

Stoye, Jörg

**Working Paper**

## A simple, short, but never-empty confidence interval for partially identified parameters

cemmap working paper, No. CWP23/21

**Provided in Cooperation with:**

Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Stoye, Jörg (2021) : A simple, short, but never-empty confidence interval for partially identified parameters, cemmap working paper, No. CWP23/21, Centre for Microdata Methods and Practice (cemmap), London,  
<https://doi.org/10.47004/wp.cem.2021.2321>

This Version is available at:

<https://hdl.handle.net/10419/241959>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A simple, short, but never-empty confidence interval for partially identified parameters

---

Jörg Stoye

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP23/21



Economic  
and Social  
Research Council

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

Jörg Stoye\*

April 22, 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.<sup>1</sup> 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,  $\theta$  is a parameter and  $\Theta_I$  is the identified set. If  $H_0$  is rejected at every  $\theta$ , 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  $\Theta_I$  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  $\pm 1.64$  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  $\pm 1.96$  of those standard errors to the average.

---

<sup>1</sup>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.<sup>2</sup> 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 formally develops the proposal and gives an intuition for why it works; the formal proof is relegated to the online appendix. Section 3 provides a numerical illustration and Section 4 an application to the data that motivated this research. Section 5 concludes.

## 2 A Misspecification-Adaptive Confidence Interval

While the interpretation of what follows is inference on a scalar parameter  $\theta$ , the only assumption is that one has well-behaved estimators of two other parameter values.

ASSUMPTION 1: *There exist estimators  $(\hat{\theta}_L, \hat{\theta}_U)$  with probability limits  $(\theta_L, \theta_U) \in \mathbf{R}^2$  such that*

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

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

The motivation is that the researcher estimates an identified set  $\Theta_I \equiv [\theta_L, \theta_U]$  containing a true parameter value  $\theta$ . Assumption 1 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  $\Theta_I$  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  $\Theta_I$ , 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  $\Theta_I$ . Heuristically, this problem is one-sided if  $\Delta$  is “large” and two-sided if it is “short,” i.e. near point identification. Ascertaining which case obtains is subject to difficulties reminiscent 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 by now considerable. Most approaches invert a test, that is, they report all values of  $\theta$  for which  $H_0 : \theta \in \Theta_I$  was not rejected.

---

<sup>2</sup>Full disclaimer: I discovered it by simulation and initially assumed a bug.

Any such confidence set can be empty; in this paper’s settings, that will happen if  $\hat{\theta}_L$  is much larger than  $\hat{\theta}_U$ , where the meaning of “much” varies across papers. This feature can be advertised as an embedded specification test but may not be wanted.<sup>3</sup> 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 will not resolve this: In this paper’s setting, the best-practice such test (Bugni, Canay, and Shi, 2015) just reports whether the test inversion interval is empty.<sup>4</sup>

Addressing this concern requires a notion of coverage for the case of misspecification, i.e. if  $\theta_L > \theta_U$ . 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,  $\Theta_I^*$  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. Note also that  $\Theta_I^*$  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*

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

Forcing coverage of  $\theta^*$  will ensure that the interval is nonempty and also that it is statistically interpretable as targeting  $\Theta_I^*$ . 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.

While the coverage notion exactly mimics Andrews and Kwon (2019), the confidence interval will be quite different. It goes “back to basics” in that, like early entries in the literature (Imbens and Manski, 2004; Stoye, 2009), it essentially just adds a certain number of standard errors to estimated bounds. An advantage is computational and conceptual simplicity; with test inversion intervals, critical values generally depend on  $\theta$  even in this simple setting and therefore must be computed many times. However, the main motivation is that the new interval performs well. Its heuristic definition is as follows:

- Compute an interval

$$CI_{\Theta_I} \equiv [\hat{\theta}_L - \frac{\hat{\sigma}_L}{\sqrt{n}} \hat{c}, \hat{\theta}_U + \frac{\hat{\sigma}_U}{\sqrt{n}} \hat{c}],$$

<sup>3</sup>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).

<sup>4</sup>This equivalence does not generalize, but Andrews and Kwon (2019) show that in “slightly misspecified” parameter regimes, spuriously precise inference generally coexists with low power of specification tests.

where  $\hat{c}$  depends on  $\alpha$  and  $\hat{\rho}$ ; see Table 1.

