Journal of Econometrics 206 (2018) 379–394

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The numerical delta method[☆]

Han Hong[a,*], Jessie Li[b]

[a] Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, United States
[b] Department of Economics, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA 95064, United States

ARTICLE INFO

Article history:
Available online 2 July 2018

JEL classification:
C12
C13
C50

Keywords:
Delta method
Numerical differentiation
Directional differentiability

ABSTRACT

This paper provides a numerical derivative based Delta method that complements the recent work by Fang and Santos (2014) and also generalizes a previous insight by Song (2014). We show that for an appropriately chosen sequence of step sizes, the numerical derivative based Delta method provides consistent inference for functions of parameters that are only directionally differentiable. Additionally, it provides uniformly valid inference for certain convex and Lipschitz functions which include all the examples mentioned in Fang and Santos (2014). We extend our results to the second order Delta method and illustrate its applicability to inference for moment inequality models.

1. Introduction

Inference on possibly nonsmooth functions of parameters has received much attention in the econometrics literature, as in Woutersen and Ham (2013) and Hirano and Porter (2012). In particular, a recent insightful paper by Fang and Santos (2014) studies inference for functions of the parameters that are only Hadamard directionally differentiable and not necessarily differentiable. Fang and Santos (2014) show that while the asymptotic distribution obtained using the bootstrap is invalid unless the target function of the parameter is differentiable, asymptotic inference using a consistent estimate of the first order directional derivative is valid as long as the target function is Hadamard directionally differentiable. In each of their examples studied, Fang and Santos (2014) constructed consistent analytical estimates of the directional derivative that are tailored to each particular case.

As an alternative to using analytical estimates, we show that numerical differentiation provides a comprehensive approach to estimating the directional derivative. The main advantage of using the numerical directional derivative is its computational simplicity and ease of implementation. In order to compute an estimate of the directional derivative, the user only needs to specify one tuning parameter (the stepsize), and she does not need to perform any additional calculations beyond evaluating the target function twice for each random draw from an approximation of the limiting distribution of the parameter estimates.

Dümbgen (1993) developed a rescaled bootstrap that was implemented for the specific problem of matrix eigenvalues. However, his Proposition 1 essentially provides pointwise consistency of the numerical delta method under directional differentiability. We build on and go beyond these initial contributions by demonstrating how to perform uniformly valid

☆ We thank Andres Santos, Joe Romano, Xiaohong Chen, Zheng Fang, Bruce Hansen, David Kaplan, Adam McCloskey, Frank Wolak, and participants at the Montreal Econometric Society World Congress and various conferences and seminars for helpful comments. In particular, we appreciate the detailed comments by the coeditor, the associate editor, and two anonymous referees that greatly helped improve the paper. We also acknowledge support by the National Science Foundation (SES 1658950), EPR and both the IRiSS and the B.F.Haley and E.S.Shaw SIEPR dissertation fellowships.

* Corresponding author.
E-mail addresses: hanhong@stanford.edu (H. Hong), jeqli@ucsc.edu (J. Li).

https://doi.org/10.1016/j.jeconom.2018.06.007
0304-4076/

inference under convexity and Lipschitz continuity. We also generalize to the second order directional delta method and study its application to a wider range of problems.

The results of this paper also complement Woutersen and Ham (2013), who provide a general inference method for functions of parameters that can be nondifferentiable and even discontinuous. In contrast, our numerical differentiation method only applies to directionally differentiable functions but can be easier to implement. We also contribute to the understanding of the statistical properties of numerical differentiation, which was analyzed in Hong et al. (2015) for different purposes. Most importantly, this paper follows up and complements the insights in Fang and Santos (2014), as well as the extensive analytic derivations in Amemiya (1985).

In some applications, the first order directional derivative may vanish on a set of parameters, which motivates the use of the second order numerical directional delta method. For example, the test statistics for moment inequality models often use the negative square test function, which has the property that the first order directional derivative is exactly zero over the null set. We demonstrate the pointwise consistency of the second order numerical directional derivative and illustrate how it can be used to conduct pointwise valid inference using the second order directional delta method.

The rest of this paper is organized as follows. Section 2 describes the setup of the model that is mostly based on summarizing Fang and Santos (2014), and describes inference based on numerical differentiation. Section 3 first discusses pointwise validity of the numerical directional delta method for all Hadamard directionally differentiable functions and then demonstrates the uniform asymptotic validity of the numerical directional delta method for convex and Lipschitz functions. Convexity and Lipschitz continuity are satisfied in all the examples provided in Fang and Santos (2014) as well as for test statistics used in certain moment inequality models. Extensions of the uniform asymptotic validity results to statistics containing nuisance parameters are discussed in Section 3.3. Section 4 describes the second order numerical directional delta method, and an application to partially identified models such as those studied in Bugni et al. (2015) is illustrated in subsection A.4 of the appendix. Section 5 reports Monte Carlo simulation results on the coverage frequencies of various types of confidence intervals obtained using the first order numerical directional delta method as well as the rejection frequencies for a moment inequalities test based on critical values obtained using the second order numerical directional delta method. Section 6 proposes a multiple point first order numerical directional derivative that could be used to reduce bias, and Section 7 concludes. The appendix contains a list of commonly used symbols, verification of convexity and Lipschitz continuity for several examples, proofs, and other technical material.

2. Numerical directional delta method

Fang and Santos (2014) study inference on a nondifferentiable mapping $\phi(\theta)$ of the parameter $\theta \in \Theta$, where θ can be either finite or infinite dimensional, under the requirement that $\theta \in \mathbb{D}_\phi$ and $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ for $\mathbb{D}$ endowed with norm $||\cdot||_{\mathbb{D}}$ and $\mathbb{E}$ endowed with norm $||\cdot||_{\mathbb{E}}$. The domain of ϕ is $\mathbb{D}_\phi$.

The true parameter is denoted θ_0, for which a consistent estimator $\hat{\theta}_n$ is available which converges in distribution at a suitable rate $r_n \to \infty$: $r_n\left(\hat{\theta}_n - \theta_0\right) \rightsquigarrow \mathbb{G}_0$ in the sense of Eq. (2.8) of Kosorok (2007),[1] where the limit distribution $\mathbb{G}_0$ is tight and is supported on $\mathbb{D}_0 \subset \mathbb{D}$. Examples of nondifferentiable $\phi(\cdot)$ functions arise in a variety of econometric applications such as moment inequalities models (Andrews and Shi, 2013; Ponomareva, 2010) and threshold regression models (Hansen, 2017). Using the notation of Fang and Santos (2014), we describe each of these examples in more detail below.

Generalization of Fang and Santos (2014) Example 2.1. Define $\phi(\theta) = a\theta^+ + b\theta^-$, where $\theta^+ = \max\{\theta, 0\}$ and $\theta^- = -\min\{\theta, 0\}$. Let $X \in \mathbb{R}$, $\theta_0 = E[X]$, and $\mathbb{D} = \mathbb{E} = \mathbb{R}$.

Generalization of Fang and Santos (2014) Example 2.2. $\theta = (\theta_1, \ldots, \theta_K)$ for $\theta_k \in \mathbb{R}^d$, $\phi(\theta) = \max(\theta_1, \ldots, \theta_K)$. $\mathbb{D} = \mathbb{R}^d \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d$ and $\mathbb{E} = \mathbb{R}$.

Fang and Santos (2014) Example 2.3. Define $\phi(\theta_0) = \sup_{f \in \mathcal{F}} E[Yf(Z)]$ as in Andrews and Shi (2013). Here, $Y \in \mathbb{R}$, $Z \in \mathbb{R}^d$, and $\theta_0 \in \ell^\infty(\mathcal{F})$. $\mathcal{F} \subset \ell^\infty\left(R^d\right)$ is a set of functions satisfying $\theta_0(f) \equiv E[Yf(Z)]$ for all $f \in \mathcal{F}$. $\mathbb{D} = \ell^\infty(\mathcal{F})$ and $\mathbb{E} = \mathbb{R}$.

Ponomareva (2010) Example. In theorem 3.5, inference is performed on $\phi(\theta_0) = \max_{x \in \mathcal{X}} E[m(Z_i)|X_i = x]$ where $\theta_0(x) = E[m(Z_i)|X_i = x]$ is the conditional expectation function, $\mathbb{D} = \ell^\infty\left(\mathbb{R}^d\right)$ and $\mathbb{E} = \mathbb{R}$.

The goal of subsequent analysis is to approximate the distribution of $\phi\left(\hat{\theta}_n\right)$, or with proper scaling and centering, that of $r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right)$, for statistical inference concerning $\phi(\theta_0)$. The asymptotic distribution bootstrap (ADB) method (coined by Woutersen and Ham (2013) and further illustrated in theorems 3 and 4 in Chernozhukov and Hong (2003)) uses the empirical distribution formed by repeated draws from

$$r_n\left(\phi\left(\hat{\theta}_n + \frac{\mathbb{Z}_n^*}{r_n}\right) - \phi\left(\hat{\theta}_n\right)\right). \tag{1}$$

[1] $X_n \rightsquigarrow X_n$ in the metric space $(\mathbb{D}, d)$ if and only if $\sup_{f \in BL_1} |E^* f(X_n) - Ef(X)| \to 0$ where BL_1 is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1.

In the above, $\mathbb{Z}_n^*$ is a function of the data and additional randomness, and its distribution given the data converges to $\mathbb{G}_1$ in probability, denoted $\mathbb{Z}_n^* \overset{\mathbb{P}}{\rightsquigarrow} \mathbb{G}_1$ in the sense of section 2.2.3 of Kosorok (2007). Here, $\mathbb{G}_1$ is an identical copy of $\mathbb{G}_0$, the random variable whose distribution is the limiting distribution of $r_n(\hat{\theta}_n - \theta_0)$. Examples of $\mathbb{Z}_n^*$ include the following:

1. Bootstrap: here $\mathbb{Z}_n^* = r_n\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, where $\hat{\theta}_n^*$ are parameter estimates obtained using multinomial, wild, or other commonly used bootstrap implementations. The bootstrap sample size can also be different from the observed sample size. For example, we can take $\mathbb{Z}_n^* = r_{m_n}\left(\hat{\theta}_{m_n}^* - \hat{\theta}_n\right)$, where $m_n \to \infty$ as $n \to \infty$, and $\hat{\theta}_{m_n}^*$ is computed from a multinomial bootstrap sample of size m_n that are i.i.d draws from the empirical distribution. Similar modifications apply to the next few methods.
2. When θ is a finite dimensional parameter, typically $r_n = \sqrt{n}$ and $\mathbb{G}_0 = N(0, \Sigma)$ for some variance covariance matrix Σ. Using a consistent estimate $\hat{\Sigma}$ of Σ, $\mathbb{Z}_n^*$ can be a random vector whose distribution given the data is given by $N\left(0, \hat{\Sigma}\right)$.
3. For correctly specified parametric models, one can use $\mathbb{Z}_n^* = r_n\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, where $\hat{\theta}_n^*$ are MCMC draws from the (pseudo) posterior distribution based on the likelihood or other objective functions (Chernozhukov and Hong, 2003).
4. In Hong and Li (2014), we propose a technique called the numerical bootstrap, which produces estimates $\theta\left(\mathcal{Z}_n^*\right)$ based on the numerical bootstrap empirical measure $\mathcal{Z}_n^* \equiv P_n + \epsilon_n\sqrt{n}\left(P_n^* - P_n\right)$, where P_n is the empirical measure, P_n^* is the bootstrap empirical measure, ϵ_n is a positive scalar step size parameter that satisfies $\epsilon_n \to 0$, and $n^\gamma \epsilon_n \to \infty$. We show that the finite sample distribution of $\mathbb{Z}_n^* = \epsilon_n^{-2\gamma}\left(\theta\left(\mathcal{Z}_n^*\right) - \theta\left(P_n\right)\right)$ converges to the same limiting distribution as that of $n^\gamma\left(\hat{\theta}_n - \theta_0\right)$ for a class of estimators that converge at rate n^γ for some $\gamma \in \left[\frac{1}{4}, 1\right)$.

Intuitively, ADB approximates the distribution of $\phi\left(\hat{\theta}_n\right)$ around $\phi(\theta_0)$ with that of $\phi\left(\hat{\theta}_n^*\right)$ around $\phi\left(\hat{\theta}_n\right)$, where $\hat{\theta}_n^*$ is a suitable version of the bootstrap in case (1); a draw from a consistent estimate of the asymptotic distribution $N\left(\hat{\theta}_n, \frac{1}{r_n^2}\hat{\Sigma}\right)$ in case (2); a draw from the MCMC chain in case (3); and a draw from $\hat{\theta}_n + r_n^{-1}\mathbb{Z}_n^*$ in case (4).

Fang and Santos (2014) showed that the ADB is asymptotically valid only if $\phi(\theta)$ is Hadamard differentiable. The delta method, however, is applicable more generally even when ADB fails, as long as $\phi(\theta)$ is Hadamard directionally differentiable even if it is not Hadamard differentiable. Fang and Santos (2014) make use of the following definition:

Definition 2.1. The map ϕ is said to be Hadamard directionally differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to a set $\mathbb{D}_0 \subset \mathbb{D}$ if there is a continuous map $\phi_\theta' : \mathbb{D}_0 \to \mathbb{E}$ such that:

$$\lim_{n\to\infty}\left\|\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi_\theta'(h)\right\|_{\mathbb{E}} = 0, \tag{2}$$

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

When $\phi(\cdot)$ is directionally differentiable in the sense defined above and when the support of the limiting distribution $\mathbb{G}_0$ is contained in $\mathbb{D}_0$, Fang and Santos (2014) showed that under suitable regularity conditions, $r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right) \rightsquigarrow \phi_{\theta_0}'(\mathbb{G}_0)$. Based on this result, Fang and Santos (2014) suggested that this limiting distribution can be consistently estimated by $\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)$, where $\mathbb{Z}_n^*$ is a consistent estimate of $\mathbb{G}_0$ (such as the bootstrap, MCMC or asymptotic normal approximation), and in particular $\hat{\phi}_n'(\cdot)$ is a consistent estimate of $\phi_{\theta_0}'(\cdot)$ in a sense that is precisely defined in their Assumption 3.3.

