cemmap

centre for microdata methods and practice

Simple misspecification adaptive inference for interval identified parameters

Jörg Stoye

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP55/20

Economic
and Social
Research Council

Simple Misspecification Adaptive Inference for Interval Identified Parameters

Jörg Stoye*

October 26, 2020

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. However, except for large positive correlation of bound estimators, the practical result will be simple: To achieve 95% coverage, compute both a simple 90% (!) confidence interval for the identified set and a standard 95% confidence interval for the pseudotrue parameter and report the union of these intervals.

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, in which improvement over existing inference methods is demonstrated. More generally, simulations suggest that the novel confidence interval has excellent length and size control.

*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. Financial support through NSF Grant SES-1824375 is gratefully acknowledged.

1 Introduction

Inference under partial identification is by now 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. In this paper's simple case of inference on an interval identified parameter on the real line, this will happen if estimates of upper and lower bounds are not ordered in the expected way. This issue is empirically highly relevant. For example, it 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 parameter.

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 intervals with the following features:

- It is never empty nor very short.
- It exhibits asymptotically guaranteed coverage uniformly over the identified set and additionally for a well-defined pseudotrue parameter.
- Because the nonemptiness can be anticipated in computing critical values, it tends to be shorter than more conventional intervals in benign cases, including in all 13 instances of the empirical application.
- It is free of tuning parameters and trivial to compute.

In general, computing the confidence interval requires to concentrate out a scalar nuisance parameter, namely the length of the identified set. For all cases except high positive correlation of estimators, the result will be the usual *one-sided* critical value. The interval then

[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.

becomes exceedingly simple to compute: For 95% coverage, add ± 1.64 standard errors to estimates of bounds, add ± 1.96 standard errors to the estimate of a pseudotrue parameter, and report the union of these intervals.

For the important special case of uncorrelated estimators and conventional coverage levels, the result can be shown completely analytically, i.e. without resorting to any numerical concentrating out. Simulations suggest remarkably tight size control and correspondingly short expected length of the new interval. Application to de Quidt, Haushofer, and Roth (2018) illustrates that the improvement is not negligible in practice.

2 A Misspecification Adaptive Confidence Interval

I first develop the new proposal heuristically and then state it formally. While the interpretation of what follows is inference on a scalar parameter θ, 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 $(\theta_L, \theta_U) \in \boldsymbol{R}^2$ 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 $\theta_0 \equiv \lambda\theta_U + (1-\lambda)\theta_L$, where $\lambda \in [0,1]$ is fixed but unknown. 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 (Chernozhukov, Lee, and Rosen, 2013).

The obvious estimator of Θ_I is $[\hat{\theta}_L, \hat{\theta}_U]$, but defining a confidence interval is more 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 the value taken by λ or by $\Delta \equiv \theta_U - \theta_L$.

The current default recommendation for this problem is a "Generalized Moment Selection" (Andrews and Soares, 2010) confidence interval. One simple version is

$$CI_{GMS} \equiv \left[\hat{\theta}_L - c\hat{\sigma}_L/\sqrt{n}, \theta_U + c\hat{\sigma}_U/\sqrt{n}\right],$$

where the critical value c equals either $\Phi^{-1}(1-\alpha)$ or $\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ depending on whether Δ appears to be "large" (in which case we really have one-sided inference) or "small" (in which case CI_{GMS} is a Bonferroni interval) and $\Phi(\cdot)$ is the standard normal distribution function.[2]

[2]The critical value can be improved by taking ρ into account, but the effect will be minimal except for large negative ρ. The refinement is implemented in Sections 3 and 4.

The catch is that we do not know which case obtains, and since Δ actually enters asymptotic approximations scaled to $\sqrt{n}\Delta$, we are unable to pre-estimate it with sufficient accuracy. However, it is easy to see that, up to here, coverage probability for given c increases in Δ. Thus, the standard remedy (suggested by Stoye (2009) for the example and by Andrews and Soares (2010), Bugni (2010), and Canay (2010) for more general settings) is to use the smaller critical value only if a pre-test suggests a large Δ.[3] The test can be either of two varieties:

(i) Letting $\hat{\Delta} \equiv \hat{\theta}_U - \hat{\theta}_L$, compare the scaled and studentized $\hat{\Delta}$ to a critical value that slowly diverges. Following Andrews and Soares (2010), a threshold of $\sqrt{\log n}$ has become the industry standard.

