Statistical Inference for Interval Identified Parameters

Jörg Stoye
New York University
j.stoye@nyu.edu

Abstract

This paper analyzes the construction of confidence intervals for a parameter θ_0 that is "interval identified," that is, the sampling process only reveals upper and lower bounds on θ_0 even in the limit. Analysis of inference for such parameters requires one to reconsider some fundamental issues. To begin, it is not clear which object – the parameter or the set of parameter values characterized by the bounds – should be asymptotically covered by a confidence region. Next, some straightforwardly constructed confidence intervals encounter problems because sampling distributions of relevant quantities can change discontinuously as parameter values change, leading to problems that are familiar from the pre-testing and model selection literatures. I carry out the relevant analyses for the simple model under consideration, but also emphasize the generality of problems encountered and connect developments to general themes in the rapidly developing literature on inference under partial identification. Results are illustrated with an application to the Survey of Economic Expectations.

Keywords. Partial identification, bounds, confidence regions, hypothesis testing, uniform inference, moment inequalities, subjective expectations.

1 Introduction

Analysis of partial identification is an area of recent growth in statistics and econometrics. To understand its premise, recall the classic definition of identification [16]: A parameter is *identified* if the mapping from its true value to population distributions of observables is invertible; thus, if we knew the latter distribution, we could back out the parameter value. In benevolent settings like those of this paper, identification implies that the parameter's true value can be learned as data accumulate.[1] In contrast, partial identification means that even in the limit, one will only learn some restrictions on this value. Somewhat more formally, if the parameter of interest is θ_0 and is contained in some parameter set Θ, then partial identification means that the population distribution of observables is consistent with any parameter value $\theta \in \Theta_0$, where Θ_0 is an *identified set* containing θ_0. Conventional identification ("point identification") obtains when $\Theta_0 = \{\theta_0\}$; the data generating process reveals nothing of interest if $\Theta_0 = \Theta$. Partial identification ("set identification") obtains in between.

Standard theories of (frequentist) estimation and inference presuppose point identification and require significant adaptation to be applicable to partially identified models. Estimation is the somewhat easier case because it is immediately clear that consistent estimators of θ_0 are unavailable, whereas the object Θ_0 itself is identified in the usual sense (if one thinks of the power set of Θ as a set of feasible parameter values). Questions that arise in estimating this set are typically more of a technical than a conceptual nature. Indeed, in many applications including this paper's, Θ_0 is a well-behaved set whose boundary can be parametrically characterized, so that consistent estimators of Θ_0 obtain straightforwardly. Theories of estimation for more general cases were provided in [5] and [9], among others.

The construction of confidence regions, on the other hand, raises a fundamental question. Should a confidence interval be constructed to cover (with some pre-specified probability) Θ_0 or rather θ_0? Beyond that, a specific technical problem emerges. Construction of confidence intervals typically requires estimation of the limiting sampling distribution of some criterion function or test statistic. These limiting distributions may change discontinuously as the shape of Θ_0 changes qualitatively, e.g. as Θ_0 loses measure.

[1] In general, identifiability is a necessary but not sufficient condition for learnability; e.g., consider incidental parameters or parameters that are discontinuous functions of population distributions.

To be uniformly valid in such critical regions, confidence regions have to implicitly or explicitly deal with a "model selection" or "pre-testing" problem.

This paper discusses these issues and illustrates their impact in a simple but, as it turns out, already quite subtle problem of inference under partial identification. I will discuss the methodological differences between confidence intervals for Θ_0 and for θ_0 and, for either case, provide confidence regions that deal with the aforementioned model selection problem as well as simple ones that do not. I also illustrate all of these in a simple application to real-world data. Parts of the paper have survey character; in particular, section 5.2 reprises results that were recently derived by this author elsewhere [28]. What's new is some technical arguments in section 5.1, the methodological discussion, the intuitions in sections 5.2 and 5.3, and the numerical examples. But to some degree, the purpose of the paper is to provide an entry point to a rapidly developing literature that might be of interest to members of the interval probabilities community.

2 The Setting

Consider the real-valued parameter $\theta_0 \equiv \theta(P_0)$ of a probability distribution $P_0(X)$; here P_0 is known a priori to lie in a set $\mathcal{P}$ that is characterized by ex ante constraints (maintained assumptions), and θ_0 is known to lie in $\Theta \equiv \theta(\mathcal{P})$. The nonstandard feature is that the random variable X is not completely observable, thus θ_0 may not be identifiable: even perfect knowledge of the observable aspects of P_0 might not reveal it. Assume, however, that those observable aspects identify bounds $\theta_l(P_0)$ and $\theta_u(P_0)$ s.t. $\theta_u > \theta_l$ and $\theta_0 \in [\theta_l, \theta_u]$ almost surely. The interval $\Theta_0 \equiv [\theta_l, \theta_u]$ will also be called *identified set*. Let $\Delta \equiv \theta_u - \theta_l$ denote its length.

Here is a motivating example that will later be analyzed numerically. Between 1994 and 1998, the Survey of Economic Expectations elicited worker expectations of job loss by asking the following question:

I would like you to think about your employment prospects over the next 12 months. What do you think is the percent chance that you will lose your job during the next 12 months?

Responses could be any number in $[0, 100]$; with extremely few exceptions near the extremal values, integers were chosen. The survey also elicited covariates, which will be ignored here. The quantity of interest is the population average of subjectively expected probability of job loss, a number that can alternatively be read as the aggregate expected fraction of jobs lost. 3688 of $n = 3860$ sample subjects answered the question, and the average subjective probability expressed by them was 14.8%. However, there was significant item nonresponse: 172 respondents refused to provide an answer. Their subjective expectations of job loss are naturally unknown, although they must lie between 0 and 100 percent. One could pin down an aggregate job loss expectation by making sufficiently strong assumptions about the missing data. For example, if it is assumed that data are missing completely at random, i.e. nonresponders entirely resemble responders other than by not responding, then the aggregate expectation is estimated as 14.8%. As the original data set contains covariates, one could – somewhat more sophisticatedly – assume that data are missing at random conditional on observables. Propensity score or other estimation methods would then lead to a somewhat different estimate that takes into account the distribution of covariates among nonresponders.[2] While they lead to sharp conclusions, these assumptions are very strong and may be accordingly controversial. Partial identification analysis seeks to avoid them, accepting that conclusions may become weaker as a result. An extreme example of this are *worst-case bounds*. In the present example, one could estimate such bounds on aggregate expectations by imputing answers of 0 respectively 100 for all missing data. Numerically, this leads to a lower bound of 14.1% and an upper one of 18.6%. In a next step, these bounds can be refined by re-introducing additional (but not fully identifying) information, and analyses of this kind now constitute a lively literature (see [18] or [19] for surveys). Worst-case bounds suffice to exhibit the inference problem, though, and I will be content with doing that here.