- Also compute the estimator

$$\hat{\theta}^* \equiv \frac{\hat{\sigma}_U \hat{\theta}_L + \hat{\sigma}_L \hat{\theta}_U}{\hat{\sigma}_L + \hat{\sigma}_U}$$

and confidence interval

$$\begin{aligned} CI_{\theta^*} &\equiv \left[ \hat{\theta}^* - \frac{\hat{\sigma}^*}{\sqrt{n}} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \hat{\theta}^* + \frac{\hat{\sigma}^*}{\sqrt{n}} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \\ \hat{\sigma}^* &\equiv \frac{\hat{\sigma}_L \hat{\sigma}_U \sqrt{2 + 2\hat{\rho}}}{\hat{\sigma}_L + \hat{\sigma}_U}. \end{aligned}$$

- Report the union  $CI_{\Theta_I} \cup CI_{\theta^*}$ .
- We will not pre-estimate  $\Delta$  but set it to its globally least favorable value. We will, however, anticipate the conservative bias ensuing from taking unions of intervals. This bias is easy to estimate; in particular, there is no parameter-on-the-boundary issue.
- One might think that concentrating out  $\Delta$  will be very conservative. It turns out that this is not so. 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  $2\hat{\sigma}^* \Phi^{-1}(1 - \alpha/2)$ . Its formal definition and theoretical justification are as follows.

DEFINITION 2: *The misspecification-adaptive confidence interval  $CI_{MA}$  is*

$$CI_{MA} \equiv \left[ \hat{\theta}_L - \frac{\hat{\sigma}_L}{\sqrt{n}} \hat{c}, \hat{\theta}_U + \frac{\hat{\sigma}_U}{\sqrt{n}} \hat{c} \right] \cup \left[ \hat{\theta}^* - \frac{\hat{\sigma}^*}{\sqrt{n}} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \hat{\theta}^* + \frac{\hat{\sigma}^*}{\sqrt{n}} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right], \quad (2.1)$$

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

$$\begin{aligned} \inf_{\Delta \geq 0} \Pr \left( Z_1 - \Delta - c \leq 0 \leq Z_2 + c \text{ or } |Z_1 + Z_2 - \Delta| \leq \sqrt{2 + 2\hat{\rho}} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right) &= 1 - \alpha, \\ \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} &\sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix} \right). \end{aligned} \quad (2.2)$$

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

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 asymptotic coverage of  $(1 - \alpha)$ .*

*Proof.* See appendix A. □

$\rho$	$\leq 0.8$	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  $\rho \leq 0.8$ , further simulations corroborate the one-sided critical value as exact solution.

Expression (2.2) is numerically evaluated for different values of  $\rho$  and target coverages in Table 1. In particular, simulation with very high accuracy suggests that  $\hat{c}$  is just the one-sided critical value for  $\rho$  up to at least .8; it then gradually increases toward the two-sided critical value, which is easily seen to solve (2.2) for  $\hat{\rho} = 1$ .<sup>5</sup>

REMARK 2: Except for large positive  $\rho$ , infimal coverage of  $(1 - \alpha)$  is attained in the limit as  $\Delta \rightarrow \infty$ . For finite  $\Delta$ ,  $CI_{MA}$  is therefore nominally conservative.

In principle, one could try to capture this by concentrating out  $\Delta$  over a more limited range, e.g. over a  $(1 - .1\alpha)$ -confidence interval with Bonferroni adjustment of subsequent inference. This could in principle lead to  $\hat{c} < \Phi^{-1}(1 - \alpha)$ . I do not advocate it because numerically, the infimum in (2.2) is well approximated for surprisingly small values of  $\Delta$ . Therefore, the “inferential cost” of a pre-test, whether through adjustment of second-stage test size or through reliance on a tuning parameter, would typically not be recovered.

REMARK 3: The literature on partial identification often focuses on uniform inference. This is because naïve inference methods may fail in cases of interest, e.g. as one approaches point identification. To prevent this, the literature has an informal requirement that inference be uniform over delicate nuisance parameters like (in this paper)  $\Delta$ ; see Molinari (2020, Section 4.3.2) for further discussion.  $CI_{MA}$  is obviously uniform in this sense because  $\Delta$  (and also the position of  $\theta$  in  $\Theta_I$ ) is set to its globally least favorable value.

To formally claim that inference is uniform over a large class of data generating processes, one would furthermore have to strengthen Assumption 1 so that consistency and asymptotic normality of bound estimators hold in a uniform sense. The exact nature of such strengthenings, and low-level assumptions that achieve them, are well understood (Andrews and Soares, 2010; Romano and Shaikh, 2008) and are omitted for brevity.

REMARK 4: The notable difference in setting to Imbens and Manski (2004) is the absence of an implicit superefficiency condition on  $\hat{\Delta}$  near true value 0. That condition turns out to

<sup>5</sup>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  $\hat{c}$ . 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  $\Delta$  and minimized as  $\Delta \rightarrow \infty$ , the same feature that is analytically proved for  $\rho = 0$  and which justifies  $\hat{c} = \Phi^{-1}(1 - \alpha)$ .



