Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models*

Federico A. Bugni †
Department of Economics
Duke University

Ivan A. Canay
Department of Economics
Northwestern University

Patrik Guggenberger
Department of Economics
University of California, San Diego

November 17, 2011

Abstract

This paper studies the behavior under local misspecification of several confidence sets (CSs) commonly used in the literature on inference in moment (in)equality models. We propose the amount of asymptotic confidence size distortion as a criterion to choose among competing inference methods. This criterion is then applied to compare across test statistics and critical values employed in the construction of CSs. We find two important results under weak assumptions. First, we show that CSs based on subsampling and generalized moment selection (Andrews and Soares, 2010) suffer from the same degree of asymptotic confidence size distortion, despite the fact that asymptotically the latter can lead to CSs with strictly smaller expected volume under correct model specification. Second, we show that the asymptotic confidence size of CSs based on the quasi-likelihood ratio test statistic can be an arbitrary small fraction of the asymptotic confidence size of CSs based on the modified method of moments test statistic.

KEYWORDS: asymptotic confidence size, moment inequalities, partial identification, size distortion, uniformity, misspecification.

---

*This paper was previously circulated under the title "Asymptotic Distortions in Locally Misspecified Moment Inequality Models". We thank the co-Editor, Jim Stock, and three referees for very helpful comments and suggestions. We also thank seminar participants at various universities, the 2010 Econometric Society World Congress, the Cemmap/Cowles "Advancing Applied Microeconometrics" conference, the Econometrics Jamboree at Duke, and the 2011 Econometric Society North American Winter Meeting for helpful comments. Bugni, Canay, and Guggenberger thank the National Science Foundation for research support via grants SES-1123771, SES-1123586, and SES-1021101, respectively. Guggenberger would also like to thank the Alfred P. Sloan Foundation for a 2009-2011 fellowship.

†Emails: federico.bugni@duke.edu; iacanay@northwestern.edu; pguggenberger@ucsd.edu.

1 Introduction

In the last couple of years there have been numerous papers in econometrics on inference in partially identified models. Many of these papers focused on inference on the identifiable parameters in models defined by moment (in)equalities of the form

$$\begin{aligned} &E_{F_0} m_j(W_i,\theta_0) \geq 0 \text{ for } j = 1,\ldots,p, \\ &E_{F_0} m_j(W_i,\theta_0) = 0 \text{ for } j = p+1,\ldots,p+v \equiv k, \end{aligned} \tag{1.1}$$

where $\theta_0 \in \Theta$ is the parameter of interest, $\{m_j(\cdot,\theta)\}_{j=1}^{k}$ are known real-valued functions, and $\{W_i\}_{i=1}^{n}$ are observed i.i.d. random vectors with joint distribution F_0. See, e.g., Imbens and Manski (2004), Chernozhukov et al. (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009b, AG from now on), and Andrews and Soares (2010).[1] As a consequence, there are currently several different testing procedures and methods to construct $(1-\alpha)$ level confidence sets (CSs) given by

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta, 1-\alpha)\}, \tag{1.2}$$

where $T_n(\theta)$ is a generic test statistic for testing the hypothesis

$$H_0 : \theta_0 = \theta \quad \text{vs.} \quad H_1 : \theta_0 \neq \theta, \tag{1.3}$$

and $c_n(\theta, 1-\alpha)$ is the critical value of the test at nominal size α. Different CSs (i.e. different combinations of tests statistics and critical values) have been compared in the literature in terms of asymptotic confidence size and asymptotic power properties (e.g. Andrews and Jia, 2008; AG; Andrews and Soares, 2010; Bugni, 2010; Canay, 2010).

In this paper we are interested in the relative robustness of CSs with respect to their distortion in asymptotic confidence size when moment (in)equalities are potentially locally violated.[2] We consider a parameter space $\mathcal{F}_n$ of (θ, F) that includes local deviations with respect to the original model in Eq. (1.1). The space $\mathcal{F}_n$ in turn enters the definition of asymptotic confidence size of CS_n in Eq. (1.2), i.e.,

$$AsySz = \liminf_{n\to\infty} \inf_{(\theta,F)\in\mathcal{F}_n} \Pr_{\theta,F}(T_n(\theta) \leq c_n(\theta, 1-\alpha)), \tag{1.4}$$

where $\Pr_{\theta,F}(\cdot)$ denotes the probability measure when the true value of the parameter is θ and the true distribution is F. Intuition might suggest that inference procedures with relatively high local power in correctly specified models suffer from relatively high distortion of asymptotic confidence size in locally misspecified models. While this intuition is supported

[1] Additional references include Pakes et al. (2005), Beresteanu and Molinari (2008), Bontemps et al. (2008), Rosen (2008), Fan and Park (2009), Galichon and Henry (2009), Stoye (2009), Bugni (2010), Canay (2010), Romano and Shaikh (2010), Galichon and Henry (2011), and Moon and Schorfheide (2011), among others.

[2] Different types of local misspecification in moment equality models have been studied by Newey (1985), Kitamura et al. (2009), and Guggenberger (2011), among others.

by several of our results, the main contributions of our paper show that the new robustness criterion can lead to conclusions that go well beyond such intuition. First, we show under mild assumptions that CSs based on subsampling and GMS critical values suffer from the same level of asymptotic size distortion, despite the fact that the latter can lead to CSs with strictly smaller expected volume under correct model specification (see Andrews and Soares, 2010). Second, we show that under certain conditions the asymptotic confidence size of CSs based on the quasi-likelihood ratio test statistic can be an arbitrary small fraction of the asymptotic confidence size of CSs based on the modified method of moments test statistic.

The novel notion of robustness proposed in this paper may provide additional discriminatory power between inference methods relative to local asymptotic power comparisons (e.g. Theorem 3.1). Consider testing the null hypothesis in Eq. (1.3), where local power is the limit of the rejection probability under a sequence of parameters that belongs to the alternative hypothesis and approaches the null hypothesis. Local power comparisons involve computing rejection probabilities of different tests under *the same sequence of local alternatives.* The test with higher limiting rejection probability under a given sequence is said to have higher local power against such particular local alternative. In the context of local misspecification, these local sequences typically belong to the parameter space $\mathcal{F}_n$ that determines the asymptotic confidence size in Eq. (1.4). The derivation of asymptotic confidence size then involves computing rejection probabilities under all sequences of parameters in $\mathcal{F}_n$ and search for the one that leads to the highest limiting rejection probability (referred to as the worst local sequence). As a result, the worst local sequence for one test might be *different* than the worst local sequence for a rival test, meaning that the behavior of these tests under *the same* sequence of local alternative parameters is insufficient to describe distortions under local misspecification. In other words, the analysis of robustness we propose is more complex than a local power analysis as it involves finding the worst case sequence in $\mathcal{F}_n$ (including local alternatives) for each of the test procedures under consideration.

The motivation behind the interest in misspecified models stems from the view that most econometric models are only approximations to the underlying phenomenon of interest and are therefore intrinsically misspecified. The partial identification approach to inference allows the researcher to conduct inference on the parameter of interest without imposing assumptions on certain fundamental aspects of the model, typically related to the behavior of economic agents. Still, for computational or analytical convenience or to obtain at least partial identification of the parameter of interest, the researcher has to impose certain other assumptions, that are typically related to functional forms or distributional assumptions.[3] Here we will not discuss the nature of a certain assumption, but rather we will take the set of moment (in)equalities as given and study how different inferential methods perform when the maintained set of assumptions is allowed to be violated (i.e. when we allow the model to

[3]See Manski (2003) and Tamer (2010) for an extensive discussion on the role of different assumptions and partial identification. Also, Ponomareva and Tamer (2011) discuss the impact of global misspecification on the set of identifiable parameters.

be misspecified).

The paper is organized as follows. Section 2 introduces the model, testing procedures, and provides an example that illustrates the nature of misspecification in our framework, Section 3 presents the theoretical results, and Section 4 concludes. The Appendix contains technical definitions, assumptions, and the proofs of the theorems and main lemma. A Supplemental Appendix (Bugni et al., 2011) includes auxiliary results and their proofs, the proof of Corollary 3.1, an additional example, verification of the assumptions in examples, and Monte Carlo simulations.

Throughout the paper we use the notation $h = (h_1, h_2)$, where h_1 and h_2 are allowed to be vectors or matrices. We also use $K^p = K \times \cdots \times K$ (with p copies) for any set K, $\infty_p = (+\infty, \ldots, +\infty)$ (with p copies), 0_p for a p-vector of zeros, I_p for a $p \times p$ identity matrix, $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{R}_{+,+\infty} = \mathbb{R}_+ \cup \{+\infty\}$, $\mathbb{R}_{+\infty} = \mathbb{R} \cup \{+\infty\}$, and $\mathbb{R}_{\pm\infty} = \mathbb{R} \cup \{\pm\infty\}$.

2 Locally Misspecified Moment (In)Equality Models

There are several CSs suggested in the literature whose asymptotic confidence size is at least equal to the nominal size. We consider CSs as in Eq. (1.2), which are determined by the choice of a test statistic $T_n(\theta)$ and a critical value $c_n(\theta, 1-\alpha)$. The test statistics include modified method of moments, quasi-likelihood ratio, and generalized empirical likelihood statistics. Critical values include plug-in asymptotic (PA), subsampling (SS), and generalized moment selection (GMS) implemented via asymptotic approximations or the bootstrap.

To assess the relative advantages of these procedures the literature has mainly focused on asymptotic size and power in correctly specified models. Bugni (2010) shows that GMS tests have more accurate asymptotic size than subsampling tests. Andrews and Soares (2010) establish that GMS tests are as powerful as subsampling tests for all sequences of local alternatives and strictly more powerful along certain sequences of local alternatives. In turn, subsampling tests are as powerful as PA tests for all sequences of local alternatives and strictly more powerful along some sequences of local alternatives. Andrews and Jia (2008) compare different combinations of tests statistics and critical values and provide a recommended test based on the quasi-likelihood ratio statistic and a refined moment selection critical value which involves a data-dependent rule for choosing the GMS tuning parameter. Additional results on power include those in Canay (2010). In this paper we are interested in ranking the resulting CSs in terms of asymptotic confidence size distortion when the moment (in)equalities in Eq. (1.1) are potentially locally violated. The following example is an illustration.

Example 2.1 (Entry Game)**.** Consider the following game. Firm $l \in \{1, 2\}$ enters a market $i \in \{1, \ldots, n\}$ whenever its profits after entry are positive. Assume the profit function is given by $\pi_{l,i}(\theta_l, W_{-l,i}) \equiv u_{l,i} - \theta_l W_{-l,i}$. Here $W_{l,i} = 1$ or 0 denotes "entering" or "not entering" market i by firm l, respectively, the subscript $-l$ denotes the decision of the other firm, the non-negative continuous random variable $u_{l,i}$ denotes the monopoly profits of firm l in market

i, and $\theta_l \in [0,1]$ is the profit reduction incurred by firm l if $W_{-l,i} = 1$. If $W_{l,i} = 0$, then $\pi_{l,i} = 0$. Thus, entering is always profitable for at least one firm.

Define $W_i = (W_{1,i}, W_{2,i})$ and $\theta_0 = (\theta_1, \theta_2)$. There are four possible outcomes in each market: (i) $W_i = (1,1)$ is the unique (Nash) equilibrium if $u_{l,i} > \theta_l$ for $l = 1,2$, (ii) $W_i = (1,0)$ is the unique equilibrium if $u_{1,i} > \theta_1$ and $u_{2,i} < \theta_2$, (iii) $W_i = (0,1)$ is the unique equilibrium if $u_{1,i} < \theta_1$ and $u_{2,i} > \theta_2$, and (iv) $W_i = (1,0)$ and $W_i = (0,1)$ are both equilibria if $u_{l,i} < \theta_l$ for $l = 1,2$. Assuming $u \sim G$ for some bivariate distribution G, the model implies

$$\begin{aligned} \Pr(W_i = (1,0)) &\le \Pr(u_{2,i} < \theta_2) \equiv G_1(\theta_0), \\ \Pr(W_i = (1,0)) &\ge \Pr(u_{1,i} > \theta_1 \ \&\ u_{2,i} < \theta_2) \equiv G_2(\theta_0), \\ \Pr(W_i = (1,1)) &= \Pr(u_{1,i} > \theta_1 \ \&\ u_{2,i} > \theta_2) \equiv G_3(\theta_0), \end{aligned} \tag{2.1}$$

where the functions $G_1(\theta_0)$, $G_2(\theta_0)$, $G_3(\theta_0)$ are implicitly defined as functions of G. Therefore

$$\begin{aligned} E_{F_0} m_1(W_i, \theta_0) &= E_{F_0}[G_1(\theta_0) - W_{1,i}(1 - W_{2,i})] \ge 0, \\ E_{F_0} m_2(W_i, \theta_0) &= E_{F_0}[W_{1,i}(1 - W_{2,i}) - G_2(\theta_0)] \ge 0, \\ E_{F_0} m_3(W_i, \theta_0) &= E_{F_0}[W_{1,i} W_{2,i} - G_3(\theta_0)] = 0, \end{aligned} \tag{2.2}$$

where F_0 denotes the true distribution of W_i compatible with the joint distribution of u_i.

