A Simple, Short, but Never-Empty Confidence Interval for Partially Identified Parameters

Jörg Stoye*

June 29, 2021

Abstract

This paper revisits the simple, but empirically salient, problem of inference on a real-valued parameter that is partially identified through upper and lower bounds with asymptotically normal estimators. A simple confidence interval is proposed and is shown to have the following properties:

- It is never empty or awkwardly short, including when the sample analog of the identified set is empty.
- It is valid for a well-defined pseudotrue parameter whether or not the model is well-specified.
- It involves no tuning parameters and minimal computation.

Computing the interval requires concentrating out one scalar nuisance parameter. In most cases, the practical result will be simple: To achieve 95% coverage, report the union of a simple 90% (!) confidence interval for the identified set and a standard 95% confidence interval for the pseudotrue parameter.

For uncorrelated estimators –notably if bounds are estimated from distinct subsamples– and conventional coverage levels, validity of this simple procedure can be shown analytically. The case obtains in the motivating empirical application (de Quidt, Haushofer, and Roth, 2018), in which improvement over existing inference methods is demonstrated. More generally, simulations suggest that the novel confidence interval has excellent length and size control. This is partly because, in anticipation of never being empty, the interval can be made *shorter* than conventional ones in relevant regions of sample space.

*Department of Economics, Cornell University, stoye@cornell.edu. Thanks to Johannes Haushofer, Jonathan de Quidt, and Chris Roth for an inquiry that motivated this work and for sharing and explaining their data. Financial support through NSF Grant SES-1824375 is gratefully acknowledged.

1 Introduction

Inference under partial identification is by now the subject of a broad literature.[1] Only recently did attention turn to the following concern: If a partially identified model is misspecified, this may manifest in either an empty or –and arguably worse– in a misleadingly small confidence region. That is, misspecified inference can be spuriously precise.

The reason is that most confidence regions used in partial identification invert tests of $H_0 : \theta \in \Theta_I$; here, θ is a parameter and Θ_I is the identified set. If H_0 is rejected at every θ, the confidence region is empty. If H_0 is barely not rejected at a few parameter values, the confidence region may be very small. This issue is empirically relevant. For example, an empty sample analog of Θ_I occurs in de Quidt, Haushofer, and Roth (2018), whose inquiry sparked the present research and whose data are reanalyzed below.

The literature on this issue is still young. Ponomareva and Tamer (2011) provide an early diagnosis. Kaido and White (2013) propose a notion of pseudotrue identified set and an estimator thereof. Molinari (2020) explains the issue in detail and highlights it as important area for further investigation. The most thorough treatment is by Andrews and Kwon (2019), who emphasize the issue's importance and provide a general inference method that avoids spurious precision and ensures coverage of a pseudotrue identified set.

The present paper is in the spirit of Andrews and Kwon (2019). I focus on the simple but empirically salient case of a scalar parameter with upper and lower bounds whose estimators are jointly asymptotically normal. That is, I revisit the setting of Imbens and Manski (2004, without their superefficiency assumption) and Stoye (2009). For this setting, I propose a confidence interval with the following features:

- It is never empty nor very short (a lower bound on its length is reported later).
- It exhibits asymptotically guaranteed coverage uniformly over the identified set and additionally for a well-defined pseudotrue parameter.
- It tends to be *shorter* than more conventional intervals in benign cases, including in the empirical application.
- It is free of tuning parameters and trivial to compute.

For target coverage of 95% and for the special case of uncorrelated estimators, e.g. in this paper's empirical application, the confidence interval can be verbally defined as follows:

- Add ± 1.64 standard errors to estimators of upper and lower bounds.
- Also compute an average of the estimators that is weighted by their standard errors, as well as the corresponding standard error. Add ± 1.96 of those standard errors to the average.

[1] See Manski (2003) for an early monograph, Tamer (2010) for a historical introductions, and Canay and Shaikh (2017) and Molinari (2020) for recent surveys that extensively cover inference.

- Report the union of the intervals.

While this paper generally proposes a somewhat less "cute" procedure with broader applicability, this specialized finding is probably the most striking part. Neither of the above two intervals is valid by itself; it is just that their coverage events are correlated in exactly the right way.

Section 2 further lays out the problem and briefly reviews important concepts. Section 2 formally develops the proposal for an asymptotic experiment and gives an intuition for why it works; the formal proof is relegated to the online appendix. Section 4 provides a numerical illustration and Section 6 an application to the data that motivated this research. Section 7 concludes.

2 A Misspecification-Adaptive Confidence Interval

Consider a setting where upper and lower bounds $\theta_L \leq \theta \leq \theta_U$ on a parameter θ can be estimated by estimators $(\hat{\theta}_L, \hat{\theta}_U)$ such that

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_L - \theta_L \\ \hat{\theta}_U - \theta_U \end{pmatrix} \stackrel{d}{\to} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & \rho\sigma_L\sigma_U \\ \rho\sigma_L\sigma_U & \sigma_U^2 \end{pmatrix}\right),$$

where $\sigma_L, \sigma_U > 0$ and consistent estimators $(\hat{\sigma}_L, \hat{\sigma}_U, \hat{\rho}) \stackrel{p}{\to} (\sigma_L, \sigma_U, \rho)$ are available.

The (asymptotic) joint normality assumption is unrestrictive if, as in the empirical application, $(\hat{\theta}_L, \hat{\theta}_U)$ are smooth functions of sample moments. It is unlikely to hold for intersection bounds (Andrews and Shi, 2013; Chernozhukov, Lee, and Rosen, 2013) and will hold for bounds that result from projecting a higher-dimensional identified set (Bugni, Canay, and Shi, 2017; Kaido, Molinari, and Stoye, 2019), including components of partially identified vectors, only in benign cases.

The obvious estimator of the *identified set* $\Theta_I \equiv [\theta_L, \theta_U]$ is $[\hat{\theta}_L, \hat{\theta}_U]$, but defining a confidence interval is delicate. Following Imbens and Manski (2004), the literature mostly focuses on confidence intervals that (asymptotically) contain the true parameter value with prespecified probability $(1-\alpha)$, irrespective of its location in Θ_I, i.e. confidence intervals that control $\inf_{\theta \in \Theta_I} \Pr(\theta \in CI)$. Finding such intervals is subtle because the nature of the testing problem qualitatively depends on the length $\Delta \equiv \theta_U - \theta_L$ of Θ_I. Heuristically, the inference problem is one-sided if Δ is "large" and two-sided if it is "short," i.e. near point identification. Ascertaining which case obtains is subject to difficulties reminiscient of post-model selection inference (Leeb and Pötscher, 2005) and parameter-on-the-boundary issues (Andrews, 2000).

The literature on how to circumvent this issue is considerable. Most approaches invert a test, that is, they report all values of θ for which $H_0 : \theta \in \Theta_I$ was not rejected. Two recurrent features of this literature are that first, accounting for a preliminary estimation step necessitates tuning parameters of Bonferroni adjustment; second, the confidence sets can be empty, namely if $\hat{\theta}_L$ is "much" larger than $\hat{\theta}_U$. This feature can be advertised as an embedded specification test but may not be wanted.[2] Arguably even more problematic is that, if the model is misspecified, a test inversion confidence interval can be short, suggesting precision when the true issue is misspecification. A specification test cannot be trusted to resolve this: Andrews and Kwon (2019) show that in "slightly misspecified" parameter regimes, spuriously precise inference generally coexists with low power of such tests.

Addressing this concern requires a more general notion of coverage. Following Andrews and Kwon (2019), define the pseudotrue identified set

$$\begin{aligned} \Theta_I^* &\equiv \Theta_I \cup \{\theta^*\} \\ \theta^* &\equiv \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U}. \end{aligned}$$

This definition is natural because $\Theta_I^* = \arg\min_\theta \max\{(\theta - \theta_U)/\sigma_U, (\theta_L - \theta)/\sigma_L, 0\}$; thus, Θ_I^* is the estimand implied by the frequent choice of $\max\{(\theta - \hat{\theta}_U)/\hat{\sigma}_U, (\hat{\theta}_L - \theta)/\hat{\sigma}_L, 0\}$ as test statistic in moment inequakity models. Note also that Θ_I^* is never empty and that $\Theta_I^* = \Theta_I$ whenever $\Theta_I \neq \emptyset$.

