\rho(C_{\phi'}\|\mathbb{G}_0 + u\|_{\mathbb{D}})] \ . \qquad (2.60)
\end{aligned}$$

For each $c \geq 0$, the set $A_c \equiv \{y : \rho(C_{\phi'}\|y\|_{\mathbb{D}}) \leq c\}$ is clearly symmetric. It is also closed since if $\{y_n\} \subset A_c$ and $y_n \to y$, then ρ being lower semicontinuous implies that

$$\rho(C_{\phi'}\|y\|_{\mathbb{D}}) \leq \liminf_{n\to\infty} \rho(C_{\phi'}\|y_n\|_{\mathbb{D}}) \leq c \ .$$

Finally, A_c is convex since if $y_1, y_2 \in A_c$, then for any $\lambda \in (0,1)$

$$\begin{aligned}
\rho(C_{\phi'}\|\lambda y_1 + (1-\lambda)y_2\|_{\mathbb{D}}) \leq& \rho(\lambda C_{\phi'}\|y_1\|_{\mathbb{D}} + (1-\lambda)C_{\phi'}\|y_2\|_{\mathbb{D}}) \\
&\leq \rho(\max\{C_{\phi'}\|y_1\|_{\mathbb{D}}, C_{\phi'}\|y_2\|_{\mathbb{D}}\}) \leq c \ .
\end{aligned}$$

Therefore $\rho(C_{\phi'}\|\cdot\|_{\mathbb{D}})$ is subconvex. We thus conclude from result (2.60) and Anderson's lemma (van der Vaart and Wellner, 1996) that

$$\begin{aligned}
&\inf_{u\in\mathbb{D}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))] \\
&\qquad\qquad \leq \inf_{u\in\mathbb{D}} E[\rho(C_{\phi'}\|\mathbb{G}_0 + u\|_{\mathbb{D}})] = E[\rho(C_{\phi'}\|\mathbb{G}_0\|_{\mathbb{D}})] < \infty \ .
\end{aligned}$$

This establishes the Lemma. ■

Proof of Corollary 2.3.1: By Theorem 2.3.1, we know that the lower bound is given by

$$\inf_{u\in\mathbb{D}}\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]\ . \tag{2.61}$$

By Assumptions 2.2.2, 2.3.4 and 2.3.3, we may conclude by Fatou's lemma that the expectation in (2.61) is lower semicontinuous in h. It follows by Lemma 2.6.5 that

$$\inf_{u\in\mathbb{D}}\sup_{c\in\overline{\theta'_0(H)}} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+c)-\phi'_{\theta_0}(c))]\ . \tag{2.62}$$

Since $\mathbb{G}_0$ is Gaussian in $\mathbb{D}\equiv\mathbf{R}^m$ with nonsingular covariance Σ_0, by Theorem 2.2.1 it must be the case that $\overline{\theta'_0(H)}=\mathbf{R}^m$. The Corollary then follows. ■

Proof of Theorem 2.3.2: Suppose first that the loss function ℓ is bounded by $M>0$. Fix $\epsilon>0$. Then there is some $u_\epsilon\in\mathbb{D}$ such that

$$\begin{aligned}&\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u_\epsilon+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]\\&\qquad\le\inf_{u\in\mathbb{D}}\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]+\frac{\epsilon}{4}\ .\end{aligned} \tag{2.63}$$

By Assumptions 2.3.3 and 2.3.4, $\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]$ is (Lipschitz) continuous in u. Thus, there is some $\delta>0$ such that

$$\begin{aligned}&\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]\\&\qquad\le\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u_\epsilon+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]+\frac{\epsilon}{4}\ ,\end{aligned}$$

whenever $\|u-u_\epsilon\|_{\mathbb{D}}<\delta$. By Assumption 2.3.6(iii) and the fact that $\tau_k\to\infty$ as $k\to\infty$, there is some $v_k\in K^k_{\tau_k}$ such that $\|u_k-u_\epsilon\|_{\mathbb{D}}<\delta$ with $u_k\equiv(\psi^k)^\intercal v_k$ for all k large enough, which in turn means that

$$\begin{aligned}&\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u_k+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]\\&\qquad\le\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u_\epsilon+\theta'_0(h))-\phi'_{\theta_0}(\theta'_0(h)))]+\frac{\epsilon}{4}\ .\end{aligned} \tag{2.64}$$

Combining results (2.63) and (2.64) we thus have for all k large enough,

$$\begin{aligned}
&\inf_{v\in K^k_{\tau_k}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(h)) - \phi'_{\theta_0}(\theta_0'(h)))] \\
&\qquad\qquad \le \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u_k + \theta_0'(h)) - \phi'_{\theta_0}(\theta_0'(h)))] \\
&\qquad\qquad \le \inf_{u\in\mathbb{D}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta_0'(h)) - \phi'_{\theta_0}(\theta_0'(h)))] + \frac{\epsilon}{2} \; . \qquad (2.65)
\end{aligned}$$

Next, for notational simplicity, define

$$B_m(v) \equiv \max_{c\in K^m_{\lambda_m}} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(g^m)^\intercal c) - \phi'_{\theta_0}(\theta_0'(g^m)^\intercal c))] \; ,$$

$$B(v) \equiv \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(h)) - \phi'_{\theta_0}(\theta_0'(h)))] \; , \; \Psi_{k,m} \equiv \arg\min_{v\in K^k_{\tau_k}} B_m(v) \; .$$

Fix k large enough so that (2.65) holds. By Assumptions 2.3.3 and 2.3.4, it is clear that both $B(v)$ and $B_m(v)$ for each $m \in \mathbf{N}$ are continuous functions on $K^k_{\tau_k}$. Moreover, $B_m(v)$ increasingly converges to $B(v)$ as $m \to \infty$ for each $v \in K^k_{\tau_k}$ with additional Assumption 2.3.6(i). To see this, fix $v \in K^k_{\tau_k}$ and pick $h_\epsilon \in H$ such that

$$B(v) - \epsilon \le E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(h_\epsilon)) - \phi'_{\theta_0}(\theta_0'(h_\epsilon)))] \le B(v) \; . \qquad (2.66)$$

By Assumptions 2.2.2, 2.3.3 and 2.3.4, $E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(h)) - \phi'_{\theta_0}(\theta_0'(h)))]$ is (Lipschitz) continuous in h, and hence by Assumption 2.3.6(i) and the fact that $\lambda_m \to \infty$ as $m \to \infty$, we have for all m sufficiently large

$$E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(h_\epsilon)) - \phi'_{\theta_0}(\theta_0'(h_\epsilon)))] \le B_m(v) + \epsilon \; . \qquad (2.67)$$

Combining previous two displays we obtain $B(v) - 2\epsilon \le B_m(v) \le B(v)$ for all m sufficiently large. This shows that $B_m(v)$ increasingly converges to $B(v)$ for each $v \in K^k_{\tau_k}$. It follows by Dini's theorem (Aliprantis and Border, 2006, Theorem 2.66) that $B_m \to B$ uniformly on $K^k_{\tau_k}$. We thus conclude that there is an m_0 such that for all $m \ge m_0$, $B(v) \le B_m(v) + \epsilon/2$

or equivalently

$$\sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+(\psi^k)^\intercal v+\theta_0'(h))-\phi'_{\theta_0}(\theta_0'(h)))]$$
$$\le \sup_{c\in K^m_{\lambda_m}} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+(\psi^k)^\intercal v+\theta_0'(g^m)^\intercal c)-\phi'_{\theta_0}(\theta_0'(g^m)^\intercal c))]+\frac{\epsilon}{2}\ , \tag{2.68}$$

for all $v\in K^k_{\tau_k}$.

Next, fix an arbitrary subsequence $\{n_\ell\}$. For fixed $m,k\in\mathbf{N}$, $B_m(\cdot)$ is continuous on $K^k_{\tau_k}$ by Assumptions 2.3.3 and 2.3.4, which, together with compactness of $K^k_{\tau_k}$, implies that $\Psi_{k,m}$ is nonempty and compact by Theorem 2.43 in Aliprantis and Border (2006). Combination of Lemma 2.6.2 and Lemma 2.6.6 then implies that there exist a further subsequence $\{n_{\ell_j}\}$ and some $v^*_{k,m}\in\Psi_{k,m}$ such that

$$\hat v_{n_{\ell_j},k,m}\xrightarrow{p} v^*_{k,m}\ , \tag{2.69}$$

as $j\to\infty$ under $\{P_{n,h}\}$ with $h\in H$, for each $k,m\in\mathbf{N}$. Result (2.69), together with Assumptions 2.3.7, 2.2.2, 2.3.1 and 2.3.2, allows us to invoke Slutsky's theorem and the Delta method to conclude that

$$r_{n_{\ell_j}}\{\phi(\hat\theta^*_{n_{\ell_j}}+\frac{\hat u_{n_{\ell_j},k,m}}{r_{n_{\ell_j}}})-\phi(\theta_{n_{\ell_j}}(h))\}\xrightarrow{L_{n_{\ell_j},h}}\phi'_{\theta_0}(\mathbb{G}_0+u^*_{k,m}+\Delta+\theta_0'(h))-\phi'_{\theta_0}(\Delta+\theta_0'(h))$$

for each $h\in H$, where $u^*_{k,m}\equiv(\psi^k)^\intercal v^*_{k,m}$. Since ℓ is bounded and continuous, it follows that for all m sufficiently large and all $k\in\mathbf{N}$,

$$\sup_{I\subset_f H}\limsup_{j\to\infty}\sup_{h\in I} E_{n_{\ell_j},h}[\ell(r_{n_{\ell_j}}\{\phi(\hat\theta_{n_{\ell_j}}+\frac{\hat u_{n_{\ell_j},k,m}}{r_{n_{\ell_j}}})-\phi(\theta_{n_{\ell_j}}(h))\})]$$
$$=\sup_{h\in H}E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+(\psi^k)^\intercal v^*_{k,m}+\theta_0'(h))-\phi'_{\theta_0}(\theta_0'(h)))]$$
$$\le\max_{v\in\Psi_{k,m}}B(v)\le\max_{v\in\Psi_{k,m}}B_m(v)+\frac{\epsilon}{2}=\inf_{v\in K^k_{\tau_k}}B_m(v)+\frac{\epsilon}{2}\ ,$$

where the first inequality is due to $v^*_{k,m}\in\Psi_{k,m}$, the second inequality is by result (2.68), while the last equality is by definition of $\Psi_{k,m}$. This implies that for all k large enough,

$$\begin{aligned}
&\limsup_{m\to\infty} \sup_{I\subset_f H} \limsup_{j\to\infty} \sup_{h\in I} E_{n_{\ell_j},h}[\ell(r_{n_{\ell_j}}\{\phi(\hat{\theta}_{n_{\ell_j}} + \frac{\hat{v}_{n_{\ell_j},k,m}}{r_{n_{\ell_j}}}) - \phi(\theta_{n_{\ell_j}}(h))\})] \\
&\qquad \le \limsup_{m\to\infty} \inf_{v\in K^k_{\tau_k}} B_m(v) + \frac{\epsilon}{2} = \inf_{v\in K^k_{\tau_k}} B(v) + \frac{\epsilon}{2} \\
&\qquad \le \inf_{u\in\mathbb{D}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))] + \epsilon \ , \qquad (2.70)
\end{aligned}$$

where the equality follows from the fact that $B_m \to B$ uniformly on $K^k_{\tau_k}$, and the last inequality holds for all k sufficiently large due to (2.65). We thus have

$$\begin{aligned}
&\limsup_{k\to\infty} \limsup_{m\to\infty} \sup_{I\subset_f H} \limsup_{j\to\infty} \sup_{h\in I} E_{n_{\ell_j},h}[\ell(r_{n_{\ell_j}}\{\phi(\hat{\theta}_{n_{\ell_j}} + \frac{\hat{v}_{n_{\ell_j},k,m}}{r_{n_{\ell_j}}}) - \phi(\theta_{n_{\ell_j}}(h))\})] \\
&\qquad \le \inf_{u\in\mathbb{D}} \sup_{h\in H} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + \theta'_0(h)) - \phi'_{\theta_0}(\theta'_0(h)))] + \epsilon \ .
\end{aligned}$$

The theorem then follows for bounded ℓ by the facts that $\{n_\ell\}$ and ϵ are arbitrary. For general loss functions ℓ, replace ℓ in the above proof with $\ell_M \equiv \ell \wedge M$ and then let $M \to \infty$. ■

Lemma 2.6.2. *Suppose that Assumptions 2.2.1, 2.2.2, 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5, 2.3.6 and 2.3.7 hold. Let $\hat{v}_{n,k,m} \in \hat{\Psi}_{k,m}$. If the loss function ℓ is bounded, then it follows that for each $k, m \in \mathbf{N}$,*

$$d(\hat{v}_{n,k,m}, \Psi_{k,m}) \xrightarrow{p} 0 \ , \qquad (2.71)$$

under $P_{n,h}$ for all $h \in H$.

PROOF: Fix $k, m \in \mathbf{N}$ throughout. For notational simplicity, write $\vartheta \equiv (v^\intercal, c^\intercal)^\intercal \in \Theta \equiv K^k_{\tau_k} \times K^m_{\lambda_m}$ and $\eta_0 \equiv (\theta'_0, \phi'_{\theta_0})$, and define the function $f_{\vartheta,\eta_0}(\cdot) : \mathbb{D} \to \mathbf{R}$ by

$$f_{\vartheta,\eta_0}(z) \equiv \ell(\phi'_{\theta_0}(z + (\psi^k)^\intercal v + \theta'_0(g^m)^\intercal c) - \phi'_{\theta_0}(\theta'_0(g^m)^\intercal c)) \ .$$

Let $\hat{\eta}_n \equiv (\hat{\theta}'_n, \hat{\phi}'_n)$ and define

$$P f_{\vartheta,\eta_0} \equiv E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta'_0(g^m)^\intercal c) - \phi'_{\theta_0}(\theta'_0(g^m)^\intercal c))] ,$$
$$P f_{\vartheta,(\theta'_0,\hat{\phi}'_n)} \equiv E[\ell(\hat{\phi}'_n(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta'_0(g^m)^\intercal c) - \hat{\phi}'_n(\theta'_0(g^m)^\intercal c))|\{X_i\}] ,$$
$$P f_{\vartheta,\hat{\eta}_n} \equiv E[\ell(\hat{\phi}'_n(\mathbb{G}_0 + (\psi^k)^\intercal v + \hat{\theta}'_n(\hat{g}^m)^\intercal c) - \hat{\phi}'_n(\hat{\theta}'_n(\hat{g}^m)^\intercal c))|\{X_i\}] ,$$
$$\mathbb{P}_n f_{\vartheta,\hat{\eta}_n} \equiv E[\ell(\hat{\phi}'_n(\hat{\mathbb{G}}^*_n + (\psi^k)^\intercal v + \hat{\theta}'_n(\hat{g}^m)^\intercal c) - \hat{\phi}'_n(\hat{\theta}'_n(\hat{g}^m)^\intercal c))|\{X_i\}] ,$$

where $P f_{\vartheta,(\theta'_0,\hat{\phi}'_n)}$, and $P f_{\vartheta,\hat{\eta}_n}$ are expectations taken with respect to $\mathbb{G}_0$ while holding $\{X_i\}_{i=1}^n$ fixed. The ensuing arguments are organized parallel to those of the consistency result in the theory of the extremum estimation, the only difference being that the set of population minimizers is possibly a nonsingleton. Therefore, we need to show a uniform convergence result and an identification condition.

<u>Uniform Convergence:</u> For each $\epsilon > 0$,

$$\sup_{\vartheta\in\Theta} |\mathbb{P}_n f_{\vartheta,\hat{\eta}_n} - P f_{\vartheta,\eta_0}| = o_p(1) , \tag{2.72}$$

under $\{P_{n,0}\}$. In turn, it suffices to show that

$$\sup_{\vartheta\in\Theta} |\mathbb{P}_n f_{\vartheta,\hat{\eta}_n} - P f_{\vartheta,\hat{\eta}_n}| = o_p(1) , \tag{2.73a}$$
$$\sup_{\vartheta\in\Theta} |P f_{\vartheta,\hat{\eta}_n} - P f_{\vartheta,(\theta'_0,\hat{\phi}'_n)}| = o_p(1) , \tag{2.73b}$$
$$\sup_{\vartheta\in\Theta} |P f_{\vartheta,(\theta'_0,\hat{\phi}'_n)} - P f_{\vartheta,\eta_0}| = o_p(1) , \tag{2.73c}$$

under $\{P_{n,0}\}$. Fix $\epsilon > 0$ and consider (2.73a). Note that for every realization of $\{X_i\}$, the real valued functions

$$\ell(\hat{\phi}'_n(\cdot + (\psi^k)^\intercal v + \hat{\theta}'_n(\hat{g}^m)^\intercal c) - \hat{\phi}'_n(\hat{\theta}'_n(\hat{g}^m)^\intercal c))$$

are bounded and Lipschitz continuous on $\mathbb{D}$ with Lipschitz constant $C_\ell C_{\hat{\phi}'}$ by Assumptions

2.3.3 and 2.3.5(iii)-(b). It then follows by Assumption 2.3.5(i) that

$$\sup_{\vartheta\in\Theta}|\mathbb{P}_n f_{\vartheta,\hat{\eta}_n} - Pf_{\vartheta,\hat{\eta}_n}| \le \sup_{f\in \mathrm{BL}_a(\mathbb{D})} |E[f(\hat{\mathbb{G}}_n^*)|\{X_i\}] - E[f(\mathbb{G}_0)]| = o_p(1) ,$$

under $\{P_{n,0}\}$, where $a \equiv \max\{M, C_\ell C_{\hat{\phi}'}\}$ with M being a upper bound of ℓ, proving (2.73a).

Next, consider (2.73b). We have

$$\begin{aligned}\sup_{\vartheta\in\Theta}|Pf_{\vartheta,\hat{\eta}_n} - Pf_{\vartheta,(\theta_0',\hat{\phi}_n')}| &\le 2C_\ell C_{\hat{\phi}'}\|\hat{\theta}_n'(\hat{g}^m)^\intercal c - \theta_0'(g^m)^\intercal c\|_{\mathbb{D}} \\ &\le 2C_\ell C_{\hat{\phi}'}\lambda_m \sum_{j=1}^m \|\hat{\theta}_n'(\hat{g}_j) - \theta_0'(g_j)\|_{\mathbb{D}} = o_p(1) ,\end{aligned}$$

under $\{P_{n,0}\}$, where the first inequality is due to Assumptions 2.3.3 and 2.3.5(iii)-(b), and the second inequality is by Assumptions 2.3.5(ii) and 2.3.6(ii). This shows (2.73b).