To do inference on θ_0, the researcher assumes G is the joint distribution of the unobserved random vector u_i.[4] Now suppose that the data comes from a local perturbation of the hypothesized model. For example, suppose that $u_i \sim G_n$, where G_n is such that

$$|G_j(\theta_0) - G_{n,j}(\theta_0)| \le r_j n^{-1/2}, \quad j = 1,2,3, \tag{2.3}$$

for some $r = (r_1, r_2, r_3)' \in \mathbb{R}^3_+$, and $G_{n,j}(\theta_0)$ is defined as $G_j(\theta_0)$ above when $u_i \sim G_n$ rather than $u_i \sim G$. Denote by F_n the true distribution of W_i that is compatible with the true joint distribution of $u_i \sim G_n$. Then, combining Eqs. (2.1) and (2.2) we obtain

$$\begin{aligned} E_{F_n} m_1(W_i, \theta_0) &= E_{F_n}[G_1(\theta_0) - W_{1,i}(1 - W_{2,i})] \ge -r_1 n^{-1/2}, \\ E_{F_n} m_2(W_i, \theta_0) &= E_{F_n}[W_{1,i}(1 - W_{2,i}) - G_2(\theta_0)] \ge -r_2 n^{-1/2}, \\ |E_{F_n} m_3(W_i, \theta_0)| &= |E_{F_n}[W_{1,i} W_{2,i} - G_3(\theta_0)]| \le r_3 n^{-1/2}. \end{aligned} \tag{2.4}$$

Thus, under the distribution F_n the moment conditions may be locally violated at θ_0.[5] ■

Remark 2.1. Note that the parameter θ_0 in the example has a meaningful interpretation independently of the potential misspecification of the model of the type considered above. However, as demonstrated, if the researcher assumes an incorrect distribution for the profits, the moment (in)equalities are potentially violated for every given sample size n at the true

[4] Note that in order to make inference on θ_0 the researcher is forced to make an assumption on G as θ_0 and G are not jointly identified. That is, without an assumption on G, θ_0 is simply not identified.

[5] For simplicity the true value θ_0 was not indexed by n even though our analysis below allows for this possibility. However, we assume throughout that the distribution G does not depend on n.

θ_0. The assumption of correct specification by the researcher of the distribution of u_i is very strong - it is therefore of critical importance to assess how robust (in terms of distortion in asymptotic size) the competing inference procedures are when the assumption fails.

Example 2.1 illustrates that local misspecification in moment inequality models can be represented by a parameter space that allows the moment conditions to be "slightly" violated, i.e., slightly negative in the case of inequalities and slightly different from zero in the case of equalities.[6] We capture this idea in the definition below, where $m(W_i, \theta) = (m_1(W_i, \theta), \ldots, m_k(W_i, \theta))$ and (θ, F) denote generic values of the parameters.

Definition 2.1 (Sequence of Parameter Spaces with Misspecification)**.** For each $n \in \mathbb{N}$, the parameter space $\mathcal{F}_n \equiv \mathcal{F}_n(r, \delta, M, \Psi)$ is the set of all tuplets (θ, F) that satisfy

$$\begin{aligned}
&(i)\ \theta \in \Theta, \\
&(ii)\ \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) \geq -r_j n^{-1/2}, \quad j = 1, \ldots, p, \\
&(iii)\ |\sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta)| \leq r_j n^{-1/2}, \quad j = p+1, \ldots, k, \\
&(iv)\ \{W_i\}_{i=1}^n \text{ are i.i.d. under } F, \\
&(v)\ \sigma_{F,j}^2(\theta) = Var_F(m_j(W_i, \theta)) \in (0, \infty), \quad j = 1, \ldots, k, \\
&(vi)\ Corr_F(m(W_i, \theta)) \in \Psi, \text{ and} \\
&(vii)\ E_F |m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M, \quad j = 1, \ldots, k,
\end{aligned} \tag{2.5}$$

where Ψ is a specified closed set of $k \times k$ correlation matrices (that may depend on the test statistic; see below), $M < \infty$ and $\delta > 0$ are fixed constants, and $r = (r_1, \ldots, r_k) \in \mathbb{R}_+^k$.

Relative to the space $\mathcal{F} \equiv \mathcal{F}_n(0_k, \delta, M, \Psi)$ in AG and Andrews and Soares (2010), the sequence of parameter spaces with misspecification depends on n and the "upper bound" on the local moment (in)equality violation r. The definitions are essentially the same, with the exception of conditions (ii)-(iii), which are modified in order to account for possible local model misspecification. We use

$$r^* \equiv \max\{r_1, \ldots, r_k\} \tag{2.6}$$

to measure the amount of misspecification.

Remark 2.2. The parameter space in Eq. (2.5) includes the space $\mathcal{F}$, which is the set of correctly specified models. The theorems in the next section continue to hold if we alternatively define $\mathcal{F}_n$ enforcing that at least one moment (in)equality is strictly locally violated. For example, one way of doing this would be to add the restriction $\sigma_{F,j}^{-1}(\theta) E_F m_j(W, \theta) = -r_j n^{-1/2}$ with $r_j > 0$ for some $j = 1, \ldots, k$.

The existing literature on inference in partially identified moment (in)equality models shows that several inference procedures achieve $AsySz \geq 1 - \alpha$ when $r^* = 0$. In this paper

[6]Example S1.1 in the Supplemental Appendix provides another illustration of this representation.

we are interested in comparing these inference procedures when there is local misspecification (i.e. $r^* > 0$). In particular, we are interested in ranking the procedures according to their distortion in asymptotic confidence size, defined as $\max\{1 - \alpha - AsySz, 0\}$. Before doing this, we present the different test statistics and critical values in the next subsection.

Remark 2.3. Alternatively one could focus on the asymptotic size distortion of the tests for the null $H_0 : \theta_0 = \theta$. The asymptotic size in that case would involve a supremum with respect to F while keeping θ fixed, i.e., $\limsup_{n\to\infty} \sup_{F:(\theta,F)\in\mathcal{F}_n} \Pr_{\theta,F}(T_n(\theta) > c_n(\theta, 1-\alpha))$. Analytically, studying such quantity is less complex than studying $AsySz$ as in the former case θ is fixed at a particular value while in the latter case θ may depend on n.

2.1 Test Statistics and Critical Values

We now present several test statistics $T_n(\theta)$ and corresponding critical values $c_n(\theta, 1-\alpha)$ to test the null hypothesis in Eq. (1.3) or, equivalently, to construct a CS as in Eq. (1.2). Define the sample moment functions $\bar{m}_n(\theta) = (\bar{m}_{n,1}(\theta), \ldots, \bar{m}_{n,k}(\theta))'$, where

$$\bar{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^{n} m_j(W_i, \theta) \text{ for } j = 1, \ldots, k. \tag{2.7}$$

Let $\hat{\Sigma}_n(\theta)$ be a consistent estimator of the asymptotic variance matrix of $n^{1/2}\bar{m}_n(\theta)$. Under our assumptions, a natural choice for this estimator is

$$\hat{\Sigma}_n(\theta) = n^{-1} \sum_{i=1}^{n} (m(W_i, \theta) - \bar{m}_n(\theta))(m(W_i, \theta) - \bar{m}_n(\theta))'. \tag{2.8}$$

The statistic $T_n(\theta)$ is defined to be of the form $T_n(\theta) = S(n^{1/2}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta))$, where S is a real-valued function on $\mathbb{R}^p_{+\infty} \times \mathbb{R}^v \times \mathcal{V}_{k\times k}$ that satisfies Assumptions A.1-A.3 and $\mathcal{V}_{k\times k}$ is the space of $k \times k$ covariance matrices. We now describe two popular choices of test functions. The first test function S is the modified method of moments given by

$$S_1(m, \Sigma) = \sum_{j=1}^{p} [m_j/\sigma_j]^2_- + \sum_{j=p+1}^{k} (m_j/\sigma_j)^2, \tag{2.9}$$

where $[x]_- = xI(x < 0)$, $m = (m_1, \ldots, m_k)'$, and σ_j^2 is the jth diagonal element of Σ. For S_1, $\Psi = \Psi_1$ in condition (vi) of Eq. (2.5), where Ψ_1 is the set of all $k \times k$ correlation matrices. Letting $\hat{\sigma}^2_{n,j}(\theta)$ be the jth diagonal element of $\hat{\Sigma}_n(\theta)$, the function S_1 leads to the test statistic

$$T_{1,n}(\theta) = n \sum_{j=1}^{p} [\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]^2_- + n \sum_{j=p+1}^{k} (\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta))^2. \tag{2.10}$$

The second test function is a Gaussian quasi-likelihood ratio function defined by

$$S_2(m, \Sigma) = \inf_{t=(t_1,0_v):t_1\in\mathbb{R}^p_{+,+\infty}} (m - t)'\Sigma^{-1}(m - t). \tag{2.11}$$

This function requires Σ to be non-singular, i.e., $\Psi = \Psi_{2,\varepsilon}$ in Eq. (2.5), where $\Psi_{2,\varepsilon} = \{\Sigma \in$

$\Psi_1 : \det(\Sigma) \geq \varepsilon\}$ for some $\varepsilon > 0$. The function S_2 leads to the test statistic

$$T_{2,n}(\theta) = \inf_{t=(t_1,0_v):t_1 \in \mathbb{R}^p_{+,+\infty}} (n^{1/2}\bar{m}_n(\theta) - t)'\hat{\Sigma}_n(\theta)^{-1}(n^{1/2}\bar{m}_n(\theta) - t). \tag{2.12}$$

The functions S_1 and S_2 satisfy Assumptions A.1-A.3 that are slight generalizations of Assumptions 1-4 in AG to our setup.

We next describe three main choices of critical values. Assuming the limiting correlation matrix of $m(W_i, \theta)$ is given by Ω and that $r^* = 0$ (i.e. correct specification), it follows that

$$T_n(\theta) \to_d S(\Omega^{1/2}Z + h_1, \Omega), \tag{2.13}$$

where $Z \sim N(0_k, I_k)$, h_1 is a k-vector with $h_{1,j} = 0$ for $j > p$ and $h_{1,j} \in [0, \infty]$ for $j \leq p$ (for details see Lemma S1.1 in the Supplemental Appendix), and $\Omega^{1/2}$ denotes a lower triangular matrix such that $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$. Ideally one would use as critical value the $1-\alpha$ quantile of $S(\Omega^{1/2}Z + h_1, \Omega)$, denoted by $c_{h_1}(\Omega, 1-\alpha)$ or, at least, a consistent estimator of it. However, h_1 cannot be estimated consistently (see AG), and so some approximation to $c_{h_1}(\Omega, 1-\alpha)$ is necessary.

Under the assumptions in the Appendix, the asymptotic distribution in Eq. (2.13) is stochastically largest over distributions in $\mathcal{F}$ (i.e. correctly specified models) when all the inequalities are binding (i.e. hold as equalities). As a result, the least favorable critical value can be shown to be $c_0(\Omega, 1-\alpha)$, the $1-\alpha$ quantile of $S(\Omega^{1/2}Z, \Omega)$ (i.e. $h_1 = 0_k$).[7] PA critical values are based on this "worst case" and are defined as consistent estimators of $c_0(\Omega, 1-\alpha)$. Let $\hat{D}_n(\theta) = Diag(\hat{\Sigma}_n(\theta))$ and define

$$\hat{\Omega}_n(\theta) = \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta). \tag{2.14}$$

The PA test rejects H_0 if $T_n(\theta) > c_0(\hat{\Omega}_n(\theta), 1-\alpha)$, where the PA critical value is

$$c_0(\hat{\Omega}_n(\theta), 1-\alpha) \equiv \inf\{x \in \mathbb{R} : \Pr(S(\hat{\Omega}_n(\theta)^{1/2}Z, \hat{\Omega}_n(\theta)) \leq x) \geq 1-\alpha\}, \tag{2.15}$$

and $Z \sim N(0_k, I_k)$ with Z independent of $\{W_i\}_{i=1}^n$.