FS Assumption 3.3. For each fixed θ_0, each compact set $K \subseteq \mathbb{D}$, and for any sequence $\delta \downarrow 0$,

$$d_{\delta,K}\left(\hat{\phi}_n'(\cdot), \phi_{\theta_0}'(\cdot)\right) \equiv \sup_{h\in K^\delta}\left\|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)\right\|_{\mathbb{E}} = o_p(1) \qquad \text{as } n \to \infty. \tag{3}$$

In the above K^δ denotes the δ-enlargement of a set K: $K^\delta \equiv \{a \in \mathbb{D} : inf_{b\in K}\|a - b\|_{\mathbb{D}} < \delta\}$. We show that the one-sided numerical derivative provides a $\hat{\phi}_n'(\cdot)$ for which this assumption holds whenever $\phi(\cdot)$ is Lipschitz. In particular, Definition 2.1 motivates the following estimate $\hat{\phi}_n'(\cdot)$ based on a one-sided finite difference formula. For $\epsilon_n \to 0$ slowly (in the sense that $r_n\epsilon_n \to \infty$, where r_n is the convergence rate of $\hat{\theta}_n$ to θ_0), define

$$\hat{\phi}_n'(h) \equiv \frac{\phi\left(\hat{\theta}_n + \epsilon_n h\right) - \phi\left(\hat{\theta}_n\right)}{\epsilon_n} \tag{4}$$

as the numerical directional derivative of ϕ in the direction of $h \in \mathbb{D}_0$. The rate requirement on the step size ϵ_n is needed to separate numerical differentiation error from the estimation error in $\hat{\theta}_n$, and serves the dual purposes of model selection and numerical differentiation.

For functions that are not Lipschitz, Section 3.1 shows that the one-sided numerical derivative will continue to consistently estimate the directional derivative as long as the function is Hadamard directionally differentiable.

The numerical directional delta method.

Given the definition in (4), the numerical directional delta method estimates the limiting distribution of $r_n\left(\phi\left(\hat{\theta}_n\right)-\phi\left(\theta_0\right)\right)$ using the distribution of the random variable:

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)\equiv\frac{\phi\left(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*\right)-\phi\left(\hat{\theta}_n\right)}{\epsilon_n} \tag{5}$$

which can be approximated by the following:

1. Draw $\mathbb{Z}_s$ from the distribution of $\mathbb{Z}_n^*$ for $s=1,\ldots,S$.
2. For the given ϵ_n, evaluate for each s:

$$\hat{\phi}_n'\left(\mathbb{Z}_s\right)\equiv\frac{\phi\left(\hat{\theta}_n+\epsilon_n\mathbb{Z}_s\right)-\phi\left(\hat{\theta}_n\right)}{\epsilon_n}. \tag{6}$$

The empirical distribution of $\hat{\phi}_n'\left(\mathbb{Z}_s\right), s=1,\ldots,S$ can then be used for confidence interval construction, hypothesis testing, or variance estimation. Consider the case when $\phi(\cdot)\in\mathbb{R}$ is a scalar. For example, a $1-\tau$ two-sided equal-tailed confidence interval for $\phi\left(\theta_0\right)$ can be formed by

$$\left[\phi(\hat{\theta})-\frac{1}{r_n}c_{1-\tau/2},\phi(\hat{\theta})-\frac{1}{r_n}c_{\tau/2}\right]$$

where $c_{\tau/2}$ and $c_{1-\tau/2}$ are the $\tau/2$ and $1-\tau/2$ empirical percentiles of $\hat{\phi}_n'\left(\mathbb{Z}_s\right)$. Symmetric confidence intervals can be formed by, where $d_{1-\tau}$ is the $1-\tau$ percentile of $|\hat{\phi}_n'(\mathbb{Z}_n^*)|$,

$$\left[\phi(\hat{\theta})-\frac{1}{r_n}d_{1-\tau},\phi(\hat{\theta})+\frac{1}{r_n}d_{1-\tau}\right]$$

Note that the random variable $\hat{\phi}_n'\left(\mathbb{Z}_s\right)$ only requires two evaluations of the $\phi(\cdot)$ function for each draw of $\mathbb{Z}_s$. The computational simplicity of the numerical derivative is one of its main advantages. In Eq. (5), $\mathbb{Z}_n^*$ can be any of the four choices discussed in the ADB method after Eq. (1). In particular, Fang and Santos (2014) recommended the bootstrap $\mathbb{Z}_n^*=r_n\left(\hat{\theta}_n^*-\hat{\theta}_n\right)$. Following the tradition of the literature (except Andrews and Buchinsky, 2000), we take $S=\infty$ in analyzing $\hat{\phi}_n'(\mathbb{Z}_n^*)$. Subsampling is also a special case of (5) when $\mathbb{Z}_n^*$ is the $\binom{n}{b}$ point discrete distribution of $r_b\left(\hat{\theta}_{n,b,i}-\hat{\theta}_n\right)$ (Eq. (2.1) page 42 of Politis et al. (1999)) and when $\epsilon_n=1/\sqrt{b}$. When all $\binom{n}{b}$ are used in subsampling, no simulation error is involved ($S=\infty$). Simulating $\mathbb{Z}_s$ from $\mathbb{Z}_n$ is only relevant when one randomly draws from the $\binom{n}{b}$ blocks.

We now give the form of $\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)$ in examples 2.1 and 2.3 of Fang and Santos (2014).

Fang and Santos (2014) Example 2.1. With $\mathbb{Z}_n^*\sim N\left(0,\hat{\sigma}_n^2\right)$ and $\hat{\sigma}_n^2$ the usual sample variance:

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)\equiv\frac{a\left(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*\right)^+ + b\left(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*\right)^- - a\hat{\theta}_n^+ + b\hat{\theta}_n^-}{\epsilon_n}.$$

Fang and Santos (2014) Example 2.3. Note that $\hat{\theta}_n(f)\equiv\theta\left(P_n\right)(f)\equiv\frac{1}{n}\sum_{i=1}^n y_if\left(z_i\right)$. Its multinomial bootstrap version is given by $\hat{\theta}_n^*(f)\equiv\theta\left(P_n^*\right)(f)\equiv\frac{1}{n}\sum_{i=1}^n y_i^*f\left(z_i^*\right)$. Alternatively the multiplier bootstrap can be used: $\theta\left(P_n^*\right)(f)\equiv\frac{1}{n}\sum_{i=1}^n\xi_i^*y_if\left(z_i\right)$ for positive random variables ξ_i^* with $E\xi_i^*=1$. In this case $\hat{\theta}_n=\theta\left(P_n\right)$, $\mathbb{Z}_n^*=\sqrt{n}\left(\theta\left(P_n^*\right)-\theta\left(P_n\right)\right)$, so that with the multinomial bootstrap,

$$\begin{aligned}\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)&\equiv\frac{\sup_{f\in\mathcal{F}}\theta\left(P_n+\epsilon_n\sqrt{n}\left(P_n^*-P_n\right)\right)(f)-\sup_{f\in\mathcal{F}}\theta\left(P_n\right)(f)}{\epsilon_n}\\&=\frac{\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n\left(y_if\left(z_i\right)+\epsilon_n\sqrt{n}\left(y_i^*f\left(z_i^*\right)-y_if\left(z_i\right)\right)\right)-\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n y_if\left(z_i\right)}{\epsilon_n},\end{aligned}$$

or with multiplier bootstrap

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)=\frac{\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n\left(y_if\left(z_i\right)+\epsilon_n\sqrt{n}\left(\xi_i^*y_if\left(z_i\right)-y_if\left(z_i\right)\right)\right)-\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^n y_if\left(z_i\right)}{\epsilon_n}.$$

A similar procedure can be applied to each of the examples in Fang and Santos (2014).

In the context of a matrix eigenvalue application and a minimum distance application, Dümbgen (1993) presented a "rescaled bootstrap method" which corresponds essentially to the numerical delta method, where the rescaling sample size is inversely related to the step size in numerical differentiation. Dümbgen (1993) showed pointwise consistency which is essentially Theorem 3.1 in Section 3.1, but did not present uniformity results. The idea of using numerical differentiation for directionally differentiable parameters also appeared in Song (2014), although Song (2014) only considered finite dimensional $\theta \in \mathbb{R}^d$ and scalar functions $\phi(\cdot) \in \mathbb{R}$ that are (1) translation equivalent: $\phi(\theta + c) = \phi(\theta) + c$ for $c \in \mathbb{R}$; and (2) scale equivalent: $\phi(\alpha\theta) = \alpha\phi(\theta)$ for $\alpha \geq 0$. Under these conditions Song (2014) gives the following more specialized form of the numerical derivative formula $\hat{\phi}_n'(\mathbb{Z}_n^*) \equiv \phi\left(\mathbb{Z}_n^* + \epsilon_n^{-1}\left(\hat{\theta}_n - \phi\left(\hat{\theta}_n\right)\right)\right)$. If $\phi(\cdot)$ is only scale equivalent as in an $\mathcal{L}_1$ version of Andrews and Soares (2010) and Bugni et al. (2015) discussed in Section 3.3, then equivalently, $\hat{\phi}_n'(\mathbb{Z}_n^*) \equiv \phi\left(\mathbb{Z}_n^* + \epsilon_n^{-1}\hat{\theta}_n\right) - \phi\left(\epsilon_n^{-1}\hat{\theta}_n\right)$.

3. Asymptotic validity

This section shows that the numerical directional delta method provides consistent inference under general conditions. We first verify pointwise consistency and then discuss uniform validity.

3.1. Pointwise asymptotic distribution

In this subsection we show pointwise consistency of the numerical delta method using the definition of Hadamard directional differentiability and (a bootstrap version of) the extended continuous mapping theorem. The first part of the following theorem is a directional delta method due to Dümbgen (1993), Fang and Santos (2014), and references therein. The second part of the theorem shows consistency of the numerical delta method. Let BL_1 be the space of Lipschitz functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1. For random variables F_1 and F_2, let $\rho_{BL_1}(F_1, F_2) = \sup_{f \in BL_1} |Ef(F_1) - Ef(F_2)|$ metrize weak convergence. As in Kosorok (2007) (pages 19–20), we use $\overset{\mathbb{P}}{\rightsquigarrow}$ to denote weak convergence in probability conditional on the data.[2]

Theorem 3.1. *Suppose $\mathbb{D}$ and $\mathbb{E}$ are Banach Spaces and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \mapsto \mathbb{E}$ is Hadamard directionally differentiable at θ_0 tangentially to $\mathbb{D}_0$. Let $\hat{\theta}_n : \{X_i\}_{i=1}^n \mapsto \mathbb{D}_\phi$ be such that for some $r_n \uparrow \infty$, $r_n\{\hat{\theta}_n - \theta_0\} \rightsquigarrow \mathbb{G}_0$ in $\mathbb{D}$, where $\mathbb{G}_0$ is tight and its support is included in $\mathbb{D}_0$. Then $r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right) \rightsquigarrow \phi_{\theta_0}'(\mathbb{G}_0)$. Let $\mathbb{Z}_n^* \overset{\mathbb{P}}{\rightsquigarrow} \mathbb{G}_0$ satisfy certain measurability assumptions stated in the appendix. Then for $\epsilon_n \to 0$, $r_n\epsilon_n \to \infty$,*

$$\hat{\phi}_n'(\mathbb{Z}_n^*) \equiv \frac{\phi\left(\hat{\theta}_n + \epsilon_n\mathbb{Z}_n^*\right) - \phi\left(\hat{\theta}_n\right)}{\epsilon_n} \overset{\mathbb{P}}{\rightsquigarrow} \phi_{\theta_0}'(\mathbb{G}_0).$$

An alternative approach to showing consistency is to use remark 3.6 and Lemma A.6 in Fang and Santos (2014), which place Lipschitz and Hölder continuity requirements on $\hat{\phi}_n'(\cdot)$, a consistent estimate of the directional derivative function. These results in Fang and Santos (2014) apply more generally to $\hat{\phi}_n'(\cdot)$ constructed using alternative methods other than numerical differentiation. The particular structure of the numerical delta method allows us to invoke the bootstrap extended continuous mapping theorem directly without having to rely on these intermediate conditions. However, establishing these conditions turns out to be important for uniform validity considerations in the next section, and are thus presented here.

Lemma 3.1 (*Fang and Santos (2014) Remark 3.6 and Lemma A.6*). *If the directional derivative estimate is Hölder continuous in the direction arguments, namely, if there exist some $\kappa > 0$ and fixed constant $C_0 < \infty$ such that for all $h_1, h_2 \in \mathbb{D}_0$ and all $n \geq 1$,*

$$\|\hat{\phi}_n'(h_1) - \hat{\phi}_n'(h_2)\|_{\mathbb{D}} \leq C_0\|h_1 - h_2\|_{\mathbb{D}}^\kappa \tag{7}$$

then Fang and Santos (2014) assumption 3.3 holds as long as pointwise for each $h \in \mathbb{D}_0$,

$$\left\|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)\right\|_{\mathbb{E}} = o_p(1). \tag{8}$$

Our first result provides the simple finding that whenever the function $\phi(\cdot)$ is Lipschitz ($\kappa = 1$), so is the one-sided numerical directional derivative.

