Impossibility Results for Nondifferentiable Functionals

Hirano, Keisuke and Porter, Jack

University of Arizona

29 June 2009

Impossibility Results for Nondifferentiable Functionals*

Keisuke Hirano
Department of Economics
University of Arizona
hirano@u.arizona.edu

Jack R. Porter
Department of Economics
University of Wisconsin
jrporter@ssc.wisc.edu

First Version: March 6, 2009
This Draft: June 29, 2009

Abstract

We examine challenges to estimation and inference when the objects of interest are nondifferentiable functionals of the underlying data distribution. This situation arises in a number of applications of bounds analysis and moment inequality models, and in recent work on estimating optimal dynamic treatment regimes. Drawing on earlier work relating differentiability to the existence of unbiased and regular estimators, we show that if the target object is not continuously differentiable in the parameters of the data distribution, there exist no locally asymptotically unbiased estimators and no regular estimators. This places strong limits on estimators, bias correction methods, and inference procedures.

1 Introduction

In bounds analysis and inference for treatment effects, certain estimands of interest are nonsmooth functionals of the underlying distribution of the data, and this creates challenges for standard estimation and inference procedures. We examine such cases, and show that nonsmoothness implies sharp limits on the performance of estimators and inference procedures. In particular, if a limiting version of the estimand is not continuously differentiable, then there exist *no* locally asymptotically unbiased estimators, and there exist *no* regular estimators, when the underlying set of distributions is a smooth family. Since no locally asymptotically unbiased estimators exist, bias correction procedures cannot completely eliminate local bias, and reducing bias too much will eventually cause the variance of the procedure to diverge. Nonexistence of regular estimators implies that standard arguments for optimality of estimators, such as the convolution theorem for semiparametric estimators, cannot be used, and that standard Wald-type inference procedures are not valid.

*Early versions of this paper were titled "Impossibility Results for Bounds Estimation." We are grateful to Gary Chamberlain, Jinyong Hahn, Bruce Hansen, Guido Imbens, Ariel Pakes, Thomas Richardson, Jamie Robins, and numerous seminar participants for helpful comments and suggestions. Porter thanks the National Science Foundation for research support under grant SES-0438123.

We use Le Cam's limits of experiments approach to provide a simple and intuitive argument for our impossibility results. Under local asymptotic normality, the multivariate normal location model serves as a limit experiment, in the sense that any sequence of estimators in the model of interest is matched by some estimator in the normal model. We show that in the normal location model, no unbiased or translation equivariant estimators exist if the functional of interest is not continuously differentiable. Therefore, there exist no sequences of estimators in the original model that are locally asymptotically unbiased or regular.

In our analysis, we draw upon Blumenthal and Cohen (1968), who showed that no unbiased estimator exists for the minimum of two independent normal means. We extend their argument to a multivariate, correlated normal model where the object of interest is a general (but nondifferentiable) function of the mean parameters, and the criterion is either unbiasedness or location equivariance. Our result on regular estimators is similar to van der Vaart (1991b), who showed that the existence of a regular estimator, combined with a further mild condition, implies that the functional is differentiable. We use a different argument, and obtain the stronger result that regularity implies continuous differentiability of the functional.

2 Examples

Before developing the theory, we begin with some examples of recent work in economics and biostatistics in which the estimand is a nondifferentiable functional of the data distribution.

Example 1 *Bounds for an Incomplete Auction Model*

Haile and Tamer (2003) showed that it is possible to obtain useful inference for valuation distributions in auction models without fully specifying the structure of the model. Suppose that bidders $i = 1, \dots, m$ draw valuations v_i independently from a distribution with cumulative distribution function (CDF) $F(v)$. Bidders make bids b_i subject to:

1. $b_i \leq v_i$
2. Bidders do not allow an opponent to win the good at a price she is willing to beat.

We do not observe v_i, only b_i. Let $G(b)$ denote the CDF of bids.

Condition 1 implies $F(v) \leq G(v)$ for all v. Haile and Tamer observed that this upper bound can be tightened as follows. Let $F_i(v)$ denote the CDF of the ith order statistic out of m. There exists a monotone mapping ϕ such that

$$F(\cdot) = \phi(F_i(\cdot); i).$$

Then Condition 1 implies:

$$F(v) \leq \min_{i=1,\dots,m} \phi(G_i(v); i) = \kappa.$$

By a similar argument, Condition 2 gives a lower bound for $F(v)$ involving a maximum of estimable quantities. Haile and Tamer noted that while the empirical analogs of the $G_i(v)$ are consistent and asymptotically jointly normally distributed, the plug-in estimator for the bound κ will be biased downward due to the convexity of the minimum operator, and that this problem can be quite severe for realistic sample sizes. They suggest a bias reduction procedure; similar issues in other bounds analyses were noted by Manski and Pepper (2000), and Kreider and Pepper (2007) suggested a bootstrap bias correction. In the analysis below, we will find that it is impossible to completely eliminate bias, and that reducing bias too much leads to large increases in variance.

