Asymptotic Distortions 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

August 10, 2010

Abstract

This paper studies the behavior under local misspecification of several tests commonly used in the literature on inference in moment inequality models. We compare across critical values and test statistics and find two important results. First, we show that confidence sets based on subsampling and generalized moment selection (GMS, Andrews and Soares (2010)) share the same level of asymptotic distortion, despite the fact that the latter leads to smaller confidence sets under correct model specification. Second, we show that the asymptotic distortion of the quasi-likelihood ratio test statistic is arbitrarily worse to that of the modified method of moments test statistic. Our results are supported by Monte Carlo simulations.

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

JEL CLASSIFICATION: C01, C12, C20.

*Email: federico.bugni@duke.edu. Website: http://econ.duke.edu/~fb32/index.html
†Email: iacanay@northwestern.edu. Website: http://faculty.wcas.northwestern.edu/~iac879.
‡E-mail: pguggenberger@ucsd.edu. Website: http://www.econ.ucla.edu/people/faculty/Guggenberger.html

1 Introduction

In the last couple of years there have been numerous papers in Econometrics on inference in partially identified models, many of which focused on inference about the identifiable parameters in models defined by moment inequalities (see, among others, Imbens and Manski (2004), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009b)(AG from now on), Andrews and Soares (2010)(AS from now on), Bugni (2010), and Canay (2010)). As a consequence, there are currently several different methods to construct confidence sets (CSs) based on test inversion that have been compared in terms of asymptotic confidence size and length properties (e.g. Andrews and Jia (2008), AG, AS, Bugni (2010) and Canay (2010)). In this paper we are interested in the relative robustness of CSs with respect to their asymptotic confidence size distortion when moment (in)equalities are potentially locally violated. Intuition might suggest that CSs that tend to be smaller under correct model specification are more size distorted under local model misspecification, that is, less robust to small perturbations of the true model.[1] We show that this is true for CSs based on plug-in asymptotic (PA) critical values compared to subsampling and generalized moment selection (GMS, see AS), as well as for the modified methods of moments (MMM) test statistic compared to the quasi likelihood ratio (QLR). However, the two main contributions of this paper lie behind two results that go beyond such intuition. First, we show that CSs based on subsampling and GMS critical values share the same level of asymptotic distortion under mild assumptions, despite the fact that the latter leads to smaller CSs under correct model specification, see AS. Second, we show that CSs based on the QLR test statistic have arbitrarily worse asymptotic confidence size distortion than CSs based on the MMM test statistic under certain conditions.

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. This is, it is typically impossible to do meaningful inference based on the data alone and therefore the researcher has no choice but to impose some structure and include some assumptions. The partial identification approach to inference (in particular, moment inequality models) allows the researcher to conduct inference on the parameter of interest in a way that is robust to certain fundamental assumptions (typically related to the behavior of economic agents), while keeping a second group of less fundamental assumptions as given (typically related to parametric functional forms). For example, in a standard simultaneous entry game where firms have profit functions given by

$$\pi_l = (u_l - \theta_l W_{-l}) I(W_l = 1), \tag{1.1}$$

[1] A test is locally more powerful if it rejects sequences of parameters that are local to the null hypothesis. Under local misspecification, these sequences are part of the perturbed set of null parameters and so rejecting them introduces size distortion. It is worth noting that the study of properties under local misspecification introduces additional complications as *all* sequences of local alternatives matter and, in particular, finding the worst sequence becomes an essential part of the analysis.

W_l is a binary entry indicator and u_l is firm's l benefit of entry, moment inequality models have been used in applied work to deal with the existence of multiple equilibria (e.g. Grieco (2009) and Ciliberto and Tamer (2010)). However, the linear structure and the parametric family of distributions for u_j are typically taken as given. One justification for this asymmetry in the way assumptions are treated lies behind the idea that there are certain assumptions that directly affect the behavior of the agents in the structural model (and partial identification aims to perform robust inference with respect to this group of assumptions), while there are other assumptions that are made out of computational and analytical convenience (i.e., functional forms and distributional assumptions). Here we will not discuss the nature of a certain assumption,[2] 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 be misspecified). There are two basic approaches to such an analysis that we briefly describe below.

First, if the nature of the misspecification remains constant throughout the sample, we say that the model is globally misspecified. In this context, the object of interest becomes a *pseudo-true value* of the parameter of interest, which is typically defined as the parameter value associated with the distribution that is closest (according to some metric) to the true data generating process (e.g., in the case of standard maximum likelihood estimation the pseudo-true value minimizes the Kullback-Leibler discrepancy between the true model and the incorrect parametric model). An extensive discussion of this type of misspecification in the context of over-identified moment equality models can be found in Hall and Inoue (2003). In the context of partially identified models, Ponomareva and Tamer (2009) discuss the impact of global misspecification on the set of identifiable parameters and show that parameters obtained from methods designed for incomplete models do not match those obtained from the same models had those been complete.

Second, if the data do not satisfy the population moment condition for any finite sample size, but do so in the limit as the sample size goes to infinity, we say that the model is locally misspecified. By its very nature, this type of analysis provides guidance in situations where the truth is just a small perturbation away from the assumed model. Such an analysis was applied in the context of over-identified moment equality models by Newey (1985). More recently Guggenberger (2009) studied the properties of linear IV estimators under local violations of exogeneity conditions, while Kitamura, Otsu, and Evdokimov (2009) considered local deviations within shrinking topological neighborhoods of point identified moment equality models and proposed an estimator that achieves optimal minimax robust properties. Since the limit of locally misspecified models equals the correctly specified model, the object of interest under local misspecification and correct specification coincides. This facilitates the interpretation relative to pseudo true values that are present in globally misspecified models.

[2] For an extensive discussion on the role of different assumptions and partial identification in general see Manski (2003) and Tamer (2009).

Therefore, if we believe that the probability law generating the observations is in fact a small perturbation of the true law, then it is of interest to seek for an inference procedure that it robust against such slight perturbation in the observed data. This is the approach that we take in this paper.

The paper is organized as follows. Section 2 introduces the model and notation, and provides two examples that illustrate the nature of misspecification we capture in our analysis. Section 3 provides asymptotic confidence size distortion results across different test statistics and critical values. Section 4 presents simulation results that support the main findings of this paper. We include the assumptions and proofs of the results in the Appendix.

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 use $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$, $\mathbb{R}_{+,+\infty} = \mathbb{R}_+ \cup \{+\infty\}$, $\mathbb{R}_{+\infty} = \mathbb{R} \cup \{+\infty\}$, $\mathbb{R}_{\pm\infty} = \mathbb{R} \cup \{\pm\infty\}$, $K^p = K \times \cdots \times K$ (with p copies) for any set K, $\infty_p = (+\infty, \ldots, +\infty)$ (with p copies), 0_k for a k-vector of zeros and I_k for a $k \times k$ identity matrix.

2 Locally Misspecified Moment Equality/Inequality Models

The moment inequality/equality model assumes the existence of a true parameter vector θ_0 ($\in \Theta \subset \mathbb{R}^d$) that satisfies the moment restrictions

$$\begin{aligned} &E_{F_0} m_j(W_i, \theta_0) \geq 0 \text{ for } j = 1, \ldots, p \text{ and} \\ &E_{F_0} m_j(W_i, \theta_0) = 0 \text{ for } j = p+1, \ldots, p+v \equiv k, \end{aligned} \tag{2.1}$$

where $\{m_j(\cdot, \theta)\}_{j=1}^k$ are known real-valued moment functions and $\{W_i\}_{i=1}^n$ are observed i.i.d. random vectors with joint distribution F_0. We consider confidence sets (CSs) for θ_0 obtained by inverting tests of the hypothesis

$$H_0 : \theta_0 = \theta \quad \text{vs.} \quad H_1 : \theta_0 \neq \theta. \tag{2.2}$$

This is, if we denote by $T_n(\theta)$ a generic test statistic for testing (2.2) and by $c_n(\theta, 1-\alpha)$ the critical value of the test at nominal size α, then the $1-\alpha$ CS for θ_0 is

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta, 1-\alpha)\}. \tag{2.3}$$

Several CSs have been suggested in the literature whose asymptotic confidence size is greater than or equal to the nominal coverage size under mild technical conditions. The test statistics include modified method of moments (MMM), quasi likelihood ratio (QLR) or generalized empirical likelihood (GEL) statistics. Critical values include plug-in asymptotic (PA), subsampling, and generalized moment selection (GMS) implemented via asymptotic approximations or the bootstrap.[3]

[3] The details of the test statistics and critical values are presented in the next section.