Theorem 3.2. *If $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is Lipschitz, satisfying $\|\phi(h_1) - \phi(h_2)\|_{\mathbb{E}} \leq C\|h_1 - h_2\|_{\mathbb{D}}$ for all $h_1, h_2 \in \mathbb{D}$, and for Lipschitz constant C that does not depend on n, then so is $\hat{\phi}_n'(h) \equiv \frac{\phi(\hat{\theta}_n + \epsilon_n h) - \phi(\hat{\theta}_n)}{\epsilon_n}$ in h for all $\epsilon_n > 0$.*

[2] $\hat{X}_n \overset{\mathbb{P}}{\rightsquigarrow} X$ means that $\hat{X}_n$ is a random function of the data and $\sup_{f \in BL_1}\left|E\left[f(\hat{X}_n)|\mathcal{X}_n\right] - Ef(X)\right| \overset{p}{\to} 0$ (where $\mathcal{X}_n$ denotes the data).

Note also that $\phi'_\theta(h)$ is Lipschitz in h for all θ whenever $\phi(\theta)$ is Lipschitz:

$$\|\phi'_\theta(h_1) - \phi'_\theta(h_2)\|_{\mathbb{E}} \le \lim_{t\downarrow 0}\left\|\frac{\phi(\theta+th_1)}{t} - \frac{\phi(\theta+th_1)}{t}\right\|_{\mathbb{E}} \le C\|h_1 - h_2\|_{\mathbb{D}}. \tag{9}$$

Theorem 3.2 and Lemma 3.1 imply that whenever the function $\phi(\cdot)$ is Lipschitz, it suffices to verify the pointwise consistency condition in (8).

Theorem 3.3. *Let the conditions in Theorem 3.1 hold for $\phi(\cdot)$ and $\hat\theta_n$. If $\epsilon_n \downarrow 0$ and $r_n\epsilon_n \to \infty$, then for $\hat\phi_n(\cdot)$ defined in (4) and for any $h \in \mathbb{D}_0$, $\left\|\hat\phi'_n(h) - \phi'_{\theta_0}(h)\right\|_{\mathbb{E}} = o_p(1)$.*

To summarize, we have shown that if the function $\phi(\cdot)$ is Lipschitz in its argument of the parameter, then so is the numerical directional derivative $\hat\phi'_n(\cdot)$ in its argument of the direction of differentiation, uniformly in the step size ϵ_n. Furthermore, we have shown that $\hat\phi'_n(h)$ converges in probability to $\phi'_{\theta_0}(h)$ for each fixed $h \in \mathbb{D}_0$. Whenever $\phi(\cdot)$ is Lipschitz, we have shown that the numerical directional derivative $\hat\phi'_n(h)$ satisfies Fang and Santos (2014) Lemma A.6, remark 3.6 and in turn Fang and Santos (2014) Assumption 3.3. Consequently, the remaining results in Fang and Santos (2014) imply that inference based on $\hat\phi'_n(\mathbb{Z}^*_n)$ is asymptotically valid, in a formal sense. Intuitively, when ϵ_n is much larger than $\frac{1}{r_n}$, the estimation error in $\hat\theta_n$ does not obscure the true direction for which the derivative is being calculated. It turns out that whenever $\phi(\cdot)$ is Lipschitz, Hadamard differentiability is equivalent to Gateaux differentiability as noted in proposition 3.5 of Shapiro (1990).[3]

Theorem 3.2 depends crucially on the function $\phi(\cdot)$ being Lipschitz in the parameter argument. This turns out to be a rather weak requirement that is satisfied by all the examples in Fang and Santos (2014). The calculations in the appendix verify that the Lipschitz condition holds for all the functions $\phi(\cdot)$ in examples 2.1–2.5, as well as the convex projection inference problem in Fang and Santos (2014). Consequently, the numerical delta method (4) provides a (pointwise) consistent asymptotic approximation for the distribution of $r_n\left(\phi\left(\hat\theta_n\right) - \phi(\theta_0)\right)$ in each of these examples, including the convex projection problem in Fang and Santos (2014).

For example, for $\phi(\theta) = \inf_{\lambda\in\Lambda}\|\theta-\lambda\|$ which defines the distance between θ and its projection onto the convex set Λ, the distribution of $r_n\left(\phi\left(\hat\theta_n\right) - \phi(\theta_0)\right)$ is accurately approximated by

$$\hat\phi'_n(\mathbb{Z}^*_n) = \frac{1}{\epsilon_n}\left(\inf_{\lambda\in\Lambda}\|\hat\theta_n + \epsilon_n\mathbb{Z}^*_n - \lambda\| - \inf_{\lambda\in\Lambda}\|\hat\theta_n - \lambda\|\right) \tag{10}$$

for some $\mathbb{Z}^*_n \overset{\mathbb{P}}{\rightsquigarrow} \mathbb{G}_0$ where $r_n\left(\hat\theta_n - \theta_0\right) \rightsquigarrow \mathbb{G}_0$. Evaluating the distribution of $\hat\phi'_n(\mathbb{Z}^*_n)$ requires solving $2\times S$ optimization routines, where S is the number of draws from $\mathbb{Z}^*_n$. This is more computationally efficient than the original solutions provided in Fang and Santos (2014), which are based on combining a model selection scheme with analytic knowledge of the function $\phi(\cdot)$. To illustrate this difference, consider again Fang and Santos (2014) example 2.1.

Fang and Santos (2014) Example 2.1.

Fang and Santos (2014) proposed to estimate $\phi'_{\theta_0}(h)$ by h if $\hat\theta_n > \kappa_n$, by $-h$ if $\hat\theta_n < -\kappa_n$, and by $|h|$ when $|\hat\theta_n| < \kappa_n$, where the selection parameter κ_n satisfies the same rate condition as the step size parameter ϵ_n: $\kappa_n \to 0$ but $\kappa_n\sqrt{n} \longrightarrow 0$. In other words, for $\phi(\theta_0) = |\theta_0|$, $\hat\phi'_n(h)$ is set to h if $\hat\theta_n$ is sufficiently positive, to $-h$ if $\hat\theta_n$ is sufficiently negative, and to $|h|$ if $\hat\theta_n$ is sufficiently close to zero.

Instead, we use the numerical directional derivative in (4):

$$\hat\phi'_n(h) \equiv \frac{\phi\left(\hat\theta_n + \epsilon_n h\right) - \phi\left(\hat\theta_n\right)}{\epsilon_n} = \frac{|\hat\theta_n + \epsilon_n h| - |\hat\theta_n|}{\epsilon_n}, \tag{11}$$

is never exactly equal to h, $-h$, or $|h|$. Instead, under the condition that $\epsilon_n \to 0$ and $\sqrt{n}\epsilon_n \to \infty$, $\hat\phi'_n(h)$ converges in probability to h when $\theta_0 > 0$, converges to $-h$ when $\theta_0 < 0$, and converges to $|h|$ when $\theta_0 = 0$. Consistent inference follows then from Slutsky's lemma.

The Lipschitz assumption can be relaxed to Hölder continuity and Fang and Santos (2014) Assumption 3.3 can still be satisfied under a stronger condition on the step size parameter, as the following theorem shows.

Theorem 3.4. *If $\phi(\cdot)$ is Holder continuous with exponent κ and $r_n^\kappa\epsilon_n \to \infty$, then for all compact $K \subset \mathbb{D} = \mathbb{R}^d$, $\sup_{h\in K}\left\|\hat\phi'_n(h) - \phi'_{\theta_0}(h)\right\|_{\mathbb{E}} = o_p(1)$.*

In finite dimension situations, K can be replaced by $K^\delta \equiv \{a \in \mathbb{D} : \inf_{b\in K}\|a-b\|_{\mathbb{D}} < \delta\}$. In general, as in Fang and Santos (2014), Fréchet directional differentiability might be needed to allow for replacement of K by K^δ.

3 We thank a referee for pointing this out.

3.2. Uniform inference

Uniform asymptotic validity over a class of distributions can be a desirable feature to establish for an inference procedure (Romano and Shaikh 2008, 2012). The Lipschitz and convexity properties of $\phi(\cdot)$ are key to establishing uniform size control in the test of $H_0 : \phi(\theta_0) \le 0$ versus $H_1 : \phi(\theta_0) > 0$.

As we show in the Appendix, the $\phi(\cdot)$ functionals considered in the examples in Fang and Santos (2014) are not only Lipschitz but also convex, so that for $\lambda \in [0, 1]$,

$$\phi(\lambda\theta_1 + (1-\lambda)\theta_2) \le \lambda\phi(\theta_1) + (1-\lambda)\phi(\theta_2).$$

We first note that convexity of the functional $\phi(\cdot)$ implies subadditivity of the directional derivative ϕ'_{θ_0}, which then implies sublinearity since the directional derivative is positively homogeneous of degree 1.

Lemma 3.2. *When $\phi(\cdot)$ is convex and Hadamard directionally differentiable at θ_0 and $\mathbb{D}_0$ is a convex set, then $\forall 0 \le \lambda \le 1$,*

$$\phi'_{\theta_0}(h_1 + h_2) \le \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2), \quad \phi'_{\theta_0}(\lambda h_1 + (1-\lambda)h_2) \le \lambda\phi'_{\theta_0}(h_1) + (1-\lambda)\phi'_{\theta_0}(h_2). \tag{12}$$

Fang and Santos (2014) use the statistic $r_n\phi\left(\hat\theta_n\right)$ to test:

$$H_0 : \phi(\theta_0) \le 0 \quad \text{against} \quad H_1 : \phi(\theta_0) > 0. \tag{13}$$

and suggested rejecting H_0 whenever $r_n\phi\left(\hat\theta_n\right) \ge \hat c_{1-\tau}$, where $\hat c_{1-\tau}$ is the $1-\tau$ quantile of $\hat\phi'_n\left(\mathbb{Z}^*_n\right)$ or its simulated version in (6). This is related to the one-sided confidence interval in Part (i) of Theorem 2.1 in Romano and Shaikh (2012):

$$P\left(r_n\left(\phi\left(\hat\theta_n\right) - \phi(\theta_0)\right) \le \hat c_{1-\tau}\right). \tag{14}$$

Whenever $\phi(\theta)$ is convex and Lipschitz in θ, using the $1-\tau$ percentile of $\hat\phi'_n\left(\mathbb{Z}^*_n\right)$ as $\hat c_{1-\tau}$ provides uniform size control for both (13) and (14) under the condition that $r_n\epsilon_n \to \infty$ without requiring $\epsilon_n \to 0$. Intuitively, convexity implies for $\epsilon_n > \frac{1}{r_n}$ and for any realization z from $\mathbb{G}_0$,

$$r_n\left(\phi\left(\theta_0 + \frac{z}{r_n}\right) - \phi(\theta_0)\right) \le \frac{1}{\epsilon_n}\left(\phi(\theta_0 + \epsilon_n z) - \phi(\theta_0)\right), \tag{15}$$

so that $\frac{1}{\epsilon_n}\left(\phi(\theta_0 + \epsilon_n\mathbb{G}_0) - \phi(\theta_0)\right)$ first order stochastically dominates $r_n\left(\phi\left(\theta_0 + \frac{\mathbb{G}_0}{r_n}\right) - \phi(\theta_0)\right)$.[4] If we denote, using notations from Romano and Shaikh (2012), the distribution functions of the two sides of (15) by $J_n(x, \mathbb{G}_0)$ and $J_{\epsilon_n}(x, \mathbb{G}_0)$, then Eq. (15) immediately implies that

$$\sup_n \sup_{x\in\mathbb{R}} \{J_{\epsilon_n}(x, \mathbb{G}_0) - J_n(x, \mathbb{G}_0)\} \le 0. \tag{16}$$

Next, $\phi(\theta)$ being Lipschitz ensures that $r_n\left(\phi\left(\theta_0 + \frac{\mathbb{G}_0}{r_n}\right) - \phi(\theta_0)\right)$ is close to $r_n\left(\phi\left(\hat\theta_n\right) - \phi(\theta_0)\right)$, whose distribution function is denoted $J_n(x, P)$, while $\frac{1}{\epsilon_n}\left(\phi(\theta_0 + \epsilon_n\mathbb{G}_0) - \phi(\theta_0)\right)$ is close to $\hat\phi'_n\left(\mathbb{Z}^*_n\right)$, whose conditional distribution function given the data is $J_{\epsilon_n}(x, P)$, so that $J_n(x, \mathbb{G}_0)$ and $J_{\epsilon_n}(x, \mathbb{G}_0)$ in (16) can be replaced by their feasible sample versions.

Uniformity statements in line with those in Romano and Shaikh (2012) are possible under the following assumptions. We focus on the finite dimensional case $\mathbb{D} = \mathbb{R}^d$ and $\mathbb{E} = \mathbb{R}$.

Assumption 3.1. Let $\mathcal{P}$ be a class of distributions such that

(i) $\lim_{n\to\infty}\sup_{P\in\mathcal{P}}\rho_{BL_1}\left(r_n\left(\hat\theta_n - \theta(P)\right), \mathbb{G}_0\right) = 0$, $\lim_{M\to\infty}\sup_{P\in\mathcal{P}}P(|\mathbb{G}_0| \ge M) = 0$;

(ii) for each $\epsilon > 0$, $\lim_{n\to\infty}\sup_{P\in\mathcal{P}}P\left(\rho_{BL_1}\left(\mathbb{Z}^*_n, \mathbb{G}_0\right) \ge \epsilon\right) = 0$.