The revised notion of validity of a confidence interval is as follows:

Definition 1: *A confidence interval CI has asymptotic coverage of* $(1-\alpha)$ *if*

$$\lim_{n\to\infty} \inf_{\theta \in \Theta_I^*} \Pr(\theta \in CI) \geq 1 - \alpha.$$

Forcing coverage of θ^* will ensure that the interval is nonempty and also that it is statistically interpretable as targeting Θ_I^*. An obvious caveat is that, as with the related literature going back to White (1982), the coverage target's substantive relevance may not be clear if the model is in fact misspecified. As Andrews and Kwon (2019) elaborate, this has to be traded off against concerns with spurious precision.

3 Inference in the Asymptotic Experiment

Because the novel confidence interval has no tuning parameters that depend on n, it is most easily introduced in the context of an asymptotic experiment. Thus, suppose the following:

Assumption 1: *The researcher observes one realization of*

$$\begin{pmatrix} \hat{\theta}_L \\ \hat{\theta}_U \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_L \\ \theta_U \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & \rho\sigma_L\sigma_U \\ \rho\sigma_L\sigma_U & \sigma_U^2 \end{pmatrix} \right).$$

[2] That was the sales pitch in Stoye (2009), but not all referees were sold on it. The embedded specification test is analyzed in more detail by Andrews and Soares (2010). The state-of-the-art specification test for this class of models is found in Bugni, Canay, and Shi (2017).

Then the interval is heuristically defined as follows:

- Compute an interval
$$CI_{\Theta_I} \equiv \left[\hat{\theta}_L - \sigma_L c, \hat{\theta}_U + \sigma_U c\right],$$
where c will be specified later.

- Also compute the estimator
$$\hat{\theta}^* \equiv \frac{\sigma_U \hat{\theta}_L + \sigma_L \hat{\theta}_U}{\sigma_L + \sigma_U}$$
and confidence interval
$$\begin{aligned} CI_{\theta^*} &\equiv \left[\hat{\theta}^* - \hat{\sigma}^* \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right), \hat{\theta}^* + \sigma^* \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right)\right] \\ \hat{\sigma}^* &\equiv \frac{\sigma_L \sigma_U \sqrt{2 + 2\hat{\rho}}}{\sigma_L \sigma_U}. \end{aligned}$$

- Report the union $CI_{MA} \equiv CI_{\Theta_I} \cup CI_{\theta^*}$.

The critical value c will be calibrated to ensure the desired coverage. An important innovation is to not pre-estimate Δ but set it to its globally least favorable value. One might be concerned that this is excessively conservative. However, in most cases, $\hat{c} = \Phi^{-1}(1-\alpha)$, i.e. we can just use the one-sided critical value, at least to extremely high simulation accuracy. If $\rho = 0$ and for conventional coverage levels, this can be shown analytically.

The new confidence interval is obviously never empty; indeed, its length cannot drop below $2\sigma^* \Phi^{-1}(1 - \alpha/2)$. Its formal definition and theoretical justification are as follows.

Definition 2: *The misspecification-adaptive confidence interval CI_{MA} is*
$$CI_{MA} \equiv \left[\hat{\theta}_L - \tfrac{\hat{\sigma}_L}{\sqrt{n}} \hat{c}, \hat{\theta}_U + \tfrac{\hat{\sigma}_U}{\sqrt{n}} \hat{c}\right] \cup \left[\hat{\theta}^* - \tfrac{\hat{\sigma}^*}{\sqrt{n}} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right), \hat{\theta}^* + \tfrac{\hat{\sigma}^*}{\sqrt{n}} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right)\right], \tag{3.1}$$
where $\hat{c}$ is the unique value of c solving
$$\begin{aligned} &\inf_{\Delta \geq 0} \; \Pr\left(Z_1 - \Delta - c \leq 0 \leq Z_2 + c \text{ or } |Z_1 + Z_2 - \Delta| \leq \sqrt{2 + 2\rho}\,\Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right)\right) = 1 - \alpha, \\ &\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \end{aligned} \tag{3.2}$$
If $\rho = 0$ and $\sqrt{2}\Phi^{-1}(1-\alpha) \geq \Phi^{-1}(1 - \alpha/2)$, then $c = \Phi^{-1}(1-\alpha)$.

Remark 1: The condition that $\sqrt{2}\Phi^{-1}(1-\alpha) \geq \Phi^{-1}(1-\alpha/2)$ holds for $\alpha < .14$, i.e. for coverage levels of 86% or higher.

Theorem 1: *The confidence interval CI_{MA} achieves coverage of $(1-\alpha)$ in the asymptotic experiment.*

Proof. See appendix A. $\square$

An intuition for the proof is as follows. For simplicity, let $\sigma_L = \sigma_{\equiv} = \sigma$ and $\rho = 0$. Also, parameterize the true parameter value as $\theta = \lambda\theta_U + (1-\lambda)\theta_L$ for some $\lambda \in [0,1]$. If c were defined by

$$\inf_{\Delta \geq 0, \lambda \in [0,1]} \Pr\left(Z_L - \tfrac{\lambda\Delta}{\sigma} - c \leq 0 \leq Z_U + \tfrac{(1-\lambda)\Delta}{\sigma} + c \text{ or } \left| Z_L + Z_U + \tfrac{(1-2\lambda)\Delta}{\sigma} \right| \leq \sqrt{2}\Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right)\right) = 1 - \alpha,$$
$$\begin{pmatrix} Z_L \\ Z_U \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \tag{3.3}$$

the claim would follow because the above probability really just spells out the event that $\lambda\theta_U + (1-\lambda)\theta_L \in CI_{MA}$.

Now fix an arbitrary $\Delta \geq 0$ and consider continuously increasing λ from 0 to 1, i.e. moving the true θ from θ_L to θ_U. Because of the symmetry of this setup, it is compelling (and also not too hard to show) that coverage as function of λ is symmetric about .5 and is equally minimized at $\lambda \in \{0,1\} \Leftrightarrow \theta \in \{\theta_L, \theta_U\}$. It follows that, to evaluate the infimal coverage over (Δ, λ), one can as well restrict attention to $\theta = \theta_U \Leftrightarrow \lambda = 1$, and substituting this into (3.3) gets one a long way toward (3.2). The final flourish is that coverage at θ_U depends on (Δ, σ) only through Δ/σ, so that for the purpose of minimizing the expression over $\Delta \geq 0$, one can set $\sigma = 1$.

It turns out that the argument can be adapted to $\sigma_L \neq \sigma_U$ and $\rho \neq 0$. This is nontrivial because the general case does not possess the above symmetry; for given fixed Δ, coverage is *not* equally minimized at both boundary points. The key observation in this proof step is that, by moving from $\lambda = 0$ to $\lambda = 1$ along any curve defined by

$$\frac{\lambda\sigma_U + (1-\lambda)\sigma_L}{\sigma_L\sigma_U}\Delta = \text{const.},$$

one can restore the symmetry. This owes to the specific choice of θ^* through two channels. First, both CI_{Θ_I} and CI_{θ^*} separately maximize coverage at the unique possible value of θ^* that lies on any one such curve; second, it ensures that coverage at $\theta = \theta_U$ does not depend on σ_U, whereas σ_L can be set to 1 just as before.

The most delicate step is that if $\rho = 0$, coverage is provably minimized as $\Delta \to \infty$, justifying use of the one-sided critical value. To appreciate this claim, consider again CI_{Θ_I} and CI_{θ^*}. For $\alpha = .05$, the coverage of CI_{Θ_I} for either θ_L or θ_U may be as low as .9 if $\Delta = 0$ and approach .95 *from below* as $\Delta \to \infty$. The coverage of CI_{θ^*} for these values is

ρ	$\leq$0.8	0.85	0.9	0.95	0.98	0.99	1.0
$\alpha = .1$	1.28	1.29	1.31	1.36	1.44	1.54	1.64
$\alpha = .05$	1.64	1.65	1.65	1.70	1.76	1.81	1.96
$\alpha = .01$	2.33	2.33	2.33	2.34	2.40	2.43	2.58

Table 1: Critical values obtained by concentrating out $\Delta \in [0, \infty)$ for different coverages and correlations. For $\rho \in (-\infty, 0.8]$, further simulations corroborate the one-sided critical value as exact solution.

.95 at $\Delta = 0$ (where they coincide with θ^*) but rapidly decreases to 0 as Δ increases. That these effects aggregate to coverage uniformly above .95 heavily relies on specific features of the bivariate Normal distribution. The proof strategy is to show, by tedious manipulation, that coverage as function of Δ first increases and then decreases.