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. AS establish that GMS tests have asymptotic power greater than or equal to that of subsampling tests and strictly greater power than subsampling tests for certain local alternatives. Furthermore, subsampling has greater than or equal power than PA tests for all local alternatives and strictly greater power for certain local alternatives. Andrews and Jia (2008) compare different combinations of tests statistics and critical values and provide a recommended test based on the QLR statistic and a refined moment selection (RMS) critical value which involves a data-dependent rule for choosing the GMS tuning parameter. Additional results on power include those in Canay (2010). Here we are interested in ranking the resulting CSs in terms of asymptotic confidence size distortion when the moment (in)equalities in (2.1) are potentially locally violated. Consider the following examples as illustrations.

Example 2.1 (Missing Data)**.** Suppose that the economic model indicates that

$$E_{F_0}\left(Y|X=x\right)=G\left(x,\theta_0\right),\forall x\in S_X, \tag{2.4}$$

where θ_0 is the true parameter value and $S_X=\{x_l\}_{l=1}^{d_x}$ is the (finite) support of X. The sample is affected by missing data on Y. Denote by Z the binary variable that takes value of one if Y is observed and zero if Y is missing. Conditional on $X=x$, Y has logical lower and upper bounds given by $Y_L\left(x\right)$ and $Y_H\left(x\right)$, respectively. When the observed data $W_i=(Y_iZ_i,Z_i,X_i)$ comes from the model in (2.4), the true θ_0 satisfies the following inequalities,

$$\begin{aligned} E_{F_0}m_{l,1}(W_i,\theta_0)&=E_{F_0}[(YZ+Y_H(x_l)(1-Z)-G(x_l,\theta_0))I(X=x_l)]\geq 0,\\ E_{F_0}m_{l,2}(W_i,\theta_0)&=E_{F_0}[(G(x_l,\theta_0)-YZ-Y_L(x_l)(1-Z))I(X=x_l)]\geq 0, \end{aligned} \tag{2.5}$$

for $l=1,\ldots,d_x$. Now suppose that in fact the data come from a local perturbation F_n of the true model F_0 such that

$$E_{F_n}\left(Y|X=x\right)=G_n\left(x,\theta_0\right),\forall x\in S_X, \tag{2.6}$$

and for a vector $r\in\mathbb{R}_+^k$

$$|G_n(x_l,\theta_0)-G(x_l,\theta_0)|\leq r_ln^{-1/2},\quad \forall l=1,2,\ldots,d_x. \tag{2.7}$$

The last condition says that the true function G_n is not too far from the model G used by the researcher. After a few manipulations, it follows that

$$\begin{aligned} E_{F_n}m_{l,1}(W_i,\theta_0)&=E_{F_n}[(YZ+Y_H(x_l)(1-Z)-G(x_l,\theta_0))I(X=x_l)]\geq -r_ln^{-1/2},\\ E_{F_n}m_{l,2}(W_i,\theta_0)&=E_{F_n}[(G(x_l,\theta_0))-YZ-Y_L(x_l)(1-Z))I(X=x_l)]\geq -r_ln^{-1/2}, \end{aligned} \tag{2.8}$$

for $l = 1, \dots, d_x$. Therefore, under the true distribution of the data the original moment conditions in (2.5) may be locally violated at θ_0. ■

Example 2.2 (Entry Game)**.** Suppose firm $l \in \{1, 2\}$ generates profits

$$\pi_{l,i}(\theta_l, W_{-l,i}) = u_{l,i} - \theta_l W_{-l,i} \tag{2.9}$$

when entering market $i \in \{1, \dots, n\}$. Here $W_{l,i} = 0$ or 1 denotes "not entering" or " entering" market i by firm l, respectively, a subscript $-l$ denotes the decision of the other firm, the random variable $u_{l,i}$ denotes the profit of firm l in market i if $W_{-l,i} = 0$, and $\theta_l \in [0, 1]$ is the profit reduction incurred by firm l if $W_{-l,i} = 1$. If firm l does not enter market i then its profit in this market is 0. We assume $(u_i)_{i=1}^n = (u_{1,i}, u_{2,i})_{i=1}^n$ are i.i.d. , with realizations restricted to $[0, 1]^2$, and continuous distribution. Therefore, if the other firm does not enter then entering is always profitable.

Define $W_i = (W_{1,i}, W_{2,i})$ and $\theta_0 = (\theta_1, \theta_2)$. There are four possible outcomes: (i) $W_i = (1, 1)$ is the unique Nash Equilibrium (NE) if $u_{l,i} > \theta_l$ for $l = 1, 2$; (ii) $W_i = (1, 0)$ is the unique NE if $u_{1,i} > \theta_1$ and $u_{2,i} < \theta_2$; (iii) $W_i = (0, 1)$ is the unique NE if $u_{1,i} < \theta_1$ and $u_{2,i} > \theta_2$, and; (iv) there are multiple equilibria if $u_{l,i} < \theta_l$ for $l = 1, 2$ as both $W_i = (1, 0)$ and $W_i = (0, 1)$ are NE. It follows that

$$\begin{aligned}
&\Pr(W_i = (1,0)) \le \Pr(\pi_{2,i} < 0) = \Pr(u_{2,i} < \theta_2) \equiv G_1(\theta_0), \\
&\Pr(W_i = (1,0)) \ge \Pr(\pi_{1,i} > 0 \;\&\; \pi_{2,i} < 0) = \Pr(u_{1,i} \ge \theta_1 \;\&\; u_{2,i} < \theta_2) \equiv G_2(\theta_0), \\
&\Pr(W_i = (1,1)) = \Pr(\pi_{1,i} > 0 \;\&\; \pi_{2,i} > 0) = \Pr(u_{1,i} \ge \theta_1 \;\&\; u_{2,i} \ge \theta_2) \equiv G_3(\theta_0),
\end{aligned} \tag{2.10}$$

where the notation $G_1(\theta_0), G_2(\theta_0), G_3(\theta_0)$ is used if $u_i \sim G$. The resulting moment (in)equalities are

$$\begin{aligned}
E_{F_0} m_1(W_i, \theta_0) &= E_{F_0}[G_1(\theta_0) - W_{1,i}(1 - W_{2,i})] \ge 0, \\
E_{F_0} m_2(W_i, \theta_0) &= E_{F_0}[W_{1,i}(1 - W_{2,i}) - G_2(\theta_0)] \ge 0, \\
E_{F_0} m_3(W_i, \theta_0) &= E_{F_0}[W_{1,i} W_{2,i} - G_3(\theta_0)] = 0,
\end{aligned} \tag{2.11}$$

where F_0 denotes the true distribution of W_i that must be compatible with the true joint distribution of u_i according to the restrictions (i)-(iv) above.

To do inference on θ_0, the econometrician assumes G is the joint distribution of the unobserved random vector u_i. Suppose the econometrician's assumption is "close to the truth but not quite the truth". More specifically, suppose that for a $r = (r_1, r_2, r_3)' \in \mathbb{R}_+^3$

$$|G_j(\theta_0) - G_{nj}(\theta_0)| \le r_j n^{-1/2}, \quad j = 1, 2, 3, \tag{2.12}$$

where G_n denotes the true distribution of u_i for sample size n and $G_{nj}(\theta_0)$ is defined as $G_j(\theta_0)$ above when $u_i \sim G_n$ rather than $u_i \sim G_n$. Denote by F_n the true distribution of

W_i that must be compatible with the true joint distribution of $u_i \sim G_n$. Then, combining Equations (2.10) and (2.11) we obtain

$$\begin{aligned} E_{F_n} m_1(W_i, \theta_0) &= E_{F_n}[G_1(\theta_0) - W_{1,i}(1 - W_{2,i})] \geq -r_1 n^{-1/2}, \\ E_{F_n} m_2(W_i, \theta_0) &= E_{F_n}[W_{1,i}(1 - W_{2,i}) - G_2(\theta_0)] \geq -r_2 n^{-1/2}, \\ |E_{F_n} m_3(W_i, \theta_0)| &= |E_{F_n}[W_{1,i} W_{2,i} - G_3(\theta_0)]| \leq -r_3 n^{-1/2}. \end{aligned} \tag{2.13}$$

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

Remark 2.1. Note that in both examples the parameter θ_0 has a meaningful interpretation independent of the potential misspecification of the model of the type considered above. However, by assuming a slightly incorrect parametric structure the moment (in)equalities are potentially violated for a given sample size n at the true θ_0.

