MISSPECIFIED MOMENT INEQUALITY MODELS: INFERENCE AND DIAGNOSTICS

By

Donald W.K. Andrews, and Soonwon Kwon

March 2017

Revised July 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2184R2

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Misspecified Moment Inequality Models: Inference and Diagnostics

Donald W. K. Andrews
Cowles Foundation
Yale University

Soonwoo Kwon
Department of Economics
Brown University

First Version: March 12, 2017
Revised: July 29, 2022*

*The authors' email addresses are donald.andrews@yale.edu and soonwoo_kwon@brown.edu. Their postal addresses are Cowles Foundation, Yale University, P.O. Box 208281, New Haven, CT, 06520-8281, USA, and Department of Economics, Brown University, Providence, R.I., 02912, USA.

This paper is an update of Andrews and Kwon (2019), which is titled "Inference in Moment Inequality Models That Is Robust to Spurious Precision under Model Misspecification." For helpful comments, the authors thank four referees, Francesca Molinari, Arie Beresteanu, Federico Bugni, Denis Chetverikov, Azeem Shaikh, seminar participants at Duke University and the University of Pittsburgh, and conference participants at the 2018 Penn State Cornell Conference on Econometrics and Industrial Organization. The first author gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-1355504 and SES-1656313.

Abstract

This paper is concerned with possible model misspecification in moment inequality models. Two issues are addressed. First, standard tests and confidence sets for the true parameter in the moment inequality literature are not robust to model misspecification in the sense that they exhibit spurious precision when the identified set is empty. This paper introduces tests and confidence sets that provide correct asymptotic inference for a pseudo-true parameter in such scenarios, and hence, do not suffer from spurious precision.

Second, specification tests have relatively low power against a range of misspecified models. Thus, failure to reject the null of correct specification does not necessarily provide evidence of correct specification. That is, model specification tests are subject to the problem that absence of evidence is not evidence of absence. This paper develops new diagnostics for model misspecification in moment inequality models that do not suffer from this problem.

Keywords: Asymptotics, confidence set, diagnostics, identification, inference, misspecification, moment inequalities, robust, spurious precision, test.

JEL Classification Numbers: C10, C12.

1 Introduction

This paper addresses two problems concerning misspecification in moment inequality models. The first is the potential for spurious precision of standard confidence sets (CS's) for a parameter θ under model misspecification. By spurious precision, we mean that the coverage probability of the CS is less than its nominal level $1 - \alpha$ for all parameter values, including the true value (if a true value is well defined) and any potential pseudo-true value. We show that one cannot rely on a model specification test to detect misspecification over a range of levels that cause substantial spurious precision. Thus, practitioners who observe a relatively short confidence interval (CI) or small CS can be mislead by spurious precision due to model misspecification. The second problem is the assessment of misspecification. Specification tests in the literature are valuable, but have the drawback that failure to reject the null hypothesis of correct specification can be due to low power, rather than to evidence in favor of the null hypothesis. Hence, alternative misspecification diagnostics are desirable.

There are reasons to worry about misspecification in moment inequality models. For example, in the hospital-HMO contract example in Pakes (2010, p. 1812), no parameter value satisfies the sample moment inequalities. The same is true in certain scenarios of the ATM cost example in Pakes, Porter, Ho, and Ishii (2015, Table I, rows 3 and 4) and in the hospital referrals study in Ho and Pakes (2014, p. 3871). As these authors discuss, this could be due to small sample effects or to misspecification of the moment inequalities. Another example is the trade participation study of Dickstein and Morales (2018, Table V) in which some specifications of the information set lead to rejection of the moment inequalities, while others do not.

To address the problem of spurious precision, we develop inference methods concerning a parameter θ that have correct asymptotic level under correct model specification and also have correct asymptotic level for a pseudo-true parameter under model misspecification. This property eliminates the problem of spurious precision under model misspecification. No procedures currently in the literature have been shown to have this property.

The approach we take is to define the identified set under model misspecification to be the set of parameter values that solve the *minimally-relaxed* moment inequalities. That is, one relaxes each moment inequality (normalized by its standard deviation) by the smallest amount $r^{\inf} \geq 0$ such that the relaxed moment inequalities hold for some parameter θ_I in the parameter space Θ. The collection of such values θ_I is defined to be the misspecification-robust (MR) identified set Θ_I^{MR}. Parameter values in the MR-identified set minimize the maximum inequality violation in the misspecified moment inequality model.

We develop tests and CS's concerning θ that are spurious-precision robust (SPUR) in that they have correct asymptotic level with respect to some $\theta_I \in \Theta_I^{MR}$ under model misspecification, just as they do under correct model specification. The method is as follows. For a test, the null hypothesis is $H_0 : \theta_0 \in \Theta_I^{MR}$. First, one estimates the nonnegative relaxation parameter $r^{\inf}$ by its sample analogue $\widehat{r}_n^{\inf}$. Then, one constructs a test statistic in the usual way, but using the sample moments relaxed by $\widehat{r}_n^{\inf}$. We refer to this statistic as a SPUR test statistic. The SPUR test statistic is combined with an extended generalized moment selection (EGMS) bootstrap critical value to yield what we call a SPUR1 test and corresponding CS.

Next, we improve the power of the SPUR1 test under correct model specification. Let $\alpha = \alpha_1 + \alpha_2$, where $\alpha_1, \alpha_2 > 0$, such as $\alpha_1 = .005$ and $\alpha_2 = .045$. We use a nominal $1 - \alpha_1$ CI for $r^{\inf}$ to construct a Bonferroni level α SPUR2 test. This test rejects the null if a level α_2 standard generalized moment selection (GMS) test, as in Andrews and Soares (2010), rejects, when the CI for $r^{\inf}$ only includes the value 0, and rejects the null if the level α_2 SPUR1 and standard GMS tests both reject, otherwise.[1] The nominal $1 - \alpha_1$ upper-bound CI for $r^{\inf}$ is obtained straightforwardly from the CI for the misspecification index $\Delta^{\inf}$, which we discuss below.

The SPUR2 test and corresponding CS are our recommended procedures because they are "adaptive." That is, they have the desirable feature that if the model is correctly specified and the identified set contains slack points for which the slackness of the inequalities is of order greater than $n^{-1/2}$, then they perform "almost" the same as the standard nonrobust GMS test and CS with probability that goes to one as $n \to \infty$ (wp$\to$1). And, if the identified set is empty, they perform "almost" the same as the robust SPUR1 test and CS wp$\to$1. By "almost," we mean that the level α SPUR2 test and CS perform the same as the level α_2 GMS test and CS in the first scenario wp$\to$1 and as the level α_2 SPUR1 test and CS wp$\to$1 in the second scenario. The difference in power is minimal—typically in [.00, .02].

An empirical illustration of the methods introduced in this paper is given for an airline entry model with 48 inequalities, 9 parameters, and sample size 7,882, as in Kline and Tamer (2016) and Kaido, Molinari, and Stoye (2019). We find that several coefficients that are significantly different from zero when using a standard test, are not when using a SPUR2 test. This suggests that the standard tests may be exhibiting spurious precision.

A possible drawback of SPUR2 procedures is that they provide valid inference for a pseudo-true parameter, but this may not be the parameter that is of greatest interest from a substantive perspective. The same drawback arises with standard maximum likelihood, least squares, and GMM methods. For example, the maximum likelihood pseudo-true parameter minimizes the Kullback-Leibler quasi-distance between the distribution of the data and the distributions in the specified

[1] In the "otherwise" scenario, the SPUR2 test typically reduces to the SPUR1 test, because the GMS test typically rejects when the SPUR1 test does.

model, whereas parameters in the MR-identified set minimize the maximum distance between the relaxed and original population moment inequalities. This is not a drawback relative to a standard non-SPUR procedure, because either the pseudo-true parameters are the same for both procedures or there are no pseudo-true parameters for the standard procedure and it exhibits spurious precision.

On the other hand, this problem can be more severe in moment inequality models. A small amount of misspecification in a moment inequality model can leave the true value far from the identified set, which does not occur in moment equality models. This occurs in the knife-edge case in which the identified set under correct specification consists of a nondegenerate set, which has positive Lebesgue measure, and an isolated point, which happens to be the true value. Under arbitrarily small misspecification, the identified set can exclude the isolated point, and hence, the true value can be far from the identified set. This is a scenario in which misspecification is not identifiable. It is an unavoidable feature of inequality models. Fortunately, it seems unlikely to arise often in practice.

To address the drawback of specification tests discussed in the first paragraph, this paper introduces a misspecification index that equals the maximum violation across moment inequalities (normalized by their standard deviations) evaluated at the parameter value θ that minimizes the maximum violation. The index is denoted by Δ^{inf}, where the inf denotes the infimum over θ. The misspecification index Δ^{inf} is positive when the model is misspecified and the identified set of θ values is empty. The value of Δ^{inf} is increasing in the magnitude of the inequality violations. It is negative or zero when the identified set is non-empty, and is negative when the identified set contains one or more slack points. When Δ^{inf} is negative, its absolute magnitude measures the size of the identified set in terms of the maximum slackness of the moment inequalities at points in the identified set.

The misspecification index Δ^{inf} is a population quantity. Its sample analogue $\widehat{\Delta}_n^{\text{inf}}$ is a consistent estimator of it, where n denotes the sample size. For $\widehat{\Delta}_n^{\text{inf}}$ to be useful, a measure of its accuracy is needed. For this, we provide a confidence interval (CI) for Δ^{inf} based on $\widehat{\Delta}_n^{\text{inf}}$. The estimator $\widehat{\Delta}_n^{\text{inf}}$ and accompanying CI provide a measure of misspecification that we recommend reporting alongside confidence sets (CS's) for θ or CI's for elements of θ. A CI for Δ^{inf} that only includes nonpositive values is strong evidence of a non-empty identified set. A CI that only includes positive values is strong evidence of misspecification. The misspecification index estimator $\widehat{\Delta}_n^{\text{inf}}$ and CI also can be used to help determine the choice of inequalities to employ. This can be done in a manner analogous to the way specification tests are used in moment equality and inequality models for this purpose. The misspecification diagnostics considered here, however, have the advantage that they avoid the problem with specification tests that failure to reject correct specification does not provide evidence in favor of correct specification because specification tests may have low power.

We also provide a test of $H_0 : \Delta^{\inf} > 0$ versus $H_1 : \Delta^{\inf} \leq 0$, which is a test of the null hypothesis that the model is misspecified and the identified set is empty against the alternative that the identified set is non-empty. These hypotheses are the reverse of the hypotheses considered by typical model specification tests, for which the null hypothesis is that the model is correctly specified and the alternative is misspecification: $H_{00} : \Delta^{\inf} \leq 0$ versus $H_{11} : \Delta^{\inf} > 0$. It is often of greater interest to consider the null hypothesis H_0 than H_{00} because rejection of H_0 provides strong evidence that the identified set is not empty.

Note that it is common in the moment inequality literature to say that the null hypothesis of a specification test is correct specification, and the alternative is misspecification. But, more precisely, the null hypothesis is a non-empty identified set, and the alternative is an empty identified set. If the identified set is empty, then the model is misspecified. But, it is quite possible for the identified set to be non-empty and the model to be misspecified. Hence, we employ the term *identifiable misspecification*, which means that the identified set is empty. If $\Delta^{\inf}$ is positive, then the model is identifiably misspecified, and if $\Delta^{\inf}$ is negative, the model is not identifiably misspecified.

In the airline entry empirical illustration, a level .05 specification test does not reject H_{00}. However, this should not be interpreted as evidence for correct specification, because the new test of H_0 also does not reject the null. The 95% CI for $\Delta^{\inf}$ is [-0.023, 0.052]. A value of .052 for $\Delta^{\inf}$ can cause a noticeable size distortion of a standard CS for θ, i.e., spurious precision.

Computation of the two-sided misspecification index CI and the SPUR2 CS (or projection CI's based on it) is more intensive than computation of the popular specification test of Bugni, Canay, and Shi (2015) (BCS) and a standard GMS CS (or projection CI's based on it), respectively. However, a large part of the computing time is spent calculating certain bootstrap values and this is easily done in parallel. Depending on the model and data set, computation may be fast and simple, or slow and difficult.

There is a fairly extensive literature on inference methods for moment inequality models, see the review papers of Canay and Shaikh (2016) and Molinari (2020) for references. In particular, see Molinari (2020, Section 5) for a discussion of misspecification in moment inequality models. Several papers provide tests of model misspecification, including Guggenberger, Hahn, and Kim (2008), Romano and Shaikh (2008), Andrews and Guggenberger (2009), Galichon and Henry (2009), Andrews and Soares (2010), Santos (2012), and BCS. Bugni, Canay, and Guggenberger (2012) analyze the behavior of standard tests for moment inequality models under local model misspecification. Ponomareva and Tamer (2011) and Kaido and White (2013) consider estimation of misspecified moment inequality models. They provide consistency results, but do not consider inference. Both employ nonparametric estimation methods. Ponomareva and Tamer (2011) focus on the linear regression model with an interval-valued outcome. Kaido and White (2013) assume that some nonparametric moment inequalities are correctly specified and misspecification is due to a parametric functional form, as opposed to, say, missing variables, mismeasured variables, or unanticipated endogeneity. Allen and Rehbeck (2019) consider a similar measure to Δ^{inf} and provide a CI for it in their study of demand based on quasilinear utility. However, in their setting, there is no unknown parameter θ, which restricts applicability and simplifies the problem considerably.

The results of Chen, Christensen, and Tamer (2018) cover inference for the identified set, as opposed to the true parameter, in moment inequality models. Their approach naturally leads to a non-empty pseudo-true identified set under misspecification, which consists of the parameter values that minimize the population objective function. In consequence, their CS's for the identified set should be consistent for this pseudo-true identified set under misspecification. Their high-level assumptions do not apply to misspecified moment inequality models, so it is an open question whether their methods yield correct asymptotic size for CS's concerning the identified set in such circumstances.

Masten and Poirier (2020) and Kédangi, Li, and Mourifié (2021) define various sets of pseudo-true parameters for misspecified partially-identified models based on different methods of relaxing the assumptions of a partially-identified model. The MR-identified set in this paper can be viewed as one such relaxation. Neither of the aforementioned papers considers inference for a set of pseudo-true parameters.

The misspecification index CI introduced here provides a new tool that can be quite useful for assessing the magnitude of model misspecification in standard moment equality models (which typically are estimated by generalized method of moments (GMM)). One forms moment inequalities by writing each equality as two inequalities and the resulting misspecification index CI gives a CI for the minimum-over-θ maximum-over-moment-functions violation of the moment equalities.

The methods introduced in this paper are robust to weak identification. The tests concerning θ apply to full vector inference. Projection can be used to obtain inference for subvectors. Alternative subvector methods are the focus of ongoing research.

The paper is organized as follows. Section 2 illustrates the spurious precision of some standard CI's in the literature and discusses the misspecification index in the context of a simple lower/upper bound moment inequality model. Section 3 describes the general moment inequality model considered in the paper and defines the MR-identified set. Section 4 introduces the SPUR1 and SPUR2 tests and CS's. Section 5 introduces the misspecification diagnostics. Section 6 illustrates the use of the SPUR2 CS's and misspecification index CI in the context of a binary entry game for airlines analyzed by Kline and Tamer (2016) and Kaido, Molinari, and Stoye (2019).

Online Appendix A provides simulation results for the misspecification index CI and for the size and power properties of the SPUR tests in lower/upper bound models and missing data models, including results concerning the sensitivity of the SPUR2 tests to the tuning parameters. Online Appendix A also provides additional results concerning the airline entry empirical illustration, including sensitivity results to the tuning parameters, power results, and a description of the initial values used in the optimization problems that deliver GMS and SPUR projection CI's. In addition, online Appendix A establishes the uniform consistency and rate of convergence of an estimator of the MR-identified set. Online Appendix B proves the results of the paper concerning the SPUR tests and CS's. Online Appendix C proves the results of the paper concerning the misspecification index. Asymptotic $n^{-1/2}$-local power and consistency results for the SPUR tests are given in the Supplemental Material to Andrews and Kwon (2019).

Let $:=$ denote "equals by definition." Let $[x]_- := \max\{-x, 0\}$ (≥ 0) for $x \in R$.

2 A Simple Example

In this section, we consider the simple lower/upper bound moment inequality model. We use it to illustrate the problem of spurious precision of standard CI's for a parameter θ and the form of the misspecification index $\Delta^{\inf}$. Suppose $\{W_i\}_{i \leq n}$ are i.i.d. with $W_i = (W_{i1}, W_{i2})' \sim N((\mu_1, \mu_2)', I_2)$, where I_2 is the 2×2 identity matrix. The unknown parameter is $\theta \in R$, the moment inequalities are $\mu_1 \leq \theta \leq \mu_2$, and the identified set is $[\mu_1, \mu_2]$. If $\mu_1 > \mu_2$, the model is identifiably misspecified.

We illustrate the problem of spurious precision of standard CI's in this model. First, consider the simple 95% CI for θ defined by $CI_{simp,n} := [\overline{W}_{n1} - 1.955/n^{1/2}, \overline{W}_{n2} + 1.955/n^{1/2}]$.[2] If the model is misspecified and $\mu_1 > \mu_2$, then the identified set is empty, $CI_{simp,n}$ is too narrow because the distribution of $\overline{W}_{n1}$ is to the right of that of $\overline{W}_{n2}$, and there is no value $\theta \in R$ that is covered by $CI_{simp,n}$ with probability .95 or greater. In this case, we say that $CI_{simp,n}$ is *spuriously precise*. It is easy to see that the value of θ for which the coverage probability is greatest in the misspecified model is $(\mu_1+\mu_2)/2$. Figure 2.1 graphs this maximum coverage probability as a function of the amount of misspecification $r^{\inf}$, which equals $(\mu_1 - \mu_2)/2$ in this model when $\mu_1 \geq \mu_2$, for $n = 250$. Figure 2.1 shows that the maximum coverage probability decreases quite rapidly as $r^{\inf}$ increases and is noticeably lower than .95 even for relatively small values of $r^{\inf}$. Figure 2.1 also graphs the expected length of $CI_{simp,n}$ as a function of $r^{\inf}$ for $n = 250$. One sees clearly that the expected length decreases as the magnitude of misspecification increases. This can be misleading

[2] Here, 1.955 is the $1 - \tau$ standard normal quantile for $\tau = (1 - .95)^{1/2}$, which gives .95 coverage when $\mu_1 = \mu_2$ and greater than .95 coverage when $\mu_1 < \mu_2$.