The example is an instance of the "mean with missing data" problem, about the simplest scenario of partial identification that one can think up.[3] In general, assume that X is supported on $[0, 1]$ and that the quantity of interest is $\mathbb{E}X$, but X is observable only if a second, binary random variable $D \in \{0, 1\}$ equals 1. Technically, the sampling process generates a random sample not of realizations x_i, but of realizations $(d_i, x_i d_i)$ which are informative about x_i only if $d_i = 1$. This sampling process identifies the following worst-case bounds:

$$\mathbb{E}(X|D=1)\Pr(D=1) \leq \mathbb{E}X \leq \mathbb{E}(X|D=1)\Pr(D=1) + 1 - \Pr(D=1).$$

These bounds are best possible without further as-

[2] The classic reference on these assumptions is [26]; for a textbook treatment, see [25].

[3] There are many natural examples in which pure identification analysis, i.e. characterization of bounds that are implied by identifiable quantities, amounts to a nontrivial optimization problem ([6], [12], [14], [27]).

sumptions; they are attained if all missing data equal 0 respectively 1.[4]

It is obvious that θ_0 cannot be estimated consistently. At the same time, I will impose assumptions that render trivial the problem of estimating Θ_0. Specifically, assume that estimators $\widehat{\theta}_l$ and $\widehat{\theta}_u$ exist and are uniformly jointly asymptotically normal:

$$\sqrt{n}\begin{bmatrix} \widehat{\theta}_l - \theta_l \\ \widehat{\theta}_u - \theta_u \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_l^2 & \rho\sigma_l\sigma_u \\ \rho\sigma_l\sigma_u & \sigma_u^2 \end{bmatrix}\right)$$

uniformly in $P \in \mathcal{P}$, where $\left(\sigma_l^2, \sigma_u^2, \rho\right)$ is known. Also, let $\widehat{\Delta} \equiv \widehat{\theta}_u - \widehat{\theta}_l$,

The full strength of $\sqrt{n}$-consistency and asymptotic joint normality of $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$ is required only to simplify the presentation. For example, $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$ could also converge at a nonparametric rate, and it would suffice for its distribution to be consistently estimated by the bootstrap. Similarly, assuming that $\left(\sigma_l^2, \sigma_u^2, \rho\right)$ is unknown but can be uniformly consistently estimated (as is the case in the numerical example) would only add notation and require some additional regularity conditions exhibited in [28]. The important substantive assumption that I do make is that the problem of estimating the asymptotic distribution of $\sqrt{n}\left[\widehat{\theta}_l - \theta_l, \widehat{\theta}_u - \theta_u\right]$ has been solved. This assumes away many issues which are not particular to partial identification problems. Note right away that in the motivating example, if one assumes that $\mathbb{E}\left(X|D=1\right)$ and $\Pr(D=1)$ are boundedly away from $\{0,1\}$, then the Berry-Esseen theorem implies uniform joint normality of the obvious estimators

$$\begin{aligned} \widehat{\theta}_l &= \frac{1}{n}\sum_{i=1}^{n} y_i d_i \\ \widehat{\theta}_u &= \frac{1}{n}\sum_{i=1}^{n} \left(y_i d_i + 1 - d_i\right) \\ \widehat{\Delta} &= 1 - \frac{1}{n}\sum_{i=1}^{n} d_i. \end{aligned}$$

In this application, Θ_0 would naturally be estimated by the plug-in estimator $\widehat{\Theta} \equiv \left[\widehat{\theta}_l, \widehat{\theta}_u\right]$, which was already discovered to numerically equal $[14.1\%, 18.6\%]$. I now turn to the difficult problem, namely how to compute confidence regions.

[4] In the specific example, the identified bounds can be seen as characterizing an interval probability for X. This generally occurs with missing data problems because these identify probability distributions up to contamination neighborhoods, and also in many but not all other settings of partial identification.

3 What Should a Confidence Region Cover?

If a parameter θ_0 is conventionally identified, one would like a confidence region CI to fulfil

$$\Pr(\theta_0 \in CI) \geq 1 - \alpha$$

for some pre-specified α, at least asymptotically as $n \to \infty$. Subject to this constraint, confidence regions should be short or fulfil some other desiderata. However, it is not obvious how to generalize this condition to situations of partial identification. The earlier strand of this literature aimed at the coverage condition

$$\Pr(\Theta_0 \subseteq CI) \geq 1 - \alpha,$$

thus the idea was to cover the identified set. The methodological contribution of [15] was to rather define coverage by

$$\inf_{\theta_0 \in \Theta_0} \Pr(\theta_0 \in CI) \geq 1 - \alpha,$$

i.e. to attempt coverage of the parameter. This has to be expressed in terms of an infimum over Θ_0 because it is not generally feasible to make coverage probabilities constant over Θ_0. For example, if Θ_0 has an interior, then under regularity conditions any reasonable (i.e. consistent in the Hausdorff metric) estimator $\widehat{\Theta}$ of Θ_0 covers any point in this interior with a limiting probability of 1. The probability limit of $(1-\alpha)$ must, therefore, apply only in some least favorable case that is typically attained on the boundary of Θ_0. Note the following, one-sided implication:

$$\left[\Theta_0 \subseteq CI \Longrightarrow \theta_0 \in CI\right], \forall \theta_0 \in \Theta_0.$$

Thus, if one is content with coverage of the parameter, then a confidence region for the identified set will be valid but generally conservative and therefore needlessly large. On the other hand, if one strives for coverage of the set, coverage of the parameter is simply not sufficient.

Before even attempting to define a confidence region, a researcher must decide which type of coverage is desired. The answer seems to be that it depends on whether Θ_0 or θ_0 is the ultimate object of interest. A reasonable case can be made for either, and I will now attempt to do so.[5]

[5] A superficial answer to this question would be that "it depends on the loss function." In general, one will want to cover the parameter if in the corresponding hypothesis testing problem, loss is incurred from falsely rejecting a null hypothesis about θ_0 as opposed to Θ_0. However, the analogy is not quite precise because coverage of Θ_0 can be justified from testing of compound nulls about θ_0, especially if one is interested in familywise control of the error rate. Also, this would only push back the methodological question by one level. Why, after all, is θ_0 and not Θ_0 in the loss function?