Example 2 *Imbens-Manski Bounds*

Imbens and Manski (2004) considered inference for a partially identified, scalar parameter, where the identification region is given by a bounded interval, $[\pi, \kappa]$. Under a uniform asymptotic normality condition for estimators of (π, κ), they propose a confidence interval method for the parameter. Ensuring uniform validity of the confidence interval requires some care for cases where π is close to κ. (See Stoye (2008) and Fan and Park (2008) for extensions of the Imbens-Manski approach.) We consider an extension of their framework where the endpoints are identified by multiple features of the data. For example, each endpoint could be determined by multiple moment inequalities. Such cases have received considerable attention in the rapidly growing moment inequality literature, see Chernozhukov, Hong, and Tamer (2007), Andrews and Soares (2007), Andrews and Guggenberger (2007), Pakes, Porter, Ho, and Ishii (2006), Bugni (2008), Canay (2008), Rosen (2008), among others, and raise further challenges to estimation and inference.

Suppose we have data $Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{P}$ for $i = 1, \ldots, n$. Let $\theta(\mathcal{P}) = (\theta_1(\mathcal{P}), \ldots, \theta_j(\mathcal{P}))'$ be a vector of functionals of the data distribution for which there exists a $\sqrt{n}$-asymptotically normal estimator. Focusing on the upper endpoint, suppose

$$\kappa = \min\{\theta_1, \ldots, \theta_j\},$$

For example, we could observe a j-dimensional data vector Y_i with $E[Y_i] = \theta$. Then κ is the minimum of the means of the components of Y_i, which is a simple moment inequality model of the form $0 \leq E(Y_i) - \ell\kappa$ with ℓ a j-vector of ones.

As we will see below, there exist no regular estimators for κ. Hence, there exist no estimators satisfying the uniform asymptotic normality condition in this extension of the Imbens-Manski framework.

Example 3 *Inference for Expected Outcomes under the Best Treatment*

Consider a randomized experiment comparing outcomes under two treatments, $T = 0, 1$. Let $Y(0)$ and $Y(1)$ denote potential outcomes under the two treatments, and define

$$\theta_0 = E[Y(0)]$$
$$\theta_1 = E[Y(1)]$$

Interest often focuses on estimating the average treatment effect $\theta_1 - \theta_0$. Recent work on optimal treatment assignment rules by Manski (2004), Dehejia (2005), Stoye (2006), Schlag (2006), Tetenov (2007), and Hirano and Porter (2008) has adopted a decision-theoretic approach for choosing the optimal treatment. Another object of interest is

$$\kappa = \max\{\theta_0, \theta_1\}.$$

This can be interpreted as the expected outcome under the best treatment. Clearly, sample analog estimators of this quantity will suffer from bias problems, just as in the previous two examples.

In addition to being of interest in their own right, objects of this form play an important role in recent work on treatment assignment problems in dynamic settings, where backward induction solutions must take into account the continuation payoffs from choosing the best treatment in later stages (Murphy (2003)). Robins (2004) noted that estimators for many such models will generally suffer from bias and lack of regularity, and develops uniform inference procedures. Moodie and Richardson (2007) and Chakraborty, Strecher, and Murphy (2008) proposed bias-correction procedures. Our results below extend the arguments in Robins (2004) to show that lack of continuous differentiability leads automatically to impossibility of locally asymptotically unbiased or regular estimators.

3 Theory

Our argument proceeds in two steps. First, we study finite sample theory in a simple normal model. Then, we use the exact results in the normal case to obtain a general asymptotic theory for the problem of estimating non-smooth functionals in a smooth family of distributions.

3.1 Exact Theory for the Multivariate Normal Location Model

Suppose we have a single observation for the k-dimensional random vector Z, where

$$Z \sim N(h, \Sigma),$$

$h = (h_1, \dots, h_k)' \in \mathbb{R}^k$, and Σ is a known variance-covariance matrix. Let $\dot{\kappa}(h)$ be some function of the parameters h.[1]

Let $U \sim \text{Uniform}[0,1]$, independently of Z, and let $T(Z,U)$ denote a scalar-valued randomized statistic. Then $T(Z,U)$ is an *unbiased* point estimator iff

$$\dot{\kappa}(h) = E_h[T(Z,U)] = \int_{[0,1]} \int_{\mathbb{R}^k} T(z,u) f(z|h,\Sigma) dz du, \quad \forall h \in \mathbb{R}^k, \tag{1}$$

where $f(z|h,\Sigma)$ denotes the multivariate normal density.

[1] The dot notation in $\dot{\kappa}$ suggests a derivative, and this will be useful for the asymptotic theory in the next subsection. Here, however, we simply view $\dot{\kappa}$ as an arbitrary function.

In addition to unbiasedness, it is useful to consider an equivariance condition. We say that $T(Z,U)$ is *translation equivariant* iff the distribution under h of

$$T(Z,U)-\dot{\kappa}(h)$$

does not depend on h.

To see how lack of smoothness in $\dot{\kappa}$ affects the possibility of unbiased or equivariant estimation, we partition Z into its first component and the remaining subvector, with parameters partitioned comformably:

$$Z=\begin{pmatrix}Z_1\\ Z_2\end{pmatrix}\sim N\left(\begin{pmatrix}h_1\\ h_2\end{pmatrix},\begin{pmatrix}\sigma_{11} & \Sigma_{1,2}\\ \Sigma_{2,1} & \Sigma_{2,2}\end{pmatrix}\right).$$

We also use the notation $\dot{\kappa}(h_1,h_2)$ for the object of interest.