Expression (3.2) is numerically evaluated for different values of ρ and target coverages in Table 1. In particular, simulation suggests that c is just the one-sided critical value for ρ up to at least .8; it then gradually increases toward the two-sided critical value, which is easily seen to solve (3.2) for $\hat{\rho} = 1$.[3]

I conclude this section by establishing that CI_{MA} is bet-proof in the sense of Müller and Norets (2016) and the statistical literature cited therein.[4] Bet-proofness refers to the property that, for *some* true parameter value, the interval exhibits at least nominal coverage almost everywhere in sample space. In current notation (and dropping the "almost everywhere" disclaimer, which we will not need):

$$\sup_{(\mu_L,\mu_U)\in \boldsymbol{R}^2,} \inf_{(\nu_L,\nu_U)\in \boldsymbol{R}^2} \inf_{\theta\in\Theta_I^*} \Pr(\theta \in CI_{MA} | \hat{\mu}_L = \nu_L, \hat{\mu}_U = \nu_U) \geq 1 - \alpha. \tag{3.4}$$

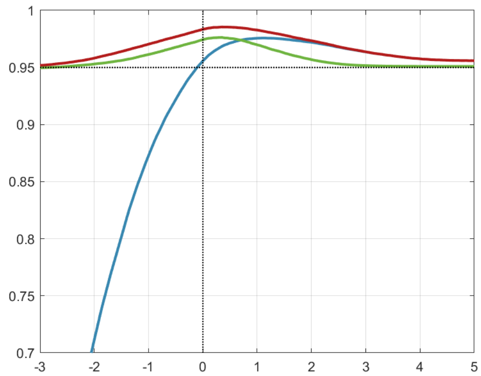
Heuristically, if this fails, an "inspector" who may bet on the event $\theta \notin CI_{MA}$ after seeing the data (but not θ) can turn an expected profit even if the bet is fair with regard to the interval's nominal coverage.[5] Even more verbally, an interval that is not bet-proof undercovers in a way that is predictable from the data. For example, no confidence interval that is empty with positive probability can be bet-proof because the inspector can bet against coverage upon seeing the empty set.

In general, bet-proofness is hard to show and appears to be an open question for some other intervals in the literature. However, an argument is available here because (i) $CI_{\theta^*} \subseteq$

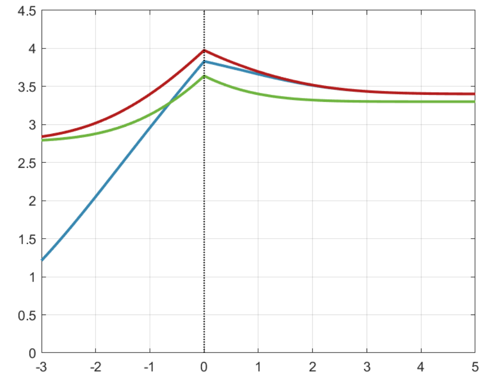
[3]The table was generated by gridding and using $B = 4000000$ simulations. This is feasible on a run-of-the-mill netbook. The relevant simulation error is the coverage error at the suggested c. Given B, it will be much smaller than what is routinely accepted in simulation-based, e.g. bootstrap, inference. For $\rho \leq .8$, further simulations establish to high accuracy that coverage is first increasing and then decreasing in Δ and minimized as $\Delta \to \infty$, the same feature that is analytically established (under conditions) for $\rho = 0$.

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

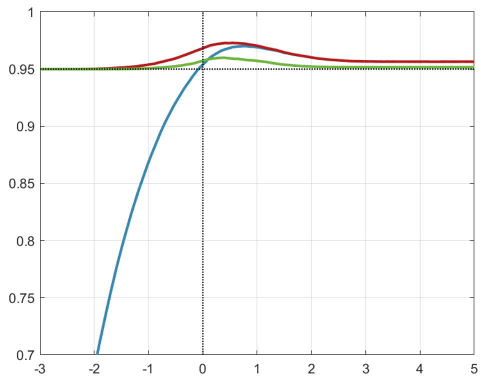
[5]Following Robinson (1977) and Müller and Norets (2016), the inspector may only bet *against* coverage. CI_{MA} tends to be conservative and so an inspector who may also bet *on* coverage can win.



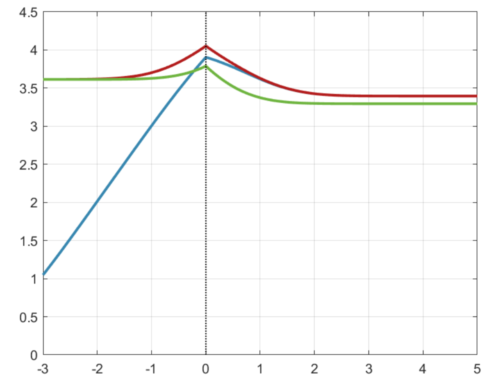
(a) Coverage when $\rho = 0$.



(b) Expected length when $\rho = 0$.



(c) Coverage when $\rho = 0.7$.



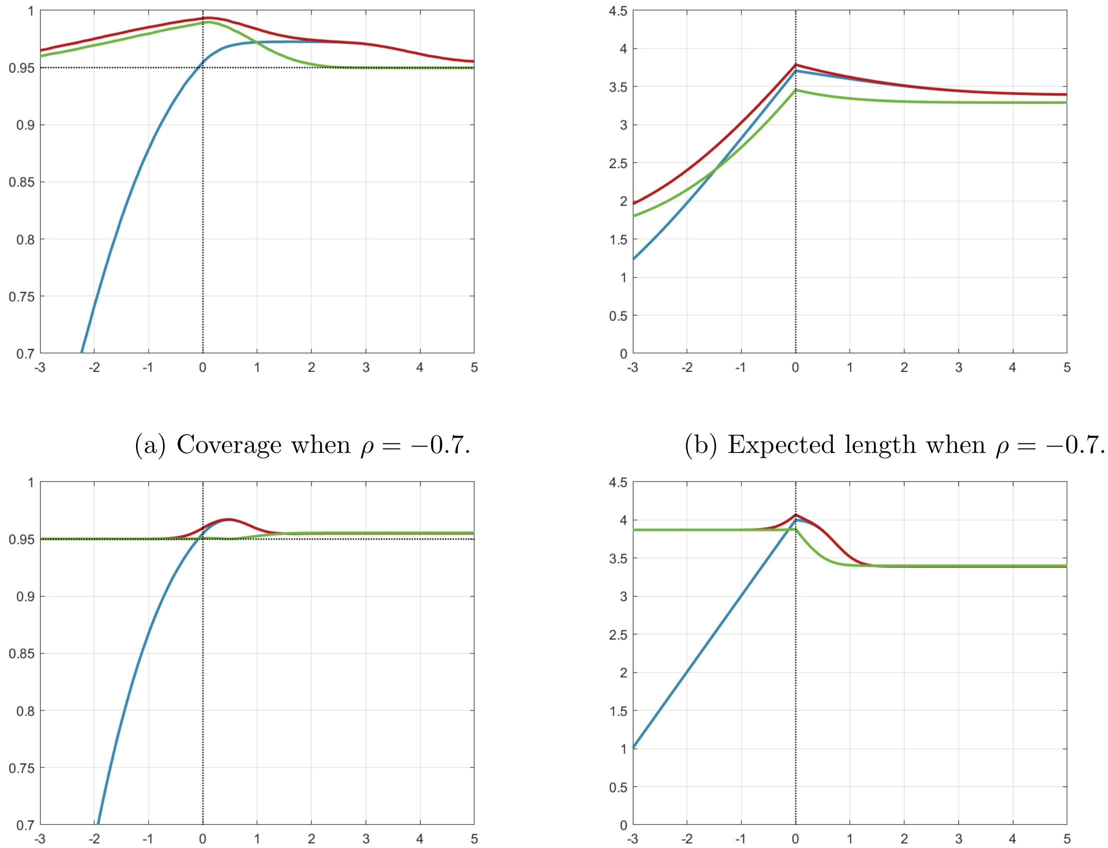
(d) Expected length when $\rho = 0.7$.

Figure 1: Coverage (left panels) and expected length (right panels; length of true interval is subtracted) of CI_{TI} (blue), $CI_{TI} \cup CI_{\theta^*}$ (red) and the new proposal CI_{MA} (green). Horizontal axis is $\Delta = \theta_U - \theta_L$; negative values indicate increasing misspecification. Nominal coverage is 95% and is indicated by a black horizontal line.

CI_{MA} and bet-proofness extends to supersets, (ii) CI_{θ^*} is bet-proof for θ^* by a result in Wallace (1959), (iii) in the point-identified case, $\Theta_I = \{\theta^*\}$. Formally:

THEOREM 2: *In the asymptotic experiment, CI_{MA} is bet-proof, i.e.* (3.4) *holds.*

Proof. See Appendix B. □



(a) Coverage when $\rho = -0.7$.