We now define the GMS critical value introduced in Andrews and Soares (2010). To this end, let $\xi_n(\theta)$ be defined as $\xi_n(\theta) \equiv \kappa_n^{-1}\hat{D}_n^{-1/2}(\theta)n^{1/2}\bar{m}_n(\theta)$, for a sequence $\{\kappa_n\}_{n=1}^\infty$ of constants such that $\kappa_n \to \infty$ as $n \to \infty$ at a suitable rate, e.g., $\kappa_n = (2\ln\ln n)^{1/2}$. For every $j = 1, \ldots, p$, the realization $\xi_{n,j}(\theta)$ is an indication of whether the jth inequality is binding or not. A value of $\xi_{n,j}(\theta)$ that is close to zero (or negative) indicates that the jth inequality is likely to be binding. On the other hand, a value of $\xi_{n,j}(\theta)$ that is positive and large, indicates that the jth inequality may not be binding. As a result, GMS tests replace the parameter h_1 in the limiting distribution with the k-vector $\varphi(\xi_n(\theta), \hat{\Omega}_n(\theta))$, where $\varphi = (\varphi_1, \ldots, \varphi_p, 0_v)' \in \mathbb{R}^k_{[+\infty]}$ is a function chosen by the researcher that is assumed to satisfy

[7] We write $c_0(\Omega, 1-\alpha)$ rather than $c_{0_k}(\Omega, 1-\alpha)$ for ease of notation.

Assumption A.4 in the Appendix. Examples include $\varphi_j(\xi,\Omega) = \infty I(\xi_j > 1)$, where we use the convention $\infty 0 = 0$, $\varphi_j(\xi,\Omega) = \max\{\xi_j, 0\}$, and $\varphi_j(\xi,\Omega) = \xi_j$ for $j = 1,\ldots,p$ (see Andrews and Soares, 2010, for additional examples). The GMS test rejects H_0 if $T_n(\theta) > \hat{c}_{n,\kappa_n}(\theta, 1-\alpha)$, where the GMS critical value is

$$\hat{c}_{n,\kappa_n}(\theta, 1-\alpha) \equiv \inf\{x \in \mathbb{R} : \Pr(S(\hat{\Omega}_n^{1/2}(\theta)Z + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta)) \le x) \ge 1-\alpha\}, \quad (2.16)$$

and $Z \sim N(0_k, I_k)$ with Z independent of $\{W_i\}_{i=1}^n$.

Finally, we consider subsampling critical values (see Politis and Romano, 1994; Politis et al., 1999). Let b_n denote the subsample size when the sample size is n. Throughout the paper we assume $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$. The number of different subsamples of size b_n is q_n (with i.i.d. observations, $q_n = n!/((n-b_n)!b_n!)$). The subsample statistics used to construct the subsampling critical value are $\{T_{n,b_n,s}(\theta)\}_{s=1}^{q_n}$, where $T_{n,b_n,s}(\theta)$ is a subsample statistic defined exactly as $T_n(\theta)$ is defined but based on the sth subsample of size b_n rather than the full sample. The subsampling test rejects H_0 if $T_n(\theta) > \hat{c}_{n,b_n}(\theta, 1-\alpha)$, where the subsampling critical value is

$$\hat{c}_{n,b_n}(\theta, 1-\alpha) \equiv \inf\{x \in \mathbb{R} : U_{n,b_n}(\theta, x) \ge 1-\alpha\}, \quad (2.17)$$

and $U_{n,b_n}(\theta, x)$ denotes the empirical distribution function of $\{T_{n,b_n,s}(\theta)\}_{s=1}^{q_n}$.

Having introduced the different test statistics and critical values typically used in the literature, we devote the next section to the analysis of the asymptotic confidence size of the different CSs under the locally misspecified models introduced in Definition 2.1.

3 Distortions of Asymptotic Confidence Size

We divide this section into two parts. First, we take the test function S as given and compare how the resulting CSs based on PA, GMS, and subsampling critical values perform under local misspecification. In this case we write $AsySz_{PA}$, $AsySz_{GMS}$, and $AsySz_{SS}$ for PA, GMS, and subsampling CSs, respectively, to make explicit the choice of critical value. Second, we take the critical value as given and compare how CSs based on the test functions S_1 and S_2 perform under local misspecification. In this case we write $AsySz_l^{(1)}$ and $AsySz_l^{(2)}$, for $l \in \{PA, GMS, SS\}$, to denote the asymptotic confidence size of the CSs based on test functions S_1 and S_2, respectively.

3.1 Comparison across Critical Values

The following theorem presents the main result of this section, which provides a ranking of PA, GMS, and subsampling CSs in terms of asymptotic confidence size distortion. In order to keep the exposition as simple as possible, we present and discuss the assumptions and technical details in the Appendix.

Theorem 3.1. *Assume the parameter space is given by $\mathcal{F}_n$ in Eq. (2.5), $0 < \alpha < 1/2$, and that S satisfies Assumptions A.1-A.3. For GMS CSs assume that $\varphi(\xi, \Omega)$ satisfies Assumption A.4, and that $\kappa_n \to \infty$ and $\kappa_n^{-1} n^{1/2} \to \infty$. For subsampling CSs suppose $b_n \to \infty$ and $b_n/n \to 0$.*

1. *It follows that $AsySz_{PA} \geq AsySz_{SS}$ and $AsySz_{PA} \geq AsySz_{GMS}$. By further assuming Assumption A.6, we have that $AsySz_l < 1 - \alpha$ for all $l \in \{PA, GMS, SS\}$.*
2. *Suppose that Assumption A.5 holds and $\kappa_n^{-1} n^{1/2} / b_n^{1/2} \to \infty$. It then follows that $AsySz_{SS} = AsySz_{GMS}$.*

Assumptions A.1-A.4 are slight modifications of the corresponding assumptions in AG and Andrews and Soares (2010). Assumptions A.5 and A.6 are introduced in this paper and ensure that the parameter space is large enough. These two new assumptions are mild and we verify them for two leading examples in the Supplemental Appendix under weak primitive conditions. Under a reasonable set of assumptions the theorem implies

$$AsySz_{GMS} = AsySz_{SS} \leq AsySz_{PA} < 1 - \alpha. \tag{3.1}$$

This equation summarizes several important results. First, it shows that under local misspecification and mild conditions all of the inferential methods are size distorted, that is, as the sample size grows, CSs may under-cover the true parameter. Second, the equation reveals that PA CSs suffer the least amount of distortion in asymptotic confidence size. This is expected as this CS uses the most conservative critical value amongst GMS and subsampling, treating each inequality as binding without using information in the data.

Eq. (3.1) also shows that GMS and subsampling CSs suffer from the same amount of distortion in asymptotic confidence size. This result illustrates a situation where a robustness analysis provides discriminatory power between tests (used to construct CSs) that supplements the local power analysis. One key difference between local power and size robustness that provides important insight lies in the type of local sequences that a local power and a robustness comparison involve.[8] Andrews and Soares (2010) show that GMS tests are as powerful as subsampling tests along *any* sequence of local alternative parameters, and strictly more powerful than subsampling tests along *some* sequences of local alternative parameters. More precisely, for any local sequence of alternative models $\{\bar{\theta}_n, \bar{F}_n\}_{n \geq 1}$ it follows that

$$\lim_{n \to \infty} \Pr_{\bar{\theta}_n, \bar{F}_n} \left(T_n \left(\bar{\theta}_n \right) \leq \hat{c}_{n,\kappa_n} (\bar{\theta}_n, 1 - \alpha) \right) \leq \lim_{n \to \infty} \Pr_{\bar{\theta}_n, \bar{F}_n} \left(T_n \left(\bar{\theta}_n \right) \leq \hat{c}_{n,b_n} (\bar{\theta}_n, 1 - \alpha) \right), \tag{3.2}$$

and with strict inequality for *some* local sequences. As true parameter sequences in a misspecified model typically include sequences of local alternatives for correctly specified models, one might then suspect that this result would translate into GMS CS having strictly larger distortion in asymptotic size than the subsampling CS. However, Eq. (3.1) shows that this is

[8]For simplicity, in the following discussion we refer to sequences rather than subsequences.

not the case. Even though the GMS and subsampling CSs differ in their asymptotic behavior along certain sequences of locally misspecified models, these sequences turn out not to be the relevant ones for the computation of the asymptotic confidence size, i.e., the ones that attain the infimum in Eq. (1.4). If we let $\{\theta_n^*, F_n^*\}_{n\geq 1}$ denote a *worst sequence for GMS* and $\{\tilde{\theta}_n, \tilde{F}_n\}_{n\geq 1}$ denote a *worst sequence for subsampling*, our result in Theorem 3.1 states that

$$\lim_{n\to\infty} \Pr_{\theta_n^*, F_n^*}(T_n(\theta_n^*) \leq \hat{c}_{n,\kappa_n}(\theta_n^*, 1-\alpha)) = \lim_{n\to\infty} \Pr_{\tilde{\theta}_n, \tilde{F}_n}(T_n(\tilde{\theta}_n) \leq \hat{c}_{n,b_n}(\tilde{\theta}_n, 1-\alpha)). \quad (3.3)$$

Thus, along the sequences of locally misspecified models that minimize their respective limiting coverage probability, the two CSs share the value of the asymptotic confidence size. Note that both tests are compared along *the same* sequence in Eq. (3.2) while they are compared along possibly *different* sequences in Eq. (3.3). This is one of the reasons why the robustness properties of a test do not follow directly from its local power properties.

Combining Theorem 3.1 with the results regarding power against local alternatives in Andrews and Soares (2010) (and their implication for the expected volume of the corresponding CSs), our results indicate that GMS CSs are preferable to subsampling CSs: there can be a reduction in expected volume under correct specification without an associated increase in the distortion of asymptotic confidence size when the model is locally misspecified.

3.2 Comparison across Test Statistics

In this section we analyze the relative distortions in asymptotic confidence size of CSs based on the test functions S_1 and S_2 defined in Eqs. (2.9) and (2.11), respectively. The next theorem states the results formally.

Theorem 3.2. *Assume the parameter space is given by $\mathcal{F}_n$ in Eq.* (2.5) *and* $0 < \alpha < 1/2$. *For GMS CSs assume that $\varphi(\xi, \Omega)$ satisfies Assumption A.4 and that $\kappa_n \to \infty$ and $\kappa_n^{-1} n^{1/2} \to \infty$. For subsampling CSs suppose $b_n \to \infty$ and $b_n/n \to 0$.*

1. *There exists $B > 0$ such that whenever r in the definition of $\mathcal{F}_n$ in Eq.* (2.5) *satisfies $r^* \leq B$, we have $AsySz_l^{(1)} > 0$ for all $l \in \{PA, GMS, SS\}$.*
2. *Suppose that Assumption A.7 holds and that r in the definition of $\mathcal{F}_n$ is such that $r^* > 0$. Then, for every $\eta > 0$ there exists an $\varepsilon > 0$ in the definition of $\Psi_{2,\varepsilon}$ such that $AsySz_l^{(2)} \leq \eta$ for all $l \in \{PA, GMS, SS\}$.*

There are several important lessons from Theorem 3.2. First, it follows that the asymptotic confidence size of the CSs based on S_1 is positive for any critical value, provided the level of misspecification is not too big, i.e., $r^* \leq B$.[9] Second, it follows that the test function S_2 leads to CSs whose asymptotic confidence size is arbitrarily small when ε in $\Psi_{2,\varepsilon}$ is small enough. This is, the asymptotic confidence size of CSs based on the test function S_2 is severely affected by the smallest amount of misspecification while CSs based on the test

[9]Section S4 in the Supplemental Appendix contains additional intuition as to why this is required.

