Review of Economic Studies (2019) **86**, 377–412 doi:10.1093/restud/rdy049

Advance access publication 11 September 2018

Inference on Directionally Differentiable Functions

ZHENG FANG
Texas A&M University

and

ANDRES SANTOS
U.C. Los Angeles

First version received December 2015*; Editorial decision August* 2018*; Accepted September* 2018 (*Eds.*)

This article 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 derived employing existing extensions to the Delta method. We show, however, that (full) 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 that remains consistent when the bootstrap fails, and is shown to provide local size control under restrictions on the directional derivative of ϕ. These results enable us to reduce potentially challenging statistical problems to simple analytical calculations—a feature we illustrate by developing a test of whether an identified parameter belongs to a convex set. We highlight the empirical relevance of our results by conducting inference on the qualitative features of trends in (residual) wage inequality in the U.S.

Key words: Delta method, Bootstrap consistency, Directional differentiability, Shape restrictions, Residual wage inequality.

JEL Codes: C1, C12, C15, J31

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

1. Interestingly, despite its importance, the origins of the Delta method remain obscure. ver Hoef (2012) attributes 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).

The editor in charge of this paper was Aureo de Paula.

models (Pakes *et al.*, 2006; Ciliberto and Tamer, 2009), the study of convex partially identified sets (Beresteanu and Molinari, 2008; Bontemps *et al.*, 2012), and the development of tests of superior predictive ability (White, 2000; Hansen, 2005), of stochastic dominance (Linton *et al.*, 2010), and of likelihood ratio ordering (Beare and Moon, 2015).

The aforementioned examples share a structure common to numerous "non-standard" inference problems in economics: the transformation ϕ is directionally (but not fully) differentiable in a local neighbourhood of θ_0. In this article, we show this common structure enables us to reduce challenging statistical questions to simple analytical considerations regarding the directional derivative of ϕ – much in the same manner the Delta method and its bootstrap counterpart fundamentally simplify the analysis of applications in which ϕ is (fully) differentiable. Concretely, we examine a setting in which θ_0 is a possibly infinite dimensional parameter and there exists 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

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

Within this framework, we study a simple unifying approach for conducting inference on the parameter $\phi(\theta_0)$ by employing $\phi(\hat{\theta}_n)$ and a suitable estimator of its asymptotic distribution—a practice common in, for example, the study of moment inequality (Andrews and Soares, 2010), conditional moment inequality (Andrews and Shi, 2013), and incomplete linear models (Beresteanu and Molinari, 2008).

As has been previously noted in the literature, the traditional Delta method generalizes to the case where ϕ is directionally differentiable. In particular, Shapiro (1991) and Dümbgen (1993) show that the Delta method may be applied whenever ϕ is Hadamard directionally differentiable at θ_0 (we review Hadamard directional differentiability in Section 2). This extension of the Delta method readily implies that

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

where ϕ'_{θ_0} denotes the directional derivative of ϕ at θ_0. The utility of the asymptotic distribution of $\phi(\hat{\theta}_n)$, however, hinges on our ability to consistently estimate it. While it is tempting in these problems to resort to resampling schemes such as the bootstrap of Efron (1979), we know by way of example that they may be inconsistent even if they are valid for the original estimator $\hat{\theta}_n$ (Bickel *et al.*, 1997; Andrews, 2000; Woutersen and Ham, 2013). In our first main result, we establish that these examples reflect a deeper underlying principle. Specifically, we establish that whenever the asymptotic distribution of $\hat{\theta}_n$ is Gaussian, full differentiability of ϕ at θ_0 is in fact a *necessary and sufficient* condition for the consistency of "standard" bootstrap methods. As a result, we obtain a purely analytical diagnostic for assessing bootstrap consistency in these settings: one need only verify whether ϕ is (fully) differentiable. An important consequence of our characterization of bootstrap consistency is that, in our setting, "standard" bootstrap methods in fact fail whenever the asymptotic distribution of $\phi(\hat{\theta}_n)$ is not Gaussian—a conclusion that yields an alternative simple way to detect the failure of the bootstrap.

Intuitively, consistently estimating the asymptotic distribution of $\phi(\hat{\theta}_n)$ requires us to adequately approximate both the distribution of $\mathbb{G}_0$ and the directional derivative ϕ'_{θ_0} (see (2)). While a consistent bootstrap procedure for $\hat{\theta}_n$ enables us to do the former, the bootstrap fails for $\phi(\hat{\theta}_n)$ due to its inability to properly estimate ϕ'_{θ_0}. These heuristics, however, readily suggest a remedy to the problem: 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 multiple superficially different resampling schemes in fact follow precisely this approach. For instance, a number of inferential procedures developed in the context of specific applications can be understood as employing an estimator $\hat{\phi}'_n$ that is derived from an analytical expression of ϕ'_{θ_0}. These include, among others, Andrews and Soares (2010) for moment inequalities, Linton *et al*. (2010) for tests of stochastic dominance, and Kaido (2016) for convex partially identified models. Additional special cases of our approach include the m out of n bootstrap of Shao (1994) and the rescaled bootstrap of Dümbgen (1993), which can be shown to implicitly rely on an estimator $\hat{\phi}'_n$ based on numerical differentiation; see Hong and Li (2017). Our results thus cast these different resampling schemes in a framework that highlights their common source of consistency and potentially eases their comparison.

Whenever ϕ is directionally differentiable at θ_0, the asymptotic distribution of $\phi(\hat{\theta}_n)$ can depend discontinuously on the value of θ_0. This sensitivity contrasts with the finite sample distribution of $\phi(\hat{\theta}_n)$, which often depends continuously on θ_0. As emphasized by Imbens and Manski (2004), such a discrepancy is cause for concern that employing the distribution of $\phi'_{\theta_0}(\mathbb{G}_0)$ as the basis for inference can result in tests with poor finite sample properties. In order to allay these concerns, we additionally study the properties of tests in a "local" asymptotic framework. Concretely, a local analysis allows us to better approximate the finite sample properties of tests when θ_0 is "close", but not equal, to a point at which ϕ is not fully differentiable. By way of example, we examine the properties of a test that employs $\phi(\hat{\theta}_n)$ as a test statistic for the hypothesis

$$H_0: \phi(\theta_0) \leq 0 \qquad\qquad H_1: \phi(\theta_0) > 0. \tag{3}$$

Under mild restrictions on $\hat{\theta}_n$, our local analysis reveals that convexity of the directional derivative suffices for establishing the ability of our procedure to (locally) control size. Thus, our results again reduce a challenging statistical problem (establishing local size control) to a simple analytical calculation (verifying convexity). We additionally argue that in problems in which the directional derivative fails to be convex, our local analysis still provides guidance on how to select a critical value for $\phi(\hat{\theta}_n)$ that results in a test with (local) size control.

In summary, our analysis provides empiricists with a simple analytical framework for conducting inference. Whenever ϕ is fully differentiable, we may rely on the standard bootstrap to obtain critical values. If on the other hand ϕ is not fully differentiable at or "near" the parameter θ_0 and $\hat{\theta}_n$ is asymptotically Gaussian, then the standard bootstrap fails. In such instances, our results show critical values can be obtained by instead employing an estimator $\hat{\phi}'_n$ of the directional derivative ϕ'_{θ_0}. The ability of the these critical values to adequately (locally) control size can easily be assessed by analytical calculations as well—for example, for testing (3) it suffices that ϕ'_{θ_0} be convex.

We illustrate the utility of our analysis with two novel applications. First, we employ our results to develop a procedure for conducting inference on partially identified linear regression models with interval valued outcomes.[2] In particular, we show the bootstrap fails in the presence of discrete regressors, and propose an inference procedure that extends the work of Beresteanu and Molinari (2008) and Bontemps *et al*. (2012) to such a setting. Second, we apply our analysis to construct a test for the general null hypothesis that a parameter θ_0 belongs to a known closed convex set. These tests enable us to examine, for instance, whether quantile treatment effect functions satisfy certain shape restrictions, or whether a vector of means belongs to a closed convex set—see, *e.g.*, Wolak (1988) and Kitamura and Stoye (2013) for examples of the latter. Our general framework

2. We are grateful to a referee for suggesting this application.

both implies that the bootstrap often fails in these problems and simultaneously suggests an alternative resampling procedure that is consistent.

Finally, we highlight the empirical relevance of our results by revisiting a large literature on the trends on wage dispersion within demographic and education groups in the United States (*i.e.* "residual" wage dispersion); see, among others, Katz and Murphy (1992), Juhn *et al.* (1993), and Autor *et al.* (2008). Specifically, our theoretical results allow us to contribute to this literature by conducting inference on different qualitative features of the residual wage dispersion trends. For instance, we build confidence regions for the year at which residual wage variance was the largest and find evidence in support of Card and DiNardo (2002) and Lemieux (2006), who argue the rise in residual wage inequality occurred primarily in the 1980s. Following Autor *et al.* (2008), however, we further examine "upper tail" and "lower tail" residual wage dispersion and uncover different trends. In particular, we find evidence indicating that "lower tail" residual wage dispersion has been decreasing monotonically for both men and women since attaining their maximum in the 1980s. In contrast, we further find that the "upper tail" residual wage dispersion for men has been increasing over our sample period, while the "upper tail" residual wage dispersion for women has remained relatively stable since the 1980s.

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). Our analysis is most closely related to the pioneering work of Dümbgen (1993), who first examined the validity of the bootstrap for estimating the asymptotic distribution of $\phi(\hat{\theta}_n)$ under a potential lack of differentiability. The results in Dümbgen (1993) imply a characterization of bootstrap consistency that, unlike ours, applies when $\mathbb{G}_0$ is not Gaussian but is harder for practitioners to verify as it concerns properties of both $\mathbb{G}_0$ and ϕ'_{θ_0}.[3] In more recent studies, 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), Song (2014), and Fang (2015), for example, explicitly employ the directional differentiability of ϕ as well, though their focus is on estimation rather than inference. Other work studying these 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.

An emerging body of research has validated the usefulness of our results by both employing and expanding on them. For instance, Seo (2018) and Beare and Shi (2018) use our framework to develop tests of stochastic monotonicity and of density ratio ordering respectively. Other applications of our results also include Jha and Wolak (2015) who estimate transaction costs in energy future markets, Lee and Bhattacharya (2015) who propose methods for estimating welfare changes in partially identified discrete choice models, Hansen (2017) who studies the asymptotic properties of regression kink models, and Masten and Poirier (2017) who conduct inference on breakdown frontiers. Finally, in a highly complementary paper, Hong and Li (2017) build on our results and propose employing a numerical derivative to obtain the estimator $\hat{\phi}'_n$.

The remainder of the article is organized as follows. Section 2 formally introduces the model we study and contains a minor extension of the Delta method for directionally differentiable functions. In Section 3 we characterize necessary and sufficient conditions for bootstrap consistency, develop an alternative method for estimating the asymptotic distribution of $\phi(\hat{\theta}_n)$, and study the local properties of this approach. Section 4 applies these results to develop a test of whether a Hilbert space valued parameter belongs to a closed convex set. Finally, in Section 5 we employ our framework to study trends in residual wage inequality in the U.S. A Supplementary

3. We discuss the relationship between these conditions in detail in Section 3.2; see Remark 3.1.

Appendix includes auxiliary results as well as examples illustrating how to verify our assumptions in specific applications.

2. SETUP AND BACKGROUND

In this section, we introduce the appropriate notion of directional differentiability and review an extension of the Delta method due to Shapiro (1991) and Dümbgen (1993).

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. 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 (*i.e.* complete normed spaces) with norms $\|\cdot\|_\mathbb{D}$ and $\|\cdot\|_\mathbb{E}$, and $\mathbb{D}_\phi$ the domain of ϕ.

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

$$r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0 \tag{4}$$

is then understood to be in $\mathbb{D}$ and with respect to the joint law of $\{X_i\}_{i=1}^n$. For instance, if $\{X_i\}_{i=1}^n$ is an i.i.d. sample and each X_i 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 our results are applicable to dependent settings as well. In addition, we note that the convergence in distribution in (4) is meant in the Hoffman-Jørgensen sense, which does not require $\hat{\theta}_n$ to be measurable—regrettably, measurability complications can arise naturally when $\mathbb{D}$ is infinite dimensional (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 Supplementary Appendix when necessary.

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

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

If in addition $\mathbf{A}$ is a compact subset of some metric space, then all continuous functions on $\mathbf{A}$ are bounded and hence belong to $\ell^\infty(\mathbf{A})$. We denote this important subspace by

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

2.1.1. Examples. In order to fix ideas, we introduce examples to which we return throughout the article to clarify our results. We defer a formal analysis to the Supplementary Appendix, where we also discuss additional applications and illustrate the verification of our assumptions.

Our first example, due to Andrews (2001), is mainly expository in nature.

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

$$\phi(\theta_0) = \max\{E[X], 0\}.$$

Here, $\theta_0 = E[X]$, $\mathbb{D} = \mathbb{E} = \mathbf{R}$, and $\phi : \mathbf{R} \to \mathbf{R}$ satisfies $\phi(\theta) = \max\{\theta, 0\}$. ‖

As our second example, we apply our analysis to obtain new results for a model previously studied by Beresteanu and Molinari (2008) and Bontemps *et al*. (2012).

Example 2.2 *(Interval Outcome Regression). Let $Y \in \mathbf{R}$, $Z \in \mathbf{R}^{d_z}$ be a set of covariates, ε be unobserved, and $\beta_0 \in \mathbf{R}^{d_z}$ be the best linear predictor coefficient satisfying*

$$Y = Z'\beta_0 + \varepsilon$$

with $E[\varepsilon Z] = 0$. Suppose Y is unobserved, but we instead see lower and upper bounds (Y_l, Y_u) satisfying $P(Y_l \leq Y \leq Y_u) = 1$; e.g., *$Y$ may be an interval censored income measure as in Manski and Tamer (2002). In such a setting, β_0 is not identified and Beresteanu and Molinari (2008) show its identified set is equal to*

$$\mathcal{B}_0 \equiv \{\beta \in \mathbf{R}^{d_z} : \beta = (E[ZZ'])^{-1}E[Z\tilde{Y}] \text{ for some } \tilde{Y} \text{ satisfying } Y_l \leq \tilde{Y} \leq Y_u \text{ a.s.}\}.$$

Since $\mathcal{B}_0$ is closed and convex, it follows that the identified set for any coordinate of β_0 is a closed interval in $\mathbf{R}$. We consider as the parameter of interest the largest endpoint of such an interval which, for a suitable choice of $p \in \mathbf{R}^{d_z}$, we may write as

$$\sup_{\beta \in \mathcal{B}_0} p'\beta \tag{5}$$

For example, $p = (1, 0, \ldots, 0)'$ yields the largest value in the identified set for the first coordinate of β_0. For our analysis, it is convenient to employ an equivalent representation of (5) due to Bontemps et al. *(2012) who show that*

$$\sup_{\beta \in \mathcal{B}_0} p'\beta = E[b_0'ZY_l + \max\{b_0'Z, 0\}(Y_u - Y_l)], \tag{6}$$

where $b_0 \equiv (E[ZZ'])^{-1}p$. Following (6) we set $\theta_0 = (b_0, \psi_0)$ for $\psi_0 : \mathbf{R}^{d_z} \to \mathbf{R}$ given by

$$\psi_0(b) \equiv E[b'ZY_l + \max\{b'Z, 0\}(Y_u - Y_l)], \tag{7}$$

and note (6) equals $\psi_0(b_0)$. Therefore, for any compact set $\mathbf{B}$ containing b_0 in its interior, we let $\mathbb{D} = \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ and define $\phi(\theta) = \psi(b)$ for any $\theta = (b, \psi) \in \mathbf{B} \times \ell^\infty(\mathbf{B})$. ‖

2.2. *Differentiability concepts*

In both of the previous examples, there exist points at which the map ϕ is not differentiable. Nonetheless, at all such points at which differentiability is lost, ϕ actually remains directionally differentiable. This is most easily seen in Example 2.1, in which the domain of ϕ is the real line. In order to address Example 2.2 and other applications of interest, however, a notion of directional differentiability that is suitable for more abstract spaces $\mathbb{D}$ is necessary. We therefore follow Shapiro (1990) and define:

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

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

$$\lim_{n \to \infty} \left\| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi_\theta'(h) \right\|_{\mathbb{E}} = 0, \tag{8}$$

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*

$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \|_{\mathbb{E}} = 0, \tag{9}$$

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 normed vector spaces (Reeds, 1976; Gill *et al*., 1989). It is therefore natural to employ an analogous approximation requirement when considering an appropriate definition of a directional derivative (compare (8) and (9)). Despite this similarity, two key differences distinguish (full) Hadamard differentiability from Hadamard directional differentiability. First, in (9) 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 (9) implies ϕ'_θ must be continuous and homogenous of degree one (Shapiro, 1990). As the next proposition shows, whether ϕ'_θ is linear or non-linear is the key property distinguishing whether ϕ is fully or directionally Hadamard differentiable.