Assumption 1 *$\dot{\kappa}$ is not continuously differentiable in h_1 at some point h_0 in the parameter space.*

The dependence of $\dot{\kappa}$ on h_2 can be arbitrary. For instance, $\dot{\kappa}$ may be a function of h_1 and a proper subset of h_2. To see why nondifferentiability can lead to problems, suppose that $T(Z,U)$ is an unbiased estimator. Through a bounding inequality for the exponential function, we can verify the uniform integrability condition that implies differentiability under the integral sign in (1). Hence, the derivative of $E_h[T(Z,U)]$ with respect to h_1 is well-defined and exists everywhere:

$$\frac{\partial}{\partial h_1}E_h[T(Z,U)]=\int_{[0,1]}\int_{\mathbb{R}^k}T(z,u)\frac{\partial}{\partial h_1}f(z|h,\Sigma)dzdu.$$

However, by assumption, $E_h[T(Z,U)]=\dot{\kappa}(h)$ is not continuously differentiable at $h=h_0$, which is a contradiction. So an unbiased estimator for $\dot{\kappa}(h)$ cannot exist. We extend this argument to obtain the following result.

Theorem 1 *Suppose Assumption 1 holds, and $T(Z,U)$ is (possibly randomized) estimator of $\dot{\kappa}$. Then T is neither unbiased nor translation equivariant.*

Proof: See Appendix.

□

The impossibility of unbiased or equivariant estimation arises from nondifferentiability of $\dot{\kappa}(h)$ at a single point h_0, even if the function is very well behaved elsewhere. In some cases where the points of nondifferentiability are isolated, one can construct estimators with arbitrarily small bias by targeting smoothed versions of $\dot{\kappa}$ or employing some sort of iterated bias correction. However, Doss and Sethuraman (1989) showed the following remarkable result: if there exists no unbiased estimator, but there exists a sequence of estimators whose bias becomes arbitrarily small (pointwise in the parameter space), then such a sequence must have variance increasing to infinity at *every* point in the parameter space. So if one reduces the bias of the estimator too much, the estimator will have arbitrarily large variance everywhere in the parameter space.

3.2 Asymptotic Theory

Although there do not exist exactly unbiased or equivariant estimators for $\dot{\kappa}(h)$ in the normal model, one could hope to construct approximately unbiased or equivariant estimators. For example, the MLE in a parametric model is not generally unbiased in finite samples, but in well-behaved settings it is asymptotically unbiased and regular. Therefore, we consider asymptotic approximations in both parametric and infinite-dimensional settings, and examine how lack of smoothness of the target functional limits the properties of estimators and inference procedures.

3.2.1 Parametric Models

First, consider a parametric family of distributions for the data. Suppose that for $i = 1, 2, \ldots, n$, the data Y_i are IID with

$$Y_i \sim G_\theta,$$

where $\theta \in \Theta \subset \mathbb{R}^k$. We assume that Θ is an open set. Let $\mathcal{Y}$ denote the support of Y_i. (The observations Y_i could be vector-valued or take values in some more general space.)

We take a standard local approximation about a point $\theta_0 \in \Theta$, and take the family of distributions to be locally asymptotically normal at θ_0 (see van der Vaart, 1998):

Assumption 2 *(a) (Differentiability in quadratic mean) There exists a function $s : \mathcal{Y} \to \mathbb{R}^m$ such that*

$$\int \left[dG_{\theta_0+h}^{1/2}(y) - dG_{\theta_0}^{1/2}(y) - \frac{1}{2} h' \cdot s(y) dG_{\theta_0}^{1/2}(y) \right]^2 = o(\|h\|^2) \quad \text{as } h \to 0;$$

(b) The Fisher information matrix $J_0 = E_{\theta_0}[ss']$ is nonsingular.

Given this assumption, it will be useful to adopt the usual local parametrization around a point θ_0,

$$\theta_{n,h} = \theta_0 + \frac{h}{\sqrt{n}}.$$

Suppose interest centers on some function of the parameters, $\kappa(\theta)$. Under conventional smoothness conditions on the sequence of experiments $\mathcal{E}^n = \{G_\theta^n : \theta \in \Theta\}$, the MLE $\hat{\theta}_{ml}$ and other estimators such as the Bayes estimator are asymptotically efficient. However, the limit distributions of derived estimators of $\kappa(\theta)$ will depend crucially on the smoothness in κ at the point θ_0. Here we want to allow κ to lie in a class of functions that includes certain non-differentiable functions, such as the min and max functions in the examples. For this purpose, define the *one-sided directional derivative* of κ at θ_0 in the direction λ as:

$$\dot{\kappa}_{\theta_0}(\lambda) = \lim_{t \downarrow 0} \frac{\kappa(\theta_0 + t\lambda) - \kappa(\theta_0)}{t}.$$

The following assumption defines the class of functions that we will consider.