Remark 2.2. In Example 2.2 the differences between $G_j(\theta_0)$ and $G_{nj}(\theta_0)$ in Equation (2.12) could arise due to misspecification in the distribution of u_i, due to misspecification in the profit function $\pi_{n,l,i}(\theta, W_{-l,i})$, or due to both. This is, let $\pi_{n,l,i}(\theta, W_{-l,i})$ be a perturbed profit function that results in $G_{n1}(\theta) \equiv \Pr(\pi_{n,2,i} < 0)$ and so on. If Equation (2.12) holds, then the representation in (2.13) would follow as in the example.

Examples 2.1 and 2.2 illustrate 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. 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 (Parameter Space)**.** The parameter space $\mathcal{F}_n \equiv \mathcal{F}_n(r, \delta, M, \Psi)$ for (θ, F) is the set of all tuplets (θ, F) that satisfy

$$\begin{aligned} &(i)\ \theta \in \Theta, \\ &(ii)\ \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) \geq -r_j n^{-1/2}, \quad j = 1, \ldots, p, \\ &(iii)\ |\sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta)| \leq r_j n^{-1/2}, \quad j = p+1, \ldots, k, \\ &(iv)\ \{W_i\}_{i=1}^n \text{ are i.i.d. under } F, \\ &(v)\ \sigma_{F,j}^2(\theta) = Var_F(m_j(W_i, \theta)) \in (0, \infty), \quad j = 1, \ldots, k, \\ &(vi)\ Corr_F(m(W_i, \theta)) \in \Psi, \text{ and}, \\ &(vii)\ E_F|m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M, \quad j = 1, \ldots, k, \end{aligned} \tag{2.14}$$

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

As made explicit in the notation, the parameter space depends on n. It also depends on the number of moment restrictions k and the "upper bound" on the local moment/moment inequality violation r. Conditions (ii)-(iii) are modifications of (3.3) in AG (or (2.2) in AS) to account for local model misspecification. Note that this captures the situations illustrated in Examples 2.1 and 2.2. Finally, we use

$$r^* \equiv \max\{r_1, \ldots, r_k\} \tag{2.15}$$

to measure the amount of misspecification.

Remark 2.3. The parameter space in (2.14) includes the space $\mathcal{F}_0 \equiv \mathcal{F}_n(0_k, \delta, M, \psi)$ for all $n \geq 1$, which is the set of correctly specified models. The content of the theorems in the next section continue to hold if we alternatively define $\mathcal{F}_n$ enforcing that at least one moment is strictly violated. One way of doing this is by adding the restriction

$$(viii)\ \sigma_{F,j}^{-1}(\theta) E_F m_j(W, \theta) = -r_j n^{-1/2} \text{ and } r_j > 0 \text{ for some } j = 1, \ldots, k. \tag{2.16}$$

The asymptotic confidence size of CS_n in Equation (2.3) is defined as

$$AsyCS = \liminf_{n \to \infty} \inf_{(\theta, F) \in \mathcal{F}_n} \Pr_{\theta,F}(T_n(\theta) \leq c_n(\theta, 1-\alpha)), \tag{2.17}$$

where $\Pr_{\theta,F}(\cdot)$ denotes the probability measure when the true value of the parameter is θ and the true distribution equals F. This is the limit of the exact size of the test, which is the magnitude one aims to control in finite samples. We know that $AsyCS \geq 1-\alpha$ for certain subsampling, PA, and GMS tests when $r^* = 0$. Here we are interested in deriving $AsyCS$ for these CSs when r^* is strictly positive and rank them according to their level of distortion, defined as

$$\max\{1 - \alpha - AsyCS, 0\} \tag{2.18}$$

Before doing this, we present the different test statistics and critical values in the next subsection.

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 (2.2) or, equivalently, to construct (2.3). Define the sample moment functions

$$\bar{m}_n(\theta) = (\bar{m}_{n,1}(\theta), \ldots, \bar{m}_{n,k}(\theta)), \text{ where}$$
$$\bar{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^{n} m_j(W_i, \theta) \text{ for } j = 1, \ldots, k. \tag{2.19}$$

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

$$\hat{\Sigma}_n(\theta) = n^{-1}\sum_{i=1}^{n}(m(W_i,\theta) - \bar{m}_n(\theta))(m(W_i,\theta) - \bar{m}_n(\theta))'. \qquad (2.20)$$

The statistic $T_n(\theta)$ is defined to be of the form

$$T_n(\theta) = S(n^{1/2}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)), \qquad (2.21)$$

where S is a real function on $\mathbb{R}^p_{+\infty} \times \mathbb{R}^v \times \mathcal{V}_{k\times k}$, the set $\mathcal{V}_{k\times k}$ is the space of $k \times k$ variance matrices and S satisfies Assumption A.1.

We now describe two popular choices of test functions. The first test function S is

$$S_1(m,\Sigma) = \sum_{j=1}^{p}[m_j/\sigma_j]^2_- + \sum_{j=p+1}^{k}(m_j/\sigma_j)^2, \qquad (2.22)$$

where $[x]_- = xI(x < 0)$, $m = (m_1, \dots, m_k)$ and σ_j^2 is the jth diagonal element of Σ. For this function, the parameter space Ψ for the correlation matrices in condition (v) of (2.14) is not restricted. That is, (2.14) holds with $\Psi = \Psi_1$, where Ψ_1 contains all $k \times k$ correlation matrices. The function S_1 leads to the test statistic

$$T_{1,n}(\theta) = n\sum_{j=1}^{p}[\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]^2_- + n\sum_{j=p+1}^{k}(\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta))^2, \qquad (2.23)$$

where $\hat{\sigma}^2_{n,j}(\theta)$ is the jth diagonal element of $\hat{\Sigma}_n(\theta)$. This statistic gives positive weight to moment inequalities only when they are violated.

The second test function is a Gaussian quasi-likelihood ratio (or minimum distance) function defined by

$$S_2(m,\Sigma) = \inf_{t=(t_1,0_v):t_1\in\mathbb{R}^p_{+,+\infty}}(m-t)'\Sigma^{-1}(m-t). \qquad (2.24)$$

This function requires Σ to be non-singular so we take $\Psi = \Psi_{2,\varepsilon}$, where $\Psi_{2,\varepsilon}$ contains all $k\times k$ correlation matrices whose determinant is greater than or equal to $\varepsilon > 0$, for some $\varepsilon > 0$. The function S_2 leads to the test statistic

$$T_{2,n}(\theta) = \inf_{t=(t_1,0_v):t_1\in\mathbb{R}^p_{+,+\infty}}(n^{1/2}\bar{m}_n(\theta) - t)'\hat{\Sigma}_n(\theta)^{-1}(n^{1/2}\bar{m}_n(\theta) - t). \qquad (2.25)$$

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.[4] AG discusses additional examples of test functions.

[4] Note $S_1(m,\Sigma)$ is increasing in $|m_j|$ for $j = p+1,\dots,k$, while $S_2(m,\Sigma)$ is not. To see this take $p = 0$, $k = 2$, and Σ with ones in the diagonal and σ_{12} off diagonal. Then $S_2(m,\Sigma) = (1-\sigma_{12}^2)^{-1}(m_1^2 + m_2^2 - 2m_1m_2\sigma_{12})$. Taking partial derivatives yields the result.

We next describe three popular choices of critical values. Assuming the limiting correlation matrix of $m(W_i, \theta)$ is given by Ω, Equations (2.1) hold, and these expectations do not change with n, it follows from Lemma B.1 that

$$T_n(\theta) \to^d S(\Omega^{1/2}Z + h_1, \Omega), \text{ where } Z \sim N(0_k, I_k), \tag{2.26}$$

and h_1 is a k-vector with jth component equal to 0 when $j > p$, and equal to 0 or ∞ for $j \leq p$ when the jth moment inequality is binding or not binding, respectively, see Lemma B.1. $\Omega^{1/2}$ denotes a lower triangular matrix such that $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$. Therefore, ideally one would like to use 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. This requires knowledge of h_1, which cannot be estimated consistently, and so some approximation to $c_{h_1}(\Omega, 1-\alpha)$ is necessary.