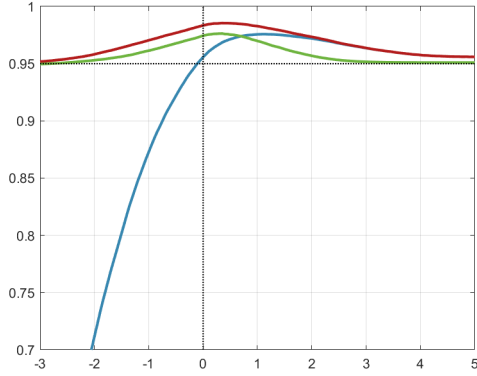
obtain if (and, in practice, only if)  $\hat{\theta}_U \geq \hat{\theta}_L$  *by construction* (Stoye, 2009, Lemma 3). This case is empirically relevant: It applies to most missing-data bounds and also bounds that rely on different truncations of observed probability measures (Horowitz and Manski, 1995; 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  $\Delta$  (because it uses superconsistency of  $\hat{\Delta}$ ), and will obviously never be empty. Not coincidentally, this case is also characterized by the possibility of  $\rho \approx 1$ ; indeed, that is how superconsistency of  $\hat{\Delta}$  arises. Whether this case applies can be ascertained before seeing any data, and I strongly suggest that users do so.

REMARK 5: I follow the bulk of the literature in focusing on uniform coverage of  $\theta \in \Theta_I^*$ . The procedure is easily adapted to coverage of the entire set  $\Theta_I^*$ . Note that, by a Bonferroni argument, a critical value of  $\hat{c} = \Phi^{-1}(1 - \alpha/2)$  would always do, and also that (as can be seen from considering large  $\Delta$ ) only a large negative value of  $\hat{\rho}$  would cause  $\hat{c}$  to be appreciably lower.

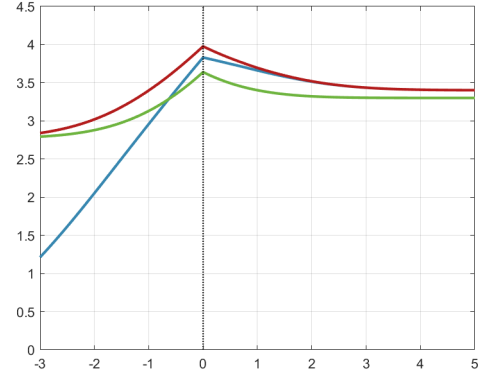
The proof of Theorem 1 contains three steps. First, it is relatively routine to show that  $CI_{MA}$  would be valid if, in line with the heuristic definition, expression (2.2) explicitly took the infimum also over values of  $(\sigma_L, \sigma_U)$  as well as  $\theta \in \Theta_I$ . In a second step, we can concentrate out all of these. In particular, one can restrict attention to one of  $\theta = \theta_L$  or  $\theta = \theta_U$ ; expression (2.2) arbitrarily chooses the latter. This finding is not obvious: For given  $\Delta$ , coverage is *not* equally minimized at the interval’s endpoints; it is only that the corresponding infima over  $\Delta \in [0, \infty)$  are the same. As final flourish in this step, it turns out that asymptotic coverage at  $\theta_U$  depends on  $(\Delta, \sigma_L, \sigma_U)$  only through  $\Delta/\sigma_L$ . For the purpose of evaluating worst-case coverage over  $\Delta \geq 0$ , we can therefore set both standard deviations to 1.

The final, and by far most delicate, step is that if  $\rho = 0$ , coverage is provably minimized as  $\Delta \rightarrow \infty$ , justifying use of the one-sided critical value  $\hat{c} = \Phi^{-1}(1 - \alpha)$ . To appreciate this claim, consider again the two components of  $CI_{MA}$  in (2.1). For  $\alpha = .05$ , the left-hand interval’s coverage for either  $\theta_L$  or  $\theta_U$  may be as low as .9 if  $\Delta = 0$  and approach .95 *from below* as  $\Delta \rightarrow \infty$ . The right-hand interval’s coverage of these values is .95 at  $\Delta = 0$  (where both coincide with  $\theta^*$ ) but rapidly decreases to 0 as  $\Delta$  increases. That these effects aggregate to coverage uniformly above .95 is far from obvious and heavily relies on specific features of the bivariate Normal distribution.

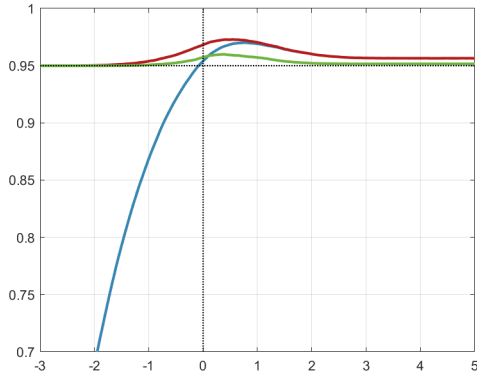
Numerically, the final step extends to moderate  $\rho$  (see again Table 1), and the proof uses conservative bounds. Some analytic result of higher generality might, therefore, be available. However, for large positive  $\rho$ , coverage is minimized at small positive  $\Delta$ . Therefore, if  $\rho$  is unknown, estimating it cannot be avoided. In particular, in view of Table 1, a pre-test for “small enough”  $\rho$  would be counterproductive: Since  $\hat{c}$  as a function of  $\rho$  is mostly completely flat, one would be unlikely to recover the inferential cost (in the sense of Remark 2) of the



