We consider estimation of bounds determined by multiple features of the data, such as bounds arising from two or more moment inequalities. In such problems, conventional estimators of the boundaries are known to be biased, and test statistics do not have standard limit distributions, at least at certain points in the parameter space. We use a limit of experiments framework to give a characterization of all attainable limit distributions of estimators, when the boundary of interest is identified as the minimum or maximum of a finite number of data features. We find that certain desirable properties—local asymptotic unbiasedness and regularity—cannot be achieved by any estimator. The problem is closely related to the problem of estimating the minimum of a set of normal means, and examining this simpler problem can suggest alternative procedures that have good risk properties.

1 Introduction

Bounds analysis has become an important tool for empirical work in economics and other fields (see for example Manski (2003)). The nonstandard behavior of statistical procedures in partially identified settings has led researchers to propose a host of estimators and inference techniques. In this paper, we examine one key source of difficulty in bounds analysis: in many applications, boundaries of identified sets are defined as minima or maxima of a set of data features. This situation arises, for example, when there are multiple moment inequalities that may determine a boundary.\footnote{A distinct issue arises when different data features determine upper and lower bounds, implying an ordering restriction on them. This plays an important role in recent work by Imbens and Manski (2004), and Stoye (2008), among others, and it is related to problem arising when a parameter is on the boundary of the parameter space (Andrews).} We examine whether certain desirable features of estimators and inference techniques are attainable in such problems.

When a boundary is defined by multiple features of the data, both estimation and inference become complicated. Manski and Pepper (2000) noted that standard estimators can be biased, and Haile and Tamer (2003), Kreider and Pepper (2007) have suggested bias-reduction techniques. Inference is also complicated because test statistics and related procedures have nonstandard limit

To understand the problem, we begin by studying it in a simple setting, where we observe a multivariate normal random vector with unknown mean, and interest centers on the minimum of the components of the mean vector. The minimum is a nondifferentiable function of the mean vector, and this nondifferentiability leads to two stark results: first, there exists no unbiased estimator for the minimum; and second, there exists no translation equivariant estimator. We extend an argument of Blumenthal and Cohen (1968), which shows nonexistence of an unbiased estimator of the minimum of two independent normal means. Our second finding is also related to Van der Vaart (1991b), which shows that in a semiparametric setting, regularity of an estimator plus a mild side condition implies differentiability of the estimand.

In addition to providing finite sample intuition, the multivariate normal model plays the role of a “limit experiment” for large sample analysis in regular parametric models. By the Asymptotic Representation Theorem Van der Vaart 1991a, any sequence of procedures in a regular parametric model is matched in the limit by some procedure in the multivariate normal model. This leads to our main findings, that there exist no estimator sequences that are locally asymptotically unbiased, or regular (locally equivariant). These impossibility results extend immediately to semiparametric models.

While we focus on estimation of boundaries, we also indicate some related findings for inference procedures. Again, studying the simple normal model, a classic argument of Fraser (1952) shows that confidence intervals satisfying certain conditions cannot be similar.

2 Finite Sample Theory for a Normal Model

We begin by studying a simple multivariate normal model, and develop some exact results. The finite sample results in this section will provide a key step in our large-sample analysis in the next section.

Suppose we observe a single observation on the vector \( Z \) with known variance-covariance \( \Sigma \),

\[
Z \sim N \left( h, \Sigma \right).
\]

Let \( h_1 \) denote the first element of the mean vector \( h \), and suppose we are interested in estimating some function of the parameters, \( \kappa(h) \). We assume that \( \kappa \) is non-differentiable in \( h_1 \) at some point \( h^0 \) in the parameter space. We are particularly interested in the following choice for \( \kappa \):

\[
\kappa(h) := \min \{ h_1, \ldots, h_k \}.
\]

This could be interpreted as a simple moment inequality problem, \( E m(Z, \gamma) \geq 0 \), where the moment function is defined as \( m(Z, \gamma) = Z - \ell \gamma \) for a \( k \)-vector of ones, \( \ell \). The (structural) parameter \( \gamma \) then
satisfies
\[ \gamma \leq E[Z_1] = h_1, \ldots, \gamma \leq E[Z_k] = h_k. \]
Under this example specification, the distribution of \( Z \) identifies the set \( \Gamma = (-\infty, \kappa(h)] \), and \( \kappa(h) \) defines the upper boundary of the identified set.