		$AsyCS^{(1)}_{PA}$	$AsyCS^{(2)}_{PA}$				$AsyCS^{(1)}_{PA}$	$AsyCS^{(2)}_{PA}$	
p	r^*		$\varepsilon = 0.10$	$\varepsilon = 0.05$	p	r^*		$\varepsilon = 0.10$	$\varepsilon = 0.05$
	0.25	88.8	63.3	40.4		0.25	84.1	72.3	63.1
2	0.50	80.8	12.4	0.8	8	0.50	64.2	26.7	8.4
	1.00	52.9	0.0	0.0		1.00	7.2	0.0	0.0
	0.25	87.4	57.7	21.6		0.25	82.5	71.7	65.9
4	0.50	74.7	5.9	0.0	10	0.50	59.2	29.9	11.8
	1.00	22.8	0.0	0.0		1.00	5.6	0.0	0.0

Table 1: Asymptotic Confidence Size (in %) for CSs based on the test functions S_1 and S_2 with a PA critical value and $\alpha = 5\%$. The numbers above were computed using the explicit formula for $AsySz$ provided in Eq. (B-1) of the Supplemental Appendix and the infimum with respect to Ω for S_1 and S_2 was carried out by minimizing over 15000 random correlation matrices in Ψ_1 and $\Psi_{2,\varepsilon}$, respectively.

function S_1 have positive asymptotic confidence size. Combining these two results, it follows that there exists $B > 0$ and $\varepsilon > 0$ in $\Psi_{2,\varepsilon}$ such that whenever $r^* \in (0, B]$,

$$AsySz^{(2)}_l < AsySz^{(1)}_l, \quad l \in \{PA, GMS, SS\}. \tag{3.4}$$

It is known from Andrews and Jia (2008) that tests based on S_2 have higher power than tests based on S_1, so intuition suggests that Eq. (3.4) should hold. However, Theorem 3.2 goes beyond this observation by showing that the cost of having a smaller expected volume under correct specification for CSs based on S_2 can be an arbitrarily low asymptotic confidence size under local misspecification.

Remark 3.1. Under certain conditions, the generalized empirical likelihood test statistics are asymptotically equivalent to $T_{2,n}(\theta)$ up to first order (see AG and Canay (2010)), and so the asymptotic confidence size of CSs based on such test statistics is equal to $AsySz^{(2)}_l$ in Theorem 3.2.

Theorem 3.2 presents an analytical result regarding the relative amount of distortion in asymptotic confidence size for different test functions. We now quantify these results by numerically computing the asymptotic confidence size of the CSs based on S_1 and S_2 using the formulas provided in Lemma B.1. Table 1 reports the cases where $p \in \{2, 4, 8, 10\}$, $k = p$, $\varepsilon \in \{0.10, 0.05\}$, and $r^* \in \{0.25, 0.50, 1.00\}$. Table 1 shows that the asymptotic confidence size of CSs based on S_2 is significantly distorted even for relatively high values of ε (i.e. $\varepsilon = 0.10$). For example, when $p = 2$ and $r^* = 0.5$, the asymptotic confidence size for the test function S_1 is 80.8% while the asymptotic confidence size for S_2 is 12.4% or lower. As suggested in Theorem 3.2, the asymptotic confidence size for S_2 is always significantly below the one for S_1 and very close to zero for $r^* \geq 0.50$.[10]

Two aspects related to the second part of Theorem 3.2 are worth mentioning. First, if we modify the test function S_2 in order to admit any matrix in the space of all correlation

[10]In Table 1 the asymptotic confidence size decreases as p grows. This is clear for S_1 but less clear for S_2. The reason is that finding the worst possible correlation matrix becomes substantially more complicated as the dimension increases, and so for $p \geq 8$ the results reported are relatively optimistic for S_2. The Supplemental Appendix explains this computations in detail.

matrices Ψ_1 (even singular ones) the result still holds. This is, suppose that for $\varepsilon > 0$ we define the test function

$$\tilde{S}_{2,\varepsilon}(m,\Sigma) = \inf_{t=(t_1,0_v):t_1\in\mathbb{R}^p_{+,+\infty}} (m-t)'\tilde{\Sigma}_\varepsilon^{-1}(m-t), \tag{3.5}$$

where $\tilde{\Sigma}_\varepsilon = \Sigma + \max\{\varepsilon - \det(\Omega), 0\}D$, $D = Diag(\Sigma)$, and $\Omega = D^{-1/2}\Sigma D^{-1/2}$. The function $\tilde{S}_{2,\varepsilon}$ is well defined on Ψ_1 and leads to the test statistic

$$\tilde{T}_{2,\varepsilon,n}(\theta) = \inf_{t=(t_1,0_v):t_1\in\mathbb{R}^p_{+,+\infty}} (n^{1/2}\bar{m}_n(\theta)-t)'\tilde{\Sigma}_{\varepsilon,n}(\theta)^{-1}(n^{1/2}\bar{m}_n(\theta)-t), \tag{3.6}$$

where $\tilde{\Sigma}_{\varepsilon,n}(\theta)$ is a consistent estimator of $\tilde{\Sigma}_\varepsilon$. This new test function coincides with S_2 when the determinant of the correlation matrix is at least ε, but it changes the weighting matrix when Ω is singular or close to singular. By construction, $\tilde{\Sigma}_\varepsilon$ has a determinant bounded away from zero. Letting $AsySz^{(\tilde{2},\varepsilon)}$ denote the asymptotic confidence size of CSs based on $\tilde{S}_{2,\varepsilon}$, the next corollary to Theorem 3.2 follows.

Corollary 3.1. *Suppose the assumptions in Theorem 3.2 hold and that $r^* > 0$. Then, for every $\eta > 0$ there exists an $\varepsilon > 0$ in the definition of $\tilde{S}_{2,\varepsilon}$ such that $AsySz_l^{(\tilde{2},\varepsilon)} \leq \eta$ for all $l \in \{PA, GMS, SS\}$.*

Second, Assumption A.7 is sufficient but not necessary for the result in Theorem 3.2 when $p > 2$. Assumption A.7 requires that at least one inequality moment restriction in Eq. (1.1) is violated and strongly negatively correlated with another inequality moment restriction that is either violated or equal to zero. When $p = 2$ it can be shown that this is actually a necessary condition to obtain the second part in Theorem 3.2. In the general case, there are alternative ways to make the parameter space large enough,[11] but Assumption A.7 has the additional advantage of making the optimization problem in Eq. (2.11) tractable. Having said this, we interpret the second part of Theorem 3.2 as a *warning* message. Unless the researcher is certain that it is impossible for inequality moment restrictions that are violated to be strongly negatively correlated with each other or with other inequality moment restrictions that are binding, the asymptotic confidence size of CSs based on S_2 could be extremely distorted.

4 Conclusion

This paper studies the behavior under local misspecification of several CSs commonly used in the literature on inference in moment inequality models. The paper proposes to use the amount of distortion in asymptotic confidence size as a criterion to choose among competing inference methods and shows that such criterion may provide additional discriminatory power to supplement local asymptotic power comparisons. In particular, we show that CSs based on

[11] In Examples 2.1 and S3.1 there are two inequality moment restrictions that are restricted in a way that when one is negative, the other one is necessarily positive. However, in Examples 2.1 this restriction is no longer present when there are more than two firms and the model includes additional covariates.

subsampling and GMS critical values suffer from the same level of asymptotic size distortion, despite the fact that the latter can lead to CSs with strictly smaller expected volume under correct model specification. We also show that the asymptotic confidence size of CSs based on the quasi-likelihood ratio test statistic can be an arbitrary small fraction of the asymptotic confidence size of CSs based on the modified method of moments test statistic.

Appendix A Additional Definitions and Assumptions

To determine the asymptotic confidence size in Eq. (1.4) we calculate the limiting coverage probability along a sequence of "worst case parameters" $\{\theta_n, F_n\}_{n\geq 1}$ with $(\theta_n, F_n) \in \mathcal{F}_n, \forall n \in \mathbb{N}$. See also Andrews and Guggenberger (2009a,b,2010a,b). We start with the following definition. Note that any Lemma or Equation that starts with the letter "S" is included in the Supplemental Appendix.

Definition A.1. For a subsequence $\{\omega_n\}_{n\geq 1}$ of $\mathbb{N}$ and $h = (h_1, h_2) \in \mathbb{R}^k_{+\infty} \times \Psi$ we denote by

$$\gamma_{\omega_n,h} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}, \tag{A-1}$$

a sequence that satisfies (i) $\gamma_{\omega_n,h} \in \mathcal{F}_{\omega_n}$ for all n, (ii) $\omega_n^{1/2}\sigma^{-1}_{F_{\omega_n,h},j}(\theta_{\omega_n,h})E_{F_{\omega_n,h}}m_j(W_i,\theta_{\omega_n,h}) \to h_{1,j}$ for $j = 1,\ldots,k$, and (iii) $Corr_{F_{\omega_n,h}}(m(W_i,\theta_{\omega_n,h})) \to h_2$ as $n\to\infty$, if such a sequence exists. Denote by H the set of points $h = (h_1, h_2) \in \mathbb{R}^k_{+\infty} \times \Psi$ for which sequences $\{\gamma_{\omega_n,h}\}_{n\geq 1}$ exist.

Denote by GH the set of points $(g_1, h) \in \mathbb{R}^k_{+\infty} \times H$ such that there is a subsequence $\{\omega_n\}_{n\geq 1}$ of $\mathbb{N}$ and a sequence $\{\gamma_{\omega_n,h}\}_{n\geq 1}$ that satisfies[12]

$$b^{1/2}_{\omega_n}\sigma^{-1}_{F_{\omega_n,h},j}(\theta_{\omega_n,h})E_{F_{\omega_n,h}}m_j(W_i,\theta_{\omega_n,h}) \to g_{1,j} \tag{A-2}$$

for $j = 1,\ldots,k$, where $g_1 = (g_{1,1},\ldots,g_{1,k})$. Denote such a sequence by $\{\gamma_{\omega_n,g_1,h}\}_{n\geq 1}$.

Denote by ΠH the set of points $(\pi_1, h) \in \mathbb{R}^k_{+\infty} \times H$ such that there is a subsequence $\{\omega_n\}_{n\geq 1}$ of $\mathbb{N}$ and a sequence $\{\gamma_{\omega_n,h}\}_{n\geq 1}$ that satisfies

$$\kappa^{-1}_{\omega_n}\omega_n^{1/2}\sigma^{-1}_{F_{\omega_n,h},j}(\theta_{\omega_n,h})E_{F_{\omega_n,h}}m_j(W_i,\theta_{\omega_n,h}) \to \pi_{1,j} \tag{A-3}$$

for $j = 1,\ldots,k$, where $\pi_1 = (\pi_{1,1},\ldots,\pi_{1,k})$. Denote such a sequence by $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$.

Our assumptions imply that elements of H satisfy certain properties. For example, for any $h \in H$, h_1 is constrained to satisfy $h_{1,j} \geq -r_j$ for $j = 1,\ldots,p$ and $|h_{1,j}| \leq r_j$ for $j = p+1,\ldots,k$, and h_2 is a correlation matrix. Note that the set H depends on the choice of S through Ψ. Note that $b/n \to 0$ implies that if $(g_1, h) \in GH$ and $h_{1,j}$ is finite $(j = 1,\ldots,k)$, then $g_{1,j} = 0$. In particular, $g_{1,j} = 0$ for $j > p$ by Eq. (2.5)(iii). Analogous statements hold for ΠH. Finally, the spaces H, GH, and ΠH for a hypothesis testing problem (see Remark 2.3) are defined analogously for a sequence $\gamma_{\omega_n,h} = \{\theta, F_{\omega_n,h}\}_{n\geq 1}$ where θ is fixed at the hypothesized value.

Lemma B.1 in the next section shows that worst case parameter sequences for PA, GMS, and subsampling CSs are of the type $\{\gamma_{n,h}\}_{n\geq 1}$, $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$, and $\{\gamma_{\omega_n,g_1,h}\}_{n\geq 1}$, respectively, and provides explicit formulas for the asymptotic confidence size of various CSs.

Definition A.2. For $h = (h_1, h_2)$, let $J_h \sim S(h_2^{1/2}Z + h_1, h_2)$, where $Z = (Z_1,\ldots,Z_k) \sim N(0_k, I_k)$. The $1-\alpha$ quantile of J_h is denoted by $c_{h_1}(h_2, 1-\alpha)$.

Note that $c_0(h_2, 1-\alpha)$ is the $1-\alpha$ quantile of the asymptotic null distribution of $T_n(\theta)$ when the moment inequalities hold as equalities and the moment equalities are satisfied.