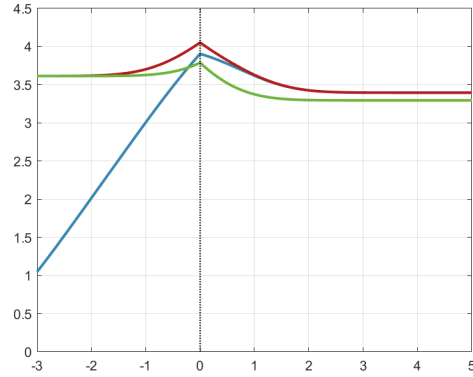
(a) Coverage when  $\rho = 0$ .



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



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



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

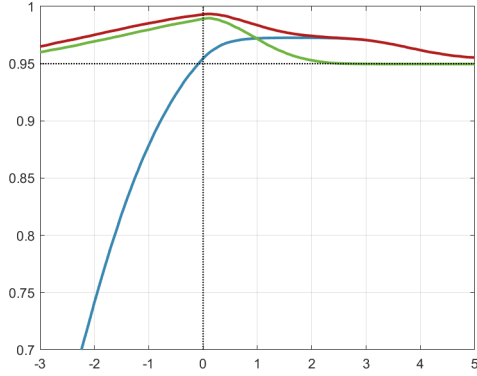
Figure 1: Coverage (left panels) and expected length (right panels; length of true interval is subtracted) of  $CI_{TI}$  (blue),  $CI_{TI} \cup CI_{\theta^*}$  (red) and the new proposal  $CI_{MA}$  (green). Horizontal axis is  $\Delta = \theta_U - \theta_L$ ; negative values indicate increasing misspecification.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

pre-test.

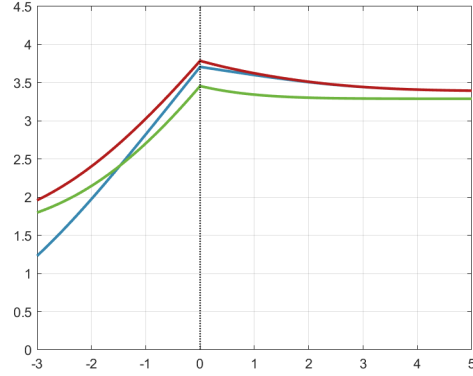
### 3 Numerical Illustration

Figures 1 and 2 compare  $CI_{MA}$  with a test inversion interval  $CI_{TI}$  that arguably reflects the state of the established literature.<sup>6</sup> 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

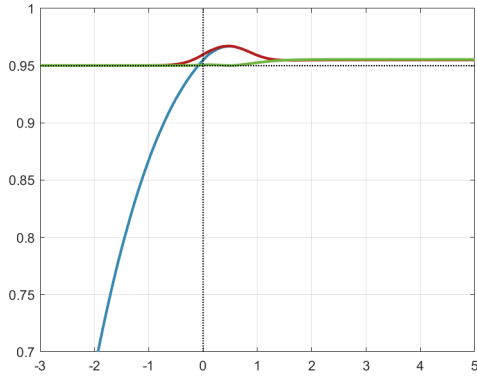
<sup>6</sup>The interval closely follows Romano, Shaikh, and Wolf (2014); other established methods (Andrews and Soares, 2010; Andrews and Barwick, 2012; Bugni, 2010; Canay, 2010) would inform similar constructions. As of writing of this manuscript, at least two rather distinct (from the preceding and 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.



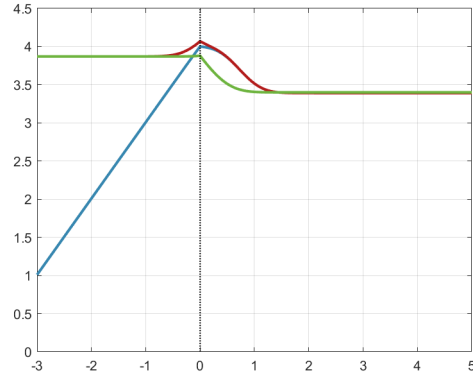
(a) Coverage when  $\rho = -0.7$ .