Under the assumptions in the Appendix, the asymptotic distribution in (2.26) is stochastically largest over distributions in $\mathcal{F}_0$ (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 $S(\Omega^{1/2}Z, \Omega)$ where $h_1 = 0_k$. PA critical values are based on this "worst case" and are defined as consistent estimators of $c_0(\Omega, 1-\alpha)$. Define

$$\hat{\Omega}_n(\theta) = \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta), \tag{2.27}$$

where $\hat{D}_n(\theta) = Diag(\hat{\Sigma}_n(\theta))$ and $\hat{\Sigma}_n(\theta)$ is defined in (2.20). The PA test rejects H_0 if $T_n(\theta) > c_0(\hat{\Omega}_n(\theta), 1-\alpha)$, where the PA critical value is

$$c_0(\hat{\Omega}_n(\theta), 1-\alpha) \equiv \inf\{x \in \mathbb{R} : \Pr(S(\hat{\Omega}_n(\theta)^{1/2}Z, \hat{\Omega}_n(\theta)) \leq x) \geq 1-\alpha\}, \tag{2.28}$$

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 AS. To this end, let

$$\xi_n(\theta) = \kappa_n^{-1}\hat{D}_n^{-1/2}(\theta)n^{1/2}\bar{m}_n(\theta), \tag{2.29}$$

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

$$\varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \tag{2.30}$$

where $\varphi = (\varphi_1, \ldots, \varphi_p, 0_v) \in \mathbb{R}^k_{[+\infty]}$ is a function chosen by the econometrician that is assumed to satisfy assumption A.4 in the Appendix. Examples include $\varphi_j^{(1)}(\xi, \Omega) = \infty I(\xi_j > 1)$,

$\varphi_j^{(2)}(\xi, \Omega) = \psi(\xi_j)$, $\varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+$, and $\varphi_j^{(4)}(\xi, \Omega) = \xi_j$ for $j = 1, \ldots, p$, where $\psi(\cdot)$ is a non-decreasing function that satisfies $\psi(x) = 0$ if $x \leq a_L$, $\psi(x) \in [0, \infty]$ if $a_L < x < a_U$, and $\psi(x) = \infty$ if $x > a_U$ for some $0 < a_L \leq a_U \leq \infty$. We use the convention $\infty 0 = 0$ for $\varphi_j^{(1)}$. See AS 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

$$\hat{c}_{n,\kappa_n}(\theta, 1-\alpha) \equiv \inf\{x \in \mathbb{R} : \Pr(S(\hat{\Omega}_n^{1/2}(\theta)Z + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta)) \leq x) \geq 1-\alpha\}, \quad (2.31)$$

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

Finally, we define subsampling critical values, see Politis and Romano (1994) and Politis, Romano, and Wolf (1999). Let $b = b_n$ denote the subsample size when the sample size is n. Throughout the paper we assume $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. The number of different subsamples of size b is q_n. With i.i.d. observations, there are $q_n = n!/((n-b)!b!)$ different subsamples of size b. The subsample statistics used to construct the subsampling critical value are $\{T_{n,b,s}(\theta)\}_{s=1}^{q_n}$, where $T_{n,b,s}(\theta)$ is a subsample statistic defined exactly as $T_n(\theta)$ is defined but based on the sth subsample of size b rather than the full sample. The empirical distribution function of $\{T_{n,b,s}(\theta)\}_{s=1}^{q_n}$ is

$$U_{n,b}(\theta, x) = q_n^{-1} \sum_{s=1}^{q_n} I(T_{n,b,s}(\theta) \leq x) \text{ for } x \in \mathbb{R}. \quad (2.32)$$

The subsampling test rejects H_0 if $T_n(\theta) > \hat{c}_{n,b}(\theta, 1-\alpha)$, where the subsampling critical value is

$$\hat{c}_{n,b}(\theta, 1-\alpha) \equiv \inf\{x \in \mathbb{R} : U_{n,b}(\theta, x) \geq 1-\alpha\}. \quad (2.33)$$

Having introduced the different test statistics and critical values, we devote the next section to the analysis of the asymptotic properties of the different tests under the locally misspecified models defined in Definition 2.1.

3 Asymptotic Confidence Size Distortions

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

Theorem 3.1. *Suppose the same assumptions hold as in Lemma B.2. Then:*

1. *We have*
$$AsyCS_{PA} \geq AsyCS_{SS} \text{ and } AsyCS_{PA} \geq AsyCS_{GMS}. \tag{3.1}$$
Therefore, PA CSs are at least as robust as GMS and subsampling CSs under local violations of the moment (in)equalities.

2. *Under Assumption A.6 we have*
$$AsyCS_{PA} < 1 - \alpha. \tag{3.2}$$
By Equation (3.1) *it follows that $AsyCS_{SS} < 1 - \alpha$ and $AsyCS_{GMS} < 1 - \alpha$.*

3. *Under Assumption A.5 and $\kappa_n^{-1} n^{1/2}/b^{1/2} \to \infty$, it follows that*
$$AsyCS_{SS} = AsyCS_{GMS}. \tag{3.3}$$
Therefore, subsampling CSs and GMS CSs are equally robust under local violations of the moment (in)equalities.

Theorem 3.1 follows as a corollary of Lemma B.2. Assumptions A.6 and A.5 ensure the model is rich enough. These are mild assumptions and we verify them for the two lead examples in Section D of the Appendix. Under a reasonable set of assumptions, the theorem concludes that
$$AsyCS_{GMS} = AsyCS_{SS} \leq AsyCS_{PA} < 1 - \alpha. \tag{3.4}$$
This equation summarizes several important results. First, it shows that, under the presence of local misspecification and relatively mild conditions, all of the inferential methods are asymptotically distorted, that is, as the sample size grows, all of the tests over-reject the null hypothesis and therefore lead to CSs that under-cover the true parameter value. Second, the equation reveals that the PA test suffers the least amount of asymptotic confidence size distortion. This is expected, since this test constructs its critical values in a conservative fashion, treating each inequality as binding without using information in the data.

Equation (3.4) also shows that the subsampling and GMS CSs share the same amount of asymptotic distortion. From the results in AS, we know that GMS tests are as powerful as subsampling tests along any sequence of local alternative models. One would then expect the

GMS CS to have greater or equal asymptotic distortion than the subsampling CS. Moreover, AS show that GMS tests are strictly more powerful than subsampling tests along some sequences of local alternative models. One might then suspect that this results would translate in the GMS CS having a strictly larger asymptotic distortion than the subsampling CS in the context of locally misspecified models. Equation (3.4) shows that this is not the case. Intuitively, even though the GMS and subsampling tests 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 sizes, i.e., the ones that attain the infimum in (2.17). In particular, 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. When combined with the results regarding power against local alternatives in AS, our results indicate that the GMS test is preferable to the subsampling test: there is a gain in asymptotic power against certain local alternatives without sacrificing in terms of asymptotic confidence size when the model is locally misspecified.

According to Equation (3.4), the PA CS is the most robust among the test considered in this section. However, PA CSs are conservative in many cases in which GMS and subsampling CSs are not, and so the price for being robust against local misspecification can be quite high if the model is correctly specified.

3.2 Comparison across Test Statistics

In this section we analyze the relative performance in terms of asymptotic confidence size of CSs based on the test functions S_1 and S_2 defined in (2.22) and (2.24), respectively. The main result of this section has two parts. First, we show that the $AsyCS$ of the test function S_1 is strictly positive for any PA, GMS or subsampling critical value. Second, we show that the $AsyCS$ of the test function S_2 can be arbitrarily close to zero, again for all critical values. The next theorem states these results formally.

Theorem 3.2. *Suppose the assumptions in Lemma B.2 hold.*

1. *There exists $B > 0$ such that whenever r in the definition of $\mathcal{F}_n$ in Equation (2.14) satisfies $r^* \leq B$*
$$AsyCS^{(1)}_{GMS} > 0. \tag{3.5}$$

2. *Suppose also that Assumption A.7 holds. Then, for every r in the definition of $\mathcal{F}_n$ with $r^* > 0$ and every $\eta > 0$, there exists $\varepsilon > 0$ in $\Psi_{2,\varepsilon}$ such that*
$$AsyCS^{(2)}_{PA} \leq \eta. \tag{3.6}$$

There are several important lessons from Theorem 3.2. First, by Theorems 3.1 and 3.2 it follows that the asymptotic confidence size of the CSs based on S_1 are positive for any critical value, provided the level of misspecification is not too big, i.e. $r^* \leq B$. Second, by

Theorems 3.1 and 3.2 it follows that the test function S_2 results in CSs whose asymptotic confidence size are arbitrarily small as ε in $\Psi_{2,\varepsilon}$ is chosen small. This is, the test function S_2 is severely affected by the smallest amount of misspecification while the test function S_1 has a positive confidence size. Combining these two results we derive the following corollary.