An interest in covering θ_0 seems to hinge on the premise that θ_0 is indeed a true parameter value in the sense of being descriptive of some feature of the real world in a way that other, observationally equivalent values $\theta \in \Theta_0$ are not. This presupposes what one might call a realist interpretation of one's statistical model, meaning that (i) different parameter values correspond to substantially different facts about the real world, (ii) we can on principle learn, at least in some approximate way, the truth about these facts, even though the data set at hand allow this only to a degree that is limited even beyond the usual issues of sampling variation. An analogy from physics for this setting might be that observations generated by a particular experiment generate very imprecise information about some object of interest, but this is because of limitations of measurement, e.g. the resolution of telescopes, and it is accepted that better experimental methods could on principle lead to more precise learning. Among the schools of thought that can be found within the interval probabilities community, this attitude might particularly appeal to researchers who think of interval probabilities mainly as a robustness or sensitivity tool.

In contrast, a statistician who accepts that Θ_0 is all that could ever be learned might find specious the aim of covering θ_0. This attitude would seem especially apt if the underspecified (e.g., interval) probabilities that partially identified models reveal in the limit correspond to fundamental limits to our ability to model underlying phenomena. An analogy from physics might be that observations are imprecise due to fundamental limitations as famously encountered in quantum physics. I conjecture that this attitude might particularly appeal to researchers who think of interval probabilities as a philosophical alternative to conventional probabilities, which they may think of as hopelessly optimistic.

I generally believe that both approaches have merit, and I will discuss both types of confidence regions below. In this paper's specific example, it is this author's feeling that coverage of θ_0 might have special merit. With item nonresponse in surveys, there is often a clear sense in which some precise answer to the item is a matter of fact; sometimes, this answer could even be gleaned from alternative data sources except for legal or practical reasons. (Income and age are salient examples.) In these cases, underidentification of θ_0 seems to stem from practical as opposed to epistemological problems; losses incurred by future policy decisions might well depend on θ_0 rather than Θ_0; and it might be reasonable to think of θ_0 as the quantity of ultimate interest.

4 A (Too) Straightforward Approach

The simplest extension of Wald-type confidence regions to inference on Θ_0 is the following construction which has been used frequently in the literature:

$$CI_{1-\alpha}(\Theta) = \left[\widehat{\theta}_l - \frac{c_\alpha \sigma_l}{\sqrt{n}}, \widehat{\theta}_u + \frac{c_\alpha \sigma_u}{\sqrt{n}}\right],$$

where $c_\alpha = \Phi^{-1}(1-\alpha/2)$ and Φ is the standard normal c.d.f.; e.g. $c_\alpha \approx 1.96$ for a 95%-confidence interval. In words, just enlarge the plug-in- estimator of Θ_0 by the usual number of standard errors. A Bonferroni argument establishes that

$$\begin{aligned}
& \lim_{n\to\infty} \Pr(\Theta_0 \subsetneq CI_{1-\alpha}(\Theta)) \\
= & \lim_{n\to\infty} \Pr\left(\widehat{\theta}_l - \frac{c_\alpha \sigma_l}{\sqrt{n}} > \theta_l \vee \widehat{\theta}_u + \frac{c_\alpha \sigma_u}{\sqrt{n}} < \theta_u\right) \\
\leq & \lim_{n\to\infty} \left(\Pr\left(\widehat{\theta}_l - \frac{c_\alpha \sigma_l}{\sqrt{n}} > \theta_l\right)\right. \\
& \left. + \Pr\left(\widehat{\theta}_u + \frac{c_\alpha \sigma_u}{\sqrt{n}} < \theta_u\right)\right) \\
= & \lim_{n\to\infty} \Pr\left(\frac{\sqrt{n}}{\sigma_l}\left(\widehat{\theta}_l - \theta_l\right) < c_\alpha\right) \\
& + \lim_{n\to\infty} \Pr\left(\frac{\sqrt{n}}{\sigma_l}\left(\widehat{\theta}_u - \theta_u\right) < -c_\alpha\right) \\
\to & 1 - \Phi(c_\alpha) + \Phi(-c_\alpha) = \alpha,
\end{aligned}$$

thus this interval appears valid (if potentially conservative). By the preceding section's reasoning, it must then be conservative for θ_0. Indeed, one can define a confidence region for θ_0 by using the above construction but lowering its confidence level. To see this, observe that

$$\begin{aligned}
& \lim_{n\to\infty} \Pr(\theta_0 \notin CI_{1-\alpha}(\Theta)) \\
= & \lim_{n\to\infty} \Pr\left(\widehat{\theta}_l - \frac{c_\alpha \sigma_l}{\sqrt{n}} > \theta_0 \vee \widehat{\theta}_u + \frac{c_\alpha \sigma_u}{\sqrt{n}} < \theta_0\right).
\end{aligned}$$

If $\theta_l < \theta_0 < \theta_u$, then both $\Pr\left(\widehat{\theta}_l - c_\alpha \sigma_l/\sqrt{n} > \theta_0\right)$ and $\Pr\left(\widehat{\theta}_u + c_\alpha \sigma_u/\sqrt{n} < \theta_0\right)$ vanish at exponential rate as $n \to \infty$, thus

$$\lim_{n\to\infty} \Pr(\theta_0 \notin CI_{1-\alpha}(\Theta)) = 0.$$

If $\theta_0 = \theta_l$, then this reasoning still holds for $\Pr\left(\widehat{\theta}_u + c_\alpha \sigma_u/\sqrt{n} < \theta_0\right)$, but one has

$$\begin{aligned}
& \lim_{n\to\infty} \Pr(\theta_l \notin CI_{1-\alpha}(\Theta)) \\
= & \lim_{n\to\infty} \Pr\left(\widehat{\theta}_l - \frac{c_\alpha \widehat{\sigma}_l}{\sqrt{n}} > \theta_0\right) = \alpha/2.
\end{aligned}$$

A similar reasoning applies if $\theta_0 = \theta_u$, thus

$$\lim_{n\to\infty} \inf_{\theta_0 \in \Theta_0} \Pr(\theta_0 \notin CI_{1-\alpha}(\Theta)) = \alpha/2,$$

and $CI_{1-\alpha}(\Theta)$ is a (non-conservative) $(1-\alpha/2)$ confidence interval for θ_0. Thus one can simply generate a $(1-\alpha)$ confidence interval for θ_0 by writing $CI_{1-\alpha}(\theta) = CI_{1-2\alpha}(\Theta)$. The intuition for this trick is that in the limit as $n \to \infty$, at least one end of the true identified set is far away from the true parameter value, so the hypothesis testing problem that corresponds to the confidence region is really one-sided.