Lastly, let us deal with (2.73c). For fixed $k, m \in \mathbf{N}$, $K_1 \equiv \{(\psi^k)^\intercal v : v \in K_{\tau_k}^k\}$ and $K_2 \equiv \{\theta_0'(g^m)^\intercal c : c \in K_{\lambda_m}^m\}$ is compact in $\mathbb{D}$ by Proposition 4.26 in Folland (1999). Fix $\epsilon, \eta > 0$. Since $\mathbb{G}_0$ is tight, there is some compact $K_0 \subset \mathbb{D}$ such that $P(\mathbb{G}_0 \notin K_0) < \eta/(2M)$. Let $K \equiv K_0 + K_1 + K_2$. Clearly, $K \subset \mathbb{D}$ is compact. We now have

$$\begin{aligned}\sup_{\vartheta\in\Theta}|Pf_{\vartheta,(\theta_0',\hat{\phi}_n')} - Pf_{\vartheta,\eta_0}| &\le \sup_{\vartheta\in\Theta} E[|\ell(\hat{\phi}_n'(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(g^m)^\intercal c) \\ &\quad - \hat{\phi}_n'(\theta_0'(g^m)^\intercal c)) - \ell(\phi_{\theta_0}'(\mathbb{G}_0 + (\psi^k)^\intercal v + \theta_0'(g^m)^\intercal c) - \phi_{\theta_0}'(\theta_0'(g^m)^\intercal c))|\,|\{X_i\}] \\ &\le \sup_{z\in K} C_\ell\|\hat{\phi}_n'(z) - \phi_{\theta_0}'(z)\|_{\mathbb{E}} + \sup_{z\in K_2} C_\ell\|\hat{\phi}_n'(z) - \phi_{\theta_0}'(z)\|_{\mathbb{E}} + 2M\cdot P(\mathbb{G}_0 \notin K_0) \\ &\le o_p(1) + \eta ,\end{aligned}$$

where the first inequality is by the triangle inequality, the second is by Assumption 2.3.3, and the last is by Lemma 2.6.7. This immediately implies (2.73c) and hence we conclude that (2.72) holds.

Identification Condition: For each $\epsilon > 0$,

$$\inf_{v\in K_{\tau_k}^k\setminus\Psi_{k,m}^\epsilon}\ \sup_{c\in K_{\lambda_m}^m} Pf_{(v,c),\eta_0} > \inf_{v\in K_{\tau_k}^k}\ \sup_{c\in K_{\lambda_m}^m} Pf_{(v,c),\eta_0} , \tag{2.74}$$

or equivalently, $\inf_{v \in K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}} B_m(v) > \inf_{v \in K^k_{\tau_k}} B_m(v)$, where $\Psi^\epsilon_{k,m} \equiv \{v \in \mathbf{R}^k : d(v, \Psi_{k,m}) \leq \epsilon\}$. To see this, fix $\epsilon > 0$ and suppose that $\inf_{u \in K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}} B_m(v) = \inf_{u \in K^k_{\tau_k}} B_m(v)$. Then we may pick a sequence $\{v_i\} \subset K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}$ such that

$$B_m(v_i) \to \inf_{v \in K^k_{\tau_k}} B_m(v) \ \text{ as } \ i \to \infty \ .$$

By passing to a subsequence if necessary, we may assume that $v_i \to v^*$ as $i \to \infty$ where $v^* \in \overline{K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}}$. Assumptions 2.3.3 and 2.3.4 imply that $B_m(v)$ is (Lipschitz) continuous in v and therefore $B_m(v^*) = \inf_{v \in K^k_{\tau_k}} B_m(v)$, meaning that $v^* \in \Psi_{k,m}$. On the other hand, $v^* \in \overline{K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}}$ implies that we may take a sequence $\{v_j\} \subset K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}$ such that $v_j \to v^*$ as $j \to \infty$, which in turn implies that

$$d(v^*, \Psi_{k,m}) = \lim_{j \to \infty} d(v_j, \Psi_{k,m}) \geq \epsilon > 0 \ ,$$

a contradiction. Therefore, (2.74) must hold.

We are now in a position to show result (2.71). Fix $\epsilon > 0$. By result (2.74), there is some $\delta > 0$ such that whenever $v \in K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}$ we have

$$B_m(v) - B_m(v^*_{k,m}) \geq \delta \ , \tag{2.75}$$

where $v^*_{k,m}$ is any element in $\Psi_{k,m}$. It follows that

$$\begin{aligned}
P_{n,0}(d(\hat{v}_{n,k,m}, \Psi_{k,m}) > \epsilon) &= P_{n,0}(\hat{v}_{n,k,m} \in K^k_{\tau_k} \setminus \Psi^\epsilon_{k,m}) \leq P_{n,0}(B_m(\hat{v}_{n,k,m}) - B_m(v^*_{k,m}) \geq \delta) \\
&= P_{n,0}(B_m(\hat{v}_{n,k,m}) - \hat{B}_m(\hat{v}_{n,k,m}) + \hat{B}_m(\hat{v}_{n,k,m}) - B_m(v^*_{k,m}) \geq \delta) \\
&\leq P_{n,0}(B_m(\hat{v}_{n,k,m}) - \hat{B}_m(\hat{v}_{n,k,m}) + \hat{B}_m(v^*_{k,m}) + \epsilon_n - B_m(v^*_{k,m}) \geq \delta) \\
&\leq P_{n,0}(2 \sup_{v \in K^k_{\tau_k}} |B_m(v) - \hat{B}_m(v)| + \epsilon_n \geq \delta) \\
&\leq P_{n,0}(2 \sup_{\vartheta \in \Theta} |\mathbb{P}_n f_{\vartheta, \hat{\eta}_n} - P f_{\vartheta, \eta_0}| + \epsilon_n \geq \delta) \to 0 \ ,
\end{aligned} \tag{2.76}$$

where the second inequality is by the definition of $\hat{v}_{n,k,m}$, and the last step is by (2.72) and the fact that $\epsilon_n = o_p(1)$ as $n \to \infty$. By Assumption 2.2.1(ii), $P_{n,h}$ and $P_{n,0}$ are mutually contiguous for each $h \in H$; see, for example, Example 6.5 in van der Vaart (1998). Result (2.71) then follows from (2.76) and Le Cam's first lemma. ■

Proof of Theorem 2.3.3: The proof follows closely that of Theorem 2.3.2. Define

$$B(u) \equiv \max_{c \in \mathbf{R}^m} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + c) - \phi'_{\theta_0}(c))] \ ,$$
$$B_\lambda(u) \equiv \max_{c \in K_\lambda^m} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + c) - \phi'_{\theta_0}(c))] \ , \ \Psi_{\tau,\lambda} \equiv \arg\min_{u \in K_\tau^m} B_\lambda(u) \ .$$

Again, consider first the case when ℓ is bounded. Fix $\epsilon > 0$ and $\tau > 0$. By Assumption 2.3.3 and 2.3.4, $B_\lambda(u)$ and $B(u)$ are both (Lipschitz) continuous in u. Moreover, it is clear that $B_\lambda(u) \uparrow B(u)$ as $\lambda \uparrow \infty$ for each $u \in K_\tau^m$. It then follows by Dini's theorem that $B_\lambda \to B$ uniformly on K_τ^m so that we may find some $\lambda > 0$ with $\lambda \geq \tau$ if necessary such that

$$B(u) \leq B_\lambda(u) + \epsilon \text{ for all } u \in K_\tau^m \ . \tag{2.77}$$

The rest of the proof is essentially the same as that of Theorem 2.3.3 by employing subsequence arguments, in view of Lemma 2.6.3 and Lemma 2.6.6. ■

Lemma 2.6.3. *Suppose that Assumptions 2.2.1, 2.2.2, 2.2.3 2.3.1, 2.3.2, 2.3.3, 2.3.4, and 2.3.5(i)(iii) hold. Let $\hat{u}_{n,\tau,\lambda} \in \hat{\Psi}_{\tau,\lambda}$. Further assume that the loss function ℓ is bounded. Then for all $\tau, \lambda > 0$ we have*

$$d(\hat{u}_{n,\tau,\lambda}, \Psi_{\tau,\lambda}) \xrightarrow{p} 0 \ , \tag{2.78}$$

under $P_{n,h}$ for each $h \in H$.

Proof: Following the proof of Lemma 2.6.2, it suffices to show a unform convergence condition and an identification condition. Since the identification condition can be shown using exactly the same arguments as before, we shall only prove the following: for fixed

$\tau, \lambda > 0$,

$$\sup_{u \in K_\tau^m} |\hat{B}_\lambda(u) - B_\lambda(u)| = o_p(1) \ , \tag{2.79}$$

under $P_{n,0}$. To this end, rewrite

$$\begin{aligned} &\sup_{u\in K_\tau^m} |\hat{B}_\lambda(u) - B_\lambda(u)| \\ &\le \sup_{u \in K_\tau^m} |\sup_{c \in K_\lambda^m} E[\ell(\hat{\phi}_n'(\hat{\mathbb{G}}_n^* + u + c) - \hat{\phi}_n'(c))|\{X_i\}] - \sup_{c\in K_\lambda^m} E[\ell(\hat{\phi}_n'(\mathbb{G}_0 + u + c) - \hat{\phi}_n'(c))|\{X_i\}]| \\ &\quad + \sup_{u \in K_\tau^m} |\sup_{c \in K_\lambda^m} E[\ell(\hat{\phi}_n'(\mathbb{G}_0 + u + c) - \hat{\phi}_n'(c))|\{X_i\}] - \sup_{c\in K_\lambda^m} E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u + c) - \phi_{\theta_0}'(c))]| \ . \end{aligned} \tag{2.80}$$

For the first term on the right hand side, we have by Assumptions 2.3.3 and 2.3.5(iii)(b):

$$\begin{aligned} &\sup_{u \in K_\tau^m} |\sup_{c \in K_\lambda^m} E[\ell(\hat{\phi}_n'(\hat{\mathbb{G}}_n^* + u + c) - \hat{\phi}_n'(c))|\{X_i\}] - \sup_{c\in K_\lambda^m} E[\ell(\hat{\phi}_n'(\mathbb{G}_0 + u + c) - \hat{\phi}_n'(c))|\{X_i\}]| \\ &\le \sup_{u\in K_\tau^m, c \in K_\lambda^m} |E[\ell(\hat{\phi}_n'(\hat{\mathbb{G}}_n^* + u + c) - \hat{\phi}_n'(c))|\{X_i\}] - E[\ell(\hat{\phi}_n'(\mathbb{G}_0 + u + c) - \hat{\phi}_n'(c))|\{X_i\}]| \\ &\le \sup_{f \in \mathrm{BL}_a(\mathbb{D})} |E[f(\hat{\mathbb{G}}_n^*)|\{X_i\}] - E[f(\mathbb{G}_0)]| = o_p(1) \ , \end{aligned} \tag{2.81}$$

where $a \equiv \max\{M, C_\ell C_{\hat{\phi}'}\}$ with M being a upper bound of ℓ, and the last equality is by Assumption 2.3.5(i). As for the second term on the right hand side of (2.80), fix $\epsilon > 0$ and choose a compact set $K_0 \subset \mathbf{R}^m$ such that $P(\mathbb{G}_0 \notin K_0) < \epsilon/(4M)$. Then

$$\begin{aligned} &\sup_{u \in K_\tau^m} |\sup_{c \in K_\lambda^m} E[\ell(\hat{\phi}_n'(\mathbb{G}_0 + u + c) - \hat{\phi}_n'(c))|\{X_i\}] - \sup_{c\in K_\lambda^m} E[\ell(\phi_{\theta_0}'(\mathbb{G}_0 + u + c) - \phi_{\theta_0}'(c))]| \\ &\le \sup_{u\in K_\tau^m, c\in K_\lambda^m} E[|\ell(\hat{\phi}_n'(\mathbb{G}_0 + u + c) - \hat{\phi}_n'(c)) - \ell(\phi_{\theta_0}'(\mathbb{G}_0 + u + c) - \phi_{\theta_0}'(c))||\{X_i\}] \\ &\le 2M \cdot P(\mathbb{G}_0 \notin K_0) + C_\ell \sup_{z \in K} \|\hat{\phi}_n'(z) - \phi_{\theta_0}'(z)\|_{\mathbb{E}} + C_\ell \sup_{z \in K_\lambda^m} \|\hat{\phi}_n'(z) - \phi_{\theta_0}'(z)\|_{\mathbb{E}} \\ &\le \frac{\epsilon}{2} + o_p(1) \ , \end{aligned} \tag{2.82}$$

where $K \equiv K_0 + K_\lambda^m + K_\tau^m$ is compact, and the last step is by Lemma 2.6.7.

Result (2.79) then follows from results (2.80), (2.81) and (2.82). The rest of the proof follows from that of Lemma (2.6.2). ■

PROOF OF THEOREM 2.3.4: The proof is essentially the same as that of Theorem 2.3.2 by combining Lemmas 2.6.4 and 2.6.6 and is thus omitted. ■

Lemma 2.6.4. *Suppose that Assumptions 2.2.1, 2.2.2, 2.2.3, 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5, 2.3.6(i)(ii), and 2.3.8 hold. Assume that the loss function ℓ is bounded. If $\hat{u}_{n,m} \in \hat{\Psi}_m$, then for each $m \in \mathbf{N}$,*

$$d(\hat{u}_{n,m}, \Psi_m) \xrightarrow{p} 0 \ , \tag{2.83}$$

under $\{P_{n,h}\}$ for each $h \in H$.

PROOF: Fix $m \in \mathbf{N}$ throughout. The proof closely follows that of Lemma 2.6.2. First, by the same arguments as before we can show the following uniform convergence result:

$$\sup_{u \in \mathbb{D}_u} |\hat{B}_m(u) - B_m(u)| \xrightarrow{p} 0 \tag{2.84}$$

under $P_{n,0}$, and the identification condition – i.e. for each $\epsilon > 0$,

$$\inf_{u \in \mathbb{D}_u \setminus \Psi_m^\epsilon} B_m(u) > \inf_{u \in \mathbb{D}_u} B_m(u) \ . \tag{2.85}$$

Fix $\epsilon > 0$. Now by the identification result (2.85), there is some $\delta > 0$ such that whenever $u \in \mathbb{D}_u \setminus \Psi_m^\epsilon$ we have

$$B_m(u) - B_m(u_m^*) \geq 2\delta \ , \tag{2.86}$$

where u_m^* is any element in Ψ_m. Moreover, by Assumptions 2.3.3 and 2.3.4, $B_m(u)$ is continuous. Then by Assumption 2.3.8(ii) and the fact that $k_n \to \infty$ as $n \to \infty$, we may pick $u_{k_n} \in \mathbb{D}_{k_n}$ such that

$$B_m(u_{k_n}) - \delta \leq B_m(u_m^*) \ , \tag{2.87}$$

for all n sufficiently large. We now have for all n sufficiently large (so that (2.87) holds):

$$\begin{aligned}
P_{n,0}(d(\hat{u}_{n,m},\Psi_m) > \epsilon) &= P_{n,0}(\hat{u}_{n,m} \in \mathbb{D}_u \backslash \Psi_m^{\epsilon}) \\
&\leq P_{n,0}(B_m(\hat{u}_{n,m}) - B_m(u_m^*) \geq 2\delta) \\
&= P_{n,0}(B_m(\hat{u}_{n,m}) - \hat{B}_m(\hat{u}_{n,m}) + \hat{B}_m(\hat{u}_{n,m}) - B_m(u_m^*) \geq 2\delta) \\
&\leq P_{n,0}(B_m(\hat{u}_{n,m}) - \hat{B}_m(\hat{u}_{n,m}) + \hat{B}_m(u_{k_n}) + \epsilon_n - B_m(u_m^*) \geq 2\delta) \\
&\leq P_{n,0}(B_m(\hat{u}_{n,m}) - \hat{B}_m(\hat{u}_{n,m}) + \hat{B}_m(u_{k_n}) + \epsilon_n - B_m(u_{k_n}) \geq \delta) \\
&\leq P_{n,0}(2 \sup_{u \in \mathbb{D}_{k_n}} |B_m(u) - \hat{B}_m(u)| + \epsilon_n \geq \delta) \\
&\leq P_{n,0}(2 \sup_{u \in \mathbb{D}_u} |B_m(u) - \hat{B}_m(u)| + \epsilon_n \geq \delta) \to 0 \ , \qquad (2.88)
\end{aligned}$$

where the second inequality is due to the definition of $\hat{u}_{n,m}$, and the third inequality is by (2.87). Result (2.83) then follows under $\{P_{n,0}\}$ by (2.88), (2.84) and $\epsilon_n = o_p(1)$ as $n \to \infty$. By Assumption 2.2.1(ii), $P_{n,h}$ and $P_{n,0}$ are mutually contiguous for each $h \in H$; see, for example, Example 6.5 in van der Vaart (1998). Then lemma follows from Le Cam's first lemma. ∎

Lemma 2.6.5. *Let $(\mathbb{D},\tau)$ be a topological space and $f : \mathbb{D} \to \overline{\mathbf{R}}$ be a lower semicontinuous function. For any $A \subset \mathbb{D}$, we have*

$$\sup_{x \in A} f(x) = \sup_{x \in \overline{A}} f(x) \ ,$$

where $\overline{A}$ denotes the closure of A relative to τ.

Proof: We only consider the nontrivial case when A is nonempty. Suppose first that $\sup_{x\in\overline{A}} f(x) = \infty$. Fix arbitrary large $M > 0$. Then there is some $x_0 \in \overline{A}$ such that $f(x_0) \geq M$. Since $x_0 \in \overline{A}$, we may pick a net $\{x_\alpha\} \subset A$ such that $x_\alpha \to x_0$ in τ. But then since f is lower semicontinuous, $\liminf_\alpha f(x_\alpha) \geq f(x_0)$. In turn, this implies that there is some α^* such that $\sup_{x\in A} f(x) \geq f(x_{\alpha^*}) > f(x_0) - 1 \geq M - 1$. Since M is arbitrary, it follows that $\sup_{x\in A} f(x) = \infty$.

