Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

7-1-2019

Inference in Moment Inequality Models That Is Robust to Spurious Precision under Model Misspecification

Donald W.K. Andrews

Soonwoo Kwon

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation

Andrews, Donald W.K. and Kwon, Soonwoo, "Inference in Moment Inequality Models That Is Robust to Spurious Precision under Model Misspecification" (2019). *Cowles Foundation Discussion Papers*. 71.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/71

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

INFERENCE IN MOMENT INEQUALITY MODELS THAT IS ROBUST TO SPURIOUS PRECISION UNDER MODEL MISSPECIFICATION

By

Donald W. K. Andrews and Soonwoo Kwon

July 2019

Revised October 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2184R

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

http://cowles.yale.edu/

Inference in Moment Inequality Models That Is Robust to Spurious Precision under Model Misspecification

Donald W. K. Andrews
Cowles Foundation
Yale University

Soonwoo Kwon
Department of Economics
Yale University

First Version: March 12, 2017
Revised: October 9, 2019*

*The authors' email addresses are donald.andrews@yale.edu and soonwoo.kwon@yale.edu. Their postal address is Cowles Foundation, Yale University, P.O. Box 208281, New Haven, CT, 06520-8281, USA.

For helpful comments, the authors thank 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

Standard tests and confidence sets 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.

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

JEL Classification Numbers: C10, C12.

1 Introduction

In the moment inequality literature, the identified set consists of all parameters that satisfy the population moment inequalities. If a model is correctly specified, the identified set is not empty. If the identified set is empty, the model is misspecified. Tests and confidence sets (CS's) in the literature are designed to have correct asymptotic level under the assumption of correct model specification. However, these methods typically lead to spurious precision under model misspecification when the identified set is empty. By spurious precision of a CS, we mean that its coverage probability 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. Practitioners who observe a relatively short confidence interval (CI) or small CS can be mislead by spurious precision. Under the assumption that the model is correct, a small CS provides considerable information. But, a small CS is misleading if it is just a by-product of model misspecification.

In this paper, we develop inference methods that are robust to model misspecification in the sense that they 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.

Misspecification is ubiquitous in empirical work because models are approximations of reality. Hence, it is desirable to use methods that are robust to model misspecification. It is well-known that standard econometric methods, such as maximum likelihood, least squares, and generalized method of moments (GMM), are robust, in a certain sense, to model misspecification. The maximum likelihood, least squares, and GMM estimators converge in probability to pseudo-true values, and tests and CS's based on these estimators have correct asymptotic level, defined with respect to the pseudo-true parameters, provided standard errors are computed appropriately.[1]

The performance of standard inference methods under misspecification is subject to the criticism that the pseudo-true parameter for a given estimation method may not be the most interesting pseudo-true parameter from a substantive empirical perspective. For example, with GMM estimators, the pseudo-true value is the parameter value that minimizes the population GMM criterion function and it may be difficult to interpret. Furthermore, different choices of the weight matrix yield different pseudo-true values.

[1] The pseudo-true value for maximum likelihood minimizes the Kullback-Leibler quasi-distance between the true distribution of the data and the distributions in the specified model. The pseudo-true value for least squares provides the best linear approximation of the true conditional mean function in terms of mean square. The pseudo-true value for GMM minimizes a population quadratic form that depends on the weight matrix employed by the GMM estimator. References include White (1982), Gallant and White (1988), Hall and Inoue (2003), and Hansen and Lee (2019), among others.

Nevertheless, standard maximum likelihood, least squares, and GMM methods, appropriately defined, are not subject to spurious precision under model misspecification. That is, these tests and CS's deliver correct asymptotic level for some pseudo-true parameter under model misspecification. Given that most, or almost all, models exhibit some amount of misspecification, these robustness properties are relied on in most, or almost all, empirical applications that employ these methods.

Standard inference methods in the literature for moment inequalities do not share the robustness property of standard maximum likelihood, least squares, and GMM methods discussed above. Yet, 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.

The misspecification of moment inequality models can arise from many sources. For example, it can be due to (i) functional form and distributional assumptions, e.g., Kawai and Watanabe (2013) specify a beta error distribution and linear functional forms (which they recognize could effect their empirical results); (ii) misspecified optimizing conditions, e.g., as seems to occur in some specifications in Dickstein and Morales (2018); (iii) some degree of non-optimal behavior when the moment inequalities are based on optimal behavior; (iv) incorrect exogeneity assumptions; (v) left-out variables; (vi) mismeasured variables; (vi) invalidity of selection-on-observables assumptions; (vii) invalidity of unconditional or conditional missing-at-random assumptions; and/or (vii) unmodelled heterogeneity.

The approach taken in this paper to moment inequality models 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 $r_F^{\inf} \geq 0$ such that the relaxed moment inequalities hold for some parameter θ_I in the parameter space Θ, where F denotes the distribution of the data. The collection of such values θ_I, which may be a singleton, is defined to be the misspecification-robust (MR) identified set Θ_I. It consists of the parameter values that solve the population moment inequality model that is closest to the misspecified moment inequality model. For example, in a model where the moment inequalities are generated by profit maximization, the minimally-relaxed moment inequalities accommodate inaccuracies in the evaluation of the returns to different choices by the firms, which may arise due to wrong beliefs or many other reasons, and/or only approximate optimization by the firms, due to computational costs.

We develop tests and CS's that are spurious-precision robust (SPUR) in that they have correct asymptotic level with respect to some $\theta_I \in \Theta_I$ under model misspecification, just as they do under correct model specification. That is, we consider inference for the true value, as in Imbens and Manski (2004), or pseudo-true value, as opposed to inference for the MR identified set. The approach we take has the attribute that different choices of the test statistic employed do not affect the definition of the MR identified set Θ_I.

Compared to standard non-robust tests and CS's for moment inequality models, the SPUR procedures in this paper have the advantage of eliminating spurious precision, which is a substantial advantage, but have two potential drawbacks. First, if the model is correctly specified, SPUR tests can sacrifice power. This is not surprising because the null hypothesis considered by SPUR tests is noticeably larger than when correct specification is assumed. On the other hand, some of the tests we develop, referred to below as SPUR2 tests, are shown to sacrifice very little power asymptotically under correct specification provided the identified set contains slack points θ for which the slackness of the inequalities is of order greater than $n^{-1/2}$.

Second, the SPUR procedures are computationally more intensive than standard non-SPUR procedures. However, for a CS, the increase in computational cost is a one-time increase. That is, once one computes a single SPUR test, the computational burden of constructing a CS by test inversion is the same as for a standard non-SPUR CS.

The SPUR procedures in this paper also have several drawbacks in an absolute sense, although these are not drawbacks relative to standard procedures and in some cases are unavoidable. First, SPUR procedures eliminate spurious precision that arises to due "identifiable" model misspecification, which leads to an empty identified set (when no relaxation is employed). But, model misspecification can be present even when this set is non-empty. In such scenarios, model specification tests have trivial power and both SPUR and non-SPUR procedures provide correct asymptotic inference for a pseudo-true value, but not necessarily for the true parameter (which may or may not be well defined under misspecification). This is an unavoidable feature of moment inequality models. It is analogous to the situation in misspecified instrumental variable models that are exactly identified (i.e., for which the number of instruments equals the dimension of the parameter).

Second, the SPUR procedures provide valid inference for a pseudo-true parameter, but this pseudo-true parameter may not be the parameter value that is of greatest interest from a substantive perspective. This is the same criticism that arises with standard maximum likelihood, least squares, and GMM methods.[2] This is not a drawback relative to standard procedures, 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.

Third, different definitions of the pseudo-true parameter could be considered. This is similar to the situation with standard estimation methods in which different choices of the form of the estimator yields different pseudo-true values. For example, one could consider a different weighting across moment functions than uniform weighting. However, uniform weighting is often natural. It is analogous to the use of a uniform prior in Bayesian analysis and has the advantage of eliminating a choice that may otherwise be somewhat arbitrary. This issue is not a drawback relative to standard procedures for the same reason as for the second point above. The extension to other weights is covered by results in the Supplemental Material.

In sum, although the SPUR procedures introduced in this paper have some drawbacks, these drawbacks are much less severe than the spurious precision property of standard moment inequality methods. Furthermore, some of the drawbacks are unavoidable and others are similar to the drawbacks of standard moment equality methods used in the literature.

There is a fairly extensive literature on inference methods for moment inequality models, see the review papers of Canay and Shaikh (2016) and Molinari (2019) for references. In particular, see Molinari (2019, 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 Bugni, Canay, and Shi (2015) (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

[2] However, the 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. No CS that is based on tests that have nontrivial power against $n^{-1/2}$ alternatives (when the population moment functions have derivatives that are bounded away from zero), as is highly desirable in general because it yields a relatively small CS, is capable of covering the true parameter in the knife-edge case with correct probability asymptotically when the misspecification of the moment functions is of order $n^{-1/2}$ or larger. When it is of order $o(n^{-1/2})$, both SPUR and non-SPUR CS's do cover the true value with correct probability asymptotically.

and misspecification is due to a parametric functional form, as opposed to, say, missing variables, mismeasured variables, or unanticipated endogeneity. A companion paper to this one, Andrews and Kwon (2019), provides a confidence interval for a measure of identifiable model misspecification in moment inequality models. Allen and Rehbeck (2018) consider a very similar measure of model misspecification to Andrews and Kwon (2019) and provide a CI for it in their study of demand based on quasilinear utility. In their setting, there is no unknown parameter θ, which simplifies the problem considerably.

We now summarize the contents of this paper. Section 2 describes the moment inequality model and defines the MR identified set, as described briefly above. In the bulk of the paper, the observations are assumed to be independent and identically distributed (i.i.d.).

For motivational purposes, Section 3 illustrates the spurious precision of some standard moment inequality CS's, namely, the generalized moment selection (GMS) CS's in Andrews and Soares (2010), under model misspecification. We determine the best-case asymptotic coverage probabilities of the CS's under sequences of distributions $\{F_n\}_{n\geq 1}$ that exhibit model misspecification of magnitude $r/n^{1/2}$ or greater for an index $r \geq 0$. We graph the decline in coverage probabilities as a function of r to illustrate the effect of spurious precision. The results indicate that fairly substantial under-coverage is possible with modest values of r. The asymptotics employed are a minor variant of those in Bugni, Canay, and Guggenberger (2012).

Section 4 introduces the SPUR tests and CS's that are considered in the paper. SPUR test statistics are constructed as follows. First, one estimates the nonnegative relaxation parameter, $r_F^{\inf}$, that is required to yield a non-empty MR identified set. Then, one constructs a test statistic in the same way as in Andrews and Soares (2010), but using the sample moments adjusted by this estimator, $\widehat{r}_n^{\inf}$, of $r_F^{\inf}$. In Andrews and Soares (2010), different S functions yield different test statistics. Any of these S functions can be employed, which yields a family of possible SPUR test statistics. The SPUR test statistics are combined with extended GMS (EGMS) bootstrap critical values to yield what we call SPUR1 tests and corresponding SPUR1 CS's.

Next, Section 4 introduces "adaptive" SPUR2 tests and CS's that 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 standard tests (that are not robust to spurious-precision) with probability that goes to one as $n \to \infty$ (wp$\to$1). And, if the model is misspecified, they perform "almost" the same as the robust SPUR1 test wp$\to$1.

The SPUR2 tests and CS's employ an upper bound CI for the measure $r_F^{\inf}$ of model misspecification that is developed in the companion paper Andrews and Kwon (2019). Let $\alpha = \alpha_1 + \alpha_2$, where $\alpha_1, \alpha_2 > 0$, such as $\alpha_1 = .005$ and $\alpha_2 = .045$. The CI for $r_F^{\inf}$ is employed to construct a Bonferroni level α SPUR2 test that equals a level α_2 standard non-SPUR GMS test when the CI only includes the value $r_F^{\inf} = 0$ and equals a level α_2 SPUR1 test otherwise. The "almost" modifier in the previous paragraph means that the level α SPUR2 test is the same as the level α_2 ($< \alpha$) standard non-SPUR test wp→1 under the conditions stated above, is the same as the level α_2 SPUR1 test wp→1 under the other conditions stated above, and is a mixture of the two otherwise.

Section 5 determines the asymptotic distribution of a SPUR test statistic under drifting sequences of distributions and parameter values that may be in the null or alternative hypothesis for models that may be correctly specified or misspecified. The most closely related asymptotics in the literature are those of BCS for their model specification test statistic and Bugni, Canay, and Shi (2017) for their subvector test statistic. Also related are the asymptotics of Chernozhukov, Hong, and Tamer (2007) (CHT) for the infimum over the parameter space of a moment inequality objective function. The most distinctive feature of our results compared to these three sets of results is that we allow for model misspecification. In addition, our results differ from those of CHT by considering drifting sequences of true distributions, rather than a fixed true distribution, in order to obtain uniform size results.

The asymptotics are obtained using a similar method to that in BCS, but the asymptotic distribution is more complicated due to possible model misspecification. Let k denote the number of moment inequalities. The asymptotic distributions depend on two R^k-valued nuisance parameter functions and a k-vector that are not consistently estimable. This complicates the construction of critical values. Section 5 discusses the form of the EGMS critical values for the SPUR1 test in light of the asymptotic null distribution of the SPUR test statistic.

Section 6 defines extended GMS (EGMS) bootstrap critical values for the SPUR test statistic. The SPUR1 tests and CS's use the EGMS critical values. The EGMS critical values are complicated because they use the data extensively to deal with the unknown nuisance parameter functions that arise in the asymptotic null distributions.

Section 7 shows that the SPUR1 and SPUR2 tests and CS's have correct asymptotic level (in a uniform sense) under correct model specification and misspecification under fairly simple and primitive conditions.

Section 8 provides simulation results for the size and power of the SPUR1 and SPUR2 tests in misspecified and correctly-specified versions of two models. In the correctly-specified versions, their power is compared to that of a standard non-SPUR GMS test from Andrews and Soares (2010). In the first model, the moment inequalities place lower and upper bounds on the value of a parameter. The second model is a version of the missing-data model considered in BCS. The simulation results reflect the discussion above. Under model misspecification, the SPUR1 and SPUR2 tests are found to have correct level, with under-rejection of the null in some scenarios, and very similar power. Under correct model specification, they have lower power than the standard non-SPUR test when the identified set is small. Under correct specification, the SPUR2 test has almost the same power as the non-SPUR test when the identified set is larger. Under correct specification, the SPUR2 test has almost equal or higher power than the SPUR1 test, with higher power occurring with larger identified sets.

Based on the above results, our recommended test and CS is the SPUR2 test and CS.

Section 9 establishes the uniform consistency under correct and incorrect model specification of an estimator of the MR identified set. Rate of convergence results for this estimator are given in the Supplemental Material using arguments similar to those in CHT.

The methods introduced in the paper cover moment equalities by writing each equality as two inequalities. The methods are robust to weak identification. They apply to full vector inference. Projection can be used to obtain inference for subvectors, see Kaido, Molinari, and Stoye (2019) for an algorithm for doing so. Alternative subvector methods are the focus of ongoing research.

An Appendix contains several assumptions that are not included in the body of the paper for ease of reading. The Supplemental Material provides asymptotic $n^{-1/2}$-local-alternative power results and consistency results under fixed and non-$n^{-1/2}$-local alternatives; shows that the "max" version of the SPUR test statistic is equivalent to a recentered test statistic; defines and provides properties of the CI for $r_F^{\inf}$ that is employed by the SPUR2 test and CS; discusses extensions of the results of the paper to tests with weighted moment inequalities, to tests without the standard-deviation normalization, and to non-i.i.d. observations; provides additional simulation results and some details of the simulation models; and contains proofs of all of the results given in the paper.

All limits are as the sample size $n \to \infty$. Let $R_{[\pm\infty]} := R \cup \{\pm\infty\}$, $R_{[+\infty]} := R \cup \{+\infty\}$, and $R_{+,\infty} := [0, \infty]$. Let $||\cdot||$ denote the Euclidean norm for vectors and the Frobenious norm for matrices. Let $[x]_- := \max\{-x, 0\}$ (≥ 0) and $[x]_+ := \max\{x, 0\}$ (≥ 0) for $x \in R$.