5 Uniform Confidence Regions

The preceding, simple constructions may be compelling at first look, but they suffer from a severe problem: Coverage fails to be uniform over interesting regions of parameter space. This is especially easy to see with respect to coverage of θ_0. While it is true for any fixed (P_0, Θ_0) that $\lim_{n\to\infty} \inf_{\theta_0 \in \Theta_0} \Pr(\theta_0 \in CI_{1-\alpha}(\Theta)) = 1-\alpha/2$, one also finds that $\Pr(\theta_0 \in CI_{1-\alpha}(\Theta)) \to 1-\alpha$ along any local sequence of parameters where $\Delta = o(n^{-1/2})$, i.e. when Δ is asymptotically small relative to sampling error. The algebraic reason is a failure, under this condition, of the above observation that $\Pr\left(\widehat{\theta}_u + c_\alpha \sigma_u/\sqrt{n} < \theta_l\right) \to 0$. The intuitive reason is that the testing problem remains two-sided in the limit. In any case, the confidence region fails to be valid precisely when conventional identifiability of θ_0 is approached, i.e. when the underlying problem actually becomes easier.

Uniformity failures are standard in statistics. Indeed, they are unavoidable if the set of distributions $\mathcal{P}$ is large enough so that the information contained in a sample cannot be bounded away from zero, as famously demonstrated in [4]. The assumption of uniform joint normality is more than sufficient to exclude such situations, however. Accordingly, the present uniformly failure has a much more avoidable cause, namely that Δ is assumed to be large relative to standard errors. If cases of near point identification are of substantive interest, as they often will be, this assumption plainly reveals an inappropriate asymptotic framework. Indeed, were one to neglect this uniformity failure, one would be led to construct confidence intervals that *shrink* as a parameter moves from point identification to slight underidentification. I therefore now turn to constructions that are valid uniformly over possible values of Δ.

The uniformity failure in the coverage argument for θ_0, and different ways to fix the construction, have received significant attention in the literature, and relevant results will be reported. Somewhat surprisingly, $CI_{1-\alpha}(\Theta)$ has seen application even though it is not uniformly valid either. The problem can be intuitively seen as follows. Suppose that $\sigma_l = 1$ but $\sigma_u = 10$. An oracle version of $CI_{95\%}(\Theta)$ that uses infeasible knowledge of these values would be

$$CI_{95\%}(\Theta) = \left[\widehat{\theta}_l - \frac{1.96}{\sqrt{n}}, \widehat{\theta}_u + \frac{19.6}{\sqrt{n}}\right],$$

but for Δ small enough, this interval is strictly contained in the standard Wald confidence region for θ_u,

$$\left[\widehat{\theta}_u - \frac{19.6}{\sqrt{n}}, \widehat{\theta}_u + \frac{19.6}{\sqrt{n}}\right],$$

thus it cannot possibly be valid for Θ_0 in such cases. The upshot is that $CI_{1-\alpha}(\Theta_0)$ is simultaneously conservative, and hence potentially too large, under pointwise asymptotics and invalid under uniform ones, a rather unsatisfactory state of affairs.

5.1 A Confidence Region for Θ_0

If CI_α is interpreted as confidence region for Θ_0, the root cause of its uniformity failure is the same one that underlies its potential conservativeness: Its construction fails to properly account for the fact that the underlying estimation problem is bivariate. This can be fixed by an alternative construction that takes just that bivariate problem – i.e., estimation of (θ_l, θ_u) – as its starting point. Thus, define an arbitrary *joint* confidence region $CI_{1-\alpha}(\theta_l, \theta_u)$ for $\{\theta_l, \theta_u\}$. Denote by $\Theta_l \subset \mathbb{R}$ the projection of this confidence region onto the θ_l-axis and by $\Theta_u \subset \mathbb{R}$ its projection onto the θ_u-axis. Then $\lim_{n\to\infty} \Pr(\theta_l \in \Theta_l, \theta_u \in \Theta_u) \geq 1-\alpha$. Let $CI'_{1-\alpha}$ be the convex hull of $\Theta_l \cup \Theta_u$, then it follows that

$$\begin{aligned} & \lim_{n\to\infty} \Pr(\theta_l \in CI'_{1-\alpha} \wedge \theta_u \in CI'_{1-\alpha}) \geq 1-\alpha \\ \Longrightarrow \quad & \lim_{n\to\infty} \Pr([\theta_l, \theta_u] \in CI'_{1-\alpha}) \geq 1-\alpha, \end{aligned}$$

where the conclusion uses convexity of $CI'_{1-\alpha}$.

This construction will be uniformly valid as long as normal approximations apply uniformly. Of course, due to the two steps of first forming projections and then computing convex hulls, it is in general conservative, and potentially very much so. This conservatism can be avoided by appropriately choosing the initial confidence region $CI_{1-\alpha}(\theta_l, \theta_u)$. In particular, one should not pick the confidence region of smallest area, i.e. the usual confidence ellipse for bivariate normal means. A better choice is the confidence region that minimizes the length of the convex hull of its projections onto the axes. This confidence region is easily identified as the smallest one to be expressed as $[a, b]^2$ for $a, b \in \mathbb{R}$, i.e. the optimal choice for $CI_{1-\alpha}(\theta_l, \theta_u)$ is

$$\begin{aligned} & CI^*_{1-\alpha}(\theta_l, \theta_u) = \arg\min\{b-a\} \\ \text{s.t.} \quad & \int_{[a,b]^2} dF_{\mathcal{N}}\left(\widehat{\theta}_l, \widehat{\theta}_u, \sigma_l n^{-1/2}, \sigma_u n^{-1/2}, \rho\right) = 1-\alpha, \end{aligned}$$

where $F_{\mathcal{N}}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ denotes a bivariate normal distribution with the specified parameters. Write $CI^*_{1-\alpha}(\theta_l, \theta_u) = [a^*, b^*]^2$, then the convex hull of the projection of this region onto the axes is $CI^*_{1-\alpha}(\Theta) = [a^*, b^*]$, and one obtains

$$\lim_{n\to\infty} \Pr([\theta_l, \theta_u] \subseteq CI^*_{1-\alpha}(\Theta)) = 1 - \alpha$$

uniformly. This construction does not seem to appear in the relevant literature, although projection techniques were used before. In particular, [8] propose to make the initial confidence region $CI_{1-\alpha}(\theta_l, \theta_u)$ balanced, that is, to equalize each parameter's contribution to noncoverage risk. A new justification for this idea in the present context will be encountered below.[6]

5.2 A Confidence Region for θ_0

Uniform confidence regions for θ_0 were recently developed in the literature, with an initial proposal by [15], some issues with which were diagnosed and alleviated in [28]. I will here provide an intuitive development that differs from the original one but connects this section to the preceding one.