[12]The definitions of the sets H and GH differ somewhat from the ones given in AG. In particular, in AG, GH is defined as a subset of $H \times H$ whereas here h_2 is not repeated. Also, the dimension of h_2 in AG is smaller than here as $vech_*(h_2)$ is replaced by h_2. We adopt this convention in order to simplify the notation.

The following Assumptions A.1-A.3 are taken from AG with Assumption 2 slightly strengthened. Assumption A.4(a)-(c) combines Assumptions GMS1 and GMS3 in Andrews and Soares (2010). Assumptions A.5-A.7 are new.

Assumption A.1. *The test function S satisfies*

(a) $S\left(\left(m_1, m_1^*\right), \Sigma\right)$ *is non-increasing in* m_1, $\forall (m_1, m_1^*) \in \mathbb{R}^p \times \mathbb{R}^v$ *and matrices* $\Sigma \in \mathcal{V}_{k\times k}$*,*

(b) $S\left(m, \Sigma\right) = S\left(\Delta m, \Delta\Sigma\Delta\right)$ *for all* $m \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k\times k}$*, and positive definite diagonal matrix* $\Delta \in \mathbb{R}^{k\times k}$*,*

(c) $S\left(m, \Omega\right) \geq 0$ *for all* $m \in \mathbb{R}^k$ *and* $\Omega \in \Psi$*, and*

(d) $S\left(m, \Omega\right)$ *is continuous at all* $m \in \mathbb{R}^p_{+\infty} \times \mathbb{R}^v$ *and* $\Omega \in \Psi$*.*

Assumption A.2. *For all* $h_1 \in [-r_j, \infty]^p_{j=1} \times [-r_j, r_j]^k_{j=p+1}$*, all* $\Omega \in \Psi$*, and* $Z \sim N\left(0_k, \Omega\right)$*, the distribution function (df) of* $S\left(Z + h_1, \Omega\right)$ *at* $x \in \mathbb{R}$ *is*

(a) continuous for $x > 0$*,*

(b) strictly increasing for $x > 0$ *unless* $p = k$ *and* $h_1 = \infty_p$*, and*

(c) less than or equal to 1/2 at $x = 0$ *when* $v \geq 1$ *or when* $v = 0$ *and* $h_{1,j} = 0$ *for some* $j = 1, \ldots, p$*.*

Assumption A.3. $S(m, \Omega) > 0$ *if and only if* $m_j < 0$ *for some* $j = 1, \ldots, p$*, or* $m_j \neq 0$ *for some* $j = p + 1, \ldots, k$*, where* $m = (m_1, \ldots, m_k)$ *and* $\Omega \in \Psi$*.*

Assumption A.4. *Let* $\xi = (\xi_1, \ldots, \xi_k)$*. For* $j = 1, \ldots, p$ *we have:*

(a) $\varphi_j(\xi, \Omega)$ *is continuous at all* $(\xi, \Omega) \in (\mathbb{R}^p_{+,+\infty} \times \mathbb{R}^v_{\pm\infty}) \times \Psi$ *for which* $\xi_j \in \{0, \infty\}$*.*

(b) $\varphi_j(\xi, \Omega) = 0$ *for all* $(\xi, \Omega) \in (\mathbb{R}^p_{+,+\infty} \times \mathbb{R}^v_{\pm\infty}) \times \Psi$ *with* $\xi_j = 0$*.*

(c) $\varphi_j(\xi, \Omega) = \infty$ *for all* $(\xi, \Omega) \in (\mathbb{R}^p_{+,+\infty} \times \mathbb{R}^v_{\pm\infty}) \times \Psi$ *with* $\xi_j = \infty$*.*

(d) $\varphi_j(\xi, \Omega) \geq 0$ *for all* $(\xi, \Omega) \in (\mathbb{R}^p_{+,+\infty} \times \mathbb{R}^v_{\pm\infty}) \times \Psi$ *with* $\xi_j \geq 0$*.*

Assumption A.5. *For any sequence* $\{\gamma_{\omega_n, g_1, h}\}_{n\geq 1}$ *in Definition A.1 there exists a subsequence* $\{\tilde{\omega}_n\}_{n\geq 1}$ *of* $\mathbb{N}$ *and a sequence* $\{\gamma_{\tilde{\omega}_n, \tilde{g}_1, h}\}_{n\geq 1}$ *such that* $\tilde{g}_1 \in \mathbb{R}^k_{+\infty}$ *satisfies* $\tilde{g}_{1,j} = \infty$ *when* $h_{1,j} = \infty$ *for* $j = 1, \ldots, p$*.*

Assumption A.6. *There exists* $h^* = (h_1^*, h_2^*) \in H$ *for which* $J_{h^*}(c_0(h_2^*, 1 - \alpha)) < 1 - \alpha$*.*

Assumption A.7. *Let* $\Xi_{l,l'}(\varepsilon) \in \mathbb{R}^{k\times k}$ *be an identity matrix except for the* (l, l') *and* (l', l) *components that are equal to* $-\sqrt{1 - \varepsilon}$ *for some* $l, l' \in \{1, \ldots, p\}$*. There exists* $h \in H$ *such that* $h_{1,l} \leq 0$*,* $h_{1,l'} \leq 0$*,* $\min\{h_{1,l}, h_{1,l'}\} < 0$*, and* $h_2 = \Xi_{l,l'}(\varepsilon)$ *for some* $l, l' \in \{1, \ldots, p\}$ *with* $l \neq l'$*.*

Assumption 4 in AG is not imposed because it is implied by the other assumptions in our paper. More specifically, note that by Assumption A.1(c) $c_0(\Omega, 1 - \alpha) \geq 0$. Also, $h_1 = 0_v$ and Assumption A.2(c) imply that the df of $S(Z, \Omega)$ is less than 1/2 at $x = 0$, which implies $c_0(\Omega, 1 - \alpha) > 0$ for $\alpha < 1/2$. Then, Assumption A.2(a) implies Assumption 4(a) in AG. Regarding Assumption 4(b) in AG, note that it is enough to establish pointwise continuity of $c_0(\Omega, 1 - \alpha)$ because by assumption Ψ is a closed set and trivially bounded. In fact, we can prove pointwise continuity of $c_{h_1}(\Omega, 1 - \alpha)$ even for a vector h_1 with $h_{1,j} = 0$ for at last one $j = 1, \ldots, k$. To do so, consider a sequence $\{\Omega_n\}_{n\geq 1}$ such that $\Omega_n \to \Omega$ for a $\Omega \in \Psi$ and a vector h_1 with $h_{1,j} = 0$ for at last one $j = 1, \ldots, k$. We need to show that $c_{h_1}(\Omega_n, 1 - \alpha) \to c_{h_1}(\Omega, 1 - \alpha)$. Let Z_n and Z be normal zero mean random vectors with covariance matrix equal to Ω_n and Ω, respectively. By Assumption A.1(d) and the continuous mapping theorem we have $S(Z_n + h_1, \Omega_n) \to_d S(Z + h_1, \Omega)$. The latter implies that $\Pr(S(Z_n + h_1, \Omega_n) \leq x) \to \Pr(S(Z + h_1, \Omega) \leq x)$ for all continuity points $x \in \mathbb{R}$ of the function $f(x) \equiv \Pr(S(Z + h_1, \Omega) \leq x)$. The convergence therefore certainly holds for all $x > 0$ by Assumption A.2(a). Furthermore, by Assumption A.2(b) f is strictly increasing for $x > 0$. By Assumption A.2(c) and $\alpha < 1/2$ it follows that $c_{h_1}(\Omega, 1-\alpha) > 0$. By an argument used in Lemma 5(a) in AG, it then follows that $c_{h_1}(\Omega_n, 1-\alpha) \to c_{h_1}(\Omega, 1-\alpha)$.

Note that S_1 and S_2 satisfy Assumption A.2 which is a strengthened version of Assumption 2 in AG. Assumption A.3 implies that $S(\infty_p, \Omega) = 0$ when $v = 0$. Assumption A.5 makes sure the parameter space is sufficiently rich. Assumption A.6 holds by Assumption A.2(a) if there exists $h^* \in H$ such that $J_{h^*}(c_0(h_2^*, 1-\alpha)) < J_{(0,h_2^*)}(c_0(h_2^*, 1-\alpha))$. Also note that by Assumption A.1(a), a $h^* \in H$ as in Assumption A.6 needs to have $h_{1,j}^* < 0$ for some $j \le p$ or $h_{1,j}^* \neq 0$ for some $j > p$. Assumptions A.5 and A.6 are verified for the two lead example in Appendix S4. Assumption A.7 guarantees two things. First, it guarantees that at least two inequalities in Eq. (1.1) are violated (or at least, one is violated and the other one is binding) and negatively correlated. Second, it guarantees that there are correlation matrices with zeros outside the diagonal except at two spots. This second part of the assumption simplifies the proof significantly but it could be replaced with alternative forms of correlation matrices.

Appendix B Proof of the Theorems and Main Lemma

Lemma B.1. *Consider CSs with nominal confidence size* $1-\alpha$ *for* $0 < \alpha < 1/2$. *Assume the nonempty parameter space is given by* $\mathcal{F}_n$ *in Eq. (2.5) for some* $r \in \mathbb{R}_+^k$, $\delta > 0$, *and* $M < \infty$. *Assume* S *satisfies Assumptions A.1-A.3. For GMS CSs assume that* $\varphi(\xi, \Omega)$ *satisfies Assumption A.4,* $\kappa_n \to \infty$, *and* $\kappa_n^{-1} n^{1/2} \to \infty$. *For subsampling CSs suppose* $b_n \to \infty$ *and* $b_n/n \to 0$. *It follows that*

$$\begin{aligned} AsySz_{PA} &= \inf_{h=(h_1,h_2)\in H} J_h(c_0(h_2, 1-\alpha)), \\ AsySz_{GMS} &\in \left[\inf_{(\pi_1,h)\in\Pi H} J_h(c_{\pi_1^*}(h_2, 1-\alpha)), \inf_{(\pi_1,h)\in\Pi H} J_h(c_{\pi_1^{**}}(h_2, 1-\alpha)) \right], \text{ and} \\ AsySz_{SS} &= \inf_{(g_1,h)\in GH} J_h(c_{g_1}(h_2, 1-\alpha)), \end{aligned} \tag{B-1}$$

where $J_h(x) = P(J_h \le x)$ *and* π_1^*, $\pi_1^{**} \in \mathbb{R}_{+\infty}^k$ *with its* j*th element defined by* $\pi_{1,j}^* = \infty I(\pi_{1,j} > 0)$ *and* $\pi_{1,j}^{**} = \infty I(\pi_{1,j} = \infty)$, $j = 1, \ldots, k$.

Proof of Lemma B.1. For any of the CSs considered in Section 2.1, there is a sequence $\{\theta_n, F_n\}_{n\ge1}$ with $(\theta_n, F_n) \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$ such that $AsySz = \liminf_{n\to\infty} \Pr_{\theta_n,F_n}(T_n(\theta_n) \le c_n(\theta_n, 1-\alpha))$. We can then find a subsequence $\{\omega_n\}_{n\ge1}$ of $\mathbb{N}$ such that

$$AsySz = \lim_{n\to\infty} \Pr_{\theta_{\omega_n},F_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n}) \le c_{\omega_n}(\theta_{\omega_n}, 1-\alpha)) \tag{B-2}$$

and condition (i) in Definition A.1 holds. Conditions (ii)-(iii) in Definition A.1 also hold for $\{\theta_{\omega_n}, F_{\omega_n}\}_{n\ge1}$ by possibly taking a further subsequence. That is, $\{\theta_{\omega_n}, F_{\omega_n}\}_{n\ge1}$ is a sequence of type $\{\gamma_{\omega_n,h}\}_{n\ge1} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\ge1}$ for a certain $h = (h_1, h_2) \in \mathbb{R}_{+\infty}^k \times \Psi$. For GMS and SS CSs, we can find subsequences $\{\omega_n\}_{n\ge1}$ (potentially different for GMS and SS CSs) such that the worst case sequence $\{\theta_{\omega_n}, F_{\omega_n}\}_{n\ge1}$ is of the type $\{\gamma_{\omega_n,\pi_1,h}\}_{n\ge1}$ or $\{\gamma_{\omega_n,g_1,h}\}_{n\ge1}$.