2 Moment Inequality Model and Misspecification-Robust Identified Set

2.1 Model and Misspecification-Robust Identified Set

We consider the moment inequality model

$$E_F m(W_i, \theta) \geq 0_k, \tag{2.1}$$

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, $\{W_i \in R^{d_W} : i = 1, ..., n\}$ are independent and identically distributed (i.i.d.) observations with distribution F, $m(\cdot, \cdot)$ is a known function from $\mathcal{W} \times \Theta \subset R^{d_W + d_\theta}$ to R^k, and E_F denotes expectation under F. The distribution F lies in a set of distributions $\mathcal{P}$. For notational 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 inequality functions are

$$\sigma^2_{Fj}(\theta) := Var_F(m_j(W, \theta)) > 0 \text{ for } j \leq k. \tag{2.2}$$

The population-standard-deviation-normalized sample moments are

$$\begin{aligned} \widetilde{m}_{nj}(\theta) &:= n^{-1} \sum_{i=1}^{n} \widetilde{m}_j(W_i, \theta), \text{ where } \widetilde{m}_j(W_i, \theta) := \frac{m_j(W_i, \theta)}{\sigma_{Fj}(\theta)} \ \forall j \leq k, \text{ and} \\ \widetilde{m}_n(\theta) &:= (\widetilde{m}_{n1}(\theta), ..., \widetilde{m}_{nk}(\theta))'. \end{aligned} \tag{2.3}$$

The moment inequality model in (2.1) can be written equivalently as $E_F \widetilde{m}(W, \theta) \geq 0_k$, where $\widetilde{m}(W, \theta) := (\widetilde{m}_1(W, \theta), ..., \widetilde{m}_k(W, \theta))'$. Note that $\widetilde{m}(W, \theta)$ depends on F, and hence, is not observed, but the dependence is suppressed for notational convenience.

Under correct (C) specification, i.e., when (2.1) holds, the identified set under F is defined by

$$\Theta^C_I(F) := \{\theta \in \Theta : E_F \widetilde{m}(W, \theta) \geq 0_k\}. \tag{2.4}$$

Under model misspecification, i.e., when (2.1) 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 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 correct specification and misspecification. Let

$$\begin{aligned} r_F(\theta) &:= \inf\{r \geq 0 : E_F \widetilde{m}(W, \theta) \geq -r 1_k\} \text{ and} \\ r_F^{\inf} &:= \inf_{\theta \in \Theta} r_F(\theta), \end{aligned} \tag{2.5}$$

where $1_k = (1, ..., 1)' \in R^k$. 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. Simple calculations show that

$$r_F(\theta) = \max_{j \leq k} r_{Fj}(\theta), \text{ where } r_{Fj}(\theta) := [E_F \widetilde{m}_j(W, \theta)]_-. \tag{2.6}$$

We define the MR identified set to be

$$\begin{aligned}\Theta_I(F) &:= \{\theta \in \Theta : r_F(\theta) = r_F^{\inf}\} \\ &= \{\theta \in \Theta : E_F \widetilde{m}(W, \theta) \geq -r_F^{\inf} 1_k\}. \end{aligned} \tag{2.7}$$

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.

For a given (known) $\theta_0 \in \Theta$, we are interested in tests of the hypotheses:

$$H_0 : \theta_0 \in \Theta_I(F) \text{ versus } H_1 : \theta_0 \notin \Theta_I(F) \tag{2.8}$$

for $F \in \mathcal{P}$, where $\mathcal{P}$ is a family of distributions that may be correctly specified or misspecified. We are also interested in CS's for parameter values θ in $\Theta_I(F)$. The CS that is obtained by inverting the test $\phi_n(\theta_0)$ is

$$CS_n := \{\theta \in \Theta : \phi_n(\theta) = 0\}. \tag{2.9}$$

2.2 Sample Statistics

The sample moments, variance estimators, and sample standard-deviation-normalized sample moments are

$$\begin{aligned}\overline{m}_{nj}(\theta) &:= n^{-1} \sum_{i=1}^{n} m_j(W_i, \theta) \ \forall j \leq k, \\ \widehat{\sigma}^2_{nj}(\theta) &:= n^{-1} \sum_{i=1}^{n} (m_j(W_i, \theta) - \overline{m}_{nj}(\theta))^2 \ \forall j \leq k, \\ \widehat{m}_{nj}(\theta) &:= \frac{\overline{m}_{nj}(\theta)}{\widehat{\sigma}_{nj}(\theta)} \ \forall j \leq k, \text{ and } \widehat{m}_n(\theta) = (\widehat{m}_{n1}(\theta), ..., \widehat{m}_{nk}(\theta))', \end{aligned} \tag{2.10}$$

where $m_j(W_i, \theta)$ denotes the jth element of $m(W_i, \theta)$. The sample variance and correlation matrices of the moments are

$$\begin{aligned}\widehat{\Sigma}_n(\theta) &:= n^{-1} \sum_{i=1}^{n} (m(W_i, \theta) - \overline{m}_n(\theta))(m(W_i, \theta) - \overline{m}_n(\theta))' \text{ and} \\ \widehat{\Omega}_n(\theta) &:= \widehat{D}_n^{-1/2}(\theta) \widehat{\Sigma}_n(\theta) \widehat{D}_n^{-1/2}(\theta), \text{ where} \\ \overline{m}_n(\theta) &:= n^{-1} \sum_{i=1}^{n} m(W_i, \theta) \text{ and } \widehat{D}_n(\theta) := Diag\{\widehat{\sigma}^2_{n1}(\theta), ..., \widehat{\sigma}^2_{nk}(\theta)\}. \end{aligned} \tag{2.11}$$

The standard-deviation-normalized sample moment and sample second-central-moment empirical processes are

$$
\begin{aligned}
\nu_n^m(\theta) &:= n^{1/2}(\widetilde{m}_n(\theta) - E_F \widetilde{m}_n(\theta)),\\
\widehat{\sigma}^2_{Fnj}(\theta) &:= n^{-1}\sum_{i=1}^n (m_j(W_i,\theta) - E_F m_j(W_i,\theta))^2,\\
\nu_{nj}^\sigma(\theta) &:= n^{1/2}\left(\frac{\widehat{\sigma}^2_{Fnj}(\theta)}{\sigma^2_{Fj}(\theta)} - 1\right) = n^{-1/2}\sum_{i=1}^n [(\widetilde{m}_j(W,\theta) - E_F\widetilde{m}_j(W,\theta))^2 - 1] \ \forall j \le k, \text{ and}\\
\nu_n(\theta) &:= \begin{pmatrix} \nu_n^m(\theta) \\ \nu_n^\sigma(\theta) \end{pmatrix}, \qquad (2.12)
\end{aligned}
$$

where the superscripts m and σ denote mean and variance, respectively, and by convention, the dependence of $\nu_n^m(\theta)$ and $\nu_n^\sigma(\theta)$ on F is suppressed for notational simplicity. Let $\nu_{nj}^m(\theta)$ and $\nu_{nj}^\sigma(\theta)$ denote the jth elements of $\nu_n^m(\theta)$ and $\nu_n^\sigma(\theta)$, respectively, for $j = 1, ..., k$.

2.3 Conditions for the I.I.D. Case

For the case of i.i.d. observations under F, we employ the following conditions. We define the covariance kernel $\Omega_F(\theta,\theta')$ of $\nu_n(\theta)$ as follows: for $\theta, \theta' \in \Theta$,

$$
\begin{aligned}
\Omega_F(\theta,\theta') &:= E_F \begin{pmatrix} \widetilde{m}(W,\theta) - E_F\widetilde{m}(W,\theta) \\ \widetilde{m}^\sigma(W,\theta) \end{pmatrix} \begin{pmatrix} \widetilde{m}(W,\theta) - E_F\widetilde{m}(W,\theta) \\ \widetilde{m}^\sigma(W,\theta) \end{pmatrix}' \in R^{2k\times 2k}, \text{ where}\\
\widetilde{m}^\sigma(W,\theta) &:= ([(\widetilde{m}_1(W,\theta) - E_F\widetilde{m}_1(W,\theta))^2 - 1], ..., [(\widetilde{m}_k(W,\theta) - E_F\widetilde{m}_k(W,\theta))^2 - 1])' \qquad (2.13)
\end{aligned}
$$

and $E_F\widetilde{m}^\sigma(W_i,\theta) = 0_k$ by the definition of $\widetilde{m}_j(W_i,\theta)$ in (2.2) and (2.3).

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 $\forall j \le k$, $\forall F \in \mathcal{P}$.

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 $\theta \in \Theta$ $\forall j \le k$, $\forall F \in \mathcal{P}$.

Assumption A.2. The empirical process $\nu_n(\cdot)$ is asymptotically ρ_F-equicontinuous on Θ uniformly in $F \in \mathcal{P}$.[3]

Assumption A.3. For some $a > 0$, $\sup_{F\in\mathcal{P}} E_F \sup_{\theta\in\Theta} ||\widetilde{m}(W,\theta)||^{4+a} < \infty$.

Assumption A.4. The covariance kernel $\Omega_F(\theta,\theta')$ satisfies: for all $F \in \mathcal{P}$,

[3] That is, $\lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{F\in\mathcal{P}} P_F^*(\sup_{\rho_F(\theta,\theta')<\delta} ||\nu_n(\theta) - \nu_n(\theta')||) = 0$, where P_F^* denotes outer probability and $\rho_F(\theta,\theta') := ||Var_F(\nu_n(\theta) - \nu_n(\theta'))||$ (which does not depend on n with i.i.d. observations).

$\lim_{\delta \to 0} \sup_{||(\theta_1, \theta_1') - (\theta_2, \theta_2')|| < \delta} ||\Omega_F(\theta_1, \theta_1') - \Omega_F(\theta_2, \theta_2')|| = 0.$

Assumption A.0 guarantees that the MR identified set $\Theta_I(F)$ in (2.7) is non-empty. Assumptions A.0(i) and A.0(ii) are the same as, and closely related to, Assumptions M.2 and M.3 of BCS, respectively. Assumptions A.1–A.4 are similar to, but somewhat stronger than, Assumptions A.1–A.4 in BCS. The former concern $\widetilde{m}_j(\cdot, \theta)$, $\widetilde{m}_j^2(\cdot, \theta)$, and $\nu_n(\cdot)$ and require $4 + a$ finite moments, whereas the latter only concern $\widetilde{m}_j(\cdot, \theta)$ and $\nu_n^m(\cdot)$ and only require $2 + a$ finite moments. The differences arise because we need to consider $(\nu_n^m(\cdot)', \nu_n^\sigma(\cdot))'$ here, rather than just $\nu_n^m(\cdot)$.

3 Spurious Precision of GMS CS's

In this section, we illustrate the spurious precision of some standard moment inequality CS's under model misspecification. Specifically, we provide some quantitative calculations of the best-case performance under misspecification of the GMS CS's in Andrews and Soares (2010). The results show that GMS CS's are too small when the identified set is empty and their volume shrinks to zero as the sample size goes to infinity. This reflects the fact that no pseudo-true value arises naturally under model misspecification for the GMS CS's. The asymptotic results upon which these calculations depend are a variant of those given in Bugni, Canay, and Guggenberger (2012).

Although we focus on GMS CS's here, other moment inequality methods 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 (2019), and Kaido, Molinari, and Stoye (2019). Methods designed for conditional moment inequalities also exhibit spurious precision under misspecification. For brevity, we do not provide references.

The method in Pakes, Porter, Ho, and Ishii (2011) is designed for correct specification when the identified set has a non-empty interior. Given the nature of this method, it would be complicated to analyze its behavior under model misspecification, but it seems quite unlikely that it would be robust. The subsampling method of CHT based on a recentered test statistic (which has its infimum over $\theta \in \Theta$ subtracted off) is probably the method in the literature that exhibits the least amount of spurious precision under misspecification. Whether it exhibits no spurious precision is an open question. It may be possible to answer this question for the "max" statistic using the asymptotic results of this paper, see Section 13 in the Supplemental Material, combined with the subsampling results in Andrews and Guggenberger (2010).[4]

Now, we illustrate the spurious precision of GMS CS's under misspecification. The standard test statistic is of the form

$$S_{n,Std}(\theta) := S\left(n^{1/2}\widehat{m}_n(\theta), \widehat{\Omega}_n(\theta)\right), \tag{3.1}$$

where $S(m,\Omega)$ is a test function defined as in Andrews and Soares (2010) with $m \in R^k$ and $\Omega \in \Psi$, and Ψ is a specified closed set of $k \times k$ correlation matrices. Here, we employ the S function $S_1(m,\Omega) := \sum_{j=1}^{k}[m_j]_-^2$. Let $\widehat{c}_n(\theta, 1-\alpha)$ denote the GMS critical value defined in Andrews and Soares (2010) using the standard GMS function $\varphi_j(\xi,\Omega) = \infty 1(\xi_j > 1)$ for $j \le k$, where $\infty \cdot 0 := 0$ by definition, and constant $\kappa_n (\ln n)^{1/2}$.

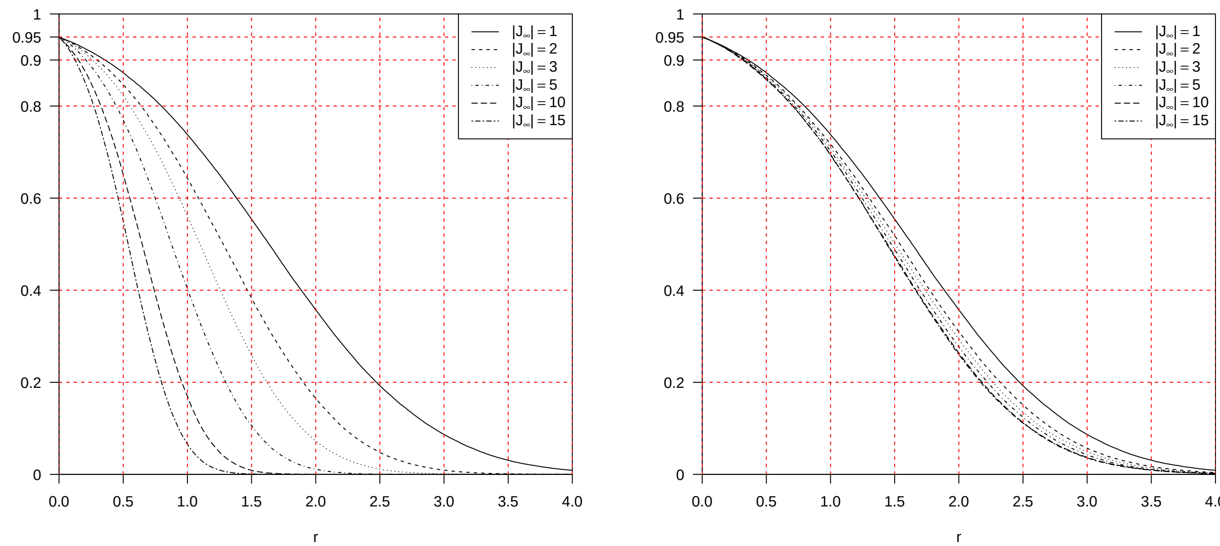
We consider a set $\mathcal{P}_n$ of distributions F for which one or more moment inequalities are violated by at least $r/n^{1/2}$, the other moment inequalities are slack by at least $d_n/n^{1/2}$ for all $\theta \in \Theta$, where $d_n\kappa_n^{-1} \to \infty$, and the correlation matrices of the moment functions are restricted to lie in some set Ψ. Lemma 17.1 in the Supplemental Material provides an expression for the maximum asymptotic coverage probability for any $\theta \in \Theta$ for the GMS CS under $\{\mathcal{P}_n\}_{n\ge1}$ misspecification, which is

$$MaxCPM(r;\Omega_\infty, J_\infty) := \limsup_{n\to\infty} \sup_{(\theta,F)\in\Theta\times\mathcal{P}_n} P_F\left(S_{n,Std}(\theta) \le \widehat{c}_n(\theta, 1-\alpha)\right), \tag{3.2}$$

where r indexes the magnitude of misspecification, Ω_∞ is the asymptotically most favorable correlation matrix of the moment functions in Ψ, and J_∞ is the asymptotically most favorable set of indices j of the moment functions that are misspecified. This lemma is proved using results in Bugni, Canay, and Guggenberger (2012). For example, if $MaxCPM(r;\Omega_\infty, J_\infty) = .70$ for a nominal 95% CS, then the asymptotic coverage probability for any potential pseudo-true value is at most .70, which indicates spurious precision of the CS.