The basic idea is the same as before, namely to start from the bivariate problem of estimating (θ_l, θ_u). The difference is that as interest is in covering θ_0 and not Θ_0, the intuitive starting point would be an interval that exhibits pre-specified coverage probability for both θ_l and θ_u, but not necessarily jointly. Some tedious algebra reveals that the shortest such construction is

$$CI^*_{1-\alpha}(\theta) \equiv \left[\widehat{\theta}_l - \frac{\sigma_l c_l}{\sqrt{n}}, \widehat{\theta}_u + \frac{\sigma_u c_u}{\sqrt{n}}\right],$$

where (c_l, c_u) minimize the length of $CI^*_{1-\alpha}(\theta)$ s.t.

$$\int_{-\infty}^{c_l} \Phi\left(\frac{\rho z + c_u + \frac{\sqrt{n}\Delta}{\sigma_u}}{\sqrt{1-\rho^2}}\right) d\Phi(z) \geq 1 - \alpha \quad (1)$$

$$\int_{-\infty}^{c_u} \Phi\left(\frac{\rho z + c_l + \frac{\sqrt{n}\Delta}{\sigma_l}}{\sqrt{1-\rho^2}}\right) d\Phi(z) \geq 1 - \alpha. \quad (2)$$

(These expressions simplify if $\rho = \pm 1$.) The constraints separately calibrate coverage probabilities at θ_l and θ_u and can be generated by writing out bivariate normal approximations to sampling distributions.

There is a catch however: Expression (1-2) includes Δ, which is not known, thus I just defined an infeasible or "oracle" confidence region. In more elementary inference problems, it is routine to initially do just that and then show that estimators can be substituted for unknown population quantities. But this does not work out here. Under the joint normality assumption, one generally has $\left(\widehat{\Delta} - \Delta\right) = O(n^{-1/2})$, thus $\sqrt{n}\widehat{\Delta}$ does not converge to $\sqrt{n}\Delta$. This will not matter if $\sqrt{n}\Delta$ diverges, in which case $\sqrt{n}\widehat{\Delta}$ diverges as well, but it renders $CI^*_{1-\alpha}(\theta)$ invalid along local parameter sequences where $\sqrt{n}\Delta$ converges.

To resolve this issue, one must ensure that the estimator Δ^* of Δ substituted into (1-2) is *superefficient at zero*. More precisely, Δ^* must have the property that there exists some sequence $\{a_n\}$ that vanishes slowly (i.e., $a_n \to 0$ but $\sqrt{n}a_n \to \infty$) s.t. if the sequence $\{\Delta_n\}$ is dominated by $\{a_n\}$, then $\sqrt{n}(\Delta^* - \Delta_n) \to 0$. Verbally, Δ^* converges at a faster rate than $n^{-1/2}$ for parameter sequences Δ_n that vanish sufficiently fast, including all sequences s.t. $\Delta_n \leq O(n^{-1/2})$.

A striking finding in [28] is that $\widehat{\Delta} \equiv \widehat{\theta}_u - \widehat{\theta}_l$ itself fulfils just this condition in a rather wide set of applications, namely whenever (i) $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$ are uniformly jointly asymptotically normal, as assumed here, and (ii) $\widehat{\Delta} \geq 0$ almost surely, e.g. $\widehat{\theta}_l \leq \widehat{\theta}_u$ by construction. Thus, if estimators of upper and lower bounds are jointly asymptotically normal and are necessarily ordered in the right way, then the implied estimator of the difference between the bounds is superefficient at zero. This condition turns out to have reasonably wide applicability. Among other things, it means that the estimator $\widehat{\Delta}$ in this paper's example – the mean with missing data – is superefficient.[7]

However, there are also many cases (e.g. in [22] and [24]) where superefficiency of $\widehat{\Delta}$ will not obtain naturally. It must then be induced artificially. A simple way to do this is to shrink $\widehat{\Delta}$ toward zero, writing

$$\Delta^* = \widehat{\Delta} \cdot \mathbb{I}\{\widehat{\Delta} \geq a_n\}, \quad (3)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function and a_n is a user-specified sequence of numbers s.t. $a_n \to 0$ but $\sqrt{n}a_n \to \infty$. One of the main results in [28] is that $CI^*_{1-\alpha}(\theta)$ is uniformly valid for θ_0 upon substitution of Δ^* for Δ in (1-2).

A second, less troublesome issue with $CI^*_{1-\alpha}(\theta)$ is that it may not be well defined as written, namely if $\widehat{\theta}_l - \sigma_l c_l/\sqrt{n} > \widehat{\theta}_u + \sigma_u c_u/\sqrt{n}$, which absent superefficiency of $\widehat{\Delta}$ is an event with nonvanishing (more precisely: not uniformly vanishing) finite sample probability. This author's proposal is to leave the interval empty in such cases. This does not affect its

[6] [13] also propose a similar construction but make it symmetric about $\left\{\widehat{\theta}_l, \widehat{\theta}_u\right\}$. [15] and [28] mention $CI_{1-\alpha}(\Theta)$ as confidence region for Θ_0; in fairness, their focus is squarely elsewhere.