Assumption 3 *κ has one-sided directional derivatives in all directions at θ_0.*

In this setting, an estimator (or estimator sequence) is a sequence of functions $T_n : \mathcal{Y}^n \to \mathbb{R}$. We focus on estimators that possess limit distributions in the sense that, for all h,

$$\sqrt{n}\,(T_n - \kappa(\theta_{n,h})) \overset{h}{\rightsquigarrow} L_h, \tag{2}$$

where $\overset{h}{\rightsquigarrow}$ indicates weak convergence under $\theta_{n,h}$. The L_h are the limiting laws of the estimator under different local sequences of parameters. These laws could, in general, be degenerate.

The standard definition of a *regular* estimator is one that has $L_h = L$ for all h, where L does not depend on h. This is a local asymptotic version of equivariance, and is intended to capture the requirement that the centered limit distributions be invariant to small perturbations of the parameters. Regularity plays an important role in conventional results on optimality of point estimators, such as semiparametric efficiency bounds and convolution theorems, and is also crucial for the uniform validity of standard inference procedures.[2] In addition, we say that T_n is *locally asymptotically unbiased* if, for all h, the laws L_h have mean 0.

Local asymptotic normality in Assumption 2 implies that the Gaussian location model

$$Z \sim N(h, J_0^{-1})$$

provides a characterization of the asymptotic behavior of our sequence of statistical models G_θ^n. In particular, by the Asymptotic Representation Theorem (van der Vaart, 1991a), the limit laws of $\sqrt{n}(T_n - \kappa(\theta_0))$ are matched by a randomized estimator $T(Z, U)$, where U is Uniform$[0,1]$ independently of Z. By Assumption 3,

$$\sqrt{n}(\kappa(\theta_0 + h/\sqrt{n})) - \kappa(\theta_0)) \longrightarrow \dot{\kappa}_{\theta_0}(h).$$

Together these conclusions imply that $T(Z,U) - \dot{\kappa}_{\theta_0}(h) \sim L_h$ for the limit laws given in (2). Then, if $\dot{\kappa}_{\theta_0}$ is not continuously differentiable at some point h_0, Theorem 1 shows that there exists no unbiased or translation equivariant estimator in the multivariate normal model. Our asymptotic impossibility result follows immediately.

Theorem 2 *Let T_n be any sequence of estimators based on $\{Y_i\}_{i=1}^n$. Suppose Assumptions 2 and 3 hold, and $\dot{\kappa}_{\theta_0}(\cdot)$ satisfies Assumption 1. Then T_n is not locally asymptotically unbiased and is not regular.*

Remarks:

1. In the Theorem, T_n can be any procedure based on the data, so the result would apply to multi-step procedures such as bias reduction following an initial estimate, procedures based on an initial moment selection step, and procedures that use resampling techniques.[3]

[2] For instance, regularity is necessary for consistency of the parametric bootstrap in LAN models; see Beran (1997).

[3] Techniques like the bootstrap often use simulation to approximate a distribution of a statistic or some other quantity that depends deterministically on the data. Typically the numerical approximation does not change the

2. Van der Vaart (1991b) contains a closely related result which shows that in a setting with a possibly infinite-dimensional parameter space, regularity and a further mild property of an estimator of some functional of the distribution of the data implies differentiability of that functional. We state and show nonexistence somewhat more simply using the finite-dimensional normal limit experiment, and obtain the same conclusion under the stronger condition of continuous differentiability.

3.2.2 Infinite-Dimensional Models

Our analysis so far has considered only smooth parametric models. A more general model for the data would have

$$Y_i \sim G \in \mathcal{G},$$

where the set of possible distributions $\mathcal{G}$ may be infinite-dimensional.

For these models, one can obtain a Gaussian process limit experiment associated with the tangent space of $\mathcal{G}$ around a centering G_0, and derive results for that case, but it is simpler to extend our results for the parametric case as follows. Suppose that $\theta(G) = (\theta_1(G), \ldots, \theta_k(G))'$ is a vector-valued functional of G that is estimable at a $\sqrt{n}$ rate and Hadamard differentiable with respect to G. As before, let $\kappa(\theta(G))$ be a scalar function of $\theta(G)$. Then, provided that there exists an LAN parametric submodel of $\mathcal{G}$ containing G_0, we can use our parametric result in Theorem 2 to conclude that there exists no locally asymptotically unbiased or regular estimator in the original semiparametric model.

The three examples in Section 2 can all be fit into this general semiparametric framework. If we localize around a measure G_0 such that at least two of the arguments in the min or max function defining the estimand are equal, then $\kappa(\theta(G_0))$ has a one-sided directional derivative that satisfies our Assumption 1, and we can conclude that there exist no locally asymptotically unbiased or regular estimators.

4 Conclusion

We have used the Le Cam limits of experiments framework to reduce the asymptotic analysis to an analysis of a multivariate normal location model. Impossibility of unbiased or equivariant estimation in the normal model implies impossibility of locally asymptotically unbiased estimation or regular estimation. As a consequence, bias reduction procedures will eventually lead to a large increase in variance, conventional arguments for optimality of estimators cannot be used, and standard Wald-type inference procedures cannot be uniformly valid.

---

limit distributions of the procedure. So our result, which holds for sequences of nonrandomized estimators, applies equally well for such resampling methods. We could also extend the result to allow the T_n to be inherently randomized, by expanding the definition of the data $\{Y_i\}$ appropriately.