Figure 3.1 graphs $MaxCPM(r;\Omega_\infty, J_\infty)$ as a function of r for the nominal 95% standard GMS CS described above. Two correlation matrices Ω_∞ are considered in which all correlations are equal to ρ for $\rho = .00$ and .75 in the two cases. For these correlation matrices, $MaxCPM(r;\Omega_\infty, J_\infty)$ depends on J_∞ only through $J_\#$, which denotes the number of elements of J_∞. Figure 3.1 considers $J_\# = 1, 2, 3, 5, 10, 15$. Note that for F in $\mathcal{P}_n$ the magnitude of misspecification for the violated

[4] It has been shown that subsampling provides correct asymptotic size of tests and CS's for the true parameter based on a class of *non-recentered* test statistics for correctly-specified moment inequality models, see Romano and Shaikh (2008) and Andrews and Guggenberger (2009). However, research on subsampling done subsequently to the publication of CHT shows that in many non-regular circumstances subsampling fails to deliver correct asymptotic size, see Andrews and Guggenberger (2010). Given the form of recentered test statistics, the potential pitfalls of subsampling are a definite concern. For recentered test statistics, it is an open question whether subsampling provides correct (uniform) asymptotic size under misspecification or even under correct specification. The answer may depend on the specific form of the moment functions. It also may depend on whether inference is for the "true" parameter or for the identified set.



(a) $\rho = 0$ (b) $\rho = .75$

Figure 3.1: Maximum Asymptotic Coverage Probabilities for any $\theta \in \Theta$ for a Standard GMS 95% Confidence Set under Model Misspecification Indexed by r: Test Function $S_1(\cdot)$, $J_{\#} = 1, 2, \cdots, 15$, and (a) $\rho = 0$ and (b) $\rho = .75$

moment functions is $E_F \widetilde{m}_j(W_i, \theta) \leq -r/n^{1/2}$, which decreases in absolute value at a $1/n^{1/2}$ rate. Hence, for a given fixed (independent of n) magnitude of misspecification, say $c < 0$, the value of r in Figure 3.1 that is relevant depends on n and equals $|n^{1/2}c|$. This implies that the effect of spurious precision due to misspecification increases significantly with the sample size.

Figure 3.1 shows substantial under-coverage of any parameter value due to model misspecification unless r is very close to zero. For example, for $r = 1$, the maximum coverage probability is .75 or less across the different scenarios considered.

For $\rho = .00$, the maximum coverage probability decreases noticeably with increases in the number $J_{\#}$ of moment conditions that are violated. For $\rho = .75$, the decrease is much less (because there is less incremental information from a additional moment condition that is highly correlated with other moment functions than when it is independent. Section 17 of the Supplemental Material provides figures for different $S(m, \Omega)$ functions and additional values of ρ.

4 SPUR Tests and Confidence Sets

In this section, 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, so both the SPUR1 and SPUR2 tests are defined here. The critical value of the SPUR1 test is complicated because of nuisance parameter functions that arise in the asymptotic distribution of the test statistic considered. To be succinct in the section and for greater clarity, we define the critical value of the SPUR1 test in Section 6 below after the asymptotic null distribution of the test statistic is provided.

4.1 SPUR1 Tests and CS's

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

$$\widehat{r}_{nj}(\theta) := [\widehat{m}_{nj}(\theta)]_- \, , \ \widehat{r}_n(\theta) := \max_{j\leq k} \widehat{r}_{nj}(\theta), \text{ and } \widehat{r}_n^{\inf} := \inf_{\theta\in\Theta} \widehat{r}_n(\theta). \tag{4.1}$$

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

$$S_n(\theta) := S\left(n^{1/2}\left(\widehat{m}_n(\theta) + \widehat{r}_n^{\inf} 1_k\right), \widehat{\Omega}_n(\theta)\right). \tag{4.2}$$

where $S(m,\Omega)$ is a test function that satisfies Assumptions S.1–S.4, which for brevity are stated in the Appendix. Examples of $S(m,\Omega)$ functions that satisfy these assumptions include

$$S_1(m,\Omega) := \sum_{j=1}^{k} [m_j]_-^2, \ S_2(m,\Omega) := \inf_{t\in R_{+,\infty}^k} (m-t)'\Omega^{-1}(m-t), \ S_4(m,\Omega) := \max_{j\leq k}[m_j]_-, \tag{4.3}$$

and $S_{2A}(m,\Omega)$ defined in Andrews and Barwick (2012), provided $\inf_{\Omega\in\Psi} \det(\Omega) > 0$ for $S_2(\cdot)$.

Let $\widehat{c}_{n,EGMS}(\theta, 1-\alpha)$ denote the $1-\alpha$ conditional quantile of a bootstrap statistic $S_{n,EGMS}^*(\theta)$ given $\{W_i\}_{i\leq n}$ for $\alpha \in (0,1)$. We refer to $\widehat{c}_{n,EGMS}(\theta, 1-\alpha)$ 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 definition of the bootstrap statistic $S_{n,EGMS}^*(\theta)$ is based on the asymptotic null distribution of the test statistic. It is complicated because the asymptotic null distribution 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, the 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 that ensures that the distribution of the bootstrap statistic $S_{n,EGMS}^*(\theta)$ is asymptotically as large as that of the asymptotic null distribution in a stochastic sense. For clarity, we define $S_{n,EGMS}^*(\theta)$ in Section 6 below after the asymptotic distribution of the test statistic has been given. The EGMS critical value can be computed by simulation.

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

$$\phi_{n,SPUR1}(\theta_0) = 1, \text{ where } \phi_{n,SPUR1}(\theta) := 1(S_n(\theta) > \widehat{c}_{n,EGMS}(\theta, 1-\alpha)). \tag{4.4}$$

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

$$CS_{n,SPUR1} := \{\theta \in \Theta : \phi_{n,SPUR1}(\theta) = 0\}. \tag{4.5}$$

An alternative to the SPUR test statistic in (4.2) is a recentered test statistic, such as considered in 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, e.g., as in (3.1). When the function S employed by the SPUR test statistic $S_n(\theta)$ defined in (4.2) is the "max" S_4 statistic, see (4.3), the recentered statistic $S_{4n,Recen}(\theta)$ is identical to the $S_{4n}(\theta)$ SPUR statistic, see Lemma 13.1 in the Supplemental Material.

4.2 Adaptive SPUR2 Tests and CS's

Now, we introduce our recommended test. It is an adaptive 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. These two tests are combined using a CI for $r_F^{\inf}$ that is introduced in Andrews and Kwon (2019). The test is adaptive in the sense that if the CI for $r_F^{\inf}$ contains only the single point 0, so the data indicate that the model is correctly specified, 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 is the SPUR1 test with level α_2, rather than α.. This test is robust to spurious precision caused by misspecification. The correct asymptotic size of this test relies on a Bonferroni argument. Simulations show that this test has good power properties relative to the SPUR1 test, see Section 8 below. 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 those scenarios.

Let $\alpha = \alpha_1 + \alpha_2 \in (0,1)$ for $\alpha_1, \alpha_2 > 0$, such as $\alpha_1 = .005$ and $\alpha_2 = .045$. The nominal $1-\alpha_1$ one-sided upper-bound CI for $r_F^{\inf}$ is

$$CI_{n,r,UP}(\alpha) := [0, \widehat{r}_{n,UP}(\alpha)]. \tag{4.6}$$

This CI equals $\{0\}$ wp$\rightarrow$1 when the model is correctly specified and the sequence of MR identified sets $\{\Theta_I(F_n)\}_{n\geq 1}$ contains slack points with slackness of order greater than $n^{-1/2}$. That is,

$\lim_{n\to\infty}\min_{j\leq k} n^{1/2}E_F m_j(W,\theta_n^I) = \infty$ 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→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→1. See Section 14 in the Supplemental Material for the definition of $CI_{n,r,UP}(\alpha)$ and these results.

Note that $\widehat{r}_{n,UP}(\alpha)$ is not based on $\widehat{r}_n^{\inf}$. Rather, it is based on a 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 properties of $CI_{n,r,UP}(\alpha)$ 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 $S_{n,Std}(\theta)$ defined in (3.1) and a GMS critical value $\widehat{c}_{n,GMS}(\theta, 1-\alpha_2)$, which is the $1-\alpha_2$ conditional quantile of $S_{n,GMS}^*(\theta)$ given $\{W_i\}_{i\leq n}$. By definition, $S_{n,GMS}^*(\theta) := S(T_{n,GMS}^*(\theta), \widehat{\Omega}_n(\theta))$, where $T_{n,GMS}^*(\theta)$ is defined as $T_{n,EGMS}^*(\theta)$ is defined as in (6.4) below, but with $sd_{1nj}^*(\theta)$ and $\widehat{r}_n(\theta)$ replaced by 1 and 0, respectively, in the definition of $\xi_{nj}(\theta)$.

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

$$\begin{aligned}\phi_{n,SPUR2}(\theta_0) := {} & 1(\widehat{r}_{n,UP}(\alpha_1) = 0)\phi_{n,GMS}(\theta_0,\alpha_2) \\ & +1(\widehat{r}_{n,UP}(\alpha_1) > 0)\min\{\phi_{n,SPUR1}(\theta_0,\alpha_2), \phi_{n,GMS}(\theta_0,\alpha_2)\}, \end{aligned} \tag{4.7}$$

where $\phi_{n,SPUR1}(\theta_0,\alpha_2)$ denotes the SPUR1 test of $H_0 : \theta \in \Theta_I(F)$ defined in (4.4), but with α_2 in place of α.[5]

The nominal level $1-\alpha$ SPUR2 CI for $\theta \in \Theta_I(F)$ is

$$CS_{n,SPUR2} := \{\theta \in \Theta : \phi_{n,SPUR2}(\theta) = 0\}. \tag{4.8}$$

By the properties of $CI_{n,r,UP}(\alpha_1)$ described above, the level α SPUR2 test has the same power properties as a level α_2 standard GMS test that is designed for correct model specification when the model is correctly specified and the identified set contains slack points θ for which the slackness of the inequalities is of order greater than $n^{-1/2}$. And, it has the same power properties as the level α_2 SPUR1 test under "large-local" or "global" model misspecification. Finite-sample simulations corroborate these asymptotic results.

We note that the SPUR2 test and CS also can be constructed using any test that has correct asymptotic size under correct model specification, such as the test in Romano, Shaikh, and Wolf (2014), not just the GMS test.

[5]Typically, the $\min\{\cdot,\cdot\}$ term in (4.7) equals $\phi_{n,SPUR2}(\theta_0,\alpha_2)$ with probability close to one.

5 Asymptotic Distribution of the SPUR Test Statistic

The EGMS critical value for the SPUR1 test defined above is constructed based on the asymptotic distribution of $S_n(\theta_0)$ under drifting sequences of null distributions $\{F_n\}_{n\geq 1}$ for which $\theta_0 \in \Theta_I(F_n)$ for $n \geq 1$. In this section, we establish this asymptotic distribution. For power properties, we also establish the asymptotic distribution under local and global alternatives as well.

One obtains a CS for $\theta \in \Theta_I(F)$ by inverting tests based on $S_n(\theta_0)$, as in (2.9). To obtain uniform asymptotic coverage probability results, we need the asymptotic distribution of $S_n(\theta_n)$ under drifting sequences of null values $\{\theta_n\}_{n\geq 1}$ and distributions $\{F_n\}_{n\geq 1}$. For this reason, in the results below, we consider the statistic

$$S_n := S_n(\theta_n) \text{ for testing } H_0 : \theta_n \in \Theta_I(F_n). \tag{5.1}$$

The results cover models that may be correctly specified or misspecified. Note that the form of the asymptotic null distribution is important in order to understand the rather complicated definition of the EGMS critical value given in Section 4.1 above.

The proofs of the asymptotic size results for SPUR tests and CS's show that it suffices to determine the asymptotic null rejection probabilities of tests under sequences or subsequences of distributions F_n that satisfy certain conditions. These conditions are Assumptions C.1–C.4, C.7, and C.8 introduced below, which depend only on deterministic quantities and can be made to hold for certain subsequences using the fact that any sequence in a compact metric set has a convergent subsequence. For this reason, we do not provide sufficient conditions for these conditions and these conditions do not appear in the statements of the asymptotic size results. For the asymptotic power results under drifting sequences of distributions given in the Supplemental Material, we employ Assumptions C.1–C.4, C.7, and C.8 as stated.

5.1 High-Level Convergence Assumptions

We write

$$\begin{aligned} S_n(\theta) &= S\left(T_n(\theta) + A_n^{\inf} 1_k, \widehat{\Omega}_n(\theta)\right), \text{ where} \\ T_n(\theta) &:= n^{1/2}\left(\widehat{m}_n(\theta) + r_{F_n}^{\inf} 1_k\right), \\ A_n^{\inf} &:= n^{1/2}\left(\widehat{r}_n^{\inf} - r_{F_n}^{\inf}\right), \end{aligned} \tag{5.2}$$

and $T_n(\theta) = (T_{n1}(\theta), ..., T_{nk}(\theta))'$. The components $T_n(\theta)$ and $A_n^{\inf}$ of $S_n(\theta)$ are recentered and rescaled such that they have asymptotic distributions. We obtain the asymptotic distribution of

$A_n^{\inf}$ using a similar approach to that in BCS. The results are also closely related to the asymptotic distribution results for the supremum of a moment inequality objective function in CHT, Theorems 4.2 and 5.2. The results given below differ from these results in that they allow for model misspecification.

As in BCS, for any $x_1, x_2 \in R_{[\pm\infty]}^{a_*}$ for some positive integer a_*, let $d(x_1, x_2) := (\sum_{j=1}^{a_*}(\Phi(x_{1,j}) - \Phi(x_{2,j}))^2)^{1/2}$, where $\Phi : R_{[\pm\infty]} \to [0,1]$, $\Phi(y)$ is the standard normal distribution function at y for $y \in R$, $\Phi(-\infty) := 0$, and $\Phi(\infty) := 1$. The space $(R_{[\pm\infty]}^{a_*}, d)$ is a compact metric space. Convergence in $(R_{[\pm\infty]}^{a_*}, d)$ to a point in R^{a_*} implies convergence under the Euclidean norm. Let $\mathcal{S}(\Theta \times R_{[\pm\infty]}^{2k})$ denote the space of non-empty compact subsets of the metric space $(\Theta \times R_{[\pm\infty]}^{2k}, d)$, where d is defined with $a_* = d_\theta + 2k$. Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes in the sense of van der Vaart and Wellner (1996). Let $\to_H$ denote convergence in Hausdorff distance (under d) for elements of $\mathcal{S}(\Theta \times R_{[\pm\infty]}^{2k})$. For any $b, \ell, m \in R^k$, including $b_n, b^*, \widetilde{b}$, ℓ_n which arise below, let b_j, ℓ_j, m_j denote the jth elements of b, ℓ, m, respectively.

To obtain the asymptotic distribution of $A_n^{\inf}$, we use the following sets:

$$\Lambda_{n,F} := \Big\{(\theta, b, \ell) \in \Theta \times R^{2k} : b_j = n^{1/2}([E_F \widetilde{m}_j(W,\theta)]_- - r_F^{\inf}), \ell_j = n^{1/2} E_F \widetilde{m}_j(W,\theta) \ \forall j \le k.\Big\} \tag{5.3}$$

for $n \ge 1$. For $(\theta, b, \ell) \in \Lambda_{n,F}$, b_j is the difference between the magnitude of violation of the jth moment at θ, $[E_F \widetilde{m}_j(W,\theta)]_-$, and the minimal relaxation, $r_F^{\inf}$, scaled by $n^{1/2}$, and ℓ_j is the jth moment at θ scaled by $n^{1/2}$. The quantities b_j and ℓ_j can be positive, negative, or zero.

For $\eta > 0$, define

$$\Theta_I^\eta(F) := \{\theta \in \Theta : \max_{j \le k}[E_F \widetilde{m}_j(W,\theta) + r_F^{\inf}]_- \le \eta/n^{1/2}\}. \tag{5.4}$$

The set $\Theta_I^\eta(F)$ is an $\eta/n^{1/2}$-expansion of the MR identified set $\Theta_I(F)$. It depends on n, but this is suppressed for simplicity. One can also write $\Theta_I^\eta(F)$ as $\{\theta \in \Theta : \max_{j \le k}[E_F \widetilde{m}_j(W,\theta)]_- - r_F^{\inf} \le \eta/n^{1/2}\}$.[6]

For $\eta > 0$, define Λ_{n,F_n}^η as in (5.3) with $\Theta_I^\eta(F_n)$ in place of Θ. By definition, $\Lambda_{n,F_n}^\eta \subset \Lambda_{n,F_n}$.

We employ the following "convergence" assumptions that apply to a drifting sequence of null values $\{\theta_n\}_{n \ge 1}$, as in (5.1), and distributions $\{F_n\}_{n \ge 1}$.

Assumption C.1. $\theta_n \to \theta_\infty$ for some $\theta_\infty \in \Theta$.

Assumption C.2. $n^{1/2} E_{F_n} \widetilde{m}_j(W, \theta_n) \to \ell_{j\infty}$ for some $\ell_{j\infty} \in R_{[\pm\infty]}$ $\forall j \le k$.