Now suppose that $\sup_{x\in\overline{A}} f(x) < \infty$. Obviously, $\sup_{x\in A} f(x) \leq \sup_{x\in\overline{A}} f(x)$. To conclude, it suffices to show that for any $\epsilon > 0$,

$$\sup_{x\in\overline{A}} f(x) \le \sup_{x\in A} f(x) + \epsilon \ . \tag{2.89}$$

First, we may pick some $x_0 \in \overline{A}$ such that $\sup_{x\in\overline{A}} f(x) \le f(x_0) + \epsilon/2$. Next, we may choose a net $\{x_\alpha\} \subset A$ such that $x_\alpha \to x_0$ in τ. Since f is lower semicontinuous, $\liminf_\alpha f(x_\alpha) \ge f(x_0)$, implying that we may find some α^* such that $f(x_0) \le f(x_{\alpha^*}) + \epsilon/2$. Combining previous two inequalities, we conclude that

$$\sup_{x\in\overline{A}} f(x) \le f(x_0) + \epsilon/2 \le f(x_{\alpha^*}) + \epsilon \le \sup_{x\in A} f(x) + \epsilon \ ,$$

proving (2.89), and we thus establish the Lemma. ■

Lemma 2.6.6. *Let $(\mathbb{D}, d)$ be a metric space and $K \subset \mathbb{D}$ a nonempty compact subset. Let $(\Omega_n, \mathcal{A}_n, P_n)$ be a sequence of probability spaces and $X_n : \Omega_n \to \mathbb{D}$ arbitrary maps such that $d(X_n, K) \xrightarrow{p} 0$ under $\{P_n\}$. Then for any subsequence $\{n_k\}$, there exist a further subsequence $\{n_{k_j}\}$ and some deterministic $c \in K$ such that $X_{n_{k_j}} \xrightarrow{p} c$ as $j \to \infty$.*

Proof: We proceed by contradiction. Fix a subsequence $\{n_k\}$ and suppose that for each $c \in K$ and every subsequence $\{n_{k_j}\}$, $X_{n_{k_j}} \overset{p}{\nrightarrow} c$ as $j \to \infty$. This implies that for each $c \in K$ there exist $\epsilon_c > 0$ and $\eta_c \in (0, 1)$ such that

$$\liminf_{k\to\infty} P_{n_k}(d(X_{n_k}, c) > 2\epsilon_c) > \eta_c \ ,$$

or equivalently,

$$\limsup_{k\to\infty} P_{n_k}(d(X_{n_k}, c) < 2\epsilon_c) < 1 - \eta_c \ . \tag{2.90}$$

Next, for each $c \in K$, let $B_c(\epsilon_c) \equiv \{c' \in K : d(c', c) < \epsilon_c\}$. Since $\{B_c(\epsilon_c)\}_{c\in K}$ is an open cover of K, compactness of K implies that there exists a finite subcover $\{B_{c_j}(\epsilon_j)\}_{j=1}^{J^*}$ with $J^* < \infty$ and $\epsilon_j \equiv \epsilon_{c_j}$ that covers K. Observe that if $d(X_{n_k}, c_j) \ge 2\epsilon_j$ for all $j = 1, \ldots, J^*$, then we must have $d(X_{n_k}, K) \ge \epsilon_0$, where $\epsilon_0 \equiv \min(\epsilon_1, \ldots, \epsilon_{J^*})$. To see this, suppose

$d(X_{n_k}, K) < \epsilon_0$ and $d(X_{n_k}, K) = d(X_{n_k}, c')$ for some $c' \in K$. Since $d(c', c_j) < \epsilon_j$ for some j, it follows that

$$d(X_{n_k}, c_j) \leq d(X_{n_k}, c') + d(c', c_j) < \epsilon_0 + \epsilon_j \leq 2\epsilon_j ,$$

a contradiction, implying that

$$P_{n_k}(d(X_{n_k}, K) \geq \epsilon_0) \geq P_{n_k}(d(X_{n_k}, c_j) \geq 2\epsilon_j, j = 1, \ldots, J^*) . \tag{2.91}$$

Elementary calculations then reveal that

$$\begin{aligned}
&\liminf_{k\to\infty} P_{n_k}(d(X_{n_k}, c_j) \geq 2\epsilon_j, j = 1, \ldots, J^*) \\
&\quad = 1 - \limsup_{k\to\infty} P_{n_k}(d(X_{n_k}, c_j) < 2\epsilon_j \text{ for some } j = 1, \ldots, J^*) \\
&\quad \geq 1 - \sum_{j=1}^{J^*} \limsup_{k\to\infty} P_{n_k}(d(X_{n_k}, c_j) < 2\epsilon_j) \geq 1 - \sum_{j=1}^{J^*}(1 - \eta_{c_j}) \equiv \eta_0 ,
\end{aligned} \tag{2.92}$$

where we may assume that $\eta_0 > 0$ by choosing η_{c_j}'s sufficiently small since we may increase each ϵ_c to make η_c arbitrarily close to 1 or $1 - \eta_c$ arbitrarily close to zero and meanwhile J^* wouldn't increase because the radius of each open ball $B_c(\epsilon_c)$ of the open cover $\{B_c(\epsilon_c)\}_{c\in K}$ increases. Combination of (2.91) and (2.92) then yields

$$\liminf_{k\to\infty} P_{n_k}(d(X_{n_k}, K) \geq \epsilon_0) \geq \eta_0 > 0 ,$$

a contradiction. This completes the proof. ■

Lemma 2.6.7. *Suppose Assumptions 2.2.1(ii) and 2.3.5(iii) hold. Then for any compact subset $K \subset \mathbb{D}$ and any $\epsilon > 0$,*

$$\sup_{I\subset_f H} \limsup_{n\to\infty} \sup_{h\in I} P_{n,h}(\sup_{z\in K} \|\hat{\phi}'_n(z) - \phi'_{\theta_0}(z)\|_{\mathbb{E}} > \epsilon) = 0 . \tag{2.93}$$

Proof: Fix a compact subset $K \subset \mathbb{D}$ and $\epsilon > 0$. Since K is compact, ϕ'_{θ_0} is continuous and hence uniformly continuous on K so that we may find a finite collection $\{z_j\}_{j=1}^{J}$ with

$J < \infty$ such that $z_j \in K$ for all j and

$$\sup_{z\in K}\min_{1\le j\le J}\max\{C_{\hat{\phi}'}\|z-z_j\|_{\mathbb{D}}, \|\phi'_{\theta_0}(z)-\phi'_{\theta_0}(z_j)\|_{\mathbb{E}}\} < \frac{\epsilon}{3}\ .$$

This, along with Assumption 2.3.5(iii)-b), implies that

$$\sup_{z\in K}\|\hat{\phi}'_n(z)-\phi'_{\theta_0}(z)\|_{\mathbb{E}} \le \max_{1\le j\le J}\|\hat{\phi}'_n(z_j)-\phi'_{\theta_0}(z_j)\|_{\mathbb{E}} + \frac{2}{3}\epsilon\ . \tag{2.94}$$

Fix a finite subset $I \subset H$. By Assumption 2.2.1(ii), $P_{n,h}$ and $P_{n,0}$ are mutually contiguous for each $h \in H$; see, for example, Example 6.5 in van der Vaart (1998). It follows from Assumption 2.3.5(iv)-a) and Le Cam's first lemma that

$$\limsup_{n\to\infty}\sup_{h\in I} P_{n,h}(\|\hat{\phi}'_n(z_j)-\phi'_{\theta_0}(z_j)\|_{\mathbb{E}} > \frac{\epsilon}{3}) = 0 \text{ for all } j = 1,\ldots,J\ . \tag{2.95}$$

Combining (2.95) and (2.95) we thus conclude that

$$\begin{aligned}
&\limsup_{n\to\infty}\sup_{h\in I}P_{n,h}(\sup_{z\in K}\|\hat{\phi}'_n(z)-\phi'_{\theta_0}(z)\|_{\mathbb{E}} > \epsilon) \\
&\qquad\le \limsup_{n\to\infty}\sup_{h\in I}P_{n,h}(\max_{1\le j\le J}\|\hat{\phi}'_n(z_j)-\phi'_{\theta_0}(z_j)\|_{\mathbb{E}} > \frac{\epsilon}{3}) \\
&\qquad\le \sum_{j=1}^{J}\limsup_{n\to\infty}\sup_{h\in I}P_{n,h}(\|\hat{\phi}'_n(z_j)-\phi'_{\theta_0}(z_j)\|_{\mathbb{E}} > \frac{\epsilon}{3}) = 0\ .
\end{aligned}$$

Since this is true for each finite $I \subset H$, the lemma then follows immediately. ■

2.6.2 Results for Examples 2.2.1 - 2.2.4

Example 2.2.1 (Best Treatment)

By Corollary 2.3.1, the lower bound when $\theta^{(1)} = \theta^{(2)}$ in this example becomes

$$\begin{aligned}
&\inf_{u\in\mathbf{R}^2}\sup_{c\in\mathbf{R}^2} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0+u+c)-\phi'_{\theta_0}(c))] \\
&\quad= \inf_{u\in\mathbf{R}^2}\sup_{c\in\mathbf{R}^2} E[(\max\{\mathbb{G}_0^{(1)}+u^{(1)}+c^{(1)}, \mathbb{G}_0^{(2)}+u^{(2)}+c^{(2)}\}-\max\{c^{(1)},c^{(2)}\})^2]\ ,
\end{aligned}$$

where $\mathbb{G}_0 \sim \mathcal{N}(0, \sigma^2 I_2)$. Replace $u^{(1)}$ and $u^{(2)}$ with $u - \Delta_u$ and u respectively; similarly define $c - \Delta_c$ and c. Since the problem is symmetric in c_1 and c_2, we may assume that $\Delta_c \geq 0$. Then we have

$$\begin{aligned}
&\inf_{u \in \mathbf{R}^2} \sup_{c \in \mathbf{R}^2} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + c) - \phi'_{\theta_0}(c))] \\
&\qquad = \inf_{u \in \mathbf{R}, \Delta_u \in \mathbf{R}} \sup_{\Delta_c \geq 0} E[(\max(\mathbb{G}_0^{(1)} - \Delta_c - \Delta_u, \mathbb{G}_0^{(2)}) + u)^2] \\
&\qquad = \inf_{u \in \mathbf{R}, \Delta_u \in \mathbf{R}} \sup_{\Delta_c \geq \Delta_u} E[(\max(\mathbb{G}_0^{(1)} - \Delta_c, \mathbb{G}_0^{(2)}) + u)^2] \ .
\end{aligned}$$

Notice that for each $u \in \mathbf{R}$,

$$\sup_{\Delta_c \geq \Delta_u} E[(\max(\mathbb{G}_0^{(1)} - \Delta_c, \mathbb{G}_0^{(2)}) + u)^2]$$

is monotonically decreasing in Δ_u, whence we have by setting $\Delta_u = \infty$ that

$$\begin{aligned}
&\inf_{u \in \mathbf{R}^2} \sup_{c \in \mathbf{R}^2} E[\ell(\phi'_{\theta_0}(\mathbb{G}_0 + u + c) - \phi'_{\theta_0}(c))] \\
&\qquad = \inf_{u \in \mathbf{R}} E[(\mathbb{G}_0^{(2)} + u)^2] = E[(\mathbb{G}_0^{(2)})^2] = \sigma^2 \ .
\end{aligned}$$

It is clear that the optimum is achieved at $u = (-\infty, 0)$ and $c = (-\infty, c^{(2)})$ with $c^{(2)} \in \mathbf{R}$ arbitrary. This is consistent with Example 6 in Song (2014). By tedious but straightforward calculations one can show that the lower bound can be also achieved at $u = 0$ and $c = 0$.

Example 2.2.2 (Interval Censored Outcome)

In this example, the identified region for ϑ is

$$\Theta_0 \equiv \{\vartheta \in \mathbf{R}^2 : E[Y_l|Z] \leq Z^\intercal \vartheta \leq E[Y_u|Z]\} \ .$$

Let's now work out $\sup_{\vartheta \in \Theta_0} \lambda^\intercal \vartheta$ for some fixed $\lambda \in \mathbf{R}^2$. We have

$$\begin{aligned}
&\sup\{\lambda^\intercal E[ZZ^\intercal]^{-1} E[ZE[Y|Z]] : E[Y_l|Z] \leq E[Y|Z] \leq E[Y_u|Z]\} \\
&\qquad = \sum_{j=-1}^{1} 1\{\lambda^\intercal E[ZZ^\intercal]^{-1} z_j \geq 0\} \lambda^\intercal E[ZZ^\intercal]^{-1} z_j E[Y_u|Z = z_j] P(Z = z_j) \\
&\qquad\quad + \sum_{j=-1}^{1} 1\{\lambda^\intercal E[ZZ^\intercal]^{-1} z_j < 0\} \lambda E[ZZ^\intercal]^{-1} z_j E[Y_l|Z = z_j] P(Z = z_j) \ ,
\end{aligned}$$

where $z_j = (1, j)^\intercal$ for $j \in \{-1, 0, 1\}$. Consider

$$\begin{aligned}
&1\{\lambda^\intercal E[ZZ^\intercal]^{-1} z_1 \geq 0\} \lambda^\intercal E[ZZ^\intercal]^{-1} z_1 \\
&\qquad = 1\{\lambda^{(1)} \frac{\theta^{(1)} + \theta^{(2)}}{\theta^{(1)} + \theta^{(2)} - (\theta^{(2)} - \theta^{(1)})^2} + \lambda^{(2)} \frac{\theta^{(1)} - \theta^{(2)}}{\theta^{(1)} + \theta^{(2)} - (\theta^{(2)} - \theta^{(1)})^2} \geq 0\} \\
&\qquad\quad \times \left[\lambda^{(1)} \frac{\theta^{(1)} + \theta^{(2)}}{\theta^{(1)} + \theta^{(2)} - (\theta^{(2)} - \theta^{(1)})^2} + \lambda^{(2)} \frac{\theta^{(1)} - \theta^{(2)}}{\theta^{(1)} + \theta^{(2)} - (\theta^{(2)} - \theta^{(1)})^2}\right] \\
&\qquad \equiv 1\{\psi(\theta) \geq 0\} \psi(\theta) \ .
\end{aligned}$$

By the chain rule for Hadamard directionally differentiable maps (see Remark 2.3.1), one can show that

$$\phi'_\theta(z) = \psi'_\theta(z) 1\{\psi(\theta) > 0\} + \max\{\psi'_\theta(z), 0\} 1\{\psi(\theta) = 0\} \ .$$

Example 2.2.3 (Incomplete Auction Model)

Lemma 2.6.8. *Let $\phi : \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R}) \to \ell^\infty(\mathbf{R})$ be given by $\phi(\theta) = \max(\theta^{(1)}, \theta^{(2)})$, and $B_i = 1\{x : \theta_i(x) > \theta_{-i}(x)\}$ for $i = 1, 2$ and $B_0 = \{x : \theta_1(x) = \theta_2(x)\}$. It follows that ϕ is Hadamard directionally differentiable at any $\theta \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ such that for any $z \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$,*

$$\phi'_\theta(z) = z^{(1)} 1_{B_1} + z^{(2)} 1_{B_2} + \max\{z^{(1)}, z^{(2)}\} 1_{B_0} \ .$$

Proof: Fix $z \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ and let $\{z_n\} \equiv \{(z_{1n}, z_{2n})\}$ be any sequence in $\ell^\infty(\mathbf{R}) \times$

$\ell^\infty(\mathbf{R})$ such that $z_n \to z$ relative to the product norm as $n \to \infty$. Take arbitrary sequence $t_n \to 0$ as $n \to \infty$. Write

$$\begin{aligned}
&t_n^{-1}[\phi(\theta + t_n z_n)(x) - \phi(\theta)(x)] \\
&\quad = t_n^{-1}\left[\max\{\theta^{(1)}(x) + t_n z_{1n}(x), \theta^{(2)}(x) + t_n z_{2n}(x)\} - \max\{\theta^{(1)}(x), \theta^{(2)}(x)\}\right] \\
&\quad = t_n^{-1}\max\{t_n z_{1n}(x), \theta^{(2)}(x) - \theta^{(1)}(x) + t_n z_{2n}(x)\}1_{B_1}(x) + \max\{z_{1n}(x), z_{2n}(x)\}1_{B_0}(x) \\
&\qquad + t_n^{-1}\max\{\theta^{(1)}(x) - \theta^{(2)}(x) + t_n z_{1n}(x), t_n z_{2n}(x)\}1_{B_2}(x)\ .
\end{aligned}$$

Consider the first term. Since $t_n = o(1)$ and $z_{1n} = z_{2n} = O(1)$, for all n sufficiently large we must have

$$\max\{t_n z_{1n}(x), \theta^{(2)}(x) - \theta^{(1)}(x) + t_n z_{2n}(x)\}1_{B_1}(x) = t_n z_{1n}(x)1_{B_1}(x)$$

uniformly in $x \in \mathbf{R}$, imply that

$$t_n^{-1}\max\{t_n z_{1n}(x), \theta^{(2)}(x) - \theta^{(1)}(x) + t_n z_{2n}(x)\}1_{B_1}(x) \to z^{(1)}1_{B_1}(x)$$

uniformly in x. The third term can be handled similarly while the second term is immediate.

■

Lemma 2.6.9. *Suppose that $\mathbb{H}$ is a separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$. Let $\{h_j\}_{j=1}^\infty$ be a complete sequence in $\mathbb{H}$ and $\mathbb{M} \subset \mathbb{H}$ a closed subspace. Let Π be the orthogonal projection onto $\mathbb{M}$. Then $\{\Pi h_j\}_{j=1}^\infty$ is complete in $\mathbb{M}$.*

PROOF: Fix $\epsilon > 0$ and $h \in \mathbb{M}$. Then by completeness of $\{h_j\}$ in $\mathbb{H}$, there exists $\lambda_1, \ldots, \lambda_n$ such that $\|h - (\lambda_1 h_1 + \cdots + \lambda_n h_n)\| < \epsilon$. It follows that

$$\begin{aligned}
\|h - (\lambda_1 \Pi h_1 + \cdots + \lambda_n \Pi h_n)\| &= \|\Pi h - (\lambda_1 \Pi h_1 + \cdots + \lambda_n \Pi h_n)\| \\
&= \|\Pi(h - (\lambda_1 h_1 + \cdots + \lambda_n h_n))\| \le \|\Pi\|_{op}\|h - (\lambda_1 h_1 + \cdots + \lambda_n h_n)\| < \epsilon\ ,
\end{aligned}$$

where the second inequality follows from $\|\Pi\|_{op} = 1$ by Conway (1990, Proposition 3.3). We thus establish the Lemma. ∎

Example 2.2.4 (Quantile Curves without Crossing)

Let $\mathbb{D} = L^2(\mathcal{T}, \nu)$ where $\mathcal{T} = [\epsilon, 1-\epsilon]$ with $0 < \epsilon < 1/2$ and ν the Lebesgue measure on $\mathcal{T}$. The set Λ of (weakly) increasing functions in $\mathbb{D}$ can be formalized as follows. As standard in L^p spaces, we consider two functions in $L^2(\mathcal{T})$ to define the same element when they are equal almost everywhere. We therefore say that $f \in L^2(\mathcal{T})$ is ν-monotone or simply monotone, if there exists a monotonic function $g : \mathcal{T} \to \mathbf{R}$ such that

$$\nu(\{t \in \mathcal{T} : f(t) \neq g(t)\}) = 0 \ .$$

We then define Λ to be the set of ν-monotone functions in $L^2(\mathcal{T})$. We first show that Λ is closed and convex so that the metric projection exists and is singleton valued.