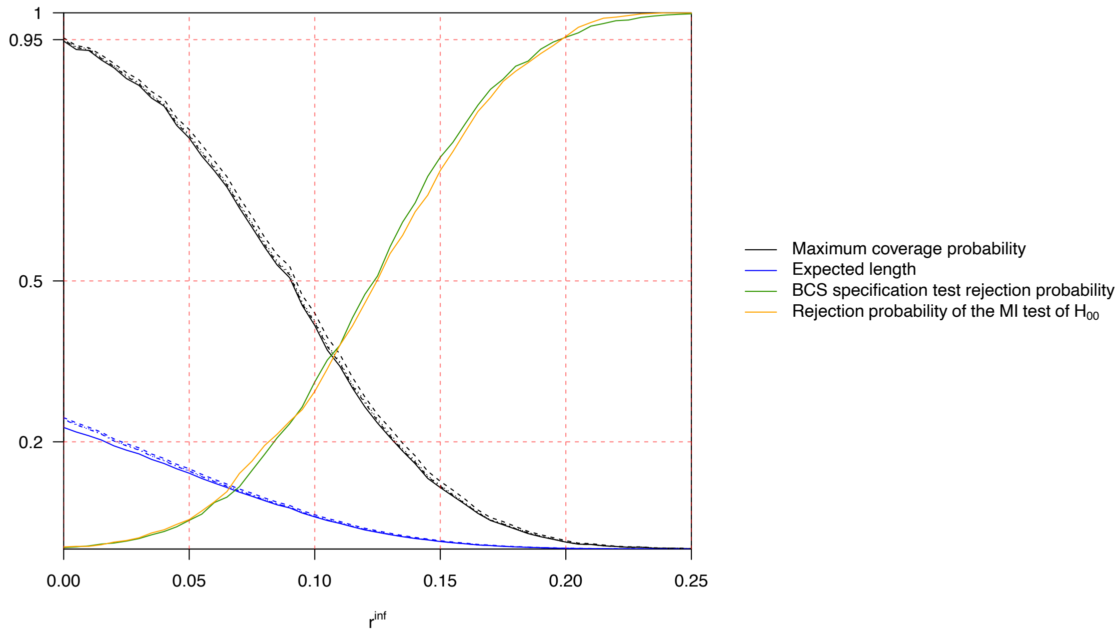

Figure 2.1: Maximum coverage probabilities for any $\theta \in \Theta$ and expected confidence interval lengths for the 95% simple confidence interval and the standard GMS, RSW, and CCK confidence intervals and rejection probabilities for the 5% BCS specification test based on the max test function and the 5% MI test of H_{00} in the lower/upper bound model with k=2 under model misspecification indexed by $r^{\inf}$. The coverage probabilities and lengths of the GMS, RSW, and CCK confidence intervals overlap too much to separately label them.

to the practitioner because it shows that a short CI can be due to misspecification, rather than to informative data.

An analogue of the simple CI $CI_{simp,n}$ is not available in more general moment inequality models. In addition, this CI over-covers the true value in the correctly specified model when $\mu_1 < \mu_2$. In consequence, a number of other CI's and CS's have been developed in the literature, including those in Andrews and Soares (2010), Romano, Shaikh, and Wolf (2014), and Chernozhukov, Chetverikov, and Kato (2019). These CI's also exhibit spurious precision, as shown in Figure 2.1.[3],[4] In fact, in this simple model, the spurious precision of all of these CI's is so similar that it is difficult to

[3]In Figure 2.1, these three CI's are all based on the "max" test statistic. The critical values employed for the Chernozhukov, Chetverikov, and Kato (2019) CI are their "self-normalized" critical values. In this simple model, their other critical values would perform similarly.

[4]Other moment inequality methods, which are not considered in Figure 2.1, also can be shown to exhibit spurious precision under misspecification. This includes the methods in Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Chiburis (2009), Galichon and Henry (2009), Bugni (2010), Canay (2010), Romano and Shaikh (2010, Ex. 2.3), Andrews and Barwick (2012), Romano, Shaikh, and Wolf (2014), Bugni, Canay, and Shi (2017), Cox and Shi (2021), and Kaido, Molinari, and Stoye (2019). Methods designed for conditional moment inequalities also exhibit spurious precision under misspecification.

distinguish the performance of one from another in Figure 2.1. In consequence, we do not label which graph corresponds to which CI. The phenomenon of spurious precision for these CS's arises in more general models than the lower/upper bound model for the same underlying reasons. For the GMS CS's in Andrews and Soares (2010), this is shown in Andrews and Kwon (2019, Sec. 3).

Figure 2.1 also graphs the power of the BCS resampling specification test. The power of the BCS test is low over a wide range of values of $r^{\inf}$ that yield substantial spurious precision of the CI's considered. For example, for $r^{\inf} = .075$, the maximum coverage of the simple CI is only .60 and the power of the BCS test to detect this level of misspecification is only .15. For $r^{\inf} = .125$, the corresponding values are .23 and .51. This clearly illustrates that one cannot rely on a model specification test to detect misspecification over a range of levels that cause substantial spurious precision.

The power of the MI test of H_{00} also is graphed in Figure 2.1. Not surprisingly, its power is essentially equal to that of the BCS test.

In the lower/upper bound model, the relaxed moment inequalities are $\theta - \mu_1 + r^{\inf} \geq 0$ and $\mu_2 - \theta + r^{\inf} \geq 0$. Let $r(\theta)$ (≥ 0) be the smallest value such that θ satisfies relaxed moment inequalities. Then, $r^{\inf} := \inf_{\theta \in R} r(\theta)$. By definition, $r(\theta) := \inf\{r \geq 0 : \theta - \mu_1 + r \geq 0, \mu_2 - \theta + r \geq 0\}$ $= \max\{\mu_1 - \theta, \theta - \mu_2, 0\}$. Some calculations give $r^{\inf} = \max\{(\mu_1 - \mu_2)/2, 0\}$. When $\mu_1 \geq \mu_2$, the MR-identified set consists of the single point $(\mu_1 + \mu_2)/2$.

To construct a CI that has correct coverage for a value in the MR-identified set, we estimate $r^{\inf}$ by its sample analogue $\widehat{r}_n^{\inf} = \max\{(\overline{W}_{n1} - \overline{W}_{n2})/2, 0\}$ and construct a CI for θ based on the relaxed sample moment inequalities $\theta - \overline{W}_{n1} + \widehat{r}_n^{\inf} \geq 0$ and $\overline{W}_{n2} - \theta + \widehat{r}_n^{\inf} \geq 0$. We employ a standard test statistic applied to these relaxed sample moments, such as $\max\{[\theta - \overline{W}_{n1} + \widehat{r}_n^{\inf}]_-, [\overline{W}_{n2} - \theta + \widehat{r}_n^{\inf}]_-\}$. We use a bootstrap critical value that takes into account the extra randomness due to $\widehat{r}_n^{\inf}$. The resulting CI is referred to as a SPUR1 CI.

To circumvent the drawbacks of specification tests, we introduce a misspecification index $\Delta^{\inf}$. It is defined to be the minimum over θ of the maximum moment inequality violation. That is, $\Delta^{\inf} := \inf_{\theta \in R} \max\{\mu_1 - \theta, \theta - \mu_2\} = (\mu_1 - \mu_2)/2$. If the model is misspecified, then $\mu_1 > \mu_2$ and $\Delta^{\inf} > 0$. If the model is correctly specified, then $\mu_1 \leq \mu_2$, $\Delta^{\inf} \leq 0$, and $-\Delta^{\inf}$ is the maximum slackness of any point in the identified set $[\mu_1, \mu_2]$. As defined, $r^{\inf} = \max\{\Delta^{\inf}, 0\}$ in this model and all other models. The sample analogue of $\Delta^{\inf}$ is $\widehat{\Delta}_n^{\inf} := \inf_{\theta \in R} \max\{\overline{W}_{n1} - \theta,$ $\theta - \overline{W}_{n2}\} = (\overline{W}_{n1} - \overline{W}_{n2})/2$, where $\overline{W}_{nj} = n^{-1}\sum_{i=1}^{n} W_{ij}$ for $j = 1, 2$. This paper introduces a nominal level α CI for $\Delta^{\inf}$ that is of the form $[\widehat{\Delta}_n^{\inf} - \widehat{c}_{n,\Delta L}(1-\alpha/2)/n^{1/2}, \widehat{\Delta}_n^{\inf} + \widehat{c}_{n,\Delta U}(1-\alpha/2)/n^{1/2}]$, where $\widehat{c}_{n,\Delta L}(1 - \alpha/2)$ and $\widehat{c}_{n,\Delta U}(1 - \alpha/2)$ are bootstrap critical values. If the CI upper bound for $\Delta^{\inf}$ is 0 or less, the CI provides evidence that the model is not identifiably misspecified. This is the reverse of a typical model specification test. If the CI lower bound for $\Delta^{\inf}$ is positive, the CI provides evidence that the model is misspecified and performs as a model specification test, but also provides information on the magnitude of misspecification.

The misspecification index CI can be used to improve the SPUR1 CI for θ. Our recommended SPUR2 CI for θ is a level $1-\alpha$ Bonferroni CI that equals a level α_2 standard GMS CI if the level $1-\alpha_1$ upper bound CI for $\Delta^{\inf}$ only includes 0 and otherwise is the union of the level $1-\alpha_2$ SPUR1 and GMS CI's, where $\alpha_1+\alpha_2=\alpha$.

3 Moment Inequality Model and MR-Identified Set

3.1 Model and Misspecification-Robust Identified Set

Here, we introduce the general moment inequality model considered in the paper:

$$E_F m(W_i,\theta) \geq 0_k \tag{3.1}$$

for $i \leq n$, where $0_k=(0,...,0)' \in R^k$, the inequality holds when the model is correctly specified and $\theta \in \Theta \subset R^{d_\theta}$ is the true value, $\{W_i \in \mathcal{W} \subset R^{d_W} : i=1,...,n\}$ are independent and identically distributed (i.i.d.) observations with distribution F, $m(\cdot,\cdot)$ is a known function from $\mathcal{W}\times\Theta$ to R^k, and E_F denotes expectation under F. The distribution F lies in a set of distributions $\mathcal{P}$. For simplicity, we let W denote a random vector with the same distribution as W_i for any $i \leq n$.

The population variances of the moment functions are

$$\sigma^2_{Fj}(\theta) := Var_F(m_j(W,\theta)) > 0 \text{ for } j \leq k, \tag{3.2}$$

where $m_j(W,\theta)$ is the jth element of $m(W,\theta)$. The population-standard-deviation-normalized moments are

$$\widetilde{m}(W,\theta) := (\widetilde{m}_1(W,\theta),...,\widetilde{m}_k(W,\theta))', \text{ where } \widetilde{m}_j(W,\theta) := \frac{m_j(W,\theta)}{\sigma_{Fj}(\theta)} \text{ for } j \leq k. \tag{3.3}$$

The moment inequality model in (3.1) can be written equivalently as $E_F\widetilde{m}(W,\theta) \geq 0_k$.

The identified set under F is

$$\Theta_I(F) := \{\theta \in \Theta : E_F\widetilde{m}(W,\theta) \geq 0_k\}, \tag{3.4}$$

which is non-empty under correct specification. Under model misspecification, i.e., when (3.1) fails to hold, this set can be empty. This can lead to inference under misspecification that is *spuriously* *precise* (i.e., a confidence set that is sufficiently small or empty such that it does not cover any parameter value with the desired coverage probability).

Now we define a minimally-relaxed identified set that is non-empty under both correct specification and misspecification. Let

$$\begin{aligned} r_F(\theta) &:= \inf\{r \geq 0 : E_F \widetilde{m}(W,\theta) + r1_k \geq 0_k\} = \max_{j \leq k} r_{Fj}(\theta), \text{ where} \\ r_{Fj}(\theta) &:= [E_F \widetilde{m}_j(W,\theta)]_-, \\ r_F^{\inf} &:= \inf_{\theta \in \Theta} r_F(\theta), \end{aligned} \tag{3.5}$$

and $1_k = (1, ..., 1)' \in R^k$. As defined, $r_F(\theta)$ is the minimal relaxation of the moment inequalities such that θ satisfies the relaxed inequalities, and $r_F^{\inf}$ is the minimal relaxation of the moment inequalities such that some $\theta \in \Theta$ satisfies the relaxed inequalities.

We define the MR-identified set to be

$$\Theta_I^{MR}(F) := \{\theta \in \Theta : r_F(\theta) = r_F^{\inf}\} = \{\theta \in \Theta : E_F \widetilde{m}(W,\theta) + r_F^{\inf} 1_k \geq 0_k\}. \tag{3.6}$$

The population quantity $r_F(\theta) - r_F^{\inf}$ is nonnegative and its zeros give the values in the MR-identified set. Under mild conditions (given in Assumption A.0 below), this MR-identified set is non-empty even under model misspecification.

As defined, the MR-identified set has the attribute that it does not depend on the choice of the test statistic that is used for inference on θ, as occurs with the definition of the pseudo-true value in GMM models.

3.2 Basic Assumptions

We employ the following assumptions on the parameter space $\mathcal{P}$ of distributions F.

Assumption A.0. (i) Θ is compact and non-empty and (ii) $E_F \widetilde{m}_j(W,\theta)$ is upper semi-continuous on Θ $\forall j \leq k$, $\forall F \in \mathcal{P}$.

Assumption A.1. The observations $W_1, ..., W_n$ are i.i.d. under F and $\{\widetilde{m}_j(\cdot,\theta) : \mathcal{W} \to R\}$ and $\{\widetilde{m}_j^2(\cdot,\theta) : \mathcal{W} \to R\}$ are measurable classes of functions indexed by $\theta \in \Theta$ $\forall j \leq k$, $\forall F \in \mathcal{P}$.

Assumption A.2. For some $a > 0$, $\sup_{F \in \mathcal{P}} E_F \sup_{\theta \in \Theta} ||\widetilde{m}(W,\theta)||^{4+a} < \infty$.

Assumption A.0 guarantees that the MR-identified set $\Theta_I^{MR}(F)$ in (3.6) is non-empty. Assumption A.2 requires $4 + a$ moments finite, rather than just $2 + a$, because estimators of $\{\sigma_{Fj}^2(\theta) : j \leq k\}$ affect the asymptotic distributions of the test statistics under misspecification, which is not the case under correct specification, and these estimators depend on sample second moments. For ease of reading, some additional equicontinuity assumptions, which are not very restrictive, are stated in Section 14 in online Appendix B.

4 SPUR Tests and Confidence Sets

In this section, we introduce tests of the hypotheses:

$$H_0 : \theta_0 \in \Theta_I^{MR}(F) \text{ versus } H_1 : \theta_0 \notin \Theta_I^{MR}(F) \tag{4.1}$$

for a given (known) $\theta_0 \in \Theta$ and unknown $F \in \mathcal{P}$. We also introduce CS's for a parameter value θ in $\Theta_I^{MR}(F)$. We introduce two tests, called SPUR1 and SPUR2 tests, and their CS counterparts, that are robust to spurious precision. The SPUR2 test and CS are our recommended test and CS. They are based on the SPUR1 test and CS, so the SPUR1 test and CS also are defined here.

4.1 SPUR1 Tests and CS's

The sample moments, variances, and standard-deviation-normalized moments are

$$\begin{aligned}
\overline{m}_{nj}(\theta) &:= n^{-1}\sum_{i=1}^{n} m_j(W_i,\theta),\ \widehat{\sigma}_{nj}^2(\theta) := n^{-1}\sum_{i=1}^{n}(m_j(W_i,\theta)-\overline{m}_{nj}(\theta))^2, \\
\widehat{m}_{nj}(\theta) &:= \frac{\overline{m}_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)} \text{ for } j \le k, \text{ and } \widehat{m}_n(\theta) = (\widehat{m}_{n1}(\theta),...,\widehat{m}_{nk}(\theta))'.
\end{aligned} \tag{4.2}$$

The sample correlation matrix of the moments is

$$\widehat{\Omega}_n(\theta) := n^{-1}\sum_{i=1}^{n}\widehat{m}_n(W_i,\theta)\widehat{m}_n(W_i,\theta)' \in R^{k\times k}, \text{ where } \widehat{m}_{nj}(W,\theta) := (m_j(W,\theta)-\overline{m}_{nj}(\theta))/\widehat{\sigma}_{nj}(\theta) \tag{4.3}$$

and $\widehat{m}_n(W,\theta)$ has jth element $\widehat{m}_{nj}(W,\theta)$ for $j \le k$.