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

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) and Kaido (2016) for examples.

The desired generalization of the Delta method only relies on two key assumptions.

Assumption 1 *(On the Map* ϕ*).*

(i) $\mathbb{D}$ *and* $\mathbb{E}$ *are Banach spaces with norms* $\|\cdot\|_{\mathbb{D}}$ *and* $\|\cdot\|_{\mathbb{E}}$.

(ii) $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$ *is Hadamard directionally differentiable at* θ_0 *tangentially to* $\mathbb{D}_0$.

Assumption 2 *(On the Estimator* $\hat{\theta}_n$*)*

(i) $\theta_0 \in \mathbb{D}_\phi$ *and* $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbb{D}_\phi$ *satisfies* $r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0$ *in* $\mathbb{D}$ *for some* $r_n \uparrow \infty$.

(ii) $\mathbb{G}_0$ *is tight and its support is included in* $\mathbb{D}_0$.

Assumption 1 formalizes our previous discussion by requiring that the map ϕ be Hadamard directionally differentiable at θ_0. In Assumption 2(i), we impose that the estimator $\hat{\theta}_n$ for θ_0 be asymptotically distributed according to some limit $\mathbb{G}_0$. The scaling r_n equals $\sqrt{n}$ in Examples 2.1

and 2.2, but may differ in non-parametric problems. Finally, Assumption 2(ii) requires that the support of the limiting process $\mathbb{G}_0$ be included in the tangential set $\mathbb{D}_0$. In addition, Assumption 2(ii) imposes that $\mathbb{G}_0$ be tight, which is a mild regularity condition that is automatically satisfied when $\mathbb{G}_0$ is finite dimensional. As previously noted by Shapiro (1991) and Dümbgen (1993), Assumptions 1 and 2 imply the validity of the Delta method for directionally differentiable maps.

Theorem 2.1 *Let Assumption 1 and Assumption 2 hold. Then, it follows that*

$$r_n\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}=\phi'_{\theta_0}(r_n\{\hat{\theta}_n-\theta_0\})+o_p(1), \tag{10}$$

and therefore $r_n\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}\xrightarrow{L}\phi'_{\theta_0}(\mathbb{G}_0)$ *in* $\mathbb{E}$.

The asymptotic distribution of $\phi(\hat{\theta}_n)$ was first established under measurability assumptions by Shapiro (1991). These measurability requirements were subsequently relaxed by Dümbgen (1993). Here, we provide a mild extension to their results by establishing the Delta method also holds "in probability" (*i.e.* result (10)), which we require for our analysis. Heuristically, the asymptotic distribution of $\phi(\hat{\theta}_n)$ follows from

$$r_n\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}\approx\phi'_{\theta_0}(r_n\{\hat{\theta}_n-\theta_0\}), \tag{11}$$

Assumption 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 (11) holds in an appropriate sense—a condition ensured by Hadamard directional differentiability. Following this insight, Theorem 2.1 can be established using the exact same arguments as in the proof of the Delta method for (fully) Hadamard differentiable maps (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, *e.g.*, Example 2.1.

Remark 2.1 *The Hadamard directional differentiability of ϕ at θ_0 only demands that ϕ'_{θ_0} be well defined on the domain $\mathbb{D}_0$; see Definition 2.1(ii). While $\mathbb{G}_0$ belongs to $\mathbb{D}_0$ with probability one by Assumption 2(ii), $r_n\{\hat{\theta}_n-\theta_0\}$ may not belong to $\mathbb{D}_0$ and thus not be in the domain of ϕ'_{θ_0}. In such instances, the expression $\phi'_{\theta_0}(r_n\{\hat{\theta}_n-\theta_0\})$ in (10) can be understood as the value a continuous extension of ϕ'_{θ_0} to $\mathbb{D}$ takes at the point $r_n\{\hat{\theta}_n-\theta_0\}$; see the proof of Theorem 2.1 for additional details.* ‖

2.3.1. Examples revisited. We next revisit our examples to illustrate how to apply Theorem 2.1.

Example 2.1 (cont.) Recall in this example $\theta_0=E[X]$ and $\phi(\theta)=\max\{\theta,0\}$ for any $\theta\in\mathbf{R}$. Given a sample $\{X_i\}_{i=1}^n$ we set $\hat{\theta}_n=\bar{X}_n\equiv\sum_{i=1}^n X_i/n$, and by direct calculation it is straightforward to verify Assumption 1 holds with $\mathbb{D}_0=\mathbf{R}$ and directional derivative

$$\phi'_{\theta_0}(h)=\begin{cases} h & \text{if } \theta_0>0 \\ \max\{h,0\} & \text{if } \theta_0=0 \\ 0 & \text{if } \theta_0<0 \end{cases} \tag{12}$$

for any $h\in\mathbf{R}$. In accord with Proposition 2.1, we note that ϕ'_{θ_0} is non-linear if and only if ϕ is not fully differentiable at θ_0 (*i.e.* $\theta_0=0$). Verifying Assumption 2 then simply requires a central limit theorem to apply, in which case we obtain

$$\sqrt{n}\{\max\{\bar{X}_n,0\}-\max\{E[X],0\}\}\xrightarrow{L}\phi'_{\theta_0}(\mathbb{G}_0)=\begin{cases}\mathbb{G}_0 & \text{if } E[X]>0\\ \max\{\mathbb{G}_0,0\} & \text{if } E[X]=0\\ 0 & \text{if } E[X]<0\end{cases} \tag{13}$$

by Theorem 2.1 and where $\mathbb{G}_0\in\mathbf{R}$ is normally distributed. ‖

Example 2.2 (cont.) In this application $\theta_0=(b_0,\psi_0)$ where $b_0=(E[ZZ'])^{-1}p\in\mathbf{R}^{d_z}$ and the function $\psi_0:\mathbf{R}^{d_z}\to\mathbf{R}$ was defined in (7). Given a sample $\{Y_{l,i},Y_{u,i},Z_i\}_{i=1}^n$ we employ as estimators $\hat{b}_n\equiv(\sum_{i=1}^n Z_iZ_i'/n)^{-1}p$, the function $\hat{\psi}_n:\mathbf{R}^{d_z}\to\mathbf{R}$ given by

$$\hat{\psi}_n(b)\equiv\frac{1}{n}\sum_{i=1}^n\{b'Z_iY_{l,i}+\max\{b'Z_i,0\}(Y_{u,i}-Y_{l,i})\},$$

and let $\hat{\theta}_n=(\hat{b}_n,\hat{\psi}_n)$. Since $E[ZZ']$ is invertible, it is possible to establish under appropriate moment conditions that for any compact $\mathbf{B}\subset\mathbf{R}^{d_z}$ the estimator $\hat{\theta}_n$ satisfies

$$\sqrt{n}\{\hat{\theta}_n-\theta_0\}\xrightarrow{L}\mathbb{G}_0=(\mathbb{G}_b,\mathbb{G}_\psi)\in\mathbf{R}^{d_z}\times\ell^\infty(\mathbf{B})$$

for some Gaussian $\mathbb{G}_0$; see the Supplementary Appendix. In particular, $\mathbb{G}_\psi$ has almost sure continuous sample paths, so that Assumption 2 holds with $\mathbb{D}_0=\mathbf{R}^{d_z}\times\mathcal{C}(\mathbf{B})$. Recall here $\phi(\theta)=\psi(b)$ for any $\theta=(b,\psi)\in\mathbf{B}\times\ell^\infty(\mathbf{B})$. Hence, setting $\mathbf{B}$ so that b_0 belongs to its interior, we obtain that $\hat{b}_n\in\mathbf{B}$ with probability tending to one and

$$\phi(\hat{\theta}_n)=\frac{1}{n}\sum_{i=1}^n\left\{p'\left(\frac{1}{n}\sum_{j=1}^n Z_jZ_j'\right)^{-1}Z_iY_{l,i}\right.$$
$$\left.+\max\left\{p'\left(\frac{1}{n}\sum_{j=1}^n Z_iZ_i'\right)^{-1}Z_i,0\right\}(Y_{u,i}-Y_{l,i})\right\}, \tag{14}$$

which equals the "plug-in" estimator proposed by Bontemps *et al.* (2012). We also note ϕ is Hadamard directionally differentiable tangentially to $\mathbb{D}_0$ at θ_0 with ϕ'_{θ_0} satisfying

$$\phi'_{\theta_0}(h)=h_\psi(b_0)+E[h_b'Z(Y_l+(Y_u-Y_l)1\{b_0'Z>0\})]+E[\max\{h_b'Z,0\}(Y_u-Y_l)1\{b_0'Z=0\}] \tag{15}$$

for any $h=(h_b,h_\psi)\in\mathbf{R}^{d_z}\times\mathcal{C}(\mathbf{B})$. By Proposition 2.1, it then follows ϕ is Hadamard differentiable at θ_0 if and only if $P(p'(E[ZZ'])^{-1}Z=0)=0$. Moreover, applying Theorem 2.1 and (15) we obtain that $\sqrt{n}\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}$ converges in distribution to

$$\mathbb{G}_\psi(b_0)+E[\mathbb{G}_b'Z(Y_l+(Y_u-Y_l)1\{b_0'Z>0\})]+E[\max\{\mathbb{G}_b'Z,0\}(Y_u-Y_l)1\{b_0'Z=0\}], \tag{16}$$

where expectations are taken over (Y_l,Y_u,Z) (but not $\mathbb{G}_0$). The asymptotic distribution in (16) can be shown to equal that obtained by Bontemps *et al.* (2012). ‖

It is also instructive to discuss an example of a map ϕ that is not Hadamard directionally differentiable. To this end, we examine a simplification of the regression kink model in Hansen (2017). Suppose that $(Y,Z)\in\mathbf{R}^2$ satisfies for some $(\gamma_0,\beta_0)\in\mathbf{R}^2$

$$Y=|Z-\gamma_0|\beta_0+\varepsilon$$

with $E[\varepsilon|Z]=0$. Let $\theta_0=(\beta_0,\gamma_0)$ and $\hat{\theta}_n=(\hat{\beta}_n,\hat{\gamma}_n)$ be the non-linear least square estimates, which Hansen (2017) shows to be asymptotically normally distributed. For some bounded $\mathbf{B}\subset\mathbf{R}$ containing γ_0 in its interior, we consider the problem of estimating $E[Y|Z=z]$ at different values of z in $\mathbf{B}$. To this end, we set $\phi:\mathbf{R}^2\to\ell^\infty(\mathbf{B})$ to equal

$$\phi(\theta)(z)=|z-\gamma|\beta$$

for any $\theta=(\beta,\gamma)$ and $z\in\mathbf{B}$ – notice $\phi(\hat{\theta}_n)$ is then a function mapping each value of $z\in\mathbf{B}$ into the corresponding forecast for Y. Further let $\phi'_{\theta_0}:\mathbf{R}^2\to\ell^\infty(\mathbf{B})$ be given by

$$\phi'_{\theta_0}(h)(z)\equiv(1\{z<\gamma_0\}\gamma+1\{z=\gamma_0\}|\gamma|-1\{z>\gamma_0\}\gamma)\beta_0+|z-\gamma_0|\beta$$

for any $h=(\beta,\gamma)\in\mathbf{R}^2$ and $z\in\mathbf{B}$. At any fixed $z\in\mathbf{B}$ we can then conclude that

$$\lim_{n\to\infty}|\frac{1}{t_n}\{\phi(\theta_0+t_nh_n)(z)-\phi(\theta_0)(z)\}-\phi'_{\theta_0}(h)(z)|$$
$$=\lim_{n\to\infty}|\frac{1}{t_n}\{|z-(\gamma_0+t_n\gamma_n)|(\beta_0+t_n\beta_n)-|z-\gamma_0|\beta_0\}-\phi'_{\theta_0}(h)(z)|=0 \quad (17)$$

for any $t_n\downarrow 0$ and $h_n=(\gamma_n,\beta_n)\to(\beta,\gamma)=h\in\mathbf{R}^2$. Hence, the forecast at a *fixed* z is a Hadamard directionally differentiable function of (β,γ). However, $\phi:\mathbf{R}^2\to\ell^\infty(\mathbf{B})$ is *not* Hadamard directionally differentiable because (17) fails to hold uniformly in $z\in\mathbf{B}$.[4] Thus, while Theorem 2.1 establishes the asymptotic distribution of a forecast at a *fixed* z, it fails to deliver the joint asymptotic distribution of the *collection* of forecasts indexed by $z\in\mathbf{B}$ (in $\ell^\infty(\mathbf{B})$). However, this is not a weakness of the Delta method: It can be shown that $\sqrt{n}\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}$ in fact does not converge in distribution in $\ell^\infty(\mathbf{B})$.

3. THE BOOTSTRAP

While Theorem 2.1 enables us to obtain an asymptotic distribution, a suitable method for estimating this limiting law is still required. In this section, we present a new result establishing necessary and sufficient conditions for the bootstrap to provide a consistent estimate of the asymptotic distribution of $r_n\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}$. We further propose an alternative to the bootstrap that generalizes existing approaches in the literature.

3.1. *Bootstrap setup*

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 conditional law of

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

4. Noting that here $\|\cdot\|_{\mathbb{E}}=\|\cdot\|_\infty$, the claim can be established by setting $h=(\beta,\gamma)$, employing a sequence $z_n=\gamma_0+\gamma t_n$, and then showing $|(\phi(\theta_0+t_nh)(z_n)-\phi(\theta_0)(z_n))/t_n-\phi'_{\theta_0}(h)(z_n)|\not\to 0$.

given the data. In order to allow for diverse resampling schemes, we simply impose that $\hat{\theta}_n^*$ be a function of the data $\{X_i\}_{i=1}^n$ and random weights $\{W_i\}_{i=1}^n$ that are independent of $\{X_i\}_{i=1}^n$. This general definition encompasses the non-parametric, Bayesian, block, m out of n, score, and weighted bootstrap as special cases.