[7] In the specific example, superefficiency of $\widehat{\Delta}$ can also be seen heuristically. The estimator $\widehat{\Delta}$ is the sample analog of a population probability $\Delta = \Pr(D = 0)$, thus it has variance $\Delta(1-\Delta)/N$, the numerator of which vanishes as $\Delta \to 0$.

validity; hence, any other fix will lead to a needlessly long interval. It can also be interpreted as an embedded specification test: Samples which induce $\widehat{\theta}_l - \sigma_l c_l/\sqrt{n} > \widehat{\theta}_u + \sigma_u c_u/\sqrt{n}$ really cast doubt on the maintained hypothesis that $\theta_u \geq \theta_l$. Having said that, some users might not like confidence sets that can be empty. They could define $CI^*_{1-\alpha}(\theta)$ in an arbitrary manner whenever $\widehat{\theta}_l - \sigma_l c_l/\sqrt{n} > \widehat{\theta}_u + \sigma_u c_u/\sqrt{n}$. A natural solution might be to proceed as if one had learned that $\theta_u = \theta_l$, thus one could write

$$CI^*_{1-\alpha}(\theta) = \left[\widehat{\theta} - \frac{c_\alpha \sigma}{\sqrt{n}}, \widehat{\theta} + \frac{\sigma c_\alpha}{\sqrt{n}}\right],$$

where $\widehat{\theta} \equiv \left(\widehat{\theta}_l/\sigma_l^2 + \widehat{\theta}_u/\sigma_u^2\right) / \left(1/\sigma_l^2 + 1/\sigma_u^2\right)$ is a variance weighted average of $\widehat{\theta}_l$ and $\widehat{\theta}_u$ and $\sigma^2 \equiv 1/\left(1/\sigma_l^2 + 1/\sigma_u^2\right)$ is its sampling variance.

5.3 Relation to Model Selection and to Moment Inequalities

To understand the workings of $CI^*_{1-\alpha}(\theta)$, it is instructive to emphasize the model selection, or "pre-testing," issue that is lurking below the surface here. Recall that confidence regions typically correspond to hypothesis tests, that is, they can be thought of as lower contour set of some test statistic, thus collecting parameter values ξ for which the data do not reject the null hypothesis $H_0 : \theta_0 = \xi$. When constructing a confidence region for θ_0, the corresponding hypothesis test appears one-sided in the pointwise limit as $n \to \infty$ for any $\Delta > 0$, thus one seemingly gets away with lower cutoff values c_α than would be required for two-sided tests. Yet the test remains two-sided if $\Delta = 0$, in which case the confidence region would surely have to be a standard Wald confidence region. The pointwise limit distributions of relevant test statistics thus change discontinuously as $\Delta \to 0$. Of course, their true finite sampling distribution are continuous in Δ for any n. It follows that for any n, the pointwise approximations must be misleading for some Δ. This is why $C_{1-\alpha}(\theta)$ fails to be uniformly valid.

This type of problem is familiar to researchers investigating model selection or pre-testing. Essentially the same issues occur at the boundary between models that a pre-test or model selection procedure aims to separate. Indeed, one can think of the present problem as one of model selection, namely as deciding whether a point identified ($\Delta = 0$) or partially identified ($\Delta > 0$) model better describes the data. The shrinkage step (3) can then be interpreted as a pre-test that decides among these models, with $\Delta^* = 0$ indicating that point identification should be presumed.[8] A general problem with pre-tests is that their sampling error must be taken into account in subsequent inference and will frequently invalidate it. To avoid this, the test underlying $CI^*_{1-\alpha}(\theta)$ has a conservative slant. Point identification requires more conservative inference in the sense of larger cutoff values, therefore one can achieve validity (at cost of having longer confidence intervals) by erring in favor of presuming point identification. This is here implemented because the sequence a_n vanishes at a rate slower than $O(n^{-1/2})$, thus along any local sequence where $\Delta \leq O(n^{-1/2})$, point identification will eventually be presumed with probability 1. The price is that $CI^*_{1-\alpha}(\theta)$ will be uniformly valid (i.e. valid along all moving parameter sequences) and pointwise exact (i.e., not conservative under asymptotics that hold true parameter values fixed), but conservative along certain local sequences. Some features of this sort are essentially unavoidable when working with pre-tests; the question is mainly whether researchers acknowledge them or not, an issue on which [17] offer some cautionary tales.

It is also noted that upper and lower bounds on a real-valued parameter θ_0 are a special case of moment inequalities, a rather general framework that recently attracted much interest ([1], [2], [3], [7], [10], [21]). Moment inequalities occur when a true parameter value θ_0 is incompletely characterized by a set of inequalities

$$\mathbb{E}(m_j(x_i, \theta_0)) \geq 0, j = 1, \ldots, J,$$

where the expectations are population expectations and the m_j are known functions. Clearly such a set of conditions generally identifies a set, e.g. a polyhedron if the m_j are linear. This paper's scenario fits this framework as the special case of two moment inequalities

$$\begin{aligned} \mathbb{E}(\theta_0 - d_i x_i) &\geq 0 \\ \mathbb{E}(d_i x_i + 1 - d_i - \theta_0) &\geq 0. \end{aligned}$$

Many of the problems encountered for moment inequalities are just more intricate versions of the ones analyzed here. In particular, the adequate definition of confidence regions will depend on which moment inequalities bind, which can potentially be determined via a pre-test; but this will encounter the problem just described. Sure enough, numerous papers on moment inequalities ([2], [3], [7], [10], [21]; see also [11] for related ideas about compound hypothesis testing more generally) contain a step in which sample analogs of moment inequalities are shrunk toward zero, i.e. they

[8] In the specific example, the discontinuity issue could also be avoided by calibrating cutoff values through subsampling [23] although not through the bootstrap [7]. See [1] for a more general analysis of subsampling and its limits in cases of partial identification.

perform the exact trick introduced in the previous subsection.[9]

5.4 Unbiasedness of Confidence Regions

I conclude the theoretical analysis with some remarks about unbiasedness of confidence intervals under partial identification.[10] Recall that a confidence region CI for θ_0 is unbiased if $\Pr(\theta \in CI)$, seen as a function of θ, is maximized at θ_0. The corresponding concept for hypothesis tests is that the probability of rejection should be minimized on the null.[11]

Unbiasedness in this sense will not apply here. Consider first coverage of θ_0 when the identified set is $[\theta_l, \theta_u]$. Any reasonable confidence region will cover points in the interior of this set with probability approaching one and thus cannot be unbiased when the truth is $\theta_0 = \theta_u$. The situation is not better regarding coverage of Θ_0. Clearly any subset of Θ_0 will be covered more frequently than Θ_0 itself. Even excluding subsets from the comparison, problems with small sets remain. For example, as long as some noncoverage risk stems from the lower end of $[\theta_l, \theta_u]$, some set of the form $[\theta_u - \sqrt{n}\gamma, \theta_u + \sqrt{n}\gamma]$ will be covered more frequently than $[\theta_l, \theta_u]$.