Estimators of $r_{Fj}(\theta)$, $r_F(\theta)$, and $r_F^{\inf}$ defined in (3.5) are

$$\widehat{r}_{nj}(\theta) := [\widehat{m}_{nj}(\theta)]_-\,,\ \widehat{r}_n(\theta) := \max_{j\le k}\widehat{r}_{nj}(\theta), \text{ and } \widehat{r}_n^{\inf} := \inf_{\theta\in\Theta}\widehat{r}_n(\theta). \tag{4.4}$$

We base a test of $H_0 : \theta_0 \in \Theta_I^{MR}(F)$ on the SPUR test statistic $S_n(\theta_0)$, where

$$S_n(\theta) := S\left(n^{1/2}\left(\widehat{m}_n(\theta)+\widehat{r}_n^{\inf}1_k\right),\widehat{\Omega}_n(\theta)\right) \tag{4.5}$$

and $S(m,\Omega)$ is a test function that satisfies certain conditions.[5] Examples of such functions are

$$S_1(m,\Omega) := \sum_{j=1}^{k}[m_j]_-^2,\ \ S_2(m,\Omega) := \inf_{t\in[0,\infty]^k}(m-t)'\Omega^{-1}(m-t),\ \ S_4(m,\Omega) := \max_{j\le k}[m_j]_-, \tag{4.6}$$

and $S_{2A}(m,\Omega)$ defined in Andrews and Barwick (2012).[6]

For testing $H_0 : \theta_0 \in \Theta_I^{MR}(F)$, the nominal level α SPUR1 test $\phi_{n,SPUR1}(\theta_0)$ rejects H_0 if

$$\phi_{n,SPUR1}(\theta_0) = 1,\ \text{ where } \phi_{n,SPUR1}(\theta) := 1(S_n(\theta) > \widehat{c}_n(\theta, 1-\alpha)) \tag{4.7}$$

and $\widehat{c}_n(\theta,1-\alpha)$ is an extended GMS (EGMS) bootstrap critical value. Because the definition of $\widehat{c}_n(\theta,1-\alpha)$ is complicated, we motivate, define, and discuss it in Sections 4.3, 4.4, and 4.5 below, rather than here.

The nominal level $1-\alpha$ SPUR1 CS for θ is

$$CS_{n,SPUR1} := \{\theta \in \Theta : \phi_{n,SPUR1}(\theta) = 0\}. \tag{4.8}$$

An alternative to the SPUR test statistic in (4.5) is a recentered test statistic, such as considered in Chernozhukov, Hong, and Tamer (2007) (CHT). It is defined to be $S_{n,Recen}(\theta) := S_{n,Std}(\theta) - \inf_{\overline{\theta}\in\Theta} S_{n,Std}(\overline{\theta})$, where $S_{n,Std}(\theta)$ is a "standard" test statistic, as in (4.5), but with $\widehat{r}_n^{\inf} = 0$. When the SPUR and recentered test statistics are based on the "max" S_4 function in (4.6), they are identical, see Section 14 in the Supplemental Material to Andrews and Kwon (2019). (Note that this is a statement about the test statistics, not about the SPUR1 and CHT tests, which use different critical values.)

4.2 SPUR2 Tests and CS's

Next, we introduce our recommended test for the hypotheses in (4.1) and corresponding CS. It is a Bonferroni test (and corresponding CS) that combines a standard GMS test that assumes correct model specification with the SPUR1 test just defined. We call it the SPUR2 test. The SPUR2 test uses a CI for $r_F^{\inf}$ that is generated by the upper bound CI for $\Delta_F^{\inf}$ in Section 5.1 below using the fact that $r_F^{\inf} = \max\{\Delta_F^{\inf}, 0\}$. Let $\alpha = \alpha_1 + \alpha_2 \in (0,1)$ for $\alpha_1, \alpha_2 > 0$, such as $\alpha_1 = .005$

[5] These conditions are: Assumption S.1. (i) $S(m,\Omega)$ is nonincreasing in $m \in R^k_{[+\infty]}$ $\forall\Omega\in\Psi$, where $\Psi := cl(\{\Omega_F(\theta) : \theta\in\Theta, F\in\mathcal{P}\})$, $cl(\cdot)$ denotes the closure of a set, and $\Omega_F(\theta) := Corr_F(m(W,\theta)) \in R^{k\times k}$, and (ii) $S(m,\Omega) \ge 0$ $\forall m\in R^k$, $\forall\Omega\in\Psi$, and (iii) $S(m,\Omega)$ is continuous at all $m \in (R\cup\{+\infty\})^k$ and $\Omega\in\Psi$. Assumption S.2. $S(m,\Omega) > 0$ iff $m_j < 0$ for some $j \le k$, $\forall\Omega\in\Psi$. Assumption S.3. For some $\chi > 0$, $S(am,\Omega) = a^\chi S(m,\Omega)$ $\forall a > 0$, $\forall m \in R^k$, $\forall\Omega\in\Psi$.

[6] The function $S_2(m,\Omega)$ satisfies the conditions in the previous footnote provided $\inf_{\Omega\in\Psi}\det(\Omega) > 0$.

and $\alpha_2 = .045$. The SPUR2 test is adaptive in the sense that if the level $1-\alpha_1$ CI for $r_F^{\inf}$ contains only the single point 0, so the data indicate that the identified set is not empty, then the test is the same as the standard GMS test with level α_2, rather than α. But, if the CI for $r_F^{\inf}$ contains positive values, then the test rejects the null if both the SPUR1 and GMS tests reject with level α_2, rather than α (which typically is equivalent to the SPUR1 test rejecting at level α_2). The SPUR2 test is robust to spurious precision caused by misspecification. Simulations show that the SPUR2 test has good power properties relative to the SPUR1 test under correct specification and misspecification, see Section 8 in online Appendix A. The SPUR2 test also has computational advantages relative to the SPUR1 test in scenarios where the CI for $r_F^{\inf}$ contains only the point 0 because it only requires the computation of the GMS test in these scenarios.

The nominal $1-\alpha_1$ one-sided upper-bound CI for $r_F^{\inf}$ that we employ is

$$CI_{n,r,UP}(\alpha) := [0, \widehat{r}_{n,UP}(\alpha)], \text{ where } \widehat{r}_{n,UP}(\alpha) := \max\{\widehat{\Delta}_{n,U}^{\inf}(\alpha), 0\} \tag{4.9}$$

and $\widehat{\Delta}_{n,U}^{\inf}(\alpha)$ is defined in (5.3) below. This CI equals $\{0\}$ wp$\to$1 when the identified set is nonempty and the sequence of identified sets $\{\Theta_I(F_n)\}_{n\geq 1}$ contains slack points with slackness of order greater than $n^{-1/2}$. That is, $\lim_{n\to\infty} \min_{j\leq k} n^{1/2} E_{F_n} m_j(W, \theta_n^I) = \infty$ for some $\{\theta_n^I \in \Theta_I(F_n)\}_{n\geq 1}$. For example, for a fixed distribution F, if $\Theta_I(F)$ contains a slack point, i.e., a point θ^I with $\min_{j\leq k} E_F m_j(W, \theta^I) > 0$, then $CI_{n,r,UP}(\alpha) = \{0\}$ wp$\to$1. On the other hand, when the model exhibits "large-local" or "global" model misspecification, i.e., when $\{F_n\}_{n\geq 1}$ is such that $n^{1/2} r_{F_n}^{\inf} \to \infty$, then $\widehat{r}_{n,UP}(\alpha) > 0$ wp$\to$1.

Note that $\widehat{r}_{n,UP}(\alpha)$ is not based on $\widehat{r}_n^{\inf}$. Rather, it is based on the statistic $\widehat{\Delta}_n^{\inf}$ that is negative when the sample moments are all slack at some value $\theta \in \Theta$ and equals $\widehat{r}_n^{\inf}$ when $\widehat{r}_n^{\inf} > 0$. This is key for the property of $CI_{n,r,UP}(\alpha)$ under correct specification described above.

Let $\phi_{n,GMS}(\theta_0, \alpha_2)$ denote a nominal level α_2 GMS test that assumes correct model specification. It is based on the test statistic $S_{n,Std}(\theta) := S(n^{1/2}\widehat{m}_n(\theta), \widehat{\Omega}_n(\theta))$ and a GMS critical value $\widehat{c}_{n,GMS}(\theta, 1-\alpha_2)$, which is the $1-\alpha_2$ sample quantile of $\{S_{n,GMS,b}^*(\theta)\}_{b\leq B}$, where $S(m,\Omega)$ is the test function considered above. By definition, $S_{n,GMS,b}^*(\theta) := S(T_{n,GMS,b}^*(\theta), \widehat{\Omega}_n(\theta))$, where $T_{n,GMS,b}^*(\theta)$ has jth element equal to $n^{1/2}(\overline{m}_{njb}^*(\theta) - \overline{m}_{nj}(\theta))/\widehat{\sigma}_{njb}^*(\theta) + \varphi(\kappa_n^{-1} n^{1/2}\widehat{m}_{nj}(\theta))$ for $j \leq k$, where $\varphi(\cdot)$ and κ_n are a standard GMS function and GMS tuning parameter, respectively, defined as in (4.19) below. By definition, $\phi_{n,GMS}(\theta, \alpha_2) := 1(S_{n,Std}(\theta) > \widehat{c}_{n,GMS}(\theta, 1-\alpha_2))$.

The nominal level α SPUR2 test of $H_0 : \theta_0 \in \Theta_I^{MR}(F)$ versus $H_1 : \theta_0 \notin \Theta_I^{MR}(F)$ is

$$\begin{aligned}\phi_{n,SPUR2}(\theta_0) := {} & 1(\widehat{r}_{n,UP}(\alpha_1) = 0)\phi_{n,GMS}(\theta_0, \alpha_2) \\ & +1(\widehat{r}_{n,UP}(\alpha_1) > 0) \min\{\phi_{n,SPUR1}(\theta_0, \alpha_2), \phi_{n,GMS}(\theta_0, \alpha_2)\},\end{aligned} \tag{4.10}$$

where $\phi_{n,SPUR1}(\theta_0, \alpha_2)$ denotes the SPUR1 test of $H_0 : \theta \in \Theta_I^{MR}(F)$ in (4.7) with α_2 in place of α.

The nominal level $1 - \alpha$ SPUR2 CS for $\theta \in \Theta_I^{MR}(F)$ is

$$CS_{n,SPUR2} := \{\theta \in \Theta : \phi_{n,SPUR2}(\theta) = 0\}. \tag{4.11}$$

The SPUR2 test $\phi_{n,SPUR2}(\theta_0)$ and CS $CS_{n,SPUR2}$ are our recommended test and CS for inference on θ.

Note that the SPUR2 test and CS also can be constructed using any test in place of the GMS test in (4.10), such as the test in Romano, Shaikh, and Wolf (2014), provided the test has correct asymptotic size under correct model specification.

4.3 Intuition Behind the SPUR1 Critical Value $\widehat{\mathbf{c}}_{\mathbf{n}}(\theta, \mathbf{1} - \alpha)$

To complete the definition of the SPUR tests and CS's, it remains to define the SPUR1 critical value $\widehat{c}_n(\theta, 1-\alpha)$ in (4.7). Because its definition is complicated, we start by providing some intuition behind the definition.

We refer to the SPUR1 critical value $\widehat{c}_n(\theta, 1-\alpha)$ as an EGMS bootstrap critical value because it is based on an extension of the GMS-type critical value employed by many tests that are designed for correct model specification. The critical value $\widehat{c}_n(\theta, 1-\alpha)$ is based on a bootstrap statistic $S_n^*(\theta)$ which, in turn, is based on the asymptotic null distribution of the test statistic $S_n(\theta)$. The test statistic $S_n(\theta)$ in (4.5) can be written as

$$S_n(\theta) := S(\underbrace{n^{1/2}(\widehat{m}_n(\theta) + r_{F_n}^{\inf} 1_k)}_{:=T_n(\theta)} + \underbrace{n^{1/2}(\widehat{r}_n^{\inf} - r_{F_n}^{\inf})}_{:=A_n} 1_k, \widehat{\Omega}_n(\theta)), \tag{4.12}$$

where F_n is the distribution of the sample. The bootstrap statistic $S_n^*(\theta)$ is based on bootstrap versions $T_{n,b}^*(\theta)$ and $A_{n,b}^*$ of $T_n(\theta)$ and A_n, respectively.

Naive definitions of the bootstrap versions of $T_n(\theta)$ and A_n would use a bootstrap sample in place of the original sample $\{W_i\}_{i \leq n}$ in the construction of $\widehat{m}_n(\theta)$ and $\widehat{r}_n^{\inf}$ and would use $\widehat{r}_n^{\inf}$ in place of $r_{F_n}^{\inf}$ in the expressions for $T_n(\theta)$ and A_n. However, such definitions would not yield a test or CS with the correct asymptotic level because (i) $\widehat{m}_n(\theta)$ is not a mean zero quantity, and (ii) the statistic $\widehat{r}_n^{\inf}$ involves the $\inf_{\theta\in\Theta}$ and $\max_{j\leq k}$ terms. Issue (i) arises with existing moment inequality-based tests and CS's that assume correct model specification. Issue (ii) yields a complicated asymptotic distribution of A_n. Both issues cause the naive bootstrap to fail. For example, see Andrews (2000) for a discussion of the failure of the naive bootstrap in a closely related problem to that of bootstrapping A_n.

The bootstrap version $T^*_{n,b}(\theta)$ of $T_n(\theta)$ that we employ has a similar form to the bootstrap statistic in a standard GMS test. The bootstrap version $A^*_{n,b}$ of A_n is complicated because the asymptotic distribution of A_n depends on several nuisance parameter functions that are not consistently estimable and a particular feature of these functions must be imposed in order to obtain a critical value that does not drift to infinity with the sample size. In contrast, a GMS critical value only has to deal with a finite-dimensional nuisance parameter that is not consistently estimable. The idea behind the EGMS critical value is to shrink estimators of the nuisance functions in a least favorable direction, which is towards $-\infty$. This ensures that the distribution of the bootstrap version of $S_n(\theta)$ is asymptotically as large as that of the asymptotic null distribution of $S_n(\theta)$ in a stochastic sense.

To provide intuition for the definition of these bootstrap counterparts, we first rewrite $T_n(\theta) = (T_{n1}(\theta), ..., T_{nk}(\theta))'$ as

$$\begin{aligned} T_{nj}(\theta) &= \widehat{\nu}_{nj}(\theta) + h_{nj}(\theta), \text{ where} \\ \widehat{\nu}_{nj}(\theta) &:= n^{1/2}(\widehat{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta)) \text{ and } h_{nj}(\theta) := n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta) + r^{\inf}_{F_n}). \end{aligned} \tag{4.13}$$

We approximate the two terms $\widehat{\nu}_{nj}(\theta)$ and $h_{nj}(\theta)$. The centered stochastic process $\{\widehat{\nu}_{nj}(\theta) : \theta \in \Theta\}$ is approximated by its bootstrap analogue $\{\widehat{\nu}^*_{njb}(\theta) : \theta \in \Theta\}$ defined in (4.17) below. The nonstochastic quantity $h_{nj}(\theta)$ is not consistently estimable. We use a GMS-type lower bound, $\varphi(\xi_{nj}(\theta))$, to bound $h_{nj}(\theta)$, where $\varphi(\cdot)$ is a GMS function defined in (4.19) below and $\xi_{nj}(\theta)$ is a rescaled estimator of $h_{nj}(\theta)$ defined in (4.19). This bound is nonnegative for θ in the null hypothesis, as desired.

Next, to motivate the definition of the bootstrap analogue $A^*_{n,b}$ of A_n, we rewrite A_n as

$$A_n = \inf_{\theta\in\Theta}\max_{j\leq k}\left([\widehat{\nu}_{nj}(\theta) + \ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_- + b_{nj}(\theta)\right), \tag{4.14}$$

where $\ell_{nj}(\theta) := n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)$ and $b_{nj}(\theta) := n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - r^{\inf}_{F_n})$. The asymptotic distribution of A_n under $\{F_n\}_{n\geq 1}$ depends on the limit of the stochastic process $\{[\widehat{\nu}_{nj}(\theta) + \ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_- + b_{nj}(\theta) : \theta \in \Theta\}$, where $\ell_{nj}(\theta)$ and $b_{nj}(\theta)$ are nonrandom functions that are not consistently estimable.

The bootstrap statistic $A^*_{n,b}$ that we employ replaces $\inf_{\theta\in\Theta}$ by $\inf_{\theta\in\widehat{\Theta}_n}$, where $\widehat{\Theta}_n$ is a consistent estimator of the MR-identified set $\Theta_I^{MR}(F)$. This replacement is valid because, roughly speaking, the set $\widehat{\Theta}_n$ includes all parameter values that are relevant for the asymptotic distribution of A_n. The use of $\widehat{\Theta}_n$, rather than Θ, typically eases computation because it substantially reduces the size of the set over which the infimum is taken, which reduces the number of initial values that needs to be considered. However, if the use of Θ is computationally easier, it can be used in place of $\widehat{\Theta}_n$ without affecting the asymptotic properties of the SPUR1 test.

To approximate $[\widehat{\nu}_{nj}(\theta)+\ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_-$, we use a bootstrap lower bound $\widehat{\chi}^*_{nj,b}(\theta)$ defined in (4.21) below.

To obtain a bootstrap lower bound for $b_{nj}(\theta)$, we first consider a quantity $\widehat{b}_{nj}(\theta)$ that shifts the sample analogue of $b_{nj}(\theta)$ towards $-\infty$, see (4.22) below. When $b_{nj}(\theta)\geq 0$, a better (GMS-type) lower bound $\varphi(\xi^A_{nj}(\theta))$ is available, see (4.23) below. However, while it is the case that $b_{nj}(\theta)\geq 0$ for some $j\leq k$, we do not know for which j this is true. Therefore, we use an estimated set $\widehat{J}_{nB}(\theta)$ that contains the value(s) j for which this better lower bound can be employed, see (4.24) below. Incorporating the better lower bound is important, because otherwise the critical value would be divergent asymptotically.

4.4 Definition of the SPUR1 Critical Value $\widehat{\mathbf{c}}_{\mathbf{n}}(\theta, \mathbf{1}-\alpha)$

The EGMS bootstrap critical value $\widehat{c}_n(\theta,1-\alpha)$ for the SPUR1 test is based on B bootstrap statistics $\{S^*_{n,b}(\theta)\}_{b\leq B}$, where $S^*_{n,b}(\theta)$ is defined following the intuition outlined in Section 4.3. Let $\{W^*_{ib}\}_{i\leq n}$ for $b=1,...,B$ denote the bootstrap samples, each one of which is an i.i.d. sample drawn with replacement from the original sample $\{W_i\}_{i\leq n}$. That is, the "nonparametric i.i.d." bootstrap is employed.