Lemma 2.6.10. *Let $\Lambda \subset L^2(\mathcal{T})$ be the set of increasing functions. Then Λ is convex and closed.*

Proof: Suppose that $f_1, f_2 \in \Lambda$. Then there exist increasing functions g_1 and g_2 such that $f_i = g_i$ almost everywhere. Since for any $a \in [0,1]$, $af_1 + (1-a)f_2 = ag_1 + (1-a)g_2$ almost everywhere, and $ag_1+(1-a)g_2$ is clearly increasing, we thus conclude that $af_1+(1-a)f_2 \in \Lambda$ and thus Λ is convex. Now take a sequence $\{f_n\} \subset \Lambda$ such that $\|f_n - f\|_{L^2} \to 0$ as $n \to \infty$. By passing to a subsequence if necessary, we may assume that $f_n \to f \in L^2(\mathcal{T})$ almost everywhere as $n \to \infty$. Since $f_n \in \Lambda$, there is an increasing function $g_n \in \Lambda$ such that $g_n = f_n$ almost everywhere. It follows that $g_n \to f$ almost everywhere. Next define, for each $t \in \mathcal{T}$,

$$\bar{f}(t) \equiv \limsup_{n\to\infty} g_n(t) \ .$$

Then $\bar{f} = f$ almost everywhere. Pick any $s, t \in \mathcal{T}$ with $s < t$, we have

$$\bar{f}(s) = \limsup_{n\to\infty} g_n(s) \leq \limsup_{n\to\infty} g_n(t) = \bar{f}(t) \ .$$

Thus f is increasing, implying that $f \in \Lambda$ and hence Λ is closed. ∎

We note that if $f \in L^2(\mathcal{T})$ is monotonically increasing on $\mathcal{T}$ except a Lebesgue null set say $E_0 \subset \mathcal{T}$, then there must exist an $\tilde{f}$ such that $\tilde{f} = f$ almost everywhere and $\tilde{f}$ is increasing everywhere on $\mathcal{T}$, meaning that $f \in \Lambda$. Specifically, we construct $\tilde{f}$ as follows:

$$\tilde{f}(t) \equiv \begin{cases} f(t) & \text{if } t \in E_1 \\ \lim_{n\to\infty} f(t_n) & \text{if } t \in E_0 \end{cases},$$

where $E_1 \equiv \mathcal{T}\backslash E_0$, and $\{t_n\} \subset E_1$ is any sequence satisfying $t_n \downarrow t$ as $n \to \infty$. Such a sequence exists because otherwise there exists a ball $B_t(r) \equiv \{t' \in \mathcal{T} : |t' - t| \leq r\}$ for some $r > 0$ such that $B_t(r) \cap E_1 = \emptyset$ and hence $B_t(r) \subset E_0$, which is impossible since then $\nu(E_0) \geq \nu(B_t(r)) > 0$. Now it is straightforward to verify that $\tilde{f}$ is increasing on the whole domain $\mathcal{T}$. One important implication out of this is that if $f \notin \Lambda$, then there exists a Lebesgue measurable set E with $\nu(E) > 0$ such that $f(s) > f(t)$ whenever $s, t \in E$ satisfy $s < t$.

We next proceed to establish the directional differentiability of metric projection onto Λ at nonboundary points. There are couple of sufficient regularity conditions in the literature towards this end. In present case, working with polyhedricity (Haraux, 1977) is easier for us.

Lemma 2.6.11. *Let $\mathbb{D} = L^2(\mathcal{T})$ and Λ the set of (weakly) increasing functions in $\mathbb{D}$. Then the projection Π_Λ is Hadamard directionally differentiable at any $\theta \in \mathbb{D}$ and the resulting derivative evaluated at $z \in \mathbb{D}$ is given by $\Pi_{C_\theta}(z)$, where*

$$C_\theta = T_{\Pi_\Lambda \theta} \cap [\theta - \Pi_\Lambda \theta]^\perp .$$

Proof: By Haraux (1977), it suffices to show that Λ is polyhedric – i.e.

$$\overline{(\Lambda + [\Pi_\Lambda \theta]) \cap [\theta - \Pi_\Lambda \theta]^\perp} = \overline{\Lambda + [\Pi_\Lambda \theta]} \cap [\theta - \Pi_\Lambda \theta]^\perp . \tag{2.96}$$

In turn, polyhedricity (2.96) is immediate if we can show that $\Lambda + [\Pi_\Lambda \theta]$ is closed. To this end, consider a sequence $\{f_n\} \subset \Lambda + [\Pi_\Lambda \theta]$ such that $\|f_n - f\|_{L^2} \to 0$ for some $f \in L^2(\mathcal{T})$.

We want to show that $f \in \Lambda + [\Pi_\Lambda \theta]$.

Let $\bar{\theta} = \Pi_\Lambda \theta$. Without loss of generality we may assume that $f_n = \lambda_n + a_n \bar{\theta}$ where $\lambda_n \in \Lambda$ is an increasing function for each $n \in \mathbf{N}$. If $\{a_n\}$ is bounded, then by passing to a subsequence if necessary we may assume that $a_n \to a \in \mathbf{R}$ as $n \to \infty$. This implies that $\lambda_n = f_n - a_n \bar{\theta} \to \lambda \equiv f - a\bar{\theta}$ in L^2 as $n \to \infty$. Since Λ is closed, we have $\lambda \in \Lambda$ and hence $f = \lambda + a\bar{\theta} \in \Lambda + [\Pi_\Lambda \theta]$. For unbounded $\{a_n\}$, by passing to a subsequence if necessary, first consider the case when $a_n \uparrow \infty$ with $a_n > 0$ for all $n \in \mathbf{N}$. Then $f_n = \lambda_n + a_n \bar{\theta} \in \Lambda$ for each $n \in \mathbf{N}$ since Λ is a convex cone. This immediately implies that $f \in \Lambda$ since Λ is closed and hence $f \in \Lambda + [\Pi_\Lambda \theta]$.

It remains to consider the case where $f_n = \lambda_n - a_n \bar{\theta}$ where $a_n \uparrow \infty$ and $a_n > 0$ for all $n \in \mathbf{N}$. Suppose that $f \notin \Lambda + [\Pi_\Lambda \theta]$. Then $f + a\bar{\theta}$ is not increasing for all $a \in \mathbf{R}$. In particular, $f + a_n \bar{\theta} + a\bar{\theta}$ is not increasing for each n and $a > 0$ – i.e. for each $n \in \mathbf{N}$, there is a subset $E_n \subset \mathcal{T}$ with $\nu(E_n) > 0$ such that for all $s, t \in E_n$ with $s < t$ we have

$$f(s) + a_n \bar{\theta}(s) + a\bar{\theta}(s) > f(t) + a_n \bar{\theta}(t) + a\bar{\theta}(t) \ . \tag{2.97}$$

Since $\|f_n - f\|_{L^2} \to 0$, by passing to a subsequence if necessary we may assume that $f_n \to f$ almost everywhere on $\mathcal{T}$ as $n \to \infty$. By Egoroff's theorem (Saks, 1937, p.19), we may write $\mathcal{T} = \bigcup_{j=0}^{\infty} F_j$ where $F_0, F_1, F_2, \ldots$ are Lebesgue measurable sets such that $\nu(F_0) = 0$, and $f_n \to f$ uniformly on each F_j for $j = 1, 2, \ldots$. Let $\tilde{E}_n = E_n \backslash F_0$ for all $n \in \mathbf{N}$. We claim that $\tilde{E}_n \supset \tilde{E}_{n+1}$ for each $n \in \mathbf{N}$. To see this, pick $s, t \in \tilde{E}_{n+1}$ with $s < t$ such that

$$f(s) + a_{n+1} \bar{\theta}(s) + a\bar{\theta}(s) > f(t) + a_{n+1} \bar{\theta}(t) + a\bar{\theta}(t) \ . \tag{2.98}$$

It follows that

$$\begin{aligned}
f(s) + a_n\bar{\theta}(s) + a\bar{\theta}(s) &= f(s) + a_{n+1}\bar{\theta}(s) + a\bar{\theta}(s) + (a_n - a_{n+1})\bar{\theta}(s) \\
&> f(t) + a_{n+1}\bar{\theta}(t) + a\bar{\theta}(t) + (a_n - a_{n+1})\bar{\theta}(s) \\
&\geq f(t) + a_{n+1}\bar{\theta}(t) + a\bar{\theta}(t) + (a_n - a_{n+1})\bar{\theta}(t) \\
&= f(t) + a_n\bar{\theta}(t) + a\bar{\theta}(t) \ ,
\end{aligned}$$

where the first inequality is by (2.97), and the second is due to the facts that $a_n \leq a_{n+1}$ and that $\bar{\theta}(s) < \bar{\theta}(t)$ by $\bar{\theta} \in \Lambda$. Clearly, $f_n \to f$ everywhere as $n \to \infty$ on $\tilde{E}_1$.

To begin with, note that if there exist $s, t \in \tilde{E}_n$ with $s < t$ for some $n \in \mathbf{N}$ such that $\bar{\theta}(s) = \bar{\theta}(t)$, then by (2.97) it must be the case that $f(s) > f(t)$. Since $f_n + \alpha_n\bar{\theta}$ is monotonically increasing, $s < t$ and $\bar{\theta}(s) = \bar{\theta}(t)$, it follows that $f_n(s) \leq f_n(t)$ for all $n \in \mathbf{N}$ and hence $f(s) \leq f(t)$ by letting $n \to \infty$, a contradiction. Therefore, we may assume without loss of generality that $\bar{\theta}(s) < \bar{\theta}(t)$ for any $s, t \in \tilde{E}_1$ with $s < t$.

We further claim that $\nu(\tilde{E}_n) \downarrow 0$ as $n \to \infty$. To see this, pick $s_1, t_1 \in \tilde{E}_1$ with $s_1 < t_1$ such that (2.97) holds. We then have

$$\begin{aligned}
&[f(s_1) + a_n\bar{\theta}(s_1) + a\bar{\theta}(s_1)] - [f(t_1) + a_n\bar{\theta}(t_1) + a\bar{\theta}(t_1)] \\
&\qquad = [f(s_1) + a\bar{\theta}(s_1)] - [f(t_1) + a\bar{\theta}(t_1)] + a_n[\bar{\theta}(s_1) - \bar{\theta}(t_1)] \\
&\qquad \to -\infty < 0 \ , \quad \text{as } n \to \infty \ .
\end{aligned}$$

Thus one of $\{s_1, t_1\}$ is not in $\bigcap_{n=1}^{\infty} \tilde{E}_n$. Continuing in this fashion, we end up with $\bigcap_{n=1}^{\infty} \tilde{E}_n$ consisting of a singleton and hence $\nu(\tilde{E}_n) \downarrow 0$ as $n \to \infty$. Since $\tilde{E}_1 \subset \bigcup_{j=1}^{\infty} F_j$, it follows that $\tilde{E}_{n_0} \subset \bigcup_{j=1}^{J} F_j$ for some n_0 and J large enough, implying that $f_n \to f$ uniformly on $\tilde{E}_{n_0}$. Thus, for all sufficiently large $n \geq n_0$ where n_0 doesn't depend on a,

$$f_n(s) + \epsilon + a_n\bar{\theta}(s) + a\bar{\theta}(s) > f_n(t) - \epsilon + a_n\bar{\theta}(t) + a\bar{\theta}(t) \ ,$$

or,

$$\lambda_n(s) + 2\epsilon + a[\bar{\theta}(s) - \bar{\theta}(t)] > \lambda_n(t) \ , \tag{2.99}$$

for $s, t \in E_n \backslash F_0$ with $s < t$. Since $\bar{\theta}(s) - \bar{\theta}(t) < 0$, by choosing $a > 0$ such that $2\epsilon + a[\bar{\theta}(s) - \bar{\theta}(t)] = 0$, we may conclude that $\lambda_n(s) > \lambda_n(t)$ for all sufficiently large $n \geq n_0$, reaching a contradiction. Hence we must have $f \in \Lambda + [\Pi_\Lambda \theta]$, meaning that $\Lambda + [\Pi_\Lambda \theta]$ is closed so that Λ is polyhedric. ∎

Chapter 3

Applications

Abstract

This chapter illustrates usefulness of the asymptotic framework developed in previous chapters. Section 3.1 constructs a test of whether a Hilbert space valued parameter belongs to a known closed convex set – a setting that includes moment inequality problems and certain tests of shape restrictions as special cases. Section 3.2 presents a new uniform limit theory for the Grenander estimator under minimal assumptions – in particular without strict concavity or bounded support. Our insight builds on the fact that the least concave majorant operator is a Hadamard directionally differentiable map.

3.1 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.

3.1.1 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

$$H_0 : \theta_0 \in \Lambda \qquad\qquad H_1 : \theta_0 \notin \Lambda \ , \tag{3.1}$$

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 3.1.1-3.1.3 below.

We formalize the introduced structure through the following assumption.

Assumption 3.1.1. (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

$$\|\theta - \Pi_\Lambda \theta\|_{\mathbb{H}} = \inf_{h \in \Lambda} \|\theta - h\|_{\mathbb{H}} \ . \tag{3.2}$$

Thus, the hypothesis testing problem in (3.1) 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.

$$H_0 : \|\theta_0 - \Pi_\Lambda \theta_0\|_{\mathbb{H}} = 0 \qquad\qquad H_1 : \|\theta_0 - \Pi_\Lambda \theta_0\|_{\mathbb{H}} > 0 \ . \tag{3.3}$$

Interpreted in this manner, it is clear that (3.3) is a special case of (1.55), with $\mathbb{D} = \mathbb{H}$,

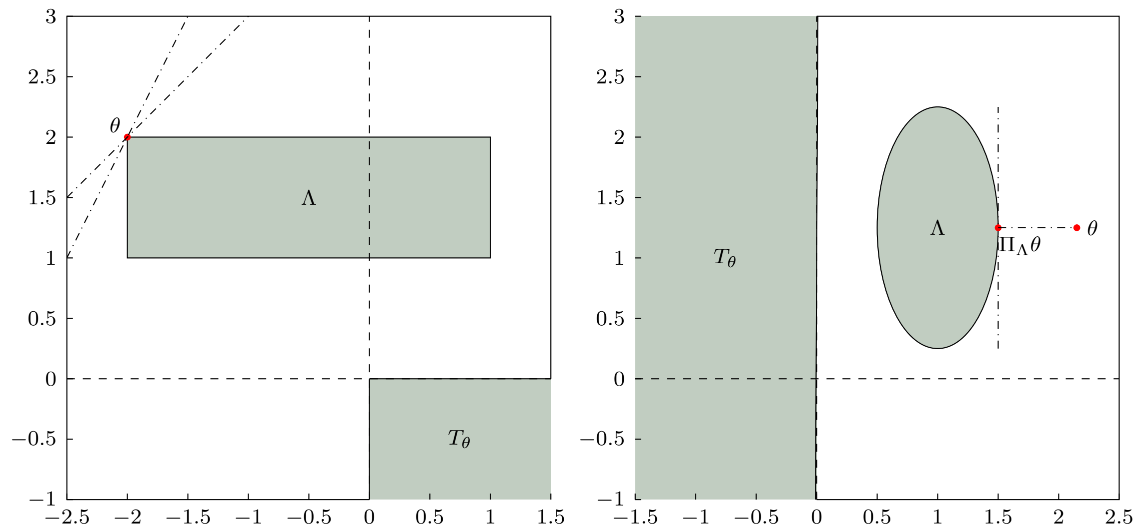

Figure 3.1: Illustrations of Tangent Cones.

$\mathbb{E} = \mathbf{R}$, 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. $r_n\phi(\hat{\theta}_n) = r_n\|\hat{\theta}_n - \Pi_\Lambda\hat{\theta}_n\|_{\mathbb{H}}$.

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

$$T_\theta \equiv \overline{\bigcup_{\alpha \geq 0} \alpha\{\Lambda - \Pi_\Lambda\theta\}}\ , \tag{3.4}$$

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 3.1 illustrates the tangent cone in two separate cases: one in which $\theta \in \Lambda$, and a second in which $\theta \notin \Lambda$.

3.1.1.1 Examples

In order to aid exposition and illustrate the generality of (3.1), we next provide examples of both well studied and new problems that fit our framework.

Example 3.1.1. Suppose $X \in \mathbf{R}^d$ and that we aim to test the moment inequalities

$$H_0 : E[X] \leq 0 \qquad\qquad H_1 : E[X] \not\leq 0 \ , \tag{3.5}$$

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

$$\phi(\theta) = \|\Pi_\Lambda \theta - \theta\|_{\mathbb{H}} = \Big\{ \sum_{i=1}^{d} (E[X^{(i)}])_+^2 \Big\}^{\frac{1}{2}} \ , \tag{3.6}$$

where $(a)_+ = \max\{a, 0\}$ and $X^{(i)}$ denotes the i^{th} coordinate of X. More generally, the hypothesis in (3.3) accommodates any regular parameter and any closed convex set in $\mathbf{R}^d$, such as the test for moment inequalities on regression coefficients proposed by Wolak (1988) and the test of random utility models developed in Kitamura and Stoye (2013). Analogously, conditional moment inequalities as in Example 1.2.3 can also 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). ■

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

$$(\theta_0(\tau), \beta(\tau)) \equiv \arg \min_{\theta \in \mathbf{R}, \beta \in \mathbf{R}^{d_z}} E[\rho_\tau(Y - D\theta - Z'\beta)] \tag{3.7}$$

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).[1] Hence, we may test shape restrictions on θ_0 by letting

$$\mathbb{H} \equiv \{\theta : [\epsilon, 1-\epsilon] \to \mathbf{R} : \langle \theta, \theta \rangle_{\mathbb{H}} < \infty\} \qquad \langle \theta_1, \theta_2 \rangle_{\mathbb{H}} \equiv \int_{\epsilon}^{1-\epsilon} \theta_1(\tau)\theta_2(\tau)d\tau \ , \tag{3.8}$$

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). ■

Our final example may be interpreted as a generalization of Example 3.1.2.