Primitive conditions for Assumption 3.1 can be found for example in the uniform central limit theorems of Romano and Shaikh (2008).

Assumption 3.2. Define for each x, a, d, $\mathcal{C}_{a,d,x} = \{g : \phi\left(d + \frac{g}{a}\right) \le x\}$. Then

$$\sup_{P\in\mathcal{P}} P\left(\mathbb{G}_0 \in \partial\mathcal{C}_{a,d,x}\right) = 0 \quad \text{for all } x, a, d,$$

where $\partial\mathcal{C}_{a,d,x}$ denotes the boundary of $\mathcal{C}_{a,d,x}$.

[4] Eq. (15) follows from rewriting it as, for $r_n\epsilon_n > 1$, $\phi\left(\theta_0 + \frac{z}{r_n}\right) \le \frac{1}{r_n\epsilon_n}\phi(\theta_0 + \epsilon_n z) + \left(1 - \frac{1}{r_n\epsilon_n}\right)\phi(\theta_0)$.

Assumption 3.2 is mainly used to invoke versions of Theorem 2.11 of Bhattacharya and Rao (1986), as in Example 3.2 of Romano and Shaikh (2012). If $\phi(\cdot)$ is scale equivariant, then it is sufficient to check all $\mathcal{C}_{d,x} \equiv \{g : \phi(d+g) \leq x\}$. Convexity is crucial in the following.

Theorem 3.5. *Define $\mathcal{P}$ to be a class of DGPs such that $r_n\left(\hat{\theta}_n - \theta(P)\right)$ is asymptotically tight uniformly over $P \in \mathcal{P}$, and Assumptions 3.1, 3.2 both hold. If $r_n\epsilon_n \to \infty$, $\epsilon_n \to 0$, and $\phi(\cdot)$ is Lipschitz and convex, then $\forall\epsilon > 0$,*

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(\sup_{x\in A} J_{\epsilon_n}(x,P) - J_n(x,P) \leq \epsilon\right) \to 1$$

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta(P))\right) \geq \hat{c}_{1-\tau}\right) \leq \tau$$

where A is any set for which $\lim_{\lambda\to 0}\sup_{P\in\mathcal{P}}\sup_{x\in A} P\left(J_{\epsilon_n}(\cdot,\mathbb{G}_0) \in (x, x+\lambda)\right) = o(1)$ and contains a neighborhood of both $J_{\epsilon_n}^{-1}(1-\tau,\mathbb{G}_0)$ and $J_n^{-1}(1-\tau,P)$ for all large n. We have used $J_{\epsilon_n}(\cdot,\mathbb{G}_0)$ to denote the random variable defined by the right hand side of (15).

According to Theorem 3.5, whenever $\phi(\cdot)$ is convex, the lower one-sided confidence interval $\left[\phi\left(\hat{\theta}_n\right) - \frac{\hat{c}_{1-\tau}}{r_n}, \infty\right)$ will have uniformly asymptotically valid coverage. Similarly, if $\phi(\cdot)$ is instead a concave function, then the same arguments will establish that the upper one-sided confidence interval of the form of $\left(-\infty, \phi\left(\hat{\theta}_n\right) - \frac{\hat{c}_{\tau}}{r_n}\right]$ has uniformly asymptotically valid coverage. Furthermore, if it is known that $\phi(\cdot) \geq 0$ (e.g. Andrews, 2000), we can use $\epsilon_n^{-1}\phi\left(\hat{\theta}_n + \epsilon_n\mathbb{Z}_n^*\right)$ in place of $\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)$ at the cost of being more conservative. Furthermore, if the least favorable null distribution is desired in hypothesis testing, then $\hat{\theta}_n$ can also be replaced by the least favorable null value θ_0 if θ_0 is known. In this case, $\hat{\phi}_n'\left(\mathbb{Z}_n^*\right) = \frac{1}{t_n}\left(\phi\left(\theta_0 + t_n\mathbb{Z}_n^*\right) - \phi(\theta_0)\right)$ consistently estimates the null distribution for any $t_n \to 0$ by the extended continuous mapping theorem. If we take $t_n = r_n^{-1}$ and use the bootstrap distribution $\mathbb{Z}_n^* = r_n\left(\hat{\theta}_n^* - \hat{\theta}_n\right)$, a *modified* bootstrap uses $r_n\left(\phi\left(\theta_0 + \hat{\theta}_n^* - \hat{\theta}_n\right) - \phi(\theta_0)\right)$ to approximate the null distribution of $r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right)$. However, it does not provide moment selection to improve the power of the test and does not offer uniform size control for $r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right)$ under drifting sequences of θ_n. In some cases, if only $\phi(\theta) = \phi_0$ but not θ_0 is known under the null, $\hat{\theta}_n$ can be either the constrained or unconstrained estimate. Note also that the only use of convexity of $\phi(\cdot)$ is the stochastic dominance condition in (15) and (16). Therefore the convexity requirement of $\phi(\cdot)$ can be replaced by the following stochastic dominance condition:

Assumption 3.3. For all θ_0, and for all $t > 0$, $\frac{\phi(\theta_0 + t\mathbb{G}_0) - \phi(\theta_0)}{t}$ is nondecreasing in t.

Even if $\phi(\theta)$ is not convex and does not satisfy Assumption 3.3, it is still possible to establish uniform size control over θ_0 under sufficient conditions for the limiting distribution of the numerical directional derivative to stochastically dominate the analytic limiting distribution over all θ_0 that lie in the null set.

Assumption 3.4. For any θ_0, for all η sufficiently close to zero and for all $t > 0$, $\frac{\phi_{\theta_0}'(\eta + t\mathbb{G}_0) - \phi_{\theta_0}'(\eta)}{t}$ is nondecreasing in t.

Clearly Assumption 3.3 (which in turn is implied by $\phi(\cdot)$ being convex) is a sufficient condition for Assumption 3.4. Assumption 3.4 is also satisfied if $\phi_{\theta_0}'(h)$ is convex in h (which in turn follows from convexity of $\phi(\cdot)$), since for $t_2 > t_1 > 0$ and any realization z from $\mathbb{G}_0$, $\frac{\phi_{\theta_0}'(\eta + t_1 z) - \phi_{\theta_0}'(\eta)}{t_1} \leq \frac{\phi_{\theta_0}'(\eta + t_2 z) - \phi_{\theta_0}'(\eta)}{t_2}$ follows from rewriting $\phi_{\theta_0}'(\eta + t_1 z) \leq \left(1 - \frac{t_1}{t_2}\right)\phi_{\theta_0}'(\eta) + \left(\frac{t_1}{t_2}\right)\phi_{\theta_0}'(\eta + t_2 z)$. Assumption 3.4 plays a similar role to (15) and (16) and implies for $\epsilon_n r_n > 1$ and any realization z from $\mathbb{G}_0$,

$$r_n\left(\phi_{\theta_0}'\left(\eta + \frac{z}{r_n}\right) - \phi_{\theta_0}'(\eta)\right) \leq \frac{\phi_{\theta_0}'(\eta + \epsilon_n z) - \phi_{\theta_0}'(\eta)}{\epsilon_n} = \phi_{\theta_0}'\left(\frac{\eta}{\epsilon_n} + z\right) - \phi_{\theta_0}'\left(\frac{\eta}{\epsilon_n}\right) \tag{17}$$

In order for $r_n\left(\phi_{\theta_0}'\left(\eta + \frac{\mathbb{G}_0}{r_n}\right) - \phi_{\theta_0}'(\eta)\right)$ to provide a good approximation to $r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right)$ and for $\phi_{\theta_0}'\left(\frac{\eta}{\epsilon_n} + \mathbb{G}_0\right) - \phi_{\theta_0}'\left(\frac{\eta}{\epsilon_n}\right)$ to provide a good approximation to $\hat{\phi}_n'\left(\mathbb{Z}_n^*\right)$, we require the following additional assumption.

Assumption 3.5. Suppose $\mathbb{D}_0$ is convex. For any $t_n \downarrow 0$, $\eta_n \to \infty$, and any given θ_0:

$$\lim_{t_n\downarrow 0, \eta_n\to\infty}\left|\frac{1}{t_n}\left(\phi(\theta_0 + \eta_n + t_n h) - \phi(\theta_0 + \eta_n)\right) - \left(\phi_{\theta_0}'\left(\frac{\eta_n}{t_n} + h\right) - \phi_{\theta_0}'\left(\frac{\eta_n}{t_n}\right)\right)\right| = 0.$$

We now state a uniformity result similar to Andrews and Soares (2010) without relying on convexity.

Theorem 3.6. *Let $\phi(\cdot)$ be Lipschitz, $r_n\epsilon_n \to \infty$, and $\epsilon_n \to 0$. Define $\mathcal{P}$ to be a class of DGPs such that $r_n\left(\hat{\theta}_n - \theta(P)\right)$ is asymptotically tight uniformly over $P \in \mathcal{P}$, Assumptions 3.1 and 3.2 hold, and for which $\phi(\cdot)$ satisfies either Assumption 3.3 or Assumptions 3.4 and 3.5. Then, $\forall \epsilon, \delta > 0$ and $x = J_n^{-1}(1-\tau-\epsilon, P)$, $\sup_{P\in\mathcal{P}}\left(J_{\epsilon_n}(x,P) \le J_n(x,P)+\epsilon\right) \ge 1-\delta$. Consequently, $\limsup_{n\to\infty}\sup_{P\in\mathcal{P}} P\left(r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta(P))\right) \ge \hat{c}_{1-\tau}\right) \le \tau$.*

It turns out that the following additional condition is also satisfied in most of the examples in Fang and Santos (2014) and in Andrews and Soares (2010): For all $v_n \to v$, $|v| = 1$, and all $|a_n| \to 0$, $\phi'_{\theta_0,v}(\cdot) = \lim_{n\to\infty}\phi'_{\theta_0+|a_n|v_n}(\cdot)$, which is the limit of the directional derivative along direction v, is well defined. It is not required for results in this section, and its only additional implication is that the asymptotic size is exact along local parameter sequences drifting sufficiently slowly: for $\epsilon_n/|\theta_0| \to 0$, $\lim_{n\to\infty} P\left(r_n\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right) \ge \hat{c}_{1-\tau}\right) = \tau$.

3.3. Dealing with nuisance parameters

Unlike conventional derivatives, directional derivatives are not generally linearly separable in different subsets of parameters unless more assumptions are made. Consider now $\phi(\theta,\alpha)$ where α are a set of nuisance parameters. In addition to requiring that $\phi(\cdot,\cdot)$ be jointly Hadamard directionally differentiable in θ,α tangentially to $\mathbb{D}_0 = \left(\mathbb{D}_{0,\theta},\mathbb{D}_{0,\alpha}\right)$, we impose the following assumption of separability and partial linearity in α:

Assumption 3.6. $\mathbb{D}_{0,\alpha}$ is convex and $\phi'_{\theta,\alpha}\left(h_\theta, h_\alpha^1 + h_\alpha^2\right) = \phi'_{\theta,\alpha}\left(h_\theta, h_\alpha^1\right) + \phi'_{\theta,\alpha}\left(0, h_\alpha^2\right)$.

This assumption holds for example in Hansen (2017) when θ is the threshold parameter and α are the regression coefficients. Under Assumption 3.6, while (5) can be used to estimate $\phi'_{\theta,\alpha}(h_\theta,h_\alpha)$ jointly in θ,α, it is also possible to estimate $\phi'_{\theta,\alpha}(h_\theta,0)$ and $\phi'_{\theta,\alpha}(0,h_\alpha)$ separately, using the numerical delta method and the bootstrap respectively. For $r_n\epsilon_n \to \infty$,

$$\begin{aligned}\hat{\phi}'_n(h_\theta,0) &= \frac{\phi\left(\hat{\theta}_n + \epsilon_n h_\theta, \hat{\alpha}_n\right) - \phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)}{\epsilon_n}\\ \hat{\phi}'_n(0,h_\alpha) &= r_n\left(\phi\left(\hat{\theta}_n, \hat{\alpha}_n + r_n^{-1}h_\alpha\right) - \phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)\right).\end{aligned} \tag{18}$$

Then (5) can be replaced by, with $\mathbb{Z}_n^* = \left(\mathbb{Z}_{n,\theta}^*, \mathbb{Z}_{n,\alpha}^*\right)$, $\hat{\phi}'_n\left(\mathbb{Z}_n^*\right) \equiv \hat{\phi}'_n\left(\mathbb{Z}_{n,\theta}^*,0\right) + \hat{\phi}'_n\left(0,\mathbb{Z}_{n,\alpha}^*\right)$. In particular, when $\mathbb{Z}_{n,\theta}^* = r_n\left(\hat{\theta}_n^* - \theta_0\right)$ and $\mathbb{Z}_{n,\alpha}^* = r_n\left(\hat{\alpha}_n^* - \alpha_0\right)$, the distribution of $r_n\left(\phi\left(\hat{\theta}_n,\hat{\alpha}_n\right) - \phi(\theta_0,\alpha_0)\right)$ is approximated by $\frac{1}{\epsilon_n}\left(\phi\left(\hat{\theta}_n + \epsilon_n r_n\left(\hat{\theta}_n^* - \hat{\theta}_n\right), \hat{\alpha}_n\right) - \phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)\right) + r_n\left(\phi\left(\hat{\theta}_n,\hat{\alpha}_n^*\right) - \phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)\right)$.