Local asymptotic normality also provides a useful way to devise alternative procedures with good properties. Any sequence of statistics with limit distributions has a matching statistic in the limiting normal model. This suggests that we could work directly in the normal model, propose alternative estimators or inference procedures, and compare their distributions under different parameters. If we find a good procedure in the normal model, it is usually possible to construct the matching sequence of estimators for the original problem of interest.

References

Andrews, D. W. K., and P. Guggenberger (2007): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," Working Paper.

Andrews, D. W. K., and G. Soares (2007): "Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection," Working Paper.

Beran, R. (1997): "Diagnosing Bootstrap Success," *Annals of the Institute of Statistical Mathematics*, 49(1), 1–24.

Blumenthal, S., and A. Cohen (1968): "Estimation of the Larger of Two Normal Means," *Journal of the American Statistical Association*, 63(323), 861–876.

Bugni, F. A. (2008): "Bootstrap Inference in Partially Identified Models," Working Paper.

Canay, I. A. (2008): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Working Paper.

Chakraborty, B., V. Strecher, and S. A. Murphy (2008): "Bias Correction and Confidence Intervals for Fitted Q-Iteration," Working Paper.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Dehejia, R. H. (2005): "Program Evaluation as a Decision Problem," *Journal of Econometrics*, 125, 141–173.

Doss, H., and J. Sethuraman (1989): "The Price of Bias Reduction when there is no Unbiased Estimate," *The Annals of Statistics*, 17(1), 440–442.

Fan, Y., and S. S. Park (2008): "Confidence Sets for Some Partially Identified Parameters," Working Paper, University of North Carolina.

Haile, P. A., and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1), 1–51.

Hirano, K., and J. R. Porter (2008): "Asymptotics for Statistical Treatment Rules," forthcoming, *Econometrica*.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Kreider, B., and J. V. Pepper (2007): "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of the American Statistical Association*, 102(478), 432–441.

Manski, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72(4), 1221–1246.

Manski, C. F., and J. V. Pepper (2000): "Monotone Instrumental Variables: with an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

Moodie, E. E. M., and T. S. Richardson (2007): "Bias Correction in Non-Differentiable Estimating Equations for Optimal Dynamic Regimes," Working Paper.

Murphy, S. A. (2003): "Optimal Dynamic Treatment Regimes," *Journal of the Royal Statistical Society, Series B*, 65, 331–366.

Pakes, A., J. R. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and Their Application," Working Paper.

Robins, J. M. (2004): "Optimal Structural Nested Models for Optimal Decisions," in *Proceedings of the Second Seattle Symposium on Biostatistics*, ed. by D. Y. Lin, and P. Heagerty, New York. Springer-Verlag.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146(1), 107–117.

Schlag, K. (2006): "Eleven – Tests Needed for a Recommendation," EUI Working Paper ECO 2006-2.

Stoye, J. (2006): "Minimax Regret Treatment Choice with Finite Samples," Working paper.

——— (2008): "More on Confidence Intervals for Partially Identified Parameters," forthcoming, *Econometrica*.

Tetenov, A. (2007): "Statistical Treatment Choice Based on Asymmetric Minmax Regret Criteria," Working paper.

Van der Vaart, A. W. (1991a): "An Asymptotic Representation Theorem," *International Statistical Review*, 59, 99–121.

——— (1991b): "On Differentiable Functionals," *The Annals of Statistics*, 19, 178–204.

——— (1998): *Asymptotic Statistics*. Cambridge University Press, New York.

Appendix

Proof of Theorem 1:

We give the argument for equivariance; the argument for unbiasedness works similarly and is omitted. Suppose that $T(Z,U)$ is equivariant. Let $\tilde{T}(z,u) = T(z,u) - \dot{\kappa}(h)$. We want to show that if the distribution of $\tilde{T}(Z,U)$ does not depend on h, then $\dot{\kappa}(h)$ must be everywhere continuously differentiable in h_1. Let $\phi_h(s) = E_h\left[e^{isT(Z,U)}\right]$ be the characteristic function of $T(Z,U)$ under h, and let $\tilde{\phi}_h(s)$ be the characteristic function of $\tilde{T}(Z,U)$ under h. We can write $\tilde{\phi}_h(s) = e^{-is\dot{\kappa}(h)}\phi_h(s)$. Since $T(Z,U)$ is equivariant,

$$e^{is\dot{\kappa}(h)} = \frac{\phi_h(s)}{\tilde{\phi}_0(s)} \tag{3}$$

where $\tilde{\phi}_0(s)$ does not depend on h. Note that $\tilde{\phi}_0(s)$ is non-zero for s in an open neighborhood of zero. By Lemma 1, $\phi_h(s)$ and hence $\frac{\phi_h(s)}{\tilde{\phi}_0(s)}$ are continuous in h for all s in a neighborhood of zero. By Lemma 3, $\dot{\kappa}$ is everywhere continuous in h.