Example 3.1.3. 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

$$E[\rho(Z, \theta_0(\tau), \tau)] = 0 \ . \tag{3.9}$$

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) or can be described by a functional lower and upper bound (Kline and Santos, 2013; Chandrasekhar et al., 2012). 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

$$\mathbb{H} \equiv \{\theta : \mathcal{T} \to \mathbf{R}^{d_\theta} : \langle \theta, \theta \rangle_{\mathbb{H}} < \infty\} \qquad \langle \theta_1, \theta_2 \rangle_{\mathbb{H}} \equiv \int_{\mathcal{T}} \theta_1(\tau)'\theta_2(\tau)w(\tau)d\tau \ . \tag{3.10}$$

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. ■

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

3.1.2 Theoretical Results

3.1.2.1 Asymptotic Distribution

Our analysis crucially relies on the seminal work of Zarantonello (1971), who established the Hadamard directional differentiability of metric projections onto convex sets in Hilbert spaces. Specifically, Zarantonello (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 3.2 illustrates a simple example in which the derivative approximation

$$\Pi_\Lambda \theta_1 - \Pi_\Lambda \theta_0 \approx \Pi_{T_{\theta_0}}(\theta_1 - \theta_0) \tag{3.11}$$

actually holds with equality.[2] 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 Zarantonello (1971), the asymptotic distribution of $r_n \phi(\hat{\theta}_n)$ can then be obtained as an immediate consequence of Theorem 1.2.1.

Proposition 3.1.1. *Let Assumption 1.2.2 and 3.1.1 hold. If $\theta_0 \in \Lambda$, then it follows that*

$$r_n \|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \xrightarrow{L} \|\mathbb{G}_0 - \Pi_{T_{\theta_0}} \mathbb{G}_0\|_{\mathbb{H}} \ . \tag{3.12}$$

In particular, Proposition 3.1.1 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

$$\phi'_{\theta_0}(h) = \|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} \ . \tag{3.13}$$

It is interesting to note that $\Lambda \subseteq T_{\theta_0}$ whenever Λ is a cone, and hence $\|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} \leq \|h - \Pi_\Lambda h\|_{\mathbb{H}}$ for all $h \in \mathbb{H}$. Therefore, the distribution of $\|\mathbb{G}_0 - \Pi_\Lambda \mathbb{G}_0\|_{\mathbb{H}}$ first order stochastically dominates that of $\|\mathbb{G}_0 - \Pi_{T_{\theta_0}} \mathbb{G}_0\|_{\mathbb{H}}$, and by Proposition 3.1.1 its quantiles may be employed

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

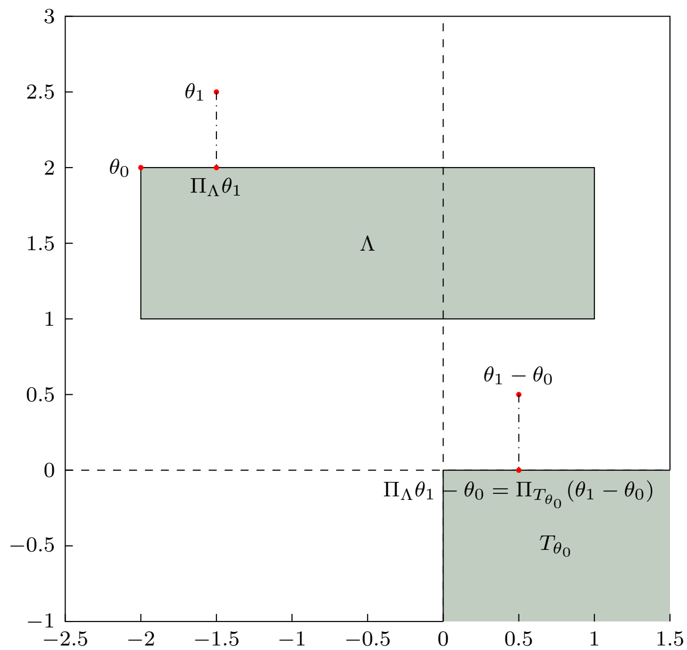

Figure 3.2: Directional Differentiability

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 3.1.1 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 3.1.1 and 3.1.2.

Remark 3.1.1. Zarantonello (1971) and Theorem 1.2.1 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 3.1.2. While we do not consider it for simplicity, it is straightforward to incorporate weight functions into the test statistic.[3] 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

[3]For instance in (3.6) we may wish to consider $\{\sum_{i=1}^{d}(E[X^{(i)}])_+^2/Var(X^{(i)})\}^{\frac{1}{2}}$ instead.

$\|\cdot\|_o$ the operator norm, we obtain by asymptotic tightness of $r_n\{\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\}$ that

$$r_n \|\hat{A}_n\{\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\}\|_{\mathbb{H}} \xrightarrow{L} \|A\{\mathbb{G}_0 - \Pi_{T_{\theta_0}} \mathbb{G}_0\}\|_{\mathbb{H}} . \tag{3.14}$$

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

3.1.2.2 Critical Values

In order to construct critical values to conduct inference, we next aim to employ Theorem 1.3.3, 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 1.3.3 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

$$\hat{\phi}'_n(h) \equiv \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}} . \tag{3.15}$$

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 1.3.3.

Proposition 3.1.2. *Let Assumptions 1.2.2, 3.1.1 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 (3.15) *satisfies Assumption 1.3.3.*

(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 1.3.3. Therefore, provided the bootstrap is consistent for the asymptotic distribution of $r_n\{\hat{\theta}_n - \theta_0\}$, Theorem 1.3.3 implies $\hat{\phi}'_n$ can be employed to construct critical values. We note that Proposition 3.1.2(i) holds irrespective of whether the null hypothesis is satisfied, which readily implies the consistency of the corresponding test.[4] In turn, Proposition 3.1.2(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 1.3.4 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 (3.15) is employed to construct critical values. Hence, we emphasize that while $\hat{\phi}'_n$ as defined in (3.15) 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 3.1.3.

Remark 3.1.3. 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 3.1.1,

$$T_{\theta_0} = \{h \in \mathbf{R}^d : h^{(i)} \leq 0 \text{ for all } i \text{ such that } E[X^{(i)}] = 0\} \ . \tag{3.16}$$

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

$$\hat{T}_n = \{h \in \mathbf{R}^d : h^{(i)} \leq 0 \text{ for all } i \text{ such that } \bar{X}^{(i)} \geq -\epsilon_n\} \tag{3.17}$$

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 1.3.3 (compare to (3.13)) and, more interestingly, that the bootstrap procedure of Theorem 1.3.3 then reduces to the generalized moment selection approach of Andrews and Soares (2010). ■

3.1.3 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 3.1.2. 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

[4] 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 $\theta_0 \in \Lambda$.

Table 3.1: Empirical Size

Bandwidth		$n = 200$								
		$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.01$		
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

$(Z^{(1)}, Z^{(2)})' \sim N(0, I)$ for I the identity matrix, and Y is related by

$$Y = \frac{\Delta}{\sqrt{n}} D \times U + Z'\beta + U \ , \tag{3.18}$$

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

$$P(Y \leq D\theta_0(\tau) + Z'\beta(\tau)|D, Z) = \tau \ , \tag{3.19}$$

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 (3.1) 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,

Table 3.2: 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

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 (3.15) with $\epsilon_n = Cn^\kappa$ for different choices of C and κ. All reported results are based on five thousand Monte Carlo replications.

Table 3.1 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 monotonically 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 1.3.4. Throughout the specifications, we see that the test effectively controls size, and Theorem 1.3.4 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 3.2, 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.

3.2 Global Limit Theory for the Grenander Estimator under Potentially Nonstrict Concavity

3.2.1 Introduction

Nonparametric estimation under shape constraints such as monotonicity, concavity and log-concavity has received increasing attention in recent years. Groeneboom and Jongbloed (2014) provide a helpful introduction to the current state of the field. As pointed out by Walther (2009), nonparametric estimation under shape constraints is attractive for two main reasons: (1) shape constraints are often implied by theoretical models or are at least plausible assumptions, and (2) nonparametric estimation under shape constraints is often feasible without the use of tuning parameters, as opposed to classical kernel or series estimators. Perhaps the best known shape constrained estimator is the Grenander estimator of a concave distribution function or nonincreasing density function. Grenander (1956) showed that, given a random sample drawn from a nonincreasing probability density, the left-derivative of the least concave majorant (lcm) of the empirical distribution function achieves the maximum likelihood among all nonincreasing densities.

In this paper we provide some new results concerning the asymptotic behavior of the Grenander estimator. Quite a lot is known already. Denote by $\hat{f}_n$ the Grenander estimator of a nonincreasing density f based on a sample of size n with empirical distribution $\mathbb{F}_n$, and by $\hat{\mathbb{F}}_n$ the Grenander estimator of the concave distribution function F, i.e. the lcm of $\mathbb{F}_n$. The pointwise asymptotic distribution of $\hat{f}_n$ was obtained by Rao (1969) at points where f

is strictly decreasing, and by Carolan and Dykstra (1999) at points where f is flat. The rate of convergence is $n^{1/3}$ in the former case and $n^{1/2}$ in the latter, and the limit distribution is non-Gaussian. Results on the global asymptotic behavior of $\hat{f}_n$ – specifically, of its L_p-risk – are provided by Groenenboom (1984), Groeneboom et al. (1999), Kulikov and Lopuhaä (2005), Durot (2007) and Durot et al. (2012). These global results require f to be strictly decreasing on its support. Kulikov and Lopuhaä (2006a) study the behavior of $\hat{f}_n$ near the boundary of its support.

Turning to the asymptotic behavior of $\hat{\mathbb{F}}_n$, it is natural to consider weak convergence of the Grenander empirical process $\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{F}}_n - F)$. A result of Kiefer and Wolfowitz (1976) implies that $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ are asymptotically equivalent when F satisfies strict concavity on its support, so that $\hat{\mathbb{G}}_n$ converges weakly to $\mathbb{G} = \mathbb{B} \circ F$, a Brownian bridge $\mathbb{B}$ composed with F. Other results on the behavior of $\hat{\mathbb{F}}_n - \mathbb{F}_n$ when F is strictly concave on its support have been provided by Wang (1994) and Kulikov and Lopuhaä (2006b, 2008). On the other hand, when F is the uniform distribution on the unit interval it is known that $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ are not asymptotically equivalent, and $\hat{\mathbb{G}}_n$ converges weakly to the lcm of $\mathbb{B}$, a process studied in detail by Carolan and Dykstra (2001). Carolan (2002) considers the more general case where F is affine over some maximal subinterval $[a, b]$ of its support, and shows that the restriction of $\hat{\mathbb{G}}_n$ to $[a, b]$ converges weakly to the lcm of the restriction of $\mathbb{G}$ to $[a, b]$.

Our first main result, Theorem 3.2.1 below, establishes the weak convergence of $\hat{\mathbb{G}}_n$ in the intermediate cases where F is concave but not necessarily strictly concave or uniform. As might be guessed from the results of Carolan (2002), the weak limit $\hat{\mathbb{G}}$ can be obtained by taking lcms of $\mathbb{G}$ over the distinct intervals on which F is affine. We do not impose bounded support or other technical conditions on F. Our proof exploits the fact that the lcm operator is Hadamard directionally differentiable (see Definition 3.2.2 below) despite not being fully Hadamard differentiable. This provides enough structure to invoke the Delta method (Shapiro, 1991; Dümbgen, 1993) and in this way derive the weak limit of $\hat{\mathbb{G}}_n$. Our result applies not only to the Grenander estimator of F, but to any estimator of F obtained by taking the lcm of an estimator $\mathbb{F}_n$ of F for which $\sqrt{n}(\mathbb{F}_n - F)$ converges weakly to a continuous process $\mathbb{G}$ vanishing at the boundaries 0 and ∞. Thus, for instance, we may take $\mathbb{F}_n$ to be a smoothed estimate of F, as in Eggermont and LaRiccia (2000), and we may allow the data used to construct $\mathbb{F}_n$ to exhibit limited serial dependence.

Our weak convergence result for $\hat{\mathbb{G}}_n$ can be useful, for example, when constructing uniform confidence bands for F. However, doing so requires consistent estimation of the weak limit $\hat{\mathbb{G}}$, which turns out to be nontrivial. The fact that the lcm operator fails to be fully Hadamard differentiable implies that the bootstrap does not produce consistent estimates of the law of $\hat{\mathbb{G}}$ (Fang and Santos, 2014). In other words, the Delta method generalizes under Hadamard directional differentiability to obtaining the weak limit of $\hat{\mathbb{G}}_n$ but not to obtaining bootstrap consistency. Our second main result, Theorem 3.2.2 below, shows how a modified bootstrap procedure may be used to approximate the law of $\hat{\mathbb{G}}$. Exploiting the fact that $\hat{\mathbb{G}}$ can be written as the composition of the Hadamard directional derivative of the lcm operator and a process $\mathbb{G}$ that can be consistently bootstrapped, we follow Fang and Santos (2014) and compose a suitable estimator of the derivative with a bootstrapped version of $\mathbb{G}$. The estimated derivative is obtained using a numerical differentiation technique along the lines pursued by Hong and Li (2014). We show how our modified bootstrap procedure may be used to construct valid confidence bands for F.

The rest of the paper is structured as follows. Section 3.2.2 establishes the uniform weak convergence of the Grenander distribution function – i.e. the least concave majorant of the empirical distribution function. Section 3.2.3 presents a consistent bootstrap for estimating the law of the weak limit derived in Section 3.2.2. Section 3.2.4 concludes. All proofs are collected in the Appendix.

Before proceeding further we introduce some additional notation. We denote by $C_0(\mathbf{R}^+)$ the set of continuous, real valued functions on $\mathbf{R}^+$ that vanish at 0 and ∞. For a convex set $T \subseteq \bar{\mathbf{R}}^+$, we let $\ell_c^\infty(T)$ denote the set of uniformly bounded, concave, real valued functions on T, and let $C_{bc}(T)$ denote the set of continuous, uniformly bounded, concave, real valued functions on T. The sets $\ell^\infty(T)$, $C_b(\mathbf{R}^+)$, $C_0(\mathbf{R}^+)$, $\ell_c^\infty(T)$ and $C_{bc}(T)$ should be understood to be equipped with the uniform norm $\|\cdot\|_\infty$, as appropriate. We let $\xrightarrow{L}$ denote weak convergence in the Hoffmann-Jørgensen sense (van der Vaart and Wellner, 1996). Finally, if T is a set equipped with a metric d, then we let $\mathrm{BL}_1(T)$ denote the set of real valued functions on T whose level and Lipschitz constant are bounded by one – i.e.,

$$\mathrm{BL}_1(T) = \left\{ f : T \to \mathbf{R} \,\middle|\, |f(t_1)| \leq 1 \text{ and } |f(t_1) - f(t_2)| \leq d(t_1, t_2) \text{ for all } t_1, t_2 \in T \right\}.$$

We write BL_1 as shorthand for $\mathrm{BL}_1(\ell^\infty(\mathbf{R}^+))$.

3.2.2 Weak Convergence

The distribution function F to be estimated is taken to be that of a nonnegative random variable. We treat it as a real valued map on $\mathbf{R}^+$ and maintain throughout that it satisfies the following condition.

Assumption 3.2.1. $F : \mathbf{R}^+ \to \mathbf{R}$ is a continuous concave distribution function on $\mathbf{R}^+$ with $F(0) = 0$.

No further technical conditions will be imposed on F. To maintain generality about the underlying sampling scheme and method of estimation, we suppose the existence of a sequence $\{\mathbb{F}_n\}_{n=1}^\infty$ as maps from the data $\{X_i\}_{i=1}^n$ into $\ell^\infty(\mathbf{R}^+)$ satisfying the following high level condition, in which $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{F}_n - F)$.

Assumption 3.2.2. $\mathbb{G}_n \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathbf{R}^+)$ for some tight random element $\mathbb{G}$ of $C_0(\mathbf{R}^+)$.

If $\mathbb{F}_n$ is the empirical distribution function of an independent and identically distributed (iid) sample of size n drawn from F then $\mathbb{G}_n$ is the usual empirical process and clearly Assumption 3.2.2 is satisfied with $\mathbb{G} = \mathbb{G}_\lambda \circ F$ and $\mathbb{G}_\lambda$ the standard Brownian bridge. More generally, we may allow the sample drawn from F to satisfy a mixing condition or related property (Dehling and Philipp, 2002), or we may take $\mathbb{F}_n$ to be a smoothed empirical distribution function (van der Vaart, 1994) or some other estimator satisfying Assumption 3.2.2 under suitable regularity conditions.

To exploit the concavity of F we propose using the estimator $\hat{\mathbb{F}}_n = \mathcal{M}\mathbb{F}_n$, where $\mathcal{M}$ is the lcm operator. If $\mathbb{F}_n$ is the empirical distribution function of an iid sample drawn from F, then the left-derivative of $\hat{\mathbb{F}}_n$ is the classical Grenander estimator of the probability density for F. The following definition of $\mathcal{M}$ is adapted from Beare and Moon (2015).

Definition 3.2.1. Given a convex set $T \subseteq \mathbf{R}^+$, the lcm over T is the operator $\mathcal{M}_T : \ell^\infty(\mathbf{R}^+) \to \ell^\infty(T)$ that maps each $\theta \in \ell^\infty(\mathbf{R}^+)$ to the function

$$\mathcal{M}_T\theta(x) = \inf\{g(x) : g \in \ell_c^\infty(T) \text{ and } \theta \leq g \text{ on } T\}, \quad x \in T.$$

We write $\mathcal{M}$ as shorthand for $\mathcal{M}_{\mathbf{R}^+}$ and refer to $\mathcal{M}$ as the lcm operator.

The definition of $\mathcal{M}_T$ given here differs from that of Beare and Moon (2015) only in that those authors took the domain of $\mathcal{M}_T$ to be $\ell^\infty([0,1])$ and required T to be a closed subinterval of the unit interval. Clearly, the image of $\mathcal{M}_T$ is $\ell_c^\infty(T)$. Other well known properties of $\mathcal{M}_T$ include monotonicity, homogeneity of degree one, and contractivity (Marshall's lemma). Beare and Moon (2015) investigated the differential properties of $\mathcal{M}$ in order to study the asymptotic behavior of a test of the monotone density ratio property proposed by Carolan and Tebbs (2005). For an application of this test to address an empirical puzzle in the financial literature, see Beare and Schmidt (2015).