Formalizing the notion of bootstrap consistency further requires us to employ a measure of distance between the limiting distribution and its bootstrap estimator. To 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 $\|\cdot\|_{\mathbf{A}}$, we define the set of functions

$$\mathrm{BL}_1(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} \text{ s.t. } |f(a)| \leq 1 \text{ and } |f(a) - f(a')| \leq \|a - a'\|_{\mathbf{A}} \text{ for all } a, a' \in \mathbf{A}\}.$$

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

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

Given the introduced notation, we can measure the distance between the conditional law of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ given $\{X_i\}_{i=1}^n$ and the limiting distribution of $r_n\{\hat{\theta}_n - \theta_0\}$ by[5]

$$\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}) | \{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|. \tag{18}$$

Employing the conditional distribution of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ given the data to approximate the distribution of $\mathbb{G}_0$ is then asymptotically justified if their distance, equivalently (18), converges in probability to zero. This type of consistency can be employed to validate the use of critical values obtained from the conditional distribution of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ given $\{X_i\}_{i=1}^n$ to conduct inference or construct confidence regions (Kosorok, 2008).

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

Assumption 3 *(On the Bootstrap $\hat{\theta}_n^*$)*

(i) $\hat{\theta}_n^* : \{X_i, W_i\}_{i=1}^n \to \mathbb{D}_\phi$ *with* $\{W_i\}_{i=1}^n$ *independent of* $\{X_i\}_{i=1}^n$.
(ii) $\hat{\theta}_n^*$ *satisfies* $\sup_{f \in BL_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}) | \{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| = o_p(1)$.
(iii) $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ *is asymptotically measurable (jointly in* $\{X_i, W_i\}_{i=1}^n$*).*
(iv) $f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})$ *is a measurable function of* $\{W_i\}_{i=1}^n$ *outer almost surely in* $\{X_i\}_{i=1}^n$ *for any continuous and bounded* $f : \mathbb{D} \to \mathbf{R}$.

Assumption 3(i) defines $\hat{\theta}_n^*$ in accord with our discussion, while Assumption 3(ii) imposes the consistency of the conditional law of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ given the data for the distribution of $\mathbb{G}_0$—*i.e.* the bootstrap "works" for the estimator $\hat{\theta}_n$. We note that in finite dimensional problems (*i.e.* $\theta_0 \in \mathbf{R}^d$), Assumption 3(ii) is equivalent to the bootstrap cdf being consistent for the cdf of $\mathbb{G}_0$ at all its continuity points. In addition, in Assumptions 3(iii)-(iv) we demand mild measurability requirements on $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$.

5. 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.

3.2. *Bootstrap failure*

Whenever ϕ is (fully) Hadamard differentiable and the bootstrap "works" for $\hat{\theta}_n$, the consistency of the bootstrap is inherited by $\phi(\hat{\theta}_n)$—*i.e.* the limiting distribution of $r_n\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}$ can be consistently estimated by the conditional law of

$$r_n\{\phi(\hat{\theta}_n^*)-\phi(\hat{\theta}_n)\} \tag{19}$$

given the data (Bickel and Freedman, 1981; van der Vaart and Wellner, 1996). We refer to the conditional law of (19) given 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 (Bickel *et al.*, 1997; Andrews, 2000). These examples serve as a warning that the standard bootstrap may fail when ϕ is not (fully) Hadamard differentiable, yet can provide little guidance as to whether the standard bootstrap is actually valid in particular applications. Our first main result establishes that these examples are in fact special cases of a deeper principle, namely that whenever $\mathbb{G}_0$ is Gaussian the standard bootstrap is consistent *if and only if* ϕ is (fully) differentiable at θ_0.

Theorem 3.1 *Let Assumptions 1, 2, and 3 hold, and suppose that $\mathbb{G}_0$ is Gaussian and its support is a vector subspace of $\mathbb{D}$. Then, it follows that ϕ is (fully) Hadamard differentiable at $\theta_0 \in \mathbb{D}_\phi$ tangentially to the support of $\mathbb{G}_0$ if and only if*

$$\sup_{f\in BL_1(\mathbb{E})} |E[f(r_n\{\phi(\hat{\theta}_n^*)-\phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n]-E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1). \tag{20}$$

A powerful implication of Theorem 3.1 is that in verifying whether the standard bootstrap is valid at a conjectured θ_0, a researcher need only verify whether ϕ is (fully) differentiable at θ_0. In effect, Theorem 3.1 thus reduces the potentially challenging statistical problem of verifying bootstrap validity to a simple and purely analytical calculation. The theorem requires both that $\mathbb{G}_0$ be Gaussian and that its support be a vector subspace of $\mathbb{D}$. The former requirement may be relaxed at the cost of additional notation, and we thus focus on the Gaussian case due to its ubiquity; see Remark 3.2. In addition, we note that under Gaussianity the condition that the support of $\mathbb{G}_0$ be a vector subspace is equivalent to zero (in $\mathbb{D}$) belonging to the support of $\mathbb{G}_0$.

A further implication of Theorem 3.1 that merits discussion follows from employing that Gaussianity of $\mathbb{G}_0$ and bootstrap consistency together imply ϕ is (fully) differentiable and hence that ϕ'_{θ_0} must be linear. 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 $\phi'_{\theta_0}(\mathbb{G}_0)$ or, equivalently, Gaussianity of the asymptotic distribution of $\phi(\hat{\theta}_n)$. Conversely, we conclude that *the standard bootstrap fails whenever the limiting distribution is not Gaussian*. Thus, the presence of a non-Gaussian limiting distribution may be viewed by practitioners as a simple yet reliable signal of the failure of the standard bootstrap. We formalize this conclusion in the following Corollary.

Corollary 3.1 *Let Assumptions 1, 2, 3 hold, and suppose $\mathbb{G}_0$ is Gaussian and its support is a vector subspace of $\mathbb{D}$. If the limiting distribution of $r_n\{\phi(\hat{\theta}_n)-\phi(\theta_0)\}$ is not Gaussian, then it follows that the standard bootstrap is inconsistent.*

Remark 3.1 *The limit of the standard bootstrap measure was first studied in Dümbgen (1993), whose results imply bootstrap consistency is equivalent to the distribution of*

$$\phi'_{\theta_0}(\mathbb{G}_0+h)-\phi'_{\theta_0}(h) \tag{21}$$

being constant in h for all h in the support of $\mathbb{G}_0$.[6] *This characterization of bootstrap consistency does not rely on Gaussianity and thus applies under more general conditions than those of Theorem 3.1. Our results complement Dümbgen (1993) by showing that, under the additional requirement that* $\mathbb{G}_0$ *be Gaussian, bootstrap consistency is in fact also equivalent to* ϕ *being (fully) differentiable at* θ_0. ‖

Remark 3.2 *Gaussianity of* $\mathbb{G}_0$ *plays an important role in the proof of Theorem 3.1 in enabling us to relate the distribution of* (21) *to that of* $\phi'_{\theta_0}(\mathbb{G}_0)$ *through the Cameron-Martin theorem. A similar insight is employed by van der Vaart (1991) and Hirano and Porter (2012) who compare characteristic functions in a limit experiment to conclude regular estimability of a functional implies its differentiability. More generally, Theorem 3.1 can be shown to hold provided the support of* $\mathbb{G}_0$ *is a vector subspace of* $\mathbb{D}$ *and the density of the distribution of* $\mathbb{G}_0+h$ *with respect to the law of* $\mathbb{G}_0$ *is suitably smooth for all h in a dense subspace of the support of* $\mathbb{G}_0$. ‖

3.3. *An alternative approach*

In what follows, we develop a resampling scheme that, unlike the standard bootstrap, remains valid when ϕ is Hadamard directionally (but not fully) differentiable. As we show, our approach can be interpreted as a generalization of certain existing methods and thus provides insights into their common structure and consistency.

3.3.1. Consistent alternative.

The Delta method, as stated in Theorem 2.1, establishes that the asymptotic distribution we aim to estimate is given by the law of $\phi'_{\theta_0}(\mathbb{G}_0)$. Crucially, the desired limiting distribution depends only on two unknowns: (i) The distribution of $\mathbb{G}_0$, and (ii) The directional derivative ϕ'_{θ_0}. By hypothesis, however, the bootstrap "works" for $\hat{\theta}_n$, and therefore the distribution of $\mathbb{G}_0$ can be consistently estimated by the conditional law of $r_n\{\hat{\theta}_n^*-\hat{\theta}_n\}$ given the data. The analogy principle thus suggests estimating the distribution of $\phi'_{\theta_0}(\mathbb{G}_0)$ by employing the conditional law given data of the statistic

$$\hat{\phi}'_n(r_n\{\hat{\theta}_n^*-\hat{\theta}_n\}), \tag{22}$$

where $\hat{\phi}'_n:\mathbb{D}\to\mathbb{E}$ is a suitable estimator of the directional derivative ϕ'_{θ_0}.[7]

In order to ensure the validity of this approach we impose the following condition:

Assumption 4 *(On the Estimator* $\hat{\phi}'_n$*)*

The map $\hat{\phi}'_n:\mathbb{D}\to\mathbb{E}$ *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*

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}P\Big(\sup_{h\in K^\delta}\|\hat{\phi}'_n(h)-\phi'_{\theta_0}(h)\|_{\mathbb{E}}>\epsilon\Big)=0. \tag{23}$$

6. See Theorem S.3.1 in the Supplementary Appendix for an analogous result under Assumptions 1–3.

7. For inference purposes, one might alternatively first obtain a confidence region $\mathcal{C}_n$ for θ_0 and employ $\phi(\mathcal{C}_n)$ as a confidence region for $\phi(\theta_0)$. Though valid under virtually no assumptions on ϕ, this "projection" approach is potentially quite conservative; see, *e.g.*, Dufour and Taamouti (2005), Romano and Shaikh (2008, 2010), and Woutersen and Ham (2013).

Intuitively, a minimum requirement to demand of the estimator $\hat{\phi}'_n$ is that $\hat{\phi}'_n(h)$ be consistent for $\phi'_{\theta_0}(h)$ at any $h \in \mathbb{D}_0$. However, since our estimand is not $\phi'_{\theta_0}(h)$ at a fixed h, but rather the distribution of $\phi'_{\theta_0}(\mathbb{G}_0)$, a certain amount of "uniformity" in the consistency of $\hat{\phi}'_n(h)$ (over h) is needed. Assumption 4 formalizes the appropriate notion of uniformity that we require. It is worth noting, however, that in many applications stronger, but simpler, conditions than (23) can be easily verified; see Remarks 3.3 and 3.4 below, and the examples discussed in the Supplementary Appendix.

Remark 3.3 *In certain applications, it is sufficient for the estimator $\hat{\phi}'_n$ to satisfy*

$$\sup_{h \in K} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1), \tag{24}$$

for any compact set $K \subseteq \mathbb{D}$. For instance, if $\mathbb{D} = \mathbf{R}^d$, then the closure of K^δ is compact for any compact K and $\delta < \infty$, and hence (24) implies (23). Alternatively, if $\mathbb{D}$ is separable and $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ is Borel measurable as a function of $\{X_i, W_i\}_{i=1}^n$, then condition (23) may be relaxed to (24) as well. ‖

Remark 3.4 *Assumption 4 greatly simplifies whenever the modulus of continuity of $\hat{\phi}'_n$ can be controlled. For instance, if $\|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)\|_{\mathbb{E}} \leq C_n \|h_1 - h_2\|_{\mathbb{D}}$ for all $h_1, h_2 \in \mathbb{D}$ and a possibly random C_n satisfying $C_n = O_p(1)$, then showing that for any $h \in \mathbb{D}_0$*

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

suffices for establishing (23) holds; see Lemma S.3.6 in the Supplementary Appendix. ‖

Given Assumption 4 we can now establish the validity of the proposed procedure.

Theorem 3.2 *If Assumptions 1, 2, 3, and 4 hold, then it follows that*

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

Theorem 3.2 shows that the conditional law of $\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\})$ given the data is indeed a consistent estimator for the asymptotic distribution of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$. Interestingly, multiple superficially different resampling schemes share the common structure in (22) but differ on their implicit choice of estimator $\hat{\phi}'_n$ of the directional derivative. For instance, in a variety of problems researchers have proposed sample analogue estimators of limiting distributions in "non-standard" problems; see, *e.g.*, Linton *et al.* (2005), Andrews and Soares (2010), and Andrews and Shi (2013). As we show in the Supplementary Appendix, these procedures may be interpreted as employing estimators $\hat{\phi}'_n$ that are implicitly inspired by an *analytical* computation of the directional derivative.

An estimator $\hat{\phi}'_n$ satisfying Assumption 4 may also be obtained by relying on *numerical* methods that avoid analytically computing ϕ'_{θ_0}. For example, building on our Theorem 3.2, Hong and Li (2017) propose estimating ϕ'_{θ_0} by employing

$$\hat{\phi}'_{s_n,n}(h) \equiv \frac{1}{s_n}\{\phi(\hat{\theta}_n + s_n h) - \phi(\hat{\theta}_n)\}, \tag{25}$$

which is a numerical derivative computed with step size s_n. Under Assumptions 1 and 2, it is possible to show that $\hat{\phi}'_{s_n,n}$ satisfies Assumption 4 provided that $r_n s_n \to \infty$; *i.e.* provided that the step size is of smaller order than the estimation uncertainty in $\hat{\theta}_n$ (see Lemma S.3.8 in the Supplementary Appendix). This numerical estimator is implicit in the rescaled bootstrap of Dümbgen (1993), which is numerically identical to composing $\hat{\phi}'_{s_n,n}$ with a non-parametric bootstrap estimator of the distribution of $\mathbb{G}_0$. Additionally, Hong and Li (2017) show the m out of n bootstrap of Shao (1994) is numerically equivalent to setting $s_n = 1/r_m$ and composing $\hat{\phi}'_{s_n,n}$ with the m out of n bootstrap estimator of the distribution of $\mathbb{G}_0$.

The wide array of possible choices for $\hat{\phi}'_n$ poses the question of what estimator $\hat{\phi}'_n$ to employ. While a conclusive answer to such question is beyond the scope of this article, we note that in our experience we have found estimators based on analytical expressions for ϕ'_{θ_0} to be preferable. In particular, employing the specific structure of ϕ'_{θ_0} it is often possible to devise estimators $\hat{\phi}'_n$ whose "tuning" parameter choices are simpler to motivate than that of the step size s_n in (25). On the other hand, such analytical estimators must be obtained on an application specific basis. In this regard, the numerical estimator offers the flexibility of being widely applicable.

It is also worth noting that the proof of Theorem 3.2 does not actually rely on ϕ'_{θ_0} being the directional derivative of ϕ at θ_0. Therefore, Theorem 3.2 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; see, *e.g.*, Chen and Fang (2015) for an application of this principle. Finally, it is important to emphasize that due to a 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 4.

Remark 3.5 *The standard bootstrap itself may also be interpreted as relying on a numerical derivative as in* (25) *with step size* $s_n = 1/r_n$ *since*

$$r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} = \hat{\phi}'_{r_n^{-1},n}(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}).$$

However, in accord with Theorem 3.1, the estimator $\hat{\phi}'_{r_n^{-1},n}$ does not *satisfy Assumption 4 whenever* ϕ *is not (fully) Hadamard differentiable. In fact, Theorem 2.1 implies that for any* $h \in \mathbb{D}$, $\hat{\phi}_{r_n^{-1},n}(h)$ *converges in distribution to* $\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(\mathbb{G}_0)$. *This conclusion underscores the importance of the condition* $s_n r_n \to \infty$ *in establishing the consistency of* $\hat{\phi}'_{s_n,n}$ *whenever* ϕ *is not (fully) Hadamard differentiable.* ‖

3.3.2. Examples revisited.

We next revisit Examples 2.1 and 2.2 to illustrate the implications of our results.