It seems more promising to take a cue from compound hypothesis testing and be content with the requirement that Θ_0 is an upper contour set of $\Pr(\theta \in CI)$. Yet even this aim seems unrealistic when Δ is allowed to be small. For example, if $\Delta = n^{-1/2}$ and σ_u sufficiently exceeds σ_l, then any convex 95% confidence region for θ_u is conservative for θ_l and hence for a parameter value locally below θ_l. Unbiasedness could then only be achieved at the price of substantial conservatism, if at all. Thus, one might further weaken the unbiasedness criterion by requiring it only to hold along parameter sequences that hold $(\Delta, \sigma_l, \sigma_u)$ fixed.

With these adjustments in place, $CI^*_{1-\alpha}(\theta)$ is (asymptotically) unbiased. In particular, (1-2) bind with probability approaching 1, and in the limit, $\Pr(\theta \in CI^*_{1-\alpha}(\theta)) \geq 1 - \alpha$ on Θ_0 but $\Pr(\theta \in CI^*_{1-\alpha}(\theta)) < 1 - \alpha$ otherwise. $CI^*_{1-\alpha}(\Theta)$, on the other hand, does not fulfil the requirement because it is based on an unbalanced simultaneous confidence region for (θ_l, θ_u). If these parameters are measured with different precision, then $CI^*_{1-\alpha}(\Theta)$ will be more likely to cover the more precisely measured one because some such allocation of noncoverage risk minimizes length. As a result, if $\sigma_u > \sigma_l$, say, then some local value of the form $\theta_l - \sqrt{n}\gamma$ is covered more frequently than θ_u. This may be acceptable because it is not obvious that a confidence region designed for Θ_0 as object of interest need be unbiased for θ_0. Having said that, such unbiasedness is achieved by the balanced construction in [8], so one arguably encounters a trade-off between unbiasedness and length of confidence regions.

6 Numerical Illustrations

This section illustrates the above findings with some numerical examples. The first one is the empirical application described in section 2; the other two use artificial data. Recall that interest was in an average subjective probability of one-year-ahead job loss. Sample size is $n = 3860$; using the notation from section 2, the sample analog of $\mathbb{E}(X|D = 1)$ is 14.8% and the sample analog of $\Pr(D = 1)$, i.e. the probability of response, is 95.5%. These numbers imply that apart from their asymptotic validity, normal approximations should be expected to work well for the given sample. Simple computations establish that furthermore

$$\left(\widehat{\theta}_l, \widehat{\theta}_u, \widehat{\Delta}, \widehat{\sigma}_l, \widehat{\sigma}_u, \widehat{\rho}\right) = (14.10, 18.55, 4.45, 23.53, 29.22, 0.714)\,.$$

The estimator of the identified set and the different confidence regions then compute as follows:

$$\begin{aligned} \widehat{\Theta} &= [14.10, 18.55] \\ CI_{95\%}(\Theta_0) &= [13.36, 19.48] \\ CI_{95\%}(\theta_0) &= [13.48, 19.33] \\ CI^*_{95\%}(\Theta_0) &= [13.33, 19.45] \\ CI^*_{95\%}(\theta_0) &= [13.48, 19.33]\,. \end{aligned}$$

The results show the expected features: $CI_{5\%}(\theta_0) \subset CI_{5\%}(\Theta_0)$ (as is the case by construction), and $CI^*_{5\%}(\Theta_0)$ differs from $CI_{5\%}(\Theta_0)$ without nesting it. Having said that, the quantitative differences are small. This comes from two facts: First, in the example, $\widehat{\Delta}$ is large relative to $\widehat{\sigma}_l/\sqrt{n-1}$, so that the uniformity issues are not salient and the fixes hence marginal; indeed $CI_{5\%}(\theta_0)$ and $CI^*_{5\%}(\theta_0)$ cannot be distinguished numerically. Second, the estimators of the bounds have strong positive correlation ($\widehat{\rho} = 0.714$), so that the construction of $CI_{5\%}(\Theta_0)$ is not all that conservative.

To bring these issues a bit more to the forefront, I also generate intervals for a hypothetical dataset in which $n = 100$, I continue to assume that $\widehat{\Delta}$ is superefficient, and

$$\left(\widehat{\theta}_l, \widehat{\theta}_u, \widehat{\Delta}, \sigma_l, \sigma_u, \rho\right) = (15, 17, 2, 20, 30, -.3)\,.$$

[9] Note that $\Delta = \mathbb{E}(1-d_i)$, thus shrinking $\widehat{\Delta}$ amounts to artificially tightening the second of the above moment inequalities.

[10] I thank a referee for raising this question.

[11] None of this can here be shown for finite samples because this paper's assumptions do not restrict finite sample distributions. I therefore mean unbiasedness to apply asymptotically as $n \to \infty$; this is a nontrivial requirement because it is understood to apply to ($\sqrt{n}$-)local alternatives.

Results then are:

$$\begin{aligned}\widehat{\Theta} &= [15,17] \\ CI_{95\%}(\Theta_0) &= [11.08, 22.88] \\ CI_{95\%}(\theta_0) &= [11.71, 21.93] \\ CI^*_{95\%}(\Theta_0) &= [10.28, 22.63] \\ CI^*_{95\%}(\theta_0) &= [11.54, 22.01]\,.\end{aligned}$$

This example is somewhat rigged to showcase the effect of Δ being small. The difference between $CI_{5\%}(\Theta_0)$ and $CI^*_{5\%}(\Theta_0)$ is much larger. The former is substantially too small at its left end and must be extended to account for the large sampling variation in $\widehat{\theta}_u$. At the same time, the negative correlation means that noncoverage at the upper and lower end of the interval are likely to occur in the same samples, thus the overall probability of noncoverage is noticeably less than the sum of those two individual probabilities. This can be exploited to make the interval shorter, and it is this effect that dominates at its right end. Finally, the higher precision of $\widehat{\theta}_l$ is exploited by $CI^*_{95\%}(\Theta_0)$ to minimize interval length at the price of unbalancedness as discussed above; a balanced version of the interval would have a higher minimum as well as maximum but be longer.