The bth EGMS bootstrap statistic $S^*_{n,b}(\theta)$ is defined by

$$S^*_{n,b}(\theta) := S\left(T^*_{n,b}(\theta)+A^*_{n,b}1_k, \widehat{\Omega}_n(\theta)\right), \tag{4.15}$$

where $T^*_{n,b}(\theta)=(T^*_{n1,b}(\theta),...,T^*_{nk,b}(\theta))'$ and $A^*_{n,b}$ are defined below. The SPUR1 bootstrap critical value $\widehat{c}_n(\theta,1-\alpha)$ is the $1-\alpha$ sample quantile of $\{S^*_{n,b}(\theta)\}_{b\leq B}$ plus a very small constant $\iota>0$.[7] See Section 4.7.1 below for the recommended choice of ι and other tuning parameters.

In the definition of the bootstrap statistics $T^*_{n,b}(\theta)$ and $A^*_{n,b}$, certain bootstrap standard deviations, $sd^*_{1njB}(\theta)$, $sd^*_{2njB}(\theta)$, and $sd^*_{3njB}(\theta)$, are used to obtain appropriate scaling. Given any bootstrap variables $\{m^*_b\}_{b\leq B}$, we denote the bootstrap sample standard deviation, modified to be

[7] The constant ι increases the critical value by a trivial amount when ι is taken to be very small. Hence, it has little to no effect on the critical value or test.

greater than or equal to a very small constant $\iota > 0$, by

$$SD_B^*(m_b^*) := \max\left\{\left(B^{-1}\sum_{b=1}^{B}\left(m_b^* - B^{-1}\sum_{c=1}^{B}m_c^*\right)^2\right)^{1/2}, \iota\right\}. \tag{4.16}$$

Define the bootstrap analogue of $\widehat{\nu}_{nj}(\theta)$ by $\widehat{\nu}_{njb}^*(\theta)$:

$$\widehat{\nu}_{njb}^*(\theta) := n^{1/2}\left(\frac{\overline{m}_{njb}^*(\theta)}{\widehat{\sigma}_{njb}^*(\theta)} - \widehat{m}_{nj}(\theta)\right),$$
$$\overline{m}_{njb}^*(\theta) := n^{-1}\sum_{i=1}^{n} m_j(W_{ib}^*, \theta), \text{ and } \widehat{\sigma}_{njb}^{*2}(\theta) := n^{-1}\sum_{i=1}^{n}(m_j(W_{ib}^*, \theta) - \overline{m}_{nj}^*(\theta))^2. \tag{4.17}$$

The bootstrap counterpart $T_{nj,b}^*(\theta)$ of $T_{nj}(\theta)$ is

$$T_{nj,b}^*(\theta) := \widehat{\nu}_{njb}^*(\theta) + \varphi(\xi_{nj}(\theta)), \tag{4.18}$$

where

$$\xi_{nj}(\theta) := (sd_{1njB}^*(\theta)\kappa_n)^{-1} n^{1/2}\left(\widehat{m}_{nj}(\theta) + \widehat{r}_n(\theta)\right),$$
$$\varphi(\xi) := \infty 1(\xi > 1) \text{ for } \xi \in R \text{ with } \infty \cdot 0 := 0 \text{ by definition}, \tag{4.19}$$

$sd_{1njB}^*(\theta) := SD_B^*\left(\frac{n^{1/2}\overline{m}_{njb}^*(\theta)}{\widehat{\sigma}_{njb}^*(\theta)} + \max_{j_1 \le k}\left[\frac{n^{1/2}\overline{m}_{nj_1b}^*(\theta)}{\widehat{\sigma}_{nj_1b}^*(\theta)}\right]_-\right)$, and κ_n is a tuning parameter that satisfies $\kappa_n \to \infty$, whose recommended value is specified in Section 4.7.1. Note that $\varphi(\cdot)$ is a standard GMS function.

To obtain a bootstrap counterpart $A_{n,b}^*$ of A_n, we first replace $\inf_{\theta\in\Theta}$ by $\inf_{\theta\in\widehat{\Theta}_n}$, where $\widehat{\Theta}_n$ is a consistent estimator of the MR-identified set:

$$\widehat{\Theta}_n := \{\theta \in \Theta : \max_{j\le k}[\widehat{m}_{nj}(\theta) + \widehat{r}_n^{\inf}]_- \le \tau_n/n^{1/2}\} \tag{4.20}$$

and τ_n is a tuning parameter that satisfies $\tau_n \to \infty$, see Section 4.7.1 for its choice.

Next, for $[\widehat{\nu}_{nj}(\theta) + \ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_-$, we use a bootstrap lower bound $\widehat{\chi}_{nj,b}^*(\theta)$ that employs the function $\chi(\nu, c_1, c_2)$:

$$\widehat{\chi}_{nj,b}^*(\theta) := \chi\left(\widehat{\nu}_{njb}^*(\theta),\ n^{1/2}\widehat{m}_{nj}(\theta) - sd_{2njB}^*(\theta)\kappa_n,\ n^{1/2}\widehat{m}_{nj}(\theta) + sd_{2njB}^*(\theta)\kappa_n\right),$$
$$\chi(\nu, c_1, c_2) := \begin{cases} \chi(\nu, c_1) & \text{if } \nu \ge 0 \\ \chi(\nu, c_2) & \text{if } \nu < 0, \end{cases} \quad \chi(\nu, c) := [\nu + c]_- - [c]_- \text{ for } \nu, c_1, c_2, c \in R, \tag{4.21}$$

and $sd^*_{2njB}(\theta) := SD^*_B(\frac{n^{1/2}\overline{m}^*_{njb}(\theta)}{\widehat{\sigma}^*_{njb}(\theta)})$.

To obtain a bootstrap lower bound for $b_{nj}(\theta)$, we first consider

$$\widehat{b}_{nj}(\theta) := n^{1/2}\left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}\right) - sd^*_{3njB}(\theta)\kappa_n, \tag{4.22}$$

where $sd^*_{3njB}(\theta) := SD^*_B\left(\left[\frac{n^{1/2}\overline{m}^*_{njb}(\theta)}{\widehat{\sigma}^*_{njb}(\theta)}\right]_- - \max_{j_1\leq k}\left[\frac{n^{1/2}\overline{m}^*_{nj_1b}(\theta)}{\widehat{\sigma}^*_{nj_1b}(\theta)}\right]_-\right)$. When $b_{nj}(\theta) \geq 0$, a better (GMS-type) lower bound is $\varphi(\xi^A_{nj}(\theta))$, where

$$\xi^A_{nj}(\theta) := (sd^*_{3njB}(\theta)\kappa_n)^{-1}n^{1/2}\left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf}\right). \tag{4.23}$$

The estimated set that contains the value(s) j for which this better lower bound applies is

$$\widehat{J}_{nB}(\theta) := \{j \in \{1,...,k\} : \widehat{r}_{nj}(\theta) \geq \widehat{r}_n(\theta) - sd^*_{3njB}(\theta)n^{-1/2}\kappa_n\}, \tag{4.24}$$

where $\widehat{r}_{nj}(\theta)$ and $\widehat{r}_n(\theta)$ are defined in (4.4).

The resulting bootstrap lower bound $A^*_{n,b}$ of A_n is

$$A^*_{n,b} := \inf_{\theta\in\widehat{\Theta}_n}\min_{j_1\in\widehat{J}_{nB}(\theta)}\max_{j\leq k}\left(\widehat{\chi}^*_{nj,b}(\theta) + 1(j\neq j_1)\widehat{b}_{nj}(\theta) + 1(j = j_1)\varphi(\xi^A_{nj}(\theta))\right). \tag{4.25}$$

As discussed in Section 4.3, $A^*_{n,b}$ is a bootstrap analogue of A_n with $\widehat{\chi}^*_{nj,b}(\theta)$ in place of $[\widehat{\nu}_{nj}(\theta) + \ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_-$ and $\widehat{b}_{nj}(\theta)$ or $\varphi(\xi^A_{nj}(\theta))$ in place of $b_{nj}(\theta)$ depending on $\widehat{J}_{nB}(\theta)$.

4.5 Technical Discussion of the SPUR1 Critical Value

The addition of $\iota > 0$ to the $1-\alpha$ sample quantile of $\{S^*_{n,b}(\theta)\}_{b\leq B}$ in the definition of $\widehat{c}_n(\theta, 1-\alpha)$ following (4.15) simplifies and weakens the assumptions needed to establish the correct asymptotic size of the SPUR1 test. It circumvents the need to impose assumptions that guarantee that the asymptotic null distribution of the test statistic is continuous at a certain quantile. This use of a very small constant $\iota > 0$ has a trivial impact on finite-sample power and $n^{-1/2}$-local and global asymptotic power, and hence, is innocuous.

The nonrandom quantity $h_{nj}(\theta)$ cannot be consistently estimated because $E_{F_n}\widetilde{m}_j(W,\theta) + r^{\inf}_{F_n}$ can be estimated only with an error of magnitude $O_p(n^{-1/2})$, which is the typical magnitude. In consequence, the error in estimation of $h_{nj}(\theta) = n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta)+r^{\inf}_{F_n})$ is $n^{1/2}O_p(n^{-1/2}) = O_p(1)$, which does not go to zero as $n \to \infty$. The nonrandom quantities $\ell_{nj}(\theta)$ and $b_{nj}(\theta)$ cannot be consistently estimated for analogous reasons.

The set estimator $\widehat{\Theta}_n$ defined in (4.20) is a uniformly (over distributions $F \in \mathcal{P}$) consistent estimator of the MR-identified set $\Theta_I^{MR}(F)$. See online Appendix A for uniform consistency and rate of convergence results for $\widehat{\Theta}_n$.

If the lower bound on $b_{nj}(\theta)$ is arbitrarily small for all j, as can occur for $\widehat{b}_{nj}(\theta)$, then the critical value can be arbitrarily large. This does not reflect the null behavior of the test statistic because $\max_{j\leq k} b_{nj}(\theta)$ can be shown to be nonnegative. In consequence, it is important for power purposes to provide a lower bound on $b_{nj}(\theta)$ that is nonnegative for the value(s) j that attain(s) $\max_{j\leq k} b_{nj}(\theta)$. This is done by using $\varphi(\xi_{nj}^A(\theta))$, which is nonnegative, and $\widehat{J}_{nB}(\theta)$. The set $\widehat{J}_{nB}(\theta)$ is designed such that it includes the value(s) of j that attain $\max_{j\leq k} b_{nj}(\theta)$ wp$\rightarrow$1.

The quantities $sd_{1njB}^*(\theta)$, $sd_{2njB}^*(\theta)$, and $sd_{3njB}^*(\theta)$ are defined so that $n^{1/2}(\widehat{m}_{nj}(\theta)+\widehat{r}_n(\theta))/sd_{1njB}^*(\theta)$, $n^{1/2}\widehat{m}_{nj}(\theta)/sd_{2njB}^*(\theta)$, and $n^{1/2}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf})/sd_{3njB}^*(\theta)$, respectively, each has an asymptotic variance of one after proper centering under correct specification and misspecification, which provides proper scaling for the tuning parameter κ_n. These quantities are defined to be greater than or equal to a very small constant $\iota > 0$ to simplify the conditions needed for asymptotic validity of the SPUR1 test. Alternatively, one could take $\iota = 0$ and impose conditions that imply that the probability limits of $sd_{anjB}^*(\theta)$ for $a = 1, 2, 3$ are necessarily positive.

The quantity $\widehat{\chi}_{nj,EGMS}^*(\theta)$ yields a lower bound on $[\widehat{\nu}_{nj}(\theta) + \ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_-$ because the function $\chi(\nu, c) := [\nu + c]_- - [c]_-$ is nondecreasing in c for $\nu \geq 0$, is zero for all c for $\nu = 0$, and is nonincreasing in c for $\nu < 0$.

We note that ignoring the A_n term in (4.12) and $A_{n,b}^*$ in (4.15) would lead to a standard GMS critical value for $S_n(\theta)$. This would not necessarily lead to correct asymptotic size under model misspecification because the A_n term, which is ignored, can be negative or positive under misspecification since $r_{F_n}^{\inf} > 0$. As noted above, for the S_4 max function, the SPUR statistic equals the recentered statistic $S_{n,Recen}(\theta)$. Hence, this argument implies that combining a recentered statistic $S_{n,Recen}(\theta)$ with the standard GMS critical value also does not necessarily lead to correct asymptotic size under model misspecification.

4.6 Asymptotic Level of the SPUR2 Tests and CS's

Next, we show that the SPUR2 tests and CS's have correct asymptotic level. For brevity, Assumptions A.3–A.6 are stated in Section 14 in online Appendix B.

Theorem 4.1 *Suppose Assumptions* A.0–A.6 *and* S.1 *hold and* $\alpha \in (0, 1)$. *The nominal level* α SPUR2 *test of* $H_0 : \theta_0 \in \Theta_I^{MR}(F)$ *and nominal level* $1 - \alpha$ SPUR2 CS *for* $\theta \in \Theta_I^{MR}(F)$ *satisfy*[8]

[8] As is standard in the literature, the asymptotics for the bootstrap are given for the case where the number of bootstrap repetitions $B = \infty$ in Theorem 5.1 and other results below. If one considered finite B, then the asymptotic results would hold provided $B \rightarrow \infty$ as $n \rightarrow \infty$.

(a) $\limsup_{n\to\infty} \sup_{F\in\mathcal{P}:\theta_0\in\Theta_I^{MR}(F)} P_F(\phi_{n,SPUR2}(\theta_0) = 1) \leq \alpha$ *and*

(b) $\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} \inf_{\theta\in\Theta_I^{MR}(F)} P_F(\theta \in CS_{n,SPUR2}) \geq 1-\alpha$, *respectively.*

Comment. The SPUR1 test and CS have correct asymptotic level under the same conditions, see online Appendix B.

Now, we state a condition under which the SPUR2 test equals the standard level α_2 GMS test wp$\to$1. We say that a value θ is slack under F if all of the moment inequalities hold strictly at θ, i.e., $\min_{j\leq k} E_F m_j(W,\theta) > 0$. If the identified sets $\{\Theta_I(F_n)\}_{n\geq 1}$ are nonempty and contain slack points for which the magnitude of slackness is of order greater than $n^{-1/2}$ (i.e., there exists a sequence $\{\theta_n^I \in \Theta_I(F_n)\}_{n\geq 1}$ for which $n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta_n^I) \to \infty \ \forall j \leq k$), then the SPUR2 test equals the standard GMS test wp$\to$1.[9]

Next, if the model is identifiably misspecified and exhibits "large-local" or "global" model misspecification in the sense that $n^{1/2}r_{F_n}^{\inf} \to \infty$, then $\widehat{r}_{n,UP}(\alpha_1) > 0$ wp$\to$1[10] and $\phi_{n,SPUR2}(\theta_0) = \min\{\phi_{n,SPUR1}(\theta_0,\alpha_2), \phi_{n,GMS}(\theta_0,\alpha_2)\}$, which typically equals $\phi_{n,SPUR1}(\theta_0,\alpha_2)$, wp$\to$1.

4.7 Tuning Parameters, Implementation, and Computation of SPUR Tests and CI's

4.7.1 Tuning Parameters

The tuning parameters for the SPUR tests and CI's are: B, ι, κ_n, τ_n, and α_1. We recommend the choices $B = 1000$, $\iota = 10^{-6}$, and $\kappa_n = \tau_n = (\ln n)^{1/2}$. For α_1, we recommend the standard Bonferroni choice of $\alpha/10$, where α is the level of the test or CS. Thus, for $\alpha = .05$, this yields $\alpha_1 = .005$.

The quantity B is the number of bootstrap repetitions. We recommend $B = 1000$ based on a combination of the recommendations and results in Efron and Tibshirani (1986, Sec. 9) and Andrews and Buchinsky (2000, 2001). This choice works well in a wide variety of bootstrapping scenarios.

The impact on the critical value $\widehat{c}_{n,\Delta U}(1-\alpha)$ of adding the constant $\iota > 0$ to the bootstrap quantile (as defined following (4.15)) is transparent. It is simply the magnitude of ι. Hence, any choice of ι that is very small will have essentially no impact on the critical value. For specificity, we recommend $\iota = 10^{-6\cdot}$. The impact of defining the bootstrap standard deviations $sd_{anjB}^*(\theta)$ for

[9] The stated condition is equivalent to Assumption SLK in Section 5.5 below. A set of sufficient conditions for Assumption SLK is given in Lemma 24.2 in online Appendix C. The stated claim holds because $\widehat{r}_{n,UP}(\alpha) := \max\{\widehat{\Delta}_{n,U}^{\inf}(\alpha), 0\}$ and Theorem 5.2(a) in Section 5.5 imply that, under Assumption SLK, $\widehat{r}_{n,UP}(\alpha_1) = 0$ wp$\to$1 and $\phi_{n,SPUR2}(\theta_0) = \phi_{n,GMS}(\theta_0,\alpha_2)$ wp$\to$1 by the definition of the SPUR2 test in (4.10).

[10] This holds by Theorem 30.1 in Section 30 in online Appendix C.

$a = 1, 2, 3$ to have a lower bound of $\iota > 0$ is very small because standard deviations are necessarily nonnegative and ι is very small.

The tuning parameter κ_n is analogous to the tuning parameter used in GMS methods. Power under the null and alternative hypotheses is decreasing in κ_n. We recommend the same value $\kappa_n = (\ln n)^{1/2}$ as is typically employed with GMS methods.

In Section 8.1.3 in online Appendix A, we simulate the effects of changes in the tuning parameters on the rejection probabilities of the SPUR2 test under the null and alternative hypotheses in a lower/upper bound model. Section 10.1 in online Appendix A reports the sensitivity of the SPUR2 CI lower and upper bounds in the empirical illustration in Section 6 below. For the tuning parameters τ_n, α_1, ι, and B, these results show that there is little sensitivity to halving or doubling the recommended tuning parameters. For the κ_n parameter, there is sensitivity in some model scenarios. The recommended value of κ_n is designed to achieve high power subject to the null rejection probabilities being less than or equal to α.