The Fang and Santos (2014) assumptions (2.1, 2.2, 2.3, 3.1, 3.2 and 3.3) are implicitly understood to hold jointly in θ,α in the rest of this section.

Theorem 3.7. *The result of Theorem 3.3 holds with (18) under Assumption 3.6.*

A special case of Assumption 3.6 is when estimating α does not affect the asymptotic distribution, as in for example the weighting matrix in moment inequality models (e.g., Andrews and Soares, 2010).

Assumption 3.7. $\phi'_{\theta,\alpha}(h_\theta,h_\alpha) = \phi'_{\theta,\alpha}(h_\theta,0)$ for all $h = (h_\theta,h_\alpha)$.

Under Assumption 3.7, it is natural to estimate $\phi'_{\theta,\alpha}(h)$ by $\hat{\phi}'_n(h_\theta,0)$, and replace $\hat{\phi}'_n\left(\mathbb{Z}_n^*\right)$ in (5) with

$$\hat{\phi}'_n\left(\mathbb{Z}_{n,\theta}^*,0\right) = \frac{\phi\left(\hat{\theta}_n + \epsilon_n\mathbb{Z}_{n,\theta}^*, \hat{\alpha}_n\right) - \phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)}{\epsilon_n}$$

Pointwise consistency of $\hat{\phi}'_n(h_\theta,0)$ for $\phi'_{\theta,\alpha}(h_\theta,0)$ follows directly from Theorem 3.3 with $h = (h_\theta,0)$. Furthermore, $\hat{\phi}'_n(h_\theta,0)$ is Lipschitz in h_θ as long as $\phi(\theta,\alpha)$ is Lipschitz in θ uniformly in α: $\left\|\hat{\phi}'_n(h_1,0) - \hat{\phi}'_n(h_2,0)\right\|_{\mathbb{E}} = \left\|\frac{\phi(\hat{\theta}_n+\epsilon_n h_1,\hat{\alpha}_n) - \phi(\hat{\theta}_n+\epsilon_n h_2,\hat{\alpha}_n)}{\epsilon_n}\right\|_{\mathbb{E}} \le C\left\|h_1 - h_2\right\|_{\mathbb{D}}$.

Under Assumption 3.7, we also obtain uniform size control with $\phi(\theta,\alpha)$ for (13) and (14), whenever $\phi(\theta,\alpha)$ is convex in θ for each α. In this case, analogous to (15), for any realization z from $\mathbb{G}_{0,\theta}$, where $\mathbb{Z}_{n,\theta}^* \overset{\mathbb{P}}{\rightsquigarrow} \mathbb{G}_{0,\theta}$,

$$r_n\left(\phi\left(\theta_0 + \frac{z}{r_n},\alpha_0\right) - \phi(\theta_0,\alpha_0)\right) \le \frac{1}{\epsilon_n}\left(\phi(\theta_0+\epsilon_n z,\alpha_0) - \phi(\theta_0,\alpha_0)\right), \tag{19}$$

so that $\frac{1}{\epsilon_n}\left(\phi\left(\theta_0 + \epsilon_n\mathbb{G}_{0,\theta},\alpha_0\right) - \phi(\theta_0,\alpha_0)\right)$ stochastically dominates $r_n\left(\phi\left(\theta_0 + \frac{\mathbb{G}_{0,\theta}}{r_n},\alpha_0\right) - \phi(\theta_0,\alpha_0)\right)$. Directional differentiability and Assumption 3.7 ensure that $r_n\left(\phi\left(\theta_0 + \frac{\mathbb{G}_{0,\theta}}{r_n},\alpha_0\right) - \phi(\theta_0,\alpha_0)\right)$ is close to $r_n\left(\phi\left(\hat{\theta}_n,\hat{\alpha}_n\right) - \phi(\theta_0,\alpha_0)\right)$ while

$\frac{1}{\epsilon_n}\left(\phi\left(\theta_0+\epsilon_n\mathbb{G}_{0,\theta},\alpha_0\right)-\phi\left(\theta_0,\alpha_0\right)\right)$ is close to $\hat{\phi}_n'\left(\mathbb{Z}_{n,\theta}^*,0\right)$. Formally, under Assumptions 3.7, 3.3 and 3.4 are only required to hold in θ_0:

Assumption 3.8. For all θ_0,α_0, and $t>0$, $\frac{\phi(\theta_0+t\mathbb{G}_{0,\theta},\alpha_0)-\phi(\theta_0,\alpha_0)}{t}$ is nondecreasing in t.

Assumption 3.9. Suppose $\mathbb{D}_{0,\theta}$ is convex. For any θ_0 and α_0, for all η and ν sufficiently close to zero, and for all $t>0$, $\frac{\phi_{\theta_0,\alpha_0}'(\eta+t\mathbb{G}_{0,\theta},\nu)-\phi_{\theta_0,\alpha_0}'(\eta,\nu)}{t}$ is nondecreasing in t. Furthermore, Assumption 3.5 holds with θ_0,α_0 and for any $h=(h_\theta,h_\alpha)=o(1)$.

Then we can state the following theorem.

Theorem 3.8. *The conclusions of Theorem 3.5 hold under its stated conditions and Assumption 3.7, where we now call $J_{\epsilon_n}(x_n,P_n)$ the distribution function of $\hat{\phi}_n'\left(\mathbb{Z}_{n,\theta}^*,0\right)$, and $J_n(x_n,P_n)$ that of $r_n\left(\phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)-\phi\left(\theta_0,\alpha_0\right)\right)$. Furthermore, the conclusions of Theorem 3.6 hold under its stated conditions and Assumption 3.7, when $\hat{c}_{1-\tau}$ refers to the $(1-\tau)$th percentile of the conditional distribution of $\hat{\phi}_n'\left(\mathbb{Z}_{n,\theta}^*,0\right)$ given the data, and if for any $\theta_0\in\Theta$, either Assumption 3.8 or Assumption 3.9 holds.*

While we have required $r_n\left(\hat{\alpha}_n-\alpha_0\right)=O_p(1)$, in many applications the weaker condition $\hat{\alpha}_n\xrightarrow{p}\alpha_0$ suffices, such as for the variance in a t-statistic and the weighting matrix for moment conditions. However, in these problems $r_n\left(\hat{\alpha}_n-\alpha_0\right)=O_p(1)$ always holds under stronger regularity conditions.

When $\phi(\cdot,\cdot)$ is fully Hadamard differentiable, Assumption 3.6 holds with

$$\phi_{\theta,\alpha}'(h_\theta,h_\alpha)=\frac{\partial}{\partial\theta}\phi_{\theta,\alpha}(h_\theta,0)+\frac{\partial}{\partial\alpha}\phi_{\theta,\alpha}(0,h_\alpha).$$

In this case the bootstrap can approximate the distribution of $r_n\left(\phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)-\phi\left(\theta_0,\alpha_0\right)\right)$ by that of $r_n\left(\phi\left(\hat{\theta}_n+r_n^{-1}\mathbb{Z}_{n,\theta}^*,\hat{\alpha}_n+r_n^{-1}\mathbb{Z}_{n,\alpha}^*\right)-\phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)\right)$, or by that of

$$r_n\left(\phi\left(\hat{\theta}_n+r_n^{-1}\mathbb{Z}_{n,\theta}^*,\hat{\alpha}_n\right)-\phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)+\phi\left(\hat{\theta}_n,\hat{\alpha}_n+r_n^{-1}\mathbb{Z}_{n,\alpha}^*\right)-\phi\left(\hat{\theta}_n,\hat{\alpha}_n\right)\right).$$

In particular, if $\phi(\cdot)$ is a model parameter itself (now denoted θ), and if θ denotes the underlying distribution (now denoted P), then the distribution of $\hat{\theta}_n-\theta_0=\theta\left(P_n,\hat{\alpha}_n\right)-\theta\left(P,\alpha_0\right)$ can be approximated by $\theta\left(P_n^*,\hat{\alpha}_n^*\right)-\theta\left(P_n,\hat{\alpha}_n\right)$, where P_n^* is the bootstrap data set and $\hat{\alpha}_n^*$ is computed on the same bootstrap data set. In some situations, if α is computed from an independent data set such that $\hat{\alpha}_n\sim N\left(\alpha,\hat{\Omega}\right)$, then $\hat{\alpha}_n^*$ can be drawn from $N\left(\hat{\alpha}_n,\hat{\Omega}\right)$. In this case an alternative approximation is $\theta\left(P_n^*,\hat{\alpha}_n\right)-\theta\left(P_n,\hat{\alpha}_n\right)+\theta\left(P_n,\hat{\alpha}_n^*\right)-\theta\left(P_n,\hat{\alpha}_n\right)$ where $\theta\left(P_n^*,\hat{\alpha}_n\right)-\theta\left(P_n,\hat{\alpha}_n\right)$ can also be replaced by any approximate distribution of $\hat{\theta}_n$ treating $\hat{\alpha}_n$ as known.

3.4. Application to partially identified models: the $\mathcal{L}_1$ version

As an application, we relate the numerical delta method to a $\mathcal{L}_1$ version of the partially identified model studied by Andrews and Soares (2010). While the current partial identification literature chooses to work with $S(x,\Sigma)=\sum_{k=1}^K\left(x_k^-\right)^2$, an alternative is to choose $S(\cdot)$ to be a L_p norm. For example, we may choose $S(x)=\min_{h\in\Lambda=R_+^k}||x-h||_p=\left(\sum_{i=1}^k\left(x_i^-\right)^p\right)^{1/p}$. For $p=2$ and when a weighting matrix W is employed,

$$S(x,W)=\min_{h\in\Lambda=R_+^k}\sqrt{(x-h)'W(x-h)}.$$

A consistent estimate $\hat{W}$ of the weighting matrix W is often available, and can be treated as a nuisance parameter that does not affect the asymptotic distribution in the sense of Assumption 3.7.

If such a L_p norm is used instead in Andrews and Soares (2010), then $S(\cdot)$ is convex and Theorem 3.5 can be applied. On the one hand, whether to take the $1/p$ root makes no difference in a point identified model since optimization is invariant to monotonic transformations. On the other hand, it implies a different directional derivative, and does make a difference in set identified models and GMS methods.

Suppose we are testing $H_0:\theta_0\geq 0$ using the sample mean $\hat{\theta}_n$. Let us consider the case of $p=2$ and a single moment equality. If we do not take the square root, we reject whenever $n\left(\hat{\theta}_n^-\right)^2$ is greater than the $(1-\alpha)$ th percentile of $\left(\left(\frac{\hat{\theta}_n}{\epsilon_n}+\mathbb{Z}_n^*\right)^-\right)^2-\left(\left(\frac{\hat{\theta}_n}{\epsilon_n}\right)^-\right)^2$, where $\mathbb{Z}_n^*$ is a normal random variable. However, if we take the square root, we reject whenever $\sqrt{n}\left(\hat{\theta}_n^-\right)$ is greater than the $(1-\alpha)$ th percentile of $\left(\frac{\hat{\theta}_n}{\epsilon_n}+\mathbb{Z}_n^*\right)^--\left(\frac{\hat{\theta}_n}{\epsilon_n}\right)^-$. The transformation for the critical values is not the same as the transformation for the test statistic, and therefore the resulting rejection areas will be different.

4. Second order numerical directional delta method

In situations in which the first order delta method limiting distribution is degenerate, the second (or higher) order delta method may provide the necessary nondegenerate large sample approximation. For example, Andrews and Soares (2010) conducts inference using $\phi(\theta)=\sum_{k=1}^{K}\left(\theta_k^-\right)^2$, which has a first order directional derivative of $\phi_\theta'(h)=-\sum_{k=1}^{K}2\theta_k^- h_k$. Under the null hypothesis of $inf_{k=1\ldots K}\theta_k\geq 0$, $\phi_\theta'(h)=0$, which leads to a degenerate first order delta method limiting distribution.

We will maintain the assumption that $\phi(\cdot)$ is first order Hadamard differentiable at θ_0. The second order Hadamard directional derivative at θ_0 in the direction h tangential to $\mathbb{D}_0\subseteq\mathbb{D}$ is defined as

$$\phi_{\theta_0}''(h)\equiv\lim_{t_n\downarrow 0,h_n\to h\in\mathbb{D}_0}\frac{\phi(\theta_0+t_nh_n)-\phi(\theta_0)-t_n\phi_{\theta_0}'(h_n)}{\frac{1}{2}t_n^2}\tag{20}$$

Sufficient conditions for the existence of $\phi_{\theta_0}''(h)$ are that $\phi(\theta)$ is Hadamard differentiable uniformly in θ around some neighborhood of θ_0 and that $\phi_\theta'(h)$ is directionally differentiable in θ at θ_0. Although the definition of the second order directional derivative contains only one direction h, in principle we can use different directions h_1 and h_2. For $g\left(t_n,h_n^1,h_n^2\right)=t_n^{-1}\left(\phi_{\theta_0+t_nh_n^1}'\left(h_n^2\right)-\phi_{\theta_0}'\left(h_n^2\right)\right)$, $\lim_{t_n\downarrow 0,\left(h_n^1,h_n^2\right)\to(h_1,h_2)}g\left(t_n,h_n^1,h_n^2\right)=\phi_{\theta_0}''(h_1,h_2)$ for $h_1\in\mathbb{D}_0$, $h_2\in\mathbb{D}_0$. In this paper, if there is only one argument in the $\phi_{\theta_0}''(\cdot)$ function, then we are assuming that $h_1=h_2$.