Example 2.1 (cont.) Recall in this example ϕ is fully Hadamard differentiable if and only if $\theta_0 \neq 0$. Hence, in accord with Andrews (2000), Theorem 3.1 implies the standard bootstrap is consistent if and only if $\theta_0 \neq 0$. Theorem 3.2, however, provides an alternative to the standard bootstrap. For instance, we may estimate ϕ'_{θ_0} employing

$$\hat{\phi}'_n(h) = \begin{cases} h & \text{if } \sqrt{n}\bar{X}_n/\hat{\sigma}_n > \kappa_n \\ \max\{h, 0\} & \text{if } |\sqrt{n}\bar{X}_n/\hat{\sigma}_n| \leq \kappa_n \\ 0 & \text{if } \sqrt{n}\bar{X}_n/\hat{\sigma}_n < -\kappa_n \end{cases}, \tag{26}$$

where $\hat{\sigma}_n$ is the sample standard deviation of $\{X_i\}_{i=1}^n$ and κ_n is a positive sequence satisfying $\kappa_n \uparrow \infty$. It is then straightforward to verify $\hat{\phi}'_n$ satisfies Assumption 4 provided $\kappa_n/\sqrt{n} \downarrow 0$. By

Theorem 3.2, we may therefore set $\hat{\theta}_n^* = \bar{X}_n^*$ for $\bar{X}_n^*$ the bootstrapped sample mean, and employ the conditional distribution given the data of

$$\hat{\phi}_n'(\sqrt{n}\{\bar{X}_n^* - \bar{X}_n\})$$

to estimate the law of $\phi_{\theta_0}'(\mathbb{G}_0)$ (as in (13)). With regards to the choice of κ_n, we note that $\hat{\phi}_n'$ may be understood as implicitly conducting a pretest of whether $\theta_0 = 0$. The condition $\kappa_n \uparrow \infty$ is equivalent to the Type I error of such pretest tending to zero, and thus we may let $\kappa_n = \Phi^{-1}(1-\alpha_n)$ for Φ the standard normal cdf and $\alpha_n \downarrow 0$. ‖

Example 2.2 (cont.) In this problem ϕ is fully Hadamard differentiable if and only if

$$P(p'(E[ZZ'])^{-1}Z = 0) = 0. \tag{27}$$

Therefore, Theorem 3.1 implies that the standard bootstrap is consistent if and only if (27) holds—a result previously conjectured by Beresteanu and Molinari (2008) and Bontemps *et al.* (2012). In order to devise a resampling scheme that remains consistent when condition (27) fails we apply Theorem 3.2. To this end, recall $\mathbb{D} = \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ and for any $h = (h_b, h_\psi) \in \mathbf{R}^{d_z} \times \ell^\infty(\mathbf{B})$ we estimate ϕ_{θ_0}' (as in (15)) by setting

$$\hat{\phi}_n'(h) = h_\psi(\hat{b}_n) + \frac{1}{n}\sum_{i=1}^n h_b' Z_i \left(Y_{l,i} + (Y_{u,i} - Y_{l,i}) 1\left\{ \frac{\sqrt{n}\hat{b}_n' Z_i}{\|Z_i\|} > \kappa_n \right\} \right) + \frac{1}{n}\sum_{i=1}^n \max\{h_b' Z_i, 0\}(Y_{u,i} - Y_{l,i}) 1\left\{ \left| \frac{\sqrt{n}\hat{b}_n' Z_i}{\|Z_i\|} \right| \leq \kappa_n \right\}, \tag{28}$$

where κ_n is a positive sequence satisfying $\kappa_n \uparrow \infty$. In the Supplementary Appendix we establish $\hat{\phi}_n'$ satisfies Assumption 4 provided $\kappa_n/\sqrt{n} \downarrow 0$. Moreover, for $\{X_i^*\}_{i=1}^n$ a sample drawn with replacement from $\{X_i\}_{i=1}^n$ and defining $\mathbb{G}_{b,n}^* \in \mathbf{R}^{d_z}$ and $\mathbb{G}_{\psi,n}^* \in \ell^\infty(\mathbf{B})$ by

$$\mathbb{G}_{b,n}^* \equiv \sqrt{n}\left\{ \left(\frac{1}{n}\sum_{i=1}^n Z_i^* Z_i^{*\prime} \right)^{-1} p - \hat{b}_n \right\}$$

$$\mathbb{G}_{\psi,n}^*(b) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \{ b' Z_i^* Y_{l,i}^* + \max\{b' Z_i^*, 0\}(Y_{u,i}^* - Y_{l,i}^*) - \hat{\psi}_n(b) \},$$

we may set $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ to equal $(\mathbb{G}_{b,n}^*, \mathbb{G}_{\psi,n}^*)$. Theorem 3.2 then implies the conditional distribution of $\hat{\phi}_n'((\mathbb{G}_{b,n}^*, \mathbb{G}_{\psi,n}^*))$ given the data is consistent for the law of $\phi_{\theta_0}'(\mathbb{G}_0)$ (as in (16)).[8] With regards to the choice of κ_n, we may follow a similar logic as in Example 2.1. In particular, since $\sup_{\|u\|=1}(\hat{b}_n - b_0)' u = \|\hat{b}_n - b_0\|$, we may set κ_n to equal the conditional $1-\alpha_n$ quantile of $\|\mathbb{G}_{b,n}^*\|$ given data for some $\alpha_n \downarrow 0$. ‖

8. To the best of our knowledge, the only previously available procedure for estimating the asymptotic distribution under a failure of (27) is subsampling (Politis *et al.*, 1999).

3.4. *Local analysis*

The asymptotic approximation implied by Theorem 2.1 may depend discontinuously on θ_0 when θ_0 is a point at which ϕ fails to be (fully) Hadamard differentiable; see, *e.g.*, Example 2.1. In contrast, the finite sample distribution of $\phi(\hat{\theta}_n)$ depends continuously on the value of the parameter θ_0. This discrepancy is cause for concern that tests justified by Theorems 2.1 and 3.2 may fail to adequately control size whenever ϕ is fully differentiable at θ_0 but θ_0 is "close" to a point at which ϕ fails to be (fully) differentiable. We next allay these concerns by conducting a complementary local analysis.

3.4.1. The local limit. The objective of our local analysis is to better model a finite sample situation in which θ_0 is "close" to a point at which ϕ fails to be (fully) differentiable. To this end, we therefore employ an asymptotic framework in which θ_0 depends on the sample size and converges to a point at which differentiability of ϕ may fail. We formalize such an approach by explicitly denoting the dependence of θ_0 on the underlying distribution of $\{X_i\}_{i=1}^n$ and letting said distribution change with the sample size.

For ease of exposition, we will assume $\{X_i\}_{i=1}^n$ is an i.i.d sample; see the Supplementary Appendix for results that apply to dependent data. We further let $\mathbf{P}$ denote the set of possible marginal distributions for X_i. In order to allow the distribution of the data to change with the sample size we next introduce the concept of a "path" (in $\mathbf{P}$). A "path" is simply an arbitrary univariate model $t \mapsto P_t \in \mathbf{P}$ for the distribution of X_i that is defined on a neighbourhood of zero and for some function g satisfies

$$\lim_{t\to 0}\int\left(\frac{1}{t}\{dP_t^{1/2}-dP_0^{1/2}\}-\frac{1}{2}g\,dP_0^{1/2}\right)^2=0, \tag{29}$$

where $dP_t^{1/2}$ denotes the square root of the density of P_t. A parametrization satisfying (29) is called quadratic mean differentiable and g is referred to as the "score".[9] We may then let the distribution of X_i change with the sample size n by setting it to equal $P_{\lambda/\sqrt{n}}$ for some $\lambda \in \mathbf{R}$ – here λ is called the "local" parameter. Notice that the distribution of X_i then converges to P_0 and, moreover, by (29) such convergence is suitably smooth.

Finally, we make the dependence of the parameter estimated by $\hat{\theta}_n$ on the distribution of X_i explicit by writing $\theta(P)$ for the value said estimand takes when X_i is distributed according to P—*i.e.* we characterize said estimand through a map $\theta:\mathbf{P}\to\mathbb{D}$. Under our local framework, $\theta(P_{\lambda/\sqrt{n}})$ then converges to $\theta(P_0)$ and, with a mild abuse of notation, we write this limiting value as $\theta_0=\theta(P_0)$. In this manner, we formalize the notion that in finite samples the parameter estimated by $\hat{\theta}_n$ (*i.e.* $\theta(P_{\lambda/\sqrt{n}})$) may be "close" to a point (*i.e.* $\theta_0=\theta(P_0)$) at which ϕ is possibly not (fully) differentiable.

The paths we consider in the described "local" construction are required to satisfy:

Assumption 5 *(Local Analysis) The path $t\mapsto P_t$ satisfies for any λ the following:*

(i) There is a $\theta'(\lambda)\in\mathbb{D}_0$ such that $\|r_n\{\theta(P_{\lambda/\sqrt{n}})-\theta(P_0)\}-\theta'(\lambda)\|_{\mathbb{D}}\to 0$.

(ii) $\theta(P_{\lambda/\sqrt{n}})\in\mathbb{D}_\phi$ and $r_n\{\hat{\theta}_n-\theta(P_{\lambda/\sqrt{n}})\}\xrightarrow{L_\lambda}\mathbb{G}_0$, where $\xrightarrow{L_\lambda}$ denotes convergence in distribution under $\{X_i\}_{i=1}^n$ i.i.d. with each X_i distributed according to $P_{\lambda/\sqrt{n}}$.

(iii) $\mathbb{G}_0$ is tight and its support is included in $\mathbb{D}_0$.

9. If $dP_t(x)$ is differentiable in t, then under appropriate conditions $g(x)=\frac{\partial}{\partial t}\log dP_t(x)|_{t=0}$.

Assumption 5(i) demands that the parameter $\theta(P_t)$ be suitably differentiable as a function of t. In turn, Assumption 5(ii) imposes that the asymptotic behaviour of $\hat{\theta}_n$ not be affected by local perturbations to the data-generating process—estimators satisfying such property for any path in $\mathbf{P}$ are often called "regular". Finally, Assumption 5(iii) clarifies that the limiting distribution $\mathbb{G}_0$ in Assumption 5(ii) must satisfy the same regularity conditions as those imposed in Assumption 2(ii). These requirements are sufficiently general to enable us to model a rich set of applications in which ϕ is fully differentiable at $\theta(P)$ but $\theta(P)$ is "close" to a point at which full differentiability fails; see our discussion of Examples 2.1 and 2.2 below.

The following Lemma is a simple modification of Theorem 2.1, and can be interpreted as extending the local analysis in Dümbgen (1993) to contiguous perturbations.

Lemma 3.1 *If Assumptions 1 and 5 hold, then for any $\lambda \in \mathbf{R}$ it follows that*

$$r_n\{\phi(\hat{\theta}_n)-\phi(\theta(P_{\lambda/\sqrt{n}}))\} \stackrel{L_\lambda}{\to} \phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda))-\phi'_{\theta_0}(\theta'(\lambda)), \tag{30}$$

where $\theta_0=\theta(P_0)$ and $\stackrel{L_\lambda}{\to}$ denotes convergence in law under $\{X_i\}_{i=1}^n \sim \bigotimes_{i=1}^n P_{\lambda/\sqrt{n}}$.

Crucially, the asymptotic distribution of Lemma 3.1 depends on the local parameter λ whenever ϕ'_{θ_0} is non-linear—*i.e.* whenever ϕ is not (fully) Hadamard differentiable at θ_0. In particular, Lemma 3.1 provides an approximation which reflects that how "close" $\theta(P)$ is to a point at which ϕ is not fully differentiable can impact the finite sample distribution of $\phi(\hat{\theta}_n)$. Such conclusion contrasts with the pointwise (in P) analysis of Theorem 2.1 and emphasizes care should be taken when applying Theorems 2.1 and 3.2 for inference. As we next argue, however, Lemma 3.1 also provides the key for understanding when our analysis delivers inference procedures with reliable size control.

3.4.2. Implications for testing.

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

$$H_0:\phi(\theta(P))\leq 0 \qquad\qquad H_1:\phi(\theta(P))>0. \tag{31}$$

A natural test statistic for this problem is given by $r_n\phi(\hat{\theta}_n)$, and a natural test is to reject for large values of said test statistic. In particular, Theorem 3.2 suggests that a level α test can be constructed by comparing $r_n\phi(\hat{\theta}_n)$ to the critical value

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

that is, $\hat{c}_{1-\alpha}$ is the $1-\alpha$ conditional quantile of $\hat{\phi}'_n(r_n\{\hat{\theta}^*_n-\hat{\theta}_n\})$ given the data. The quantile $\hat{c}_{1-\alpha}$ can be approximated via simulation by computing multiple draws of $\hat{\theta}^*_n$.

In order to evaluate the ability of the proposed test to provide size control, we evaluate its rejection probability along a "local" sequence of distributions. In particular, we focus on sequences of distributions that converge to a distribution P_0 such that $\phi(\theta(P_0))=0$—*i.e.* such that P_0 is on the "boundary" of the null hypothesis. Our next result builds on Lemma 3.1 to characterize the desired limiting rejection probability.

Theorem 3.3 *Let $\{X_i\}_{i=1}^n$ be i.i.d. with X_i distributed according to $P_{\lambda/\sqrt{n}}$ for some path $t\mapsto P_t$, and set $\theta_0=\theta(P_0)$. If Assumptions 1, 3, 4, 5 hold, $\phi(\theta_0)=0$, and the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ is continuous* *and increasing at its* $1-\alpha$ *quantile, denoted* $c_{1-\alpha}$*, then*

$$\limsup_{n\to\infty} P^n_{\lambda/\sqrt{n}}(r_n\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) \geq P(\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda)) > c_{1-\alpha}). \tag{32}$$

Moreover: (i) Result (32) holds with equality whenever $c_{1-\alpha}$ *is a continuity point of the cdf of* $\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda))$*, and (ii) The limiting rejection probability equals* α *when* $\lambda=0$*.*

Because the local sequence may approach P_0 from the "null" (*i.e.* $\phi(\theta(P_{\lambda/\sqrt{n}}))\leq 0$) or the "alternative" (*i.e.* $\phi(\theta(P_{\lambda/\sqrt{n}}))>0$), Theorem 3.3 can be employed to study both the local size and power of the proposed test. The presence of an inequality in (32), rather than an equality, is due to convergence in distribution only implying convergence of the corresponding cdfs at continuity points of the limit. Indeed, as noted by Theorem 3.3, result (32) holds with equality whenever $c_{1-\alpha}$ is a continuity point of the cdf of $\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda))$. Notably, since $c_{1-\alpha}$ is by hypothesis a continuity point of the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$, Theorem 3.3 implies the limiting rejection probability is equal to α whenever $\lambda=0$—*i.e.* whenever P does not change with the sample size. In other words, Theorem 3.3 implies "pointwise" (in P) size control of the proposed test.