Corollary 3.1. *Suppose all the assumptions in Theorems 3.1 and 3.2 hold. Then, for $l \in \{PA, GMS, SS\}$ there exists $B > 0$ and $\epsilon > 0$ such that whenever $r^* \leq B$,*

$$AsyCS_l^{(2)} < AsyCS_l^{(1)}. \tag{3.7}$$

The corollary states that the test function S_1 results in CSs that are more robust than those based on the test function S_2 for any PA, GMS and subsampling critical value. 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 Equation (3.7) should hold. However, Theorem 3.2 quantifies this relationship by showing that the cost of having higher power under correct specification is an arbitrarily low asymptotic confidence size under local misspecification.

Remark 3.1. Given that generalized empirical likelihood (GEL) test statistics are asymptotically equivalent to $T_{2,n}(\theta)$ up to first order (see AG and Canay (2010)), the results from Theorem 3.2 also hold for CSs based on GEL test statistics.

To understand the intuition behind Theorem 3.2 it is enough to consider the case where $p = k = 2$ together with the limit of the PA critical value $c_0(\Omega, 1-\alpha)$. In this simple case, it follows from Lemma B.1 that

$$AsyCS_{PA}^{(1)} \leq \Pr([Z_1^\star - r_1]_-^2 + [-Z_1^\star]_-^2 \leq c_0(\Omega, 1-\alpha)), \tag{3.8}$$

where $Z^\star \sim N(0, \Omega)$ and $\Omega \in \Psi_1$ is a correlation matrix with off-diagonal elements $\rho = -1$. Theorem 3.2 shows that $AsyCS_{PA}^{(1)}$ is strictly positive provided the amount of misspecification is not too big (i.e., $r^* \leq B$). The reason why some condition on r^* must be placed is evident: if the amount of misspecification is really big there is no way to bound the asymptotic distortion. To illustrate this, suppose $r_1 > (2c_0(\Omega, 1-\alpha))^{1/2}$ and let $A \equiv [Z_1^\star - r_1]_-^2$ and $B \equiv [-Z_1^\star]_-^2$ so that the RHS of Equation (3.8) is $\Pr(A + B \leq c_0(\Omega, 1-\alpha))$. On the one hand, if $Z_1^\star \notin [0, r_1]$ it follows that either $B = 0$ and $A > c_0(\Omega, 1-\alpha)$ or $A = 0$ and $B > c_0(\Omega, 1-\alpha)$. On the other hand, if $Z_1^\star \in [0, r_1]$, $A + B = (Z_1^\star - r_1)^2 + Z_1^{\star 2} \geq r_1^2/2 > c_0(\Omega, 1-\alpha)$. We can then conclude that

$$\Pr([Z_1^\star - r_1]_-^2 + [-Z_1^\star]_-^2 \leq c_0(\Omega, 1-\alpha)) = 0, \tag{3.9}$$

meaning that $AsyCS_{PA}^{(1)} = 0$ when $r^* > (2c_0(\Omega, 1-\alpha))^{1/2}$. For this level of r^*, $AsyCS_{PA}^{(2)} = 0$ as well so both test statistics suffer from the maximum amount of distortion. Therefore, in order to get meaningful results we must restrict the magnitude of r^* as in Theorem 3.2.

In addition, Theorem 3.2 shows that $AsyCS_{PA}^{(2)}$ can be arbitrarily close to zero when ε in the space $\Psi_{2,\varepsilon}$ is small. To illustrate this, consider the case where $p = k = 2$ together with the limit of the PA critical value $c_0(\Omega, 1-\alpha)$. By $\Omega \in \Psi_{2,\varepsilon}$, the off-diagonal element ρ of the correlation matrix Ω has to lie in $[-(1-\varepsilon)^{1/2}, (1-\varepsilon)^{1/2}]$. It follows from Lemma B.1 that

$$AsyCS_{PA}^{(2)} \leq \Pr(\tilde{S}_2(Z^\star, r_1, \Omega) \leq c_0(\Omega, 1-\alpha)), \tag{3.10}$$

where $Z^\star \sim N(0, \Omega)$, $\Omega \in \Psi_{2,\varepsilon}$ with $\rho = -(1-\varepsilon)^{1/2}$ and

$$\tilde{S}_2(Z^\star, r_1, \Omega) = \frac{1}{\varepsilon} \inf_{t \in \mathbb{R}^2_{+,+\infty}} \{\sum_{j=1}^{2} (Z_j^\star - r_1 - t_j)^2 + 2(1-\varepsilon)^{1/2}(Z_1^\star - r_1 - t_1)(Z_2^\star - r_1 - t_2)\}. \tag{3.11}$$

The solution to the above optimization problem can be divided in four cases (see Lemma B.3 for details), depending on the value of the realizations (z_1, z_2) of $(Z_1^\star, Z_2^\star)$. However, there exists a set $A \subset \mathbb{R}^2$ such that for all $(z_1, z_2) \in A$

$$\tilde{S}_2(z, r_1, \Omega) = \tilde{S}_2(z, 0, \Omega) + \frac{2}{1-(1-\varepsilon)^{1/2}}[-r_1(z_1 + z_2 - r_1)], \tag{3.12}$$

$[-r_1(z_1 + z_2 - r_1)] > 0$ and $\Pr(Z^\star \in A) \to 1$ as $\varepsilon \to 0$. It is immediate to see from the second term in Equation (3.12) that when ε is small (i.e., correlation close to -1), small distortions will shift the location of $\tilde{S}_2(Z^\star, r_1, \Omega)$ upwards arbitrarily on A. As a result

$$\Pr(\tilde{S}_2(Z^\star, r_1, \Omega) \leq c_0(\Omega, 1-\alpha) | A) \to 0, \tag{3.13}$$

as $\varepsilon \to 0$ since it can be shown that the critical value $c_0(\Omega, 1-\alpha)$ is uniformly bounded in $\Omega \in \Psi_2$. Therefore, Equation (3.10) implies that CSs based on S_2 have asymptotic confidence size arbitrarily close to zero when ε is small.

We quantify the results in Theorem 3.2 by numerically computing the asymptotic confidence size of the CSs based on S_1 and S_2 using the formulas provided in Lemma B.2. Table 1 reports the cases where $p \in \{2, 4, 6, 8, 10\}$, $k = 0$, $\varepsilon \in \{0.10, 0.05\}$ and $r^* \in \{0.25, 0.50, 1.00\}$. Since the computation of AsyCS involves taking an infimum over correlation matrices (see Lemma B.2), we compute this magnitude numerically by taking the infimum over 15000 randomly generated correlation matrices in Ψ_1 and $\Psi_{2,\varepsilon}$.

Table 1 shows that S_2 is significantly distorted even for values of ε that are big, i.e., $\varepsilon = 0.1$. For example, when $p = 2$ and $r^* = 0.5$, the AsyCS of the test function S_1 is 0.80 while the AsyCS of S_2 is 0.10 at best. Note that the coverage should deteriorate as p increases. Table 1 shows this clearly for S_1, but less clear for S_2. The reason is that finding the worst possible correlation matrix becomes more complicated as the dimension increases, and so for $p \geq 8$ the results are relatively optimistic for S_2. However, even for these cases the AsyCS of S_2 is significantly below that of S_1.

		$AsyCS_{PA}^{(1)}$	$AsyCS_{PA}^{(2)}$	
p	r^*		$\varepsilon = 0.10$	$\varepsilon = 0.05$
2	0.25	0.888	0.637	0.351
	0.50	0.800	0.101	0.003
	1.00	0.502	0.000	0.000
4	0.25	0.866	0.588	0.314
	0.50	0.739	0.071	0.001
	1.00	0.256	0.000	0.000
6	0.25	0.847	0.631	0.347
	0.50	0.674	0.091	0.002
	1.00	0.153	0.000	0.000
8	0.25	0.830	0.713	0.441
	0.50	0.617	0.134	0.009
	1.00	0.082	0.000	0.000
10	0.25	0.804	0.720	0.461
	0.50	0.571	0.124	0.010
	1.00	0.050	0.000	0.000

Table 1: Asymptotic Confidence Size for CSs based on the test functions S_1 and S_2 with a PA critical value. Computations were carried over 15000 random correlation matrices in Ψ_1 and $\Psi_{2,\varepsilon}$, respectively.

4 Numerical Simulations

To be added soon.

Appendices

Appendix A Additional Notation and Assumptions

To determine the asymptotic confidence size (2.17) of the CSs 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). The following definition provides the details.