Therefore, in order to determine the asymptotic confidence size of the CSs we only have to consider the limiting coverage probabilities under sequences of the type $\{\gamma_{\omega_n,h}\}_{n\ge1}$ for PA, $\{\gamma_{\omega_n,\pi_1,h}\}_{n\ge1}$ for GMS, and $\{\gamma_{\omega_n,g_1,h}\}_{n\ge1}$ for SS. From Lemma S1.1 in the Supplement, the limiting distribution of the test statistic under a sequence $\{\gamma_{\omega_n,h}\}_{n\ge1}$ is $J_h \sim S(Z_{h_2} + h_1, h_2)$. By Assumption A.1(a), for given h_2 the $1-\alpha$ quantile of J_h does not decrease as $h_{1,j}$ decreases (for $j = 1, \ldots, p$).

PA critical value: The PA critical value is given by $c_0(\hat{h}_{2,\omega_n}, 1-\alpha)$, where

$$\hat{h}_{2,\omega_n} = \hat{\Omega}_{\omega_n}(\theta_{\omega_n,h}) \tag{B-3}$$

and $\hat{\Omega}_s(\theta) = (\hat{D}_s(\theta))^{-1/2} \hat{\Sigma}_s(\theta) (\hat{D}_s(\theta))^{-1/2}$. From Eq. (S2.2)(iii) we know that under $\{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\ge1}$, we have $\hat{h}_{2,\omega_n} \to_p h_2$. This together with Assumption A.1 implies $c_0(\hat{h}_{2,\omega_n}, 1-\alpha) \to_p c_0(h_2, 1-\alpha)$. Furthermore, by Assumption A.2(c), $c_0(h_2, 1-\alpha) > 0$ and by Assumption A.2(a), J_h is continuous for $x > 0$. Using the proof of Lemma 5(ii) in AG (and its subsequent comments), we have $\Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n}) \leq c_0(\hat{h}_{2,\omega_n}, 1-\alpha)) \to J_h(c_0(h_2, 1-\alpha))$ and therefore also $\lim_{n\to\infty} \Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n}) \leq c_0(\hat{h}_{2,\omega_n}, 1-\alpha)) = J_h(c_0(h_2, 1-\alpha))$. As a result, $AsySz_{PA} = J_h(c_0(h_2, 1-\alpha))$ for some $h \in H$, which implies $AsySz_{PA} \geq \inf_{h\in H} J_h(c_0(h_2, 1-\alpha))$. However, Eq. (B-2) implies that $AsySz_{PA} = \inf_{h\in H} \lim_{n\to\infty} \Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n}) \leq c_0(\hat{h}_{2,\omega_n}, 1-\alpha))$. This expression equals $\inf_{h=(h_1,h_2)\in H}\ J_h(c_0(h_2, 1-\alpha))$, completing the proof.

GMS critical value: To simplify notation, we write $\{\gamma_{\omega_n}\} = \{\theta_{\omega_n}, F_{\omega_n}\}$ instead of $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1} = \{\theta_{\omega_n,\pi_1,h}, F_{\omega_n,\pi_1,h}\}_{n\geq 1}$. Recall that the GMS critical value $\hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)$ is the $1-\alpha$ quantile of $S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n})$ for $Z \sim N(0_k, I_k)$. We first show the existence of random variables $c^*_{\omega_n}$ and $c^{**}_{\omega_n}$ such that under $\{\gamma_{\omega_n}\}$

$$\begin{aligned} \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha) &\geq c^*_{\omega_n} \to_p c_{\pi_1^*}(h_2, 1-\alpha), \\ \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha) &\leq c^{**}_{\omega_n} \to_p c_{\pi_1^{**}}(h_2, 1-\alpha). \end{aligned} \tag{B-4}$$

We begin by showing the first line in Eq. (B-4). Suppose $c_{\pi_1^*}(h_2, 1-\alpha) = 0$, then, $\hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha) \geq 0 = c_{\pi_1^*}(h_2, 1-\alpha)$ under $\{\gamma_{\omega_n}\}_{n\geq 1}$ by Assumption A.1(c). Now suppose $c_{\pi_1^*}(h_2, 1-\alpha) > 0$. For given $\pi_1 \in \mathbb{R}^k_{+,\infty}$ and for $(\xi, \Omega) \in \mathbb{R}^k \times \Psi$ let $\varphi^*(\xi, \Omega)$ be the k-vector with jth component given by

$$\varphi^*_j(\xi,\Omega) = \begin{cases} \varphi_j(\xi,\Omega) & \text{if } \pi_{1,j} = 0 \text{ and } j \leq p, \\ \infty & \text{if } \pi_{1,j} > 0 \text{ and } j \leq p, \\ 0 & \text{if } j = p+1,\ldots,k. \end{cases} \tag{B-5}$$

Define $c^*_{\omega_n}$ as the $1-\alpha$ quantile of $S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi^*(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n})$. As $\varphi^*_j \geq \varphi_j$ it follows from Assumption A.1(a) that $c^*_{\omega_n} \leq \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)$ a.s. $[Z]$ under $\{\gamma_{\omega_n}\}_{n\geq 1}$. Furthermore, by Lemma 2(a) in the Supplemental Appendix of Andrews and Soares (2010) we have $c^*_{\omega_n} \to_p c_{\pi_1^*}(h_2, 1-\alpha)$ under $\{\gamma_{\omega_n}\}_{n\geq 1}$. This completes the proof of the first line in Eq. (B-4).

Now consider the second line in Eq. (B-4). Suppose either $v \geq 1$ or $v = 0$ and $\pi_1^{**} \neq \infty_p$. Define

$$\varphi^{**}_j(\xi,\Omega) = \begin{cases} \min\{0, \varphi_j(\xi,\Omega)\} & \text{if } \pi_{1,j} < \infty \text{ and } j \leq p, \\ \varphi_j(\xi,\Omega) & \text{if } \pi_{1,j} = \infty \text{ and } j \leq p, \\ 0 & \text{if } j = p+1,\ldots,k, \end{cases} \tag{B-6}$$

and define $c^{**}_{\omega_n}$ as the $1-\alpha$ quantile of $S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi^{**}(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n})$. Note that the definition of $\varphi^{**}_j(\xi,\Omega)$ implies that $\varphi^{**}_j \leq \varphi_j$. The same steps as in the proof of (Andrews and Soares, 2010, Lemma 2(a)) can be used to prove the second line of Eq. (B-4). In particular, by Assumption A.4 $\varphi^{**}(\xi,\Omega) \to \varphi^{**}(\pi_1, \Omega_0)$ for any sequence $(\xi,\Omega) \in \mathbb{R}^k_{+\infty} \times \Psi$ for which $(\xi,\Omega) \to (\pi_1, \Omega_0)$ and $\Omega_0 \in \Psi$.

Suppose now that $v = 0$ and $\pi_1^{**} = \infty_p$. It follows that $c_{\pi_1^{**}}(h_2, 1-\alpha) = 0$ by Assumption A.3 and that $\pi_1 = \infty_p$. In that case define $c^{**}_{\omega_n} = \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)$ which converges to zero in probability because by Assumption A.3, $\pi_1 = \infty_p$, and by Assumption A.4, $0 \leq S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n}) \to_p 0$. This implies the second line in Eq. (B-4).

Having proven Eq. (B-4), we now prove the second line in Eq. (B-1). Consider first the case $(\pi_1, h) \in \Pi H$ such that $c_{\pi_1^*}(h_2, 1-\alpha) > 0$. It then follows from Eq. (B-4) and Lemma 5 in AG that

$$\begin{aligned} \liminf_{n\to\infty} \Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n}) \leq \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)) &\leq \liminf_{n\to\infty} \Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n}) \leq c^{**}_{\omega_n}) \\ &= J_h(c_{\pi_1^{**}}(h_2, 1-\alpha)). \end{aligned} \tag{B-7}$$

Likewise $\liminf_{n\to\infty} \Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n}) \leq \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)) \geq J_h(c_{\pi_1^*}(h_2, 1-\alpha))$.

Next consider the case $(\pi_1, h) \in \Pi H$ such that $c_{\pi_1^*}(h_2, 1-\alpha) = 0$. By Assumption A.2(c) and $\alpha < 1/2$, this implies $v = 0$ and $\pi^*_{1,j} > 0$ for all $j = 1,\ldots,p$. By definition of π_1^*, it follows that $\pi_{1,j} > 0$ for all $j = 1,\ldots,p$ and so, $\kappa_n \to \infty$ implies $h_1 = \infty_p$. Under any sequence $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$

with $h = (\infty_p, h_2)$ we have

$$1 \geq \liminf_{n\to\infty} \Pr_{\gamma_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n}) \leq \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)) \geq \liminf_{n\to\infty} \Pr_{\gamma_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n}) \leq 0) = J_h(0) = 1, \quad \text{(B-8)}$$

where we apply the argument in Eq. (A.12) of AG for the first equality and use Assumption A.3 for the second equality. Therefore, $\liminf_{n\to\infty} \Pr_{\gamma_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n}) \leq \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)) = 1$. Note that when $h_1 = \infty_p$, $J_h(c) = 1$ for any $c \geq 0$. The last statement and Eqs. (B-2), (B-7), and (B-8) complete the proof of the lemma.

Subsampling critical value: Instead of $\{\gamma_{\omega_n,g_1,h}\}_{n\geq 1} = \{\theta_{\omega_n,g_1,h}, F_{\omega_n,g_1,h}\}_{n\geq 1}$ we write $\{\gamma_{\omega_n}\} = \{\theta_{\omega_n}, F_{\omega_n}\}$ to simplify notation. We first verify Assumptions A0, B0, C, D, E0, F, and G0 in AG. Following AG, define a vector of (nuisance) parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ where $\gamma_3 = (\theta, F)$, $\gamma_1 = \{\sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta)\}_{j=1}^k \in \mathbb{R}^k$, and $\gamma_2 = Corr_F(m(W_i, \theta)) \in \mathbb{R}^{k\times k}$ for (θ, F) introduced in the model defined in (2.5). Then, Assumption A0 in AG clearly holds. With $\{\gamma_{\omega_n,h}\}_{n\geq 1}$ and H defined in definition A.1, Assumption B0 then holds by Lemma S1.1. Assumption C holds by assumption on the subsample blocksize b. Assumptions D, E0, F, and G0 hold by the same argument as in AG using the strengthened version of Assumption A.2(b) and (c) for the argument used to verify Assumption F. Therefore, Theorem 3(ii) in AG applies with their GH replaced by our GH and their GH^* (defined on top of Eq. (9.4) in AG) which is the set of points $(g_1, h) \in GH$ such that for all sequences $\{\gamma_{w_n,g_1,h}\}_{n\geq 1}$

$$\liminf_{n\to\infty} \Pr_{\gamma_{w_n,g_1,h}}(T_{w_n}(\theta_{w_n,g_1,h}) \leq c_{w_n,b_{w_n}}(\theta_{w_n,g_1,h}, 1-\alpha)) \geq J_h(c_{g_1}(h_2, 1-\alpha)). \quad \text{(B-9)}$$

By Theorem 3(ii) in AG and continuity of J_h at positive arguments, it is then enough to show that the set $\{(g_1, h) \in GH\backslash GH^*; c_{g_1}(h_2, 1-\alpha) = 0\}$ is empty. To show this, note that by Assumption A.2(c) $c_{g_1}(h_2, 1-\alpha) = 0$ implies that $v = 0$ and by Assumption A.1(a) it follows that $c_{h_1}(h_2, 1-\alpha) = 0$. Using the same argument as in AG, namely the paragraph including Eq. (A.12) with their LB_h equal to 0, shows that any $(g_1, h) \in GH$ with $c_{g_1}(h_2, 1-\alpha) = 0$ is also in GH^*. □

Proof of Theorem 3.1. Part 1. Note that for $h \in H$ and $\kappa_n \to \infty$, there exists a subsequence $\{\omega_n\}_{n\geq 1}$ and a sequence $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$ for some $\pi_1 \in \mathbb{R}_\infty^k$ with $\pi_{1,j} \geq 0$ for $j = 1, \ldots, p$ and $\pi_{1,j} = 0$ for $j = p+1, \ldots, k$. By definition, $\pi_1^{**} \geq 0$. Assumption A.1(a) then implies that $c_0(h_2, 1-\alpha) \geq c_{\pi_1^{**}}(h_2, 1-\alpha)$ and so $AsySz_{PA} \geq AsySz_{GMS}$. The result for subsampling CSs is analogous. Finally, note that $AsySz_{PA} = \inf_{h=(h_1,h_2)\in H} J_h(c_0(h_2, 1-\alpha)) \leq J_{h^*}(c_0(h_2^*, 1-\alpha)) < 1-\alpha$.