In accord with the spirit of the Delta method, a purely analytical condition suffices for leveraging Theorem 3.3 to show the proposed test can also provide local size control. Concretely, whenever ϕ'_{θ_0} is convex and $\mathbb{G}_0$ is Gaussian, the conclusion of Theorem 3.3 can be strengthened to show that if $P_{\lambda/\sqrt{n}}$ satisfies the null hypothesis for all n, then

$$\lim_{n\to\infty} P^n_{\lambda/\sqrt{n}}(r_n\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) = P(\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda)) > c_{1-\alpha}); \tag{33}$$

that is, convexity of ϕ'_{θ_0} enables us to verify continuity of the limiting cdf at $c_{1-\alpha}$. As a result of (33), it follows that the proposed test delivers local size control if and only if

$$P(\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda)) > c_{1-\alpha}) \leq \alpha \tag{34}$$

for *any* sequence $\{P_{\lambda/\sqrt{n}}\}_{n=1}^{\infty}$ satisfying the null hypothesis. This potentially daunting to verify condition is fortunately also implied by convexity of ϕ'_{θ_0}. Intuitively, since $\phi(\theta(P_{\lambda/\sqrt{n}}))$ converges to $\phi(\theta(P_0))$ and $\phi(\theta(P_0))$ takes the largest value in the null hypothesis (*i.e.* zero), the derivative $\phi'_{\theta_0}(\theta'(\lambda))$ must be negative.[10] Hence, since ϕ'_{θ_0} is homogeneous of degree one by construction, convexity implies

$$\begin{aligned}
\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda)) &= 2\phi'_{\theta_0}\left(\frac{1}{2}\mathbb{G}_0+\frac{1}{2}\theta'(\lambda)\right) && \text{(homogeneity)}\\
&\leq \phi'_{\theta_0}(\mathbb{G}_0)+\phi'_{\theta_0}(\theta'(\lambda)) && \text{(convexity)}\\
&\leq \phi'_{\theta_0}(\mathbb{G}_0) && (\phi'_{\theta_0}(\theta'(\lambda))\leq 0).
\end{aligned}$$

In other words, the distribution of $\phi'_{\theta_0}(\mathbb{G}_0)$ is the "least favorable" (in the sense of having the largest quantiles) among all local limits. Since $c_{1-\alpha}$ is precisely the $1-\alpha$ quantile of this "least favorable" distribution, it follows that convexity of ϕ'_{θ_0} suffices for establishing the local size control of the proposed test (formally, (34) holds as required). In summary, the potentially challenging task of verifying local size control can be reduced to the simple to verify requirement that the directional derivative be convex.

The following Corollary formalizes our preceding discussion.

10. More precisely, we are employing that $\phi'_{\theta_0}(\theta'(\lambda)) = \lim_{n\to\infty} r_n\{\phi(\theta(P_{\lambda/\sqrt{n}}))-\phi(\theta(P_0))\}\leq 0$.

Corollary 3.2 *Let the conditions of Theorem 3.3 hold, $\mathbb{G}_0$ be Gaussian, $\mathbb{D}_0$ be a convex set, and $\phi'_{\theta_0}:\mathbb{D}_0\to\mathbf{R}$ be a convex function. If $\phi(\theta(P_{\lambda/\sqrt{n}}))\leq 0$ for all n, then*

$$\limsup_{n\to\infty} P^n_{\lambda/\sqrt{n}}(r_n\phi(\hat{\theta}_n)>\hat{c}_{1-\alpha})\leq\alpha.$$

While ϕ'_{θ_0} is convex in most of the examples we study, we emphasize that this condition can fail to hold. It is thus important to note that when ϕ'_{θ_0} is not convex, a test that rejects whenever $r_n\phi(\hat{\theta}_n)$ exceeds $\hat{c}_{1-\alpha}$ may fail to provide local size control. Nonetheless, Theorem 3.3 still gives us the means for obtaining a suitable critical value in such instances. In particular, in order for a test that rejects whenever $r_n\phi(\hat{\theta}_n)$ exceeds a critical value $c^\star$ to provide local size control, the critical value $c^\star$ must satisfy

$$P(\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda))>c^\star)\leq\alpha$$

whenever $\phi(\theta(P_{\lambda/\sqrt{n}}))\leq 0$ for all n. While the desired $c^\star$ may not equal $c_{1-\alpha}$ when ϕ'_{θ_0} is not convex, it is sometimes still possible to find $c^\star$ by employing the analytical expression for the directional derivative; see, *e.g.*, our discussion of Example 2.2 below. Finally, we emphasize that even though we have focused on hypothesis testing, the local analysis in this Section is equally applicable to the construction of confidence regions.

Remark 3.6 *The proof of Theorem 3.3 further implies that whenever $c_{1-\alpha}$ is a continuity point of the cdf of $\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda))$ we must have that*

$$\lim_{n\to\infty} P^n_{\lambda/\sqrt{n}}(r_n\phi(\hat{\theta}_n)>\hat{c}_{1-\alpha})=P(\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda))>c_{1-\alpha}). \tag{35}$$

Applying result (35) to sequences $\{P_{\lambda/\sqrt{n}}\}_{n=1}^\infty$ that approach P_0 from the alternative (i.e. $\phi(\theta(P_{\lambda/\sqrt{n}}))>0$), allows us to characterize the local power of the test. It is worth noting that (35) reduces to a familiar expression whenever ϕ is (fully) differentiable at θ_0. In such a setting, ϕ'_{θ_0} is linear, which implies $\mathbb{Z}_0\equiv\phi'_{\theta_0}(\mathbb{G}_0)$ is Gaussian, and therefore

$$P(\phi'_{\theta_0}(\mathbb{G}_0+\theta'(\lambda))>c_{1-\alpha})=P(\mathbb{Z}_0+\phi'_{\theta_0}(\theta'(\lambda))>c_{1-\alpha});$$

that is, the local power is determined by the probability that the sum of a Gaussian variable $\mathbb{Z}_0$ and a Pitman drift $\phi'_{\theta_0}(\theta'(\lambda))$ exceeds the $1-\alpha$ quantile of $\mathbb{Z}_0$. ‖

3.4.3. Examples revisited.

We return to Examples 2.1 and 2.2 to illustrate the preceding local analysis.

Example 2.1 (cont.) Notice in this application $\theta(P)=E_P[X]$, where E_P denotes the expectation when X is distributed according to P. To examine the impact on the distribution of $\max\{\bar{X},0\}$ of $\theta(P)$ being "close" to zero, we consider a path such that $\theta(P_{\lambda/\sqrt{n}})=E_{P_{\lambda/\sqrt{n}}}[X]=\lambda/\sqrt{n}$. Assumption 5 then holds with $\theta'(\lambda)=\lambda$, and hence

$$\sqrt{n}\{\phi(\hat{\theta}_n)-\phi(\theta(P_{\lambda/\sqrt{n}}))\}\overset{L_\lambda}{\to}\max\{\mathbb{G}_0+\lambda,0\}-\max\{\lambda,0\} \tag{36}$$

by Lemma 3.1. In particular, result (36) reflects that the finite sample distribution of $\max\{\bar{X},0\}$ depends on the distance of $E_P[X]$ to zero. Nonetheless, the directional derivative $\phi'_{\theta_0}(h)=\max\{h,0\}$ is convex, and hence Corollary 3.2 implies we may employ the bootstrap procedure of Theorem 3.2 to obtain a test of (31) that locally controls size.[11] This observation has previously been extensively employed by the literature on models defined by moment inequalities (Andrews and Soares, 2010). ‖

Example 2.2 (cont.) Recall in this application $\phi(\theta(P))$ is the largest value in the identified set of a coordinate of β_0, which Bontemps *et al.* (2012) show equals

$$\phi(\theta(P)) = E_P[b_0(P)'ZY_l + \max\{b_0(P)'Z, 0\}(Y_u - Y_l)]$$

with $b_0(P) \equiv (E_P[ZZ'])^{-1}p$. It is straightforward to verify that ϕ'_{θ_0}, as characterized in (15), is convex and therefore Corollary 3.2 provides us with a test of (31) that delivers local size control. In this application, however, it is also interesting to test

$$H_0 : \phi(\theta(P)) \geq c_0 \qquad\qquad H_1 : \phi(\theta(P)) < c_0, \tag{37}$$

which, through test inversion over c_0, enables us to obtain an upper one-sided confidence region for the largest value in the identified set of a coordinate of β_0. When $P(b_0(P)'Z = 0)$ is positive, and hence ϕ fails to be (fully) Hadamard differentiable, it is possible to show that a test that rejects (37) whenever $\sqrt{n}\phi(\hat{\theta}_n)$ is smaller than $\hat{c}_\alpha$ can fail to provide local size control. Nonetheless, Theorem 3.3 can be used to show that any $c^\star$ satisfying

$$P(\mathbb{G}_\psi(b_0) + \mathbb{G}_b' E[Z(Y_l + (Y_u - Y_l)(1\{b_0(P)'Z > 0\} + 1\{Z \in A\}))] < c^\star) \leq \alpha$$

for all sets $A \subset \{z : z'b_0(P) = 0\}$ can be used as a critical value to obtain a test that provides local size control; see the Supplementary Appendix for additional discussion. ‖

4. CONVEX SET PROJECTIONS

In this section, we demonstrate the usefulness of our results by constructing a test of whether a Hilbert space valued parameter belongs to a known closed convex set—a setting that encompasses tests of moment inequalities, shape restrictions, and the validity of random utility models. Despite the generality of the problem, we show our results enable us to develop a valid test by relying only on analytical calculations.

4.1. *Projection setup*

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

$$H_0 : \theta_0 \in \Lambda \qquad\qquad H_1 : \theta_0 \notin \Lambda, \tag{38}$$

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

11. The cdf of $\max\{\mathbb{G}_0, 0\}$ is continuous and strictly increasing at $c_{1-\alpha}$ if and only if $\alpha < 0.5$ and $\mathrm{Var}\{\mathbb{G}_0\} > 0$.

To map this problem into the framework of Section 3.4.2, we define the projection map $\Pi_\Lambda : \mathbb{H} \to \Lambda$ which for each $\theta \in \mathbb{H}$ returns the unique element $\Pi_\Lambda \theta \in \Lambda$ satisfying

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

We then let $\phi : \mathbb{H} \to \mathbf{R}$ be the function mapping each $\theta \in \mathbb{H}$ into its distance to Λ; *i.e.*

$$\phi(\theta) \equiv \|\theta - \Pi_\Lambda \theta\|_{\mathbb{H}}. \tag{39}$$

Employing this notation, it follows that the null hypothesis in (38) is satisfied if and only if $\phi(\theta_0) \leq 0$. Thus, we may view the hypothesis testing problem defined in (38) as a special case of the general testing problems examined in Section 3.4.2 (see (31)).

As a final piece of notation, we need to introduce the tangent cone of Λ at a $\theta \in \mathbb{H}$, which plays a crucial 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

$$T_\theta \equiv \overline{\bigcup_{\alpha \geq 0} \alpha\{\Lambda - \Pi_\Lambda \theta\}}. \tag{40}$$

Heuristically, T_θ constitutes a local approximation to the set Λ at the point $\Pi_\Lambda \theta$. To appreciate this connection, note that $\{\Lambda - \Pi_\Lambda \theta\}$ consists of all vectors h such that a movement from $\Pi_\Lambda \theta$ by h remains in Λ (*i.e.* $\Pi_\Lambda \theta + h \in \Lambda$). Up to closure, T_θ then consists of all vectors that point in the same direction as some $h \in \{\Lambda - \Pi_\Lambda \theta\}$ (*i.e.* αh for some $\alpha \geq 0$). Loosely speaking, the boundary of T_θ is as a result determined solely by the "local" region of the boundary of Λ that $\Pi_\Lambda \theta$ is in; see Figure 1.

4.1.1. Examples. In order to aid exposition, we next provide examples of both well studied and new applications that are a special case of the introduced hypothesis testing problem.

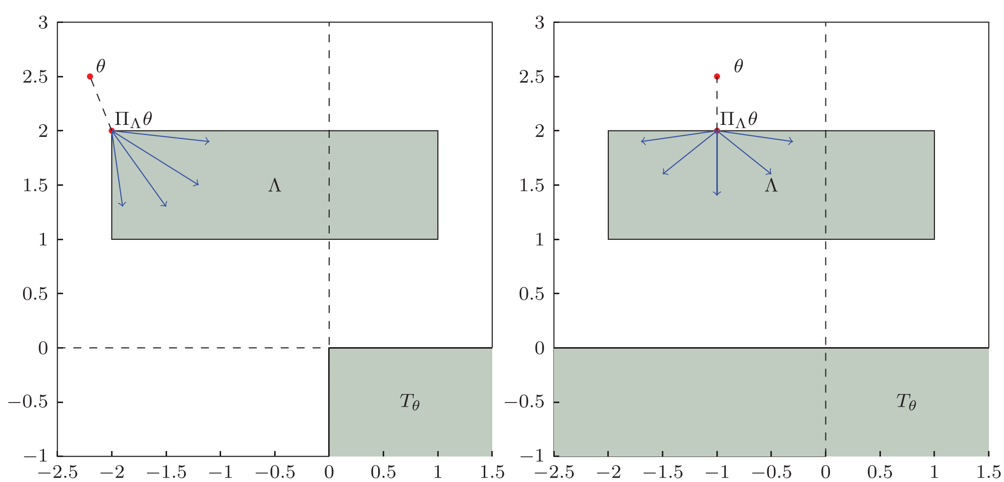

FIGURE 1

Illustrations of tangent cones. Arrows represent elements of $\Lambda - \Pi_\Lambda \theta$, while T_θ is (up to closure) the set of vectors αh for some $\alpha \geq 0$ and $h \in \Lambda - \Pi_\Lambda \theta$. T_θ equals $\{(x,y) : x \geq 0, y \leq 0\}$ and $\{(x,y) : y \leq 0\}$ in the first and second figures respectively.

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

$$H_0 : E[X] \leq 0 \qquad\qquad H_1 : E[X] \nleq 0,$$

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*

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

where $(a)_+ = \max\{a, 0\}$ *and* $X^{(i)}$ *denotes the* i^{th} *coordinate of* X*. Related special cases of the hypothesis in* (38) *include tests of inequalities on regression coefficients (Wolak, 1988) and tests of random utility models (Kitamura and Stoye, 2013).* ‖

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

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

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

where $\rho_\tau(u) = (\tau - 1\{u \leq 0\})u$ *and* $\tau \in (0,1)$*. Under appropriate restrictions, the sample analogue estimator of* θ_0 *converges in distribution in* $\ell^\infty([\epsilon, 1-\epsilon])$ *for any* $\epsilon \in (0, 1/2)$ *(Angrist* et al.*, 2006). In this setting, it is often of interest to test for shape restrictions on* θ_0*, which we may accomplish by setting* $\mathbb{H}$ *to equal the Hilbert space*

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

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.* ‖

As our final example we introduce a generalization of Example 4.2 that we will apply to study qualitative features of the trend in residual U.S. wage inequality; see Section 5.

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

$$E[\rho(Z, \theta_0(\tau), \tau)] = 0.$$

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; Chandrasekarh* et al.*, 2013).*

Escanciano and Zhu (2013) derive an estimator of θ_0 which, for an integrable function $w:\mathcal{T}\to \mathbf{R}_+$, converges in distribution in the space

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

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

4.2. *Theoretical results*

4.2.1. The test statistic. Following the analysis of Section 3.4.2, we test the null hypothesis that the parameter θ_0 belongs to the set Λ by rejecting for large values of the test statistic

$$r_n\phi(\hat{\theta}_n);$$

that is, by rejecting whenever the distance of the estimator $\hat{\theta}_n$ to Λ is "too large".

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

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

actually holds with equality. 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.

Since $\phi(\theta_0)=0$ whenever $\theta_0\in\Lambda$, the asymptotic distribution of $r_n\phi(\hat{\theta}_n)$ under the null hypothesis is an immediate consequence of Theorem 2.1 and Zarantonello (1971).