(ii) Use a fixed pre-test size, commonly set to $.1\alpha$, and replace α with $.9\alpha$ in subsequent inference (Andrews and Barwick, 2012; Romano, Shaikh, and Wolf, 2014). This Bonferroni adjustment reduces reliance on asymptotic approximation. It will be used when computing CI_{GMS} below.

Again, CI_{GMS} and its close relatives can be empty. This can be advertised an embedded specification test but may not be wanted.[4] Arguably even more problematic is that, if the model is misspecified, CI_{GMS} can be spuriously short. Also, a specification test will not be helpful: In this paper's setting, the best-practice such test (Bugni, Canay, and Shi, 2015) agrees with the embedded one, i.e. it just reports whether CI_{GMS} is empty.[5]

Addressing this concern requires a notion of coverage for the case of misspecification, i.e. if $\theta_L > \theta_U$. Taking a cue from Andrews and Kwon (2019), define the pseudotrue parameter

$$\theta^* \equiv \frac{\sigma_U \theta_L + \sigma_L \theta_U}{\sigma_L + \sigma_U}$$

and define validity of a confidence interval as follows:

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

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

Here, the order of lim and inf enforces that coverage is attained uniformly over $\Theta_I \cup \{\theta^*\}$. For uniformity over data generating processes, see Remark 3 below. Forcing coverage of θ^* will ensure that the interval is nonempty and also that it is statistically interpretable. The obvious caveat is that, as with pseudotrue parameters elsewhere, the substantive relevance of θ^* may be dubious.

[3]There are many variations on this, e.g. smooth shrinkage of $\hat{\Delta}$, that will not make much difference in the example. I refer to Andrews and Soares (2010) for details.

[4]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).

[5]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.

$\boldsymbol{\rho}$	$\leq$0.8	0.85	0.9	0.95	0.98	1.0
$\boldsymbol{\alpha = .1}$	1.28	1.29	1.31	1.36	1.44	1.64
$\boldsymbol{\alpha = .05}$	1.64	1.65	1.65	1.70	1.76	1.96
$\boldsymbol{\alpha = .01}$	2.33	2.33	2.33	2.34	2.40	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.

The pseudotrue parameter θ^* has natural 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

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

Validity of CI_{θ^*} for θ^* is obvious. Hence, one can achieve overall validity by reporting $CI_{GMS} \cup CI_{\theta^*}$. This will be rather conservative if Δ is nonnegative but short; see the red curves in the left panels of Figure 1 for a preview. One may want to downward adjust c in anticipation of this. However, the correct adjustment depends on how likely CI_{θ^*} is to be contained in CI_{GMS}, and this again depends on $\sqrt{n}\Delta$. Importantly, while large Δ previously increased coverage, its effect through this novel channel is in the opposite direction, and the net effect is not monotonic. This makes the problem delicate and is why Andrews and Kwon (2019) shrink a generalized notion of $\hat{\Delta}$ in different directions, using potentially different tuning parameters, for different moving parts of the inference problem.

The new proposal builds on these ideas but adds some important twists. First, I avoid any pretests and tuning parameters and globally concentrate out Δ. That is, I calibrate c s.t. the smallest coverage over $\Delta \in [0, \infty)$ (as well as $\lambda \in [0, 1]$) is estimated to be $(1 - \alpha)$. This may seem like a step backwards –global concentrating out of similar parameters appears in Rosen (2008) and in the statistics literature referenced therein– but its effect on inference turns out to be small.