4.7.2 Implementation

See Algorithm 1 for pseudo-code for computing the SPUR1 test.

4.7.3 Computation

Computation of the SPUR1 test requires computing $A^*_{n,b}$, and then, $S^*_{n,b}(\theta)$ for a given null value θ. Given $A^*_{n,b}$, $S^*_{n,b}(\theta)$ has a closed form expression and is very quick to compute.

To compute $A^*_{n,b}$, one can use standard nonlinear constrained optimization software (e.g., `slsqp` in the R library `nloptr` or `fmincon` in Matlab).[11] The objective function and constraints are

$$\begin{aligned} A^*_{n,b}(\theta) &:= \min_{j_1 \in \widehat{J}_{nB}(\theta)} \max_{j \le k} (\widehat{\chi}^*_{nj,b}(\theta) + 1(j \neq j_1)\widehat{b}_{nj}(\theta) + 1(j = j_1)\varphi(\xi^A_{nj}(\theta))) \text{ and} \\ -\widehat{m}_{nj}(\theta) &\le \tau_n / n^{1/2} + \widehat{r}^{\inf}_n \text{ for } j \le k. \end{aligned} \tag{4.26}$$

The constraints correspond the requirement that $\theta \in \widehat{\Theta}_n$.[12] The advantage of employing k constraints, rather than the single constraint in $\widehat{\Theta}_n$, is that it makes the constraints differentiable in many contexts, including the empirical illustration in Section 6 and the simulations in online Appendix A. In the case where the Jacobian of $\widehat{m}_n(\theta)$ can be calculated analytically, providing this to the optimization algorithm typically results in faster and more stable calculations. In some cases, as in the empirical illustration and the lower bound/upper bound model in the simulations,

[11] We use the former. The `nlopt` library also is callable from C, C++, Fortran, Matlab, Python, and Julia.

[12] These constraints are equivalent to the constraints that define $\widehat{\Theta}_n$ in (4.20) because, for $b, c \ge 0$, $[a + b]_- \le c$ if and only if $[a]_- - b \le c$.

Algorithm 1 SPUR1 test for $H_0 : \theta_0 \in \Theta_I^{MR}$

Inputs: $\{W_i\}_{i \leq n}$, $m(\cdot)$, $S(\cdot)$, Θ, α, B, κ_n, τ_n, ι
▷ recommend $B = 1{,}000$, $\kappa_n = \tau_n = (\ln n)^{1/2}$, $\iota = 10^{-6}$

Output: SPUR1 test, $\phi_{SPUR1,n}(\theta_0)$

function $\widehat{r}_{nj}(\theta)$ ▷ for $j = 1, \ldots, k$
  $\overline{m}_{nj}(\theta) \leftarrow n^{-1} \sum_{i=1}^n m_j(W_i, \theta)$
  $\widehat{\sigma}_{nj}^2(\theta) \leftarrow n^{-1} \sum_{i=1}^n (m_j(W_i, \theta) - \overline{m}_{nj}(\theta))^2$
  $\widehat{m}_{nj}(\theta) \leftarrow \frac{\overline{m}_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}$
  return $[\widehat{m}_{nj}(\theta)]_-$ ▷ $[x]_- = \max\{-x, 0\}$
end function

$\widehat{r}_n(\theta) \leftarrow \max_{j \leq k} \widehat{r}_{nj}(\theta)$
$\widehat{r}_n^{\inf} \leftarrow \inf_{\theta \in \Theta} \widehat{r}_n(\theta)$

function $A_n^*(\theta, \{W_{ib}^*\}_{i \leq n})$ ▷ $\{W_{ib}^*\}_{i \leq n}$: generic bootstrap sample
  for $j = 1, \ldots, k$ **do**
    $\overline{m}_{njb}^*(\theta) \leftarrow n^{-1} \sum_{i=1}^n m_j(W_{ib}^*, \theta)$,
    $\widehat{\sigma}_{njb}^{*2}(\theta) \leftarrow n^{-1} \sum_{i=1}^n (m_j(W_{ib}^*, \theta) - \overline{m}_{njb}^*(\theta))^2$
    $\widehat{\nu}_{njb}^*(\theta) \leftarrow n^{1/2} \left(\frac{\overline{m}_{njb}^*(\theta)}{\widehat{\sigma}_{njb}^*(\theta)} - \widehat{m}_{nj}(\theta) \right)$
    if $\widehat{\nu}_{njb}^*(\theta) \geq 0$ **then**
      $\widehat{\chi}_{nj,b}^*(\theta) \leftarrow [\widehat{\nu}_{njb}^*(\theta) + n^{1/2}\widehat{m}_{nj}(\theta) - sd_{2njB}^*(\theta)\kappa_n]_- - [n^{1/2}\widehat{m}_{nj}(\theta) - sd_{2njB}^*(\theta)\kappa_n]_-$
    else if $\widehat{\nu}_{njb}^*(\theta) < 0$ **then** ▷ see (4.21) for definition of $sd_{2njB}^*(\theta)$
      $\widehat{\chi}_{nj,b}^*(\theta) \leftarrow [\widehat{\nu}_{njb}^*(\theta) + n^{1/2}\widehat{m}_{nj}(\theta) + sd_{2njB}^*(\theta)\kappa_n]_- - [n^{1/2}\widehat{m}_{nj}(\theta) + sd_{2njB}^*(\theta)\kappa_n]_-$
    end if
    $\xi_{nj}^A(\theta) \leftarrow (sd_{3njB}^*(\theta)\kappa_n)^{-1} n^{1/2} \left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf} \right)$
    $\widehat{b}_{nj}(\theta) \leftarrow n^{1/2} \left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}_n^{\inf} \right) - sd_{3njB}^*(\theta)\kappa_n$
  end for ▷ see (4.22) for definition of $sd_{3njB}^*(\theta)$
  $\widehat{J}_{nB}(\theta) \leftarrow \{j \in \{1, ..., k\} : \widehat{r}_{nj}(\theta) \geq \widehat{r}_n(\theta) - sd_{3njB}^*(\theta) n^{-1/2} \kappa_n\}$
  return $\min_{j_1 \in \widehat{J}_{nB}(\theta)} \max_{j \leq k} \left(\widehat{\chi}_{nj,b}^*(\theta) + 1(j \neq j_1)\widehat{b}_{nj}(\theta) + 1(j = j_1)\varphi(\xi_{nj}^A(\theta)) \right)$
end function ▷ $\varphi(\xi) = \infty 1(\xi > 1)$

function Spur1Test(θ_0, $\{W_i\}_{i \leq n}$, B, α)
  $\widehat{\Theta}_n \leftarrow \{\theta \in \Theta : \max_{j \leq k}[\widehat{m}_{nj}(\theta) + \widehat{r}_n^{\inf}]_- \leq \tau_n / n^{1/2}\}$
  for $b = 1, ..., B$ **do**
    Draw bootstrap sample $\{W_{ib}^*\}_{i \leq n}$
    for $j = 1, \ldots, J$ **do**
      $T_{nj}^*(\theta_0)[b] \leftarrow \widehat{\nu}_{njb}^*(\theta_0) + \varphi((sd_{1njB}^*(\theta_0)\kappa_n)^{-1} n^{1/2} \left(\widehat{m}_{nj}(\theta_0) + \widehat{r}_n(\theta_0) \right))$
    end for ▷ see (4.19) for definition of $sd_{1njB}^*(\theta_0)$
    $A_n^*[b] \leftarrow \inf_{\theta \in \widehat{\Theta}_n} A_n^*(\theta, \{W_{ib}^*\}_{i \leq n}))$
    $S_n^*(\theta_0)[b] \leftarrow S\left(T_n^*(\theta_0)[b] + A_n^*[b] 1_k, \widehat{\Omega}_n(\theta_0) \right)$
  end for
  $\widehat{c}_n(\theta, 1 - \alpha) \leftarrow \text{quantile}(\{S_n^*(\theta_0)\}, 1 - \alpha) + \iota$ ▷ critical value
  return $\phi_{SPUR1,n}(\theta_0) \leftarrow 1(S(n^{1/2} \left(\widehat{m}_n(\theta_0) + \widehat{r}_n^{\inf} 1_k \right), \widehat{\Omega}_n(\theta_0)) > \widehat{c}_n(\theta, 1 - \alpha))$
end function

$m(W, \theta)$ is additive in the sense that it can be written as $m(W, \theta) = g(W) + f(\theta)$. If so, the Jacobian of $\widehat{m}_n(\theta)$ is the same as the Jacobian of $f(\theta)$, which in many cases is not difficult to calculate and does not depend on b, so it only needs to be computed once, not B times.[13]

For the initial value for the optimization problem, we recommend using θ^{init}, which is defined in Section 5.6.3 below, rather than here, because it utilizes notation introduced in Section 5). It works well in the empirical illustration and in the simulations.[14]

Parallel computation of $\{A^*_{n,b}\}_{b \leq B}$ is straightforward and reduces the computation time considerably. The bootstrap statistics $\{A^*_{n,b}\}_{b \leq B}$ do not depend on θ, so they only need to be computed once when one is computing a SPUR1 CS by test inversion.

As an example, we discuss computation in the empirical illustration in Section 6. The model is an airline binary entry model with 48 inequalities and 9 parameters, and the sample size is 7,882. All computations for the empirical illustration are performed on a computer with an Intel Xeon Gold 6240 processor, which has 18 cores with double threads.[15] Hence, when we say "run in parallel" here and below, it means that the computation was run in parallel on the 36 threads. We use the `slsqp` command in the R package `nloptr` for all of the optimization problems. The analytical Jacobian of $\widehat{m}_n(\theta)$ is passed to the algorithm. Note that the computation of non-convex optimization problems is hard in general. The computation times in the empirical illustration may not be indicative of computation times in other models with other data sets.

In the empirical illustration, computation of $\{A^*_{n,b}\}_{b \leq B}$ is done in parallel over the 36 threads on the Intel 6240 processor and it takes approximately 17.5 minutes for $B = 1{,}000$.

In practice, the empirical researcher often is interested in a CI for $p'\theta$ for some $p \in R^{d_\theta}$. For example, $p = (1, 0, \ldots, 0)'$ corresponds to the case where the researcher is interested in the first element of θ. As is standard in practice, one can report the projection CI of $p'\theta$ in this case.[16] To calculate the projection CI for the SPUR1 test, one solves two nonlinear constrained optimization problems for the lower and upper bounds. For example, for the upper bound of the SPUR1 projection CI, one solves

$$\max_{\theta \in \Theta} p'\theta \quad \text{subject to} \quad S_n(\theta) \leq \widehat{c}_n(\theta, 1 - \alpha_2). \tag{4.27}$$

[13]If the moment functions are not additive, then computational speed can be increased by taking the number of bootstrap repetitions used to compute $sd^*_{2njB}(\theta)$ and $sd^*_{3njB}(\theta)$ to be smaller than B, such as 250 (but using $B = 1000$ everywhere else). The choice of 250 is based on the recommendations and results in Efron and Tibshirani (1986, Sec. 9) and Andrews and Buchinsky (2000, 2001) for bootstrap standard error estimators, which require fewer bootstrap repetitions than bootstrap quantiles to be accurate.

[14]The computations were checked by using additional initial values that satisfy the constraints, and the results obtained were essentially identical.

[15]The computation times reported below for the empirical illustration are for the case of an unknown correlation parameter ρ. Section 6 also reports some results when ρ is taken to be a fixed value.

[16]By definition, the SPUR1 projection CI for the jth component of θ is $[\inf_{\theta \in CS_{n,SPUR1}} \theta_j, \sup_{\theta \in CS_{n,SPUR1}} \theta_j]$.

As above, we use standard software (the `slsqp` command from the R package `nloptr` or `fmincon` in Matlab) to solve this optimization problem.[17] We use a number of different initial values, which are described in detail in Section 10.4 in online Appendix A.

In the empirical illustration, the 9 SPUR1 projection CI's take about 50 minutes to compute given $\{A^*_{n,b} : b \leq B\}$. Hence, the total time for computing the 9 SPUR1 projection CI's is approximately $17.5 + 50 = 67.5$ minutes.

Next, we discuss computation of SPUR2 CS's. The statistic $\widehat{r}_{n,UP}(\alpha_1)$ in (4.9) is the same for all tests $\{\phi_{n,SPUR2}(\theta) : \theta \in \Theta\}$ that yield the SPUR2 CS. Hence, it only needs to be computed once. When $\widehat{r}_{n,UP}(\alpha_1) > 0$, the level $1-\alpha$ SPUR2 CS can be written as the union of the level $1-\alpha_2$ SPUR1 and GMS CS's, and we find that it is quickest to compute these CS's separately. The same is true for SPUR2 projection CI's. The GMS projection CI's are computed as in (4.27) with $S_{n,Std}(\theta)$ and $\widehat{c}_{n,GMS}(\theta, 1-\alpha_2)$ in place of $S_n(\theta)$ and $\widehat{c}_n(\theta, 1-\alpha_2)$, respectively, using the same software. Again, we use a number of different initial values as described in Section 10.4 in online Appendix A. When $\widehat{r}_{n,UP}(\alpha_1) = 0$, the level $1-\alpha$ SPUR2 CS is the level α_2 GMS CS, and the level $1-\alpha_2$ SPUR2 projection CI's are the level α_2 GMS projection CI's.

Computing $\widehat{r}_{n,UP}(\alpha_1) = \max\{\widehat{\Delta}^{\inf}_{n,U}(\alpha_1), 0\}$ is the same as computing $\widehat{\Delta}^{\inf}_{n,U}(\alpha_1)$, which is discussed in Section 5.6.3 below. In the empirical illustration, it takes about 22 minutes. As discussed above, in the empirical illustration, computing the 9 SPUR1 projection CI's takes about 67.5 minutes. In addition, computation of the 9 GMS projection CI's takes about 21 minutes.[18] Hence, the total time for computing the 9 SPUR2 projection CI's is $22 + 67.5 + 21 = 110.5$ minutes using a single computer with Intel 6240 processor with 36 threads.

5 Misspecification Diagnostics

This section introduces the model misspecification diagnostics.

5.1 Misspecification Index, Confidence Intervals, and Tests

Define the misspecification index $\Delta^{\inf}_F$ by

$$\Delta^{\inf}_F := \inf_{\theta \in \Theta} \max_{j \leq k} \Delta_{Fj}(\theta), \text{ where } \Delta_{Fj}(\theta) := -E_F \widetilde{m}_j(W, \theta) \text{ for } j \leq k. \tag{5.1}$$

[17] Alternatively, one could use the E-A-M algorithm of Kaido, Molinari, and Stoye (2019), which may reduce the projection computation time.

[18] Some of the initial values used for calculating the GMS projection CI's overlap with those used for the SPUR1 projection CI's. For such values, we include their computation time in that for the SPUR1 projection CI's.

When positive, $\Delta_{Fj}(\theta)$ is the magnitude of the violation of moment j when evaluated at θ. When negative, $-\Delta_{Fj}(\theta)$ is the slackness of moment j when evaluated at θ. When the identified set $\Theta_I(F)$ is empty, $\Delta_F^{\inf}$ is positive and is increasing in the amount of misspecification, as measured by the minimum over Θ of the maximum inequality violation over the k moments. When $\Theta_I(F)$ is nonempty, $\Delta_F^{\inf}$ is nonpositive and $-\Delta_F^{\inf}$ is increasing in the size of $\Theta_I(F)$, as measured by the maximum over $\theta \in \Theta$ of the minimum slackness of the k moments. In short, $\Delta_F^{\inf}$ is a misspecification index (MI) for positive values and a measure of the size of the identified set when negative. Note that when $\Delta_F^{\inf} \geq 0$, $r_F^{\inf} = \Delta_F^{\inf}$, and when $\Delta_F^{\inf} < 0$, $r_F^{\inf} = 0$. Thus, $r_F^{\inf} = \max\{\Delta_F^{\inf}, 0\}$ and $\Delta_F^{\inf}$ is more informative than $r_F^{\inf}$.

A consistent estimator of $\Delta_F^{\inf}$ is

$$\widehat{\Delta}_n^{\inf} := \inf_{\theta \in \Theta} \max_{j \leq k} \widehat{\Delta}_{nj}(\theta), \text{ where } \widehat{\Delta}_{nj}(\theta) := -\widehat{m}_{nj}(\theta). \tag{5.2}$$

For the estimator $\widehat{\Delta}_n^{\inf}$ to be informative, one needs a measure of its accuracy. For this, we provide nominal level $1-\alpha$ upper-bound, lower-bound, and two-sided CI's for the MI $\Delta_F^{\inf}$ based on $\widehat{\Delta}_n^{\inf}$:

$$\begin{aligned} CI_{n,\Delta U}(\alpha) &:= \left(-\infty, \widehat{\Delta}_{n,U}^{\inf}(\alpha)\right] \text{ for } \widehat{\Delta}_{n,U}^{\inf}(\alpha) := \widehat{\Delta}_n^{\inf} + \frac{\widehat{c}_{n,\Delta U}(1-\alpha)}{n^{1/2}}, \\ CI_{n,\Delta L}(\alpha) &:= \left[\widehat{\Delta}_{n,L}^{\inf}(\alpha),\ \infty\right) \text{ for } \widehat{\Delta}_{n,L}^{\inf}(\alpha) := \widehat{\Delta}_n^{\inf} - \frac{\widehat{c}_{n,\Delta L}(1-\alpha)}{n^{1/2}}, \text{ and} \\ CI_{n,\Delta}(\alpha) &:= [\widehat{\Delta}_{n,L}^{\inf}(\alpha/2),\ \widehat{\Delta}_{n,U}^{\inf}(\alpha/2)], \end{aligned} \tag{5.3}$$

respectively, where $\widehat{c}_{n,\Delta U}(1-\alpha)$ and $\widehat{c}_{n,\Delta L}(1-\alpha)$ are bootstrap critical values defined in Section 5.3 below. Implementation and computation of these CI's is discussed in Section 5.6 below.