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 $\{\gamma_{\omega_n,h}\}_{n\geq 1} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}$ that satisfies[5]

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

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 $\{\gamma_{\omega_n,h}\}_{n\geq 1} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}$ 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 constrained to satisfy the conditions on the correlations matrix. Note that the set H depends on the choice of S through Ψ. Note that $b/n \to 0$ implies that for $(g_1, h) \in GH$ it follows that $g_{1,j} = 0$ whenever $h_{1,j}$ is finite ($j = 1,\ldots,k$). In particular, $g_{1,j} = 0$ for $j > p$ by (2.14)(iii). Analogous statements hold for ΠH.

Lemma B.2 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

$$J_h \sim S(h_2^{1/2}Z + h_1, h_2) \tag{A-4}$$

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.

The following Assumptions A.1-A.3 are taken from AG with Assumption 2 strengthened. Assumption A.4(a)-(c) combines Assumptions GMS1 and GMS3 in AS. In the assumptions below, the set Ψ is as in condition (v) of Equation (2.14).

[5] Note that the definitions of the sets H and GH differ somewhat from the ones given in AG. E.g. in AG, GH is defined as a subset of $H \times H$ whereas here, we do not repeat the component h_2. Also, the dimension of h_2 in AG is smaller than here replacing h_2 by $vech_*(h_2)$ (which denotes the vector of elements of h_2 that lie below the main diagonal). We use h_2 as defined because it simplifies notation.

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

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

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

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

(d) $S(m, \Omega)$ *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* $Z \sim N(0_k, \Omega)$, *the distribution function (df) of* $S(Z + h_1, \Omega)$ *at* $x \in \mathbb{R}$ *is,*

(a) continuous for $x > 0$,

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

(c) less than or equal to 1/2 at $x = 0$ *whenever* $v \geq 1$ *or* $v = 0$ *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 is a* $h^* = (h_1^*, h_2^*) \in H$ *for which* $J_{h^*}(c_0(h_2^*, 1 - \alpha)) < 1 - \alpha$.

Assumption A.7. *Let* $k \geq 2$. *For any* $\Omega \in \Psi_{2,\varepsilon}$ *with smallest off diagonal element equal to* $-\sqrt{1 + \varepsilon}$, *there exists a subsequence* $\{\omega_n\}_{n\geq 1}$ *of* $\mathbb{N}$ *and a sequence* $\{\theta_{\omega_n}, F_{\omega_n}\}_{n\geq 1}$ *with* $(\theta_{\omega_n}, F_{\omega_n}) \in \mathcal{F}_{\omega_n}$ *for every* $n \in \mathbb{N}$, *such that* $Corr_{F_{\omega_n}}(m(W, \theta_{\omega_n})) \to \Omega$ *as* $n \to \infty$, $\omega_n^{1/2} \sigma^{-1}_{F_{\omega_n}, j}(\theta_{\omega_n}) E_{F_{\omega_n}}(m_j(W, \theta_{\omega_n})) \to h_{1,j}$ *as* $n \to \infty$, *and* $h_{1,j} < 0$ *for at least two values of* $j = 1, \ldots, k$.

We do not impose Assumption 4 in AG 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, by Assumption A.2(c) $c_0(\Omega, 1 - \alpha) = 0$ implies $v = 0$ and $h_{1,j} \neq 0$ for $j = 1, \ldots, p$ and therefore this case does not occur in Assumption 4 in AG. Therefore, Assumption 4(a) in AG follows from our Assumption A.2(a). 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. To do so, consider a sequence $\{\Omega_n\}_{n\geq 1}$ such that $\Omega_n \to \Omega$ for a $\Omega \in \Psi$. We need to show that $c_0(\Omega_n, 1 - \alpha) \to c_0(\Omega, 1 - \alpha)$. Let Z_n and Z 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 $S(Z_n, \Omega_n) \to_d S(Z, \Omega)$. The latter implies that $\Pr(S(Z_n, \Omega_n) \leq x) \to \Pr(S(Z, \Omega) \leq x)$ for all continuity points $x \in \mathbb{R}$ of the function $f(x) \equiv \Pr(S(Z, \Omega) \leq x)$. The convergence therefore certainly holds for all $x > 0$. Furthermore, by Assumption A.2(b) f is strictly increasing for $x > 0$. By Assumption A.2(c) it follows that $c_0(\Omega, 1 - \alpha) > 0$. By Lemma 5(a) in AG, it then follows that $c_0(\Omega_n, 1 - \alpha) \to c_0(\Omega, 1 - \alpha)$.

Note that S_1 and S_2 satisfy Assumption A.2 which is a strengthened version of Assumption 2 from AG using the same proof as in AG. Assumption 3 implies that $S(\infty_p, \Omega) = 0$ when $v = 0$. Assumption A.6 holds by Assumption A.1 if $J_{h^*}(c_0(h_2^*, 1 - \alpha)) < J_{(0,h_2^*)}(c_0(h_2^*, 1 - \alpha))$ for a $h^* \in H$. 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 $j \leq p$ or $h^*_{1,j} \neq 0$ for some $j > p$. Assumption A.7 guarantees that correlation matrices with strong negative correlations can be achieved by a sequence that implies some level of misspecification. Requiring that any Ω can be achieved is stronger than needed but simplifies the proofs significantly.

Appendix B Auxiliary Lemmas

Lemma B.1. *Assume the parameter space is given by $\mathcal{F}_n$ in Equation* (2.14) *and S satisfies Assumption A.1. Under any sequence $\{\gamma_{\omega_n,h}\}_{n\geq 1} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}$ defined in definition A.1 for a subsequence $\{\omega_n\}_{n\geq 1}$ and $h = (h_1, h_2)$, it follows*

$$T_{\omega_n}(\theta_{\omega_n,h}) \to_d J_h \sim S(h_2^{1/2} Z + h_1, h_2), \tag{B-1}$$

where $T_n(\cdot)$ is the test statistic associated with S and $Z = (Z_1, \ldots, Z_k) \sim N(0_k, I_k)$.

Lemma B.2. *Consider confidence intervals with nominal confidence size $1-\alpha$ for $0 < \alpha < 1/2$. Assume the nonempty parameter space is given by $\mathcal{F}_n$ in* (2.14) *for some $r \in \mathbb{R}^k$ with nonnegative components, $\delta > 0$, and $M < \infty$. Assume S satisfies Assumptions A.1-A.3. For GMS type 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 $b_n/n \to 0$. It follows that*

$$\begin{aligned} AsyCS_{PA} &= \inf_{h=(h_1,h_2)\in H} J_h(c_0(h_2, 1-\alpha)), \\ AsyCS_{GMS} &\in \left[\inf_{(\pi_1,h)\in \Pi H} J_h(c_{\pi_1^*}(h_2, 1-\alpha)), \inf_{(\pi_1,h)\in \Pi H} J_h(c_{\pi_1^{**}}(h_2, 1-\alpha))\right], \\ AsyCS_{SS} &= \inf_{(g_1,h)\in GH} J_h(c_{g_1}(h_2, 1-\alpha)), \end{aligned} \tag{B-2}$$

where $J_h(x) = P(J_h \leq x)$ and π_1^*, $\pi_1^{**} \in \mathbb{R}^k_{+\infty}$ with j-th element defined by

$$\pi_{1,j}^* = \infty I(\pi_{1,j} > 0) \quad \text{and} \quad \pi_{1,j}^{**} = \infty I(\pi_{1,j} = \infty), \quad j = 1, \ldots, k. \tag{B-3}$$

Lemma B.3. *For any $a \in (0,1)$ and $\rho \in [-1+a, 1-a]$ let $f(z_1, z_2, \rho)$ be defined as follows,*

$$f(z_1, z_2, \rho) = (1-\rho^2)^{-1} \min_{(u_1,u_2)\in\mathbb{R}^2_{+,\infty}} \left\{(z_1-u_1)^2 + (z_2-u_2)^2 - 2\rho(z_1-u_1)(z_2-u_2)\right\}. \tag{B-4}$$

Then $f(z_1, z_2, \rho)$ takes values according to the following four cases:

1. *$z_1 \geq 0$, $z_2 \geq 0$. Then, $f(z_1, z_2, \rho) = 0$.*
2. *$z_1 \geq 0$, $z_2 < 0$. If $\rho \leq z_1/z_2$, then*

   $$f(z_1, z_2, \rho) = (1-\rho^2)^{-1}[z_1^2 + z_2^2 - 2\rho z_1 z_2]. \tag{B-5}$$