(b) Expected length when $\rho = -0.7$.

(c) Coverage when $\rho = 0.95$.

(d) Expected length when $\rho = 0.95$.

Figure 2: Continuation of Figure 1. The last case ($\rho = .95$) illustrates a setting where $\Delta \to \infty$ is not least favorable and where $\hat{c} > 1.64$.

4 Numerical Illustration

Figures 1 and 2 compare CI_{MA} with a test inversion interval CI_{TI} that arguably reflects the state of the established literature.[6] It inverts a test of $H_0 : \theta \leq \theta_U, \theta \geq \theta_L$ by taking the maximum (studentized) violation as test statistic, i.e. the same test statistic that generally implies Θ_I^* as pseudotrue identified set. The critical value is based on a pre-test –specifically, a one-sided $(.1\alpha)$-Wald test– that potentially discards one of the inequality constraints as nonbinding. Depending on the pre-test's result, the critical value is then either a simple one-sided critical value or computed by a simulation that takes ρ into account. In either case, the second-stage test is of size $.9\alpha$, so that the pre-test is accounted for by Bonferroni correction.

[6]The interval closely follows Romano, Shaikh, and Wolf (2014); other established methods (Andrews and Soares, 2010; Andrews and Barwick, 2012; Bugni, 2010; Canay, 2010) would inform similar constructions. As of writing of this manuscript, at least two rather distinct (from the preceding and from each other) proposals are in the pipeline (Andrews, Roth, and Pakes, 2019; Cox and Shi, 2020). Both invert a test and can be empty; Andrews, Roth, and Pakes (2019) also has a tuning parameter. They are compared in Cox and Shi (2020). A comparison of all these approaches in simple examples might be worthwhile.

The resulting test is inverted, and the critical value is recomputed, as θ changes, making the interval considerably shorter than early entries in the literature (Imbens and Manski, 2004; Stoye, 2009). Compared to CI_{MA}, test inversion adds orders of magnitude of computational cost, though at a very low absolute level. I abstract from asymptotic approximation by drawing estimators straight from limiting distributions and taking $(\sigma_L, \sigma_U, \rho)$ to be known. Interval length Δ is denominated in estimator standard errors because $\sqrt{n}\sigma_L = \sqrt{n}\sigma_U = 1$ throughout.

The comparison is extended into the misspecified range by letting Δ take on negative values. The test inversion interval obviously undercovers in that range. To clarify comparisons, I also compute $CI_{TI} \cup CI_{\theta^*}$. Recall that CI_{MA} can be loosely intuited as refining this construction by adjusting the critical value to account for union-taking. Nominal coverage is 95% throughout.

Figure 1 illustrates the results for $\rho = 0$ (top panels) and $\rho = .7$ (bottom panels); Figure 2 extends the exercise to $\rho = -.7$ and finally to $\rho = .95$. The last case is arguably contrived but serves to illustrate that $\Delta \to \infty$ is not always least favorable. By the same token, this is the only case in which $\hat{c} > \Phi^{-1}(.95)$.

With one caveat discussed below, the figures suggest dominating performance of CI_{MA}: It is shorter, and this is also reflected in more precise size control and thereby more power of the implied test. The advantage is especially apparent for small positive Δ. What happens here is that the correction provided by CI_{θ^*} allows CI_{MA} to transition to just adding 1.64 standard errors considerably more quickly than a pre-test could justify. Indeed, for $\rho \leq .4$, this transition occurs at a *negative* estimated interval length $\hat{\Delta}$; that is, CI_{MA} just adds 1.64 standard errors to bounds estimates whenever these are ordered in the expected way. The slight advantage of CI_{MA} for large Δ reflects that CI_{TI} accounts for a pre-test.

One might wonder how Andrews and Kwon (2019) would perform in the example. While the exact answer depends on choice of multiple tuning parameters, some qualitative considerations are as follows. Their interval starts from CI_{TI} and expands it in order to avoid spurious precision.[7] As a result, it will be bounded from below in both length and coverage by the blue curves in Figures 1 and 2. In an initial refinement, Andrews and Kwon (2019) form the union between CI_{TI} and a never-empty confidence interval. Their preferred confidence interval does this only if an additional pre-test fails to reject misspecification. While this mitigates the effect of expanding CI_{TI}, the final confidence interval still contains CI_{TI} and considerably exceeds it for small positive Δ (see their Section 8.1, whose setting resembles the present one). This will obviously be reflected in its statistical performance. Conversely, an intriguing feature of CI_{MA} is that it "spends" the "coverage capital" gained from ensuring nonemptiness by being shorter than CI_{TI} for interesting values of Δ. In fairness to Andrews and Kwon (2019), it appears far from obvious how to implement such a feature in their much

[7] Andrews and Kwon (2019) implement CI_{TI} through Andrews and Soares (2010) but point out that Romano, Shaikh, and Wolf (2014) could be used instead. The difference will be small in the present setting.

more general setting.

The advantage of CI_{MA} fades out, and even reverses, in the special case where $\rho \to 1$ but *not* $\Delta \to 0$. In that limit, c will converge to the two-sided critical value, whereas a pre-test will eventually recommend a one-sided test. In practice, if an *estimated* correlation coefficient were very high, I would advise a user to consider whether Remark 2 below applies.

5 The Interval with Estimated Parameters

A typical difficulty in partial identification is that correctly linking the finite sample setting to an asymptotic experiment is subtle: Nuisance parameters like Δ cannot be estimated with sufficient accuracy for that pre-estimation step to be ignorable. This frequently gives rise to tuning parameters. It is not an issue here because Δ is globally concentrated out. As a result, the content of this section is routine, but this fact is in itself part of the contribution.

Assume:

Assumption 2: *The d.g.p. F is contained in a set $\mathcal{F}$ such that there exist estimators $(\hat{\theta}_L, \hat{\theta}_U)$ with*

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_L - \theta_L \\ \hat{\theta}_U - \theta_U \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & \rho\sigma_L\sigma_U \\ \rho\sigma_L\sigma_U & \sigma_U^2 \end{pmatrix}\right),$$

uniformly over $F \in \mathcal{F}$, where $\sigma_L, \sigma_U \geq \underline{\sigma} > 0$ (here, underlineσ cannot depend on F) and uniformly consistent estimators $(\hat{\sigma}_L, \hat{\sigma}_U, \hat{\rho}) \xrightarrow{p} (\sigma_L, \sigma_U, \rho)$ are available.

The definition of CI_{MA} is adapted in the natural way.

Definition 3: *The misspecification-adaptive confidence interval CI_{MA} is*

$$CI_{MA} \equiv \left[\hat{\theta}_L - \tfrac{\hat{\sigma}_L}{\sqrt{n}}\hat{c}, \hat{\theta}_U + \tfrac{\hat{\sigma}_U}{\sqrt{n}}\hat{c}\right] \cup \left[\hat{\theta}^* - \tfrac{\hat{\sigma}^*}{\sqrt{n}}\Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right), \hat{\theta}^* + \tfrac{\hat{\sigma}^*}{\sqrt{n}}\Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right)\right], \tag{5.1}$$

where $\hat{c}$ is the unique value of c solving

$$\inf_{\Delta \geq 0} \ \Pr\left(Z_1 - \Delta - c \leq 0 \leq Z_2 + c \ \text{or} \ |Z_1 + Z_2 - \Delta| \leq \sqrt{2 + 2\hat{\rho}}\Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right)\right) = 1 - \alpha,$$
$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}\right). \tag{5.2}$$

If $\rho = 0$ is known and $\sqrt{2}\Phi^{-1}(1-\alpha) \geq \Phi^{-1}(1-\alpha/2)$, then $\hat{c} = \Phi^{-1}(1-\alpha)$.

Theorem 3: *The confidence interval CI_{MA} achieves uniform asymptotic coverage of $(1 - \alpha)$:*

$$\inf_{F \in \mathcal{F}} \inf_{\theta \in \Theta_I^*} \Pr(\theta \in CI_{MA}) = 1 - \alpha.$$

.