If $\widehat{\Delta}_{n,U}^{\inf}(\alpha) < 0$, the upper-bound CI, $CI_{n,\Delta U}(\alpha)$, provides evidence that the identified set is not empty. Consider the null hypothesis H_0 that the identified set is empty (and hence, the model is identifiably misspecified) and the alternative hypothesis H_1 that the identified set is not empty (and the model is not identifiably misspecified). These hypotheses are:

$$H_0 : \Delta_F^{\inf} > 0 \text{ versus } H_1 : \Delta_F^{\inf} \leq 0. \tag{5.4}$$

The test that rejects the null when

$$\widehat{\Delta}_{n,U}^{\inf}(\alpha) < 0, \text{ or equivalently, when } \widehat{\Delta}_n^{\inf} < -\widehat{c}_{n,\Delta U}(1-\alpha)/n^{1/2}, \tag{5.5}$$

is a nominal level $1-\alpha$ test of H_0 versus H_1.

The hypotheses H_0 and H_1 are the reverse of the usual model specification hypotheses H_{00} :

$\Delta_F^{\inf} \leq 0$ versus $H_{11} : \Delta_F^{\inf} > 0$, in which the null hypothesis is that the identified set is not empty. A drawback of a test of H_{00} versus H_{11} is that failure to reject the null H_{00} may be due to low power, and hence, does not provide evidence that the identified set is nonempty. In contrast, for a test of H_0 versus H_1, falsely concluding that the identified set is nonempty, i.e., falsely rejecting H_0, is controlled and occurs with probability α or less.

The above test of H_0 versus H_1 has power against alternatives for which $\Delta_F^{\inf} < 0$, which corresponds to the case where the correctly-specified model has slack points. The test has no power (i.e., power is α or less) against an alternative for which $\Delta_F^{\inf} = 0$, because such an alternative is on the boundary of the null hypothesis. The "larger" is the identified set, in the sense of having a more negative value of $\Delta_F^{\inf}$, the higher is the power of the test.

The test that rejects H_{00} if $\widehat{\Delta}_{n,L}^{\inf}(\alpha) > 0$ is quite similar to the BCS resampling specification test based on the "max" test function. However, the lower-bound CI, $CI_{n,\Delta L}(\alpha)$, is more informative than this test because it indicates how large $\Delta_F^{\inf}$ is.[19]

Section 5.5 below shows that when the sequence of identified sets $\{\Theta_I(F_n)\}_{n\geq 1}$ under distributions $\{F_n\}_{n\geq 1}$ contains slack points with slackness greater than $n^{-1/2}$ (i.e., $n^{1/2}E_{F_n}\widetilde{m}_j(W, \theta_n^I) \to \infty$ $\forall j \leq k$ for some sequence $\{\theta_n^I \in \Theta_I(F_n)\}_{n\geq 1}$), then $CI_{n,\Delta U}(\alpha) \subset (-\infty, 0)$ wp$\to$1. Thus, in such cases, one can detect that the identified set is not empty wp$\to$1. Section 5.5 also shows that if the model exhibits "large-local" or "global" model misspecification (i.e., $n^{1/2}\Delta_{F_n}^{\inf} \to \infty$), then the lower-bound CI $CI_{n,\Delta L}(\alpha) \subset (0, \infty)$ wp$\to$1. Hence, in such cases, identifiable misspecification can be detected wp$\to$1.

We recommend that an empirical researcher report $\widehat{\Delta}_n^{\inf}$ and $CI_{n,\Delta}(\alpha)$ to provide information on model specification. In addition, these statistics can be useful when determining which moment inequalities to employ, analogously to the use of the J test of misspecification in over-identified GMM models to help determine which moment equalities to employ. The statistic $\widehat{\Delta}_n^{\inf}$ has the helpful feature that it indicates which moment inequality yields the largest violation. One can see the effect of a particular inequality j or a set of inequalities by dropping these inequalities and seeing how $\widehat{\Delta}_n^{\inf}$ and $CI_{n,\Delta}(\alpha)$ change. For examples, to see their impact on the misspecification index, one can drop inequalities that (i) rely on more assumptions (on the underlying economic model) than other inequalities, or (ii) are key in terms of the underlying economic model, or (iii) are believed to be correctly specified, or (iv) are more likely than other inequalities to be misspecified. Of course, using $\widehat{\Delta}_n^{\inf}$ and $CI_{n,\Delta}(\alpha)$ in these ways would raise post-model selection issues, as occurs in GMM models when one selects the moments using the data.

[19] The CI $CI_{n,\Delta L}(\alpha)$ cannot be constructed from the BCS results, because the latter do not cover null hypothesis values with $\Delta_F^{\inf} > 0$.

5.2 Intuition Behind the Critical Value $\widehat{\mathbf{c}}_{\mathbf{n},\boldsymbol{\Delta}\mathbf{U}}(\mathbf{1}-\alpha)$

The definition of the MI critical value $\widehat{c}_{n,\Delta U}(1-\alpha)$ is complicated. In consequence, prior to stating its definition, we provide intuition in this section for the form that it takes.

The MI critical value $\widehat{c}_{n,\Delta U}(1-\alpha)$ in (5.3) is a bootstrap critical value. A naive definition of the bootstrap statistic used to construct $\widehat{c}_{n,\Delta U}(1-\alpha)$ would exactly mimic the form of the recentered and rescaled estimator $\widehat{\Delta}_n^{\inf}$ of $\Delta_{F_n}^{\inf}$, denoted by $A_{n,\Delta} := n^{1/2}(\widehat{\Delta}_n^{\inf} - \Delta_{F_n}^{\inf})$. That is, it would use a bootstrap sample in place of the original sample $\{W_i\}_{i\leq n}$ and use $\widehat{\Delta}_n^{\inf}$ in place of $\Delta_{F_n}^{\inf}$ in the expression for $A_{n,\Delta}$. However, such a definition would not yield a test or CS with the correct asymptotic level because the statistic $\widehat{\Delta}_n^{\inf}$ involves the $\inf_{\theta\in\Theta}$ and $\max_{j\leq k}$ terms. The latter yield a complicated asymptotic distribution of $A_{n,\Delta}$ that involves terms that are not consistently estimable and cause the naive bootstrap to fail.

Instead of the naive bootstrap, we employ a bootstrap version of $A_{n,\Delta}$ that takes account of the nonregular form of the asymptotic distribution of $A_{n,\Delta}$. To motivate the form of the bootstrap version of $A_{n,\Delta}$, we rewrite $A_{n,\Delta}$ as

$$A_{n,\Delta} := n^{1/2}(\widehat{\Delta}_n^{\inf} - \Delta_{F_n}^{\inf}) = \inf_{\theta\in\Theta} \max_{j\leq k}(\underbrace{n^{1/2}(\widehat{\Delta}_{nj}(\theta) - \Delta_{F_nj}(\theta))}_{:=-\widehat{\nu}_{nj}(\theta)} + \underbrace{n^{1/2}(\Delta_{F_nj}(\theta) - \Delta_{F_n}^{\inf})}_{:=e_{nj}(\theta)}), \tag{5.6}$$

where $\widehat{\nu}_{nj}(\theta)$ is a properly centered stochastic process indexed by θ and $e_{nj}(\theta)$ is a nonrandom function that is not consistently estimable.

We employ a bootstrap version of $A_{n,\Delta}$ that uses a bootstrap stochastic process to approximate $\widehat{\nu}_{nj}(\theta)$ and replaces $e_{nj}(\theta)$ by an estimated lower bound that is designed to impose the condition $\max_{j\leq k} e_{nj}(\theta) \geq 0$. Using a lower bound ensures that the critical value $\widehat{c}_{n,\Delta U}(1-\alpha)$ is sufficiently large to yield an asymptotic coverage probability that is at least $1-\alpha$.[20]

The bootstrap statistic $A^*_{n,\Delta U,b}$ that we employ replaces $\inf_{\theta\in\Theta}$ by $\inf_{\theta\in\widehat{\Theta}_{\min,n}}$, where $\widehat{\Theta}_{\min,n}$ is a consistent estimator of the set of values $\theta\in\Theta$ that minimize $\max_{j\leq k}\Delta_{Fj}(\theta)$ over $\theta\in\Theta$. This replacement is valid because, roughly speaking, the set $\widehat{\Theta}_{\min,n}$ includes all parameter values that are relevant for the asymptotic distribution of $A_{n,\Delta}$. The set $\widehat{\Theta}_{\min,n}$ could be replaced by Θ without affecting the asymptotic properties of the CI $CI_{n,\Delta U}(\alpha)$, but $\widehat{\Theta}_{\min,n}$ eases computation because it substantially reduces the size of the set over which the infimum is taken, which reduces the number of initial values that needs to be considered.

[20] A lower bound is required for the following reason. We have $\Delta_{F_n}^{\inf} \notin CI_{n,\Delta U}(\alpha)$ iff $A_{n,\Delta} := n^{1/2}(\widehat{\Delta}_n^{\inf} - \Delta_{F_n}^{\inf}) \leq -\widehat{c}_{n,\Delta U}(1-\alpha)$ by (5.3). In consequence, a larger value of $\widehat{c}_{n,\Delta U}(1-\alpha) > 0$ leads to a smaller non-coverage event. By the definition that follows, $\widehat{c}_{n,\Delta U}(1-\alpha)$ is the $1-\alpha$ sample quantile of $\{-A^*_{n,\Delta U,b}\}_{b\leq B}$ and smaller values of $\{A^*_{n,\Delta U,b}\}_{b\leq B}$ lead to a larger value of $\widehat{c}_{n,\Delta U}(1-\alpha)$. Thus, $e_{nj}(\theta)$ needs to be lower bounded in $A^*_{n,\Delta U,b}$ to ensure the asymptotic coverage probability is at least $1-\alpha$.

The centered stochastic process $\widehat{\nu}_{nj}(\theta)$ is estimated by the bootstrap stochastic process $\widehat{\nu}^*_{njb}(\theta)$ defined in (4.17).

For $e_{nj}(\theta)$, we consider an asymptotic lower bound $\widehat{e}_{nj}(\theta)$ that is obtained by shifting the sample analogue of $e_{nj}(\theta)$ toward $-\infty$ (defined in (5.8) below). However, this lower bound does not incorporate the fact that $\max_{j\leq k} e_{nj}(\theta) \geq 0$. To do so, we employ $\widehat{J}_{neB}(\theta)$, which is an estimator of those value(s) of j for which $e_{nj}(\theta) \geq 0$. For $j \in \widehat{J}_{neB}(\theta)$, we employ a better (GMS-type) lower bound on $e_{nj}(\theta)$ given by $\varphi(\xi^e_{nj}(\theta))$ (defined in (5.12) below). As with the SPUR1 critical value $\widehat{c}_n(\theta, 1-\alpha)$, incorporating the better lower bound is important. Otherwise, the critical value would be divergent asymptotically.

5.3 Definitions of the Critical Values $\widehat{\mathbf{c}}_{\mathbf{n},\boldsymbol{\Delta}\mathbf{U}}(\mathbf{1}-\alpha)$ and $\widehat{\mathbf{c}}_{\mathbf{n},\boldsymbol{\Delta}\mathbf{L}}(\mathbf{1}-\alpha)$

The critical value $\widehat{c}_{n,\Delta U}(1-\alpha)$ in (5.3) is based on B bootstrap statistics $\{A^*_{n,\Delta U,b}\}_{b\leq B}$, where $A^*_{n,\Delta U,b}$ is defined following the intuition outlined in Section 5.2. Let $\{W^*_{ib}\}_{i\leq n}$ for $b = 1, ..., B$ denote the bootstrap samples defined in Section 4.4.

The bootstrap analogue of $\widehat{\nu}_{nj}(\theta)$ is $\widehat{\nu}^*_{njb}(\theta)$, defined in (4.17).

Let

$$\widehat{\Theta}_{\min,n} := \{\theta \in \Theta : \max_{j\leq k} \widehat{\Delta}_{nj}(\theta) \leq \widehat{\Delta}^{\inf}_n + \tau_n/n^{1/2}\}, \tag{5.7}$$

where τ_n is a constant for which $\tau_n \to \infty$. See Section 5.6.1 below for the recommended choice of τ_n and other tuning parameters.

Define the shifted lower bound on $e_{nj}(\theta)$ by $\widehat{e}_{nj}(\theta)$:

$$\widehat{e}_{nj}(\theta) := n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}^{\inf}_n\right) - \widehat{sd}_{njB}(\theta)\kappa_n, \tag{5.8}$$

where κ_n is a tuning parameter that satisfies $\kappa_n \to \infty$ and $\widehat{sd}_{njB}(\theta)$ is a simulated standard deviation estimator that is used to obtain appropriate scaling under model misspecification, as discussed in Section 5.4 below. The standard deviation estimator $\widehat{sd}_{njB}(\theta)$ is defined as follows. Let $Z_s \sim iid\ N(0_{2k}, I_{2k})$ for $s = 1, ..., B$. Let c_j denote the jth elementary k-vector. By definition,

$$\begin{aligned}
&\widehat{sd}_{njB}(\theta) := \max\left\{V^{1/2}_{njB}(\theta), \iota\right\}, \text{ where } V_{njB}(\theta) := B^{-1}\sum_{s=1}^{B}(Q_{njs}(\theta) - \overline{Q}_{njB}(\theta))^2,\\
&\overline{Q}_{njB}(\theta) := B^{-1}\sum_{s=1}^{B} Q_{njs}(\theta),\ Q_{njs}(\theta) := \widehat{G}^{m\sigma}_{njs}(\theta) - \max_{j_1\leq k}\widehat{G}^{m\sigma}_{nj_1s}(\theta),\\
&\widehat{G}^{m\sigma}_{njs}(\theta) := (c'_j, -(1/2)\widehat{m}_{nj}(\theta)c'_j)\widehat{\Omega}^{1/2}_{n+}(\theta)Z_s,
\end{aligned} \tag{5.9}$$

and ι is a very small positive constant. In (5.9), $\widehat{\Omega}_{n+}(\theta)$ is a consistent estimator of the variance matrix $\Omega_{F+}(\theta)$ of the population-standard-deviation-normalized moment functions and recentered second central moment functions. The latter are defined by

$$\begin{aligned} \widetilde{m}_j(W,\theta) &:= \frac{m_j(W,\theta)}{\sigma_{Fj}(\theta)},\ \widetilde{m}_j^\sigma(W,\theta) := (\widetilde{m}_j(W,\theta) - E_F\widetilde{m}_j(W,\theta))^2 - 1 \text{ for } j \leq k, \text{ and} \\ \Omega_{F+}(\theta) &:= Var_F(\widetilde{m}(W,\theta)', \widetilde{m}^\sigma(W,\theta)') \in R^{2k\times 2k}, \end{aligned} \tag{5.10}$$

where $\widetilde{m}(W,\theta)$ and $\widetilde{m}^\sigma(W,\theta)$ have jth elements $\widetilde{m}_j(W,\theta)$ and $\widetilde{m}_j^\sigma(W,\theta)$, respectively.

The estimator $\widehat{\Omega}_{n+}(\theta)$ of $\Omega_{F+}(\theta)$ is defined by

$$\begin{aligned} \widehat{\Omega}_{n+}(\theta) &:= n^{-1}\sum_{i=1}^n \left(\begin{array}{c} \widehat{m}_n(W_i,\theta) \\ \widehat{m}_n^\sigma(W_i,\theta) \end{array}\right)\left(\begin{array}{c} \widehat{m}_n(W_i,\theta) \\ \widehat{m}_n^\sigma(W_i,\theta) \end{array}\right)' \in R^{2k\times 2k}, \text{ where} \\ \widehat{m}_{nj}^\sigma(W,\theta) &:= \widehat{m}_{nj}^2(W,\theta) - 1, \end{aligned} \tag{5.11}$$

and $\widehat{m}_n(W,\theta)$ and $\widehat{m}_n^\sigma(W,\theta)$ have jth elements $\widehat{m}_{nj}(W,\theta)$ and $\widehat{m}_{nj}^\sigma(W,\theta)$, respectively. The upper left $k \times k$ block of $\widehat{\Omega}_{n+}(\theta)$ is the sample correlation matrix of the moments $\widehat{\Omega}_n(\theta)$, defined in (4.3).

Next, define the better lower bound on $e_{nj}(\theta)$ by

$$\varphi(\xi_{nj}^e(\theta)),\ \text{where } \xi_{nj}^e(\theta) := (\widehat{sd}_{njB}(\theta)\kappa_n)^{-1}n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}_n^{\inf}\right) \tag{5.12}$$

and $\varphi(\cdot)$ is defined in (4.19). This lower bound holds for $j \in \widehat{J}_{neB}(\theta)$, where

$$\widehat{J}_{neB}(\theta) := \{j \in \{1,...,k\} : \widehat{\Delta}_{nj}(\theta) \geq \max_{j_1 \leq k} \widehat{\Delta}_{nj_1}(\theta) - \widehat{sd}_{njB}(\theta)n^{-1/2}\kappa_n\}. \tag{5.13}$$

The "upper" EGMS bootstrap statistic $A_{n,\Delta U,b}^*$ is

$$A_{n,\Delta U,b}^* := \inf_{\theta\in\widehat{\Theta}_{\min,n}} \min_{j_1 \in \widehat{J}_{neB}(\theta)} \max_{j\leq k}\left(-\widehat{\nu}_{njb}^*(\theta) + 1(j \neq j_1)\widehat{e}_{nj}(\theta) + 1(j = j_1)\varphi(\xi_{nj}^e(\theta))\right). \tag{5.14}$$

As discussed in Section 5.2, $A_{n,\Delta U,b}^*$ is a bootstrap analogue of $A_{n,\Delta}$ with $\widehat{\nu}_{njb}^*(\theta)$ in place of $\widehat{\nu}_{nj}(\theta)$ and $\widehat{e}_{nj}(\theta)$ or $\varphi(\xi_{nj}^e(\theta))$ in place of $e_{nj}(\theta)$ depending on $\widehat{J}_{neB}(\theta)$.