Now, we show differentiability of $\dot{\kappa}$ in h_1. Take a nonzero s in a neighborhood of zero. Both $\cos(\cdot)$ and $\sin(\cdot)$ are differentiable at $s\dot{\kappa}(h_0)$, and one of these derivatives must be non-zero. Without loss of generality, assume $cos'(s\dot{\kappa}(h_0)) = -\sin(s\dot{\kappa}(h_0)) \neq 0$. The real part of the right hand side of (3) is:

$$\frac{Re(\phi_h(s))Re(\tilde{\phi}_0(s)) + Im((\phi_h(s))Im(\tilde{\phi}_0(s))}{[Re(\tilde{\phi}_0(s))]^2 + [Im(\tilde{\phi}_0(s))]^2}$$

From Lemma 4, both real and imaginary parts of $\phi_h(s)$ are differentiable in h_1 at h_0, yielding differentiability of the above expression in h_1. We've already established that $s\dot{\kappa}$ is continuous at h_0 in h, so Lemma 6 gives the desired differentiability. Lastly we show continuous differentiability of $\dot{\kappa}$ in h_1 at h_0. Differentiating equation (3) for nonzero s in a neighborhood of zero, we get

$$\frac{\partial}{\partial h_1}\dot{\kappa}(h) = -ie^{-is\dot{\kappa}(h)}\frac{\frac{\partial}{\partial h_1}\phi_h(s)}{s\tilde{\phi}_0(1)}.$$

So, given the continuity of $\dot{\kappa}$, continuity of $\frac{\partial}{\partial h_1}\phi_h(s)$ suffices to give the result. Lemma 7 shows the needed continuity. □

Proof of Theorem 2:

If the estimator sequence does not possess limit distributions as in Equation (2), then it cannot be regular or locally asymptotically unbiased by definition, and the conclusion of the theorem holds trivial. Now consider the case where Equation (2) holds for the estimator sequence. By the argument preceding the statement of the theorem,

$$\sqrt{n}(T_n - \kappa(\theta_0 + h/\sqrt{n})) \overset{h}{\rightsquigarrow} T - \dot{\kappa}_{\theta_0}(h).$$

The result follows by Theorem 1. □

Lemmas

Let

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix} \right).$$

Here Z_1 is scalar, but Z_2 is a vector of arbitrary finite-dimensional size. Let $f(z|h, \Sigma)$ denote the density of a k-dimensional multivariate normal distribution with mean h and positive definite variance-covariance Σ.

Lemma 1 *$\phi_h(s) = E_h\left[e^{isT(Z,U)}\right]$ is continuous in h for all s.*

Proof: We show continuity for the real component below. The proof for the imaginary component follows similarly. Consider some point h_0. Let $\zeta = \sup_{h:\|h-h_0\|\leq 1} f(z|h,\Sigma)$ and note $\zeta < \infty$. Set $C = 2$ and $\delta = 1$ in Lemma 2, and let q be as given by the Lemma 2. Then

$$\begin{aligned} &\int_0^1 \int \sup_{h:\|h-h_0\|\leq 1} |\cos(sT(z,u))[f(z|h,\Sigma) - f(z|h_0,\Sigma)]| \; dz \; du \\ &\leq \; 2\int_0^1 \int \sup_{h:\|h-h_0\|\leq 1} f(z|h,\Sigma) dz du \\ &\leq \; 2\int_0^1 \int_{z:\|z-h_0\|\leq 2} \zeta \; dz \; du + 2q^{-k^2/2} \int_0^1 \int_{z:\|z-h_0\|>2} f(z|h_0, q^{-1}\Sigma) \; dz \; du \\ &\leq \; 2 \cdot 4^k \zeta + 2q^{-k^2/2} \; < \; \infty \end{aligned}$$

For all z and u, $|\cos(sT(z,u))[f(z|h,\Sigma) - f(z|h_0,\Sigma)]| \longrightarrow 0$ as $h \longrightarrow h_0$, so by dominated convergence

$$|Re(\phi_h(s)) - Re(\phi_{h_0}(s))| \leq \int_0^1 \int |\cos(sT(z,u))[f(z|h) - f(z|h_0)]| \; dz \; du \longrightarrow 0.$$

□

Lemma 2 *Given $C > \delta > 0$ and some h_0, there exists $q > 0$ such that*

$$\sup_{h:\|h-h_0\|\leq \delta} f(z|h,\Sigma) \leq q^{-k^2/2} f(z|h_0, q^{-1}\Sigma)$$

for z such that $\|z - h_0\| \geq C$.

Proof: Let $\underline{\alpha} = \inf_{\|t\|=1} t'\Sigma^{-1}t$ and $\bar{\alpha} = \sup_{\|t\|=1} t'\Sigma^{-1}t$. Σ^{-1} is positive definite, so $\underline{\alpha} > 0$. Set $q = \frac{\underline{\alpha}(C-\delta)^2}{\bar{\alpha}C^2}$. For z such that $\|z - h_0\| \geq C$,

$$\begin{aligned} \inf_{h:\|z-h\|\geq\|z-h_0\|-\delta} (z-h)'\Sigma^{-1}(z-h) &= \inf_{d:d\geq\|z-h_0\|-\delta} d^2 \inf_{t:\|\frac{t}{d}\|=1} \left(\frac{t}{d}\right)' \Sigma^{-1} \left(\frac{t}{d}\right) \\ &= \; (\|z-h_0\| - \delta)^2 \underline{\alpha} \geq \|z-h_0\|^2 q\bar{\alpha} \geq q(z-h_0)'\Sigma^{-1}(z-h_0). \end{aligned}$$