Game	$[\hat{\theta}_L, \hat{\theta}_U]$	CI_{MA}	CI_{TI}	**rel. length**
Ambiguity Aversion	[0.499,0.557]	[0.459,0.597]	[0.458,0.598]	0.97
Effort: 1 cent bonus	[0.469,0.484]	[0.448,0.503]	[0.448,0.504]	0.97
Effort: 0 cent bonus*	[0.343,0.331]	[0.318,0.356]	[0.315,0.358]	0.91
Lying**	[0.530,0.537]	[0.512,0.556]	[0.508,0.560]	0.83
Time**	[0.766,0.770]	[0.722,0.814]	[0.712,0.824]	0.82
Trust Game 1	[0.430,0.455]	[0.388,0.493]	[0.387,0.495]	0.96
Trust Game 2	[0.348,0.398]	[0.328,0.426]	[0.327,0.427]	0.97
Ultimatum Game 1	[0.443,0.470]	[0.422,0.493]	[0.422,0.494]	0.97
Ultimatum Game 2	[0.362,0.413]	[0.342,0.436]	[0.341,0.436]	0.97

Table 2: Confidence intervals applied to data in de Quidt, Haushofer, and Roth (2018, compare select columns of their Table 1). Relative length refers to relative (of CI_{MA} over CI_{TI}) excess length beyond $\max\{\hat{\Delta}, 0\}$. Of special interest: Case (*) has inverted bound estimators, displayed with abuse of interval notation. Cases (**) are short (near point identified) estimated intervals.

Proof. See appendix C. □

Remark 2: This section's setting is essentially the same as (the corresponding special case of) much of the literature. A notable outlier in this regard is Imbens and Manski (2004), who avoid tuning parameters by assuming superefficient estimation of Δ near true value 0. That condition applies if (and, in practice, only if) $\hat{\theta}_U \geq \hat{\theta}_L$ *by construction* (Stoye, 2009, Lemma 3). This case is empirically relevant: It applies to most missing-data bounds and also bounds that rely on different truncations of observed probability measures (Horowitz and Manski, 1995; Lee, 2009), unless further refinements turn these into intersection bounds. If it obtains and other regularity conditions hold, the confidence interval in Imbens and Manski (2004) is valid, is expected to be rather efficient for small Δ (because it uses superconsistency of $\hat{\Delta}$), and will obviously never be empty.[8] Fortunately, whether this case applies can be ascertained before seeing any data.

6 Empirical Application

De Quidt, Haushofer, and Roth (2018) estimate upper and lower bounds on behavioral parameters from different treatments in a between-subjects design, meaning that estimators are uncorrelated. At the same time, bounds can and did in fact invert, triggering an inquiry by the authors that led to the present paper.

Table 2 displays estimated bounds, CI_{MA}, and CI_{TI} for selected instances of the "weak bounds" data. This refers to a baseline setting before inducing experimenter demand. For

[8]Not coincidentally, this case is also characterized by the possibility of $\rho \approx 1$; indeed, that is how superconsistency of $\hat{\Delta}$ arises.

more details, I refer to de Quidt, Haushofer, and Roth (2018), particularly Figure 1 and corresponding explanations. The last column divides the length of CI_{MA} by the length of CI_{TI}, subtracting $\max\{\hat{\Delta}, 0\}$ from both. Both intervals make full use of $\rho = 0$ being known.

The comparison is between CI_{MA} and CI_{TI}; obviously, $CI_{TI} \cup CI_{\theta^*}$ would be larger than CI_{TI}. The data include one case (*) where bound estimators are inverted and where ex post, $CI_{MA} = CI_{\theta^*}$.[9] There are also two cases (**) of short estimated intervals (relative to standard errors), i.e. of near point identification. Because CI_{MA} cannot be empty, one might have conjectured it to be the longer one in these cases. In fact, it is noticeably shorter in all of them – the effect of "spending coverage capital" from the nonemptiness correction dominates. In all other cases, both intervals effectively add 1.64 standard errors.[10]

7 Conclusion

For a simple, but empirically relevant, partial identification problem, I propose a confidence interval that has competitive size control and length including in the misspecified case, while being extremely easy to compute. The most striking finding is that in many cases, a seemingly crude fix to a nominal 90% confidence interval ensures 95% coverage at little cost in terms of interval length and with practically zero computation. Simulations are encouraging, and the confidence interval improves on current best practice in application to recent lab experiments.

The approach is complementary to Andrews and Kwon (2019), from whom I take the broad motivation as well as the novel coverage requirement. Of course, their approach applies far beyond the present paper's simple setting. On the other hand, it has several tuning parameters and expands a conventional confidence interval, whereas the present proposal is tuning parameter free and compensates for expanding the conventional interval by reducing its standalone nominal coverage. A question of obvious interest, but also beyond my current reach, is whether this last feature can be usefully generalized. As it stands, the present proposal is limited to a specific setting but appears both practical and powerful when that setting obtains.

---

[9] This case would not have led any specification test to reject the model, even before taking multiple hypothesis testing into account.

[10] In those cases, the small differences favoring CI_{MA} reflect Bonferroni adjustment for pre-tests, i.e. the specifics of Romano, Shaikh, and Wolf (2014). In cases where $[\theta_L, \theta_U]$ is obviously "long," researchers will in practice be tempted to claim an asymptotic pre-test and just use 1.64 standard errors.

A Proof of Theorem 1

Step 1: Preliminaries. Validity of CI_{MA} for θ^*, including in the misspecified case, is obvious because CI_{θ^*} is valid for θ^* and is contained in CI_{MA}. We can therefore restrict attention to the case where $\theta_L \leq \theta_U$. For the rest of this prof, θ denotes a true parameter value and $\lambda \in [0,1]$ is defined by $\theta = \lambda\theta_U + (1-\lambda)\theta_L$.

Write

$$CI_{MA} = \left[\hat{\theta}_L - \sigma_L c, \hat{\theta}_U + \sigma_U c\right] \cup \left[\frac{\sigma_L\hat{\theta}_U + \sigma_U\hat{\theta}_L}{\sigma_L + \sigma_U} - \sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \frac{\sigma_L\hat{\theta}_U + \sigma_U\hat{\theta}_L}{\sigma_L + \sigma_U} + \sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right],$$

where $\sigma^* \equiv \sqrt{2+2\rho}\sigma_L\sigma_U/(\sigma_L+\sigma_U)$ is the standard deviation of $(\lambda^*\hat{\theta}_U + (1-\lambda^*)\hat{\theta}_L - \theta^*)$ and $\lambda^* \equiv \sigma_L/(\sigma_L+\sigma_U)$ is the mixture weight characterizing θ^*. Also write $\hat{\theta}_L = \theta_L + \sigma_L Z_L$ and $\hat{\theta}_U = \theta_U + \sigma_U Z_U$, where

$$\begin{pmatrix} Z_L \\ Z_U \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

We have that $\theta \in CI_{MA}$ if either

$$\begin{aligned}
& \hat{\theta}_L - \sigma_L c \leq \lambda\theta_U + (1-\lambda)\theta_L \leq \hat{\theta}_U + \sigma_U c \\
\iff\ & \hat{\theta}_L - \theta_L \leq \lambda\Delta + \sigma_L c, \quad \hat{\theta}_U - \theta_U \geq -(1-\lambda)\Delta - \sigma_U c \\
\iff\ & Z_L \leq \frac{\lambda}{\sigma_L}\Delta + c, \quad Z_U \geq -\frac{1-\lambda}{\sigma_U}\Delta - c
\end{aligned}$$

or

$$\begin{aligned}
& \frac{\sigma_L\hat{\theta}_U + \sigma_U\hat{\theta}_L}{\sigma_L+\sigma_U} - (\lambda\theta_U + (1-\lambda)\theta_L) \in \left[-\sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right] \\
\iff\ & \frac{\sigma_L\left(\theta_U + \sigma_U Z_U\right) + \sigma_U\left(\theta_L + \sigma_L Z_L\right)}{\sigma_L+\sigma_U} - (\lambda\theta_U + (1-\lambda)\theta_L) \in \left[-\sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right] \\
\iff\ & \frac{\sigma_L\sigma_U}{\sigma_L+\sigma_U}(Z_L+Z_U) + \left(\frac{\sigma_L\theta_U + \sigma_U\theta_L}{\sigma_L+\sigma_U} - (\lambda\theta_U + (1-\lambda)\theta_L)\right) \in \left[-\sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sigma^*\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right] \\
\iff\ & Z_L + Z_U + \frac{\sigma_L\theta_U + \sigma_U\theta_L - (\sigma_L+\sigma_U)(\lambda\theta_U + (1-\lambda)\theta_L)}{\sigma_L\sigma_U} \\
& \qquad \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right] \\
\iff\ & Z_L + Z_U + \frac{(1-\lambda)\sigma_L - \lambda\sigma_U}{\sigma_L\sigma_U}\Delta \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right].
\end{aligned}$$

In sum,

$$
\begin{aligned}
& \Pr(\theta \in CI_{MA}) \\
= \; & \Pr\left(\left\{Z_L - \frac{\lambda}{\sigma_L}\Delta \leq c^* \cap Z_U + \frac{1-\lambda}{\sigma_U}\Delta \geq -c^*\right\}\right. \\
& \left. \cup \left\{Z_L + Z_U + \frac{\lambda\sigma_U + (1-\lambda)\sigma_L}{\sigma_L\sigma_U}\Delta \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\right) \\
\geq \; & \inf_{\Delta \geq 0, \lambda \in [0,1]} \Pr\left(\left\{Z_L - \frac{\lambda}{\sigma_L}\Delta \leq c^* \cap Z_U + \frac{1-\lambda}{\sigma_U}\Delta \geq -c^*\right\}\right. \\
& \left. \cup \left\{Z_L + Z_U + \frac{(1-\lambda)\sigma_L - \lambda\sigma_U}{\sigma_L\sigma_U}\Delta \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\right).
\end{aligned}
$$

Step 2 shows that the above infimum equals (3.2). Step 3 shows that if $\rho = 0$, then the infimum is approximated as $\Delta \to \infty$.