(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  $\Delta \rightarrow \infty$  is not least favorable and where  $\hat{c} > 1.64$ .

implies  $\Theta_I^*$  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  $\rho$  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. The resulting test is inverted, and the critical value is recomputed, as  $\theta$  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  $\Delta$  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  $\Delta$  take on negative values. The test inversion interval obviously undercovers in that range. To clarify compar-

isons, 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  $\Delta \rightarrow \infty$  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  $\Delta$ . What happens here is that the correction provided by  $CI_{\theta^*}$  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  $\Delta$  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.<sup>7</sup> 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  $\Delta$  (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  $\Delta$ . In fairness to Andrews and Kwon (2019), it appears far from obvious how to implement such a feature in their much more general setting.

The advantage of  $CI_{MA}$  fades out, and even reverses, in the special case where  $\rho \rightarrow 1$  but *not*  $\Delta \rightarrow 0$ . In that limit,  $\hat{c}$  will converge to the two-sided critical value, whereas a pre-test will eventually recommend a one-sided test. While such scenarios can obviously be simulated, they arguably are contrived. The possibility of high  $\rho$  and correspondingly precise estimation of  $\Delta$  is empirically relevant, but it corresponds to the superefficiency case discussed in Remark 4 and therefore to small  $\Delta$  as well as to a case distinction that can be decided in pre-data analysis. Also, one could in principle fix this issue by layering a pre-test on top of  $CI_{MA}$ ; however, as general advice in this matter, I stand by Remark 2.

---

<sup>7</sup>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.

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.

## 4 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 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,  $CI_{MA} = CI_{\theta^*}$ .<sup>8</sup> 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.<sup>9</sup>

<sup>8</sup>This case would not have led any specification test to reject the model, even before taking multiple hypothesis testing into account.

<sup>9</sup>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.

## 5 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.

## A Proof of Theorem 1

Validity in case of  $\Delta < 0$  is obvious – only coverage of  $\theta^*$  is required in this case, and  $CI_{\theta^*}$  achieves that by itself. Since for  $\Delta \geq 0$ , we have  $\Theta_I^* = \Theta_I$ , it remains to show coverage of  $\theta \in [\theta_L, \theta_U]$  assuming that  $\theta_U \geq \theta_L$ . For the remainder of this proof, express the true value of  $\theta$  as  $\theta = \lambda\theta_U + (1 - \lambda)\theta_L$  for some  $\lambda \in [0, 1]$ . Consider initially the idealized confidence interval  $CI_{MA}^*$ , which is just like  $CI_{MA}$  except that, rather than by (2.2), a critical value  $c^*$  is defined by setting  $\inf_{\Delta \geq 0, \lambda \in [0, 1]} \Pr(E_{\Delta, \lambda, c^*}) = 1 - \alpha$ , where

$$\begin{aligned} E_{\Delta, \lambda, c} &= \left\{ Z_1^* - \frac{\lambda}{\sigma_L} \Delta \leq c \cap Z_2^* + \frac{1 - \lambda}{\sigma_U} \Delta \geq -c \right\} \\ &\cup \left\{ Z_1^* + Z_2^* + \left( \frac{1 - \lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \Delta \in \left[ -\sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \right\}, \\ &\quad \left( \begin{matrix} Z_1^* \\ Z_2^* \end{matrix} \right) \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \end{aligned} \quad (\text{A.1})$$

Note two differences to (2.2): The construction explicitly minimizes over both  $\Delta$  and  $\lambda$ , and it is infeasible in that population values of  $(\sigma_L, \sigma_U, \rho)$  are used.

Step 1 of the proof establishes validity of  $CI_{MA}^*$ . Step 2 shows that  $\lambda$  can always be set to 1, transforming the above into (2.2). Step 3 establishes that if  $\rho = 0$ , one can furthermore take the limit as  $\Delta \rightarrow \infty$ . The argument that  $(\sigma_L, \sigma_U, \rho)$  can be replaced with consistent estimators is omitted for brevity.

**Step 1: Validity of  $CI_{MA}^*$ .** Write

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

where  $\sigma^* \equiv \sqrt{2 + 2\rho} \sigma_L \sigma_U / (\sigma_L + \sigma_U)$  is the asymptotic standard deviation of  $\sqrt{n}(\lambda^* \hat{\theta}_U + (1 - \lambda^*) \hat{\theta}_L - \theta^*)$  and  $\lambda^* \equiv \sigma_L / (\sigma_L + \sigma_U)$  is the mixture weight characterizing  $\theta^*$ .