As pointed out by Beare and Moon (2015), $\mathcal{M}$ fails to be Hadamard differentiable and hence violates the assumptions made in standard treatments of the Delta method (van der Vaart and Wellner, 1996); however, $\mathcal{M}$ does satisfy a certain form of directional differentiability studied by Shapiro (1990). Quite remarkably, the Delta method is valid under this weaker notion of differentiability (Shapiro, 1991; Dümbgen, 1993).

Definition 3.2.2. Let $\mathbb{D}$ and $\mathbb{E}$ be normed spaces equipped with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$ respectively. A map $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$ is said to be *Hadamard directionally differentiable* at $\theta \in \mathbb{D}_\phi$ tangentially to a set $\mathbb{D}_0 \subset \mathbb{D}$ if there is a map $\phi'_\theta : \mathbb{D} \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \right\|_{\mathbb{E}} = 0 \tag{3.20}$$

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 with various notions of differentiability in the literature, Hadamard directional differentiability can be understood by looking at the restrictions imposed on the approximating map (i.e. the derivative) and the way the approximation error is controlled (Averbukh and Smolyanov, 1967, 1968). Specifically, let

$$\mathrm{Rem}_\theta(h) \equiv \phi(\theta + h) - \{\phi(\theta) + \phi'_\theta(h)\}, \tag{3.21}$$

where $\phi(\theta) + \phi'_\theta(h)$ can be viewed as the first order approximation of $\phi(\theta + h)$. Hadamard directional differentiability of ϕ then amounts to requiring the approximation error $\mathrm{Rem}_\theta(h)$ satisfy that $\mathrm{Rem}_\theta(th)/t$ tends to zero uniformly in $h \in K$ for any compact set K – i.e.,

$$\sup_{h\in K} \left\| \frac{\mathrm{Rem}_\theta(th)}{t} \right\|_{\mathbb{E}} \to 0 \quad \text{as} \quad t \downarrow 0.$$

However, unlike Hadamard differentiability that requires the approximating map ϕ'_θ to be linear and continuous, linearity of the directional counterpart is often lost, even though continuity is assured (Shapiro, 1990, Proposition 3.1). In fact, linearity of the derivative is the exact gap between these two notions of differentiability (Fang and Santos, 2014, Proposition 2.1).

Proposition 3.2.1. *The lcm operator $\mathcal{M} : \ell^\infty(\mathbf{R}^+) \to \ell^\infty(\mathbf{R}^+)$ is Hadamard directionally differentiable at any $\theta \in C_{bc}(\mathbf{R}^+)$ tangentially to $C_0(\mathbf{R}^+)$. Its directional derivative $\mathcal{M}'_\theta$: $C_0(\mathbf{R}^+) \to \ell^\infty(\mathbf{R}^+)$ is uniquely determined as follows: for any $h \in C_0(\mathbf{R}^+)$ and $x \in \mathbf{R}^+$, we have $\mathcal{M}'_\theta h(x) = \mathcal{M}_{T_{\theta,x}} h(x)$, where $T_{\theta,x}$ is the union of all convex subsets of $\mathbf{R}^+$ that contain x and over which θ is affine.*

Proposition 3.2.1 establishes Hadamard directional differentiability of the lcm operator. Similar to Lemma 3.2 in Beare and Moon (2015), the derivative $\mathcal{M}'_\theta$ is nonlinear in general and in fact linear if and only if θ is strictly concave, in which case $\mathcal{M}'_\theta$ is the identity map. If θ is affine in a neighborhood of x, the set $T_{\theta,x}$ is a either a closed interval $[a_{\theta,x}, b_{\theta,x}]$ or a half-line $[a_{\theta,x}, \infty)$. If θ is not affine in a neighborhood of x, we have $T_{\theta,x} = \{x\}$ and $\mathcal{M}'_\theta h(x) = h(x)$. The derivative $\mathcal{M}'_\theta$ therefore behaves like a hybrid of the lcm and identity operators: for any direction $h \in C_0(\mathbf{R}^+)$, it majorizes h by concave functions on regions over which θ is affine but acts like an identity map elsewhere.

Let $\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{F}}_n - F)$, the Grenander empirical process, and let $\hat{\mathbb{G}} = \mathcal{M}'_F\mathbb{G}$. With Hadamard directional differentiability of $\mathcal{M}$ in hand, we obtain our first main result by employing the Delta method.

Theorem 3.2.1. *Under Assumptions 3.2.1 and 3.2.2 we have* $\hat{\mathbb{G}}_n \xrightarrow{L} \hat{\mathbb{G}}$ *in* $\ell^\infty(\mathbf{R}^+)$*.*

Theorem 3.2.1 can be viewed as an extension of a result of Carolan (2002), who showed in the proof of his Theorem 5 that when F is affine over a maximal interval $[a, b] \subseteq \mathbf{R}^+$, the restriction of $\hat{\mathbb{G}}_n$ to $[a, b]$ converges weakly to the lcm of the restriction of $\mathbb{G}$ to $[a, b]$. Our result extends his to obtain weak convergence of the entire process $\hat{\mathbb{G}}_n$ even when F may have multiple affine segments separated by kinks, or by intervals over which it is strictly concave. Further, since our proof is an application of the Delta method, it is simple for us to consider general estimators $\mathbb{F}_n$ satisfying Assumption 3.2.2, whereas Carolan requires $\mathbb{F}_n$ to be the empirical distribution based on iid draws from F. Note that when F is strictly concave we have $\hat{\mathbb{G}} = \mathbb{G}$, and so from Theorem 3.2.1 we recover the asymptotic equivalence of $\hat{\mathbb{F}}_n$ and $\mathbb{F}_n$ implied by results of Kiefer and Wolfowitz (1976), but here under milder conditions.

3.2.3 The Bootstrap

Parallel to the level of generality adopted in our treatment of the estimator $\mathbb{F}_n$ in the previous section, we here maintain a high degree of generality with respect to the method used to obtain a bootstrap version $\mathbb{F}_n^*$ of $\mathbb{F}_n$. For example, for the standard nonparametric bootstrap,

$$\mathbb{F}_n^*(x) = \frac{1}{n}\sum_{i=1}^n 1\{X_i^* \leq x\} = \frac{1}{n}\sum_{i=1}^n W_{ni} 1\{X_i \leq x\} \ , \, x \in [0, \infty) \ ,$$

where $\{X_i^*\}_{i=1}^n$ are an i.i.d. sample from $\mathbb{F}_n$ conditional on the data $\{X_i\}_{i=1}^n$, and $(W_{n1}, \ldots, W_{nn})$ is a multinomial vector independent of $\{X_i\}_{i=1}^n$ with n categories and probabilities $(1/n, \ldots, 1/n)$. To rigorously formalize our discussion, we need to be careful about the measurability issues. Throughout, $\{X_n\}_{n=1}^\infty$ are defined as the coordinate projections on the first "∞" coordinates in the canonical probability space $(\prod_{n=1}^\infty \mathbf{R}^+ \times \mathcal{Z}, \prod_{n=1}^\infty \mathcal{B}(\mathbf{R}^+) \times \mathcal{C}, \prod_{n=1}^\infty P \times Q)$ where

$\mathcal{B}(\mathbf{R}^+)$ is the Borel σ-algebra on $\mathbf{R}^+$ and P is the probability measure associated with F, and the bootsrtap weights depend on the last factor only.

Given the above setup, one might expect $\sqrt{n}\{\mathcal{M}\mathbb{F}_n^* - \mathbb{F}_n\}$ to be a consistent estimate. Unfortunately, this would not work and necessarily produces inconsistent results (Fang and Santos, 2014). However, the form of the weak limit $\mathcal{M}'_F(\mathbb{G})$ in Theorem 3.2.1 suggests a solution of composing some bootstrapped approximation $\mathbb{G}_n^*$ of $\mathbb{G}$ with a suitable estimator $\hat{\mathcal{M}}'_n$ of the derivative $\mathcal{M}'_F$ as in Fang and Santos (2014). Bootstrapping $\mathbb{G}$ is easy and can be implemented, for example, by the nonparametric bootstrap. Estimating $\mathcal{M}'_F$ is tricky. The simple plug-in estimator $\mathcal{M}'_{\mathbb{F}_n}$ would not work, which is actually easy to see. Recall from Proposition 3.2.1 that $\mathcal{M}'_F$ is linear in h when h is strictly concave but otherwise nonlinear. Heuristically, there exist concave functions having affine sections arbitrarily close to any given strictly concave function, implying that local perturbations from a strictly concave function will result in dramatic changes of the derivatives. This type of "discontinuity" of the directional derivatives is closely related to the bootstrap inconsistency under nonsmooth transformations.

For the sake of bootstrap consistency, Fang and Santos (2014) impose the high level condition that an appropriate derivative estimator would converge in probability to the truth uniformly on any δ-enlargement of compact sets (see Assumption 3.3 in Fang and Santos (2014)). There are at least two ways to accomplish this. As in Seo (2014) and Beare and Shi (2015), we may employ a linear isometry approximation to represent the derivative $\mathcal{M}'_F$ as a supremum over a set – a set that is similar to the contact set in Linton et al. (2010). As such, estimating the derivative boils down to estimating the contact set. Alternatively, we may also follow the numerical approach proposed by Hong and Li (2014) which we adopt in this paper.

Define $\hat{\mathcal{M}}'_n : \ell^\infty(\mathbf{R}^+) \to \ell^\infty(\mathbf{R}^+)$ by

$$\hat{\mathcal{M}}'_n(h) = \frac{\mathcal{M}(\mathbb{F}_n + t_n h) - \mathcal{M}(\mathbb{F}_n)}{t_n} \ , \ h \in \ell^\infty(\mathbf{R}^+) \ , \tag{3.22}$$

where t_n is a sequence of positive scalars approaching zero. We now aim to show that

$\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})$ can consistently estimate the law of $\mathcal{M}'_F(\mathbb{G})$ where $\mathbb{F}^*_n : \{X_i, W_i\}_{i=1}^n \to \ell^\infty(\mathbf{R}^+)$ is a bootstrap analog $\mathbb{F}^*_n$ of $\mathbb{F}_n$ with $\{W_i\}_{i=1}^n$ the bootstrap weights. To this end, we formalize bootstrap consistency of $\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}$ as follows.

Assumption 3.2.3. (i) $\mathbb{F}^*_n : \{X_i, W_i\}_{i=1}^n \to \ell^\infty(\mathbf{R}^+)$ where $\{W_i\}_{i=1}^n$ are the bootstrap weights independent of $\{X_i\}_{i=1}^n$; (ii) $\mathbb{F}^*_n$ satisfies

$$\sup_{h \in \mathrm{BL}_1} |E_W[h(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})] - E[h(\mathbb{G})]| = o_p(1) ,$$

where E_W denotes (outer) expectation with respect to $\{W_i\}$ holding $\{X_i\}$ fixed.

Assumption 3.2.4. $E_W[h(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})^*] - E_W[h(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})_*] \to 0$ almost surely for all $h \in C_b(\ell^\infty(\mathbf{R}^+))$, where $h(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})^*$ and $h(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})_*$ denote minimal measurable majorant and maximal measurable minorant respectively with respect to $\{X_i, W_i\}$ jointly.

Assumption 3.2.3(i) defines the bootstrap analog $\mathbb{F}^*_n$ of $\mathbb{F}_n$, while Assumption 3.2.3(ii) simply imposes bootstrap consistency of $\mathbb{F}^*_n$ which accommodates block bootstrap and general exchangeable bootstrap as well as nonparametric bootstrap. Assumption 3.2.4 demands mild measurability requirement and is automatically satisfied for nonparametric bootstrap.[5]

We are now in a position to establish the bootstrap consistency of $\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})$.

Theorem 3.2.2. *Let Assumptions 3.2.1, 3.2.2, 3.2.3 and 3.2.4 hold. If $\{t_n\}$ satisfies $t_n \downarrow 0$ and $\sqrt{n}t_n \to \infty$ as $n \to \infty$, then*

$$\sup_{h \in BL_1} |E_W[h(\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}))] - E[h(\mathcal{M}'_F(\mathbb{G}))]| = o_p(1) .$$

Theorem 3.2.2 confirms that the law of $\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})$ is consistent for the law of the weak limit $\mathcal{M}'_F(\mathbb{G})$ in Theorem 3.2.1 conditional on the data $\{X_i\}_{i=1}^n$. It is worth noting that estimation of the derivative entails choice of a tuning parameter t_n which approaches zero at a rate slower than $\sqrt{n}$, a phenomenon prominent in irregular models.

[5]Note that Assumption 3.2.4(ii) is stronger than asymptotic measurability of $\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\}$, and closely related to the measurability part of the definition of bootstrap consistency as in van der Vaart and Wellner (1996) and Kosorok (2008).

To illustrate usefulness of Theorem 3.2.2, we next construct a confidence band for F at confidence level $1-\alpha$ with $\alpha \in (0,1)$.

Corollary 3.2.1. *Let Assumptions 3.2.1, 3.2.2, 3.2.3 and 3.2.4 hold, and*

$$\hat{c}_{1-\alpha} \equiv \inf\{c : P_W(\|\hat{\mathcal{M}}'_n(\sqrt{n}\{\mathbb{F}^*_n - \mathbb{F}_n\})\|_\infty \leq c) \geq 1-\alpha\} \ .$$

It follows that:

(i) $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$ *where* $c_{1-\alpha}$ *is the* $1-\alpha$ *quantile of the distribution function of* $\|\mathcal{M}'_F(\mathbb{G})\|_\infty$*;*

(ii) $\{\hat{\mathbb{F}}_n(t) \pm \hat{c}_{1-\alpha}/\sqrt{n} : t \in [0,\infty)\}$ *is a confidence band for* F *with asymptotic confidence level* $1-\alpha$ *– i.e.,*

$$\liminf_{n\to\infty} P\left(F(t) \in \hat{\mathbb{F}}_n(t) \pm \frac{\hat{c}_{1-\alpha}}{\sqrt{n}} \text{ for all } t \in [0,\infty)\right) \geq 1-\alpha \ .$$

Corollary 3.2.1 shows that our proposed bootstrap provides valid critical values and confidence regions. Note that the distribution function of $\|\mathcal{M}'_F(\mathbb{G})\|_\infty$ is strictly increasing everywhere on its support; see Lemma 3.4.1. In practice, $\hat{c}_{1-\alpha}$ is infeasible but can be computed by simulation methods. As a final remark, we note that the confidence band above is valid at least locally in view of the facts that F is a regular parameter in the sense of van der Vaart and Wellner (1996) and that $\|\mathcal{M}'_F(\mathbb{G})\|_\infty$ is a subadditive functional of the Gaussian process $\mathbb{G}$ (Fang and Santos, 2014). The former is a well known fact (Bickel et al., 1993, Example 5.3.1). To see the latter, by Lemma 6.10 in Aliprantis and Border (2006) we may write

$$\|\mathcal{M}'_F(\mathbb{G})\|_\infty = \sup_{g\in\mathbb{S}^*} |g(\mathcal{M}'_F(\mathbb{G}))| = \sup_{g\in\mathbb{S}^*} g(\mathcal{M}'_F(\mathbb{G})) \ ,$$

where $\mathbb{S}^*$ is the unit circle of the topological dual space $\ell^\infty(\mathbf{R}^+)^*$ of $\ell^\infty(\mathbf{R}^+)$ – i.e. $\mathbb{S}^* \equiv \{g \in \ell^\infty(\mathbf{R}^+)^* : \|g\|_{op} = 1\}$ with $\|\cdot\|_{op}$ the corresponding operator norm; see also the proof of Lemma 3.4.1.

3.2.4 Conclusion

In this paper, we have derived the uniform weak convergence of the least concave majorant of the empirical distribution function under minimal assumptions, in particular without bounded support condition and adaptive to strictness of concavity. The derivation built on the fact that least concave majorant operator is Hadamard directionally differentiable. Since the bootstrap consistency of bootstrap is necessarily lost under such transformation, we proposed a consistent one which involved estimating the directional derivative. We illustrated the usefulness of our bootstrap by constructing valid critical values and confidence bands. In fact, our results can be used to build up statistics for testing concavity, or equivalently monotonicity of density in the context of the Grenander problem, and construct pointwise critical values which avoid usage of those based on least favorable curves. We leave the study of this problem to future research.

3.3 Acknowledgement

Section 3.1 is part of the paper: "Inference on Directoinally Differentiable Functions," coauthored with Andres Santos, while Section 3.2 is a coauthored work, under the same title, with Brendan K. Beare.