[6] This holds because for $b, c \ge 0$, $[a+b]_- \le c$ if and only if $[a]_- - b \le c$, see (23.77) in the Supplemental Material. The set $\Theta_I^\eta(F)$ in (5.4) equals the set $\Theta_I^\eta(F)$ in BCS—which depends on a function $S(m, \Omega)$—only when the model is correctly specified (i.e., $r_F^{\inf} = 0$) and when BCS's function $S(m,\Omega)$ equals $\max_{j \le k}[m_j]_-$.

Assumption C.3. $n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n)+r_{F_n}^{\inf})\to h_{j\infty}$ for some $h_{j\infty}\in R_{[\pm\infty]}$ $\forall j\leq k$.

Assumption C.4. $\sup_{\theta\in\Theta}||E_{F_n}\widetilde{m}(W,\theta)-\widetilde{m}(\theta)||\to 0$ for some nonrandom bounded continuous R^k-valued function $\widetilde{m}(\cdot)$ on Θ.

Assumption C.5. $\nu_n(\cdot):=(\nu_n^m(\cdot)',\nu_n^\sigma(\cdot)')'\Rightarrow G(\cdot):=(G^m(\cdot)',G^\sigma(\cdot)')'$ as $n\to\infty$, where $\{G(\theta):\theta\in\Theta\}$ is a mean zero R^{2k}-valued Gaussian process with bounded continuous sample paths a.s. and $G^m(\theta),G^\sigma(\theta)\in R^k$.

Assumption C.6. $\widehat{\Omega}_n(\theta_n)\to_p\Omega_\infty$ for some $\Omega_\infty\in\Psi$.

Assumption C.7. $\Lambda_{n,F_n}\to_H\Lambda$ for some non-empty set $\Lambda\in\mathcal{S}(\Theta\times R_{[\pm\infty]}^{2k})$.

Assumption C.8. $\Lambda_{n,F_n}^{\eta_n}\to_H\Lambda_I$ for some non-empty set $\Lambda_I\in\mathcal{S}(\Theta\times R_{[\pm\infty]}^{2k})$, where $\{\eta_n\}_{n\geq 1}$ is a sequence of positive constants for which $\eta_n\to\infty$.

All of the limit quantities above, θ_∞, $\{\ell_{j\infty}\}_{j\leq k}$, etc., depend on $\{\theta_n\}_{n\geq 1}$ and $\{F_n\}_{n\geq 1}$. Lemma 24.1 in the Supplemental Material shows that Assumptions A.1–A.4, C.1, and uniform convergence of the covariance kernel $\Omega_{F_n}(\cdot,\cdot)$ to a continuous limit function $\Omega_\infty(\cdot,\cdot)$ are sufficient conditions for Assumptions C.5 and C.6 for the case of i.i.d. observations. Assumption C.7 is a generalization of assumption (iii) in Theorem 3.1 of BCS to allow for model misspecification. Assumption C.8 is used to simplify the asymptotic distribution of S_n.[7]

Let

$$\widetilde{m}_{j\infty}=\widetilde{m}_j(\theta_\infty)\text{ for }j\leq k\text{ and }\widetilde{m}(\theta)=(\widetilde{m}_1(\theta),...,\widetilde{m}_k(\theta))'. \tag{5.5}$$

The limit values $\ell_{j\infty}$, $h_{j\infty}$, and $\widetilde{m}_{j\infty}$ in Assumptions C.2 and C.3 and (5.5) have the following properties.

Lemma 5.1 (a) *Under Assumption* C.3, *if* $\theta_n\in\Theta_I(F_n)$ *for all* n *large, then* $h_{j\infty}\geq 0$ $\forall j\leq k$, (b) *under Assumptions* C.2 *and* C.3, $\ell_{j\infty}\leq h_{j\infty}$ $\forall j\leq k$, (c) *under Assumptions* C.1, C.2, *and* C.4, $|\widetilde{m}_{j\infty}|\leq|\ell_{j\infty}|$ *and if* $|\ell_{j\infty}|<\infty$, *then* $\widetilde{m}_{j\infty}=0$ $\forall j\leq k$, *and* (d) *under Assumptions* C.1–C.4, *if* $\theta_n\in\Theta_I(F_n)$ *for all* n *large and the model is correctly specified, then* $h_{j\infty}=\ell_{j\infty}$ *and* $h_{j\infty},\ell_{j\infty},\widetilde{m}_{j\infty}\geq 0$ $\forall j\leq k$.

Comment. By Lemma 5.1(a), under the null hypothesis H_0 in (2.8), $h_{j\infty}\geq 0$ $\forall j\leq k$.

The elements (θ,b,ℓ) of Λ in Assumption C.7 have the following properties.

[7] BCS use a sequence $\{\eta_n\}_{n\geq 1}$ as in Assumption C.8 and a sequence $\{\kappa_n\}_{n\geq 1}$ that enters their GMS procedure. Their results hold for $\eta_n=\ln\kappa_n$, where $\kappa_n\to\infty$ and $\kappa_n/n^{1/2}\to 0$. In contrast, in our results, $\{\eta_n\}_{n\geq 1}$ in Assumption C.8 and the sequence $\{\kappa_n\}_{n\geq 1}$ that enters our EGMS procedure are unrelated.

Lemma 5.2 *Under* $\{F_n\}_{n\geq 1}$, (a) $\max_{j\leq k} b_{nj}(\theta) \geq 0$ $\forall\theta \in \Theta$, $\forall n \geq 1$, *where* $b_{nj}(\theta) := n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta)]_- - r_{F_n}^{\inf})$, (b) $\forall(\theta, b, \ell) \in \Lambda$, $\max_{j\leq k} b_j \geq 0$ *provided Assumption* C.7 *holds,* (c) $\exists\widetilde{\theta}_n \in \Theta$ *with* $\max_{j\leq k} b_{nj}(\widetilde{\theta}_n) = 0$ $\forall n \geq 1$ *provided Assumption* A.0 *holds,* (d) $\exists(\widetilde{\theta}, \widetilde{b}, \widetilde{\ell}) \in \Lambda$ *with* $\max_{j\leq k} \widetilde{b}_j = 0$ *provided Assumptions* A.0 *and* C.7 *hold, and* (e) $\forall(\theta, b, \ell) \in \Lambda$, $|\ell_j| < \infty$ *implies* $\widetilde{m}_j(\theta) = 0$ $\forall j \leq k$ *provided Assumptions* C.4 *and* C.7 *hold.*

Comment. Lemma 5.2(a)–(d) is used to show that the asymptotic distribution of $A_n^{\inf}$ is in R a.s. Lemma 5.2(a) and (b) are key properties that are utilized when constructing a stochastic lower bound on the asymptotic distribution of $A_n^{\inf}$. Lemma 5.2(c) implies that the MR identified set is non-empty under Assumption A.0 for all $n \geq 1$. Lemma 5.2(e) is used to show that the asymptotic distribution of $A_n^{\inf}$ simplifies somewhat in some scenarios.

Next, we state assumptions that specify whether $\{\theta_n\}_{n\geq 1}$ is a sequence of parameter values (i) in the MR identified set or $n^{-1/2}$-local to the MR identified set, i.e., a null or $n^{-1/2}$-local alternative (NLA) sequence, or (ii) non-$n^{-1/2}$-local to the MR identified set, which yields a consistent alternative (CA) sequence.

Assumption NLA. $\min_{j\leq k} h_{j\infty} > -\infty$.

Assumption CA. $\min_{j\leq k} h_{j\infty} = -\infty$.

Assumption N. $\theta_n \in \Theta_I(F_n)$ $\forall n \geq 1$.

Assumption N implies Assumption NLA. Assumption NLA also covers $n^{-1/2}$-local alternatives, see Assumption LA in the Appendix. A sufficient condition for Assumption CA is that $(\theta_n, F_n) = (\theta_*, F_*)$ does not depend on n and $E_{F_*}\widetilde{m}_j(W, \theta_*) + r_{F_*}^{\inf} < 0$ for some $j \leq k$, which is Assumption FA in the Appendix. See Lemma 20.1 in the Supplemental Material for these results.

5.2 Asymptotic Distribution of $\mathbf{S_n}$

For notational simplicity, we use the following conventions: for any scalars $\nu \in R$ and $c = \pm\infty$, where ν may be deterministic or random and c is deterministic, we let

$$\nu + c = c,\ [\nu + c]_- - [c]_- = 0 \text{ when } c = +\infty, \text{ and } [\nu + c]_- - [c]_- = -\nu \text{ when } c = -\infty.^8 \quad (5.6)$$

Let $G_j^m(\theta)$, $G_j^\sigma(\theta)$, $\nu_{nj}^m(\theta)$, and $\nu_{nj}^\sigma(\theta)$ denote the jth elements of $G^m(\theta)$, $G^\sigma(\theta)$, $\nu_n^m(\theta)$, and

[8] This notation is motivated by the fact that for finite deterministic scalar constants ν and c, for ν fixed, $\lim_{c\to\pm\infty}(\nu + c) = \lim_{c\to\pm\infty} c$, $\lim_{c\to+\infty}([\nu + c]_- - [c]_-]) = 0$, and $\lim_{c\to-\infty}([\nu + c]_- - [c]_-) = -\nu$, and analogous convergence in probability results hold when ν is random.

$\nu_n^\sigma(\theta)$, respectively. Let

$$\begin{aligned} G_{j\infty}^m &:= G_j^m(\theta_\infty),\ G_{j\infty}^\sigma := G_j^\sigma(\theta_\infty),\ G_{j\infty}^{m\sigma} := G_{j\infty}^m - \frac{1}{2}\widetilde{m}_{j\infty}G_{j\infty}^\sigma, \\ G_j^{m\sigma}(\theta) &:= G_j^m(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)G_j^\sigma(\theta),\ \text{and} \\ \nu_{nj}^{m\sigma}(\theta) &:= \nu_{nj}^m(\theta) - \frac{1}{2}\widetilde{m}_j(\theta)\nu_{nj}^\sigma(\theta) \end{aligned} \tag{5.7}$$

for $j \leq k$ and θ_∞ as in Assumption C.1. Define

$$T_{j\infty} := G_{j_\infty}^{m\sigma} + h_{j\infty} \text{ for } j \leq k \text{ and } T_\infty := (T_{1\infty}, ..., T_{k\infty})', \tag{5.8}$$

where we employ the notational convention in (5.6). Thus, we have: $T_{j\infty} = \infty$ if $\ell_{j\infty} = \infty$ (because $h_{j\infty} \geq \ell_{j\infty} = \infty$ by Lemma 5.1(c)), $T_{j\infty} = G_{j_\infty}^m + h_{j\infty}$ if $|\ell_{j\infty}| < \infty$ (because $|\ell_{j\infty}| < \infty$ implies that $\widetilde{m}_{j\infty} = 0$ by Lemma 5.1(c)), and $T_{j\infty}$ is finite and as in (5.8) with $\widetilde{m}_{j\infty} \neq 0$ if $\ell_{j\infty} = -\infty$ and $|h_{j\infty}| < \infty$. As noted above, under H_0, $h_{j\infty} \geq 0$ for $j \leq k$.

If the model is correctly specified and $\theta_n \in \Theta_I(F_n)$ for n large, then T_{j_∞} simplifies to

$$T_{j_\infty} = G_{j_\infty}^m + \ell_{j\infty} \tag{5.9}$$

because, in this case, $h_{j\infty} = \ell_{j\infty}$ (by Lemma 5.1(d)), $\ell_{j\infty} \in [-\infty, 0)$ cannot occur (because $\ell_{j\infty} \geq 0$ by Lemma 5.1(d)), $|\ell_{j\infty}| < \infty$ implies that $\widetilde{m}_{j\infty} = 0$ (by Lemma 5.1(c)), and $\ell_{j\infty}$ $(= h_{j\infty}) = \infty$ implies $G_{j_\infty}^m - (\widetilde{m}_{j\infty}/2)G_{j_\infty}^\sigma + h_{j\infty} = \infty = G_{j_\infty}^m + \ell_{j\infty}$ (by the notational convention in (5.6)).

The following quantities arise with the asymptotic distribution of $A_n^{\inf}$:

$$\begin{aligned} A_n^{\inf}(\Lambda_{n,F_n}) &:= \inf_{(\theta,b,\ell)\in\Lambda_{n,F_n}} \max_{j\leq k} \left([\nu_{nj}^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_- + b_j\right) \text{ and} \\ A_\infty^{\inf}(\Lambda) &:= \inf_{(\theta,b,\ell)\in\Lambda} \max_{j\leq k} \left([G_j^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_- + b_j\right). \end{aligned} \tag{5.10}$$

We show that $A_n^{\inf} = A_n^{\inf}(\Lambda_{n,F_n}) + o_p(1) \to_d A_\infty^{\inf}(\Lambda)$ as $n \to \infty$ in Lemma 21.1 in the Supplemental Material and Theorem 5.3 below. The term in parentheses in the definition of $A_\infty^{\inf}(\Lambda)$ equals b_j when $\ell_j = +\infty$ (because $[\nu + c]_- - [c]_- = 0$ for $\nu \in R$ and $c = +\infty$ by definition in (5.6)); equals $[G_j^m(\theta) + \ell_j]_- - [\ell_j]_- + b_j$ when $|\ell_j| < \infty$ (because $|\ell_j| < \infty$ implies $\widetilde{m}_j(\theta) = 0$ for $(\theta, b, \ell) \in \Lambda$ by Lemma 5.2(e)); and equals $-G_j^{m\sigma}(\theta) + b_j$ when $\ell_j = -\infty$ (because $[\nu + c]_- - [c]_- = -\nu$ for $\nu \in R$ and $c = -\infty$ by definition in (5.6)).

The asymptotic distribution of the SPUR statistic S_n under the null hypothesis and $n^{-1/2}$-local alternatives is the distribution of

$$S_\infty := S(T_\infty + A_\infty^{\inf}(\Lambda)1_k, \Omega_\infty), \text{ which is equal to } S_{I\infty} := S(T_\infty + A_\infty^{\inf}(\Lambda_I)1_k, \Omega_\infty) \qquad (5.11)$$

under Assumption C.8.

Theorem 5.3 (a) *Under* $\{F_n\}_{n\geq 1}$ *and Assumptions* C.1–C.5, $T_n(\theta_n) \to_d T_\infty$,

(b) *under* $\{F_n\}_{n\geq 1}$ *and Assumptions* A.0, C.4, C.5, *and* C.7, $A_n^{\inf} \to_d A_\infty^{\inf}(\Lambda)$,

(c) *under Assumptions* A.0 *and* C.7, $A_\infty^{\inf}(\Lambda) \in R$ *a.s.,*

(d) *under Assumptions* C.1–C.5 *and* NLA*,* $T_{j\infty} > -\infty$ *a.s.* $\forall j \leq k$,

(e) *under* $\{F_n\}_{n\geq 1}$ *and Assumptions* A.0, C.1–C.7, NLA, *and* S.1(iii), $S_n \to_d S_\infty$,

(f) *under Assumptions* A.0, C.1, *and* C.3–C.8, $A_\infty^{\inf}(\Lambda) = A_\infty^{\inf}(\Lambda_I)$ *a.s. and* $S_\infty = S_{I\infty}$ *a.s.,*

(g) *under Assumptions* C.1–C.5, *and* CA, $T_{j\infty} = -\infty$ *a.s. for some* $j \leq k$,

(h) *under* $\{F_n\}_{n\geq 1}$ *and Assumptions* A.0, C.1–C.7, CA, S.1(iii), S.2, *and* S.3, $S_n \to_p \infty$, *and*

(i) *the convergence results in parts* (a)–(e) *hold jointly.*

Comments. (i). Under correct model specification, $r_F^{\inf} = 0$, $A_n^{\inf} = n^{1/2}\widehat{r}_n^{\inf}$ (see (5.2)), $n^{1/2}\widehat{r}_n^{\inf}$ is the same as the model specification test statistic in BCS when their function $S(m, \Omega)$ equals $\max_{j\leq k}[m_j]_-$, and the asymptotic distribution of $A_n^{\inf}$ given in Theorem 5.3(b) can be shown to reduce to the same distribution as the asymptotic null distribution of the specification test statistic given in Theorem 3.1 of BCS. In addition, in the correctly specified case, $A_n^{\inf} = n^{1/2}\widehat{r}_n^{\inf}$ equals CHT's statistic $\inf_{\theta\in\Theta} a_n Q_n(\theta)$ for moment inequality models when $Q_n(\theta)$ is the "max" sample objective function defined by $\max_{j\leq k}[\widehat{m}_{nj}(\theta)]_-$ (and $a_n = n^{1/2}$) and CHT provide the asymptotic distribution of $\inf_{\theta\in\Theta} a_n Q_n(\theta)$ under correct specification and for a fixed true distribution (rather than a drifting sequence of distributions as in Theorem 5.3(b)).[9] Theorem 5.3(b) extends these results to allow for model misspecification.