Step 2: Concentrating out λ. Consider the reparameterization

$$
(X_1, X_2) \equiv \left(\frac{Z_U + Z_L}{\sqrt{2}}, \frac{Z_U - Z_L}{\sqrt{2}}\right) \Longleftrightarrow (Z_L, Z_U) = \left(\frac{X_1 - X_2}{\sqrt{2}}, \frac{X_1 + X_2}{\sqrt{2}}\right) \tag{A.1}
$$

and observe that (X_1, X_2) are uncorrelated. Algebraic manipulation yields

$$
\begin{aligned}
= \; & \Pr\left(\left\{Z_L - \frac{\lambda}{\sigma_L}\Delta \leq c^* \cap Z_U + \frac{1-\lambda}{\sigma_U}\Delta \geq -c^*\right\}\right. \\
& \left. \cup \left\{Z_L + Z_U + \frac{(1-\lambda)\sigma_L - \lambda\sigma_U}{\sigma_L\sigma_U}\Delta \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\right) \\
= \; & \Pr\left(\left\{X_1 - X_2 - \frac{\lambda}{\sigma_L}\sqrt{2}\Delta \leq \sqrt{2}c \cap X_1 + X_2 + \frac{1-\lambda}{\sigma_U}\sqrt{2}\Delta \geq -\sqrt{2}c\right\}\right. \\
& \left. \cup \left\{X_1 + \frac{(1-\lambda)\sigma_L - \lambda\sigma_U}{\sigma_L\sigma_U}\frac{\Delta}{\sqrt{2}} \in \left[-\sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\right) \\
= \; & \Pr\left(X_1 \in \left[-X_2 - \sqrt{2}c - \frac{1-\lambda}{\sigma_U}\sqrt{2}\Delta, X_2 + \sqrt{2}c + \frac{\lambda}{\sigma_L}\sqrt{2}\Delta\right]\right. \\
& \left. \cup \left[-\sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right) - \frac{(1-\lambda)\sigma_L - \lambda\sigma_U}{\sigma_L\sigma_U} \cdot \frac{\Delta}{\sqrt{2}}, \sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right) - \frac{(1-\lambda)\sigma_L - \lambda\sigma_U}{\sigma_L\sigma_U} \cdot \frac{\Delta}{\sqrt{2}}\right]\right)
\end{aligned}
$$

with the understanding that the first interval in the last expression above is empty for large negative X_2. Next consider minimizing this probability subject to the constraint that

$$
\Delta = \frac{\sigma_L\sigma_U}{\lambda\sigma_U + (1-\lambda)\sigma_L}\beta
$$

for some fixed value $\beta \geq 0$. By first substituting the constraint into the objective and then rearranging terms, we find that the minimand equals

$$\begin{aligned}
& \Pr\left(X_1 \in \left[-X_2 - \sqrt{2}c - \frac{(1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L}\sqrt{2}\beta, X_2 + \sqrt{2}c + \frac{\lambda\sigma_U}{\lambda\sigma_U + (1-\lambda)\sigma_L}\sqrt{2}\beta\right]\right. \\
& \quad \left.\cup \left[\frac{\lambda\sigma_U - (1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L} \cdot \frac{\beta}{\sqrt{2}} - \sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \frac{\lambda\sigma_U - (1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L} \cdot \frac{\beta}{\sqrt{2}} + \sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right) \\
= \; & \Pr\left(X_1 \in \left[\frac{\lambda\sigma_U - (1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L} \cdot \frac{\beta}{\sqrt{2}} - X_2 - \sqrt{2}(\beta + c), \frac{\lambda\sigma_U - (1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L} \cdot \frac{\beta}{\sqrt{2}} + X_2 + \sqrt{2}(\beta + c)\right]\right. \\
& \quad \left.\cup \left[\frac{\lambda\sigma_U - (1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L} \cdot \frac{\beta}{\sqrt{2}} - \sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \frac{\lambda\sigma_U - (1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L} \cdot \frac{\beta}{\sqrt{2}} + \sqrt{1+\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right)
\end{aligned}$$

The last expression reveals that both intervals are centered at $\frac{\lambda\sigma_U - (1-\lambda)\sigma_L}{\lambda\sigma_U + (1-\lambda)\sigma_L} \cdot \frac{\beta}{\sqrt{2}}$, an expression that increases in λ and takes value 0 at $\lambda = \lambda^*$. The intervals' length does not depend on λ, and their union coincides with the larger of the two (whose identity depends on X_2).

Now condition the above probability on X_2. Because X_1 and X_2 are uncorrelated, the conditional distribution of X_1 is Normal, centered at 0, and does not change with X_2. By log-concavity of the Normal distribution (or by taking derivatives), the conditional probability increases in λ up to λ^* and decreases in λ thereafter. Furthermore, substituting in $\lambda = 0$ or $\lambda = 1$ reveals symmetry about 0: The expressions are identical up to switching the sign of X_1. It follows that conditionally on any value of X_2 (and therefore unconditionally), the above probability is equally minimized at $\lambda \in \{0, 1\}$. If both of (Δ, λ) are concentrated out globally, one can then restrict attention to one of $\lambda = 0$ or $\lambda = 1$.

Finally, we arbitrarily choose $\lambda = 1$ and undo the reparameterization to find that the probability to be minimized is

$$\Pr\left(\left\{Z_L - \frac{\Delta}{\sigma_L} \leq c \cap Z_U \geq -c\right\} \cup \left\{Z_L + Z_U - \frac{\Delta}{\sigma_L} \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\right).$$

Since we take the infimum of this expression over $\Delta \geq 0$, we can simplify it by setting $\sigma_L = 1$, leading to ().

Step 3: For $\rho = 0$, concentrating out Δ. For the remainder of this proof, suppose $\rho = 0$. In view of step 2, also restrict attention to $\lambda = 1$ and $\sigma_L = 1$. This step's claim is that for any fixed $c \geq 0$ and α subject to condition (), the infimum

$$\begin{aligned}
& \inf_{\Delta \geq 0} \psi(\Delta), \\
& \psi(\Delta) \equiv \Pr\left(\left\{Z^L - \Delta \leq c \cap Z^U \geq -c\right\} \cup \left\{Z^L + Z^U - \Delta \in \left[-\sqrt{2}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\right)
\end{aligned}$$

is attained as $\Delta \to \infty$. The difficult part of the argument is to show that $\psi(\cdot)$ is first increasing and then decreasing (possibly, although not in fact, all increasing or all decreasing) in $\Delta \geq 0$.

Suppose this is true, then it follows that $\inf_{\Delta\geq 0}\psi(\Delta)$ is attained either at $\Delta = 0$ or as $\Delta \to \infty$. In the former case, $\theta_U = \theta^*$, so that CI_{MA} is obviously conservative. The latter limit is easily seen to equal $1-\alpha$, and this is indeed the (unattained) infimal coverage.

To show the more intricate step, write $\gamma = \sqrt{2}\Phi^{-1}(1-\alpha/2)$, then we have

$$\psi(\Delta) = \Pr\Big(\underbrace{\big\{Z_L - \Delta \leq c \cap Z_U \geq -c\big\} \cup \big\{Z_L + Z_U - \Delta \in [-\gamma, \gamma]\big\}}_{E}\Big),$$

where (Z_L, Z_U) is bivariate standard Normal. Note that the condition on critical values translates as $2c \geq \gamma$.