Note that $\phi_{\theta_0}''(h)$ is continuous with respect to $h\in\mathbb{D}_0$, and it is also positively homogeneous of degree 2: $\phi_{\theta_0}''(ch)=c^2\phi_{\theta_0}''(h)$ for all $c\geq 0$ and $h\in\mathbb{D}_0$. A simple illustrative example is $\phi(\theta)=\left(\theta^-\right)^2$. For this function, the first order directional derivative is $\phi_\theta'(h)=-2\theta^-h$, which is identically zero for $\theta\geq 0$. The second order directional derivative is $\phi_{\theta_0}''(h)=2\left(h^-\right)^2 1\left(\theta_0=0\right)+2h^2 1\left(\theta_0<0\right)$.

The first part of the following theorem is due to Römisch (2005) and Shapiro (2000); in the second part we incorporate the numerical directional derivative.[5]

Theorem 4.1 (*Second Order Directional Delta Method*). *Suppose $\mathbb{D}$ and $\mathbb{E}$ are Banach Spaces and $\phi:\mathbb{D}_\phi\subseteq\mathbb{D}\mapsto\mathbb{E}$ is second order Hadamard directionally differentiable at θ_0 tangentially to $\mathbb{D}_0$. Let $\hat\theta_n:\{X_i\}_{i=1}^n\mapsto\mathbb{D}_\phi$ be such that for some $r_n\uparrow\infty$, $r_n\{\hat\theta_n-\theta_0\}\rightsquigarrow\mathbb{G}_0$ in $\mathbb{D}$ and assume the support of $\mathbb{G}_0$ is included in $\mathbb{D}_0$. Then,*

$$r_n^2\left[\phi(\hat\theta_n)-\phi(\theta_0)-\phi_{\theta_0}'(\hat\theta_n-\theta_0)\right]\rightsquigarrow\mathcal{J}\equiv\frac{1}{2}\phi_{\theta_0}''(\mathbb{G}_0)\tag{21}$$

Let $\epsilon_n\to 0$, $r_n\epsilon_n\to\infty$, and $\mathbb{Z}_n^*\overset{\mathbb{P}}{\rightsquigarrow}\mathbb{G}_0$. Then if $\phi_{\theta_0}'(h)\equiv 0\ \forall h\in\mathbb{D}_0$,

$$\frac{\phi\left(\hat\theta_n+\epsilon_n\mathbb{Z}_n^*\right)-\phi\left(\hat\theta_n\right)}{\epsilon_n^2}\overset{\mathbb{P}}{\rightsquigarrow}\mathcal{J}\equiv\frac{1}{2}\phi_{\theta_0}''(\mathbb{G}_0).\tag{22}$$

Pointwise asymptotic validity of the numerical directional delta method is justified by (22). There are several alternatives for approximating $\frac{1}{2}\phi_{\theta_0}''(\mathbb{G}_0)$. First, the left hand side of (22) can be replaced by $\hat\phi_n''\left(\mathbb{Z}_n^*\right)$ where the second order directional derivative can be estimated by

$$\hat\phi_n''(h)\equiv\frac{\phi(\hat\theta_n+2\epsilon_nh)-2\phi(\hat\theta_n+\epsilon_nh)+\phi(\hat\theta_n)}{\epsilon_n^2}\tag{23}$$

Theorem 4.2. *Under convexity of $\mathbb{D}_0$ and the same conditions as in Theorem 4.1, except without $\phi_{\theta_0}'(h)\equiv 0$, for $\hat\phi_n''(h)$ in (23), $\hat\phi_n''\left(\mathbb{Z}_n^*\right)\overset{\mathbb{P}}{\rightsquigarrow}\phi_{\theta_0}''(\mathbb{G}_0)$.*

If the first derivative $\phi_\theta'(h)$ is analytically known, as in Andrews and Soares (2010), another alternative is to estimate the second order directional derivative (21) by

$$\bar\phi_n''(h_1,h_2)\equiv\frac{\phi_{\hat\theta_n+\epsilon_nh_1}'(h_2)-\phi_{\hat\theta_n}'(h_2)}{\epsilon_n}\tag{24}$$

Theorem 4.3. *For $\bar\phi_n''(h,h)$ defined in (24), $\bar\phi_n''(\mathbb{Z}_n^*,\mathbb{Z}_n^*)\overset{\mathbb{P}}{\rightsquigarrow}\phi_{\theta_0}''(\mathbb{G}_0)$.*

We can show that $\bar\phi_n''(h,h)=\frac{\phi_{\hat\theta_n+\epsilon_nh}'(h)-\phi_{\hat\theta_n}'(h)}{\epsilon_n}$ is Lipschitz whenever $\phi_\theta'(h)$ is.

Theorem 4.4. *If $\phi_\theta'(h):\mathbb{D}_\phi\to\mathbb{E}$ is Lipschitz in θ and h, then for all $\epsilon_n\downarrow 0$, $\bar\phi_n''(h,h)=\frac{\phi_{\hat\theta_n+\epsilon_nh}'(h)-\phi_{\hat\theta_n}'(h)}{\epsilon_n}$ is Lipschitz in h.*

[5] Recent independent work by Chen and Fang (2015) also studies inference under first order degeneracy.

Theorem 4.1 applies when $\phi'_{\theta_0}(h) \equiv 0$, in which case $r_n^2\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_0)\right) \rightsquigarrow \mathcal{J}$. By Theorems 4.1–4.3, $\frac{\phi\left(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*\right)-\phi\left(\hat{\theta}_n\right)}{\epsilon_n^2}$ in (22), $\hat{\phi}''_n\left(\mathbb{Z}_n^*\right)$ in (23) and $\bar{\phi}''_n\left(\mathbb{Z}_n^*, \mathbb{Z}_n^*\right)$ in (24) converge to the same limiting distribution $\mathcal{J} = \frac{1}{2}\phi''_{\theta_0}(\mathbb{G}_0)$ under fixed θ_0 asymptotics and under a local drifting sequence of parameters θ_n where $r_n(\theta_n - \theta_0) \to c$ for $||c|| < \infty$. In the latter case, let $\mathbb{Z}_n = r_n\left(\hat{\theta}_n - \theta_n\right) \rightsquigarrow \mathbb{G}_0$. Then $r_n^2\left(\phi(\hat{\theta}_n) - \phi(\theta_n)\right)$ satisfies

$$r_n^2\left(\phi\left(\frac{1}{r_n}\left(r_n(\theta_n-\theta_0)+\mathbb{Z}_n\right)\right)-\phi(\theta_0)\right) - r_n^2\left(\phi\left(\frac{1}{r_n}\left(r_n(\theta_n-\theta_0)\right)\right)-\phi(\theta_0)\right) \rightsquigarrow \frac{1}{2}\phi''_{\theta_0}(c+\mathbb{G}_0) - \frac{1}{2}\phi''_{\theta_0}(c).$$

The equalities follow from $r_n(\theta_n-\theta_0)+\mathbb{Z}_n \rightsquigarrow c+\mathbb{G}_0$, $r_n(\theta_n-\theta_0) \rightsquigarrow c$, and the definition of the second order delta method.

The behaviors of $\hat{\phi}''_n(\mathbb{Z}_n^*)$, $\bar{\phi}''_n(\mathbb{Z}_n^*, \mathbb{Z}_n^*)$ and $\frac{\phi\left(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*\right)-\phi\left(\hat{\theta}_n\right)}{\epsilon_n^2}$ differ under a more distant local drifting sequence of parameters $\frac{\theta_n-\theta_0}{\epsilon_n} \to c$, when $0 < ||c|| < \infty$, which implies different finite sample behaviors.

On the one hand, $\frac{1}{\epsilon_n^2}\left(\phi(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*) - \phi(\hat{\theta}_n)\right) \rightsquigarrow \frac{1}{2}\phi''_{\theta_0}(c+\mathbb{G}_0) - \frac{1}{2}\phi''_{\theta_0}(c)$. On the other hand, for (23),

$$\begin{aligned}\frac{1}{2}\hat{\phi}''_n(\mathbb{Z}_n^*) =& \frac{1}{2}\frac{1}{\epsilon_n^2}\left[\phi\left(\epsilon_n\left(\frac{\theta_n-\theta_0}{\epsilon_n}+\frac{\mathbb{Z}_n}{r_n\epsilon_n}+2\mathbb{Z}_n^*\right)\right)-\phi(\theta_0)\right] - \frac{1}{\epsilon_n^2}\left[\phi\left(\epsilon_n\left(\frac{\theta_n-\theta_0}{\epsilon_n}+\frac{\mathbb{Z}_n}{r_n\epsilon_n}+\mathbb{Z}_n^*\right)\right)-\phi(\theta_0)\right]\\ &+\frac{1}{2}\frac{1}{\epsilon_n^2}\left[\phi\left(\epsilon_n\left(\frac{\theta_n-\theta_0}{\epsilon_n}+\frac{\mathbb{Z}_n}{r_n\epsilon_n}\right)\right)-\phi(\theta_0)\right] \rightsquigarrow \frac{1}{4}\phi''_{\theta_0}(c+2\mathbb{G}_0) - \frac{1}{2}\phi''_{\theta_0}(c+\mathbb{G}_0) + \frac{1}{4}\phi''_{\theta_0}(c).\end{aligned}$$

It can also be shown that for (24),

$$\frac{1}{2}\bar{\phi}''_n(\mathbb{Z}_n^*, \mathbb{Z}_n^*) \equiv \frac{\phi'_{\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*}\left(\mathbb{Z}_n^*\right) - \phi'_{\hat{\theta}_n}\left(\mathbb{Z}_n^*\right)}{2\epsilon_n} \rightsquigarrow \frac{1}{2}\phi''_{\theta_0}(c+\mathbb{G}_0, \mathbb{G}_0) - \frac{1}{2}\phi''_{\theta_0}(c, \mathbb{G}_0).$$

The differences between various methods of estimating the second order derivative when $\frac{\theta_n-\theta_0}{\epsilon_n} \to c$ can be illustrated using a simple test of $H_0 : \theta_0 \geq 0$ against $H_1 : \theta_0 < 0$, which is converted to $H_0 : \phi(\theta_0) = 0$ against $H_1 : \phi(\theta_0) > 0$ using the test function $\phi(\theta) = \left(\theta^-\right)^2$, which has $\phi'_\theta(h) = -2\theta^- h$ and $\phi''_\theta(h) = 2\left(h^-\right)^2 1(\theta = 0) + 2h^2 1(\theta < 0)$. Consider a level α test with rejection region $\{r_n^2\phi(\hat{\theta}_n) \geq d_{1-\alpha}\}$, where $d_{1-\alpha}$ is the $1-\alpha$ percentile of one of the following four distributions: (1) $\frac{1}{\epsilon_n^2}\phi(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*)$; (2) $\frac{1}{2}\frac{1}{\epsilon_n}\left(\phi'_{\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*}(\mathbb{Z}_n^*) - \phi'_{\hat{\theta}_n}(\mathbb{Z}_n^*)\right)$; (3) $\frac{1}{\epsilon_n^2}\left(\phi(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*) - \phi(\hat{\theta}_n)\right)$; (4) $\frac{1}{2}\hat{\phi}''_n(\mathbb{Z}_n^*)$. Let $\theta_0 = 0$ and $\frac{\theta_n}{\epsilon_n} \to c$. The corresponding limiting distributions are

(1) $\frac{1}{2}\phi''_0(c+\mathbb{G}_0) = \left((\mathbb{G}_0+c)^-\right)^2$
(2) $\frac{1}{2}\left(\phi'_{\mathbb{G}_0+c}(\mathbb{G}_0) - \phi'_c(\mathbb{G}_0)\right) = -(\mathbb{G}_0+c)^-\mathbb{G}_0 + c^-\mathbb{G}_0$
(3) $\frac{1}{2}\phi''_0(c+\mathbb{G}_0) - \frac{1}{2}\phi''_0(c) = \left((\mathbb{G}_0+c)^-\right)^2 - \left(c^-\right)^2$
(4) $\frac{1}{4}\phi''_0(c+2\mathbb{G}_0) - \frac{1}{2}\phi''_0(c+\mathbb{G}_0) + \frac{1}{4}\phi''_0(c) = \frac{1}{2}\left((2\mathbb{G}_0+c)^-\right)^2 - \left(\mathbb{G}_0^-\right)^2 + \frac{1}{2}\left(c^-\right)^2$

First consider the case of $c > 0$, which corresponds to size control. In this case it is not difficult to see that (4) $\succeq$ (2) $\succeq$ (1) $=$ (3) in descending order of first order stochastic dominance. Furthermore, (1) through (4) all stochastically dominate the distribution of the test statistic under the null of $\theta_0 > 0$, which is $\lim_{h\to\infty}\frac{1}{2}\phi''_{\theta_0}(h+\mathbb{G}_0) - \frac{1}{2}\phi''_{\theta_0}(h) = 0$ because $r_n(\theta_n - \theta_0) \to \infty$ when $\frac{\theta_n-\theta_0}{\epsilon_n} \to c$. By imposing a zero first order derivative under the null, (2) and (4) provide better finite sample size control. However, comparing the finite sample powers of these tests when $\frac{\theta_n}{\epsilon_n} \to c < 0$ does not give a conclusive ranking. While it is clear that the recentered version (3) is always more powerful than the nonrecentered version (1), there does not seem to be a uniform ranking among (2)–(4). The ranking might depend on the range of the alternative hypothesis.

5. Monte Carlo simulations

In this section we report two finite sample simulations. The first uses a simple parametric example to show consistency of the first order numerical delta method, while the second applies the second order numerical delta method to the moment inequalities setup in Andrews and Soares (2010).