Hence, for z such that $\|z-h_0\| \geq C$,

$$\begin{aligned}\sup_{h:\|h-h_0\|\leq\delta} f(z|h,\Sigma) &\leq \sup_{h:\|z-h\|\geq\|z-h_0\|-\delta} q^{-k^2/2}\left(2\pi|q^{-1}\Sigma|\right)^{-k/2}\exp\left(-\frac{1}{2}(z-h)'\Sigma^{-1}(z-h)\right)\\ &= q^{-k^2/2}f(z|h_0,q^{-1}\Sigma).\end{aligned}$$

□

Lemma 3 *If $e^{isk(h)}$ is continuous in h for all s, then $k(h)$ is continuous in h.*

Proof: The function e^{ist} is continuous and periodic in t So, $\lim_{h\longrightarrow h_0} e^{isk(h)} = e^{isk(h_0)}$ implies $\lim_{h\longrightarrow h_0} sk(h) = sk(h_0) + r_s 2\pi$ for some integer r_s. This can only hold for all s close to zero if $r_s = 0$. Hence k is continuous at h_0.

□

Lemma 4 *$\phi_h(s) = E_h\left[e^{isT(Z,U)}\right]$ is everywhere differentiable in h_1 for all s, and*

$$\frac{\partial}{\partial h_1}\phi_h(s) = \int_0^1\int e^{isT(z,u)}\frac{\partial}{\partial h_1}f(z|h)\ dz\ du.$$

Proof: We show differentiability for the real component below. The proof for the imaginary component follows similarly. First we show

$$\left\{\left|\cos(sT(z,u))\frac{f(z|(h_1+\Delta,h_2),\Sigma)-f(z|(h_1,h_2),\Sigma)}{\Delta}\right| : |\Delta|\leq 1\right\}$$

is uniformly integrable, and the result will follow by the Dominated Convergence Theorem. Let $\bar{h}(z_2,h) = h_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(z_2-h_2)$ and $\bar{\sigma} = \sigma_{11} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$, so $Z_1|\ Z_2 = z_2 \sim N(\bar{h}(z_2,h),\bar{\sigma}^2)$. By the mean value theorem,

$$f(z|(h_1+\Delta,h_2),\Sigma)-f(z|(h_1,h_2),\Sigma) = \frac{1}{\bar{\sigma}^2\sqrt{2\pi\bar{\sigma}^2}}(z_1-\bar{h}(z_2,h)-\bar{\Delta})\exp\left(-\frac{1}{2\bar{\sigma}^2}(z_1-\bar{h}(z_2,h)-\bar{\Delta})^2\right)\Delta f(z_2|h_2)$$

for some $\bar{\Delta}$ between zero and Δ. Then,

$$\begin{aligned}
&\sup_{|\Delta|\le 1}\left|\frac{f(z|(h_1+\Delta,h_2),\Sigma)-f(z|(h_1,h_2),\Sigma)}{\Delta}\right| \\
\le\ & \sup_{|\Delta|\le 1}\frac{1}{\sqrt{2\pi\bar{\sigma}^2}}\frac{\left|z_1-\bar{h}(z_2,h)-\Delta\right|}{\bar{\sigma}^2}\exp\left(-\frac{1}{2\bar{\sigma}^2}|z_1-\bar{h}(z_2,h)-\Delta|^2\right)f(z_2|h_2,\Sigma_{2,2}) \\
\le\ & \sup_{|\Delta|\le 1}\frac{1}{\sqrt{2\pi\bar{\sigma}^2}}e^{\frac{1}{2\bar{\sigma}^2}}\exp\left(-\frac{1}{2\bar{\sigma}^2}(|z_1-\bar{h}(z_2,h)-\Delta|-1)^2\right)f(z_2|h_2,\Sigma_{2,2}) \qquad \text{(by Lemma 5)} \\
\le\ & \begin{cases} \frac{1}{\sqrt{2\pi\bar{\sigma}^2}}e^{\frac{1}{2\bar{\sigma}^2}}\exp\left(-\frac{1}{2\bar{\sigma}^2}(z_1-\bar{h}(z_2,h)-2)^2\right)f(z_2|h_2,\Sigma_{2,2}) & \text{if } z_1-\bar{h}(z_2,h)\ge 2 \\ \frac{1}{\sqrt{2\pi\bar{\sigma}^2}}e^{\frac{1}{2\bar{\sigma}^2}}\exp\left(-\frac{1}{2\bar{\sigma}^2}(z_1-\bar{h}(z_2,h)+2)^2\right)f(z_2|h_2,\Sigma_{2,2}) & \text{if } z_1-\bar{h}(z_2,h)\le -2 \\ \sup_{|\Delta|\le 1}\frac{1}{\sqrt{2\pi\bar{\sigma}^2}}e^{\frac{1}{2\bar{\sigma}^2}}f(z_2|h_2,\Sigma_{2,2}) & \text{if } -2<z_1-\bar{h}(z_2,h)<2 \end{cases}
\end{aligned}$$

The dominance condition sufficient for uniform integrability follows immediately, and the result follows by dominated convergence.