Using $\Phi(\cdot)$ and $\phi(\cdot)$ for the standard normal distribution and density functions, write

$$\Pr(E|Z_U = z_U) = \begin{cases} \Phi(\gamma + \Delta - z_U) - \Phi(-\gamma + \Delta - z_U), & z_U < -c \\ \Phi(\gamma + \Delta - z_U), & -c \leq z_U \leq -c + \gamma \\ \Phi(\Delta + c), & z_U > -c + \gamma \end{cases}$$

and therefore (the last step below will be elaborated after the display)

$$\begin{aligned} & \frac{d\psi(\Delta)}{d\Delta} \\ = \; & \int_{-\infty}^{-c+\gamma} \phi(\gamma + \Delta - z_U)\phi(z_U)dz_U - \int_{-\infty}^{-c} \phi(-\gamma + \Delta - z_U)\phi(z_U)dz_U + \int_{-c+\gamma}^{\infty} \phi(\Delta + c)\phi(z_U)dz_U \\ = \; & \underbrace{\sqrt{2}\left(\phi\left(\frac{\gamma+\Delta}{\sqrt{2}}\right) - \phi\left(\frac{-\gamma+\Delta}{\sqrt{2}}\right)\right)\Phi\left(\frac{\gamma - \Delta - 2c}{\sqrt{2}}\right)}_{A} + \underbrace{\phi(\Delta + c)\Phi(c - \gamma)}_{B}. \end{aligned} \tag{A.2}$$

To see the last step, note first that $\int_{-c+\gamma}^{\infty} \phi(\Delta+c)\phi(z_U)dz_U$ simplifies to B. Next, the vectors $(\gamma + \Delta - z_u, z_U)$ and $((\gamma+\Delta)/\sqrt{2}, (2z_u - \gamma - \Delta)/\sqrt{2})$ are of the same length, so that[11]

$$\begin{aligned} \int_{-\infty}^{-c+\gamma} \phi(\gamma + \Delta - z_U)\phi(z_U)dz_U &= \int_{-\infty}^{-c+\gamma} \phi\left(\frac{\gamma+\Delta}{\sqrt{2}}\right)\phi\left(\frac{2z_U - \gamma - \Delta}{\sqrt{2}}\right)dz_U \\ &= \phi\left(\frac{\gamma+\Delta}{\sqrt{2}}\right)\int_{-\infty}^{(\gamma-\Delta-2c)/\sqrt{2}} \sqrt{2}\phi(t)dt = \sqrt{2}\phi\left(\frac{\gamma+\Delta}{\sqrt{2}}\right)\Phi\left(\frac{\gamma-\Delta-2c}{\sqrt{2}}\right). \end{aligned}$$

A similar computation for $\int_{-\infty}^{-c} \phi(-\gamma + \Delta - z_U)\phi(z_U)dz_U$ and rearrangement of terms yield term A in (A.2).

Term A equals zero at $\Delta = 0$ and then becomes negative. Term B is positive throughout. Because all terms vanish as $\Delta \to \infty$, it is not useful to take further derivatives. However, we can compare the terms' relative magnitude. In particular, we will see that $|A|/|B|$ increases

[11] While not essential to the argument, it migtht be helpful to note that this transformation corresponds to a reparameterization in terms of (X_1, X_2) as in Step 2.

in Δ, hence $d\psi(\Delta)/d\Delta$ has at most one sign change and that sign change (if it occurs) is from positive to negative as claimed.

To see monotonicity of $|A|/|B|$, write

$$\begin{aligned}
\frac{|A|}{|B|} &= \sqrt{2}\cdot\frac{\phi\left(\frac{-\gamma+\Delta}{\sqrt{2}}\right)-\phi\left(\frac{\gamma+\Delta}{\sqrt{2}}\right)}{\phi(\Delta+c)}\cdot\frac{\Phi\left(\frac{\gamma-\Delta-2c}{\sqrt{2}}\right)}{\Phi(c-\gamma)}\\
&= \sqrt{2}\cdot\frac{\exp\left(-\frac{1}{4}\left(\gamma^2+\Delta^2-2\gamma\Delta\right)\right)-\exp\left(-\frac{1}{4}\left(\gamma^2+\Delta^2+2\gamma\Delta\right)\right)}{\exp\left(-\frac{1}{2}\left(\Delta^2+c^2+2\Delta c\right)\right)}\cdot\frac{\Phi\left(\frac{\gamma-\Delta-2c}{\sqrt{2}}\right)}{\Phi(c-\gamma)}\\
&= \left(\exp\left(\frac{\Delta^2}{4}+\Delta c+\frac{\gamma\Delta}{2}\right)-\exp\left(\frac{\Delta^2}{4}+\Delta c-\frac{\gamma\Delta}{2}\right)\right)\Phi\left(\frac{\gamma-\Delta-2c}{\sqrt{2}}\right)\cdot\text{const.},
\end{aligned}$$

where "const." absorbs terms that do not depend on Δ. The derivative of this expression with respect to Δ (and dropping the multiplicative constant) is

$$\begin{aligned}
&\underbrace{\left(\frac{\Delta+2c+\gamma}{2}\exp\left(\frac{\Delta^2}{4}+\Delta c+\frac{\gamma\Delta}{2}\right)-\frac{\Delta+2c-\gamma}{2}\exp\left(\frac{\Delta^2}{4}+\Delta c-\frac{\gamma\Delta}{2}\right)\right)}_{C}\Phi\left(\frac{\gamma-\Delta-2c}{\sqrt{2}}\right)\\
&-\frac{1}{\sqrt{2}}\left(\exp\left(\frac{\Delta^2}{4}+\Delta c+\frac{\gamma\Delta}{2}\right)-\exp\left(\frac{\Delta^2}{4}+\Delta c-\frac{\gamma\Delta}{2}\right)\right)\phi\left(\frac{\gamma-\Delta-2c}{\sqrt{2}}\right)\\
\geq\ &\left(\frac{\Delta+2c+\gamma}{2}\exp\left(\frac{\Delta^2}{4}+\Delta c+\frac{\gamma\Delta}{2}\right)-\frac{\Delta+2c-\gamma}{2}\exp\left(\frac{\Delta^2}{4}+\Delta c-\frac{\gamma\Delta}{2}\right)\right)\frac{\frac{\Delta+2c-\gamma}{\sqrt{2}}}{\left(\frac{\Delta+2c-\gamma}{\sqrt{2}}\right)^2+1}\phi\left(\frac{\Delta+2c-\gamma}{\sqrt{2}}\right)\\
&-\frac{1}{\sqrt{2}}\left(\exp\left(\frac{\Delta^2}{4}+\Delta c+\frac{\gamma\Delta}{2}\right)-\exp\left(\frac{\Delta^2}{4}+\Delta c-\frac{\gamma\Delta}{2}\right)\right)\phi\left(\frac{\gamma-\Delta-2c}{\sqrt{2}}\right),
\end{aligned}$$

using that $C\geq 0$ and $\Phi(-t)\geq\frac{t}{t^2+1}\phi(t)$. In order to sign this, divide through by $\phi(\ldots)$ (both are the same by symmetry of $\phi(\cdot)$) as well as by $\exp\left(\frac{\Delta^2}{4}+\Delta c-\frac{\gamma\Delta}{2}\right)$, multiply through by $\sqrt{2}$ as well as $\left(\frac{(\Delta+2c-\gamma)^2}{2}+1\right)$, and rearrange terms to conclude that the last expression above has the same sign as

$$\begin{aligned}
&\left(\frac{\Delta+2c+\gamma}{\sqrt{2}}\times\frac{\Delta+2c-\gamma}{\sqrt{2}}-\left(\frac{(\Delta+2c-\gamma)^2}{2}+1\right)\right)\exp(\gamma\Delta)\\
&\qquad-\left(\frac{\Delta+2c-\gamma}{\sqrt{2}}\times\frac{\Delta+2c-\gamma}{\sqrt{2}}-\left(\frac{(\Delta+2c-\gamma)^2}{2}+1\right)\right)\\
=\ &\left(\frac{(\Delta+2c+\gamma)(\Delta+2c-\gamma)}{2}-\frac{(\Delta+2c-\gamma)^2}{2}-1\right)\exp(\gamma\Delta)+1\\
=\ &(\gamma(\Delta+2c-\gamma)-1)\exp(\gamma\Delta)+1.
\end{aligned}$$

At $\Delta=0$, this simplifies to $\gamma(2c-\gamma)$ and therefore is nonnegative if $2c\geq\gamma$. But one can also write

$$\begin{aligned} & \frac{d}{d\Delta}\big((\gamma(\Delta+2c-\gamma)-1)\exp(\gamma\Delta)+1\big) \\ = \quad & \gamma\exp(\gamma\Delta)+(\gamma(\Delta+2c-\gamma)-1)\gamma\exp(\gamma\Delta) \\ = \quad & \gamma^2(\Delta+2c-\gamma)\exp(\gamma\Delta), \end{aligned}$$

which is again nonnegative if $2c \geq \gamma$. This concludes the proof.