This can be seen in Table 1, which displays the recommended value of c for different target coverages and values of ρ. (It will turn out that this value does not depend on (σ_L, σ_U).) Strikingly, for a wide range of parameter values, the recommendation is simply to use the usual *one-sided* critical value. For all entries with $\rho \leq .8$, simulations strongly suggest that this recommendation is exact.[6] The new proposal will, therefore, frequently amount to the

[6] In these cases, simulation at $c = \Phi^{-1}(1 - \alpha)$ suggests that worst-case coverage occurs in the limit as

trivial procedure advertised in the abstract and introduction. Indeed, it requires very high correlation to deviate from this procedure in a quantitatively meaningful way.

The final twist occurs if $\rho = 0$ is known, notably when bounds are estimated from different samples. For this special case and for conventional levels of desired coverage, a proof that $\Delta \to \infty$ minimizes coverage is available, resulting in a completely analytic justification for the simple procedure.

Due to some further simplifications that will be explained later, the formal definition of the novel confidence interval is as follows.

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

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

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

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

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

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

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

Proof. See appendix A. □

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

In principle, one could try to capture this by concentrating out Δ over a more limited range, e.g. over a $(1 - .1\alpha)$-confidence interval with Bonferroni adjustment of subsequent inference. I do not advocate this because numerically, even a value of $\Delta \approx 3\sigma_L/\sqrt{n}$, which would be hard to exclude, is similar to $\Delta = \infty$. Therefore, the "inferential cost" of a pre-test, whether through adjustment of second-stage test size or through reliance on a tuning parameter, would not be recovered.

REMARK 2: The condition that $\sqrt{2}\Phi^{-1}(1-\alpha) \geq \Phi^{-1}(1-\alpha/2)$ holds for $\alpha < .14$, i.e. for conventional coverage levels. Simulations suggest that the result holds true for all α.

$\Delta \to \infty$ and therefore equals $(1-\alpha)$. Since this limit must be below $(1-\alpha)$ if c is below the one-sided critical value, such a recommendation is tight.

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) Δ and λ; see Molinari (2020, Section 4.3.2) for further discussion. CI_{MA} is obviously uniform in this sense because both Δ and λ are set to globally least favorable values.

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 since Andrews and Soares (2010) 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 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 Δ (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. In contrast, the perfect application of CI_{MA} is the polar opposite case in which bounds are independently estimated.

Remark 5: I follow the bulk of the literature in focusing on uniform coverage of $\theta_0 \in \{\theta^*\} \cup \Theta_I$. However, the procedure could be adapted to coverage of the entire set $\{\theta^*\} \cup \Theta_I$. 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 Δ) 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, recall that there are two delicate nuisance parameters, Δ and λ.[7] The proof shows first that the procedure would be valid if expression (2.2) explicitly concentrated out both. Next, λ can be concentrated out analytically. In particular, one can restrict attention to one of θ_L or θ_U; expression (2.2) arbitrarily chooses the latter. While not unexpected, this finding is not completely obvious: For given Δ, 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 θ_U depends on $(\Delta, \sigma_L, \sigma_U)$ only through Δ/σ_L. For the purpose of

[7] There is also $(\sigma_L, \sigma_U, \rho)$, but these can be handled in entirely standard ways. In contrast, both Δ and λ are scaled by $\sqrt{n}$ in the asymptotic approximation, not to mention that λ is not identified.

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 with the further restriction that $\rho = 0$, coverage is provably minimized as $\Delta \to \infty$, 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 θ_L or θ_U will be .9 if $\Delta = 0$ and approach .95 *from below* as $\Delta \to \infty$. The right-hand interval's coverage of these values is .95 at $\Delta = 0$ (where both coincide with θ^*) but rapidly decreases to 0 as Δ increases. That these effects aggregate to coverage uniformly above .95 is not obvious and maybe even unexpected – certainly to this author, who first discovered it in simulations.