5.1. Confidence intervals in a basic model

Consider a simple set up of i.i.d data $X_i \overset{iid}{\sim} N(\theta_n, 1)$ and $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i \equiv \bar{X}$. The function of interest is $\phi(\theta) = a\theta^+ + b\theta^-$, where $\theta^+ = \max\{\theta, 0\}$ and $\theta^- = -\min\{\theta, 0\}$. Functions of this type appear in Hansen (2017)'s continuous threshold regression model and in moment inequality inference models. We approximate the distribution of $r_n(\phi(\hat{\theta}_n) - \phi(\theta_n))$ using $\hat{\phi}'_n\left(\mathbb{Z}_n^*\right) = \frac{\phi\left(\hat{\theta}+\epsilon_n\mathbb{Z}_n^*\right)-\phi\left(\hat{\theta}\right)}{\epsilon_n}$. where $\mathbb{Z}_n^* \overset{\mathbb{P}}{\rightsquigarrow} \mathbb{G}_0$ and $r_n(\hat{\theta}_n - \theta_n) \rightsquigarrow \mathbb{G}_0$. We use $\mathbb{Z}_n^* = N(0, \hat{\sigma})$, where $\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^n(X_i - \bar{X})^2}$.

For c_α denoting the α quantile of $\hat{\phi}_n'(\mathbb{Z}_n^*)$ and d_α denoting the α quantile of $|\hat{\phi}_n'(\mathbb{Z}_n^*)|$, we report (1) a symmetric two sided interval $\left[\phi(\hat{\theta}_n) - \frac{1}{r_n} d_{1-\alpha}, \phi(\hat{\theta}_n) + \frac{1}{r_n} d_{1-\alpha}\right]$; (2) an equal-tailed two-sided interval $\left[\phi(\hat{\theta}_n) - \frac{1}{r_n} c_{1-\alpha/2}, \phi(\hat{\theta}_n) - \frac{1}{r_n} c_{\alpha/2}\right]$; (3) an upper one-sided confidence interval $\left(-\infty, \phi(\hat{\theta}_n) - \frac{1}{r_n} c_\alpha\right]$; (4) a lower one-sided confidence interval $\left[\phi(\hat{\theta}_n) - \frac{1}{r_n} c_{1-\alpha}, \infty\right)$.

For $a > 0, b > 0$ or $a > 0, b < 0, a > |b|$ or $a < 0, b > 0, |a| < b$, $\phi(\theta)$ is a convex function of θ. Then Theorem 3.5 implies that the lower one-sided interval is uniformly valid at least conservatively. Both the upper one-sided interval and as a result the equal-tailed two sided interval are only valid under fixed asymptotics, but can undercover for local drifting parameter sequences between orders of $1/\sqrt{n}$ and ϵ_n.

Analogously, for $a < 0, b < 0$ or $a < 0, b > 0, a < |b|$ or $a > 0, b < 0, |a| > b$, $\phi(\theta)$ is a concave function of θ. Then Theorem 3.5 implies that the upper one-sided interval is uniformly valid at least conservatively. Both the lower one-sided interval and as a result the equal-tailed two sided interval are only valid under fixed asymptotics, but can undercover for local drifting parameter sequences between orders of $1/\sqrt{n}$ and ϵ_n.

For the two sided symmetric interval, note that in this model, the directional derivative $\phi_\theta'(h)$ is given by (1) ah if $\theta > 0$; (2) $-bh$ if $\theta < 0$; (3) $ah^+ + bh^-$ if $\theta = 0$. It satisfies the condition that

$$|\phi_\theta'(h_1 + h_2) - \phi_\theta'(h_2)| \leq |\phi_\theta'(h_1)|, \tag{25}$$

Note that $|\phi_\theta'(\mathbb{G}_0 + c) - \phi_\theta'(c)|$ and $|\phi_\theta'(\mathbb{G}_0)|$ are, respectively, the analytic limit and numerical delta method limit under the Fang and Santos (2014) local sequence $\theta_n = c/\sqrt{n}$. Therefore (25) implies that the symmetric two sided interval is at least conservatively valid under the local sequence of $\theta_n = c/\sqrt{n}$. The two sided symmetric interval may undercover, however, for the local parameter sequence of $\theta_n = c\epsilon_n$. In other words, when $\sqrt{n}\epsilon_n \to \infty$, neither the symmetric nor the equal-tailed two sided intervals are uniformly valid, but the symmetric interval is valid for a wider range of local parameter sequences than the equal-tailed interval.

The set of tables titled "Monte Carlo Simulations for the Normal Mean Model" show empirical coverage frequencies for $a = 1.5, b = 0.5$, which corresponds to convex $\phi(\theta)$. Results for concave $\phi(\theta)$ are analogous and omitted for brevity. Empirical coverage frequencies are computed for four different values of ϵ_n: $n^{-1/6}, n^{-1/3}, n^{-1/2}, n^{-1}$; and eleven different values of θ_n: $-2, -n^{-1/6}, -n^{-1/3}, 0, n^{-1}, n^{-1/1.5}, n^{-1/2}, n^{-1/3}, n^{-1/6}, n^{-1/10}$, and 2. The empirical coverage frequencies for the four different kinds of confidence intervals (symmetric two-sided, equal-tailed two-sided, upper one-sided, and lower one-sided) when $\epsilon_n = n^{-1/6}, \epsilon_n = n^{-1/3}, \epsilon_n = n^{-1/2}$, and $\epsilon_n = n^{-1}$ are summarized in tables 1 through 4, tables 5 through 8, tables 9 through 12, and tables 13 through 16 respectively. The nominal coverage frequency is 95%.

When $\sqrt{n}\epsilon_n \to \infty$, the symmetric two-sided confidence intervals have an empirical coverage frequency close to the nominal frequency in the regions $\theta_n \in \{0, n^{-1}, n^{-1/1.5}, n^{-1/2}\}$ and $\frac{\theta_n}{\epsilon_n} \to \pm\infty$. The empirical coverage frequency is below the nominal frequency when $\frac{\theta_n}{\epsilon_n} \to c$ for $0 < c < \infty$. The equal-tailed two-sided confidence intervals have an empirical coverage frequency close to the nominal frequency in the regions $\theta_n \in \{0, n^{-1}\}$ and $\frac{\theta_n}{\epsilon_n} \to \pm\infty$. In the region where $\theta_n\sqrt{n} \to c_1$ for $0 < |c_1| \leq \infty$ and $\frac{\theta_n}{\epsilon_n} \to c_2$ for $0 \leq |c_2| < \infty$, the empirical coverage frequency is far below the nominal frequency.

When $\sqrt{n}\epsilon_n \to \infty$, the lower one-sided confidence intervals provide conservatively valid coverage for all values of θ_n, which is to be expected given the theoretical results. On the other hand, the upper one-sided confidence intervals undercover for values of θ_n that satisfy $\theta_n\sqrt{n} \to c_1$ for $|c_1| > 0$ and $\frac{\theta_n}{\epsilon_n} \to c_2$ for $0 \leq |c_2| < \infty$ while providing coverage close to the nominal frequency for the other values of θ_n.

5.2. Small step size in the basic example

While the theory in the previous sections is provided for larger step sizes ($\sqrt{n}\epsilon_n \to \infty$), it turns out that in the example above a small step size might also be a possible choice for constructing confidence intervals in some situations. In this section we let $\sqrt{n}\epsilon_n \to 0$ and examine the consequences for the numerical delta method. Let $\mathbb{Z}_n = \sqrt{n}\left(\hat{\theta}_n - \theta_n\right)$ so that $\left(\mathbb{Z}_n^*, \mathbb{Z}_n\right) \rightsquigarrow (\mathbb{G}_1, \mathbb{G}_0)$, where $\mathbb{G}_1 \sim N(0, 1)$, $\mathbb{G}_0 \sim N(0, 1)$, $\mathbb{G}_1 \perp \mathbb{G}_0$. Also note that $\phi(\theta) = a\theta^+ + b\theta^-$ is homogeneous of degree one. We can write down the following heuristic calculations.

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*\right) = \phi\left(\frac{\hat{\theta}_n}{\epsilon_n} + \mathbb{Z}_n^*\right) - \phi\left(\frac{\hat{\theta}_n}{\epsilon_n}\right) = \phi\left(\frac{\mathbb{Z}_n}{\sqrt{n}\epsilon_n} + \frac{\theta_n}{\epsilon_n} + \mathbb{Z}_n^*\right) - \phi\left(\frac{\mathbb{Z}_n}{\sqrt{n}\epsilon_n} + \frac{\theta_n}{\epsilon_n}\right)$$

Also note that $\sqrt{n}\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_n)\right) = \phi\left(\mathbb{Z}_n + \sqrt{n}\theta_n\right) - \phi\left(\sqrt{n}\theta_n\right)$. We now consider three regimes separately.

Case 1:. If $\sqrt{n}\theta_n \to 0$, then $\sqrt{n}\left(\phi\left(\hat{\theta}_n\right) - \phi(\theta_n)\right) \rightsquigarrow a\mathbb{G}_0^+ + b\mathbb{G}_0^-$. Also,

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*\right) \rightsquigarrow W = \begin{cases} a\mathbb{G}_1^+ & \text{with probability} \quad P(\mathbb{G}_0 > 0) \\ -b\mathbb{G}_1^- & \text{with probability} \quad P(\mathbb{G}_0 < 0) \end{cases}$$

It can be verified that $|W|$ and $a\mathbb{G}_0^+ + b\mathbb{G}_0^-$ have the same distribution, so that two sided symmetric intervals are valid. By symmetry, so are the two sided equal-tailed intervals.

Case 2:. If $\sqrt{n}\theta_n = a_n \to \pm\infty$, both two sided intervals are valid since the analytic limit and the numeric limit have the same distribution:

$$\sqrt{n}\left(\phi\left(\hat{\theta}_n\right) - \phi\left(\theta_n\right)\right) \rightsquigarrow \begin{cases} a\mathbb{G}_0 & \text{if} \quad a_n > 0 \\ -b\mathbb{G}_0 & \text{if} \quad a_n < 0 \end{cases} \qquad \hat{\phi}_n'\left(\mathbb{Z}_n^*\right) \rightsquigarrow \begin{cases} a\mathbb{G}_1 & \text{if} \quad a_n > 0 \\ -b\mathbb{G}_1 & \text{if} \quad a_n < 0 \end{cases}$$

Case 3:. If $\sqrt{n}\theta_n \to c$, where $0 < |c| < \infty$, then the two distributions differ, and two sided intervals are generally invalid since

$$\sqrt{n}\left(\phi\left(\hat{\theta}_n\right) - \phi\left(\theta_n\right)\right) \rightsquigarrow a(c+\mathbb{G}_0)^+ + b(c+\mathbb{G}_0)^- - ac^+ - bc^-,$$

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*\right) \rightsquigarrow \begin{cases} a\mathbb{G}_1 & \text{with probability} \quad P\left(\mathbb{G}_0 > -c\right) \\ -b\mathbb{G}_1 & \text{with probability} \quad P\left(\mathbb{G}_0 < -c\right) \end{cases}$$

However, in a special case of case 3, when $a = b = 1$, the analytic limit becomes $|\mathbb{G}_0 + c| - |c|$ and the numeric limit becomes $\mathbb{G}_1$. Since $|\mathbb{G}_1|$ first order stochastically dominates $||\mathbb{G}_0 + c| - |c||$, symmetric two sided intervals are at least conservatively valid.

The knife-edge case of $\epsilon_n = n^{-1/2}$ corresponds essentially to the bootstrap. With the bootstrap,

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*\right) \rightsquigarrow \phi\left(\mathbb{G}_0 + \mathbb{G}_1 + \lim\sqrt{n}\theta_n\right) - \phi\left(\mathbb{G}_0 + \lim\sqrt{n}\theta_n\right)$$

Comparing this to $\sqrt{n}\left(\phi\left(\hat{\theta}_n\right) - \phi\left(\theta_n\right)\right) = \phi\left(\mathbb{Z}_n + \sqrt{n}\theta_n\right) - \phi\left(\sqrt{n}\theta_n\right) \rightsquigarrow \phi\left(\mathbb{G}_0 + \lim\sqrt{n}\theta_n\right) - \phi\left(\lim\sqrt{n}\theta_n\right)$ shows that when $\theta_n = 0$, the analytic limit is $\phi\left(\mathbb{G}_0\right)$ and the numerical limit is $\phi\left(\mathbb{G}_0 + \mathbb{G}_1\right)$ - $\phi\left(\mathbb{G}_0\right)$. Since $|\phi\left(\mathbb{G}_0\right)|$ first order stochastically dominates $|\phi\left(\mathbb{G}_0 + \mathbb{G}_1\right) - \phi\left(\mathbb{G}_0\right)|$, the bootstrap symmetric two-sided interval will undercover. However, when $\sqrt{n}|\theta_n|$ is larger (e.g. when $\sqrt{n}\theta_n \to \infty$), the bootstrap symmetric two-sided interval will not undercover.