Proposition 4.1 *Let $\mathbb{D}=\mathbb{H}$ for $\mathbb{H}$ a Hilbert space with norm $\|\cdot\|_{\mathbb{H}}$, and $\Lambda\subseteq\mathbb{H}$ be a closed convex set. If $\theta_0\in\Lambda$, then $\phi:\mathbb{H}\to\mathbf{R}$ as in (39) satisfies Assumption 1 with*

$$\phi'_{\theta_0}(h)=\|h-\Pi_{T_{\theta_0}}h\|_{\mathbb{H}} \tag{41}$$

and $\mathbb{D}_0=\mathbb{H}$. Therefore, for any estimator $\hat{\theta}_n$ satisfying Assumption 2 it follows that

$$r_n\phi(\hat{\theta}_n)=r_n\|\hat{\theta}_n-\Pi_\Lambda\hat{\theta}_n\|_{\mathbb{H}}\xrightarrow{L}\|\mathbb{G}_0-\Pi_{T_{\theta_0}}\mathbb{G}_0\|_{\mathbb{H}}=\phi'_{\theta_0}(\mathbb{G}_0). \tag{42}$$

The asymptotic distribution obtained in Proposition 4.1 can depend discontinuously in θ_0 when θ_0 is on the boundary of Λ. This discontinuity is cause for concern that the asymptotic distribution in Proposition 4.1 may be a poor approximation to the finite sample distribution of $r_n\phi(\hat{\theta}_n)$ whenever θ_0 is "near" the boundary of Λ. For instance, in the moment inequalities application of Example 4.1, θ_0 is "near" the boundary of Λ precisely when the moments inequalities are "close" to binding. Fortunately, for the purposes of testing the null hypothesis in (38), Corollary 3.2 reassures us that said discontinuity in the asymptotic distribution may be immaterial. In particular, Corollary 3.2 implies that employing the quantiles of the asymptotic distribution in Proposition 4.1 as critical values will result in a test with local size control provided ϕ'_{θ_0} is convex.

Our next result verifies this important convexity requirement on ϕ'_{θ_0}.

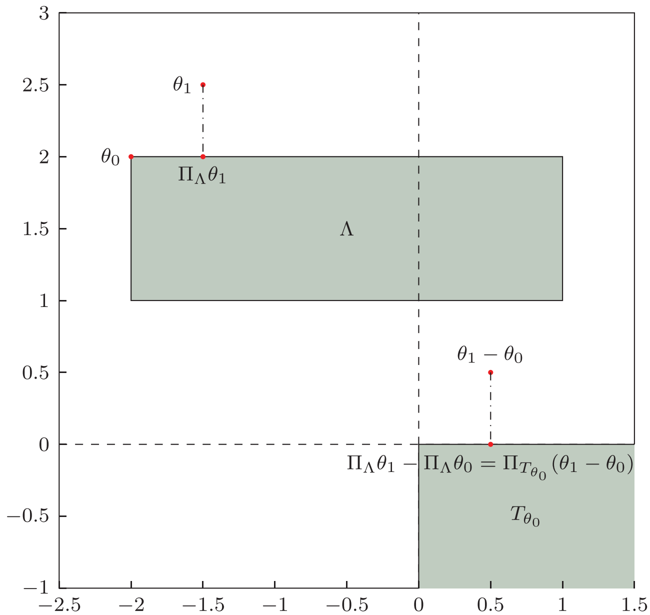

FIGURE 2

Illustration of directional differentiability. Here, $\Pi_\Lambda\theta_1-\Pi_\Lambda\theta_0$ is approximated without error by the derivative $\Pi_{T_{\theta_0}}(\theta_1-\theta_0)$. Note $\Pi_\Lambda\theta_0=\theta_0$ since $\theta_0\in\Lambda$.

Proposition 4.2 *Let $\mathbb{D}=\mathbb{H}$ for $\mathbb{H}$ a Hilbert space with norm $\|\cdot\|_{\mathbb{H}}$, and $\Lambda\subseteq\mathbb{H}$ be a closed convex set. Then it follows that $\phi'_{\theta_0}:\mathbb{H}\to\mathbf{R}$ (as in (41)) is convex.*

Together, Propositions 4.1 and 4.2 and Corollary 3.2 imply employing the quantiles of $\phi'_{\theta_0}(\mathbb{G}_0)$ as critical values will deliver size control even if θ_0 is "near" the boundary of Λ; see also Remark 4.2 below. Alternatively, 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 asymptotic distribution of Proposition 4.1 is first order stochastically dominated by the distribution of $\|\mathbb{G}_0-\Pi_\Lambda\mathbb{G}_0\|_{\mathbb{H}}$ and thus the quantiles of the latter may be employed for potentially conservative inference—an approach that may be viewed as analogous to assuming all moments are binding in the moment inequalities literature.

4.2.2. The critical value.

In order to construct critical values for the test statistic $r_n\phi(\hat{\theta}_n)$ we next aim to employ Theorem 3.2 to devise a consistent estimator of the quantiles of $\phi'_{\theta_0}(\mathbb{G}_0)$. To this end, we employ the analytical characterization of ϕ'_{θ_0} obtained in Proposition 4.1 to obtain an estimator $\hat{\phi}'_n$ that satisfies Assumption 4 under no additional requirements.

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

$$\hat{\phi}'_n(h)\equiv\sup_{\theta\in\Lambda:r_n\|\theta-\Pi_\Lambda\hat{\theta}_n\|_{\mathbb{H}}\leq\kappa_n}\|h-\Pi_{T_\theta}h\|_{\mathbb{H}} \tag{43}$$

(compare to (41)). Heuristically, we estimate $\phi'_{\theta_0}(h)$ 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 κ_n/r_n neighbourhood of $\Pi_\Lambda\hat{\theta}_n$. It is evident from this construction that provided $\kappa_n\uparrow\infty$ at an appropriate rate, the shrinking neighbourhood 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 an upper bound for $\phi'_{\theta_0}(h)$. As the following Proposition shows, however, $\hat{\phi}'_n$ is not only an upper bound but is in fact also consistent for ϕ'_{θ_0} in the sense required by Theorem 3.2.

Proposition 4.3 *If the conditions of Proposition 4.1 are satisfied, and $\kappa_n \uparrow \infty$ satisfies $\kappa_n/\sqrt{n} \downarrow 0$, then it follows that $\hat{\phi}'_n$ as defined in (43) satisfies Assumption 4.*

Theorem 3.2 therefore implies that, provided the bootstrap is consistent for the asymptotic distribution of $r_n\{\hat{\theta}_n - \theta_0\}$, we may employ the $1-\alpha$ quantile of

$$\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) = \sup_{\theta \in \Lambda : r_n \|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n} \|r_n\{\hat{\theta}^*_n - \hat{\theta}_n\} - \Pi_{T_\theta}\{r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}\}\|_{\mathbb{H}}$$

(conditional on $\{X_i\}_{i=1}^n$) as a critical value for the test statistic $r_n\phi(\hat{\theta}_n)$. While the proposed estimator $\hat{\phi}'_n$ is appealing due to its general applicability, we note that in specific applications computationally simpler approaches may be available; see Remark 4.2. Alternatively, critical values may also be obtained by employing numerical estimators of ϕ'_{θ_0} as in Hong and Li (2017). In accord with our discussion in Section 3.3, however, we find that the choice of tuning parameter is simpler to motivate when employing estimators $\hat{\phi}'_n$ that leverage the analytical computation of ϕ'_{θ_0}.

Remark 4.1 *In selecting κ_n in (43), it is helpful to view $\{\theta \in \Lambda : r_n \|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le \kappa_n\}$ as a confidence region for θ_0 whose confidence level tends to one. Moreover, since*

$$r_n \|\theta_0 - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \le r_n \|\theta_0 - \hat{\theta}_n\|_{\mathbb{H}} \xrightarrow{L} \|\mathbb{G}_0\|_{\mathbb{H}},$$

due to $\theta_0 \in \Lambda$, it follows that setting κ_n to equal the $1-\alpha_n$ quantile of $\|\mathbb{G}_0\|_{\mathbb{H}}$ results in a confidence level of $1-\alpha_n$. This observation suggests selecting κ_n to equal the $1-\alpha_n$ conditional quantile of $\|r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}\|_{\mathbb{H}}$ given $\{X_i\}_{i=1}^n$ for some sequence $\alpha_n \downarrow 0$. ‖

Remark 4.2 *In certain applications, the tangent cone T_{θ_0} can be easily estimated and as a result so can ϕ'_{θ_0}. For instance, suppose Λ is a polyhedron so that it satisfies*

$$\Lambda = \{h \in \mathbb{H} : \langle h, a_j \rangle_{\mathbb{H}} \le b_j \text{ for all } 1 \le j \le J\},$$

where $a_j \in \mathbb{H}$, $b_j \in \mathbf{R}$, and $J < \infty$. For any $\theta_0 \in \Lambda$, the tangent cone T_{θ_0} then equals

$$T_{\theta_0} = \{h \in \mathbb{H} : \langle h, a_j \rangle_{\mathbb{H}} \le 0 \text{ for all } j \text{ s.t. } \langle \theta_0, a_j \rangle_{\mathbb{H}} = b_j\};$$

that is T_{θ_0} is determined by the inequality constraints that "bind" at θ_0. This characterization suggests the following simple estimator for the tangent cone:

$$\hat{T}_n \equiv \left\{ h \in \mathbb{H} : \langle h, a_j \rangle_{\mathbb{H}} \le 0 \text{ for all } j \text{ s.t. } \frac{r_n\{\langle a_j, \hat{\theta}_n \rangle_{\mathbb{H}} - b_j\}}{\hat{\sigma}_j} \ge -\kappa_n \right\} \tag{44}$$

for $\kappa_n \uparrow \infty$ satisfying $\kappa_n / r_n \downarrow 0$ and $\hat{\sigma}_j^2$ any estimator of the asymptotic variance of $r_n \langle a_j, \hat{\theta}_n - \theta_0 \rangle_{\mathbb{H}}$. It is then straightforward to verify that $\hat{\phi}'_n(h) = \|h - \Pi_{\hat{T}_n} h\|_{\mathbb{H}}$ satisfies Assumption 4 (compare to

(41)). *Moreover, following Remark 4.1, we may set*

$$\kappa_n = \inf\left\{c : P\left(\max_{1\le j\le J} \frac{r_n\langle a_j, \hat{\theta}_n - \hat{\theta}_n^*\rangle_{\mathbb{H}}}{\hat{\sigma}_j} \le c | \{X_i\}_{i=1}^n\right) \ge 1-\alpha_n\right\} \tag{45}$$

for some sequence $\alpha_n \downarrow 0$ – such a choice allows us to interpret $\hat{T}_n$ as a confidence region for the set of binding constraints at θ_0 with confidence level $1-\alpha_n$. ‖

4.3. *Simulation evidence*

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

$$Y = \frac{\Delta}{\sqrt{n}} D\times U + Z'\beta + U, \tag{46}$$

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) satisfies

$$P(Y \le D\theta_0(\tau)+Z'\beta(\tau)|D,Z)=\tau, \tag{47}$$

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 test whether the quantile treatment effect $\theta_0(\tau)$ is monotonically increasing in τ, which corresponds to the special case of (38) 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. Critical values 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 $\hat{\phi}_n'(\sqrt{n}\{\hat{\theta}_n^*-\hat{\theta}_n\})$. Since in this problem Λ is a polyhedron, we set $\hat{\phi}_n'(h)=\|h-\Pi_{\hat{T}_n}h\|_{\mathbb{H}}$ for $\hat{T}_n$ as in (44) and $\hat{\sigma}_j^2$ the bootstrap estimate of the asymptotic variance of $r_n\langle a_j,\theta_0-\hat{\theta}_n\rangle_{\mathbb{H}}$. In order to explore the sensitivity of our results to the choice of κ_n we additionally select κ_n according to (45) for different values of α_n. All reported results are based on five thousand Monte Carlo replications. Computational costs are modest, with a single replication of sample size five hundred being completed in 12.47 seconds when running MATLAB on a single Intel i7-7700K 4.2 Ghz core.

Table 1 reports the empirical rejection rates for different values of the local parameter $\Delta\in\{0,1,2\}$ – recall that since $\theta_0(\tau)=\tau\Delta/\sqrt{n}$, the null hypothesis that θ_0 is monotonically increasing is satisfied for all such Δ. For the explored sample sizes of two and five hundred observations, we observe little sensitivity to the value of the bandwidth α_n defining the estimator $\hat{\phi}_n'$ (as in (45)). In addition, the row labelled "Theoretical" reports the rejection rates we should expect according to the local asymptotic approximation of Theorem 3.3. Throughout the specifications, we see that the test effectively controls size even for the "aggressive" choice of $\alpha_n=0.9$ (theory requires $\alpha_n\downarrow 0$). In addition, the theoretical predictions of Theorem 3.3 provide an adequate approximation to the empirical rejection probabilities across the different values of the local parameter.

TABLE 1
Empirical rejection probabilities for the null hypothesis that θ_0 is monotonically increasing

Bandwidth	$n=200$								
	Level $=0.1$			Level $=0.05$			Level $=0.01$		
α_n	$\Delta=0$	$\Delta=1$	$\Delta=2$	$\Delta=0$	$\Delta=1$	$\Delta=2$	$\Delta=0$	$\Delta=1$	$\Delta=2$
0.9	0.071	0.033	0.015	0.039	0.018	0.006	0.010	0.003	0.001
0.5	0.046	0.019	0.006	0.025	0.010	0.003	0.005	0.001	0.001
0.1	0.042	0.016	0.005	0.023	0.009	0.002	0.005	0.001	0.000
Theoretical	0.100	0.042	0.015	0.050	0.018	0.006	0.010	0.003	0.001
Bandwidth	$n=500$								
	Level $=0.1$			Level $=0.05$			Level $=0.01$		
α_n	$\Delta=0$	$\Delta=1$	$\Delta=2$	$\Delta=0$	$\Delta=1$	$\Delta=2$	$\Delta=0$	$\Delta=1$	$\Delta=2$
0.9	0.076	0.036	0.017	0.042	0.019	0.009	0.011	0.005	0.003
0.5	0.053	0.021	0.009	0.028	0.010	0.004	0.006	0.002	0.000
0.1	0.051	0.019	0.008	0.026	0.009	0.003	0.006	0.002	0.000
Theoretical	0.100	0.042	0.015	0.050	0.018	0.006	0.010	0.003	0.001

Notes: Δ represents the local parameter (as in (46)) and α_n the bandwidth employed in computing $\hat{\phi}'_n(h)=\|h-\Pi_{\hat{T}_n}h\|_{\mathbb{H}}$ (as in (44) and (45)). The row labelled "Theoretical" displays the rejection probabilities implied by Theorem 3.3

TABLE 2
Empirical local power for a level 5% test of the null hypothesis that θ_0 is weakly increasing

Bandwidth	$n=200$							
α_n	$\Delta=-1$	$\Delta=-2$	$\Delta=-3$	$\Delta=-4$	$\Delta=-5$	$\Delta=-6$	$\Delta=-7$	$\Delta=-8$
0.9	0.089	0.190	0.362	0.590	0.801	0.938	0.991	1.000
0.5	0.062	0.156	0.326	0.559	0.783	0.933	0.991	1.000
0.1	0.060	0.152	0.322	0.557	0.780	0.933	0.991	1.000
Theoretical	0.120	0.245	0.423	0.623	0.796	0.911	0.970	0.992
Bandwidth	$n=500$							
α_n	$\Delta=-1$	$\Delta=-2$	$\Delta=-3$	$\Delta=-4$	$\Delta=-5$	$\Delta=-6$	$\Delta=-7$	$\Delta=-8$
0.9	0.098	0.209	0.387	0.610	0.807	0.931	0.983	0.997
0.5	0.075	0.180	0.354	0.582	0.792	0.926	0.982	0.997
0.1	0.072	0.178	0.352	0.580	0.790	0.925	0.982	0.997
Theoretical	0.120	0.245	0.423	0.623	0.796	0.911	0.970	0.992

Notes: Δ represents the local parameter (as in (46)) and α_n the bandwidth employed in computing $\hat{\phi}'_n(h)=\|h-\Pi_{\hat{T}_n}h\|_{\mathbb{H}}$ (as in (44) and (45)). The row labelled "Theoretical" displays the rejection probabilities implied by Theorem 3.3.