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

3 Numerical Illustration

Figure 1 compares CI_{MA} (green) with CI_{GMS} (blue) and also with $CI_{GMS} \cup CI_{\theta^*}$ (red). The latter comparison illustrates that the effect of adjusting c is not negligible. The comparison is extended into the misspecified range by letting Δ take on negative values. The impact of asymptotic approximations, which will be comparable across methods, is minimized by simulating straight from normal distributions of $(\hat{\theta}_L, \hat{\theta}_U)$ and taking $(\sigma_L, \sigma_U, \rho)$ to be known. To be scrupulously fair, knowledge of ρ is also built into construction of CI_{GMS}, very slightly lowering its critical values. Nominal coverage is 95% throughout. The figure illustrates $\rho = 0$ (top panels) and $\rho = .5$ (bottom panels), with $\sqrt{n}\sigma_L = \sqrt{n}\sigma_U = 1$ throughout, so that Δ is denominated in estimator standard errors.

The figure suggests 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 $\rho = 0$, 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_{GMS} accounts for a pre-test.

The considerable advantage of CI_{MA} fades out, and even reverses, as ρ gets very close

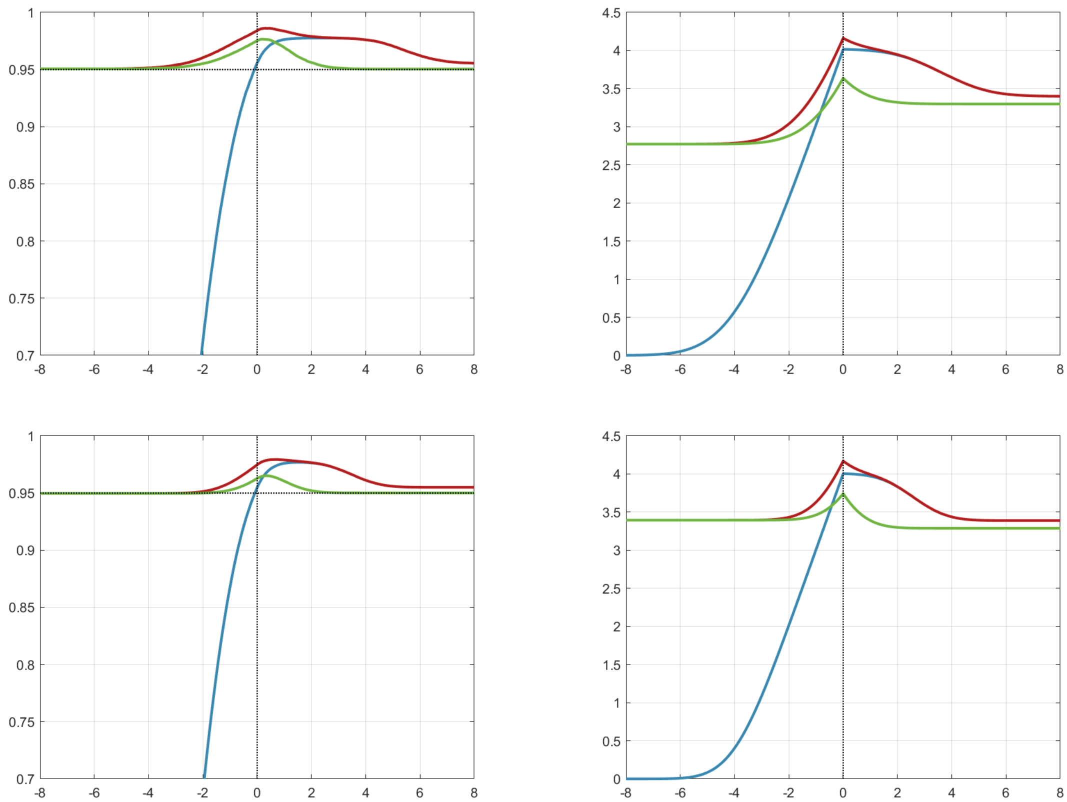

Figure 1: Coverage (left panels) and expected length (right panels; length of true interval is subtracted) of CI_{GMS} (blue), $CI_{GMA} \cup CI_{\theta^*}$ (red) and the new proposal CI_{MA} (green). Horizontal axis is $\Delta = \theta_U - \theta_L$. Nominal coverage is 95% and is indicated by a black horizontal line. Top panels: $\rho = 0$; bottom panels: $\rho = 0.5$.

to 1. This is because $\hat{c}$ then approaches the usual two-sided critical value, driven by the possibility that Δ is small. If the data actually suggest a large Δ, CI_{GMS} will comparatively benefit. A look at Table 1 confirms that the estimators need to be very correlated indeed (roughly, $\rho > .95$) for this to be salient. To any practitioner who encounters such a sample correlation, I would advise to investigate whether the setting is in fact the one of Remark 4. 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 1 above.