Define also standardized estimation errors

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

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

$$\begin{aligned} &\hat{\theta}_L - \frac{\sigma_L}{\sqrt{n}} c^* \leq \lambda \theta_U + (1 - \lambda) \theta_L \leq \hat{\theta}_U + \frac{\sigma_U}{\sqrt{n}} c^* \\ \iff &\hat{\theta}_L - \theta_L \leq \lambda \Delta + \frac{\sigma_L}{\sqrt{n}} c^*, \quad \hat{\theta}_U - \theta_U \geq -(1 - \lambda) \Delta - \frac{\sigma_U}{\sqrt{n}} c^* \\ \iff &\bar{\varepsilon}_L \leq \frac{\lambda}{\sigma_L} \sqrt{n} \Delta + c^*, \quad \bar{\varepsilon}_U \geq -\frac{1 - \lambda}{\sigma_U} \sqrt{n} \Delta - c^* \end{aligned}$$

or

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

In sum,

$$\begin{aligned}
& \Pr(\theta \in CI_{MA}^*) \\
= & \Pr \left( \left\{ \bar{\varepsilon}_L - \frac{\lambda}{\sigma_L} \sqrt{n} \Delta \leq c^* \cap \bar{\varepsilon}_U + \frac{1 - \lambda}{\sigma_U} \sqrt{n} \Delta \geq -c^* \right\} \right. \\
& \quad \left. \cup \left\{ \bar{\varepsilon}_L + \bar{\varepsilon}_U + \left( \frac{1 - \lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \sqrt{n} \Delta \in \left[ -\sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \right\} \right) \\
\rightarrow & \Pr \left( \left\{ Z_1^* - \frac{\lambda}{\sigma_L} \sqrt{n} \Delta \leq c^* \cap Z_2^* + \frac{1 - \lambda}{\sigma_U} \sqrt{n} \Delta \geq -c^* \right\} \right. \\
& \quad \left. \cup \left\{ Z_1^* + Z_2^* + \left( \frac{1 - \lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \sqrt{n} \Delta \in \left[ -\sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \right\} \right) \\
\geq & \inf_{\Delta \geq 0, \lambda \in [0, 1]} \Pr \left( \left\{ Z_1^* - \frac{\lambda}{\sigma_L} \Delta \leq c^* \cap Z_2^* + \frac{1 - \lambda}{\sigma_U} \Delta \geq -c^* \right\} \right. \\
& \quad \left. \cup \left\{ Z_1^* + Z_2^* + \left( \frac{1 - \lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \Delta \in \left[ -\sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \sqrt{2 + 2\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \right\} \right) \\
= & 1 - \alpha,
\end{aligned}$$

where the convergence uses Assumption 1 and the next step uses the definition of  $c^*$  and also observes that, since we take an infimum over  $\Delta \geq 0$ , we can drop the  $\sqrt{n}$  premultiplying  $\Delta$ .

**Step 2: Concentrating out  $\lambda$ .** We first concentrate out  $\lambda$ , for which  $\{0, 1\}$  are equally least favorable if  $\Delta$  is unrestricted. To see this, consider the reparameterization

$$(X_1, X_2) \equiv \left( \frac{Z_2^* + Z_1^*}{\sqrt{2}}, \frac{Z_2^* - Z_1^*}{\sqrt{2}} \right) \iff (Z_1^*, Z_2^*) = \left( \frac{X_1 - X_2}{\sqrt{2}}, \frac{X_1 + X_2}{\sqrt{2}} \right) \quad (\text{A.2})$$



and observe that  $(X_1, X_2)$  are uncorrelated. Simple algebra yields

$$E_{\Delta, \lambda, c} = \left\{ X_1 - X_2 - \frac{\lambda}{\sigma_L} \sqrt{2} \Delta \leq \sqrt{2} c \cap X_1 + X_2 + \frac{1-\lambda}{\sigma_U} \sqrt{2} \Delta \geq -\sqrt{2} c \right\} \\ \cup \left\{ X_1 + \left( \frac{1-\lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \frac{\Delta}{\sqrt{2}} \in \left[ -\sqrt{1+\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \sqrt{1+\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \right\}.$$

Consider minimizing  $\Pr(E_{\Delta, \lambda, c})$  subject to the constraint that

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

for some fixed value  $\beta \geq 0$ . This is without loss of generality since one can minimize over  $\beta$  in a second step and every value of  $(\Delta, \lambda) \in [0, \infty) \times [0, 1]$  is consistent with some  $\beta \geq 0$ . Also rearranging expressions to be of form “ $\dots \leq X_1 \leq \dots$ ”, one can write

$$E_{\Delta, \lambda, c} \Big|_{\Delta = \frac{\sigma_L \sigma_U}{\lambda \sigma_U + (1-\lambda) \sigma_L} \beta} \\ = \left\{ -X_2 - \sqrt{2} c - \frac{(1-\lambda) \sigma_L}{\lambda \sigma_U + (1-\lambda) \sigma_L} \sqrt{2} \beta \leq X_1 \leq X_2 + \sqrt{2} c + \frac{\lambda \sigma_U}{\lambda \sigma_U + (1-\lambda) \sigma_L} \sqrt{2} \beta \right\} \\ \cup \left\{ \frac{\lambda \sigma_U - (1-\lambda) \sigma_L}{\lambda \sigma_U + (1-\lambda) \sigma_L} \times \frac{\beta}{\sqrt{2}} - \sqrt{1+\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \leq X_1 \leq \frac{\lambda \sigma_U - (1-\lambda) \sigma_L}{\lambda \sigma_U + (1-\lambda) \sigma_L} \times \frac{\beta}{\sqrt{2}} + \sqrt{1+\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right\}$$