The critical value $\widehat{c}_{n,\Delta U}(1-\alpha)$ is the $1-\alpha$ sample quantile of $\{-A_{n,\Delta U,b}^*\}_{b\leq B}$ plus a very small constant $\iota > 0$.

The critical value $\widehat{c}_{n,\Delta L}(1-\alpha)$ is defined analogously to the discard relaxation critical value in Bugni, Canay, and Shi (2017). It is based on B bootstrap statistics $\{A_{n,\Delta L,b}^*\}_{b\leq B}$, where $A_{n,\Delta L,b}^*$

is defined by

$$A^*_{n,\Delta L,b} := \inf_{\theta \in \widehat{\Theta}_{\min,L,n}} \max_{j \leq k} \left(-\widehat{\nu}^*_{njb}(\theta) - \varphi(-\xi^e_{nj}(\theta)) \right), \tag{5.15}$$

where $\widehat{\Theta}_{\min,L,n}$ is the same as $\widehat{\Theta}_{\min,n}$ in (5.7), but with $\tau_n = 0$. Then, $\widehat{c}_{n,\Delta L}(1-\alpha)$ is the $1-\alpha$ sample quantile of $\{A^*_{n,\Delta L,b}\}_{b \leq B}$ plus $\iota > 0$ for ι as above. See Bugni, Canay, and Shi (2017) for the intuition behind the definition of $A^*_{n,\Delta L,b}$.

5.4 Technical Discussion of the $\widehat{\mathbf{c}}_{\mathbf{n},\boldsymbol{\Delta}\mathbf{U}}(\mathbf{1}-\alpha)$ Critical Value

The nonrandom quantity $e_{nj}(\theta)$ is not consistently estimable for the same reason that $h_{nj}(\theta)$ is not, see Section 4.5.

If the lower bound on $e_{nj}(\theta)$ is arbitrarily small for all j, as can occur for $\widehat{e}_{nj}(\theta)$, then the critical value can be arbitrarily large. This does not reflect the null behavior of the test statistic because $\max_{j \leq k} e_{nj}(\theta)$ can be shown to be nonnegative. Hence, for power purposes it is important to provide a lower bound on $e_{nj}(\theta)$ that is nonnegative for the value(s) j that attain(s) $\max_{j \leq k} e_{nj}(\theta)$. This is accomplished by using $\varphi(\xi^e_{nj}(\theta))$, which is nonnegative, and $\widehat{J}_{neB}(\theta)$. The set $\widehat{J}_{neB}(\theta)$ is designed such that it includes the value(s) of j that attain $\max_{j \leq k} e_{nj}(\theta)$ wp$\rightarrow$1.

The quantity $\widehat{sd}_{njB}(\theta)$ is employed to provide proper scaling for the tuning parameter κ_n, analogously to $sd^*_{1njB}(\theta)$, $sd^*_{2njB}(\theta)$, and $sd^*_{3njB}(\theta)$ in the definition of the SPUR1 critical value, see Section 4.5.

5.5 Asymptotic Level of the Misspecification Index CI's

First, we show that the upper- and lower-bound CI's for $\Delta^{\inf}_F$ have correct asymptotic level. For ease of reading, we state some additional assumptions on the parameter space $\mathcal{P}$ of distributions F, viz., Assumption A.3–A.8, in Section 14 in online Appendix B. These assumptions are not very restrictive.

Theorem 5.1 *The nominal level* $1-\alpha$ *CI's* $CI_{n,\Delta U}(\alpha)$ *and* $CI_{n,\Delta L}(\alpha)$ *satisfy*

(a) $\liminf_{n\to\infty} \inf_{F \in \mathcal{P}} P_F(\Delta^{\inf}_F \in CI_{n,\Delta U}(\alpha)) \geq 1-\alpha$ *under Assumptions* A.0–A.6 *and*

(b) $\liminf_{n\to\infty} \inf_{F \in \mathcal{P}} P_F(\Delta^{\inf}_F \in CI_{n,\Delta L}(\alpha)) \geq 1-\alpha$ *under Assumptions* A.0–A.5, A.7, *and* A.8.

Comment. Theorem 5.1 implies that the two-sided CI $CI_{n,\Delta}(\alpha)$ in (5.3) has correct asymptotic level of $1-\alpha$.

Now, we give a condition under which $CI_{n,\Delta U}(\alpha) \subset (-\infty, 0)$ wp$\rightarrow$1. This condition requires that the identified set is nonempty and contains slack (SLK) points.

Assumption SLK. The sequence $\{F_n\}_{n\geq 1}$ is such that $n^{1/2}\Delta^{\inf}_{F_n} \to -\infty$.

Assumption SLK holds for a fixed distribution F (that does not depend on n) if $\Theta_I(F)$ contains a slack point. More generally, for a sequence of identified sets $\{\Theta_I(F_n)\}_{n\geq 1}$ that may depend on n, Assumption SLK requires that the sets contain slack points for which the magnitude of slackness is of order greater than $n^{-1/2}$. Assumption SLK is equivalent to the existence of a sequence $\{\theta^I_n \in \Theta_I(F_n)\}_{n\geq 1}$ for which $n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta^I_n) \to \infty\ \forall j \leq k$.[21]

Next, we give a condition under which $CI_{n,\Delta L}(\alpha) \subset (0,\infty)$ wp$\to$1. This condition defines "large-local" or "global" model misspecification (MM).

Assumption MM. The sequence $\{F_n\}_{n\geq 1}$ is such that $n^{1/2}\Delta^{\inf}_{F_n} \to \infty$.

Assumption MM is equivalent to $n^{1/2}r^{\inf}_{F_n} \to \infty$, since $r^{\inf}_F = \max\{\Delta^{\inf}_F, 0\}$.

Theorem 5.2 *Suppose Assumptions* A.0–A.6 *hold.* (a) *For sequences* $\{F_n\}_{n\geq 1}$ *that satisfy Assumption* SLK, $\liminf_{n\to\infty} P_{F_n}(\widehat{\Delta}_{n,\Delta U}(\alpha) < 0) = 1$. (b) *For sequences* $\{F_n\}_{n\geq 1}$ *that satisfy Assumption* MM, $\liminf_{n\to\infty} P_{F_n}(\widehat{\Delta}_{n,\Delta L}(\alpha) > 0) = 1$.

The asymptotic level and consistency properties of the test in (5.5) of $H_0: \Delta^{\inf}_F > 0$ follow from the properties of the CI $CI_{n,\Delta U}(\alpha)$ given in Theorems 5.1(a) and 5.2(a).

Corollary 5.3 *The nominal level* α *test in* (5.5) *of* $H_0 : \Delta^{\inf}_F > 0$ *satisfies* (a) $\limsup_{n\to\infty} \sup_{F\in\mathcal{P}:\Delta^{\inf}_F>0} P_F(n^{1/2}\widehat{\Delta}^{\inf}_n < -\widehat{c}_{n,\Delta U}(1-\alpha)) \leq \alpha$ *under Assumptions* A.0–A.6 *and*

(b) $\liminf\limits_{n\to\infty} P_{F_n}(n^{1/2}\widehat{\Delta}^{\inf}_n < -\widehat{c}_{n,\Delta U}(1-\alpha)) = 1$ *for sequences* $\{F_n\}_{n\geq 1}$ *that satisfy Assumption* SLK, *under Assumptions* A.0–A.6.

5.6 Tuning Parameters, Implementation, and Computation of the Misspecification Index CI's

5.6.1 Tuning Parameters

The tuning parameters for the misspecification index CI's are: B, ι, κ_n, and τ_n. We recommend the same choices $B = 1000$, $\iota = 10^{-6}$, $\kappa_n = \tau_n = (\ln n)^{1/2}$ as in Section 4.7.1 and for the same reasons.

Section 8.1.3 in online Appendix A reports results concerning the sensitivity to the tuning parameters of the rejection probabilities of the MI test in a lower/upper bound model. Section 10.1 in online Appendix A reports the sensitivity of the MI CI lower and upper bounds in the empirical illustration in Section 6 below. For the tuning parameters τ_n, ι, and B, there is little sensitivity to

[21] A set of sufficient conditions for Assumption SLK is given in Lemma 24.2 in online Appendix C.

Algorithm 2 Calculating the upper CI for $\Delta_F^{\inf}$, $CI_{n,\Delta U}(\alpha)$.

Inputs: $\{W_i\}_{i\le n}$, $m(\cdot)$, Θ, α, B, κ_n, τ_n, ι
$\triangleright$ recommend $B = 1{,}000$, $\kappa_n = \tau_n = (\ln n)^{1/2}$, $\iota = 10^{-6}$

Output: Upper CI for $\Delta_F^{\inf}$

function $\widehat{\Delta}_{nj}(\theta)$ $\triangleright$ for $j = 1, \ldots, k$
    $\overline{m}_{nj}(\theta) \leftarrow n^{-1}\sum_{i=1}^{n} m_j(W_i, \theta)$
    $\widehat{\sigma}_{nj}^2(\theta) \leftarrow n^{-1}\sum_{i=1}^{n}(m_j(W_i,\theta) - \overline{m}_{nj}(\theta))^2$
    $\widehat{m}_{nj}(\theta) \leftarrow \frac{\overline{m}_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)}$
    return $-\widehat{m}_{nj}(\theta)$
end function
$\widehat{\Delta}_n^{\inf} \leftarrow \inf_{\theta\in\Theta}\max_{j\le k}\widehat{\Delta}_{nj}(\theta)$ $\triangleright$ estimator of misspecification index
function $A_{n,\Delta U}^*(\theta, \{W_{ib}^*\}_{i\le n})$ $\triangleright$ $\{W_{ib}^*\}_{i\le n}$: generic bootstrap sample
    for $j = 1, \ldots, k$ **do**
        $\overline{m}_{njb}^*(\theta) \leftarrow n^{-1}\sum_{i=1}^{n} m_j(W_{ib}^*, \theta)$,
        $\widehat{\sigma}_{njb}^{*2}(\theta) \leftarrow n^{-1}\sum_{i=1}^{n}(m_j(W_{ib}^*,\theta) - \overline{m}_{nj}^*(\theta))^2$
        $\widehat{\nu}_{njb}^*(\theta) \leftarrow n^{1/2}\left(\frac{\overline{m}_{njb}^*(\theta)}{\widehat{\sigma}_{njb}^*(\theta)} - \widehat{m}_{nj}(\theta)\right)$
        $\widehat{e}_{nj}(\theta) \leftarrow n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}_n^{\inf}\right) - \widehat{sd}_{njB}(\theta)\kappa_n$
        $\triangleright$ See (5.9) and (5.11) for definition of $\widehat{sd}_{njB}(\theta)$
        $\xi_{nj}^e(\theta) \leftarrow (\widehat{sd}_{njB}(\theta)\kappa_n)^{-1} n^{1/2}\left(\widehat{\Delta}_{nj}(\theta) - \widehat{\Delta}_n^{\inf}\right)$
    end for
    $\widehat{J}_{neB}(\theta) \leftarrow \{j \in \{1, ..., k\} : \widehat{\Delta}_{nj}(\theta) \ge \max_{j_1\le k}\widehat{\Delta}_{nj_1}(\theta) - \widehat{sd}_{njB}(\theta) n^{-1/2}\kappa_n\}$
    return $\min_{j_1\in\widehat{J}_{neB}(\theta)}\max_{j\le k}\left(-\widehat{\nu}_{njb}^*(\theta) + 1(j \ne j_1)\widehat{e}_{nj}(\theta) + 1(j = j_1)\varphi(\xi_{nj}^e(\theta))\right)$
end function $\triangleright$ $\varphi(\xi) = \infty 1(\xi > 1)$
function UpperCI($\{W_i\}_{i\le n}$, B, α)
    $\widehat{\Theta}_{\min,n} \leftarrow \{\theta \in \Theta : \max_{j\le k}\widehat{\Delta}_{nj}(\theta) \le \widehat{\Delta}_n^{\inf} + \tau_n/n^{1/2}\}$
    for $b = 1, ..., B$ **do**
        Draw bootstrap sample $\{W_{ib}^*\}_{i\le n}$
        $A_{n,\Delta U}^*[b] \leftarrow \inf_{\theta\in\widehat{\Theta}_{\min,n}} A_{n,\Delta U}^*(\theta, \{W_{ib}^*\}_{i\le n})$
    end for
    $\widehat{c}_{n,\Delta U}(1-\alpha) \leftarrow \text{quantile}(-A_{n,\Delta U}^*, 1-\alpha) + \iota$ $\triangleright$ critical value
    $\widehat{\Delta}_{n,U}^{\inf}(\alpha) \leftarrow \widehat{\Delta}_n^{\inf} + \frac{\widehat{c}_{n,\Delta U}(1-\alpha)}{n^{1/2}}$
    return $CI_{n,\Delta U}(\alpha) \leftarrow (-\infty, \widehat{\Delta}_{n,U}^{\inf}(\alpha)]$
end function

halving or doubling the recommended tuning parameters. For the κ_n parameter, there is sensitivity in some model scenarios, but it is less than that of the SPUR2 tests.

5.6.2 Implementation

See Algorithm 2 for pseudo-code for computing the MI CI upper bound. Analogous pseudo-code for the MI CI lower bound is much simpler and, for brevity, is not provided.

5.6.3 Computation

Computation of the MI CI's in (5.3) requires computing $\widehat{\Delta}_n^{\inf}$ and the critical values $\widehat{c}_{n,\Delta L}(1-\alpha)$ and $\widehat{c}_{n,\Delta U}(1-\alpha)$. We recommend computing $\widehat{\Delta}_n^{\inf}$ by solving the problem: $\inf_{\gamma \in R, \theta \in \Theta} \gamma$ subject to $\widehat{\Delta}_{nj}(\theta) \leq \gamma \, \forall j \leq k$ using standard nonlinear optimization software (e.g., `slsqp` in the R library `nloptr`, which is what we use, or `fmincon` in Matlab). When $\widehat{\Delta}_{nj}(\theta)$ is differentiable, this formulation makes both the objective function and constraints differentiable. If an analytic expression for the Jacobian of $\widehat{m}_n(\theta)$ is available, this can be passed to the algorithm to increase the speed of the computation.

In the empirical illustration, the analytical Jacobian of $\widehat{m}_n(\theta)$ is passed to the algorithm. To compute $\widehat{\Delta}_n^{\inf}$, we use 100 initial values drawn according to a Sobol sequence on Θ. The computation is run in parallel over these initial values. The computation takes less than 10 seconds.

The critical value $\widehat{c}_{n,\Delta U}(1-\alpha)$ requires the computation of $\{A^*_{n,\Delta U,b}\}_{b \leq B}$. To compute $A^*_{n,\Delta U,b}$, we use constrained optimization with the objective function and constraints being

$$\begin{aligned} A^*_{n,\Delta U,b}(\theta) &:= \min_{j_1 \in \widehat{J}_{neB}(\theta)} \max_{j \leq k} \left(-\widehat{\nu}^*_{njb}(\theta) + 1(j \neq j_1)\widehat{e}_{nj}(\theta) + 1(j = j_1)\varphi(\xi^e_{nj}(\theta))\right) \text{ and} \\ -\widehat{m}_{nj}(\theta) &\leq \widehat{\Delta}_n^{\inf} + \tau_n / n^{1/2} \text{ for } j \leq k, \end{aligned} \tag{5.16}$$

respectively. The constraints correspond the requirement that $\theta \in \widehat{\Theta}_{\min,n}$. When the Jacobian of $\widehat{m}_n(\theta)$ can be calculated analytically, providing this to the optimization algorithm typically results in faster and more stable calculations.[22]

For the initial value for the optimization problem, we recommend using $\theta^{\text{init}} \in \arg\min_{\theta \in \Theta} \max_{j \leq k} \widehat{\Delta}_{nj}(\theta)$, which is typically unique. It works well in the empirical illustration and in the simulations. This is expected because the set $\widehat{\Theta}_{\min,n}$ is a small expansion of $\arg\min_{\theta \in \Theta} \max_{j \leq k} \widehat{\Delta}_{nj}(\theta)$. Since one has to calculate $\widehat{\Delta}_n^{\inf}$ anyway to compute the $CI_{n,\Delta U}(\alpha)$ CI, this initial value is readily available.

Computation of $\{A^*_{n,\Delta U,b}\}_{b \leq B}$ is easily done in parallel, which greatly reduces the time required to carry out the computations. In the empirical illustration, computation of $\{A^*_{n,\Delta U,b}\}_{b \leq B}$ is done in parallel over the 36 threads on the Intel 6240 processor and it takes approximately 22 minutes for $B = 1{,}000$. Thus, the total computation time for the $CI_{n,\Delta U}(\alpha)$ CI is approximately 22 minutes.

Computation of $CI_{n,\Delta U}(\alpha)$ for alternative values of α, such as $\alpha/2$ for the upper bound of the two-sided CI $CI_{n,\Delta}(\alpha)$ and a value α_1 that is used by the SPUR2 CS, takes essentially no additional time because each only requires an additional sample quantile calculation.

The values $\{A^*_{n,\Delta L,b}\}_{b \leq B}$ are computed in the same way as $\{A^*_{n,\Delta U,b}\}_{b \leq B}$, see (5.16), using

[22] If $m(W,\theta)$ is additive in the sense that it can be written as $m(W,\theta) = g(W) + f(\theta)$, then $\widehat{\Omega}_{n+}(\theta)$ in (5.11), which appears in the definition of $\widehat{sd}_{njB}(\theta)$ in (5.9), does not depend on θ. Hence, $\widehat{\Omega}_{n+}(\theta)$ only needs to be calculated once.

the same initial value, but with the simpler objective function $A^*_{n,\Delta L,b}(\theta) := \max_{j\leq k}(-\widehat{\nu}^*_{njb}(\theta) - \varphi(-\xi^e_{nj}(\theta)))$ and the k inequality constraints $-\widehat{m}_{nj}(\theta) \leq \widehat{\Delta}^{\inf}_n$ for $j \leq k$, which correspond to $\theta \in \widehat{\Theta}_{\min,L,n}$. In the empirical illustration, computation of $\{A^*_{n,\Delta L,b}\}_{b\leq B}$ takes about 30 seconds using the Intel 6420 processor with 36 threads running in parallel for $B = 1{,}000$. So, computation of the $CI_{n,\Delta L}(\alpha)$ CI and the lower bound of the two-sided CI $CI_{n,\Delta}(\alpha)$ takes about 40 seconds (including the time to compute $\widehat{\Delta}^{\inf}_n$).