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, so that the issue of spurious precision may arise.

Game	Group	$\hat{\theta}_L$	$\hat{\theta}_U$	CI_{MA}	CI_{GMS}	rel. length
Ambiguity	5	0.499	0.557	[0.459,0.597]	[0.451,0.601]	0.822
DG	2	0.325	0.389	[0.309,0.405]	[0.308,0.406]	0.970
DG	4	0.502	0.575	[0.473,0.603]	[0.473,0.604]	0.970
Effort: incentive	6	0.469	0.484	[0.448,0.503]	[0.443,0.508]	0.822
Effort: no incentive	6	0.343	0.331*	[0.318,0.356]	[0.315,0.358]	0.888
Lying	5	0.530	0.537	[0.512,0.556]	[0.508,0.560]	0.822
Risk	2	0.531	0.556	[0.507,0.582]	[0.501,0.587]	0.822
Risk	4	0.570	0.633	[0.535,0.665]	[0.528,0.672]	0.822
Time	5	0.766	0.770	[0.722,0.814]	[0.712,0.824]	0.822
Trust	5	0.430	0.455	[0.388,0.493]	[0.379,0.501]	0.822
Trustworthiness	5	0.348	0.398	[0.328,0.426]	[0.323,0.433]	0.822
UG 1	5	0.443	0.470	[0.422,0.493]	[0.418,0.498]	0.822
UG 2	5	0.362	0.413	[0.342,0.436]	[0.341,0.436]	0.970

Table 2: GMS confidence interval and new proposal for data in de Quidt, Haushofer, and Roth (2018). Relative length refers to excess length over $\hat{\Delta}$ or, in the case (*) of inverted bounds, over 0.

I focus on bounds for a baseline setting before inducing experimenter demand because these are on the short side and therefore raise interesting econometric issues. Table 2 displays estimated bounds, CI_{MA}, and CI_{GMS} for the "weak bounds" data.[8] The last column divides the length of CI_{MA} by the length of CI_{GMS} (subtracting the width of the estimated bounds if it is positive). As with the numerical illustration, CI_{GMS} uses that $\rho = 0$ is known.

Each row of the table corresponds to one of three cases. A relative length of .822 indicates that in practice, CI_{MA} adds 1.64 standard errors whereas CI_{GMS} adds 2.00 because the pre-test did not reject a short interval (and also accounting for the pre-test). A relative length of .970 indicates that the pre-test suggested a long interval, in which case CI_{GMS} uses 1.70 standard errors. Finally, the table's fifth row reports a relative length of .888. In this experiment, bound estimators were inverted: $(\hat{\theta}_L, \hat{\theta}_U) = (.343, .331)$ with standard errors $(.0136, .0139)$. In this row only, CI_{MA} ex post coincides with CI_{θ^*}. That is, the misspecification adaptive correction is binding and yet the novel confidence interval is still the shorter one.

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

[8] The computations correspond to material in the paper's online appendix. Figure 1 in the main paper visualizes the data analyzed here, and the inverted bounds are visible in its 5^{th} entry.

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 construction heavily relies on the structure of the specific case. For a more general but also more complicated (including in relatively simple instances) treatment, I refer to Andrews and Kwon (2019). For reasons alluded to in section 2, their treatment is delicate and involves numerous tuning parameters. Of course, it is also more widely applicable. The approaches are, therefore, very much complementary. The present proposal is strictly limited to a simple instance of the problem but appears extremely attractive when that instance applies.