2.1 Nonexistence of an Unbiased Estimator

First, consider point estimation of \( \kappa(h) \). Let \( T(Z) \) denote an estimator of \( \kappa \) based on \( Z \). The estimator is unbiased if
\[ E_h[T(Z)] = \kappa(h) \quad \forall h \in \mathbb{R}^k. \]
For the case where \( Z \) is a vector of two independent normals, Blumenthal and Cohen (1968) showed that no unbiased estimator exists. Their argument extends straightforwardly to the multivariate case.

Suppose that there exists some unbiased estimator \( T(Z) \), and consider its expectation under \( h \):
\[ E_h[T(Z)] = \int T(z) f(z|h)dz, \]
where \( f(z|h) \) denotes the multivariate normal density with mean vector \( h \) and variance-covariance matrix \( \Sigma \). Through a bounding inequality for the exponential function, we can verify the uniform integrability condition that implies differentiability under the integral sign.

Hence, we may consider the derivative of \( E_h[T(Z)] \) with respect to \( h_1 \) through an interchange of the order of differentiation and integration,
\[ \frac{\partial}{\partial h_1} E_h[T(Z)] = \int T(z) \frac{\partial}{\partial h_1} f(z|h)dz. \]
It will follow that the derivative is well-defined and exists everywhere. However, if \( E_h[T(Z)] = \kappa(h) \), then it cannot be differentiable at \( h = h^0 \), which is a contradiction. So, an unbiased estimator for \( \kappa(h) \) cannot exist.\(^2\)

The impossibility of unbiased estimation arises from nondifferentiability of \( \kappa(h) \) at a single point in the parameter space. This might suggest that we could find estimators with arbitrarily small bias. However, this will come at a steep price. Doss and Sethuraman (1989) show the following remarkable result when no unbiased estimator exists: if 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).

2.2 Nonexistence of an Equivariant Estimator

A related property of estimators that will be particularly relevant in the asymptotic analysis is equivariance. We will call an estimator \( T(Z) \) of \( \kappa(h) \) equivariant (or translation equivariant) if, for any \( h \), with \( Z \sim N(h, \Sigma) \),
\[ T(Z) - \kappa(h) \sim F, \]
\(^2\)In the Appendix, we provide the formal argument that also allows for randomized estimators.
where the law $F$ does not depend on $h$. This requirement is the finite-sample analog of regularity of estimators. It turns out that a variant of the Blumenthal-Cohen argument can be used to show that no equivariant estimators exist.

Suppose that $T(Z)$ is equivariant, and consider its characteristic function under $h$: $\phi_h(u) = E_h[e^{iuT(Z)}]$. Let $\tilde{T}(z) = T(z) - \kappa(h)$. We want to show that if the distribution of $\tilde{T}(z)$ does not depend on $h$, then $\kappa(h)$ must be differentiable in $h$.

Let $\tilde{\phi}_h(u)$ be the characteristic function of $\tilde{T}(z)$ under $h$. We can write

$$\tilde{\phi}_h(u) = E_h\left[e^{iu\tilde{T}(z)}\right] = E_h\left[e^{iu(T(z) - \kappa(h))}\right] = e^{-iu\kappa(h)}\phi_h(u)$$

If $T(z)$ is equivariant, then $\tilde{\phi}_h(u)$ does not depend on $h$, so

$$\tilde{\phi}_0(u) = e^{-iu\kappa(h)}\phi_h(u)$$

where $\tilde{\phi}_0(u)$ does not depend on $h$. Write this as:

$$e^{iu\kappa(h)} = \frac{\phi_h(u)}{\tilde{\phi}_0(u)}$$

After verifying that we can differentiate under the integral sign, it is easy to show that $\phi_h(u)$ is differentiable in $h_1$, and since the left hand side is $\cos(u\kappa(h)) + i\sin(u\kappa(h))$, clearly $\kappa(h)$ must be differentiable. Thus, no translation equivariant estimator exists for $\kappa(h)$.

### 3 Local Asymptotic Theory

Although there do not exist exactly unbiased or equivariant estimators for $\kappa(h)$ 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 and examine whether estimators of boundaries can have good properties.

#### 3.1 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

$$Y_i \sim G_\pi,$$

where the $m$-dimensional parameter $\pi$ can be partitioned as $\pi = (\theta, \lambda)$, with $\theta \in \Theta \subset \mathbb{R}^k$ and $\lambda \in \Lambda \subset \mathbb{R}^{m-k}$, where $k \leq m$. We assume that both $\Theta$ and $\Lambda$ are open sets. 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.) As before, suppose interest centers on

$$\kappa(\pi) = \min\{\theta_1, \ldots, \theta_k\},$$
For example, suppose that \( Y_i \) is \( k \)-dimensional, and suppose we have parametrized the model so that \( E[Y_i] = \theta \). Then \( \kappa(\pi) \) is the minimum of the means of the components of \( Y_i \).