The second hypothetical example features a large Δ but a very negative correlation between estimators, implying that the Bonferroni construction $CI_{95\%}(\Theta_0)$ is quite conservative. With $n = 100$ and

$$\left(\widehat{\theta}_l, \widehat{\theta}_u, \widehat{\Delta}, \sigma_l, \sigma_u, \rho\right) = (10, 20, 10, 20, 20, -.9)\,,$$

one accordingly gets

$$\begin{aligned}\widehat{\Theta} &= [10,20] \\ CI_{95\%}(\Theta_0) &= [6.08, 23.92] \\ CI_{95\%}(\theta_0) &= [6.71, 23.29] \\ CI^*_{95\%}(\Theta_0) &= [6.40, 23.59] \\ CI^*_{95\%}(\theta_0) &= [6.71, 23.29]\end{aligned}$$

and $CI^*_{95\%}(\Theta_0)$ is noticeably smaller than $CI_{95\%}(\Theta_0)$.

7 Summary and Outlook

Analysis of partial identification aims to provide conclusions which are robust, even at the price of not always being very strong. It is close in spirit and in methods to much work on interval probabilities (and also to robust Bayesian approaches). The systematic analysis of estimation and inference under partial identification is the object of a currently active literature. One general finding is that compared to well known methods that apply to conventionally identified methods, basic questions about inference have to be asked anew, and findings become substantially more nuanced.

This paper illustrated some of these issues in the very simple setting of an interval identified real-valued parameter. Inference toward an expected value when some data are missing served as motivating example that was carried out with real-world data. The issues encountered along the way range from the methodological or even philosophical to the pragmatic and quite technical. In particular, it was seen that simple asymptotic frameworks can inform misleading results, and that there are some nontrivial complications which link the inference problem to the large and growing literature on post model selection estimation and inference. Work on much more general settings than the one investigated here is under way; it encounters essentially the same problems, and then some. It is hoped that once these general theories are in place, thinking in terms of partial identification, rather than assuming away all identification problems, becomes part of many statisticians' and applies researchers' toolkit.

Acknowledgements

I thank Nick Kiefer for a question that ultimately led to the construction of $CI^*_{1-\alpha}(\Theta)$ and two referees as well as a seminar audience at Yale's statistics department for helpful comments. This paper was written while the author visited the Cowles Foundation at Yale University, whose hospitality is gratefully acknowledged. Financial support from a University Research Challenge Fund, New York University, is gratefully acknowledged.

References

[1] D.W.K. Andrews and P. Guggenberger. Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities, *Econometric Theory*, forthcoming.

[2] D.W.K. Andrews and P. Jia. Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure. Cowles Foundation Discussion Paper 1676, 2008.

[3] D.W.K. Andrews and G. Soares. Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection, Cowles Foundation Discussion Paper 1631, 2007.

[4] R.R. Bahadur and L.J. Savage. The Nonexistence of Certain Statistical Procedures in Nonparamet-

ric Problems. *Annals of Mathematical Statistics* 25:1115–1122, 1955.

[5] A. Beresteanu and F. Molinari. Asymptotic Properties for a Class of Partially Identified Models. *Econometrica*, 76:763-814, 2008.

[6] A. Beresteanu, I. Molchanov, and F. Molinari. Sharp Identification Regions in Games. CEMMAP working paper 15/08, 2008.

[7] F.A. Bugni. Bootstrap Inference in Partially Identified Models. Preprint, Northwestern University, 2007.

[8] J. Cheng and D.S. Small. Bounds on Causal Effects in Three-Arm Trials with Non-Compliance. *Journal of the Royal Statistical Society, Series B*, 68:815-836, 2006.

[9] V. Chernozhukov, H. Hong, and E.T. Tamer. Parameter Set Inference in a Class of Econometric Models. *Econometrica*, 75:1243-1284, 2007.

[10] Y. Fan and S.S. Park. Confidence Intervals for Some Partially Identified Parameters. Preprint, Vanderbilt University, 2007.

[11] P.R. Hansen. Asymptotic Tests of Composite Hypotheses. Preprint, Brown University, 2003.

[12] B.E. Honoré and E.T. Tamer. Bounds on Parameters in Panel Dynamic Discrete Choice Models. *Econometrica* 74:611–629, 2006.

[13] J.L. Horowitz and C.F. Manski. Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data. *Journal of the American Statistical Association* 95:77–84, 2000.

[14] J.L. Horowitz, C.F. Manski, M. Ponomareva, and J. Stoye: Computation of Bounds on Population Parameters When the Data are Incomplete. *Reliable Computing* 9:419-440, 2003.

[15] G. Imbens and C.F. Manski. Confidence Intervals for Partially Identified Parameters. *Econometrica*, 72:1845-1857, 2004.

[16] T. Koopmans. Identification Problems in Economic Model Construction. *Econometrica* 17:125-144, 1949.

[17] H. Leeb and B. Pötscher. Model Selection and Inference: Facts and Fiction. *Econometric Theory* 21:21-59, 2005.

[18] C.F. Manski. *Partial Identification of Probability Distributions*. Springer Verlag, 2003.

[19] C.F. Manski. *Identification for Prediction and Decision*. Harvard University Press, 2007.

[20] C.F. Manski and J. Straub. Worker Perceptions of Job Insecurity in the Mid-1990s: Evidence from the Survey of Economic Expectations. *Journal of Human Resources* 35:447-479, 2000.

[21] K. Menzel. Estimation and Inference with Many Moment Inequalities. Preprint, Massachussetts Institute of Technology, 2008.

[22] A. Pakes, J. Porter, K. Ho, and J. Ishii. Moment Inequalities and their Application. Preprint, Harvard University, 2006.

[23] J.P. Romano and A.M. Shaikh. Inference for Identifiable Parameters in Partially Identified Econometric Models. *Journal of Statistical Planning and Inference*, 138:2786-2807, 2008.

[24] A.M. Rosen. Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities. *Journal of Econometrics*, 146:107-117, 2008.

[25] P.R. Rosenbaum. *Observational Studies (2nd Edition)*. Springer Verlag, 2002.

[26] D.B. Rubin. Inference and Missing Data. *Biometrika* 63:581–592, 1976.

[27] J. Stoye. Partial Identification of Spread Parameters. Preprint, New York University, 2005.

[28] J. Stoye. More on Confidence Intervals for Partially Identified Parameters. *Econometrica*, forthcoming.