B Proof of Theorem 2

The interval CI_{θ^*} is bet-proof for θ^*. Because bet-proofness is inherited by supersets, this implies that CI_{MA} is bet-proof for θ (and even for Θ_I) if $\Delta = 0 \Leftrightarrow \theta_L = \theta_U \Leftrightarrow \Theta_I = \{\theta^*\}$. Bet-proofness of CI_{θ^*}, in turn, follows from Wallace (1959, Theorem 2) upon observing that (i) CI_{θ^*} is Bayesian $(1-\alpha)$ credible with respect to the uniform improper prior on (θ_L, θ_U), (ii) this uniform prior is admissible in the sense of Wallace (1959, in words, the prior density is everywhere finite), so that the theorem applies.

C Proof of Theorem 3

To guarantee uniformity of the result, argue along a moving parameter sequence $(\theta_{L,n}, \theta_{U,n}, \sigma_{L,n}, \sigma_{U,n}, \rho_n)$. The inference problem is location invariant, so we can set $\theta_{L,n} = 0$. In case that $\theta_{U,n} - \theta_{L,n} \to \infty$, it is easy to see that coverage converges to $(1-\alpha)$ at both bounds and to a higher limit in between, thus assume that $\theta_{U,n} \leq \overline{\Delta} < \infty$. But now $(\theta_{L,n}, \theta_{U,n}, \sigma_{L,n}, \sigma_{U,n}, \rho_n)$ lives on a compact set, so that by extracting subsequences, we can restrict attention to sequences converging to accumulation points $(\theta_{L,n}, \theta_{U,n}, \sigma_{L,n}, \sigma_{U,n}, \rho_n) \to (\theta_L, \theta_U, \sigma_L, \sigma_U, \rho_n)$ where furthermore $\sigma_L, \sigma_U > 0$. As final preliminary step, if $\rho = -1$ then the choice of $c = \Phi^{-1}(1-\alpha)$ guarantees coverage because in this case, $\Pr(\Theta_I \subseteq CI_{\Theta_I}) \to 1-\alpha$ due to perfect correlation of noncoverage events at the upper and lower bound. Since the definition of c forces $c \geq \Phi^{-1}(1-\alpha)$, we can assume $\rho > -1$.

Define c as before by

$$\begin{aligned} \inf_{\Delta \geq 0, \lambda \in [0,1]} \Pr \quad & \left(\left\{ Z_L - \frac{\lambda}{\sigma_L}\Delta \leq c^* \cap Z_U + \frac{1-\lambda}{\sigma_U}\Delta \geq -c^* \right\} \right. \\ & \left. \cup \left\{ Z_L + Z_U + \left(\frac{1-\lambda}{\sigma_U} - \frac{\lambda}{\sigma_L}\right)\Delta \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right] \right\} \right), \\ & \begin{pmatrix} Z_L \\ Z_U \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \end{aligned} \tag{C.1}$$

Because we excluded $\rho = -1$, all random variables in the above expression are nondegenerate normal and all sets have interiors. Also taking into account that $\sigma_L, \sigma_U > 0$, uniform consistency of estimators implies that $\hat{c} \xrightarrow{p} c^*$ uniformly in F. Next, define

$$(\bar{\varepsilon}_L, \bar{\varepsilon}_U) \equiv \sqrt{n}\left(\frac{\hat{\theta}_L - \theta_L}{\sigma_L}, \frac{\hat{\theta}_U - \theta_U}{\sigma_U}\right)$$

and observe uniform convergence in distribution of $(\bar{\varepsilon}_L, \bar{\varepsilon}_U)$ to (Z_L, Z_U).

Borrowing some algebra from Step 1 of the proof of Theorem 1, we can then write

$$\begin{aligned}
& \Pr(\theta \in CI_{MA}) \\
= \; & \Pr\Bigg(\left\{\bar{\varepsilon}_L - \frac{\lambda}{\hat{\sigma}_L}\sqrt{n}\Delta \leq \hat{c} \cap \bar{\varepsilon}_U + \frac{1-\lambda}{\hat{\sigma}_U}\sqrt{n}\Delta \geq -c^*\right\} \\
& \cup \left\{\bar{\varepsilon}_L + \bar{\varepsilon}_U + \left(\frac{1-\lambda}{\hat{\sigma}_U} - \frac{\lambda}{\hat{\sigma}_L}\right)\sqrt{n}\Delta \in \left[-\sqrt{2+2\hat{\rho}}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\hat{\rho}}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\Bigg) \\
\to \; & \Pr\Bigg(\left\{Z_L - \frac{\lambda}{\sigma_L}\sqrt{n}\Delta \leq c^* \cap Z_U + \frac{1-\lambda}{\sigma_U}\sqrt{n}\Delta \geq -c^*\right\} \\
& \cup \left\{Z_L + Z_U + \left(\frac{1-\lambda}{\sigma_U} - \frac{\lambda}{\sigma_L}\right)\sqrt{n}\Delta \in \left[-\sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right]\right\}\Bigg) \\
\geq \; & 1 - \alpha,
\end{aligned}$$

where the convergence uses Assumption 2 as well as restrictions on $(\sigma_L, \sigma_U, \rho)$ and the next step uses the definition of c^* and observes that $\sqrt{n}\Delta$ is itself a possible value of Δ in that definition.

References

ANDREWS, D. W. K. (2000): "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space," *Econometrica*, 68(2), 399–405.

ANDREWS, D. W. K., AND P. J. BARWICK (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80(6), 2805–2826.

ANDREWS, D. W. K., AND S. KWON (2019): "Inference in Moment Inequality Models That Is Robust to Spurious Precision under Model Misspecification," *Cowles Foundation Discussion Paper CFDP 2184R*.

ANDREWS, D. W. K., AND X. SHI (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

ANDREWS, I., J. ROTH, AND A. PAKES (2019): "Inference for Linear Conditional Moment Inequalities," arXiv:1909.10062.

BUGNI, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

BUGNI, F. A., I. A. CANAY, AND X. SHI (2017): "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8(1), 1–38.

CANAY, I. (2010): "EL inference for partially identified models: large deviations optimality and bootstrap validity," *Journal of Econometrics*, 156(2), 408–425.

CANAY, I. A., AND A. M. SHAIKH (2017): *Practical and Theoretical Advances in Inference for Partially Identified Models*vol. 2 of *Econometric Society Monographs*, pp. 271–306. Cambridge University Press.

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

COX, G., AND X. SHI (2020): "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," arXiv:1907.06317.

DE QUIDT, J., J. HAUSHOFER, AND C. ROTH (2018): "Measuring and Bounding Experimenter Demand," *American Economic Review*, 108(11), 3266–3302.

Horowitz, J. L., and C. F. Manski (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63(2), 281–302.

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

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87(4), 1397–1432.

Kaido, H., and H. White (2013): *Estimating Misspecified Moment Inequality Models*pp. 331–361. Springer New York, New York, NY.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 76(3), 1071–1102.

Leeb, H., and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21(1), 21–59.

Manski, C. F. (2003): *Partial Identification of Probability Distributions (Springer Series in Statistics)*. Springer-Verlag, Berlin.

Müller, U. K., and A. Norets (2016): "Credibility of Confidence Sets in Nonstandard Econometric Problems," *Econometrica*, 84(6), 2183–2213.

Molinari, F. (2020): "Microeconometrics with Partial Identification," Handbook of econometrics, forthcoming.

Ponomareva, M., and E. Tamer (2011): "Misspecification in moment inequality models: back to moment equalities?," *The Econometrics Journal*, 14(2), 186–203.

Robinson, G. K. (1977): "Conservative Statistical Inference," *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(3), 381–386.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82(5), 1979–2002.

Stoye, J. (2009): "More on Confidence Regions for Partially Identified Parameters," *Econometrica*, 77(4), 1299–1315.

Tamer, E. (2010): "Partial Identification in Econometrics," *Annual Review of Economics*, 2(1), 167–195.

Wallace, D. L. (1959): "Conditional Confidence Level Properties," *The Annals of Mathematical Statistics*, 30(4), 864 – 876.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1–25.