(ii). The asymptotic distributions in Theorem 5.3 depend on the localization parameters $h_{j\infty}$ and $\ell_{j\infty}$, which are not consistently estimable, and $\widetilde{m}_{j\infty}$, which is consistently estimable. Under the null hypothesis H_0 in (2.8), $h_{j\infty} \geq 0$ for all $j \leq k$. The asymptotic distribution also depends on the (b_j, ℓ_j) values, which appear in the limit sets Λ and Λ_I, and are not consistently estimable.

[9] The asymptotic distribution of Chernozhukov, Hong, and Tamer's (2007) statistic $\inf_{\theta\in\Theta} a_n Q_n(\theta)$ is given in their Theorems 4.2(2) and 5.2(2) by the difference between $\mathcal{C}$ in their (4.8) and (4.7) or the difference between $\mathcal{C}(\theta)$ in their (5.6) and (5.5). Their definition of the identified set on p. 1265 assumes correct model specification, as do their equation (4.5) and Assumption M.2. The function $\xi(\theta)$ in their Theorem 4.2 only takes values of $-\infty$ or 0 due to their asymptotics being for a fixed true distribution, as opposed to a drifting sequence of distributions. Because Chernozhukov, Hong, and Tamer (2007) consider "$\leq$" inequalities, whereas the present paper considers "$\geq$" inequalities, the sample moments enter the statistics with different signs in the two papers.

(iii). Theorem 5.3(c) is important because it implies that adding $A_\infty^{\inf}(\Lambda)$ to $T_{j\infty}$ cannot result in adding $+\infty$ to $-\infty$ or $-\infty$ to $+\infty$.

(iv). Theorem 5.3(f) is important because it implies that parameters $(\theta, b, \ell) \in \Lambda\backslash\Lambda_I$ do not contribute to the infimum in $A_\infty^{\inf}(\Lambda)$. This means that when constructing a critical value for a test based on S_n one only needs to find a lower bound on $A_\infty^{\inf}(\Lambda_I)$.

(v). The stochastic process $G_j^\sigma(\cdot)$ enters S_∞ (through $G_j^{m\sigma}(\cdot)$). Thus, the asymptotic distribution of S_n depends on the randomness due to the estimation of the standard deviation of the jth sample moment by $\widehat{\sigma}_{nj}(\theta)$ for $j \leq k$. Under correct model specification, this is not the case.

(vi). For any subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$, Theorem 5.3 and its proof hold with q_n in place of n throughout, including the assumptions.

(vii). To prove Theorem 5.3(b), we use a similar proof to the proof of Theorem 3.1 of BCS with $S(m, \Omega) = \max_{j\leq k}[m_j]_-$ in their proof. The statistic $A_n^{\inf}(\Lambda_{n,F_n})$ depends on $b_{nj}(\theta) := n^{1/2}[E_{F_n}\widetilde{m}_j(W,\theta)]_- - n^{1/2}r_{F_n}^{\inf}$, $\ell_{nj}(\theta) := n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta)$, $\nu_{nj}^m(\theta)$, and $\nu_{nj}^\sigma(\theta)$, whereas the statistic in BCS depends on $\ell_{nj}(\theta)$ and $\nu_{nj}^m(\theta)$.

(viii). The proof of Theorem 5.3(b) uses the fact that the function $\chi(\nu, c) := [\nu + c]_- - [c]_-$ for $\nu, c \in R$ satisfies $|\chi(\nu, c)| \leq |\nu|$ (see (21.7) in the Supplemental Material), which implies that $\sup_{\theta\in\Theta}\sup_{\ell_j\in R}\left|[\nu_{nj}^{m\sigma}(\theta) + \ell_j]_- - [\ell_j]_-\right| = O_p(1)$.

(ix). For the purposes of inference (i.e., obtaining a critical value), one needs a stochastic lower bound on the distribution of the vector sum $T_\infty + A_\infty^{\inf}1_k$ for the case when $h_{j\infty} \geq 0$ for all $j \leq k$.

6 EGMS Bootstrap Critical Value

The bootstrap statistic $S_{n,EGMS}^*(\theta)$, which is used to construct the critical value for the SPUR1 test in Section 4.1, is defined as follows. Let $\{W_i^*\}_{i\leq n}$ be an i.i.d. bootstrap sample drawn with replacement from the original sample $\{W_i\}_{i\leq n}$. Define

$$\begin{aligned}
\overline{m}_{nj}^*(\theta) &:= n^{-1}\sum_{i=1}^n m_j(W_i^*, \theta),\\
\widehat{\sigma}_{nj}^{*2}(\theta) &:= n^{-1}\sum_{i=1}^n (m_j(W_i^*, \theta) - \overline{m}_{nj}^*(\theta))^2 \ \forall j \leq k,\\
\widehat{\nu}_{nj}(\theta) &:= n^{1/2}\left(\widehat{m}_{nj}(\theta) - E_{F_n}\widetilde{m}_j(W,\theta)\right) \text{ and}\\
\widehat{\nu}_{nj}^*(\theta) &:= n^{1/2}\left(\frac{\overline{m}_{nj}^*(\theta)}{\widehat{\sigma}_{nj}^*(\theta)} - \widehat{m}_{nj}(\theta)\right) \ \forall j \leq k.^{10} \qquad (6.1)
\end{aligned}$$

Let $Var^*(\cdot)$ denote the bootstrap variance based on the nonparametric i.i.d. bootstrap in (6.1).

By Theorem 5.3 and the definition of $\Lambda_{n,F}$ in (5.3), the asymptotic distribution of the test statistic S_n depends on the limits of the following quantities that cannot be consistently estimated:

$$\begin{aligned} h_{nj} &:= n^{1/2}(E_{F_n}\widetilde{m}_j(W,\theta_n) + r_{F_n}^{\inf}), \\ b_{nj}(\theta) &:= n^{1/2}([E_{F_n}\widetilde{m}_j(W,\theta))]_- - r_{F_n}^{\inf}), \text{ and} \\ \ell_{nj}(\theta) &:= n^{1/2}E_{F_n}\widetilde{m}_j(W,\theta) \end{aligned} \tag{6.2}$$

for $\theta \in \Theta_I^{\eta_n}(F_n)$, where θ_n is the null value. GMS methods in the literature are concerned with the behavior of $\ell_{nj}(\theta_n)$. But here, we need methods that apply to h_{nj}, $b_{nj}(\theta)$, and $\ell_{nj}(\theta)$ for $\theta \in \Theta_I^{\eta_n}(F_n)$. In addition, the set $\Theta_I^{\eta_n}(F_n)$, which is an expansion of the MR identified set, is unknown. This set enters the asymptotic distribution of S_n because its Hausdorff limit, $\Theta_{\Lambda_I} := \{\theta : (\theta, b, \ell) \in \Lambda_I$ for some $b, \ell \in R^k\}$, is part of Λ_I, which arises in Theorem 5.3(f). We estimate $\Theta_I^{\eta_n}(F_n)$ using a set estimator $\widehat{\Theta}_n$ ($\subset \Theta$) that is designed to contain $\Theta_I^{\eta_n}(F_n)$ wp$\to$1 under drifting sequences of distributions $\{F_n\}_{n\geq 1}$.

Define

$$S_{n,EGMS}^*(\theta) := S\left(T_{n,EGMS}^*(\theta) + A_{n,EGMS}^{*\inf}1_k, \widehat{\Omega}_n(\theta)\right). \tag{6.3}$$

where $T_{nj,EGMS}^*(\theta)$ and $A_{n,EGMS}^{*\inf}$ are bootstrap analogues of $T_n(\theta)$ and $A_n^{\inf}$ in (5.2), respectively, with the null hypothesis is imposed in $T_{nj,EGMS}^*(\theta)$. We define

$$\begin{aligned} T_{nj,EGMS}^*(\theta) &:= \widehat{\nu}_{nj}^*(\theta) + \varphi_j(\xi_n(\theta), \widehat{\Omega}_n(\theta)) \ \forall j \leq k, \text{ where} \\ \xi_{nj}(\theta) &:= (sd_{1nj}^*(\theta)\kappa_n)^{-1} n^{1/2}\left(\widehat{m}_{nj}(\theta) + \widehat{r}_n(\theta)\right) \ \forall j \leq k, \ \xi_n(\theta) = (\xi_{n1}(\theta), ..., \xi_{nk}(\theta))', \end{aligned} \tag{6.4}$$

$\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), ..., \varphi_k(\xi, \Omega))'$ is a specified GMS function that satisfies Assumption A.5, which is stated in the Appendix for brevity, with the leading choice being $\varphi_j(\xi, \Omega) = \infty 1(\xi_j > 1)$ for $j \leq k$, where $\infty \cdot 0 := 0$ by definition. In (6.4), κ_n is a positive tuning parameter that satisfies $\kappa_n \to \infty$, which is Assumption A.6 in the Appendix, with the leading choice being $\kappa_n = (\ln n)^{1/2}$, as in Andrews and Soares (2010) and BCS. In (6.4), $sd_{1nj}^*(\theta) := \max\{Var^*(n^{1/2}(\widehat{m}_{nj}(\theta) + \widehat{r}_n(\theta)))^{1/2}, 1\}$ is a bootstrap standard deviation that is used to obtain appropriate scaling. We scale κ_n by $sd_{1nj}^*(\theta)$ because the asymptotic variance of $n^{1/2}\widehat{m}_{nj}(\theta)$ is one under correct specification, so no scaling is typically done with GMS critical values, but here the asymptotic variance of $n^{1/2}(\widehat{m}_{nj}(\theta) + \widehat{r}_n(\theta))$ can be larger, especially under model misspecification. Analogous scaling of certain quantities by $sd_{anj}^*(\theta)$ for $a = 2, ...4$ is employed below. Explicit expressions for these bootstrap quantities are given in Section 15 in the Supplemental Material.

The quantity $\xi_{nj}(\theta_n)$ in (6.4) multiplied by $sd_{1nj}^*(\theta_n)\kappa_n$ equals $n^{1/2}(\widehat{m}_{nj}(\theta_n) + \widehat{r}_n(\theta_n))$, which is an estimator of $h_{nj} := n^{1/2}(E_{F_n}m_j(W,\theta_n) + r_{F_n}^{\inf})$, whose limit $h_{j\infty}$ (see Assumption C.3) appears in the asymptotic null distribution of S_n (see Theorem 5.3(e), (5.8), and (5.11)) and is nonnegative under H_0. Thus, $\xi_{nj}(\theta_n)$ is an estimator of $n^{1/2}(E_{F_n} m_j(W,\theta_n) + r^{\inf}_{F_n})$ that is shrunk towards 0 by $(sd^*_{1nj}(\theta_n)\kappa_n)^{-1}$.

Next, we define $A^{*\inf}_{n,EGMS}$. We employ the following estimator of the expansion $\Theta_I^{\eta_n}(F_n)$ of the MR identified set:

$$\widehat{\Theta}_n := \{\theta \in \Theta : \max_{j \le k}[\widehat{m}_{nj}(\theta) + \widehat{r}^{\inf}_n]_- \le \tau_n/n^{1/2}\}, \tag{6.5}$$

where $\{\tau_n\}_{n\ge1}$ is a sequence of positive constants that satisfies $\tau_n \to \infty$. As with $\{\kappa_n\}_{n\ge1}$, one can employ the BIC choice $\tau_n = (\ln n)^{1/2}$.

The asymptotic distribution of $A^{\inf}_n$ depends on the asymptotic behavior of $[\widehat{\nu}_{nj}(\theta) + \ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_-$. The EGMS bootstrap lower bound version of this quantity, $\widehat{\chi}^*_{nj,EGMS}(\theta)$, is defined as follows. For $\nu \in R$ and $c_1, c_2, c \in R_{[\pm\infty]}$, let

$$\chi(\nu, c_1, c_2) := \begin{cases} \chi(\nu, c_1) & \text{if } \nu \ge 0 \\ \chi(\nu, c_2) & \text{if } \nu < 0, \end{cases} \quad \text{where } \chi(\nu, c) := [\nu + c]_- - [c]_- \tag{6.6}$$

and $\chi(\nu, c)$ is defined for $c = \pm\infty$ as in (5.6). Note that $\chi(\nu, c_1, c_2)$ is continuous on $R \times R^2_{[\pm\infty]}$ under d because $\chi(\nu, c)$ is continuous on $R \times R_{[\pm\infty]}$ under d and $\chi(0, c) = 0$ for all $c \in R_{[\pm\infty]}$. Define

$$\widehat{\chi}^*_{nj,EGMS}(\theta) := \chi\left(\widehat{\nu}^*_{nj}(\theta),\ n^{1/2}\widehat{m}_{nj}(\theta) - sd^*_{2nj}(\theta)\kappa_n,\ n^{1/2}\widehat{m}_{nj}(\theta) + sd^*_{2nj}(\theta)\kappa_n\right), \tag{6.7}$$

where $\{\kappa_n\}_{n\ge1}$ are as in (6.4) and $sd^*_{2nj}(\theta) := \max\{Var^*(n^{1/2}\widehat{m}_{nj}(\theta))^{1/2}, 1\}$ for $j \le k$. Roughly speaking, $\widehat{\chi}^*_{nj,EGMS}(\theta)$ yields a lower bound on $[\widehat{\nu}_{nj}(\theta) + \ell_{nj}(\theta)]_- - [\ell_{nj}(\theta)]_-$ uniformly over $\theta \in \Theta$ wp$\to$1 because the function $\chi(\nu, c) := [\nu + c]_- - [c]_-$ is nondecreasing in c for $\nu \ge 0$, is zero for all c for $\nu = 0$, and is nonincreasing in c for $\nu < 0$, and the distribution of $\widehat{\nu}^*_{nj}(\theta)$ approximates that of $\widehat{\nu}_{nj}(\theta)$, which converges in distribution to $G^{m\sigma}_j(\theta)$.[11]

The EGMS bootstrap version of $A^{\inf}_n$ also requires asymptotic lower bounds on the $b_{nj}(\theta)$ quantities in (6.2). We replace $b_{nj}(\theta)$ by its sample analogue and shift it towards $-\infty$ by a scaled version of the constant κ_n introduced above. Specifically, we replace $b_{nj}(\theta)$ by

$$\widehat{b}_{nj,EGMS}(\theta) := n^{1/2}\left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}^{\inf}_n\right) - sd^*_{3nj}(\theta)\kappa_n, \tag{6.8}$$

where $sd^*_{3nj}(\theta) := \max\{Var^*(n^{1/2}([\widehat{m}_{nj}(\theta)]_- - \widehat{r}^{\inf}_n))^{1/2}, 1\}$ for $j \le k$ and Θ is replaced by $\widehat{\Theta}_n$ in the $Var^*(\cdot)$ bootstrap version of $\widehat{r}^{\inf}_n$.

[11] The function $\chi(\nu, c)$ satisfies these monotonicity properties because, (i) for $\nu > 0$, $\chi(\nu, c) := -\nu\ (< 0)$ for $c < -\nu$, $\chi(\nu, c) := c\ (< 0)$ for $c \in [-\nu, 0)$, and $\chi(\nu, c) := 0$ for $c \ge 0$, and (ii) for $\nu < 0$, $\chi(\nu, c) := -\nu\ (> 0)$ for $c < 0$, $\chi(\nu, c) := -\nu - c\ (> 0)$ for $c \in [0, \nu)$, and $\chi(\nu, c) := 0$ for $c \ge -\nu$.

Note, however, that the lower bound $\widehat{b}_{nj,EGMS}(\theta)$ does not exploit the key information that $\max_{j\leq k} b_{nj}(\theta) \geq 0$ by Lemma 5.2(a). So, the lower bound $\widehat{b}_{nj,EGMS}(\theta)$ by itself is not adequate—it would yield a critical value that slowly diverges in probability to ∞ as $n \to \infty$.

Let

$$\xi^b_{nj}(\theta) := (sd^*_{3nj}(\theta)\kappa_n)^{-1} n^{1/2} \left([\widehat{m}_{nj}(\theta)]_- - \widehat{r}^{\inf}_n \right) \ \forall j \leq k \text{ and } \xi^b_n(\theta) = (\xi^b_{n1}(\theta), ..., \xi^b_{nk}(\theta))', \quad (6.9)$$

where κ_n and $sd^*_{3nj}(\theta)$ are as above. If j_1 is an index for which $b_{nj_1}(\theta) \geq 0$, of which there is at least one, then we use $\varphi_{j_1}(\xi^b_n(\theta), \widehat{\Omega}_n(\theta))$ as the lower bound on $b_{nj_1}(\theta)$ (for the same GMS-type reasons that motivate the use of $\varphi_j(\xi_n(\theta_n), \widehat{\Omega}_n(\theta_n))$ above.)