In Table 2, we examine the local power of a 5% level test by considering values of $\Delta\in\{-1,\ldots,-8\}$. For these 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 find the theoretical local power to align more closely with the empirical rejection rates for larger values of the local parameter Δ. Overall, we find the simulation results encouraging, though certainly limited in their scope.

5. RESIDUAL WAGE INEQUALITY

A large literature examines trends in residual wage inequality using the Current Population Survey (CPS); see, among others, Katz and Murphy (1992), Juhn *et al.* (1993), and Autor *et al.* (2008). Existing work has largely focused on the qualitative features of the estimates of these trends—*e.g.*, by examining whether these estimates are decreasing or increasing, or determining the year at which residual wage inequality is the largest. These qualitative features are often directionally differentiable functions of the trends, and as a result statistical inference suffers from the challenges addressed in Sections 3 and 4. We thus next complement the existing literature by going beyond point estimates and employing our results to conduct formal statistical inference.

5.1. *Data and background*

We follow the analysis of Lemieux (2006) and Autor *et al.* (2008) in examining inequality measures based on the residuals from a Mincer (1974) regression. Specifically, we employ the Merged Outgoing Rotation Groups (ORG) supplements to the 1979 to 2016 CPS to estimate, separately by year and gender, the regression

$$Y_i = X_i'\beta_0 + \varepsilon_i \qquad\qquad E[X_i\varepsilon_i] = 0, \tag{48}$$

where Y_i denotes log hourly wages and X_i includes a constant, dummy variables for twenty three age categories and eight education categories, and the interaction of the Z-score for age and its square with the education dummy variables.[12] Dispersion measures for the distribution of ε are then understood as capturing wage dispersion among workers with similar age and education profiles. Following the literature, we refer to these dispersion measures as *residual wage inequality* and study their evolution through time. As noted by Juhn *et al.* (1993), the growth in residual wage inequality accounts for the majority of the growth in overall wage inequality.

A challenge in examining trends in residual wage inequality is that compositional changes in the labour force can mechanically generate increased dispersion in the distribution of ε. For instance, the variance of wages among uneducated older workers is considerably larger than among uneducated younger workers. Thus, an increase in the share of uneducated older workers can mechanically generate an increment in residual wage variance. To address this concern, we follow Lemieux (2006) in controlling for these compositional changes by employing the reweighting methodology developed by DiNardo *et al.* (1996) to keep labour force composition constant at its 1979 level.[13]

For our sample, we restrict attention to currently employed workers who are between sixteen and sixty-four years old. Wages are price deflated, and observations with hourly earnings below one or above one hundred in 1979 dollars are dropped. All observations with imputed wage observations are dropped from the sample. Between January 1994 and August 1995, flags for imputed wages are missing, and thus we follow Lemieux (2006) and Autor *et al.* (2008) in dropping these observations as well; see Hirsch and Schumacher (2004) for additional discussion. All calculations employ weights equal to the product of the CPS sampling weights and hours worked in the previous week.

12. Employing the Z-score of age makes the design matrix easier to invert due to higher powers of Z-score age being better scaled than higher powers of age. Employing Z-scores of age instead of age itself does not otherwise impact our analysis, since all statistics depend on the residuals in (48) and replacing age with its Z-score does not change the column span of the regressors.

13. The reweighting function is estimated by a linear probability model that includes the same covariates employed in (48). Predicted probabilities are truncated to lie between 0.01 and 0.99. For men, this truncation binds in 0% of the observations in the full sample and 0.000002% of the observations in the bootstrap samples. For women, the corresponding numbers are 0.0000014% and 0.000002%.

5.2. *Empirical results*

Figure 3 depicts the point estimates for residual wage variance (*i.e.* the variance of ε in (48)) between 1979 and 2016 for both men and women. In agreement with Lemieux (2006) we find that adjusting for changes in labour force composition is important. In particular, the composition adjusted estimates appear to be stable since 1990, while the composition unadjusted estimates have steadily increased during the same time period.

We next go beyond the point estimates and examine a series of hypothesis tests that fit the framework of Sections 3 and 4. To this end, we think of the functions in Figure 3 as the point estimates $\hat{\theta}_n$ of their population counterparts θ_0—*e.g.* $\theta_0(t)$ is the population value of the composition adjusted residual wage variance at time t and θ_0 is the corresponding set of population values between 1979 and 2016. The randomness in the estimator $\hat{\theta}_n$ is then the result of the random sampling conducted by the CPS. On average, each year approximately 28,000 households are introduced into our sample. Due to the sampling structure, however, each household appears in

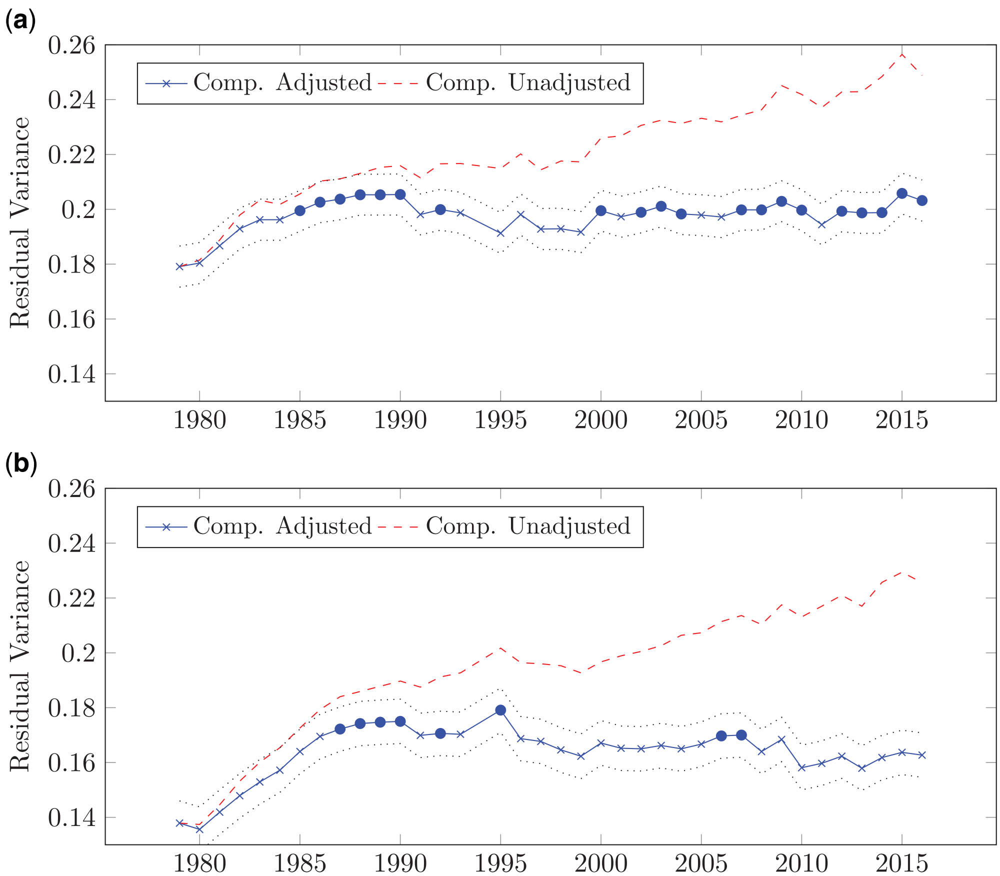

FIGURE 3

Composition adjusted series keep labour force composition constant at its 1979 level employing the reweighing scheme of DiNardo *et al*. (1996), while the composition unadjusted series omits such reweighing. Dots indicate year belongs to a 95% confidence region for the point in time at which composition residual variance was the largest (obtained through test inversion of (49)). Dotted lines form 95% uniform confidence bands for composition adjusted residual variance. (a) Residual variance for men; (b) Residual variance for women.

two consecutive years. For inference, we employ a block bootstrap that clusters at the household level to compute the bootstrapped version $\hat{\theta}_n^*$ of $\hat{\theta}_n$. Many of the hypotheses tests we examine fit the framework of Remark 4.2. For such applications, we conduct inference by employing the results in Section 3 with $\hat{\phi}_n'$ as specified in Remark 4.2 employing $\alpha_n = 0.01$ (see (45)). We view such a choice for α_n as "conservative" in that the corresponding estimator $\hat{\phi}_n'$ can then be interpreted as based on a 99%-confidence region for the set of binding constraints; see Remark 4.2. As in the simulations of Section 4.3, our conclusions are not overly sensitive to the choice of α_n.

Through test inversion, we first build a 95% confidence region for the years at which the composition adjusted residual wage variances of men and women attained their largest values. Specifically, for each $t_0 \in \{1979, \ldots, 2016\}$ we test the null hypothesis

$$H_0 : \theta_0(t_0) - \theta_0(t) \geq 0 \text{ for all } t \neq t_0 \qquad H_1 : \theta_0(t_0) - \theta_0(t) < 0 \text{ for some } t \neq t_0. \tag{49}$$

Of course, even if t_0 is indeed the point at which residual wage inequality was the largest, the point estimate $\hat{\theta}_n$ may not exhibit its largest value at t_0 due to sampling uncertainty. We therefore formally test the null hypothesis in (49) by employing the test statistic

$$\inf_{\theta} \sqrt{n} \|\hat{\theta}_n - \theta\| \text{ s.t. } \theta(t_0) \geq \theta(t) \text{ for all } t \neq t_0;$$

that is, we measure the (Euclidean) distance between the estimate $\hat{\theta}_n$ and the set of functions satisfying the null hypothesis. Our analysis in Section 3.4 implies "standard" bootstrap methods fail when the value of $\theta_0(t)$ is "close" to $\theta(t_0)$ at some $t \neq t_0$ (here, "close" is meant relative to sampling uncertainty). We therefore obtain critical values by applying the bootstrap methodology of Section 4 instead. Specifically, since the null hypothesis in (49) can be expressed as θ_0 belonging to a set defined by linear constraints, we follow Remark 4.2 by employing as critical values the quantiles (conditional on the data) of

$$\inf_{\theta} \|\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\} - \theta\| \text{ s.t. } \theta(t) \leq \theta(t_0) \text{ for all } t \text{ satisfying } \frac{\sqrt{n}\{\hat{\theta}_n(t) - \hat{\theta}_n(t_0)\}}{\hat{\sigma}_t} \geq -\kappa_n,$$

where $\hat{\theta}_n^*$ is the bootstrapped version of $\hat{\theta}_n$ and $\hat{\sigma}_t^2$ is a bootstrap estimate of the asymptotic variance of $\sqrt{n}\{(\hat{\theta}_n(t_0) - \hat{\theta}_n(t)) - (\theta_0(t_0) - \theta_0(t))\}$. As previously stated, the bandwidth κ_n was chosen by setting α_n to equal 0.01 in (45), which in this application translates into setting κ_n to equal the 0.99 quantile (conditional on the data) of

$$\max_{t \neq t_0} \frac{\sqrt{n}\{(\hat{\theta}_n^*(t) - \hat{\theta}_n^*(t_0)) - (\hat{\theta}_n(t) - \hat{\theta}_n(t_0))\}}{\hat{\sigma}_t}.$$

In Figure 3 we mark with a dot the members of the described confidence region for the year at which the composition adjusted residual wage variance attained its largest value. Both confidence regions support the analysis of Card and Dinardo (2002), who argue that a significant and permanent increment in the residual wage variance took place in the 1980s. To examine this claim, we further test the null hypothesis that the residual wage variance was larger than its 1979 level at all years since 1990. We find the p-values for this hypothesis to equal 0.95 and 0.99 for men and women, respectively. However, residual wage variance has not been constant since 1990.

The p-values for the null hypothesis that it has been constant since 1990 equal zero for both men and women.

As emphasized by Autor *et al.* (2008), the relative stability in the estimates of residual wage variance after the 1980s hides important changes in other aspects of the residual wage distribution. Figure 4 depicts the difference between the 90^{th} and 50^{th} quantiles (known as the 90-50 Gap) and the difference between the 50^{th} and 10^{th} quantiles (known as the 50-10 Gap) of the residual wage distribution for men and women. Given the importance of the changes in the labour force composition, all point estimates in Figure 4 keep labour force composition constant at its 1979 level. The point estimates reaffirm the analysis in Autor *et al.* (2008), who find that the relatively stable behaviour of the residual wage variance estimates after the 1980s masks important differences in the estimates of dispersion within the upper tail (90-50 Gap) and the lower tail (50-10 Gap) of the residual wage distribution.

In order to quantify the statistical uncertainty in the qualitative features of our point estimates, we next conduct a series of hypothesis tests. Starting with the 50-10 Gap, we build a 95% confidence region for the year at which the 50-10 Gap was the largest. Members of the confidence regions for men and for women are marked by a dot in Figure 4. We further examine whether there is a clear downward trend by testing whether the 50-10 Gap has been monotonically decreasing since attaining its largest value. To this end, we employ the results of Section 4 which enable us to test whether a trend has been monotonically decreasing since time period t_0 by using the test statistic

$$\inf_{\theta} \sqrt{n}\|\hat{\theta}_n-\theta\| \text{ s.t. } \theta(t+1)\leq\theta(t) \text{ for all } t\geq t_0; \tag{50}$$

that is, we measure the distance between $\hat{\theta}_n$ and the set of functions that are monotonically decreasing since t_0 (see Figure 4 for the plot of the function minimizing (50)). The results of Section 3 imply "standard" bootstrap methods fail when $\theta_0(t)$ and $\theta_0(t+1)$ are "close" for some $t\geq t_0$. Hence, we instead rely on the critical value of Section 4, which in this application corresponds to the quantiles (conditional on the data) of

$$\begin{aligned}&\inf_{\theta}\|\sqrt{n}\{\hat{\theta}^*_n-\hat{\theta}_n\}-\theta\|\\ &\qquad \text{s.t. } \theta(t+1)\leq\theta(t) \text{ for all } t\geq t_0 \text{ satisfying } \frac{\sqrt{n}\{\hat{\theta}_n(t+1)-\hat{\theta}_n(t)\}}{\hat{\sigma}_t}\geq-\kappa_n,\end{aligned}$$

where $\hat{\sigma}^2_t$ is a bootstrap estimate of the asymptotic variance of $\sqrt{n}\{(\hat{\theta}_n(t+1)-\hat{\theta}_n(t))-(\theta_0(t+1)-\theta_0(t))\}$. We again select κ_n by setting $\alpha_n=0.01$ in (45), which here corresponds to letting κ_n equal the 0.99 quantile (conditional on the data) of

$$\max_{t\geq t_0}\frac{\sqrt{n}\{(\hat{\theta}^*_n(t+1)-\hat{\theta}^*_n(t))-(\hat{\theta}_n(t+1)-\hat{\theta}_n(t))\}}{\hat{\sigma}_t}.$$

Employing the described test, we find evidence in favour of the 50-10 Gaps for both men and women exhibiting a downward trend since attaining their largest values. In particular, we find a p-value of 0.806 for the null hypothesis that the 50-10 Gap for men has been monotonically decreasing since 1989[14] —by way of comparison, we find a p-value of zero for the null hypothesis

14. This test employs the test statistic in (50) with $t_0=1989$. We chose $t_0=1989$ since 1989 is the largest year in the confidence region for the year at which the 50-10 Gap for men was the largest. Similarly, in evaluating the 50-10 Gap for women we set $t_0=1990$.