A Proof of Theorem 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

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

Note two differences to (2.2): The construction explicitly minimizes over both Δ and λ, 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 λ 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 $\Delta \to \infty$. The argument that $(\sigma_L, \sigma_U, \rho)$ can be replaced with consistent estimators is omitted for brevity. Note also that validity in case of $\Delta < 0$ is obvious.

Step 1: Validity of CI^*_{MA}. Write

$$CI^*_{MA} = \left[\hat\theta_L - \frac{\sigma_L}{\sqrt{n}}c^*, \hat\theta_U + \frac{\sigma_U}{\sqrt{n}}c^*\right] \cup \left[\frac{\sigma_L\hat\theta_U + \sigma_U\hat\theta_L}{\sigma_L+\sigma_U} - \frac{\sigma^*}{\sqrt{n}}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right), \frac{\sigma_L\hat\theta_U + \sigma_U\hat\theta_L}{\sigma_L+\sigma_U} + \frac{\sigma^*}{\sqrt{n}}\Phi^{-1}\left(1-\tfrac{\alpha}{2}\right)\right],$$

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 θ^*.

Define also standardized estimation errors

$$(\bar\varepsilon_L, \bar\varepsilon_U) \equiv \sqrt{n}\left(\frac{\hat\theta_L - \theta_L}{\sigma_L}, \frac{\hat\theta_U - \theta_U}{\sigma_U}\right).$$

Recalling that the true parameter value equals $\theta_0 = \lambda\theta_U + (1-\lambda)\theta_L$ for some $\lambda \in [0,1]$, we have that $\theta_0 \in CI^*_{MA}$ if either

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

or

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

In sum,

$$\begin{aligned}
& \Pr(\theta_0 \in CI^*_{MA}) \\
= \quad & \Pr\left(\left\{ \bar{\varepsilon}_L - \frac{\lambda}{\sigma_L}\sqrt{n}\Delta \leq c^* \cap \bar{\varepsilon}_U + \frac{1-\lambda}{\sigma_U}\sqrt{n}\Delta \geq -c^* \right\} \right. \\
& \left. \cup \left\{ \bar{\varepsilon}_L + \bar{\varepsilon}_U + \left(\frac{1-\lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \sqrt{n}\Delta \in \left[-\sqrt{2+2\rho} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right), \sqrt{2+2\rho} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right) \right] \right\} \right) \\
\to \quad & \Pr\left(\left\{ Z_1^* - \frac{\lambda}{\sigma_L}\sqrt{n}\Delta \leq c^* \cap Z_2^* + \frac{1-\lambda}{\sigma_U}\sqrt{n}\Delta \geq -c^* \right\} \right. \\
& \left. \cup \left\{ Z_1^* + Z_2^* + \left(\frac{1-\lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \sqrt{n}\Delta \in \left[-\sqrt{2+2\rho} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right), \sqrt{2+2\rho} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right) \right] \right\} \right) \\
\geq \quad & \inf_{\Delta \geq 0, \lambda \in [0,1]} \Pr\left(\left\{ Z_1^* - \frac{\lambda}{\sigma_L}\Delta \leq c^* \cap Z_2^* + \frac{1-\lambda}{\sigma_U}\Delta \geq -c^* \right\} \right. \\
& \left. \cup \left\{ Z_1^* + Z_2^* + \left(\frac{1-\lambda}{\sigma_U} - \frac{\lambda}{\sigma_L} \right) \Delta \in \left[-\sqrt{2+2\rho} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right), \sqrt{2+2\rho} \Phi^{-1}\left(1 - \tfrac{\alpha}{2}\right) \right] \right\} \right) \\
= \quad & 1 - \alpha,
\end{aligned}$$

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 Δ.