The constraint $\max_{j\leq k} b_{nj}(\theta) \geq 0$ given in Lemma 5.2(a) implies that for some $j_1 \leq k$, $b_{nj_1}(\theta) \geq 0$ (where j_1 typically depends on θ). The index j_1 is unknown and cannot be consistently estimated. However, the following sets $\widehat{J}_n(\theta)$ can be shown to contain the value(s) j_1 that maximize $b_{nj}(\theta)$ (for all $\theta \in \Theta$) wp→1:

$$\widehat{J}_n(\theta) := \{j \in \{1, ..., k\} : \widehat{r}_{nj}(\theta) \geq \widehat{r}_n(\theta) - sd^*_{4nj}(\theta) n^{-1/2} \kappa_n\}, \quad (6.10)$$

where $\widehat{r}_{nj}(\theta)$ and $\widehat{r}_n(\theta)$ are defined in (4.1) and $sd^*_{4nj}(\theta) := \max\{Var^*(n^{1/2}(\widehat{r}_{nj}(\theta) - \widehat{r}_n(\theta)))^{1/2}, 1\}$.

We define the EGMS bootstrap version, $A^{*\inf}_{n,EGMS}$, of $A^{\inf}_n$ to be

$$A^{*\inf}_{n,EGMS} := \inf_{\theta \in \widehat{\Theta}_n} \min_{j_1 \in \widehat{J}_n(\theta)} \max_{j \leq k} \left(\widehat{\chi}^*_{nj,EGMS}(\theta) + 1(j \neq j_1) \widehat{b}_{nj,EGMS}(\theta) + 1(j = j_1) \varphi_j(\xi^b_n(\theta), \widehat{\Omega}_n(\theta)) \right). \quad (6.11)$$

The idea behind the definition of $A^{*\inf}_{n,EGMS}$ is as follows. The constraint $\max_{j\leq k} b_{nj}(\theta) \geq 0$ implies that for some $j_1 \in \widehat{J}_n(\theta)$, $b_{nj_1}(\theta) \geq 0$ (wp→1). Since $\widehat{J}_n(\theta)$ is not necessarily a singleton, we allow j_1 to be any of the values in $\widehat{J}_n(\theta)$ and take a minimum over $j_1 \in \widehat{J}_n(\theta)$ to get a lower bound. Under the presumption that j_1 is a value for which $b_{nj_1}(\theta) \geq 0$, we use a lower bound on $b_{nj}(\theta)$ that equals $\varphi_j(\xi^b_n(\theta), \widehat{\Omega}_n(\theta))$ for $j = j_1$ and equals the (typically) smaller value $\widehat{b}_{nj,EGMS}(\theta)$ for $j \neq j_1$. This definition then incorporates the constraint that $\max_{j\leq k} b_{nj}(\theta) \geq 0$.

Given $T^*_{n,EGMS}(\theta)$ and $A^{*\inf}_{n,EGMS}$, the definition of $S^*_{n,EGMS}(\theta)$ in (6.3) is complete and the EGMS bootstrap test statistic $\widehat{c}_{n,EGMS}(\theta, 1-\alpha)$ is the $1-\alpha$ conditional quantile of $S^*_{n,EGMS}(\theta)$ given $\{W_i\}_{i\leq n}$ for $\alpha \in (0,1)$, which can be computed by simulation.

7 Asymptotic Level of the SPUR1 and SPUR2 Tests

Here we show that the SPUR1 and SPUR2 tests and CS's have correct asymptotic level under a set of relatively primitive conditions with i.i.d. observations.

The following theorem uses two assumptions, Assumptions A.7 and A.8, which are stated in the Appendix, for brevity. Assumption A.7 is a continuity condition on the asymptotic distribution S_∞ and is closely related to Assumption A.7 in BCS. Assumption A.8 requires $E_F\widetilde{m}(W,\theta)$ to be equicontinuous on Θ over $F \in \mathcal{P}$, which is not restrictive.

Theorem 7.1 *Under Assumptions* A.0–A.8 *and* S.1, *for* $\alpha \in (0,1)$,

(a) *the nominal level* α SPUR1 *and* SPUR2 *tests of* $H_0 : \theta_0 \in \Theta_I(F)$ *satisfy*

$$\limsup_{n\to\infty} \sup_{F\in\mathcal{P}:\theta_0\in\Theta_I(F)} P_F(\phi_{n,SPUR}(\theta_0) = 1) \leq \alpha \textit{ for } SPUR = SPUR1,\ SPUR2, \textit{ and}$$

(b) *the nominal level* $1-\alpha$ SPUR1 *and* SPUR2 CS's *for* $\theta \in \Theta_I(F)$ *satisfy*

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} \inf_{\theta\in\Theta} P_F(\theta \in CS_{n,SPUR}) \geq 1-\alpha \textit{ for } SPUR = SPUR1,\ SPUR2.$$

Comment. Theorem 7.1 does not require Assumption A.6 of BCS, which is imposed in their main result Theorem 4.1, or its sufficient condition Assumption A.8 of BCS. BCS's Assumption A.8 imposes a minorant condition on the population criterion function that is used to construct their test statistic and convexity of Θ, which could be restrictive. Assumption A.6 (or A.8) of BCS is not needed in Theorem 7.1 because the testing problem here differs from that in BCS.

Asymptotic power results for the SPUR1 and SPUR2 tests are given in Section 12 in the Supplemental Material. These include power for $n^{-1/2}$-local alternatives and consistency for non-$n^{-1/2}$-local alternatives, including fixed alternatives.

8 Simulation Results

In this section, we provide Monte Carlo simulation results that illustrate the performance of the SPUR1 and SPUR2 tests. When the model under consideration is correctly specified, we compare these tests to the standard GMS test. We consider two simple models under various levels of misspecification (i.e., different values of $r^{\inf}_{F_n}$). All simulation results are based on 1,000 simulation repetitions, 500 bootstrap replications, a sample size of $n = 250$, $\kappa_n = \tau_n = (\ln n)^{1/2}$, and $S(\cdot) = S_1(\cdot)$. The GMS function $\varphi(\cdot) = (\varphi_1(\cdot), \ldots, \varphi_k(\cdot))'$ employed is $\varphi_j(\xi, \Omega) = \infty 1(\xi_j > 1)$

for $j \leq k$. The significance level is fixed at $\alpha = .05$ with $\alpha_1 = .005$ and $\alpha_2 = .045$ for the SPUR2 test.

8.1 Lower/Upper Bound Model

First, we consider a simple model where the mean of the data imposes lower and upper bounds on a scalar parameter. The data $\{W_i\}_{i \leq n}$ are i.i.d. with $W_i = (W_{i1}, \ldots, W_{ik})' \sim N(\mu, I_k)$, where $\mu = (\mu_1, \ldots, \mu_k)' \in R^k$ and I_k denotes the $k \times k$ identity matrix. We consider $k = 2$ and 4. The parameter space Θ is taken to be $[-20, 20]$.

For $k = 2$, the population moment inequalities are

$$E_F W_{i1} \leq \theta \,\text{and}\, \theta \leq E_F W_{i2}. \tag{8.1}$$

The model is misspecified if and only if $\mu_1 > \mu_2$; and $r_F^{\inf} = [\mu_1 - \mu_2]_+/2$. For $k = 4$, the moment inequalities are

$$E_F W_{i1} \leq \theta,\ E_F W_{i2} \leq \theta,\ \theta \leq E_F W_{i3},\ \text{and}\, \theta \leq E_F W_{i4}. \tag{8.2}$$

Misspecification arises if and only if $\max\{\mu_1, \mu_2\} > \min\{\mu_3, \mu_4\}$; and $r_F^{\inf} = [\max\{\mu_1, \mu_2\} - \min\{\mu_3, \mu_4\}]_+/2$.

We consider various configurations of μ. Note that when $r_F^{\inf} > 0$, the MR identified set is always a singleton in this model, but it may have different lengths when $r_F^{\inf} = 0$. Accordingly, when $r_F^{\inf} = 0$ we consider configurations that correspond to different lengths of the MR identified set. For $k = 2$, we take $\mu = (r, -r)'$ for each $r \in \{.5, 1, 2, 5\}$ as the misspecified cases. We have $r_F^{\inf} = r$ and $\Theta_I(F) = \{0\}$ in these cases. Figure 8.1 gives the simulated rejection probabilities, i.e., power, of the SPUR1 and SPUR2 tests for a range of null values $\theta_0 \geq 0$ based on these mean vectors.[12] For the correctly-specified cases, we take $\mu = (-\ell, 0)'$ for each $\ell \in \{0, .5, 1, 2\}$. Here the MR identified set is $\Theta_I(F) = [-\ell, 0]$, which has length ℓ. For each value of ℓ, Figure 8.2 provides the simulated rejection probabilities of the SPUR1, SPUR2, and standard GMS tests in these correctly-specified models for fixed $\Theta_I(F) = [-\ell, 0]$ and a range of null hypothesis values $\theta_0 \geq 0$.

For $k = 4$, many different configurations of μ are possible for a given value of $r_F^{\inf} > 0$ or a given length of the MR identified set when $r_F^{\inf} = 0$. Accordingly, we consider several scenarios for $k = 4$. For the misspecified cases, we consider five different scenarios: "binding," "almost binding," "somewhat slack," "very slack," and "slack/almost binding."[13] In each scenario, we consider $r_F^{\inf} =$

[12] That is, Figure 8.1 reports power for the true θ being 0, which is in $\Theta_I(F) = \{0\}$, and the null being $\theta_0 > 0$ for a range of θ_0 values. This differs from, but is no less informative than, a conventional power function that considers a fixed null value and a range of true alternative values.

[13] For given $r > 0$, the mean vectors μ in the five misspecified scenarios are (i) "binding": $\mu = (r, r, -r, -r)'$, (ii)

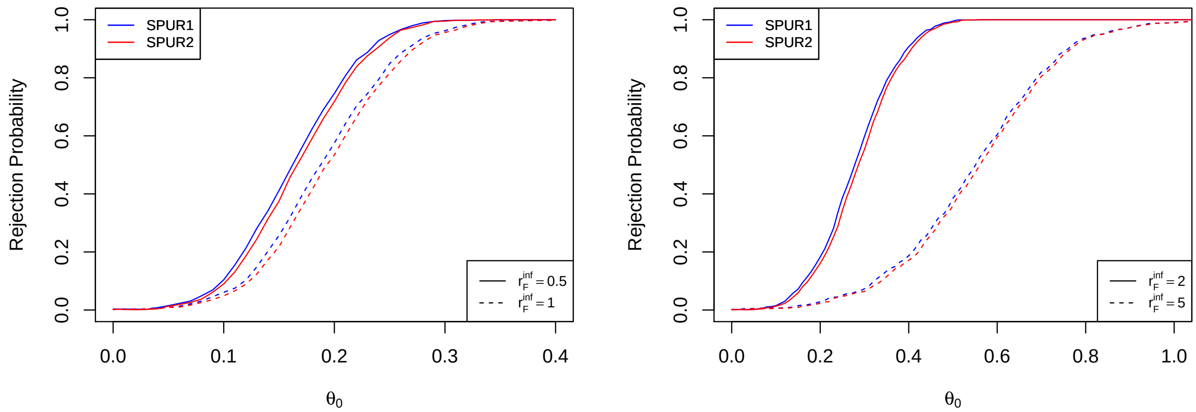


Figure 8.1: Rejection probabilities for misspecified cases for $k = 2$. Each plot shows, for different values of $r_F^{\inf}$, the rejection probabilities of the SPUR1 and SPUR2 tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values and fixed identified set $\Theta_I(F) = \{0\}$.

.5 and 1. Regardless of the scenario and the value of $r_F^{\inf}$, the MR identified set is $\Theta_I(F) = \{0\}$. Figure 8.3 gives the simulation results under the "binding," "almost binding," and "somewhat slack" scenarios. The results for the "very slack" and "slack/almost binding" cases are given in Section 18 of the Supplemental Material.

For the correctly-specified cases and $k = 4$, we consider the same five scenarios as for the misspecified cases. However, the definitions are slightly different in the correctly-specified cases.[14] The MR identified set takes the form $\Theta_I(F) = [-\ell, 0]$ for each $\ell \in \{0, .5, 1\}$. The simulation results for these cases are given in Figure 8.4.

Figures 8.1 and 8.3 show that the performance of the two tests, SPUR1 and SPUR2, is quite similar under misspecification (i.e., $r_F^{\inf} > 0$), which is what we expect given the discussion in Section 4.2. Looking at the rejection probability at $\theta_0 = 0$, we see that both tests have correct size, but under-reject with the null rejection probabilities being close to 0. The rejection probabilities increase to 1 fairly quickly as the distance between the null value and the MR identified set increases. The tests perform better when $r_F^{\inf}$ is smaller, but they perform reasonably well even when $r_F^{\inf}$ is as large as 5, which is five times the standard deviation of the moment functions. Additionally,

"almost binding": $\mu = (r, r - .1, -r + .1, -r)'$, (iii) "somewhat slack": $\mu = (r, r - .5, -r + .5, -r)'$, (iv) "very slack": $\mu = (r, r - 1, -r + 1, -r)'$, and (v) "slack/almost binding": $\mu = (r, r - .1, -r + 1, -r)'$. In each scenario, $r_F^{\inf} = r$ and the identified set is $\Theta_I(F) = \{0\}$.

[14] For given $\ell > 0$, the mean vectors μ in the five correctly-specified scenarios are (i) "binding": $\mu = (-\ell, -\ell, 0, 0)'$, (ii) "almost binding": $\mu = (-\ell - .1, -\ell, 0, .1)'$, (iii) "somewhat slack": $\mu = (-\ell - .5, -\ell, 0, .5)'$, (iv) "very slack": $\mu = (-\ell - 1, -\ell, 0, 1)'$, and (v) "slack/almost binding": $\mu = (-\ell - 1, -\ell, 0, .1)'$. In all scenarios, $\Theta_I(F) = [-\ell, 0]$ and the identified set has length ℓ.

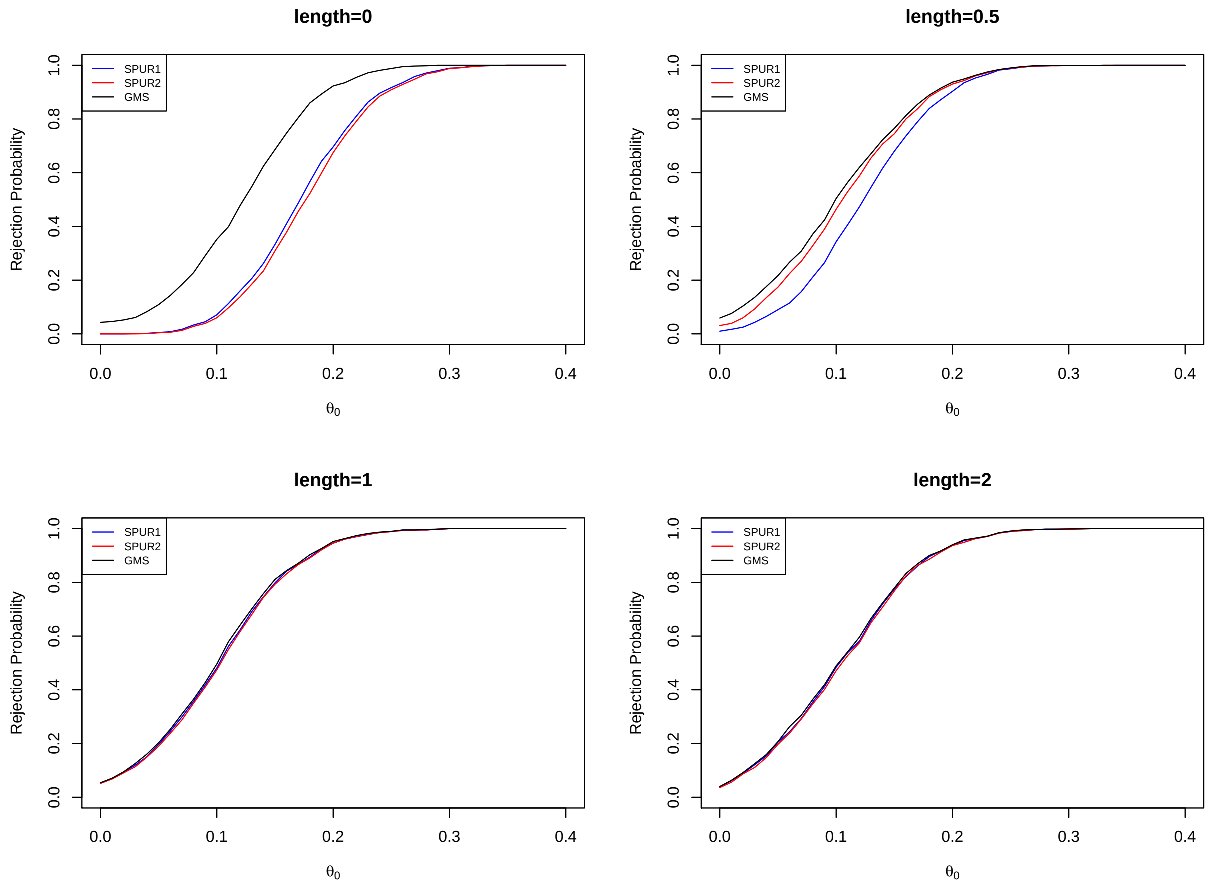


Figure 8.2: Rejection probabilities for correctly specified cases for $k = 2$. Each plot shows, for different lengths ℓ of the identified set, the rejection probabilities of the SPUR1, SPUR2, and standard GMS tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values and identified set $\Theta_I(F) = [-\ell, 0]$.

for the cases with $k = 4$, we see that the performance of the tests does not differ much across the different scenarios.