3.4 Appendix

3.4.1 Proofs of Section 3.1

Proof of Proposition 3.1.1: We proceed by verifying Assumptions 1.2.1, 1.2.2, and 1.2.3, and then employing Theorem 1.2.1 to obtain (3.12). 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:

$$r_n\|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} = r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \ . \tag{3.23}$$

By Lemma 4.6 in Zarantonello (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

$$\phi'_{\theta_0}(h) = \|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} \ , \tag{3.24}$$

for any $h \in \mathbb{H}$. Thus, (3.24) verifies Assumption 1.2.1 and, because in this case $\mathbb{D} = \mathbb{D}_0 = \mathbb{H}$, we conclude Assumption 1.2.3 holds as well. Since Assumption 1.2.2 was directly imposed, the Proposition then follows form Theorem 1.2.1. ■

Proof of Proposition 3.1.2: 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:

$$\begin{aligned} \hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2) &\leq \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n} \{\|h_1 - \Pi_{T_\theta} h_1\|_{\mathbb{H}} - \|h_2 - \Pi_{T_\theta} h_2\|_{\mathbb{H}}\} \\ &\leq \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n} \{\|h_1 - \Pi_{T_\theta} h_2\|_{\mathbb{H}} - \|h_2 - \Pi_{T_\theta} h_2\|_{\mathbb{H}}\} \leq \|h_1 - h_2\|_{\mathbb{H}} \ , \end{aligned} \tag{3.25}$$

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 (3.25) 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 $h_1, h_2 \in \mathbb{H}$:

$$|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)| \leq \|h_1 - h_2\|_{\mathbb{H}} \ . \tag{3.26}$$

Thus, by Lemma 1.6.6, in verifying $\hat{\phi}'_n$ satisfies Assumption 1.3.3 it suffices to show that:

$$|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| = o_p(1) \tag{3.27}$$

for all $h \in \mathbb{H}$. To this end, note that convexity of Λ and Proposition 46.5(2) in Zeidler (1990) 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 1.2.2 and $r_n\epsilon_n \uparrow \infty$ by hypothesis, we conclude that:

$$\liminf_{n\to\infty} P(\|\Pi_\Lambda\theta_0 - \Pi_\Lambda\hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n) \geq \liminf_{n\to\infty} P(r_n\|\theta_0 - \hat{\theta}_n\|_{\mathbb{H}} \leq r_n\epsilon_n) = 1 \; . \tag{3.28}$$

Moreover, the same arguments as in (3.28) and the triangle inequality further imply that:

$$\liminf_{n\to\infty} P(\|\theta - \Pi_\Lambda\theta_0\|_{\mathbb{H}} \leq 2\epsilon_n \text{ for all } \theta \in \Lambda \text{ s.t. } \|\theta - \Pi_\Lambda\hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n)$$
$$\geq \liminf_{n\to\infty} P(\|\Pi_\Lambda\theta_0 - \Pi_\Lambda\hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n) = 1 \; . \tag{3.29}$$

Hence, from the definition of $\hat{\phi}_n'$ and results (3.28) and (3.29) we obtain for any $h \in \mathbb{H}$:

$$\liminf_{n\to\infty} P(\|h - \Pi_{T_{\theta_0}}h\|_{\mathbb{H}} \leq \hat{\phi}_n'(h) \leq \sup_{\theta\in\Lambda:\|\theta-\Pi_\Lambda\theta_0\|_{\mathbb{H}}\leq 2\epsilon_n} \|h - \Pi_{T_\theta}h\|_{\mathbb{H}}) = 1 \; . \tag{3.30}$$

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:

$$\limsup_{n\to\infty}\{\sup_{\theta\in\Lambda:\|\theta-\Pi_\Lambda\theta_0\|_{\mathbb{H}}\leq 2\epsilon_n} \|h - \Pi_{T_\theta}h\|_{\mathbb{H}}\} = \lim_{n\to\infty} \|h - \Pi_{T_{\theta_n}}h\|_{\mathbb{H}} \; . \tag{3.31}$$

By Theorem 4.2.2 in Aubin and Frankowska (2009), 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 $\tilde{h}_n \in T_{\theta_n}$ for all n and $\|\Pi_{T_{\theta_0}}h - \tilde{h}_n\|_{\mathbb{H}} = o(1)$. Thus,

$$\limsup_{n\to\infty}\{\sup_{\theta\in\Lambda:\|\theta-\Pi_\Lambda\theta_0\|_{\mathbb{H}}\leq 2\epsilon_n} \|h - \Pi_{T_\theta}h\|_{\mathbb{H}}\}$$
$$= \lim_{n\to\infty} \|h - \Pi_{T_{\theta_n}}h\|_{\mathbb{H}} \leq \lim_{n\to\infty} \|h - \tilde{h}_n\|_{\mathbb{H}} = \|h - \Pi_{T_{\theta_0}}h\|_{\mathbb{H}} \; , \tag{3.32}$$

where the first equality follows from (3.31), 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 (3.30) and (3.32) we conclude that (3.27) holds, and by Lemma 1.6.6 and (3.26) that $\hat{\phi}_n'$ satisfies Assumption 1.3.3.

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 (1990), 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:

$$\|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|_{\mathbb{H}}^2 = \langle h_1 + h_2, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\rangle_{\mathbb{H}} \ . \tag{3.33}$$

However, Proposition 46.5(4) in Zeidler (1990) further implies that $\langle c, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\rangle \leq 0$ 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 (3.33) and the Cauchy Schwarz inequality that

$$\begin{aligned}\|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|_{\mathbb{H}}^2 &\leq \langle h_1 - \Pi_{T_{\theta_0}} h_1 + h_2 - \Pi_{T_{\theta_0}} h_2, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\rangle_{\mathbb{H}} \\ &\leq \|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|_{\mathbb{H}} \times \|(h_1 - \Pi_{T_{\theta_0}} h_1) + (h_2 - \Pi_{T_{\theta_0}} h_2)\|_{\mathbb{H}} \ . \end{aligned} \tag{3.34}$$

Thus, the Proposition follows from (3.34) and the triangle inequality. ■

3.4.2 Proofs of Section 3.2

Proof of Proposition 3.2.1: Fix $f \in \ell^\infty_c(\mathbf{R}^+)$ and let $\{h_n\}$ be a sequence in $\ell^\infty(\mathbf{R}^+)$ such that $\|h_n - h\|_\infty \to 0$ as $n \to \infty$ where $h \in C_0(\mathbf{R}^+)$, and $t_n \downarrow 0$ as $n \to \infty$. We want to show that

$$\begin{aligned}&\|\frac{\mathcal{M}(f + t_n h_n) - \mathcal{M}(f)}{t_n} - \mathcal{M}'_f(h)\|_\infty \\ &\qquad = \|\frac{\mathcal{M}(f + t_n h_n) - f}{t_n} - \mathcal{M}'_f(h)\|_\infty \to \infty \ . \end{aligned} \tag{3.35}$$

By Lemma 2.2 in Durot and Tocquet (2003), we have

$$\|\frac{\mathcal{M}(f + t_n h_n) - \mathcal{M}(f)}{t_n} - \frac{\mathcal{M}(f + t_n h) - \mathcal{M}(f)}{t_n}\|_\infty \leq \|h_n - h\|_\infty \to 0 \ . \tag{3.36}$$

Thus, in verifying (3.35) it suffices to show that

$$\|\frac{\mathcal{M}(f + t_n h) - f}{t_n} - \mathcal{M}'_f(h)\|_\infty \to \infty \ . \tag{3.37}$$

For notational simplicity, define $g_n, g : \mathbf{R}^+ \to \mathbf{R}$ by

$$g_n(x) \equiv \frac{\mathcal{M}(f + t_n h)(x) - f(x)}{t_n} \ , \ g(x) \equiv \mathcal{M}'_f(h)(x) \ , \forall\, x \in \mathbf{R}^+ \ .$$

In order to show (3.37) or $g_n \to g$ uniformly, we show $g_n \to g$ pointwise, compactify $\mathbf{R}^+$ to embed everything into $\ell^\infty(\bar{\mathbf{R}}^+)$, and then conclude uniformity under suitable continuity and monotonicity by Dini's theorem.

Step 1: For each $x \in \mathbf{R}^+$, $g_n(x) \downarrow g(x)$ as $n \to \infty$. The proof here closely follows that of Lemma 3.2 in Beare and Moon (2015). Fix $x \in (0, \infty)$. Since f is concave on $[0, \infty)$, we have by Theorem 7.23 in Aliprantis and Border (2006) that there is an affine function $\xi : \mathbf{R}^+ \to \mathbf{R}$ such that $f(x) = \xi(x)$ and $\xi \geq f$ on $\mathbf{R}^+$. For any f_1, f_2 with f_2 affine, it is an easy consequence of Lemma 2.1 in Durot and Tocquet (2003) that $\mathcal{M}(f_1 + f_2) = \mathcal{M}(f_1) + f_2$. Thus, together with the fact that $\mathcal{M}$ is positively homogeneous of degree one, we may rewrite

$$g_n(x) = t_n^{-1}\{\mathcal{M}(f + t_n h - \xi)(x)\} = \mathcal{M}(h + t_n^{-1}\{f - \xi\})(x) \ . \tag{3.38}$$

Since $\xi \geq f$ and $t_n \downarrow 0$, it is clear from representation (3.38) that $g_n(x)$ is decreasing in n.

Next, we will show that for $h_{f,n} \equiv h + t_n^{-1}\{f - \xi\}$ and any fixed $\delta > 0$,

$$g_n(x) = \mathcal{M}_{I_x}(h_{f,n})(x) \equiv \mathcal{M}_{T_x(\delta)}(h + t_n^{-1}\{f - \xi\})(x) \ , \tag{3.39}$$

where $T_x(\delta) \equiv [(a_x - \delta) \vee 0, b_x + \delta]$ for all n sufficiently large. By Lemma 1 in Carolan (2002), we may rewrite (3.38) as

$$g_n(x) = \sup_{0 \leq u \leq x} \sup_{x \leq v \leq \infty} \frac{(v - x)h_{f,n}(u) + (x - u)h_{f,n}(v)}{v - u} \ , \tag{3.40}$$

where $0/0$ is defined as $h_{f,n}(x) = h(x)$ if $u = v = x$. For $u \neq v$, substituting $h_{f,n} \equiv$

$h + t_n^{-1}\{f - \xi\}$ back into the objective function in (3.40) yields that

$$\frac{(v-x)h_{f,n}(u) + (x-u)h_{f,n}(v)}{v-u} = \frac{(v-x)h(u) + (x-u)h(v)}{v-u} + t_n^{-1}\left[\frac{(v-x)f(u) + (x-u)f(v)}{v-u} - f(x)\right] , \tag{3.41}$$

where we used the fact that ξ is affine with $\xi(x) = f(x)$. By concavity of f and definitions of a_x and b_x, the term in the brackets of (3.41) is negative and bounded away from zero as u and v with $u \neq v$ range over the complement of $[(a_x - \delta) \vee 0, x] \cup [x, (b_x + \delta)]$. Thus, by choosing n sufficiently large, we may restrict the suprema in (3.40) to $u \in [(a_x - \delta) \vee 0, x]$ and $v \in [x, (b_x + \delta)]$:

$$g_n(x) = \sup_{u\in[(a_x-\delta)\vee 0,x]} \sup_{v\in[x,(b_x+\delta)]} \frac{(v-x)h_{f,n}(u) + (x-u)h_{f,n}(v)}{v-u} = \mathcal{M}_{T_x(\delta)}h_{f,n}(x) , \tag{3.42}$$

proving (3.39), where the second equality is again by Lemma 1 in Carolan (2002).

In what follows, define $\mathcal{M}_{[a_x,b_x]}h(x) = h(x)$ if $a_x = b_x = x$. Now by (3.39) we have

$$\limsup_{n\to\infty} |g_n(x) - g(x)| \leq \limsup_{n\to\infty} |\mathcal{M}_{[(a_x-\delta)\vee 0,b_x+\delta]}h_{f,n}(x) - \mathcal{M}_{[a_x,b_x]}h(x)| . \tag{3.43}$$

Since $h_{f,n}(x') \leq h(x')$ for all $x' \in (0, \infty)$ with equality if $a_x \leq x' \leq b_x$, it follows that

$$\mathcal{M}_{[a_x,b_x]}h(x) \leq \mathcal{M}_{[(a_x-\delta)\vee 0,b_x+\delta]}h_{f,n}(x) \leq \mathcal{M}_{[(a_x-\delta)\vee 0,b_x+\delta]}h(x) . \tag{3.44}$$

Combining (3.43) and (3.44) we may conclude that

$$\limsup_{n\to\infty} |g_n(x) - g(x)| \leq \limsup_{n\to\infty} |\mathcal{M}_{[(a_x-\delta)\vee 0,b_x+\delta]}h(x) - \mathcal{M}_{[a_x,b_x]}h(x)| . \tag{3.45}$$

It follows that $g_n(x) \downarrow g(x)$ as $n \to \infty$ for $x \in (0, \infty)$ by letting $\delta \downarrow 0$ in (3.45). For $x = 0$, it is clear that $g_n(0) = 0$ since $h(0) = 0$, in which case we trivially have $g_n(0) \downarrow g(0) = 0$ as $n \to \infty$.

<u>Step 2:</u> Embedding into $\ell^\infty(\bar{\mathbf{R}}^+)$. Let $\bar{\mathbf{R}}^+ \equiv [0, \infty]$ be the one point compactification of

$\mathbf{R}^+$. Since $h \in C_0(\mathbf{R}^+)$, we also have $h \in C_0(\bar{\mathbf{R}}^+)$ by setting $h(\infty) = 0$. Moreover, we may set $h_n(\infty) = 0$ so that $h_n(\infty) \to h(\infty)$ as $n \to \infty$. In this way, we still have $\|h_n - h\|_\infty \to 0$ in $\ell^\infty(\bar{\mathbf{R}}^+)$. Finally, we may also identify $f \in \ell^\infty(\bar{\mathbf{R}}^+)$ by setting $f(\infty) = \sup_{x \in \mathbf{R}^+} f(x)$.[6] To distinguish, we denote the uniform norm in $\ell^\infty(\bar{\mathbf{R}}^+)$ by $\|\cdot\|_{\infty,\bar{\mathbf{R}}^+}$.

<u>STEP 3:</u> Conclude by invoking Dini's theorem. To verify (3.37), we now aim for the following stronger result:

$$\|g_n - g\|_{\infty,\bar{\mathbf{R}}^+} = \|\frac{\mathcal{M}(f + t_n h) - f}{t_n} - \mathcal{M}'_f(h)\|_{\infty,\bar{\mathbf{R}}^+} \to \infty \ , \tag{3.46}$$

where we set $\mathcal{M}'_f(h)(\infty) = h(\infty) = 0$. Since $g_n(\infty) = g(\infty) = 0$, together with results in Step 1, this implies that $g_n \downarrow g$ pointwise on $\bar{\mathbf{R}}^+$.

Notice that g_n is continuous on $\bar{\mathbf{R}}^+$ for each $n \in \mathbf{N}$. To see this, note that g_n is automatically continuous as a concave function on $[0, \infty)$ by right continuity of f at 0; also as $x \to \infty$, we have by $h \in C_0(\mathbf{R}^+)$ and Lemma 2.2 in Durot and Tocquet (2003),

$$|g_n(x)| \equiv |\frac{\mathcal{M}(f + t_n h)(x) - f(x)}{t_n}| \le \sup_{t \in [x,\infty)} |h(t)| \to 0 = g_n(\infty) \ .$$

Second, g is continuous on $\bar{\mathbf{R}}^+$ as well by definition. It follows by Dini's theorem (Aliprantis and Border, 2006, Theorem 2.66) that $g_n \to g$ uniformly on $\bar{\mathbf{R}}^+$, proving (3.46) and we are done. ■

PROOF OF THEOREM 3.2.1: By F being concave, we may rewrite $\sqrt{n}\{\hat{\mathbb{F}}_n - F\} = \sqrt{n}\{\mathcal{M}\mathbb{F}_n - \mathcal{M}F\}$, which together with Assumption 3.2.2, Proposition 3.2.1, and $\mathbb{G}$ being tight with $P(\mathbb{G} \in C_0(\mathbf{R}^+)) = 1$, allows us to invoke Theorem 2.1 in Fang and Santos (2014) to conclude. ■

PROOF OF THEOREM 3.2.2: We proceed by verifying conditions of Theorem 3.3 in Fang and Santos (2014). First, we show that our proposed estimator $\hat{\mathcal{M}}'_n : \ell^\infty(\mathbf{R}^+) \to \ell^\infty(\mathbf{R}^+)$ meets Assumption 3.3 in Fang and Santos (2014). Note that $\hat{\mathcal{M}}'_n$ is Lipschitz continuous

[6] A function $f : \bar{\mathbf{R}}^+ \to \mathbf{R}$ is said to be concave if $f(tx + (1-t)y) \ge tf(x) + (1-t)f(y)$ for all $t \in (0, 1)$ and all $x, y \in \bar{\mathbf{R}}^+$.

surely for if $h_1, h_2 \in \ell^\infty(\mathbf{R}^+)$, then by Lemma 2.2 in Durot and Tocquet (2003),

$$\begin{aligned}\|\hat{\mathcal{M}}_n'(h_1) - \hat{\mathcal{M}}_n'(h_2)\|_\infty &= \|\frac{\mathcal{M}(\mathbb{F}_n + t_n h_1) - \mathcal{M}(\mathbb{F}_n)}{t_n} - \frac{\mathcal{M}(\mathbb{F}_n + t_n h_2) - \mathcal{M}(\mathbb{F}_n)}{t_n}\|_\infty \\ &= \|\frac{\mathcal{M}(\mathbb{F}_n + t_n h_1) - \mathcal{M}(\mathbb{F}_n + t_n h_2)}{t_n}\|_\infty \leq \|h_1 - h_2\|_\infty .\end{aligned}$$

Next, fix $h \in C_0(\mathbf{R}^+)$ and we want to show the pointwise consistency of $\hat{\mathcal{M}}_n'$. Since $C_0(\mathbf{R}^+)$ is closed in $\ell^\infty(\mathbf{R}^+)$, $\mathcal{M}_F'$ can be extended continuously to the entire space $\ell^\infty(\mathbf{R}^+)$ by Theorem 4.1 in Dugundji (1951). Now rewrite

$$\begin{aligned}\hat{\mathcal{M}}_n'(h) &= \frac{\mathcal{M}(\mathbb{F}_n + t_n h_1) - \mathcal{M}(\mathbb{F}_n)}{t_n} \\ &= \frac{\mathcal{M}(F + t_n h_n) - \mathcal{M}F}{t_n} - (\sqrt{n}t_n)^{-1} \cdot \sqrt{n}\{\mathcal{M}(\mathbb{F}_n) - \mathcal{M}F\} ,\end{aligned} \tag{3.47}$$

where $h_n \equiv (\sqrt{n}t_n)^{-1}\sqrt{n}\{\mathbb{F}_n - F\} + h$. By Assumption 3.2.2, $\sqrt{n}\{\mathbb{F}_n - F\} \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathbf{R}^+)$ where $\mathbb{G}$ is tight and hence separable. Also, $\sqrt{n}t_n \to \infty$ as $n \to \infty$ by assumption. It follows by Example 1.4.7 and Lemma 1.10.2 in van der Vaart and Wellner (1996) that $h_n \xrightarrow{p} h$. Now by Proposition 3.2.1 and the extended continuous mapping theorem,

$$\frac{\mathcal{M}(F + t_n h_n) - \mathcal{M}F}{t_n} \xrightarrow{p} \mathcal{M}_F'(h) . \tag{3.48}$$

Theorem 3.2.1 and $\sqrt{n}t_n \to \infty$ as $n \to \infty$ together imply again by Example 1.4.7 and Lemma 1.10.2 in van der Vaart and Wellner (1996) that

$$(\sqrt{n}t_n)^{-1} \cdot \sqrt{n}\{\mathcal{M}(\mathbb{F}_n) - \mathcal{M}F\} \xrightarrow{p} 0 . \tag{3.49}$$

Combination of results (3.47), (3.48) and (3.49) then leads to $\hat{\mathcal{M}}_n'(h) \xrightarrow{p} \mathcal{M}_F'(h)$ as $n \to \infty$. We may now conclude by Lemma A.6 in Fang and Santos (2014) that Assumption 3.3 in Fang and Santos (2014) holds with θ, ϕ, $\mathbb{D}$, $\mathbb{D}_0$ and $\mathbb{E}$ replaced by F, $\mathcal{M}$, $\ell^\infty(\mathbf{R}^+)$, $C_0(\mathbf{R}^+)$ and $\ell^\infty(\mathbf{R}^+)$ respectively. The conclusion follows by Theorem 3.3 in Fang and Santos (2014).