Step 2: Concentrating out λ. (This step is visualized in Appendix B.) We first concentrate out λ, for which $\{0,1\}$ are equally least favorable if Δ is unrestricted. To see this, consider the reparameterization

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

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

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

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 β. Also rearranging expressions to be of form "$\cdots \leq X_1 \leq \ldots$", one can write

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

and therefore

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

with the understanding that the first of the two intervals 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 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). 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 λ up to λ^* and decreases in λ thereafter conditionally on any x_2, hence also unconditionally. Furthermore, the expression is symmetric in λ and therefore equally minimized at $\lambda \in \{0, 1\}$ (although these minima correspond to different Δ), meaning that if both of (Δ, λ) are concentrated out globally, one can restrict attention to one of $\lambda = 0$ or $\lambda = 1$.

We finally observe that the way in which σ_L enters

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

allows the simplification

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

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 $\lambda = 1$. 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 $\Delta \to \infty$. 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.

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

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

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

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

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 $\Delta\to\infty$, 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 Δ, 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

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

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

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

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 regroup terms to conclude that the last expression above has the same sign as

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

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

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

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

References

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 Disucssion Paper CFDP 2184R*.

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

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.

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.

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., 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.

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., A. M. SHAIKH, AND M. WOLF (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82(5), 1979–2002.

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

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.

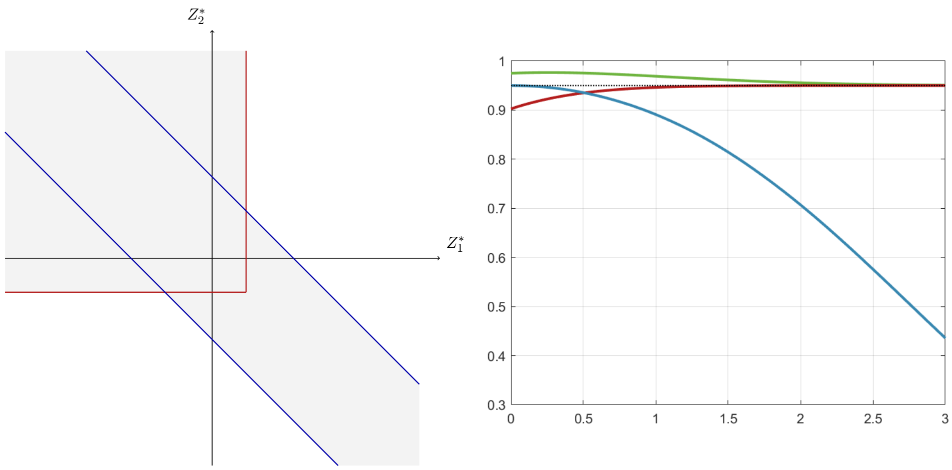

Figure 2: Left panel: Event $E_{0,1,c}$ from the proof. Red contour indicates coverage of θ_U by $\left[\hat{\theta}_L - c, \hat{\theta}_U + c\right]$, blue contour outlines coverage by CI_{θ^*}. Changing (Δ, λ) can translate the event arbitrarily but only to the southeast. The picture is to scale for $c = \Phi^{-1}(1-\alpha)$ and $\rho = 0$. Right panel: Probability of "red" event, "blue" event, and their union ($= E_{\Delta,1,c}$, green) as Δ increases from 0 to 5.

B Proof Heuristic (Not for Publication)

The following interpretive remarks may help digest the proof. For $\rho = 0$, the event $E_{0,1,\Phi^{-1}(1-\alpha)}$ is drawn to scale in the left panel of Figure 2. By increasing Δ and/or λ, one can translate it by an arbitrary vector in the fourth quadrant; that is, the set can be arbitrarily shifted downward and rightward. The aim is to bound the event's probability uniformly over these translations. The reparameterization in (A.2) corresponds to tilting the panel clockwise by 45°, upon which it has a vertical axis of symmetry. The further reparameterization in Step 2 is useful because level sets of β (but not Δ) are horiontal lines in the tilted figure. These features usefully interact with the fact that (X_1, X_2) are uncorrelated (although not standardized). They are used in Step 2 and early on (in deriving term A) in Step 3.

The figure's right panel illustrates how coverage (green), as well as coverage by the two "component intervals" in isolation (colors corresponding to left panel), changes as Δ increases. Step 3 of the proof verifies that the green curve first increases and then decreases. Numerically, this is true far more generally.