For the correctly-specified cases, we focus on the comparison of the SPUR1 and SPUR2 tests with the standard GMS test, which is known to perform well in such cases. From the discussion in Section 4.2, we expect the SPUR2 and standard GMS tests to exhibit similar performance when the length of the identified set is large enough. Indeed, in Figure 8.2, we see that when the length of the identified set is .5 the rejection probabilities of the two tests are very close to each other, and when the length is greater than .5 all three tests are essentially indistinguishable. We can also see that the SPUR2 test catches up to the standard GMS test under shorter identified sets than the SPUR1 test does, which shows its adaptive nature. However, when the identified set is a singleton, the SPUR1 and SPUR2 tests are more conservative than the standard GMS test under the null

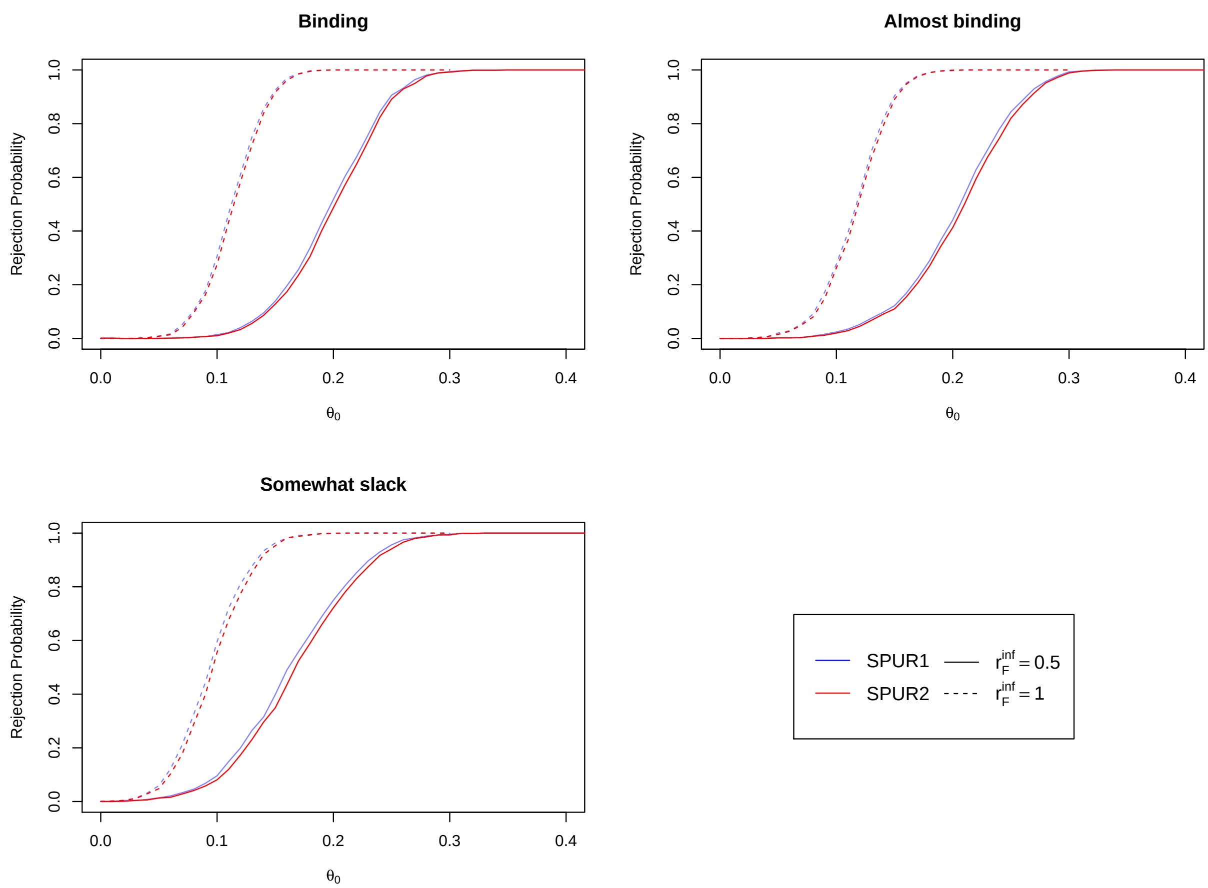


Figure 8.3: Rejection probabilities for misspecified cases for $k = 4$. Each plot shows, under different scenarios, the rejection probabilities of the SPUR1 and SPUR2 tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values, identified set $\Theta_I(F) = \{0\}$, and two different values of r_F^{inf}.

and have lower power over a wide range of positive θ_0 values. Essentially the same occurs when $k = 4$. That is, for each of the scenarios, the SPUR1 and SPUR2 tests are more conservative when the identified set has length 0, the SPUR2 test performs similarly to the standard GMS test when the length is .5, and all three tests are indistinguishable when the length is greater than .5. Again, this exhibits the adaptive nature of the SPUR2 test. When $k = 4$, the discrepancy between the standard GMS test and the SPUR1 and SPUR2 tests is largest in the "binding" scenario.

Section 18 in the Supplemental Material provides analogous results to those given above, but for $k = 8$. The same qualitative results are found to hold for $k = 8$ as for $k = 4$.

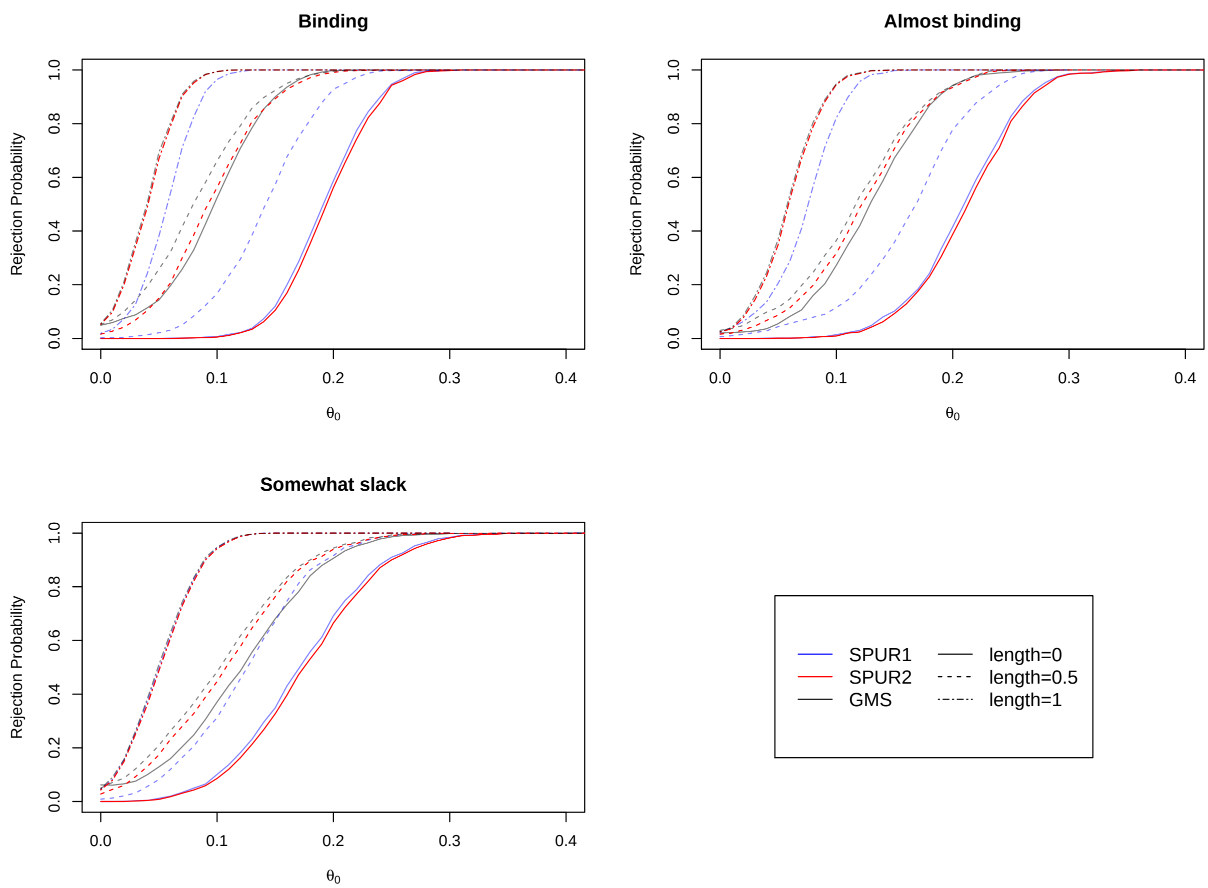


Figure 8.4: Rejection probabilities for correctly specified cases for $k = 4$. Each plot shows, under different scenarios, the rejection probabilities of the SPUR1, SPUR2, and standard GMS tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_0 values and different lengths ℓ of the identified set $\Theta_I(F) = [-\ell, 0]$.

8.2 Missing Data Model

In this subsection, we revisit the missing data model that BCS use in their simulations. The specification of the model closely follows BCS, but we consider a somewhat different data generating process.[15] Example 2.1 of BCS provides motivation for the model. Let $\{W_i = (Y_i Z_i, Z_i, X_i)\}_{i \le n}$ be the i.i.d data. Here, $Z_i \sim Bernoulli(p_z)$ is the indicator of whether the outcome variable Y_i is missing. It is independent of $(Y_i, X_i)'$. The conditional distribution of Y_i given X_i is

$$Y_i|X_i = x_1 \sim N(0,1),\ Y_i|X_i = x_2 \sim N((1+\widetilde{r})/p_z, 1), \text{ and } Y_i|X_i = x_3 \sim N(0,1), \tag{8.3}$$

[15] A different data generating process is employed to ensure that the random variable YZ is nonnegative, which is an implication of the structure of the missing data model.

with $P(X_i = x_1) = P(X_i = x_2) = P(X_i = x_3) = 1/3$. The parameter space is $\Theta = [-20, 20] \times [-20, 20]$. The moment functions are

$$\begin{aligned} m_1(W_i, \theta) &= (\theta_1 - YZ)1\{X = x_1\}, \\ m_2(W_i, \theta) &= (1 - \theta_1 - YZ)1\{X = x_2\}, \text{ and} \\ m_3(W_i, \theta) &= (\theta_2 - YZ)1\{X = x_3\} \text{ for } \theta = (\theta_1, \theta_2)'. \end{aligned} \tag{8.4}$$

The value of $\widetilde{r}$ determines whether the model is misspecified. When $\widetilde{r} \leq 0$, the model is correctly specified, which implies that $r_F^{\inf} = 0$, and the MR identified set is $\Theta_I(F) = [0, -\widetilde{r}] \times [0, \infty)$. When $\widetilde{r} > 0$, the model is misspecified and some calculations show that

$$r_F^{\inf} = \left(\frac{\widetilde{r}^2/3}{\left(p_z^{1/2} + ((1+\widetilde{r})^2(1/p_z - 1) + p_z)^{1/2}\right)^2 + 2\widetilde{r}^2/3} \right)^{1/2}. \tag{8.5}$$

For $\widetilde{r} > 0$, it can be shown that the MR identified set is $\Theta_I(F) = \{\theta_1^I(\widetilde{r})\} \times [\theta_1^I(\widetilde{r}), \infty)$, where

$$\theta_1^I(\widetilde{r}) := -\frac{p_z^{1/2}\widetilde{r}}{p_z^{1/2} + ((1+\widetilde{r})^2(1/p_z - 1) + p_z)^{1/2}}. \tag{8.6}$$

See Section 19 in the Supplemental Material for the derivations of (8.5) and (8.6).

We take $p_z = .8$ throughout. We consider values of $\widetilde{r}$ that cover both misspecified and correctly-specified cases. As above, we simulate rejection probabilities for a fixed data generating process and a range of null hypothesis values $\theta_0 = (\theta_{01}, \theta_{02})'$, where $H_0 : \theta = \theta_0$. For the null values, we consider θ_{02} fixed at $\theta_1^I(\widetilde{r})$ when $\widetilde{r} > 0$ and at 0 when $\widetilde{r} \leq 0$, and we consider a range of θ_{01} values. Accordingly, the x–axes in Figures 8.5 and 8.6 correspond to the first element of the null vector.

Figure 8.5 reports the simulated rejection probabilities for the misspecified cases with $\widetilde{r} = .1$, .2, .5, and 1.[16] Here, the MR identified set is $\{0\} \times [0, \infty)$. As in the lower/upper bound model, the SPUR1 and SPUR2 tests perform quite similarly, as expected. Also, the rejection probabilities increase to 1 fairly quickly as the distance between the null value and the MR identified set increases, and the performance is better for smaller values of $\widetilde{r}$ (or, equivalently, smaller values of $r_F^{\inf}$).

Figure 8.6 provides the results under correct specification. Here, we see that when $\widetilde{r} = 0$, which implies that the identified set contains no slack points, the standard GMS test performs better than the SPUR1 and SPUR2 tests, which is expected. In this case, the SPUR1 and SPUR2 tests have almost identical rejection probabilities. Also, the difference between the standard GMS test and

[16] By (8.5), these $\widetilde{r}$ values correspond (approximately) to $r_F^{\inf} = .03$, .07, .14, and .24, respectively.

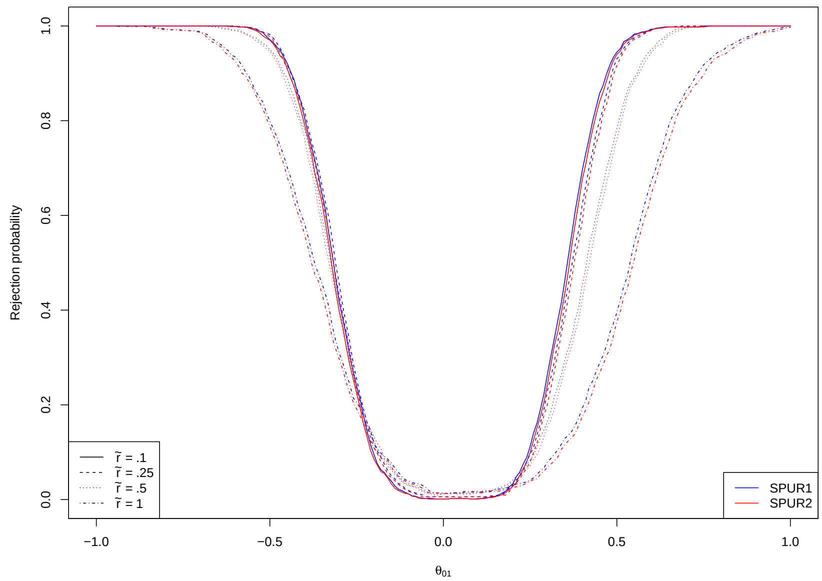


Figure 8.5: Rejection probabilities under misspecification for the missing data model. The figure shows the rejection probabilities of the SPUR1 and SPUR2 tests for the null hypothesis $H_0 : \theta = \theta_0$ for a range of θ_{01} values and a fixed identified set, for four different $\widetilde{r}$ values.

the SPUR2 test decreases quickly as the identified set gets larger (i.e., as $\widetilde{r}$ become more negative) and, hence, contains more slack points. The SPUR2 test is essentially on par with the standard GMS test when $\widetilde{r}$ is -1. The difference in power between the standard GMS test and the SPUR1 test also decreases to some extent as the identified set get larger. But, the SPUR1 test has lower power (similar to the $\widetilde{r} = -1$ case) even for $\widetilde{r}$ values in the range of $[-2, -5]$ (based on results not reported in Figure 8.6). Overall, the four plots show how the SPUR2 test adapts, and eventually behaves very much like the standard GMS test as the identified set gets larger.