Under conventional smoothness conditions on the sequence of experiments \( \mathcal{E}^n = \{ G^n_\pi : \pi \in \Theta \times \Lambda \} \), 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 whether some of the \( \theta_1, \ldots, \theta_k \) are equal. For this reason, we adopt a local parametrization in which, loosely speaking, the data do not perfectly indicate which of the constraints is binding in determining the boundary. Let \( \theta_0 = (\rho_0, \ldots, \rho_0)' \) belong to \( \Theta \), and consider sequences of parameters \( \pi_n(h) \) of the form

\[
\pi_{n,h} = \left( \frac{\theta_0}{\sqrt{n}}, \lambda_0 + \frac{h(2)}{\sqrt{n}} \right),
\]

where \( h(1) \in \mathbb{R}^k \) and \( h(2) \in \mathbb{R}^{m-k} \), and \( h = (h(1)', h(2)')' \). Then

\[
\kappa(\pi_{n,h}) = \rho_0 + \frac{1}{\sqrt{n}} \min\{h_1, \ldots, h_k\}.
\]

(Our argument would also work for cases where at least two of the \( \theta_1, \ldots, \theta_k \) are local to each other, with the remaining \( \theta_j \) larger in value.)

Our approach chooses the centering parameter value \( \pi_0 \) to lie in a special subspace of the parameter, and deserves some further explanation. If we instead chose a centering such that all the elements of \( \theta_0 \) were distinct, then for a sufficiently large sample size, it would be trivial to determine which element of \( \theta \) determined \( \kappa(\theta) \), and any reasonable estimator would be unbiased and regular. This would not adequately capture the behavior of the estimator when at least some of the elements of \( \theta \) are close to each other, relative to sampling variability. Our choice of localization preserves the potential for bias in large samples.\(^3\) As we will see below, this localization exactly reproduces the special case of normally distributed data, similar to the way that the weak instrument asymptotics of Staiger and Stock (1997) reproduce the finite-sample distribution of estimators under normality. When \( h_1, \ldots, h_k \) are far apart, our approximation can capture situations where the data are very informative about which element of \( \theta \) determines \( \kappa \). We also note that this type of localization is the key case to consider for uniform inference properties Andrews and Guggenberger 2009.

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

\[
\sqrt{n}(T_n - \kappa(\pi_{n,h})) \xrightarrow{h} L_h, \tag{1}
\]

where \( \xrightarrow{h} \) indicates weak convergence under \( \pi_{n,h} \). 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

\(^3\)For a similar reason, Hirano and Porter (2008) consider a localization of a treatment assignment problem around parameter values such that the better treatment cannot be perfectly learned as sample size grows.
parameters. Many conventional results on optimality for point estimators, such as semiparametric efficiency bounds and convolution theorems, are stated for regular estimators.

In addition, we say that $T_n$ is **locally asymptotically unbiased** if, for all $h$, the laws $L_h$ have mean 0.

In order to establish formal results, we assume a standard smoothness condition for the parametric model:

**Assumption 1** (Differentiability in quadratic mean) There exist a function $s_y : \mathcal{Y} \to \mathbb{R}^m$ such that

$$\int \left[ dG_{\pi_0 + h}(y) - dG_{\pi_0}^{1/2}(y) - \frac{1}{2} h' \cdot s_y(y) dG_{\pi_0}^{1/2}(y) \right]^2 = o(\|h\|^2) \quad \text{as } h \to 0;$$

The differentiability in quadratic mean assumption implies that the model for $Y_i$ is locally asymptotically normal (see van der Vaart, 1998), with Fisher information matrix $J_0 = E_{\pi_0}[ss']$. For simplicity, assume $J_0$ is nonsingular. Informally, local asymptotic normality implies that the model

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

provides a characterization of the possible limit distributions of estimator sequences in the asymptotic problem. More precisely, by the Asymptotic Representation Theorem (van der Vaart, 1991), for any estimator $T_n$ with limit laws $L_h$ as in (1), there exists a randomized estimator $T(Z, U)$, where $Z \sim N(h, J_0^{-1})$ and $U$ is Uniform[0,1] independently of $Z$, such that

$$L_h = \mathcal{L}_h[T(Z, U) - \min\{h_1, \ldots, h_k\}],$$

where $\mathcal{L}_h[\cdot]$ denotes the law of the enclosed statistic in the normal experiment. However, a simple extension of the arguments in Section 2 (see the Appendix) show that there exist no estimators in the $N(h, J_0^{-1})$ experiment that are unbiased or equivariant. This leads to our main result:

**Theorem 2** Let $T_n$ be any sequence of estimators based on $\{Y_i\}_{i=1}^n$ that possess limit distributions as in Equation (1). Suppose Assumption 1 holds and $J_0$ is nonsingular. Then $T_n$ is not locally asymptotically unbiased and is not regular.