■

Proof of Corollary 3.2.1: Note that $\mathcal{M}_F'(\mathbb{G})$ is tight since $\mathbb{G}$ is tight and $\mathcal{M}_F'$ is continuous as a Hadamard directional derivative (Shapiro, 1990, Propositioin 3.1). By Proposition 10.7 in Kosorok (2008), Theorem 3.2.2 and $\|\cdot\|_\infty$ being Lipschitz continuous, we conclude that

$$\sup_{h\in\mathrm{BL}_1(\mathbf{R})} |E_W[h(\|\hat{\mathcal{M}}_n'(\sqrt{n}\{\mathbb{F}_n^* - \mathbb{F}_n\})\|_\infty)] - E[h(\|\mathcal{M}_F'(\mathbb{G})\|_\infty)]| = o_p(1) \ . \tag{3.50}$$

Part (i) then follows from the same proof as that of Corollary 3.2 in Fang and Santos (2014) by noticing that the distribution function of $\|\mathcal{M}_F'(\mathbb{G})\|_\infty$ is strictly increasing at the $1-\alpha$ quantile with $\alpha \in (0,1)$ in view of Lemma 3.4.1.

As for part (ii), Theorem 3.2.1 implies by the continuous mapping theorem that

$$\|\sqrt{n}\{\hat{\mathbb{F}}_n - F\}\|_\infty \xrightarrow{L} \|\mathcal{M}_F'(\mathbb{G})\|_\infty \ . \tag{3.51}$$

Then the portmanteau theorem, part (i) and $c_{1-\alpha}$ being a continuity point imply that

$$\begin{aligned}
\liminf_{n\to\infty} P(F(t) \in \hat{\mathbb{F}}_n(t) \pm \frac{\hat{c}_{1-\alpha}}{\sqrt{n}} \text{ for all } t\in[0,\infty)) \\
&= \liminf_{n\to\infty} P(\|\sqrt{n}\{\hat{\mathbb{F}}_n - F\}\|_\infty \le \hat{c}_{1-\alpha}) \\
&= P(\|\mathcal{M}_F'(\mathbb{G})\|_\infty \le c_{1-\alpha}) = 1-\alpha \ .
\end{aligned} \tag{3.52}$$

We thus proved part (ii). ∎

Lemma 3.4.1. *The distribution function of $\|\mathcal{M}_F'(\mathbb{G})\|_\infty$ for all concave F is absolutely continuous and strictly increasing on its support.*

Proof: First of all, notice that $\mathcal{M}_F'$ is convex on $C_0(\mathbf{R}^+)$. To see this, note that $\mathcal{M}$ is convex for if $a\in(0,1)$ and $f,g\in\ell^\infty(\mathbf{R}^+)$, then $a\mathcal{M}(f)+(1-a)\mathcal{M}(g)$ is clearly a concave majorant of $af+(1-a)g$ and hence

$$\mathcal{M}(af+(1-a)g) \le a\mathcal{M}(f) + (1-a)\mathcal{M}(g) \ .$$

Now let $a\in(0,1)$ and $f,g\in C_0(\mathbf{R}^+)$. Take $\{f_n,g_n\}\subset\ell^\infty(\mathbf{R}^+)$ such that $f_n\to f$ and

$g_n \to g$. Then

$$\begin{aligned}
\mathcal{M}'_F(af+(1-a)g) &= \lim_{n\to\infty} \frac{\mathcal{M}(F+t_n(af_n+(1-a)g_n))-F}{t_n} \\
&\le \lim_{n\to\infty} \frac{a\mathcal{M}(F+t_nf_n)+(1-a)\mathcal{M}(F+t_ng_n)-F}{t_n} \\
&= a\cdot \lim_{n\to\infty} \frac{\mathcal{M}(F+t_nf_n)-F}{t_n} + (1-a)\cdot \lim_{n\to\infty} \frac{\mathcal{M}(F+t_ng_n))-F}{t_n} \\
&= a\mathcal{M}'_F(f)+(1-a)\mathcal{M}'_F(g)\ . \qquad (3.53)
\end{aligned}$$

Next, let $\mathbb{S}^*$ be the unit circle of the topological dual space $\ell^\infty(\mathbf{R}^+)^*$ of $\ell^\infty(\mathbf{R}^+)$ – i.e. $\mathbb{S}^* \equiv \{g \in \ell^\infty(\mathbf{R}^+)^* : \|g\|_{op} = 1\}$ where $\|\cdot\|_{op}$ is the corresponding operator norm. Then by Lemma 6.10 in Aliprantis and Border (2006) we may write

$$\|\mathcal{M}'_F(\mathbb{G})\|_\infty = \sup_{g\in\mathbb{S}^*} |g(\mathcal{M}'_F(\mathbb{G}))| = \sup_{g\in\mathbb{S}^*} g(\mathcal{M}'_F(\mathbb{G}))\ .$$

Define $T : C_0(\mathbf{R}^+) \to \mathbf{R}$ by

$$T(f) \equiv \sup_{g\in\mathbb{S}^*} g(\mathcal{M}'_F(f))\ , \qquad (3.54)$$

so that $\|\mathcal{M}'_F(\mathbb{G})\|_\infty = T(\mathbb{G})$. Inspecting (3.53) and (3.54) we see that T is convex. Second, T is obviously continuous. Third, $\mathbb{G}$ is separable by Lemma 1.3.2 in van der Vaart and Wellner (1996)and $\mathbb{G}$ being tight, which in turn implies by Theorem 7.1.7 in Bogachev (2007) that $\mathbb{G}$ is Radon. Putting all these pieces together we are then able to conclude by Theorem 11.1 in Davydov et al. (1998) that the distribution function H of $\|\mathcal{M}'_F(\mathbb{G})\|_\infty$ is absolutely continuous and has a positive and continuous derivative on the support of H except on an at most countable set, which in turn implies that H is strictly increasing on the support. ■

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. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 3rd ed. Springer Verlag.

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 inequalities. *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. D. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *The American Economic Review*, **80** 313–336.

Angrist, J. D., Chernozhukov, V. and Fernández-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. (2014). Multiscale adaptive inference on conditional moment inequalities. Working paper.

Aubin, J.-P. and Frankowska, H. (2009). *Set-Valued Analysis*. Springer.

Averbukh, V. I. and Smolyanov, O. G. (1967). The theory of differentiation in linear topological spaces. *Russian Mathematical Surveys*, **22** 201–258.

Averbukh, V. I. and Smolyanov, O. G. (1968). The various definitions of the derivative in linear topological spaces. *Russian Mathematical Surveys*, **23** 67.

Bassett, J., Gilbert and Koenker, R. (1982). An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association*, **77** 407–415.

Beare, B. and Shi, X. (2015). An improved bootstrap test of density ratio ordering. Working paper.

Beare, B. K. and Moon, J.-M. (2015). Nonparametric tests of density ratio ordering. *Econometric Theory*, **31** 471–492.

Beare, B. K. and Schmidt, L. D. W. (2015). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics* forthcoming.

Begun, J. M., Hall, W. J., Huang, W.-M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, **11** 432–452.

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. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, **7** 1–31.

Bickel, P. J., Klaassen, C. A., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press.

Blackwell, D. (1951). Comparison of experiments. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. 93–102.

Blumenthal, S. and Cohen, A. (1968a). Estimation of the larger of two normal means. *Journal of the American Statistical Association*, **63** 861–876.

Blumenthal, S. and Cohen, A. (1968b). Estimation of the larger translation parameter. *The Annals of Mathematical Statistics*, **39** 502–516.

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

Bogachev, V. I. (2007). *Measure Theory*. Springer-Verlag.

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: Large deviations optimality and bootstrap validity. *Journal of Econometrics*, **156** 408 – 425.

Carolan, C. and Dykstra, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *The Canadian Journal of Statistics*, **27** 557–566.

Carolan, C. and Dykstra, R. (2001). Marginal densities of the least concave majorant of Brownian motion. *The Annals of Statistics*, **29** 1732–1750.

Carolan, C. A. (2002). The least concave majorant of the empirical distribution function. *Canadian Journal of Statistics*, **30** 317–328.

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.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, **34** 305 – 334.

Chandrasekhar, A., Chernozhukov, V., Molinari, F. and Schrimpf, P. (2012). Inference for best linear approximations to set identified functions. Working paper.

Chen, X., Tamer, E. and Torgovitsky, A. (2011). Sensitivity analysis in semiparametric likelihood models. Working paper.

Chernoff, H. (1956). Large-sample theory: Parametric case. *The Annals of Mathematical Statistics*, **27** 1–22.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, **78** 1093–1125.

Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, **73** 245–261.

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, **132** 491 – 525.

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. 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.

Conway, J. B. (1990). *A Course in Functional Analysis*. 2nd ed. Springer-Verlag.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998). *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society.

Dehling, H. and Philipp, W. (2002). *Empirical Process Techniques for Dependent Data*. Springer.

Dorfman, R. (1938). A note on the δ-method for finding variance formulae. *The Biometric Bulletin*, **1** 92.

Doss, H. and Sethuraman, J. (1989). The price of bias reduction when there is no unbiased estimate. *The Annals of Statistics*, **17** 440–442.

Dudley, R. M. (1966). Convergence of Baire measures. *Studia Mathematica*, **27** 251–268.

DUDLEY, R. M. (1968). Distances of probability measures and random variables. *The Annals of Mathematical Statistics*, **39** 1563–1572.

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.

DUROT, C. (2007). On the $\mathbb{L}_p$ -error of monotonicity constrained estimators. *The Annals of Statistics*, **35** 1080–1104.

DUROT, C., KULIKOV, V. N. and LOPUHAÄ, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *The Annals of Statistics*, **40** 1578–1608.

DUROT, C. and TOCQUET, A.-S. (2003). On the distance between the empirical process and its concave majorant in a monotone regression framework. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, **39** 217 – 240.

DVORETZKY, A., WALD, A. and WOLFOWITZ, J. (1950). Elimination of randomization in certain problems of statistics and of the theory of games. *Proceedings of the National Academy of Sciences of the United States of America*, **36** 256–260.

DVORETZKY, A., WALD, A. and WOLFOWITZ, J. (1951). Elimination of randomization in certain statistical decision procedures and zero-sum two-person games. *The Annals of Mathematical Statistics*, **22** 1–21.

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

EGGERMONT, P. P. B. and LARICCIA, V. N. (2000). Maximum likelihood estimation of smooth monotone and unimodal densities. *The Annals of Statistics*, **28** 922–947.

ESCANCIANO, J. C. and ZHU, L. (2013). Set inferences and sensitivity analysis in semi-parametric conditionally identied models. Working paper.

FANG, Z. and SANTOS, A. (2014). Inference on directionally differentiable functions. Working paper.

FEINBERG, E. and PIUNOVSKIY, A. (2006). On the Dvoretzky-Wald-Wolfowitz theorem on nonrandomized statistical decisions. *Theory of Probability & Its Applications*, **50** 463–466.

FOLLAND, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*. 2nd ed. Wiley & Sons.

GALLANT, A. R. and NYCHKA, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, **55** 363–390.

GRENANDER, U. (1956). On the theory of mortality measurement. *Scandinavian Actuarial Journal*, **1956** 70–96.

GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1999). Asymptotic normality of the L_1 error of the grenander estimator. *The Annals of Statistics*, **27** 1316–1347.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge University Press.

Groenenboom, P. (1984). Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer* (R. A. O. Lucien M. Le Cam, ed.), vol. II. Wadsworth, Monterey, CA, 539–555.

Haile, P. A. and Tamer, E. (2003). Inference with an incomplete model of English auctions. *Journal of Political Economy*, **111** 1–51.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **14** 323–330.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, 175–194.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

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

Haraux, A. (1977). How to differentiate the projection on a convex set in Hilbert space. some applications to variational inequalities. *Journal of the Mathematical Society of Japan*, **29** 615–631.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, **51** 186–192.

Hirano, K. and Porter, J. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, **80** 1769–1790.

Hong, H. and Li, J. (2014). The numerical directional Delta method. Working paper, Stanford University.

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

Huber, P. J. (1966). Strict efficiency excludes superefficiency. *The Annals of Mathematical Statistics*, **37** 1425.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer-Verlag.

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

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, **13** 433–451.

Jeganathan, P. (1981). On a decomposition of the limit distribution of a sequence of estimators. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, **43** 26–36.

JEGANATHAN, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, **44** 173–212.

KAIDO, H. (2013a). Asymptotically efficient estimation of weighted average derivatives with an interval censored variable. Working paper.

KAIDO, H. (2013b). A dual approach to inference for partially identified econometric models. Working paper.

KAIDO, H. and SANTOS, A. (2014). Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica*, **82** 387–413.

KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **34** 73–85.

KITAMURA, Y. and STOYE, J. (2013). Nonparametric analysis of random utility models: Testing. Working paper.

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.

KOSHEVNIK, Y. and LEVIT, B. (1976). On a non-parametric analogue of the information matrix. *Theory of Probability & Its Applications*, **21** 738–753.

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

KULIKOV, V. N. and LOPUHAÄ, H. P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *The Annals of Statistics*, **33** 2228–2255.

KULIKOV, V. N. and LOPUHAÄ, H. P. (2006a). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *The Annals of Statistics*, **34** 742–768.

KULIKOV, V. N. and LOPUHAÄ, H. P. (2006b). The limit process of the difference between the empirical distribution function and its concave majorant. *Statistics & Probability Letters*, **76** 1781 – 1786.

KULIKOV, V. N. and LOPUHAÄ, H. P. (2008). Distribution of global measures of deviation between the empirical distribution function and its concave majorant. *Journal of Theoretical Probability*, **21** 356–377.

LE CAM, L. M. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. In *University of California Publications in Statistics*, vol. 1. 277–330.

LE CAM, L. M. (1955). An extension of Wald's theory of statistical decision functions. *The Annals of Mathematical Statistics*, **26** 69–81.

LE CAM, L. M. (1964). Sufficiency and approximate sufficiency. *The Annals of Mathematical Statistics*, **35** 1419–1455.

LE CAM, L. M. (1972). Limits of experiments. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab*, vol. 1. University of California Press, 245–261.

LE CAM, L. M. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*. Springer.

LEE, Y. (2009). Efficiency bounds for semiparametric estimation of quantile regression under misspecification. Working paper.

LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*. 2nd ed. Springer.

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer Verlag.

LEVIT, B. (1978). Infinite-dimensional informational bounds. *Theory of Probability and Its Applications*, **23** 371–377.

LINTON, O., SONG, K. E. 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.

MANSKI, C. F. and PEPPER, J. V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, **68** 997–1010.

MANSKI, C. F. and PEPPER, J. V. (2009). More on monotone instrumental variables. *Econometrics Journal*, **12** S200–S216.

MANSKI, C. F. and TAMER, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, **70** 519–546.

MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. vol. 976 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, 75–265.

MILLAR, P. W. (1985). Non-parametric applications of an infinite dimensional convolution theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **68** 545–556.

MURALIDHARAN, K. and SUNDARARAMAN, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, **119** 39–77.

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65** 331–355.

NEWEY, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In *Econometrics* (H. G.S.Maddala, C.R.Rao, ed.), vol. 11 of *Handbook of Statistics*, chap. 16. Amsterdam: North-Holland, 419 – 454.

NEWEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, **71** 1565–1578.

Oosterhoff, J. and van Zwet, W. (1979). A note on contiguity and Hellinger distance. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume* (J. Jurecková, ed.). Reidel, Dordrecht, 157–166.

Pakes, A., Porter, J., Ho, K. and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, **83** 315–334.

Pfanzagl, J. and Wefelmeyer, W. (1982). *Contributions to a General Asymptotic Statistical Theory*. Springer-Verlag.

Rao, B. L. S. P. (1969). Estimation of a unimodal density. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, **31** 23–36.

Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, **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.

Saks, S. (1937). *Theory of the Integral*. Courier Dover Publications.

Santos, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica*, **80** 213–275.

Seo, J. (2014). Tests of stochastic monotonicity with improved size and power properties. Working paper.

Severini, T. A. and Tripathi, G. (2001). A simplified approach to computing bounds in semiparametric models. *Journal of Econometrics*, **102** 23 – 66.

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 on Optimization*, **4** 130–141.

Song, K. (2014). Local asymptotic minimax estimation of nonregular parameters with translation-scale equivariant maps. *Journal of Multivariate Analysis*, **125** 136 – 158.

Song, K. (2015). Minimax estimation of nonregular parameters and discontinuity in minimax risk. In *Essays in Honor of Peter C. B. Phillips*. 557–585.

Strasser, H. (1982). Local asymptotic minimax properties of Pitman estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **60** 223–247.

van der Vaart, A. W. (1988a). Estimating a real parameter in a class of semiparametric models. *The Annals of Statistics*, **16** 1450–1474.

van der Vaart, A. W. (1988b). *Statistical Estimation in Large Parameter Spaces*. CWI Tracts 44, Centrum voor Wiskunde en Informatica, Amsterdam.

van der Vaart, A. W. (1989). On the asymptotic information bound. *The Annals of Statistics*, **17** 1487–1500.

van der Vaart, A. W. (1991a). An asymptotic representation theorem. *International Statistical Review*, **59** 97–121.

van der Vaart, A. W. (1991b). Efficiency and Hadamard differentiability. *Scandinavian Journal of Statistics*, **18** 63–75.

van der Vaart, A. W. (1991c). On differentiable functionals. *The Annals of Statistics*, **19** 178–204.

van der Vaart, A. W. (1992). Asymptotic linearity of minimax estimators. *Statistica Neerlandica*, **46** 179–194.

van der Vaart, A. W. (1994). Weak convergence of smoothed empirical processes. *Scandinavian Journal of Statistics*, **21** 501–504.

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

van der Vaart, A. W. and Wellner, J. A. (1990). Prohorov and continuous mapping theorems in the Hoffmann-Jørgensen weak convergence theory, with application to convolution and asymptotic minimax theorems. Tech. Rep. 157, Department of Statistics, University of Washington, Seattle.

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

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

Wald, A. (1950). *Statistical Decision Functions*. Wiley.

Walther, G. (2009). Inference and modeling with log-concave distributions. *Statistical Science*, **24** 319–327.

Wang, Y. (1994). The limit distribution of the concave majorant of an empirical distribution function. *Statistics & Probability Letters*, **20** 81 – 84.

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.

Zarantonello, 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. (1990). *Nonlinear Functional Analysis and Its Applications*, vol. III. Springer.