Part 2. First, assume $(g_1, h) \in GH$. By Assumption A.1(a), $AsySz_{SS} \geq AsySz_{GMS}$ follows from showing that there exists a $(\pi_1, h) \in \Pi H$ with $\pi_{1,j}^{**} \geq g_{1,j}$ for all $j = 1, \ldots, p$. We have $g_{1,j} \geq 0$ for $j = 1, \ldots, p$ and $g_{1,j} = 0$ for $j = p+1, \ldots, k$. By definition, there exists a subsequence $\{\omega_n\}_{n\geq 1}$ and a sequence $\{\gamma_{\omega_n,g_1,h}\}_{n\geq 1}$. Because $\kappa_n^{-1} n^{1/2}/b_n^{1/2} \to \infty$ it follows that there exists a subsequence $\{v_n\}_{n\geq 1}$ of $\{\omega_n\}_{n\geq 1}$ such that under $\{\gamma_{v_n,g_1,h}\}_{n\geq 1}$

$$\kappa_{v_n}^{-1} v_n^{1/2} \sigma_{F_{v_n,h},j}^{-1}(\theta_{v_n,h}) E_{F_{v_n,h}} m_j(W_i, \theta_{v_n,h}) \to \pi_{1,j}, \quad \text{(B-10)}$$

for some $\pi_{1,j}$ such that for $j = 1, \ldots, p$, $\pi_{1,j} = \infty$ if $g_{1,j} > 0$ and $\pi_{1,j} \geq 0$ if $g_{1,j} = 0$, and $\pi_{1,j} = 0$ for $j = p+1, \ldots, k$. This proves the existence of a sequence $\{\gamma_{v_n,\pi_1,h}\}_{n\geq 1}$. For $j = 1, \ldots, k$, if $\pi_{1,j} = \infty$ then by definition $\pi_{1,j}^{**} = \infty$ and if $\pi_{1,j} \geq 0$ then $\pi_{1,j}^{**} \geq 0$. Therefore, $\pi_{1,j}^{**} \geq g_{1,j}$ for all $j = 1, \ldots, p$ and so $AsySz_{SS} \geq AsySz_{GMS}$.

Second, assume $(\pi_1, h) \in \Pi H$ so that $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$ exists. To show $AsySz_{SS} \leq AsySz_{GMS}$ it is enough to show that there exists $\{\gamma_{\tilde{\omega}_n,\tilde{g}_1,h}\}_{n\geq 1}$ such that $\pi_{1,j}^* \leq \tilde{g}_{1,j}$ for $j = 1, \ldots, k$. Note that it is possible to take a further subsequence $\{v_n\}_{n\geq 1}$ of $\{\omega_n\}_{n\geq 1}$ such that on $\{v_n\}_{n\geq 1}$ the sequence $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$ is a sequence $\{\gamma_{v_n,g_1,h}\}_{n\geq 1}$ for some $g_1 \in \mathbb{R}^k$. By Assumption A.5 there then exists a sequence $\{\gamma_{\tilde{\omega}_n,\tilde{g}_1,h}\}_{n\geq 1}$ for some subsequence $\{\tilde{\omega}_n\}_{n\geq 1}$ of $\mathbb{N}$ and a $\tilde{g}_1$ that satisfies $\tilde{g}_{1,j} = \infty$ when $h_{1,j} = \infty$ and $\tilde{g}_{1,j} \geq 0$ for $j = 1, \ldots, k$. Clearly, for all $j = 1, \ldots, p$ for which $h_{1,j} = \infty$ this implies $\pi_{1,j}^* \leq \tilde{g}_{1,j} = \infty$. In addition, if $h_{1,j} < \infty$ it follows that $\pi_{1,j} = 0$ and thus, by definition, $\pi_{1,j}^* = 0 \leq \tilde{g}_{1,j}$. This is, for $j = 1, \ldots, k$ we have that $\pi_{1,j}^* \leq \tilde{g}_{1,j}$ and, as a result, $AsySz_{SS} \leq AsySz_{GMS}$. This completes the proof.

□

Proof of Theorem 3.2. Part 1. By Lemma B.1

$$AsySz^{(1)}_{GMS} \geq \inf_{(\pi_1,h)\in\Pi H} \Pr\left(S_1(h_2^{1/2}Z + h_1, h_2) \leq c_{\pi_1^*}(h_2, 1-\alpha)\right), \tag{B-11}$$

where $Z \sim N(0_k, I_k)$, $h_2 \in \Psi_1$, $c_{\pi_1^*}(h_2, 1-\alpha)$ is the $1-\alpha$ quantile of $S_1(h_2^{1/2}Z + \pi_1^*, h_2)$, and π_1^* is defined in Lemma B.1. Recall that

$$S_1(h_2^{1/2}Z + h_1, h_2) = \sum_{j=1}^{p}[h_2^{1/2}(j)Z + h_{1,j}]_-^2 + \sum_{j=p+1}^{k}(h_2^{1/2}(j)Z + h_{1,j})^2, \tag{B-12}$$

where $h_2^{1/2}(j) \in \mathbb{R}^k$ denotes the jth row of $h_2^{1/2}$. If we denote by $h_2^{1/2}(j,s)$ the sth element of the vector $h_2^{1/2}(j)$, the following properties hold for all $j \geq 1$

$$\sum_{s=1}^{k}(h_2^{1/2}(j,s))^2 = 1, \quad h_2^{1/2}(j,s) = 0,\ \forall s > j, \quad |h_2^{1/2}(j,s)| \leq 1,\ \forall s \geq 1. \tag{B-13}$$

The properties in Eq. (B-13) follow by h_2 having ones in the main diagonal and $h_2^{1/2}$ being lower triangular. We use Eq. (B-13) and the Cauchy-Schwarz inequality to derive the following three useful inequalities. For any $z \in \mathbb{R}^k$ and $j = 1, \ldots, k$,

$$(h_2^{1/2}(j)z + h_{1,j})^2 \leq \sum_{m=1}^{j}(h_2^{1/2}(j,m))^2 \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2 = \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \tag{B-14}$$

$$[h_2^{1/2}(j)z + h_{1,j}]_-^2 \leq \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \text{ and provided } h_{1,j} \in (0,\infty), \tag{B-15}$$

$$[h_2^{1/2}(j)z + h_{1,j}]_-^2 \leq [h_2^{1/2}(j)z]_-^2 \leq \sum_{s=1}^{j} z_s^2. \tag{B-16}$$

For every $z \in \mathbb{R}^k$ and $h \in H$ define

$$\begin{aligned}\tilde{S}_1(z,h) = {} & \sum_{j=1}^{p}\sum_{s=1}^{j} z_s^2 I(h_{1,j} \in (0,\infty)) + \sum_{j=1}^{p}\sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2 I(h_{1,j} \leq 0) \\ & + \sum_{j=p+1}^{k}\sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2. \end{aligned}\tag{B-17}$$

It follows from Eqs. (B-14), (B-15), and (B-16) that $\tilde{S}_1(z,h) \geq S_1(h_2^{1/2}z + h_1, h_2)$ for all $z \in \mathbb{R}^k$. Therefore, for all $h \in H$ and $x \in \mathbb{R}$

$$\Pr(S_1(h_2^{1/2}Z + h_1, h_2) \leq x) \geq \Pr(\tilde{S}_1(Z,h) \leq x). \tag{B-18}$$

Let $B > 0$ and define $A_B \equiv \{z \in \mathbb{R} : |z| \leq B\}$ and $A_B^k = A_B \times \cdots \times A_B$. Since A_B has positive length on $\mathbb{R}$, it follows that for $Z \sim N(0_k, I_k)$,

$$\Pr(Z \in A_B^k) = \prod_{s=1}^{k}\Pr(Z_s \in A_B) > 0. \tag{B-19}$$

Let $\{\pi_{1,l}, h_l\}_{l\geq 1}$ for $h_l = (h_{1,l}, h_{2,l})$ be a sequence in ΠH such that

$$\inf_{(\pi,h)\in\Pi H}\Pr(S_1(h_2^{1/2}Z + h_1, h_2) \leq c_{\pi_1^*}(h_2, 1-\alpha)) = \lim_{l\to\infty}\Pr(S_1(h_{2,l}^{1/2}Z + h_{1,l}, h_{2,l}) \leq c_{\pi_{1,l}^*}(h_{2,l}, 1-\alpha)), \tag{B-20}$$

and define the sequence $\{B_l\}_{l\geq 1}$ by $B_l = (c_{\pi^*_{1,l}}(h_{2,l}, 1-\alpha)/(2k(k+1)))^{1/2}$.

Define $B = \liminf_{l\to\infty} B_l$. Note that $B \geq 0$. We first consider the case $B > 0$ and then the case $B = 0$. When $B > 0$, assume $r^* \leq B$. Then, there exists a subsequence $\{\omega_l\}_{l\geq 1}$ such that $B_{\omega_l} \geq B$ for all ω_l and thus $r^* \leq B_{\omega_l}$ along the subsequence. By multiplying out, it follows that for all $z_s \in A_{B_{\omega_l}}$ and $j = 1, \ldots, k$, $(z_s + h_2^{1/2}(j,s)h_{1,j})^2 \leq B^2_{\omega_l} + r^{*2} + 2B_{\omega_l} r^*$, when $|h_{1,j}| \leq r_j$. Then, for all $z \in A^k_{B_{\omega_l}}$

$$\tilde{S}_1(z, h_l) \leq \sum_{j=1}^{k}\sum_{s=1}^{j} 4B^2_{\omega_l} = 2k(k+1)B^2_{\omega_l} = c_{\pi^*_{1,\omega_l}}(h_{2,\omega_l}, 1-\alpha). \tag{B-21}$$

As a result, when $r^* \leq B$

$$\Pr(\tilde{S}_1(Z, h_{\omega_l}) \leq c_{\pi^*_{1,\omega_l}}(h_{2,\omega_l}, 1-\alpha)) \geq \Pr(Z \in A^k_{B_{\omega_l}}) > 0. \tag{B-22}$$

It follows from Eqs. (B-11), (B-18), (B-19), (B-20), and (B-22) that

$$\begin{aligned} AsySz^{(1)}_{GMS} &\geq \inf_{(\pi,h)\in\Pi H} \Pr(S_1(h_2^{1/2}Z + h_1, h_2) \leq c_{\pi^*_1}(h_2, 1-\alpha)) \\ &= \lim_{l\to\infty} \Pr(S_1(h_{2,l}^{1/2}Z + h_{1,l}, h_{2,l}) \leq c_{\pi^*_{1,l}}(h_{2,l}, 1-\alpha)) \\ &\geq \liminf_{l\to\infty} \Pr(\tilde{S}_1(Z, h_{\omega_l}) \leq c_{\pi^*_{1,\omega_l}}(h_{2,\omega_l}, 1-\alpha)) \\ &\geq \liminf_{l\to\infty} \Pr(Z \in A^k_{B_{\omega_l}}) > 0. \end{aligned} \tag{B-23}$$

Now consider the case $B = 0$. It follows that there exists a subsequence $\{\omega_l\}_{l\geq 1}$ of $\mathbb{N}$ such that $\lim_{l\to\infty} c_{\pi^*_{1,\omega_l}}(h_{2,\omega_l}, 1-\alpha) = 0$. Let π^*_{1,j,ω_l} denote the jth element of π^*_{1,ω_l}. Since $\pi^*_{1,j,\omega_l} \in \{0, \infty\}$ for $j = 1, \ldots, p$ and $\pi^*_{1,j,\omega_l} = 0$ for $j = p+1, \ldots, k$, there exists a further subsequence $\{\tilde{\omega}_l\}_{l\geq 1}$ such that $\pi^*_{1,\tilde{\omega}_l} = \bar{\pi}^*_1$ for some vector $\bar{\pi}^*_1 \in \mathbb{R}^k_{+,+\infty}$ whose first p components are all in $\{0,\infty\}$ and $h_{2,\tilde{\omega}_l} \to h_2$. Assume that $\bar{\pi}^*_{1,j} = 0$ for some $j = 1, \ldots, k$. By Assumption A.2(c) and $\alpha < 1/2$, it follows that $c_{\bar{\pi}^*_1}(h_2, 1-\alpha) > 0$. Also, by pointwise continuity of $c_{\bar{\pi}^*_1}(h_2, 1-\alpha)$ in h_2 it follows that $\lim_{l\to\infty} c_{\bar{\pi}^*_1}(h_{2,\tilde{\omega}_l}, 1-\alpha) = c_{\bar{\pi}^*_1}(h_2, 1-\alpha) > 0$, which is a contradiction. Therefore, it must be that $k = p$ and $\bar{\pi}^*_1 = \infty_p$. It then follows that $h_{1,\tilde{\omega}_l} = \infty_p$ and $S_1(h^{1/2}_{2,\tilde{\omega}_l}Z + h_{1,\tilde{\omega}_l}, h_{2,\tilde{\omega}_l}) = 0$ a.s. along the subsequence. Therefore the expression on the right hand side of Eq. (B-20) equals 1 in this case. Finally, by the proof of Theorem 3.1, $AsySz^{(1)}_{PA} \geq AsySz^{(1)}_{SS} \geq AsySz^{(1)}_{GMS}$, completing the proof.