and therefore

$$\Pr(E_{\Delta, \lambda, c} | X_2 = x_2) \Big|_{\Delta = \frac{\sigma_L \sigma_U}{\lambda \sigma_U + (1-\lambda) \sigma_L} \beta} \\ = \Pr \left( X_1 \in \left[ -x_2 - \sqrt{2} c - \frac{(1-\lambda) \sigma_L}{\lambda \sigma_U + (1-\lambda) \sigma_L} \sqrt{2} \beta, x_2 + \sqrt{2} c + \frac{\lambda \sigma_U}{\lambda \sigma_U + (1-\lambda) \sigma_L} \sqrt{2} \beta \right] \right. \\ \left. \cup \left[ \frac{\lambda \sigma_U - (1-\lambda) \sigma_L}{\lambda \sigma_U + (1-\lambda) \sigma_L} \times \frac{\beta}{\sqrt{2}} - \sqrt{1+\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right), \frac{\lambda \sigma_U - (1-\lambda) \sigma_L}{\lambda \sigma_U + (1-\lambda) \sigma_L} \times \frac{\beta}{\sqrt{2}} + \sqrt{1+\rho} \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \right] \right)$$

with the understanding that the first interval above is empty for small enough  $x_2$ .

Irrespective of the value taken by  $x_2$ , both intervals are centered at  $\frac{\lambda \sigma_U - (1-\lambda) \sigma_L}{\lambda \sigma_U + (1-\lambda) \sigma_L} \times \frac{\beta}{\sqrt{2}}$ , an expression that increases in  $\lambda$  and takes value 0 at  $\lambda = \lambda^*$ . The intervals' length does not depend on  $\lambda$ , and their union coincides with the larger of the two (whose identity depends on  $x_2$ ). Again irrespective of the value of  $x_2$ ,  $X_1$  is distributed normally around 0. By log-concavity of the Normal distribution (or by taking derivatives), the above probability therefore increases in  $\lambda$  up to  $\lambda^*$  and decreases in  $\lambda$  thereafter conditionally on any  $x_2$ , hence also unconditionally. Furthermore, plugging in  $\lambda \in \{0, 1\}$  reveals symmetry about 0: Switching  $\lambda$  from 0 to 1 is equivalent to leaving  $\lambda$  unchanged but replacing  $X_1$  with  $-X_1$ . The probabilities of all intervals in the above display are, therefore, equally minimized at  $\lambda \in \{0, 1\}$  (although these minima correspond to different  $\Delta$ ). This establishes that, if both of  $(\Delta, \lambda)$  are concentrated out globally, one can restrict attention to one of  $\lambda = 0$  or  $\lambda = 1$ .

We finally observe that the way in which  $\sigma_L$  enters

$$E_{\Delta,1,c} = \left\{ Z_1^* - \frac{\Delta}{\sigma_L} \leq c \cap Z_2^* \geq -c \right\} \cup \left\{ Z_1^* + Z_2^* - \frac{\Delta}{\sigma_L} \in \left[ -\sqrt{2+2\rho}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right] \right\}$$

allows the simplification

$$\begin{aligned} & \inf_{\Delta \geq 0} \Pr(E_{\Delta,1,c}) \\ &= \inf_{\Delta \geq 0} \Pr \left( \left\{ Z_1^* - \Delta \leq c \cap Z_2^* \geq -c \right\} \cup \left\{ Z_1^* + Z_2^* - \Delta \in \left[ -\sqrt{2+2\rho}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \sqrt{2+2\rho}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right] \right\} \right). \end{aligned}$$

**Step 3: For  $\rho = 0$ , concentrating out  $\Delta$ .** For the remainder of this proof, suppose  $\rho = 0$ . In view of step 2, also restrict attention to  $\lambda = 1$ . This step's main claim is that  $\Pr(E_{\Delta,1,c})$  is first increasing and then decreasing (possibly, although not in fact, all increasing or all decreasing) in  $\Delta \geq 0$ . Suppose the claim is true, then it follows that  $\inf_{\Delta \in [0, \infty)} \Pr(E_{\Delta,1,c})$  is attained either at  $\Delta = 0$  or as  $\Delta \rightarrow \infty$ . 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.