9 Uniform Consistency of $\widehat{\Theta}_n$

The following result shows that the set estimator $\widehat{\Theta}_n$ defined in (6.5) is uniformly consistent for the MR identified set $\Theta_I(F)$ over $F \in \mathcal{P}$ with respect to the Hausdorff metric d_H. The result is similar to results in Theorem 3.1 of CHT except that it applies under both correct model specification and misspecification, and it establishes uniform, rather than pointwise, consistency.

For $\theta \in \Theta$ and $A \subset \Theta$, define the distance between θ and A as $d(\theta, A) := \inf_{\theta' \in A} ||\theta - \theta'||$. For any $\varepsilon > 0$ and $F \in \mathcal{P}$, define $\Theta_{I,\varepsilon}(F) := \{\theta \in \Theta : d(\theta, \Theta_I(F)) \leq \varepsilon\}$. The set $\Theta_{I,\varepsilon}(F)$ is an ε-expansion of the MR identified set $\Theta_I(F)$.

For any $F \in \mathcal{P}$, $\inf_{\theta \in \Theta \backslash \Theta_{I,\varepsilon}(F)} \max_{j \leq k} [E_F \widetilde{m}_j(W_i, \theta)]_- - r_F^{\inf} > 0$ for all $\varepsilon > 0$ under Assumption

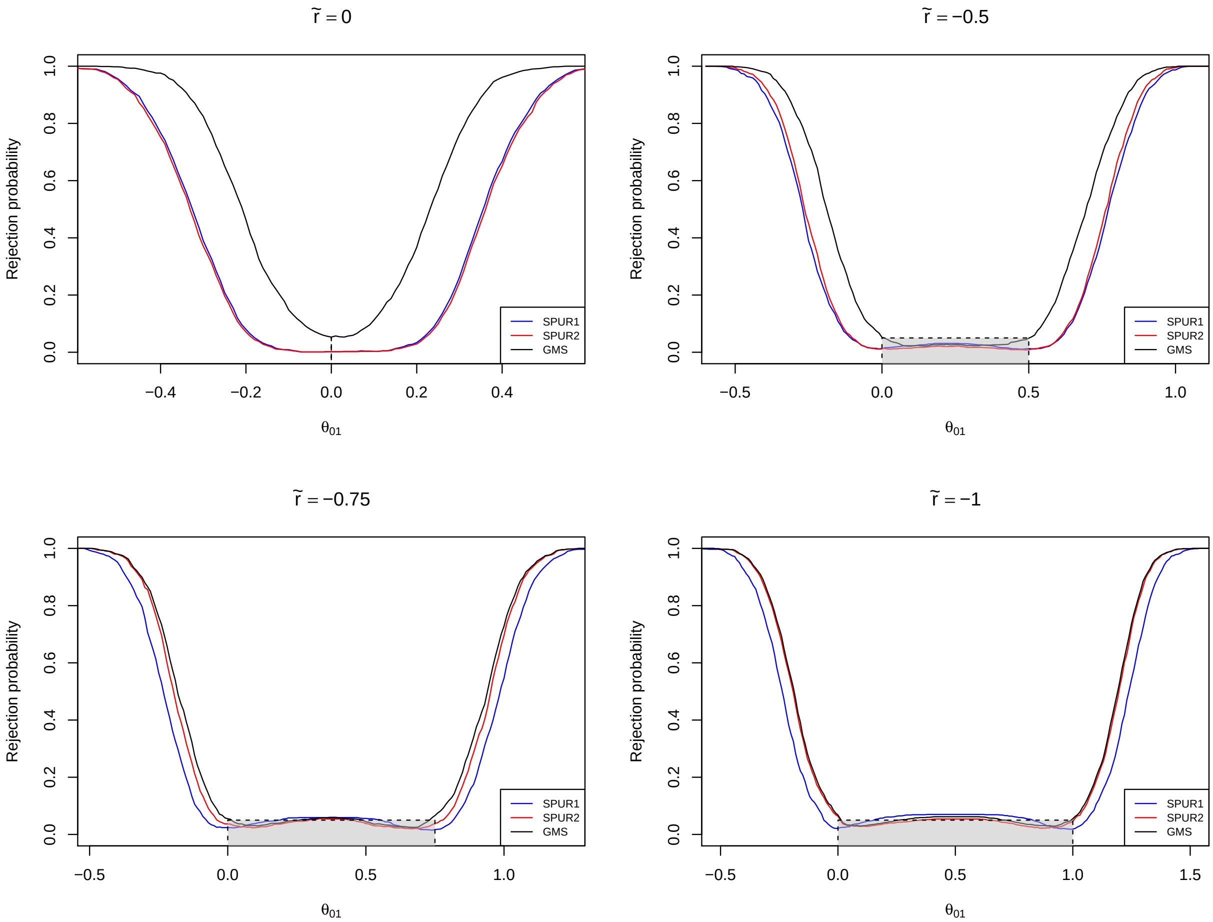


Figure 8.6: Rejection probabilities under correct specification for the missing data model. Each plot shows the rejection probabilities of the SPUR1, SPUR2, and standard GMS tests for the null hypothesis $H_0 : \theta = \theta_0$ and a range of θ_{01} values, for one of the four $\widetilde{r}$ values considered. The shaded region in each plot delineates the identified set.

A.0 by the definitions of $r_F^{\inf}$ and $\Theta_{I,\varepsilon}(F)$. The following Assumption A.9 requires that this positive quantity is bounded away from zero over $F \in \mathcal{P}$.

Assumption A.9. For all $\varepsilon > 0$, $\inf_{F\in\mathcal{P}} \inf_{\theta\in\Theta\backslash\Theta_{I,\varepsilon}(F)} \max_{j\le k} \, [E_F \widetilde{m}_j(W_i,\theta)]_- - r_F^{\inf} > 0$.

Uniform consistency of $\widehat{\Theta}_n$ for $\Theta_I(F)$ is established in the following theorem.

Theorem 9.1 *Suppose Assumptions* A.0–A.4, A.8, *and* A.9 *hold and the positive constants* $\{\tau_n\}_{n\ge1}$ *that appear in* (BC.5) *satisfy* $\tau_n \to \infty$ *and* $\tau_n/n^{1/2} = o(1)$. *Then, for all* $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{F\in\mathcal{P}} P_F(d_H(\widehat{\Theta}_n, \Theta_I(F)) > \varepsilon) = 0.$$

Comments. (i). If Assumption A.9 fails to hold, the result of Theorem 9.1 holds with $\mathcal{P}_U$ in place of $\mathcal{P}$ for any $\mathcal{P}_U \subset \mathcal{P}$ for which Assumption A.9 holds with $\mathcal{P}_U$ in place of $\mathcal{P}$. In particular, for a fixed distribution $F \in \mathcal{P}$, the result of Theorem 9.1 holds with $\mathcal{P}_U = \{F\}$ in place of $\mathcal{P}$ because Assumption A.9 automatically holds in this case.

(ii) Lemma 26.1(b) in the Supplemental Material provides rate of convergence results for the set estimator $\widehat{\Theta}_n$.

10 Appendix: Additional Assumptions

The following four assumptions concern the test function $S(m, \Omega)$ introduced in Section 3.

Assumption S.1. (i) $S(m, \Omega)$ is nonincreasing in $m \in R^k_{[+\infty]}$ $\forall \Omega \in \Psi$.

(ii) $S(m, \Omega) \geq 0$ $\forall m \in R^k$, $\forall \Omega \in \Psi$.

(iii) $S(m, \Omega)$ is continuous at all $m \in R^k_{[+\infty]}$ and $\Omega \in \Psi$.

Assumption S.2. $S(m, \Omega) > 0$ iff $m_j < 0$ for some $j \leq k$, $\forall \Omega \in \Psi$.

Assumption S.3. For some $\chi > 0$, $S(am, \Omega) = a^\chi S(m, \Omega)$ $\forall a > 0$, $\forall m \in R^k$, $\forall \Omega \in \Psi$.

Assumption S.4. For all $h \in (-\infty, \infty]^k$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the distribution function of $S(Z + h, \Omega)$ at $x \in R$ is (i) continuous for $x > 0$, (ii) strictly increasing for $x > 0$ unless $h = (\infty, \ldots, \infty)' \in R^k_{[\pm\infty]}$, and (iii) less than 1/2 for $x = 0$ if $h_j = 0$ for some $j \leq k$.[17]

The following assumptions define $n^{-1/2}$-local alternatives and fixed alternatives.

Assumption LA. The null values $\{\theta_n\}_{n \geq 1}$ and distributions $\{F_n\}_{n \geq 1}$ satisfy: (i) $||\theta_n - \theta_{In}|| = O(n^{-1/2})$ for some sequence $\{\theta_{In} \in \Theta_I(F_n)\}_{n \geq 1}$, (ii) $n^{1/2}(E_{F_n} \widetilde{m}_j(W, \theta_{In}) + r^{\inf}_{F_n}) \to h_{Ij\infty}$ for some $h_{Ij\infty} \in R_{[\pm\infty]}$ $\forall j \leq k$, and (iii) $E_F \widetilde{m}(W, \theta)$ is Lipschitz on Θ uniformly over $\mathcal{P}$, i.e., there exists a constant $K < \infty$ such that $||E_F \widetilde{m}(W, \theta_1) - E_F \widetilde{m}(W, \theta_2)|| \leq K||\theta_1 - \theta_2||$ $\forall \theta_1, \theta_2 \in \Theta$, $\forall F \in \mathcal{P}$.

Assumption FA. (i) $(\theta_n, F_n) = (\theta_*, F_*) \in \Theta \times \mathcal{P}$ does not depend on $n \geq 1$ and (ii) $E_{F_*} \widetilde{m}_j(W, \theta_*) + r^{\inf}_{F_*} < 0$ for some $j \leq k$.

The following assumption concerns the GMS function $\varphi = (\varphi_1, ..., \varphi_k)'$, which appears in (6.4).

Assumption A.5. Given the function $\varphi : R^k_{[+\infty]} \times \Psi \to R^k_{[+\infty]}$, there is a function $\varphi^* : R^k_{[+\infty]} \to R^k_{[+\infty]}$ that takes the form $\varphi^*(\xi) = (\varphi^*_1(\xi_1), ..., \varphi^*_k(\xi_k))'$ and $\forall j \leq k$, (i) $\varphi^*_j(\xi_j) \geq \varphi_j(\xi, \Omega) \geq 0$ $\forall (\xi, \Omega) \in R^k_{[+\infty]} \times \Psi$, (ii) φ^*_j is nondecreasing and continuous under the metric d, and (iii) $\varphi^*_j(\xi_j) = 0$ $\forall \xi_j \leq 0$ and $\varphi^*_j(\infty) = \infty$.

[17] Assumption S.1(i), (ii), and (iii), S.2, and S.3 correspond to Assumptions 1(a), (c), and (d), 3, and 6 in Andrews and Guggenberger (2009) and Andrews and Soares (2010) and M.4(a), (c), and (d), M.7, and M.8 in BCS, respectively. Assumption S.4 is a variation of Assumption 2 in Andrews and Soares (2010).

The function $\varphi_j(\xi, \Omega) = \infty 1(\xi_j > 1)$ for $j \leq k$, where $\infty \cdot 0 := 0$, satisfies Assumption A.5 with $\varphi_j^*(\xi_j) = \infty 1(\xi_j \geq 1) + (\xi_j/(1-\xi_j))1(0 \leq \xi_j < 1)$. For other choices of φ, including one that depends on Ω, see Andrews and Soares (2010).

The following are the conditions on κ_n and τ_n, which appear in (6.4) (and elsewhere) and (6.5), respectively.

Assumption A.6. (i) $\kappa_n \to \infty$. (ii) $\tau_n \to \infty$.[18]

The asymptotic size of a nominal level $1-\alpha$ CS based on a test $\phi_n(\cdot)$ is $\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} \inf_{\theta\in\Theta_I(F)} P_F(\phi_n(\theta) = 0)$. It is determined using subsequence arguments as follows. There always exist sequences $\{F_n\}_{n\geq 1}$ and $\{\theta_n \in \Theta_I(F_n)\}_{n\geq 1}$ and a subsequence $\{q_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ such that

$$\liminf_{n\to\infty} \inf_{F\in\mathcal{P}} \inf_{\theta\in\Theta_I(F)} P_F(\phi_n(\theta) = 0) = \liminf_{n\to\infty} P_{F_n}(\phi_n(\theta_n) = 0) = \lim P_{F_{q_n}}(\phi_{q_n}(\theta_{q_n}) = 0). \tag{10.1}$$

Hence, to establish correct asymptotic level, it suffices to show that the right-hand side of (10.1) is $1-\alpha$ or greater. For the subsequences $\{F_{q_n}\}_{n\geq 1}$ and $\{\theta_{q_n} \in \Theta_I(F_{q_n})\}_{n\geq 1}$, the test statistic S_{q_n} has asymptotic distribution S_∞ defined in (5.11). Let $c_\infty(1-\alpha)$ denote the $1-\alpha$ quantile of S_∞. Note that $c_\infty(1-\alpha) \geq 0$. We impose the following assumption on the distribution function of S_∞ at $c_\infty(1-\alpha)$. This assumption is only employed in conjunction with Assumption N, i.e., when S_∞ is an asymptotic null distribution of S_n.

Assumption A.7. Under $\{F_{q_n}\}_{n\geq 1}$ and $\{\theta_{q_n}\}_{n\geq 1}$, (i) if $c_\infty(1-\alpha) > 0$, then $P(S_\infty = c_\infty(1-\alpha)) = 0$, and (ii) if $c_\infty(1-\alpha) = 0$, then $\limsup_{n\to\infty} P_{F_{q_n}}(S_{q_n} > 0) \leq \alpha$.

When testing $H_0 : \theta_0 \in \Theta_I(F)$, $\{\theta_{q_n}\}_{n\geq 1}$ in Assumption A.7 is replaced by $\{\theta_0\}_{n\geq 1}$. Assumption A.7 is closely related to Assumption A.7 in BCS.

In the asymptotic results in Theorem 7.1, Assumption A.7(ii) can be eliminated if one defines the critical value to be $\max\{\widehat{c}_{n,EGMS}(\theta, 1-\alpha), \zeta\}$ for $\theta = \theta_n$ or $\theta = \theta_0$ for some very small constant $\zeta > 0$, such as 10^{-6}. In the vast majority of scenarios, this has no effect on the test or CS in finite samples or asymptotically (because $\widehat{c}_{n,EGMS}(\theta, 1-\alpha)$ determines the maximum).

Assumption A.8. $E_F\widetilde{m}(W,\theta)$ is equicontinuous on Θ over $F \in \mathcal{P}$. That is, $\lim_{\delta\downarrow 0} \sup_{F\in\mathcal{P}} \sup_{||\theta-\theta'||<\delta} ||E_F\widetilde{m}(W,\theta) - E_F\widetilde{m}(W,\theta')|| = 0$.

[18] Assumption A.6, as well as Assumptions A.8 defined below, are unrelated to Assumptions A.6 and A.8 in BCS. Assumptions A.1–A.5 and A.7 are related to assumptions with the same names in BCS.

References

Allen R., and J. Rehbeck (2018), "Assessing Misspecification and Aggregation for Structured Preferences," unpublished manuscript, Department of Economics, University of Western Ontario.

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

——— (2010), "Asymptotic Size and a Problem with Subsampling and with the m Out of n Bootstrap," *Econometric Theory,* 26, 426–468.

Andrews, D. W. K., and S. Han (2009), "Invalidity of the Bootstrap and the m Out of n Bootstrap for Confidence Interval Endpoints Defined by Moment Inequalities," *Econometrics Journal,* 12, S172–S199.

Andrews, D. W. K., and S. Kwon (2019), "Confidence Intervals for Measures of Misspecification in Moment Inequality and/or Equality Models," unpublished manuscript, Cowles Foundation, 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., 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.

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.

Hansen, B. E., and S. Lee (2019), "Inference for Iterated GMM under Misspecification," unpublished working paper, Department of Economics, University of Wisconsin, Madison.

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

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

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.

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

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

Molinari, F. (2019): "Econometrics with Partial Identification," cemmap working paper CWP25/19, Institute for Fiscal Studies, University College London.

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.