Part 2. By Lemma B.1

$$AsySz^{(2)}_{PA} = \inf_{h\in H} \Pr\left(S_2(h_2^{1/2}Z + h_1, h_2) \leq c_0(h_2, 1-\alpha)\right), \tag{B-24}$$

where $h_2^{1/2}Z \sim N(0_k, h_2)$, $c_0(h_2, 1-\alpha)$ is the $1-\alpha$ quantile of $S_2(h_2^{1/2}Z, h_2)$, and H is the space defined in Definition A.1. For $\varepsilon > 0$, let $h^\star_{2,\varepsilon} = \Xi_{1,2}(\varepsilon) \in \mathbb{R}^{k\times k}$, where $\Xi_{1,2}(\varepsilon)$ is defined in Assumption A.7. By Assumption A.7 and without loss of generality, there exists $h_1 \in \mathbb{R}^k$ with $h_{1,1} \leq 0$, $h_{1,2} \leq 0$, and $\min\{h_{1,1}, h_{1,2}\} < 0$ such that $(h_1, h^\star_{2,\varepsilon}) \in H$. It follows that $\det(h^\star_{2,\varepsilon}) = \varepsilon$ and

$$(h^\star_{2,\varepsilon})^{-1} = \begin{bmatrix} i_\varepsilon & 0_{2\times(k-2)} \\ 0_{(k-2)\times 2} & I_{k-2} \end{bmatrix}, \quad \text{where} \quad i_\varepsilon = \frac{1}{1-\rho^2_\varepsilon}\begin{bmatrix} 1 & -\rho_\varepsilon \\ -\rho_\varepsilon & 1 \end{bmatrix}, \tag{B-25}$$

$0_{l,s}$ denotes a $l \times s$ matrix of zeros, and $\rho_\varepsilon \equiv -\sqrt{1-\varepsilon}$. Let $Z^\varepsilon \sim N(0_k, h^\star_{2,\varepsilon})$. Then

$$\begin{aligned} S_2(Z^\varepsilon + h_1, h^\star_{2,\varepsilon}) = \inf_{t\in\mathbb{R}^p_{+,+\infty}} &\{(1-\rho^2_\varepsilon)^{-1}[(Z^\varepsilon_1 + h_{1,1} - t_1)^2 + (Z^\varepsilon_2 + h_{1,2} - t_2)^2 \\ &- 2\rho_\varepsilon(Z^\varepsilon_1 + h_{1,1} - t_1)(Z^\varepsilon_2 + h_{1,2} - t_2)] + \sum_{j=3}^{p}(Z^\varepsilon_j + h_{1,j} - t_j)^2\} + \sum_{j=p+1}^{k}(Z^\varepsilon_j + h_{1,j})^2. \end{aligned} \tag{B-26}$$

At the infimum, $t_j = \max\{Z_j^\varepsilon + h_{1,j}, 0\}$ for $j = 3, \ldots, p$ and so

$$
\begin{aligned}
S_2(Z^\varepsilon + h_1, h_{2,\varepsilon}^\star) = \inf_{t \in \mathbb{R}^2_{+,+\infty}} & \big\{(1-\rho_\varepsilon^2)^{-1}[(Z_1^\varepsilon + h_{1,1} - t_1)^2 + (Z_2^\varepsilon + h_{1,2} - t_2)^2 \\
& -2\rho_\varepsilon (Z_1^\varepsilon + h_{1,1} - t_1)(Z_2^\varepsilon + h_{1,2} - t_2)]\big\} + \sum_{j=3}^{p} [Z_j^\varepsilon + h_{1,j}]_-^2 + \sum_{j=p+1}^{k} (Z_j^\varepsilon + h_{1,j})^2. \\
& \geq \underline{S_2}((Z_1^\varepsilon + h_{1,1}, Z_2^\varepsilon + h_{1,2}), \rho_\varepsilon), \qquad \text{(B-27)}
\end{aligned}
$$

where $\underline{S_2}((z_1, z_2), \rho_\varepsilon) : \mathbb{R}^2 \times (0,1) \mapsto \mathbb{R}_+$ is defined as follows

$$
\underline{S_2}((z_1, z_2), \rho_\varepsilon) = \inf_{t \in \mathbb{R}^2_{+,+\infty}} \big\{(1-\rho_\varepsilon^2)^{-1}[(z_1 - t_1)^2 + (z_2 - t_2)^2 - 2\rho_\varepsilon (z_1 - t_1)(z_2 - t_2)]\big\}. \qquad \text{(B-28)}
$$

Let $h_{1,1} < 0$ without loss of generality (since $\min\{h_{1,1}, h_{1,2}\} < 0$). For small $\beta > 0$, $(h_{1,1}, h_{1,2}) \in H_\beta$ where the set $H_\beta \subseteq \mathbb{R}^2$ is defined in Lemma S1.3. By Eq. (B-27) and Lemma S1.3, there exists a function $\tau_\varepsilon((z_1, z_2), (h_{1,1}, h_{1,2})) : A_{\beta,\varepsilon} \times H_\beta \mapsto \mathbb{R}_+$ such that

$$
S_2(z + h_1, h_{2,\varepsilon}^\star) \geq \underline{S_2}((z_1, z_2), \rho_\varepsilon) + \frac{\tau_\varepsilon((z_1, z_2), (h_{1,1}, h_{1,2}))}{1 - \rho_\varepsilon^2}, \qquad \text{(B-29)}
$$

for all $z \in \mathbb{R}^k$ with $(z_1, z_2) \in A_{\beta,\varepsilon}$ and for the particular $(h_1, h_{2,\varepsilon}^\star)$ under consideration.

Next, note that by Lemma S1.2 it follows that with probability one

$$
S_2(Z^\varepsilon, h_{2,\varepsilon}^\star) = \sum_{j=3}^{p} [Z_j^\varepsilon]_-^2 + \sum_{j=p+1}^{k} (Z_j^\varepsilon)^2 + f(Z_1^\varepsilon, Z_2^\varepsilon, \rho_\varepsilon) \leq \sum_{j=3}^{p} [Z_j^\varepsilon]_-^2 + \sum_{j=p+1}^{k} (Z_j^\varepsilon)^2 + (Z_1^\varepsilon)^2 + W^2, \qquad \text{(B-30)}
$$

where $W = (Z_2^\varepsilon - \rho_\varepsilon Z_1^\varepsilon)/\sqrt{1-\rho_\varepsilon^2}$ and hence $Z_1^\varepsilon \perp W \sim N(0,1)$, and $f(\cdot)$ is defined in Lemma S1.2 and satisfies $f(Z_1^\varepsilon, Z_2^\varepsilon, \rho_\varepsilon) \leq (Z_1^\varepsilon)^2 + W^2$ with probability one for all $\varepsilon > 0$. As a result, the $1-\alpha$ quantile of $S_2(Z^\varepsilon, h_{2,\varepsilon}^\star)$, denoted by $c_0(h_{2,\varepsilon}^\star, 1-\alpha)$, is bounded above by the $1-\alpha$ quantile of the RHS of Eq. (B-30), which does not depend on ε. It then follows that $c_0(h_{2,\varepsilon}^\star, 1-\alpha) \leq C < \infty$, where C denotes the $(1-\alpha)$ quantile of the RHS of Eq. (B-30). By Lemma S1.3 we have that $\forall \eta > 0$, $\exists \varepsilon > 0$ such that

$$
\Pr\left(\frac{\tau_\varepsilon((Z_1^\varepsilon, Z_2^\varepsilon), (h_{1,1}, h_{1,2}))}{1 - \rho_\varepsilon^2} > C, (Z_1^\varepsilon, Z_2^\varepsilon) \in A_{\beta,\varepsilon}\right) \geq 1 - \eta. \qquad \text{(B-31)}
$$

We can conclude that $\forall \eta > 0$, $\exists \varepsilon > 0$ such that

$$
\begin{aligned}
AsySz_{PA}^{(2)} & \leq \Pr\left(S_2^\varepsilon(Z^\varepsilon + h_1, h_{2,\varepsilon}^\star) \leq c_0(h_{2,\varepsilon}^\star, 1-\alpha)\right) \\
& \leq 1 - \Pr\left(S_2^\varepsilon(Z^\varepsilon + h_1, h_{2,\varepsilon}^\star) > C\right) \\
& \leq 1 - \Pr\left(S_2^\varepsilon(Z^\varepsilon + h_1, h_{2,\varepsilon}^\star) > C, (Z_1^\varepsilon, Z_2^\varepsilon) \in A_{\beta,\varepsilon}\right) \\
& \leq \eta, \qquad \text{(B-32)}
\end{aligned}
$$

where the first inequality follows from $(h_1, h_{2,\varepsilon}^\star) \in H$, the second inequality from $c_0(h_{2,\varepsilon}^\star, 1-\alpha) \leq C$, the third inequality from $A_{\beta,\varepsilon} \subseteq \mathbb{R}^2$, the fourth one from $\underline{S_2}((z_1, z_2), \rho_\varepsilon) \geq 0$ $\forall (z_1, z_2) \in \mathbb{R}^2$ and Eqs. (B-29) and (B-31). By Theorem 3.1, $AsySz_{PA}^{(2)} \geq AsySz_{SS}^{(2)} \geq AsySz_{GMS}^{(2)}$ and this completes the proof. □

References

ANDREWS, D. W. K. AND P. GUGGENBERGER (2009a): "Hybrid and Size-Corrected Subsample Methods," *Econometrica*, 77, 721–762.

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

——— (2010a): "Applications of Hybrid and Size-Corrected Subsampling Methods," *Journal of Econometrics*, 158, 285–305.

——— (2010b): "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 P. JIA (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," Manuscript, Yale University.

ANDREWS, D. W. K. AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–158.

BERESTEANU, A. AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2008): "Set Identified Linear Models," *Econometrica*, forthcoming.

BUGNI, F., I. A. CANAY, AND P. GUGGENBERGER (2011): "Supplement to 'Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models'," *Econometrica Supplemental Material*.

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

CANAY, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

FAN, Y. AND S. S. PARK (2009): "Partial Identification of the Distribution of Treatment Effects and its Confidence Sets," in *Nonparametric Econometric Methods (Advances in Econometrics)*, ed. by T. B. Fomby and R. C. Hill, United Kingdom: Emerald Group Publishing Limited, vol. 25, 3–70.

GALICHON, A. AND M. HENRY (2009): "Dilation Bootstrap: A Methodology for Constructing Confidence Regions with Partially Identified Models," Manuscript, University of Montreal.

——— (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78, 1264–1298.

GUGGENBERGER, P. (2011): "On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Assumption," *Econometric Theory*, doi:10.1017/S0266466611000375. Published online by Cambridge University Press 13 September 2011.

IMBENS, G. AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kitamura, Y., T. Otsu, and K. Evdokimov (2009): "Robustness, Infinitesimal Neighborhoods, and Moment Restrictions," CFDP 1720.

Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer-Verlag, New York.

Moon, H. R. and F. Schorfheide (2011): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica,* forthcoming.

Newey, W. K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2005): "Moment Inequalities and Their Applications," Manuscript, Harvard University.

Politis, D. N. and J. P. Romano (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22, 2031–2050.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*, Springer, New York.

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

Romano, J. P. and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786–2807.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–212.

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

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

Tamer, E. (2010): "Partial Identification in Econometrics," *Annual Reviews of Economics*, forthcoming.