It remains to show the main claim. Write  $\gamma = \sqrt{2}\Phi^{-1}(1 - \alpha/2)$ , then (also using  $\rho = 0$ ) we have

$$E_{\Delta,1,c} = \{Z_1^* - \Delta \leq c \cap Z_2^* \geq -c\} \cup \{Z_1^* + Z_2^* - \Delta \in [-\gamma, \gamma]\},$$

where  $(Z_1^*, Z_2^*)$  is bivariate standard Normal. We will henceforth think of  $\Pr(E_{\Delta,1,c})$  as function of  $\Delta$  with  $(c, \gamma)$  fixed. 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

$$\Pr(E_{\Delta,1,c} | Z_2^* = z_2) = \begin{cases} \Phi(\gamma + \Delta - z_2) - \Phi(-\gamma + \Delta - z_2), & z_2 < -c \\ \Phi(\gamma + \Delta - z_2), & -c \leq z_2 \leq -c + \gamma \\ \Phi(\Delta + c), & z_2 > -c + \gamma \end{cases}$$

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

$$\begin{aligned} & \frac{d\Pr(E_{\Delta,1,c})}{d\Delta} \\ &= \frac{d \int_{-\infty}^{\infty} \Pr(E_{\Delta,1,c} | Z_2^* = z_2) \phi(z_2) dz_2}{d\Delta} \\ &= \int_{-\infty}^{-c+\gamma} \phi(\gamma + \Delta - z_2) \phi(z_2) dz_2 - \int_{-\infty}^{-c} \phi(-\gamma + \Delta - z_2) \phi(z_2) dz_2 + \int_{-c+\gamma}^{\infty} \phi(\Delta + c) \phi(z_2) dz_2 \\ &= \underbrace{\sqrt{2} \left( \phi\left(\frac{\gamma + \Delta}{\sqrt{2}}\right) - \phi\left(\frac{-\gamma + \Delta}{\sqrt{2}}\right) \right)}_A \Phi\left(\frac{\gamma - \Delta - 2c}{\sqrt{2}}\right) + \underbrace{\phi(\Delta + c)\Phi(c - \gamma)}_B. \end{aligned} \tag{A.3}$$

To see the last step, note first that  $\int_{-c+\gamma}^{\infty} \phi(\Delta + c)\phi(z_2)dz_2$  simplifies to  $B$ . Next,

$$(Z_1^*, Z_2^*) = (\gamma + \Delta - z_2, z_2) \Leftrightarrow (X_1, X_2) = \left( \frac{\gamma + \Delta}{\sqrt{2}}, \frac{2z_2 - \gamma - \Delta}{\sqrt{2}} \right),$$

where  $(X_1, X_2)$  is as in (A.2). Because  $\rho = 0$  implies that  $(X_1, X_2)$  is standard normal, we have

$$\begin{aligned} \int_{-\infty}^{-c+\gamma} \phi(\gamma + \Delta - z_2)\phi(z_2)dz_2 &= \int_{-\infty}^{-c+\gamma} \phi\left(\frac{\gamma + \Delta}{\sqrt{2}}\right) \phi\left(\frac{2z_2 - \gamma - \Delta}{\sqrt{2}}\right) dz_2 \\ &= \sqrt{2} \int_{-\infty}^{(\gamma - \Delta - 2c)/\sqrt{2}} \phi\left(\frac{\gamma + \Delta}{\sqrt{2}}\right) \phi(t)dt = \sqrt{2} \phi\left(\frac{\gamma + \Delta}{\sqrt{2}}\right) \Phi\left(\frac{\gamma - \Delta - 2c}{\sqrt{2}}\right). \end{aligned}$$

A similar computation for  $\int_{-\infty}^{-c} \phi(-\gamma + \Delta - z_2)\phi(z_2)dz_2$  and rearrangement of terms yield term  $A$  in (A.3).

Term  $A$  equals zero at  $\Delta = 0$  and then becomes negative. Term  $B$  is positive throughout. Because all terms vanish as  $\Delta \rightarrow \infty$ , it is not useful to directly take further derivatives. However, we can compare the terms' relative magnitude. In particular, we will see that  $|A|/|B|$  increases in  $\Delta$ , hence  $d\Pr(E_{\Delta,1,c})/d\Delta$  has at most one sign change and that sign change (if it occurs) is from positive to negative, establishing the claim.

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

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

where “const.” absorbs terms that do not depend on  $\Delta$ . The derivative of this expression with respect to  $\Delta$  (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  $\phi(\dots)$  (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

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

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

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

which is again nonnegative if  $2c \geq \gamma$ . Thus,  $|A|/|B|$  is nondecreasing in  $\Delta$  for all  $\Delta \geq 0$ , concluding the proof.

## 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 (2015): “Specification tests for partially identified models defined by moment inequalities,” *Journal of Econometrics*, 185(1), 259–282.
- (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.
- 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.
- ROMANO, J. P., AND A. M. SHAIKH (2008): “Inference for identifiable parameters in partially identified econometric models,” *Journal of Statistical Planning and Inference*, 138(9), 2786 – 2807.
- 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.
- WHITE, H. (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50(1), 1–25.