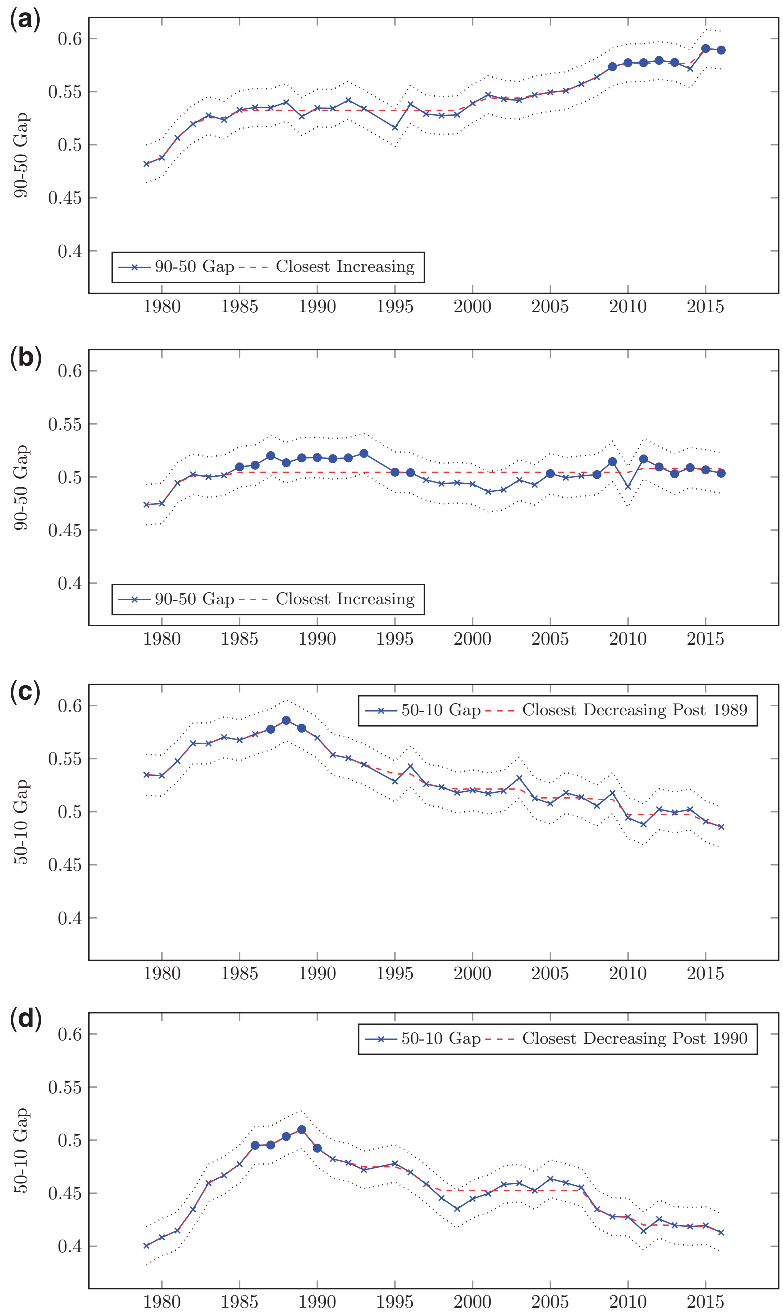

FIGURE 4

Differences between the 90th and 50th quantiles (90-50 Gap) and between the 50th and 10th quantiles (50-10 Gap) of residual wage inequality. Series are adjusted to keep the labour force composition at its 1979 level. Dots indicate year belongs to a 95% confidence interval for the point in time at which Gap series was the largest. Dotted lines represent 95% uniform confidence bands. (a) Men 90-50 Gap; (b) women 90-50 Gap; (c) Men 50-10 Gap; (d) women 50-10 Gap.

that the 50-10 Gap for men has been monotonically increasing during the same time period. The 50-10 Gap for women exhibits similar behaviour: The p-value for the null hypothesis that the 50-10 Gap for women has been monotonically decreasing since 1990 equals 0.19, while the p-value for the null hypothesis that it has been monotonically increasing is equal to zero.

The similarities in the 50-10 Gap series for men and women stand in stark contrast to the differences in the 90-50 Gap series. In particular, the point estimate for the 90-50 Gap for men exhibits an upward trend while the 90-50 Gap for women has remained relatively more stable. These differences were not as evident in previous studies employing ORG files up to only 2005, such as, for example, Lemieux (2006) and Autor *et al*. (2008). In order to quantify this contrast in the 90-50 Gap series, we test the null hypothesis that the 90-50 Gap for men has been monotonically increasing over the entire sample period and find a p-value of 0.772. In sharp contrast, the p-value for the null hypothesis that the 90-50 Gap for women has been monotonically increasing is equal to zero. It is worth noting that we reject the null hypothesis that the 90-50 Gap for women is increasing despite a monotonically increasing function fitting between the uniform confidence bands for the 90-50 Gap (Figure 4). This observation simply reflects that a test of monotonicity that is based on examining whether a monotonic function fits between the uniform confidence bands can have low power.

In summary, our analysis supports previous conclusions that the variance of residual wage inequality increased dramatically in the 1980s for both men and women. The residual wage dispersion in the lower tail (as measured by the 50-10 Gap) also behaves similarly for men and women: Both show a clear downward trends since attaining their largest values in the late 1980s. In contrast, the 90-50 Gaps for men and women exhibit significantly different qualitative features.

6. CONCLUSION

In this article, we have developed a general asymptotic framework for conducting inference in an important class of applications. 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}. Our results thus facilitate an understanding of potentially challenging statistical problems, such as bootstrap consistency or the local behavior of tests, through simple analytical calculations. We hope our results are therefore of use to theorists and empirical researchers alike in easily diagnosing and addressing "non-standard" inference problems.

Acknowledgments. We thank Aureo de Paula and three anonymous referees for suggestions that helped greatly improve the article. We are also indebted to Brendan Beare, Qihui Chen, Xiaohong Chen, Victor Chernozhukov, Bruce Hansen, Han Hong, Hiroaki Kaido, Patrick Kline, and numerous seminar participants for their valuable comments. Research supported by NSF Grant SES-1426882.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

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.

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. (2001), "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis", *Econometrica*, **69**, 683–734.

ANDREWS, D. W. K. and SHI, X. (2013), "Inference Based on Conditional Moment Inequatlies", *Econometrica*, **81**, 609–666.

ANDREWS, D. W. K. and SOARES, G. (2010), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection", *Econometrica*, **78**, 119–157.

ANGRIST, J., CHERNOZHUKOV, V. and FERNANDEZ-Val, I. (2006), "Quantile Regression under Misspecification, With an Application to the U.S. Wage Structure", *Econometrica*, **74**, 539–563.

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

AUTOR, D. H., KATZ, L. F. and KEARNEY, M. S. (2008), "Trends in US Wage Inequality: Revising the Revisionists", *The Review of Economics and Statistics*, **90**, 300–323.

BEARE, B. and MOON, J. (2015), "Nonparametric Tests of Density Ratio Ordering", *Econometric Theory*, **31**, 471–492.

BEARE, B. and SHI, X. (2018), "An Improved Bootstrap Test of Density Ratio Ordering", *Econometrics and Statistics*, forthcoming.

BERESTEANU, A. and MOLINARI, F. (2008), "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*, **76**, 763–814.

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

BICKEL, P. J. and FREEDMAN, D. A. (1981), "Some Asymptotic Theory for the Bootstrap", *The Annals of Statistics*, **9**, 1196–1217.

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

BUCHINSKY, M. (1994), "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression", *Econometrica*, **62**, 405–458.

BUGNI, F. A. (2010), "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set", *Econometrica*, **78**, 735–753.

CANAY, I. A. (2010), "El Inference for Partially Identified Models Defined by Moment Inqeualities: Coverage of the Identified Set", *The Journal of Econometrics*, **156**, 408–425.

CARD, D. and DINARDO, J. E. (2002), "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles", *Journal of Labor Economics*, **20**, 733–783.

CHANDRASEKARH, A., CHERNOZHUKOV, V., MOLINARI, F. *ET AL.* (2013) "Inference for Best Linear Approximations to Set Identified Functions", (Working Paper, Massachusetts Institute of Technology).

CHEN, Q. and FANG, Z. (2015), "Inference on Functionals Under First Order Degeneracy" (Working Paper, University of California – San Diego).

CHEN, X., TAMER, E. and TORGOVITSKY, A. (2011), "Sensitivity Analysis in Partially Identified Semiparametric Models" (Cowles Foundation Discussion Paper No. 1836).

CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, **75**, 1243–1284.

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

CILIBERTO, F. and TAMER, E. (2009), "Market Structure and Multiple Equilibria in Airline Markets", *Econometrica*, **77**, 1791–1828.

DINARDO, J., FORTIN, N. M. and LEMIEUX, T. (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica*, **64**, 1001–1044.

DORFMAN, R. (1938), "A Note on the δ-method for Finding Variance Formulae", *The Biometric Bulletin*, **1**, 129–137.

DUFOUR, J.-M. and TAAMOUTI, M. (2005), "Projection-based Statistical Inference in Linear Structural Models with Possibly Weak Instruments", *Econometrica*, **73**, 1351–1365.

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

EFRON, B. (1979), "Bootstrap Methods: Another Look at the Jacknife", *Annals of Statistics*, **7**, 1–26.

ESCANCIANO, J. C. and ZHU, L. (2013), "Inference in Semiparametric Partially Identified Models" (Working paper, Indiana University).

FANG, Z. (2015), "Optimal Plug-in Estimators of Directionally Differentiable Functionals" (Working paper, University of California - San Diego).

GILL, R. D., WELLNER, J. A. and PRÆSTGAARD, J. (1989), "Non-and Semi-Parametric Maximum Likelihood Estimators and the Von Mises Method (Part 1)[with Discussion and Reply]", *Scandinavian Journal of Statistics*, **16**, 97–128.

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

HANSEN, B. E. (2017), "Regression Kink with an Unknown Threshold", *Journal of Business & Economic Statistics*, **35**, 228–240.

HANSEN, P. R. (2005), "A Test for Superior Predictive Ability", *Journal of Business and Economic Statistics*, **23**, 365–380.

HIRANO, K. and PORTER, J. R. (2012), "Impossibility Results for Nondifferentiable Functionals", *Econometrica*, **80**, 1769–1790.

HIRSCH, B. T. and SCHUMACHER, E. J. (2004), "Match Bias in Wage Gap Estimates Due to Earnings Imputation", *Journal of labor economics*, **22**, 689–722.

HONG, H. and LI, J. (2017), "The Numerical Directional Delta Method and Bootstrap" (Working paper, Stanford University).

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

IMBENS, G. W. and MANSKI, C. F. (2004), "Confidence Intervals for Partially Identified Parameters", *Econometrica*, **72**, 1845–1857.

JHA, A. and WOLAK, F. A. (2015), "Testing for Market Efficiency with Transaction Costs: An Application to Financial Trading in Wholesale Electricity Markets", (Working paper, Stanford university).

JUHN, C., MURPHY, K. M. and PIERCE, B. (1993), "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, **101**, 410–442.

KAIDO, H. (2016), "A Dual Approach to Inference for Partially Identified Econometric Models", *Journal of Econometrics*, **192**, 269–290.

KATZ, L. F. and MURPHY, K. M. (1992), "Changes in Relative Wages, 1963–1987: Supply and Demand Factors", *The Quarterly Journal of Economics*, **107**, 35–78.

KITAMURA, Y. and STOYE, J. (2013), "Nonparametric Analysis of Random Utility Models: Testing" (Cemmap Working Paper (CWP36/13). Center for Microdata Methods and Practice).

KLINE, P. and SANTOS, A. (2013), "Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the U.S. Wage Structure", *Quantitative Economics*, **4**, 231–267.

KOSOROK, M. (2008), *Introduction to Empirical Processes and Semiparametric Inference* (New York, USA: Springer).

LEE, Y.-Y. and BHATTACHARYA, D. (2015), "Welfare Analysis for Discrete Choice with Interval-Data on Income (Technical report, Working Paper, Oxford University).

LEMIEUX, T. (2006), "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review*, **96**, 461–498.

LINTON, O., MAASOUMI, E. and WHANG, Y.-J. (2005), "Consistent Testing for Stochastic Dominance under General Sampling Schemes", *The Review of Economic Studies*, **72**, pp. 735–765.

LINTON, O., SONG, E., K. and WHANG, Y.-J. (2010), "An Improved Bootstrap Test of Stochastic Dominance", *Journal of Econometrics*, **154**, 186–202.

MANSKI, C. F. (2003), *Partial Identification of Probability Distributions*. (New York: Springer).

MANSKI, C. F. and TAMER, E. (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome", *Econometrica*, **70**, 519–546.

MASTEN, M. A. and POIRIER, A. (2017), "Inference on Breakdown Frontiers", (Working paper, Duke University).

MINCER, J. A. (1974), "Schooling, Experience, and Earnings", *NBER Books*.

MURALIDHARAN, K. and SUNDARARAMAN, V. (2011), "Teacher Performance Pay: Experimental Evidence from India", *Journal of Political Economy*, **9**, 39–77.

PAKES, A., PORTER, J., HO, K. *ET AL*. (2006), "Moment Inequalities and Their Application" (Working Paper, Harvard University).

POLITIS, D. N., ROMANO, J. and WOLF, M. (1999), *Subsampling* (New York: Springer).

REEDS, J. A. (1976), *On the Definition of von Mises Functionals*. Thesis, Harvard University.

ROMANO, J. P. and SHAIKH, A. M. (2008), "Inference for Identifiable Parameters in Partially Identified Econometric Models", *Journal of Statistical Planning and Inference – Special Issue in Honor of Ted Anderson*, **138**, 2786–2807.

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

SEO, J. (2018), "Tests of Stochastic Monotonicity with Improved Size and Power Properties", *Journal of Econometrics*, forthcoming.

SHAO, J. (1994), "Bootstrap Sample Size in Nonregular Cases", *Proceedings of the American Mathematical Society*, **122**, 1251–1262.

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.

SONG, K. (2014), "Local Asymptotic Minimax Estimation of Nonregular Parameters with Translation-Scale Equivariant Maps", *Journal of Multivariate Analysis*, **125**, 136–158.

VAN DER VAART, A. (1991), "On Differentiable Functionals", *The Annals of Statistics*, **19**, 178–204.

VAN DER VAART, A. (1998), *Asymptotic Statistics* (Cambridge, UK: Cambridge University Press).

VAN DER VAART, A. and WELLNER, J. (1996), *Weak Convergence and Empirical Processes* (Springer).

VER HOEF, J. M. (2012), "Who Invented the Delta Method?" *The American Statistician*, **66**, 124–127.

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.

WOUTERSEN, T. and HAM, J. C. (2013), "Calculating Confidence Intervals for Continuous and Discontinuous Functions of Parameters" (Cemmap Working Paper (CWP23/13). Center for Microdata Methods and Practice).

ZARANTONELLO, E. H. (1971), "Projections on Convex Sets and Hilbert Spaces and Spectral Theory", in E. H. Zaranotello, ed. *Contributions to Nonlinear Functional Analysis* (New York, USA: Academic Press).