6 Empirical Illustration

We revisit the analysis by Kline and Tamer (2016) of entry behavior in airline markets and examine the potential effect of misspecification on the results of the study. The same empirical setting has been considered by Kaido, Molinari, and Stoye (2019) using a non-Bayesian approach. Details on the data and definitions of the covariates can be found in Section 8 of Kline and Tamer (2016).

6.1 Model

For each market $i = 1, \ldots, n$, which is defined as a trip between two airports, there are two types of entrants; low cost carriers (LCC) and other airlines (OA). All airlines in each market are aggregated into these two groups, which simplifies the entry game to a two-player game. A binary random variable $Y_{i,t}$ indicates entrance of player $t \in \{\text{LCC}, \text{OA}\}$ into market i, where t denotes type. The profit that player t makes in market i is given by

$$\Pi_{i,t} = \begin{cases} X'_{i,t}\beta_t + Y_{i,-t}\gamma_t + \varepsilon_{i,t} & \text{if } Y_{i,t} = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6.1}$$

where $-t$ denotes the opponent of t, $X_{i,t}$ is a vector of observable covariates specified below, and $\varepsilon_i := (\varepsilon_{i,\text{LCC}}, \varepsilon_{i,\text{OA}})' \sim N\left(0, \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)\right)$ is a vector of unobserved (to the econometrician) profit shifters independent of the observed covariates and across markets. The parameters of the model are $\theta := (\beta'_{\text{LCC}}, \gamma_{\text{LCC}}, \beta'_{\text{OA}}, \gamma_{\text{OA}}, \rho)'$, where we assume $\gamma_{\text{LCC}}, \gamma_{\text{OA}} \leq 0$ and $\rho \geq 0$, as in Kline and Tamer (2016, Section 7.1).[23] The observable covariates are specified as $X_{i,t} := (1, X^{\text{size}}_i, X^{\text{pres}}_{i,t})'$ with coefficient $\beta_t := (\beta^{\text{const}}_t, \beta^{\text{size}}_t, \beta^{\text{pres}}_t)'$, where X^{size}_i is the indicator of whether the size of market i is

[23] For some of the analysis, we treat ρ as known to see how the misspecification index changes as we vary ρ. In this case, with some abuse of notation, the parameters of the model are understood to be $\theta := (\beta'_{\text{LCC}}, \gamma_{\text{LCC}}, \beta'_{\text{OA}}, \gamma_{\text{OA}})'$. The motivation for considering known ρ is that ρ is poorly identified and we want to ensure that this does not effect the results. On the other hand, we expect that it will not effect the results because the misspecification index, GMS, and SPUR2 methods are all robust to weak, or lack of, identification.

greater than the median size across all markets and $X_{i,t}^{\text{pres}}$ is a measure of an airline's presence in the two airports associated with market i. The precise definition of these variables can be found in Kline and Tamer (2016).

We assume complete information, so that the players observe ε_i in addition to everything the econometrician observes, and that the market outcome is determined by a pure strategy Nash equilibrium. Given this, there are two conditional moment equalities and two conditional moment inequalities. Writing the two moment equalities as four moment inequalities, the model can be written as six conditional moment inequalities. Since X_i is discrete with its support $\mathcal{X}$ consisting of $2^3 = 8$ different values, the six conditional moment inequalities can be transformed into $k = 48$ unconditional moment inequalities. The sample size is $n = 7{,}882$.

Let $Y_i := (Y_{i,\text{LCC}}, Y_{i,\text{OA}})'$ and $X_i := (X_{i,\text{LCC}}', X_{i,\text{OA}}')'$. Let $P_{00}(x, \theta)$ denote the model probability that $Y_i = (0, 0)$ when $X_i = x$ and the parameter equals θ. Define $P_{11}(x, \theta)$ analogously. Let $\underline{P}_{01}(x, \theta)$ and $\overline{P}_{01}(x, \theta)$ denote the model lower and upper bounds on the probabilities that $Y_i = (0, 1)$ and $Y_i = (1, 0)$, respectively, given $X_i = x$ and the parameter θ. Explicit expressions for $P_{00}(x, \theta)$, $\underline{P}_{01}(x, \theta)$, etc. are given in Section 10 in online Appendix A. Let $p_x := P(X_i = x)$. Following Kaido, Molinari, and Stoye (2019), we take p_x to be known.[24] This yields the following moment functions that are additively separable in the data and parameters:

$$\begin{aligned}
&E[1(Y_i = (0,0)', X_i = x) - P_{00}(x,\theta)p_x] \geq 0, \\
&E[P_{00}(x,\theta)p_x - 1(Y_i = (0,0)', X_i = x)] \geq 0, \\
&E[1(Y_i = (0,1)', X_i = x) - \underline{P}_{01}(x,\theta)p_x] \geq 0, \\
&E[\overline{P}_{01}(x,\theta)p_x - 1(Y_i = (0,1)', X_i = x)] \geq 0, \\
&E[1(Y_i = (1,1)', X_i = x) - P_{11}(x,\theta)p_x] \geq 0, \\
&E[P_{11}(x,\theta)p_x - 1(Y_i = (1,1)', X_i = x)] \geq 0, \qquad (6.2)
\end{aligned}$$

for $x \in \mathcal{X}$. In practice, we take the empirical distribution of X_i to be the true distribution and use it in place of p_x, as in Kaido, Molinari, and Stoye (2019). The parameter spaces for β_t, γ_t, and ρ are $[-8, 2] \times [-2, 3] \times [-2, 10]$, $[0, 4]$, and $[0, 0.85]$, respectively, for $t =$LCC, OA, as in Kaido, Molinari, and Stoye (2019).

If the model is misspecified, the MR-identified set consists of the parameter values that satisfy a minimally-relaxed version of these inequalities across the 8 covariate values, as defined in Section 3.

[24] If p_x is unknown, one can use inequalities given in Section 10 in online Appendix A.

6.2 Results

We diagnose whether the model is misspecified using the method described in Section 5, and compute SPUR2 projection CI's for each of the model parameters. We use the "sum-of-squares" test function S_1 defined in (4.6), $\alpha_1 = .005$ and $\alpha_2 = .045$ for the SPUR2 CI's, bootstrap sample size $B =$ 1,000, and tuning parameters $\kappa_n = \tau_n = (\ln n)^{1/2}$, as recommended in Sections 4.7.1 and 5.6.1.

Table 1: CI's for $\Delta_F^{\inf}$. The first column gives the specified values of ρ. The second through fifth columns give the corresponding estimators $\widehat{\Delta}_n^{\inf}$, lower-bound CI's, upper-bound CI's, and two-sided CI's for $\Delta_F^{\inf}$, respectively. All of the CI's have nominal 95% confidence level.

ρ	$\widehat{\Delta}_n^{\inf}$	Lower-Bound CI	Upper-Bound CI	Two-sided CI
0.0	0.023	[-0.013, ∞)	($-\infty$, 0.056]	[-0.021, 0.058]
0.2	0.021	[-0.015, ∞)	($-\infty$, 0.054]	[-0.024, 0.056]
0.4	0.021	[-0.015, ∞)	($-\infty$, 0.055]	[-0.021, 0.056]
0.6	0.021	[-0.016, ∞)	($-\infty$, 0.054]	[-0.019, 0.055]
0.75	0.019	[-0.017, ∞)	($-\infty$, 0.052]	[-0.022, 0.053]
0.85	0.018	[-0.018, ∞)	($-\infty$, 0.050]	[-0.023, 0.052]
unknown	0.018	[-0.018, ∞)	($-\infty$, 0.050]	[-0.024, 0.052]

Table 1 provides 95% lower-bound, upper-bound, and two-sided CI's for $\Delta_F^{\inf}$ for different values of ρ, when ρ is treated as an known parameter, as well as for ρ unknown. The lower-bound CI's include 0 for all ρ values, which implies that one cannot reject the null hypothesis of correct specification (i.e., $H_{00} : \Delta_F^{\inf} \leq 0$) at the 5% level for any fixed ρ or ρ unknown. The upper-bound CI's also include 0 for all ρ values, which implies that one cannot reject the null hypothesis of misspecification (i.e., $H_0 : \Delta_F^{\inf} > 0$) at the 5% level for any fixed ρ or ρ unknown. Hence, standard inference methods in the literature may give spuriously precise CI's for the model parameters. All of the 95% two-sided CI's for $\Delta_F^{\inf}$ include 0, which implies that correct specification and misspecification is consistent with the data.

For each model parameter, Table 2 reports the CI that is obtained by projecting the standard GMS CS (second column) and SPUR2 CS (third column).[25] [26] It is clear that the SPUR2 CI's are

[25] By definition, the projection CI for the jth component of θ is $[\inf_{\theta \in CS} \theta_j, \sup_{\theta \in CS} \theta_j]$, where CS denotes the CS obtained by inverting the standard GMS test or the SPUR2 test.

[26] The first column of Table 2 can be compared with Table 1 of Kaido, Molinari, and Stoye (2019) (KMS). They use "calibrated projection" and the "max" test statistic, rather than the standard projection method and the S_1 test statistic used here. This results in some differences between the KMS and GMS CI's. The nine KMS CI's corresponding to those in Table 2 are [-2.060, -.851], [.188, .403], [1.751, 1.995], [-1.442, -.188], [.396, .590], [.338, .565], [.397, .581], [-1.470, -.766], and [.186, .850], respectively.

noticeably different from (and wider than) the standard GMS CI's. The difference is significant in the sense that six of the seven parameters that are statistically significantly different from zero using the standard GMS procedure are insignificant using the SPUR2 procedure. This also suggests that the standard GMS CI's may be spuriously precise (and thus misleading) in this empirical context.

Table 2: Projection CI's for model parameters obtained from the standard GMS and SPUR2 95% confidence sets for θ. Here, ρ is treated as an unknown parameter.

Parameter	Standard GMS	SPUR2
$\beta_{\text{LCC}}^{\text{const}}$	[-2.177, -0.910]	[-8.000, 0.411]
$\beta_{\text{LCC}}^{\text{size}}$	[0.139, 0.416]	[-0.659, 1.183]
$\beta_{\text{LCC}}^{\text{pres}}$	[1.691, 1.991]	[0.927, 9.264]
γ_{LCC}	[-1.356, 0.000]	[-2.269, 0.000]
$\beta_{\text{OA}}^{\text{const}}$	[0.440, 0.583]	[0.056, 1.002]
$\beta_{\text{OA}}^{\text{size}}$	[0.362, 0.544]	[-0.280, 1.265]
$\beta_{\text{OA}}^{\text{pres}}$	[0.403, 0.590]	[-0.441, 1.189]
γ_{OA}	[-1.461, -0.392]	[-2.253, 0.000]
ρ	[0.000, 0.850]	[0.000, 0.850]

The results above show why it can be misleading to use a standard specification test followed by a moment inequality confidence set that is not robust to spurious precision under misspecification. For instance, suppose an empirical researcher uses the resampling test of Bugni, Canay, and Shi (2015) as a first-stage specification test. This test fails to reject the null hypothesis of correct specification (or more precisely, the null hypothesis that the identified set is nonempty) at nominal level 5% for all values of ρ.[27] Hence, the empirical researcher proceeds by using the standard GMS CS to construct the CI's, obtaining what is in the second column of Table 2. The researcher will consider these CI's to be "correct" because the model "passed" the specification test, but, as discussed earlier, these CI's differ considerably with the SPUR2 CI's.

A sensitivity analysis of the results in Tables 1 and 2 to the choice of tuning parameters is given in Section 10.1 in online Appendix A. There is very little sensitivity to τ_n, α_1, ι, and B, but some sensitivity to κ_n. Power results for the MI and SPUR2 tests, which are inverted to obtain the CI's in Tables 1 and 2, in a simplified version of the airline entry model are provided in Section 10.2 in online Appendix. The results show that these tests have reasonable power.

The computation of, and the computation times for, the results in Tables 1 and 2 are discussed above in Sections 4.7.3 and 5.6.3. In sum, using a computer with an Intel Xeon Gold 6240 processor,

[27]This test is implemented using the "sum-of-squares" test function S_1 and $\kappa_n = (\ln n)^{1/2}$.

which has 18 cores with double threads, and using the `slsqp` command in the R package `nloptr` for all of the constrained optimization problems, the results in Table 1 for unknown ρ and Table 2 take about 10 seconds to compute $\widehat{\Delta}_n^{\inf}$, 22 minutes for the MI CI upper bounds, .5 additional minutes for the MI CI lower bounds, and 88.5 additional minutes for the nine SPUR2 projection CI's of which 67.5 and 21 minutes are for the SPUR1 and GMS projection CI's, respectively. The total time is 110.5 minutes.

References

Allen R., and J. Rehbeck (2019), "Assessing Misspecification and Aggregation for Structured Preferences," Department of Economics Research Report #2019-4, Western University, London, Ontario, Canada.

Andrews, D. W. K., (2000), "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space," *Econometrica,* 68, 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, 2805–2826.

Andrews D. W. K. and M. Buchinsky (2000), "A Three-step Method for Choosing the Number of Bootstrap Repetitions," *Econometrica,* 68, 2000, 23–51

Andrews D. W. K. and M. Buchinsky (2001), "Evaluation of a Three-step Method for Choosing the Number of Bootstrap Repetitions," *Journal of Econometrics,* 103, 345–386.

Andrews, D. W. K., and P. Guggenberger (2009), "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory,* 25, 669–709.

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 Number 2184, Yale University.

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

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

Bugni, F. A., I. A. Canay, and P. Guggenberger (2012), "Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models," *Econometrica,* 80, 1741–1768.

Bugni, F. A., I. A. Canay, and X. Shi (2015), "Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics,* 185, 259–282.

——— (2017), "Inference for Subvectors and Other Functions of Partially Identified Parameters," *Quantitative Economics,* 8, 1–38.

Canay, I. A. (2010), “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” *Journal of Econometrics,* 156, 408-425.

Canay, I. A., and A. M. Shaikh (2016), “Practical and Theoretical Advances for Inferences in Partially Identified Models,” in *Advances in Economics and Econometric: Eleventh World Congress,* Vol. 2, 271–306, ed. by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson. New York: Cambridge University Press.

Chernozhukov, V., D. Chetverikov, and K. Kato (2019), “Inference on Causal and Structural Parameters Using Moment Inequalities,” *Review of Economic Studies,* 86, 1867–1900.

Chen, X., T. M. Christensen, and E. Tamer (2018), “Monte Carlo Confidence Sets for Identified Sets,” *Econometrica,* 86, 1965–2018.

Chernozhukov, V., H. Hong, and E. Tamer (2007), “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica,* 75, 1243–284.

Chiburis, R. (2009), “Approximately Most Powerful Tests for Moment Inequalities,” unpublished manuscript, Department of Economics, University of Texas, Austin.

Cox, G. and X. Shi (2019), “A Simple Uniformly Valid Test for Inequalities,” unpublished manuscript, Department of Economics, National University of Singapore.

Dickstein, M. J. and E. Morales (2018), “What Do Exporters Know?” *Quarterly Journal of Economics,* 133, 1753–1801.

Efron, B. and R. Tibshirani (1986), “Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy,” *Statistical Science,* 1, 54–77.

Galichon, A. and M. Henry (2009), “A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters,” *Journal of Econometrics,* 152, 186–196.

Gallant, A. R., and H. White (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell.

Guggenberger, P., J. Hahn, and K. Kim (2008), “Specification Testing under Moment Inequalities,” *Economics Letters,* 2, 375–378.

Hall, A. R., and A. Inoue (2003), “The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models,” *Journal of Econometrics,* 114, 361–394. Corrigendum in *Journal of Econometrics,* 2007, 141, 1418.

Ho, K., and A. Pakes (2014), "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review,* 104, 3841–3884.

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

Kaido H., and H. White (2013), "Estimating Misspecified Moment Inequality Models," in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis,* ed. by X. Chen and N. R. Swanson. New York: Springer.

Kawai, K., and Y. Watanabe (2013), "Inferring Strategic Voting," *American Economic Review,* 103, 624–662.

Kédangi, D., L. Li, and I. Mourifié (2021), "Discordant Relaxations of Misspecified Models," arXiv:2012.11679v3.

Kline, B. and E. Tamer (2016), "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics,* 7, 329–366.

Masten, M. A. and A. Poirier (2020), "Salvaging Falsified Instrumental Variables Models," *Econometrica,* 89, 1449–1469.

Molinari, F. (2020): "Microeconometrics with Partial Identification," arxiv.org/pdf/2004_11751 /pdf.

Pakes, A. (2010), "Alternative Models for Moment Inequalities," *Econometrica,* 78, 1783–1822.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2011), "Moment Inequalities and Their Application," unpublished manuscript, Department of Economics, Harvard University.

——— (2015), "Moment Inequalities and Their Application," *Econometrica,* 83, 315–334.

Ponomareva, M., and E. Tamer (2011), "Misspecification in Moment Inequality Models: Back to Moment Equalities?," *Econometrics Journal,* 14, 186–203.

Romano, J. P., and A. Shaikh (2008), "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Inference and Planning* (*Special Issue in Honor of T. W. Anderson*), 138, 2786–2807.

——— (2010), "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica,* 78, 169–211.

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

Rosen, A. M. (2008), "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics,* 146, 107-117.

Santos, A. (2012), "Inference in Nonparametric Instrumental Variables with Partial Identification," *Econometrica,* 80, 213–275.

van der Vaart, A., and J. Wellner (1996), *Weak Convergence and Empirical Processes.* New York: Springer.

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