5.3. Second order numerical derivative

The purpose of these Monte Carlo simulations is to investigate the power and size of moment inequality tests of the form $H_0 : inf_{j=1...J}\theta_{n,j} \geq 0$ and $H_1 : inf_{j=1...J}\theta_{n,j} < 0$. Let $\phi(\theta) = \sum_{j=1}^{J}\left(\theta_j^-\right)^2 = \sum_{j=1}^{J}(-min\{\theta_j, 0\})^2$ and $\phi_\theta'(h) = -\sum_{j=1}^{J}2\theta_j^- h_j$. Data are drawn from $X_i \overset{iid}{\sim} N(\theta_n, I_2)$ and $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}X_i \equiv \bar{X}$. We reject when $r_n^2\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}$, where $\hat{c}_{1-\alpha}$ is the $1-\alpha$ quantile of one of the following four ways of estimating the second order numerical derivative:

1. Andrews and Soares (2010) with 4th GMS function: $\frac{1}{\epsilon_n^2}\phi(\hat{\theta}_n + \epsilon_n\mathbb{Z}_n^*)$
2. Derivative of Analytic First Order Derivative: $\frac{1}{2}\frac{1}{\epsilon_n}\left(\phi_{\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*}'(\mathbb{Z}_n^*) - \phi_{\hat{\theta}_n}'(\mathbb{Z}_n^*)\right)$
3. Numerical Second Order Derivative 1: $\frac{1}{\epsilon_n^2}\left(\phi(\hat{\theta}_n + \epsilon_n\mathbb{Z}_n^*) - \phi(\hat{\theta}_n)\right)$
4. Numerical Second Order Derivative 2: $\frac{1}{2}\hat{\phi}_n''(\mathbb{Z}_n^*) = \frac{1}{2}\frac{\phi(\hat{\theta}_n+2\epsilon_n\mathbb{Z}_n^*)-2\phi(\hat{\theta}_n+\epsilon_n\mathbb{Z}_n^*)+\phi(\hat{\theta}_n)}{\epsilon_n^2}$

We take $\mathbb{Z}_n^* = N(0, \hat{\sigma})$, where $\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$. We use four different choices of ϵ_n: $\sqrt{\log(n)}/\sqrt{n}$, $n^{-1/6}$, $n^{-1/3}$, $n^{-1/2}$ and eleven different choices of θ_n: $-n^{-1/6}$, $-n^{-1/3}$, $-n^{-1/2}$, $-n^{-1/1.5}$, $-n^{-1}$, 0, n^{-1}, $n^{-1/1.5}$, $n^{-1/2}$, $n^{-1/3}$, and $n^{-1/6}$. The choice of $\epsilon_n = \sqrt{\log(n)}/\sqrt{n}$ is the one proposed by Andrews and Soares (2010). The set of tables titled "Monte Carlo Simulations for the Second Order Directional Delta Method" show the empirical rejection frequencies for the four different tests.

We can see that when $\epsilon_n = \frac{\sqrt{\log(n)}}{\sqrt{n}}$, the Andrews and Soares (2010) test has lower power than the other three tests for alternatives of the form $\theta_n \in \{-n^{-1/3}, -n^{-1/2}, -n^{-1/1.5}, -n^{-1}\}$. The Andrews and Soares (2010) test also has worse size control than all of the other tests except for the numerical second order derivative 1 test. The tests using the derivative of the analytic first order derivative and the numerical second order derivative 2 have the highest power against all alternatives and exhibit good size control.

As we go from $\epsilon_n = \frac{\sqrt{\log(n)}}{\sqrt{n}}$ to $\epsilon_n = n^{-1/6}$, the power of the Andrews and Soares (2010) test increases so that it is approximately equal to the power of the tests using the derivative of the analytic first order derivative and the numerical second order derivative 2 for all alternatives except $\theta_n = -n^{-1/2}$, in which case the Andrews and Soares (2010) test has lower power. The Andrews and Soares (2010) test has slightly better size control than the tests using the derivative of the analytic first order derivative and the numerical second order derivative 2 when $\theta_n \in \{0, n^{-1}\}$.

As we decrease ϵ_n from $n^{-1/6}$ to $n^{-1/2}$, the power of the Andrews and Soares (2010) test for alternatives of the form $\theta_n \in \{-n^{-1/6}, -n^{-1/3}, -n^{-1/2}\}$ decreases dramatically, and the size for $\theta_n \in \{n^{-1}, n^{-1/1.5}, n^{-1/2}\}$ increases to above the nominal size. In contrast, for the test using the numerical second order derivative 2, the power for alternatives of the form $\theta_n \in \{-n^{-1/6}, -n^{-1/3}, -n^{-1/2}, -n^{-1/1.5}\}$ and the size for all nonnegative θ_n are not greatly affected. The power of the test using the derivative of the analytic first order derivative is not greatly affected for $\theta_n \in \{-n^{-1/6}, -n^{-1/3}\}$ but the power does decrease dramatically for alternatives drifting faster to zero. The size of the test using the derivative of the analytic first order derivative decreases to almost 0 when $\epsilon_n = n^{-1/2}$ while the size of the test using the numerical second order derivative 2 is not greatly affected.

Note that for a given value of ϵ_n and any value of θ_n in the alternative, the power of the Andrews and Soares (2010) test is always no greater than the power of the test using the numerical second order derivative 1. This is consistent with our prediction at the end of Section 4. Moreover, for all values of θ_n in the alternative and for $\epsilon_n \in \{\sqrt{\log(n)}/\sqrt{n}, n^{-1/6}, n^{-1/3}\}$, the power of the test using the numerical second order derivative 2 is the greatest among the four tests. Only when $\epsilon_n = n^{-1/2}$ and only for alternatives $\theta_n \in \{-n^{-1/1.5}, -n^{-1}\}$ drifting very quickly to zero is its power lower than that of the Andrews and Soares (2010) test and the test using the numerical second order derivative 1, while still having higher power than the test using the derivative of the analytic first order derivative.

6. Bias reduction

If the functional of interest $\phi(\theta)$ admits a higher order directional Taylor expansion with a nondegenerate first order derivative, it is possible to modify the first order numerical directional delta method to make use of a higher order multiple point differentiation formula to reduce the bias in approximating the first order directional derivative numerically (Hong et al., 2015). Estimating the first derivative using multiple point numerical differentiation is akin to the use of (one sided) higher order kernel and local polynomial methods for bias reduction. Specifically, assume that, for $\phi_\theta^{(j)}(h)$ being functionals of h that are homogeneous of degree j, for $h_n \to h$,

$$\phi(\theta + th_n) = \sum_{j=0}^{r} \frac{1}{j!}\phi_\theta^{(j)}(h_n)\, t^j + O\left(t^{r+1}\right), \qquad \phi_\theta^{(j)}(h_n) - \phi_\theta^{(j)}(h) = O(h_n - h) = o(1). \tag{26}$$

Consider a p-point operator for estimating the first order directional derivative, with $p \le r$,

$$\begin{aligned} L_{\theta,p}^{\epsilon_n}(h) = \frac{1}{\epsilon_n}\sum_{l=0}^{p} a_l \phi(\theta + \epsilon_n l h) &= \frac{1}{\epsilon_n}\sum_{l=0}^{p} a_l \left[\sum_{j=0}^{r}\frac{1}{j!}\phi_\theta^{(j)}(h)\,\epsilon_n^j l^j + O\left(\epsilon_n^{r+1}\right)\right] \\ &= \sum_{j=0}^{p}\phi_\theta^{(j)}(h)\frac{\epsilon_n^{j-1}}{j!}\sum_{l=0}^{p} a_l l^j + O\left(\epsilon_n^p\right) \end{aligned}$$

The coefficients a_l, $l = 0, \ldots, p$ are determined by the system of equations:

$$\sum_{l=0}^{p} a_l l^j = \begin{cases} 1 & \text{for} \quad j = 1 \\ 0 & \text{for} \quad j \ne 1, j \le p. \end{cases} \tag{27}$$

Using these choices for a_l and $\epsilon_n \to 0$ leads to

$$L_{\theta,p}^{\epsilon_n}(h) = \phi_\theta^{(1)}(h) + O\left(\epsilon_n^p\right) \tag{28}$$

The p-point first order numerical derivative is

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*; p\right) \equiv L_{\hat{\theta},p}^{\epsilon_n}\left(\mathbb{Z}_n^*\right) \tag{29}$$

For example, $\hat{\phi}_n'\left(\mathbb{Z}_n^*\right) = \frac{\phi\left(\hat{\theta}_n + \epsilon_n \mathbb{Z}_n^*\right) - \phi\left(\hat{\theta}_n\right)}{\epsilon_n}$ corresponds to $p = 1$, $a_0 = -1$, $a_1 = 1$. When $p = 2$, $a_0 = -\frac{3}{2}$, $a_1 = 2$, $a_2 = -\frac{1}{2}$:

$$\hat{\phi}_n'\left(\mathbb{Z}_n^*; 2\right) \equiv \frac{-\frac{1}{2}\phi\left(\hat{\theta}_n + 2\epsilon_n\mathbb{Z}_n^*\right) + 2\phi\left(\hat{\theta}_n + \epsilon_n\mathbb{Z}_n^*\right) - \frac{3}{2}\phi\left(\hat{\theta}_n\right).}{\epsilon_n} \tag{30}$$

It is straightforward to generalize Theorem 3.1 to show consistency of (29).

Theorem 6.1. *Let (26) and the conditions in Theorem 3.1 hold. Then* $\hat{\phi}_n'\left(\mathbb{Z}_n^*; p\right) \overset{\mathbb{P}}{\rightsquigarrow} \phi_{\theta_0}'(\mathbb{G}_0)$.

7. Conclusion

We have proposed a one-sided finite difference numerical directional derivative as a computationally simple estimator for the directional directive developed in Fang and Santos (2014). We have demonstrated that when the $\phi(\cdot)$ function is Lipschitz, the numerical directional derivative is a consistent estimator for the directional derivative. Additionally, we have shown how to conduct uniformly valid inference using the first order directional delta method when $\phi(\cdot)$ is a convex and Lipschitz function. Lastly, we have demonstrated how to consistently estimate the second order directional derivative and use the second order directional delta method to conduct pointwise valid inference.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2018.06.007.

References

Amemiya, Takeshi, 1985. Advanced Econometrics. Harvard University Press.
Andrews, Donald W.K., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68 (2), 399–405.
Andrews, Donald W.K., Buchinsky, Moshe, 2000. A three-step method for choosing the number of bootstrap repetitions. Econometrica 68 (1), 23–51.
Andrews, Donald W.K., Soares, Gustavo, 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78 (1), 119–157.
Andrews, Donald W.K., Shi, Xiaoxia, 2013. Nonparametric inference based on conditional moment inequalities. J. Econometrics.
Bhattacharya, Rabindra N., Rao, Ramaswamy Ranga, 1986. Normal Approximation and Asymptotic Expansions, vol. 64. SIAM.
Bugni, Federico A., Canay, Ivan A., Shi, Xiaoxia, 2015. Specification tests for partially identified models defined by moment inequalities. J. Econometrics 185 (1), 259–282.
Chen, Qihui, Fang, Zheng, 2015. Inference on Functionals Under First Order Degeneracy. Available at SSRN.
Chernozhukov, Victor, Hong, Han, 2003. A MCMC approach to classical estimation. J. Econometrics 115 (2), 293–346.
Dümbgen, Lutz, 1993. On nondifferentiable functions and the bootstrap. Probab. Theory Related Fields 95 (1), 125–140.
Fang, Zheng, Santos, Andres, 2014. Inference on Directionally Differentiable Functions. arXiv preprint arXiv:1404.3763.
Hansen, Bruce E., 2017. Regression kink with an unknown threshold. J. Bus. Econom. Statist. 35 (2), 228–240.
Hirano, Keisuke, Porter, Jack R., 2012. Impossibility results for nondifferentiable functionals. Econometrica 80 (4), 1769–1790.
Hong, Han, Li, Jessie, 2014. The Numerical Bootstrap. Working Paper, Stanford University.
Hong, Han, Mahajan, Aprajit, Nekipelov, Denis, 2015. Extremum estimation and numerical derivatives. J. Econometrics 188 (1), 250–263.
Kosorok, Michael R., 2007. Introduction to Empirical Processes and Semiparametric Inference. Springer.
Politis, D., Romano, J., Wolf, M., 1999. Subsampling. In: Springer Series in Statistics.
Ponomareva, Maria, 2010. Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates. Working Paper.
Romano, Joseph P., Shaikh, Azeem M., 2008. Inference for identifiable parameters in partially identified econometric models. J. Statist. Plann. Inference 138 (9), 2786–2807.
Romano, Joseph P., Shaikh, Azeem M., 2012. On the uniform asymptotic validity of subsampling and the bootstrap. Ann. Statist. 40 (6), 2798–2822.
Römisch, Werner, 2005. Delta method, infinite dimensional. Encycl. Stat. Sci..
Shapiro, Alexander, 1990. On concepts of directional differentiability. J. Optim. Theory Appl. 66 (3), 477–487.
Shapiro, Alexander, 2000. Statistical inference of stochastic optimization problems. In: Probabilistic Constrained Optimization. Springer, pp. 282–307.
Song, Kyungchul, 2014. Local asymptotic minimax estimation of nonregular parameters with translation-scale equivariant maps. J. Multivariate Anal. 125, 136–158.

Further Reading

Bonnans, J. Frederic, Shapiro, Alexander, 2013. Perturbation Analysis of Optimization Problems. Springer Science & Business Media.
Bugni, Federico, Canay, Ivan, Shi, Xiaoxia, 2017. Inference for functions of partially identified parameters in moment inequality models. Quant. Econ. 8 (1), 1–38.
van der Vaart, A.W., Wellner, Jon, 1996. Weak Convergence and Empirical Processes. Springer.
Woutersen, Tiemen, Ham, John C., 2013. Calculating Confidence Intervals for Continuous and Discontinuous Functions of Parameters. Cemmap Working Paper, Centre for Microdata Methods and Practice.