   If $\rho > z_1/z_2$, then $f(z_1, z_2, \rho) = z_2^2$.
3. *$z_1 < 0$, $z_2 \geq 0$. If $\rho \leq z_2/z_1$, then Equation* (B-5) *holds. Otherwise $f(z_1, z_2, \rho) = z_1^2$.*
4. *$z_1 < 0$, $z_2 < 0$. If $\rho \leq \min\{z_1/z_2, z_2/z_1\}$, then Equation* (B-5) *holds. Otherwise $f(z_1, z_2, \rho) = \max\{z_1^2, z_2^2\}$.*

Lemma B.4. *Suppose that $k = p = 2$ and for any $\beta > 0$ let $\bar{H}_\beta$ be defined as,*

$$\bar{H}_\beta \equiv \{(h_1, h_2) \in \mathbb{R}^2 \times \Psi_1 : h_{1,1} \leq -\beta,\ h_{1,2} \leq -\beta,\ h_2 = (1, \rho; \rho, 1),\ \rho \leq -\beta\}. \tag{B-6}$$

Also, define the set $A_{h_1,\rho} \equiv A^a_{h_1,\rho} \cup A^b_{h_1,\rho} \subseteq \mathbb{R}^2$, where

$$A^a_{h_1,\rho} = \{z \in \mathbb{R}^2 : z_1 \geq 0, z_2 < 0, z_1 - \rho z_2 \leq -h_{1,1} + \rho h_{1,2}\} \tag{B-7}$$

$$A^b_{h_1,\rho} = \{z \in \mathbb{R}^2 : z_1 < 0, z_2 \geq 0, z_2 - \rho z_1 \leq -h_{1,2} + \rho h_{1,1}\}. \tag{B-8}$$

Then,

1. *$\forall \eta > 0$, $\exists \bar{\rho} > -1$ such that $\inf_{(h_1, h_2 = \bar{h}_2)\in \bar{H}_\beta} \Pr(Z_{\bar{h}_2} \in A_{h_1,\bar{\rho}}) \geq 1-\eta$, where $\bar{h}_2 = (1, \bar{\rho}; \bar{\rho}, 1)$.*

2. There exists a function $\tau(z,h_1,h_2)$ such that $\forall z \in A_{h_1,\rho}$, $\inf_{(h_1,h_2)\in \bar{H}_\beta} \tau(z,h_1,h_2) > 0$ and

$$S_2(z+h_1,h_2) = S_2(z,h_2) + \frac{1}{1-\rho^2}\tau(z,h_1,h_2). \tag{B-9}$$

Appendix C Proof of the Lemmas and Theorems

Proof of Lemma B.1. The proof follows along the lines of the proof of Theorem 1 in AG. By Lemma 1 in AG we have for any $s \in N$

$$T_s(\theta_s) = S\Big(\hat{D}_s^{-1/2}(\theta_s)s^{1/2}\bar{m}_s(\theta_s),\ \hat{D}_s^{-1/2}(\theta_s)\hat{\Sigma}_s(\theta_s)\hat{D}_s^{-1/2}(\theta_s)\Big). \tag{C-1}$$

For $j = 1,\dots,k$, define $A_{s,j} = s^{1/2}(\bar{m}_{s,j}(\theta_s) - E_{F_s}\bar{m}_{s,j}(\theta_s))/\sigma_{F_s,j}(\theta_s)$. As in Lemma 2 in AG, applied to (A.3)(x) in that paper, we have that

$$\begin{aligned}
&\text{(i) } A_{\omega_n} = (A_{\omega_n,1},\dots,A_{\omega_n,k})' \to_d Z_{h_2} = (Z_{h_2,1},\dots,Z_{h_2,k})' \sim N(0_k,h_2) \text{ as } n\to\infty,\\
&\text{(ii) } \hat{\sigma}_{\omega_n,j}(\theta_{\omega_n,h})/\sigma_{F_{\omega_n,h},j}(\theta_{\omega_n,h}) \to_p 1 \text{ as } n\to\infty \text{ for } j=1,\dots,k,\\
&\text{(iii) } \hat{D}_{\omega_n}^{-1/2}(\theta_{\omega_n,h})\hat{\Sigma}_{\omega_n}(\theta_{\omega_n,h})\hat{D}_{\omega_n}^{-1/2}(\theta_{\omega_n,h}) \to_p h_2 \text{ as } n\to\infty.
\end{aligned} \tag{C-2}$$

under any sequence $\gamma_{\omega_n,h} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\ge 1}$. These results follow after completing the subsequence $\gamma_{\omega_n,h} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\ge 1}$. For $s \in \mathbb{N}$ define the sequence $\{\theta_s, F_s\}_{s\ge 1}$ as follows. For any $s \le \omega_1$, $(\theta_s, F_s) = (\theta_{\omega_1,h}, F_{\omega_1,h})$. For any $s > \omega_1$ and since $\{\omega_n\}_{n\ge 1}$ is a subsequence of $\mathbb{N}$, there exists a unique $m \in \mathbb{N}$ such that $\omega_{m-1} < s \le \omega_m$. For every such s, set $(\theta_s, F_s) = (\theta_{\omega_m,h}, F_{\omega_m,h})$. Now let $\{W_i\}_{i\le n}$ be i.i.d. under F_s. By construction, $\forall s \in \mathbb{N}$, $(\theta_s, F_s) \in \mathcal{F}_{\omega_m}$ for some $m \in \mathbb{N}$ and $Corr_{F_s}(m(W_i,\theta_s)) \to h_2$. Then, the results (i)-(iii) of Equation (C-2) hold from standard CLT and LLN with ω_n, $\theta_{\omega_n,h}$, and $F_{\omega_n,h}$ replaced by s, θ_s, and F_s respectively. But the convergence results along $\{\theta_s, F_s\}_{s\ge1}$ then imply convergence along the subsequence $\{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\ge 1}$ as by construction the latter coincides with the former on indices $s = \omega_n$.

From (C-2), the jth element of $\hat{D}_{\omega_n}^{-1/2}(\theta_{\omega_n,h})\omega_n^{1/2}\bar{m}_{\omega_n}(\theta_{\omega_n,h})$ equals $(A_{\omega_n,j} + \omega_n^{1/2}E_{F_{\omega_n,h}}\bar{m}_{\omega_n,j}(\theta_{\omega_n,h})/\sigma_{F_{\omega_n,h},j}(\theta_{\omega_n,h}))\ \times(1+o_p(1))$. We next consider a k-vector-valued function of $\hat{D}_{\omega_n}^{-1/2}(\theta_{\omega_n,h})\omega_n^{1/2}\bar{m}_{\omega_n}(\theta_{\omega_n,h})$ that converges in distribution whether or not some elements of h_1 equal ∞. Write the right-hand side of (C-1) as a continuous function of this k-vector and apply the continuous mapping theorem. Let $G(\cdot)$ be a strictly increasing continuous df on $\mathbb{R}$, such as the standard normal df, and let $G(\infty) = 1$. For $j = 1,\dots,k$, we have

$$\begin{aligned}
G_{\omega_n,j} &\equiv G\left(\hat{\sigma}_{\omega_n,j}^{-1}(\theta_{\omega_n,h})\omega_n^{1/2}\bar{m}_{\omega_n,j}(\theta_{\omega_n,h})\right)\\
&= G\left(\hat{\sigma}_{\omega_n,j}^{-1}(\theta_{\omega_n,h})\sigma_{F_{\omega_n,h},j}(\theta_{\omega_n,h})\left[A_{\omega_n,j} + \omega_n^{1/2}E_{F_{\omega_n,h}}\bar{m}_{\omega_n,j}(\theta_{\omega_n,h})/\sigma_{F_{\omega_n,h},j}(\theta_{\omega_n,h})\right]\right).
\end{aligned} \tag{C-3}$$

If $h_{1,j} < \infty$ then

$$G_{\omega_n,j} \to_d G\left(Z_{h_2,j} + h_{1,j}\right) \tag{C-4}$$

by (C-3), (C-2), the definition of $\gamma_{\omega_n,h}$, and the continuous mapping theorem. If $h_{1,j} = \infty$ (which can only happen for $j = 1,\dots,p$), then

$$G_{\omega_n,j} = G\left(\hat{\sigma}_{\omega_n,j}^{-1}(\theta_{\omega_n,h})\omega_n^{1/2}\bar{m}_{\omega_n,j}(\theta_{\omega_n,h})\right) \to_p 1 \tag{C-5}$$

by (C-3), $A_{\omega_n,j} = O_p(1)$, and $G(x) \to 1$ as $x \to \infty$. The results in (C-4)-(C-5) hold jointly and combine to give

$$G_{\omega_n} \equiv (G_{\omega_n,1},\dots,G_{\omega_n,k})' \to_d (G(Z_{h_2,1}+h_{1,1}),\dots,G(Z_{h_2,k}+h_{1,k}))' \equiv G_\infty, \tag{C-6}$$

where $G(Z_{h_2,j} + h_{1,j}) = 1$ by definition when $h_{1,j} = \infty$. Let G^{-1} denote the inverse of G. For $x =$