□

Lemma 5 *Suppose $c>0$. For all $u\ge 0$, $\frac{u}{c}e^{-\frac{u^2}{2c}} < e^{\frac{1}{2c}}e^{-\frac{(u-1)^2}{2c}}$.*

Proof: Note that $e^{\frac{1}{2c}}e^{-\frac{(u-1)^2}{2c}} = e^{\frac{u}{c}}e^{-\frac{u^2}{2c}}$, and, since $e^v > v$ for $v\ge 0$, the last term is strictly greater than $\frac{u}{c}e^{-\frac{u^2}{2c}}$.

□

Lemma 6 *Suppose (a) w is differentiable at x_0 and $w'(x_0)\neq 0$; (b) k is continuous at h_0 and $x_0 = k(h_0)$; (c) $w\circ k$ is differentiable at h_0. Then, k is differentiable at h_0. Moreover,*

$$\frac{\partial}{\partial h_1}k(h_0) = \frac{\frac{\partial(w\circ k)}{\partial h_1}(h_0)}{w'(k(h_0))}.$$

Proof: By (c),

$$\frac{\partial(w\circ k)}{\partial h_1}(h_0) = \lim_{\Delta\longrightarrow 0}\frac{w(k(h_{0,1}+\Delta,h_{0,2}))-w(k(h_0))}{k(h_{0,1}+\Delta,h_{0,2})-k(h_0)}\cdot\frac{k(h_{0,1}+\Delta,h_{0,2})-k(h_0)}{\Delta}$$

By (a) and (b),

$$w'(k(h_0)) = \lim_{\Delta\longrightarrow 0}\frac{w(k(h_{0,1}+\Delta,h_{0,2}))-w(k(h_0))}{k(h_{0,1}+\Delta,h_{0,2})-k(h_0)}$$

The result follows by taking the limit of a quotient as a quotient of the limits where the denominator exists.

□

Lemma 7 *$\frac{\partial}{\partial h_1}\phi_h(s)$ is everywhere continuous in h for all s.*

Proof: We prove continuity for the real part, and a similar argument yields continuity for the imaginary part. Consider some point h_0. Let $\zeta = \sup_{h:\|h-h_0\|\le 1} f(z|h,\Sigma)$, $\varsigma = \sup_{z:\|z-h_0\|\le 2} |z_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}z_2|$ and $\varrho = \sup_{h:\|h-h_0\|\le 1} |h_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}h_2|$. Set $C = 2$ and $\delta = 1$ in Lemma 2, and let q be as given by Lemma 2.

$$\begin{aligned}
&\int_0^1 \int \sup_{h:\|h-h_0\|\le 1} |\cos(sT(z,u))\frac{1}{\bar{\sigma}^2}[(z_1 - \bar{h}(z_2,h))f(z|h,\Sigma) - (z_1 - \bar{h}(z_2,h_0))f(z|h_0,\Sigma)]|\ dz\ du \\
&\le \frac{1}{\bar{\sigma}^2}\int_0^1 \int \sup_{h:\|h-h_0\|\le 1} \Big|[(z_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}z_2) - (h_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}h_2)](f(z|h,\Sigma) - f(z|h_0,\Sigma)) \\
&\qquad\qquad - (h_1 - h_{0,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}(h_2 - h_{0,2}))f(z|h_0,\Sigma)\Big| dzdu \\
&\le \frac{1}{\bar{\sigma}^2}\int_0^1 \int |z_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}z_2| \sup_{h:\|h-h_0\|\le 1} |f(z|h,\Sigma) - f(z|h_0,\Sigma)|\ dzdu \\
&\quad + \frac{\varrho}{\bar{\sigma}^2}\int_0^1 \int \sup_{h:\|h-h_0\|\le 1} |f(z|h,\Sigma) - f(z|h_0,\Sigma)|\ dz\ du + \frac{2\varrho}{\bar{\sigma}^2}\int_0^1 \int f(z|h_0,\Sigma)\ dz\ du
\end{aligned}$$

The second term is bounded by the argument in the proof of Lemma 1. The third term is equal to $2\varrho/\bar{\sigma}^2$. So, we need only bound the first term.

$$\begin{aligned}
&\frac{1}{\bar{\sigma}^2}\int_0^1 \int |z_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}z_2| \sup_{h:\|h-h_0\|\le 1} |[f(z|h,\Sigma) - f(z|h_0,\Sigma)]|\ dz\ du \\
&\le \frac{2}{\bar{\sigma}^2}\int_0^1 \int |z_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}z_2| \sup_{h:\|h-h_0\|\le 1} f(z|h,\Sigma) dzdu \\
&\le \frac{2}{\bar{\sigma}^2}\int_0^1 \int_{z:\|z-h_0\|\le 2} |z_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}z_2|\zeta\ dz\ du \\
&\le \frac{2}{\bar{\sigma}^2}\cdot 4^k\varsigma\zeta + \frac{2}{\bar{\sigma}^2}q^{-k^2/2}\int_0^1 \int_{z:\|z-h_0\|>2} |z_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}z_2| f(z|h_0,q^{-1}\Sigma)\ dz\ du
\end{aligned}$$

The last integral expression is bounded by the absolute moment of a linear combination of multivariate normals, which exists and is bounded. The boundedness of the whole expression follows. The conclusion follows by the Dominated Convergence Theorem.

□