**Remark 3** 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.

---

4See Theorem 4.1 of van der Vaart (1991). Note that the “differentiability in the limit” condition for that theorem is satisfied by our mapping $\kappa$, even though it is not differentiable in the ordinary sense.

5Techniques like the bootstrap often use simulation to as a numerical approximation to a particular distribution of a statistic or other quantity that depends deterministically on the data. Typically the numerical approximation does not change the 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.
Remark 4 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. Because we work here in a parametric setting, we can state and show nonexistence somewhat more simply using the finite-dimensional normal limit limit experiment.

3.2 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. In this case, 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 \( G \), with \( \kappa(\theta(G)) = \min\{\theta_1(G), \ldots, \theta_k(G)\} \) as before. Choose the centering \( G_0 \) so that at least two of the \( \theta_1(G_0), \ldots \theta_k(G_0) \) are equal and equal to \( \kappa(\theta(G_0)) \). Then, provided that there exists an LAN parametric submodel of \( \mathcal{G} \) containing \( G_0 \), we can conclude that there is no locally asymptotically unbiased or regular estimator.

4 Some Implications for Inference

4.1 Nonexistence of a Similar Upper Confidence Bound

Although our main interest is in estimation, we briefly indicate a related result for inference. We start with the bivariate normal model from section 2. Suppose we want to find a statistic \( T(Z) \) satisfying:

\[ P_h \left[ T(Z) \geq \kappa(h) \right] = \beta \quad \forall h. \quad (2) \]

Then the set \( (-\infty, T(Z)] \) would be a similar (one-sided) confidence region.

Condition (2) can be satisfied by a trivial choice for \( T \). For instance, we could define \( T \) to be positive or negative infinity depending on the outcome of a \( \text{Bernoulli}(\beta) \) random variable independent of \( Z \). So, to examine the possibility of nontrivial similar confidence regions, we restrict the class of functions that \( T \) must lie in.

Note that increasing either a single component of \( E(h)(Z) \) or both components will (weakly) increase \( \kappa \). Suppose that we make a corresponding requirement of \( T \).

1. For all \( z_1, z_2, T(z_1 + \delta, z_2 + \delta) \) is monotone nondecreasing in \( \delta \).

2. If \( z_1 = \min\{z_1, z_2\} \), then

\[ T(z_1, z_2) \leq T(z_1 + \delta, z_2) \quad \text{for any } \delta > 0. \]
If \( z_2 = \min\{z_1, z_2\} \), then

\[ T(z_1, z_2) \leq T(z_1, z_2 + \delta) \text{ for any } \delta > 0. \]

Fraser (1952) shows that there exists no statistic \( T \) satisfying Conditions 1 and 2 and having the similarity property (2). (See also Dudewicz (1970) for a closely related result.) Thus, in the bivariate normal model, a valid confidence bound satisfying the mild conditions 1-2 must have greater than \( \beta \) coverage for some values of \((h_1, h_2)\).

[To be added: Asymptotic connection to Andrews and Soares (2007) class of confidence sets]

5 Conclusion

When multiple estimated constraints may determine a boundary, we find that no asymptotically unbiased or regular estimators exist. An implication is that focusing solely on bias reduction may be counterproductive, because it will eventually lead to a large bias-variance tradeoff. In addition, some of the standard concepts for defining optimal procedures, such as the convolution theorem and the concept of a best regular estimator, will not be useful, because no regular estimators exist. However, one can still compare risk functions of procedures, and order procedures by their minmax or average risk.

The normal limit experiment can be quite useful in developing alternative procedures. One can propose procedures in the simple normal model and study their risk functions. For example, Blumenthal and Cohen (1968) propose a number of alternative estimators in the normal model, and show that some of them have better bias and MSE properties. Since asymptotically sufficient statistics (such as the MLE) will match the “observed data” in the normal limit experiment, it is typically possible to construct procedures whose limit distributions match this proposal.

References


6 Appendix: Finite Sample Theory for a Multivariate Normal Model

[To Be Added]