$(x_1, \ldots, x_k)' \in \mathbb{R}^p_{+\infty} \times \mathbb{R}^v$, let $G_{(k)}(x) = (G(x_1), \ldots, G(x_k))' \in (0,1]^p \times (0,1)^v$. For $y = (y_1, \ldots, y_k)' \in (0,1]^p \times (0,1)^v$, let $G^{-1}_{(k)}(y) = (G^{-1}(y_1), \ldots, G^{-1}(y_k))' \in \mathbb{R}^p_{+\infty} \times \mathbb{R}^v$. Define $S^*(y, \Omega) = S(G^{-1}_{(k)}(y), \Omega)$ for $y \in (0,1]^p \times (0,1)^v$ and $\Omega \in \Psi$. By Assumption 1(d) in AG, $S^*(y, \Omega)$ is continuous at all (y, Ω) for $y \in (0,1]^p \times (0,1)^v$ and $\Omega \in \Psi$. We now have

$$
\begin{aligned}
T_{\omega_n}(\theta_{\omega_n,h}) &= S\Big(G^{-1}_{(k)}(G_{\omega_n}),\ \hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\hat{\Sigma}_{\omega_n}(\theta_{\omega_n,h})\hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\Big) \\
&= S^*\Big(G_{\omega_n},\ \hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\hat{\Sigma}_{\omega_n}(\theta_{\omega_n,h})\hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\Big) \\
&\to_d \ S^*(G_\infty, h_2) \\
&= S(G^{-1}_{(k)}(G_\infty), h_2) \\
&= S(Z_{h_2} + h_1, h_2) \sim J_h, \qquad \text{(C-7)}
\end{aligned}
$$

where the convergence holds by (C-2), (C-6), and the continuous mapping theorem, the last equality holds by the definitions of $G^{-1}_{(k)}$ and G_∞ and the last line hold by definition of J_h. □

Proof of Lemma B.2. For any of the CSs considered in Section 2.1, there is a sequence $\{\theta_n, F_n\}_{n\geq 1}$ with $(\theta_n, F_n) \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$ such that $AsyCS = \liminf_{n\to\infty} \Pr_{\theta_n, F_n}(T_n(\theta_n) \leq c_n(\theta_n, 1-\alpha))$. We can then find a subsequence $\{\omega_n\}_{n\geq 1}$ of $\mathbb{N}$ such that

$$
\lim_{n\to\infty} \Pr_{\theta_{\omega_n}, F_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n}) \leq c_{\omega_n}(\theta_{\omega_n}, 1-\alpha)) = AsyCS \qquad \text{(C-8)}
$$

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

This proves that in order to determine the asymptotic confidence size of the CSs, we only have to be concerned about the limiting coverage probabilities under sequences of the type $\{\gamma_{\omega_n,h}\}_{n\geq 1}$ for PA, $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$ for GMS, and $\{\gamma_{\omega_n,g_1,h}\}_{n\geq 1}$ for SS. From Lemma B.1 we know that the limiting distribution of the test statistic under a sequence $\{\gamma_{\omega_n,h}\}_{n\geq 1}$ is $J_h \sim S(Z_{h_2} + h_1, h_2)$. By Assumption A.1(a) it follows that for given h_2 the $1-\alpha$ quantiles of J_h do 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

$$
\hat{h}_{2,\omega_n} = \hat{\Omega}_{\omega_n}(\theta_{\omega_n,h}) \qquad \text{(C-9)}
$$

and $\hat{\Omega}_s(\theta) = (\hat{D}_s(\theta))^{-1/2}\hat{\Sigma}_s(\theta)(\hat{D}_s(\theta))^{-1/2}$. From (C-2)(iii) we know that under $\{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}$, 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) and the comment to Lemma 5 in AG, 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, $AsyCS_{PA} = J_h(c_0(h_2, 1-\alpha))$ for some $h \in H$, which implies $AsyCS_{PA} \geq \inf_{h\in H} J_h(c_0(h_2, 1-\alpha))$. However, Equation (C-8) implies that $AsyCS_{PA} = \inf_{h\in H} \lim_{n\to\infty} \Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n}) \leq c_0(\hat{h}_{2,\omega_n}, 1-\alpha))$. 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 $S(\hat{h}^{1/2}_{2,\omega_n} Z + \varphi(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n})$ 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 $\{\gamma_{\omega_n}\}$

$$\begin{aligned}\hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n},1-\alpha) &\geq c^*_{\omega_n} \to_p c_{\pi_1^*}(h_2,1-\alpha),\\ \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n},1-\alpha) &\leq c^{**}_{\omega_n} \to_p c_{\pi_1^{**}}(h_2,1-\alpha).\end{aligned} \tag{C-10}$$

We begin by showing the first line in Equation (C-10). 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 j-th component given by

$$\varphi^*_j(\xi,\Omega)=\begin{cases}\varphi_j(\xi,\Omega) & \text{if } \pi_{1,j}=0 \text{ and } j\leq p,\\ \infty & \text{if } \pi_{1,j}>0 \text{ and } j\leq p,\\ 0 & \text{if } j=p+1,\ldots,k.\end{cases} \tag{C-11}$$

Define $c^*_{\omega_n}$ as the $1-\alpha$ quantile of $S(\hat{h}^{1/2}_{2,\omega_n}Z+\varphi^*(\xi_{\omega_n}(\theta_{\omega_n},\hat{h}_{2,\omega_n})),\hat{h}_{2,\omega_n})$. 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 Supplementary Appendix of AS 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 Equation (C-10).

Next consider the second line in Equation (C-10). Suppose that either $v\geq 1$ or $v=0$ and $\pi_1^{**}\neq\infty_p$. Define

$$\varphi^{**}_j(\xi,\Omega)=\begin{cases}\min\{0,\varphi_j(\xi,\Omega)\} & \text{if } \pi_{1,j}<\infty \text{ and } j\leq p,\\ \varphi_j(\xi,\Omega) & \text{if } \pi_{1,j}=\infty \text{ and } j\leq p,\\ 0 & \text{if } j=p+1,\ldots,k,\end{cases} \tag{C-12}$$

and define $c^{**}_{\omega_n}$ as the $1-\alpha$ quantile of $S(\hat{h}^{1/2}_{2,\omega_n}Z+\varphi^{**}(\xi_{\omega_n}(\theta_{\omega_n},\hat{h}_{2,\omega_n})),\hat{h}_{2,\omega_n})$. Note that the definition of $\varphi^{**}_j(\xi,\Omega)$ implies that $\varphi^{**}_j\leq\varphi_j$. The same steps as in the proof of Lemma 2(a) of AS can be used to prove the second line of Equation (C-10). In particular, note that 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, $0\leq S(\hat{h}^{1/2}_{2,\omega_n}Z+\varphi(\xi_{\omega_n}(\theta_{\omega_n},\hat{h}_{2,\omega_n})),\hat{h}_{2,\omega_n})\to_p 0$. This implies the second line in Equation (C-10).

Having proven Equation (C-10), we now prove the second line in Equation (B-2). Consider first the case $(\pi_1,h)\in\Pi H$ such that $c_{\pi_1^*}(h_2,1-\alpha)>0$. In this case, it follows from (C-10) and Lemma 5 in AG that

$$\begin{aligned}\liminf_{n\to\infty}\Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n})\leq\hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n},1-\alpha)) &\leq \liminf_{n\to\infty}\Pr_{\gamma_{\omega_n,h}}(T_{\omega_n}(\theta_{\omega_n})\leq c^{**}_{\omega_n})\\ &= J_h(c_{\pi_1^{**}}(h_2,1-\alpha)).\end{aligned} \tag{C-13}$$

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 $\alpha<0.5$, 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, since $\kappa_n\to\infty$, this 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

$$1\geq\liminf_{n\to\infty}\Pr_{\gamma_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n})\leq\hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n},1-\alpha))\geq\liminf_{n\to\infty}\Pr_{\gamma_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n})\leq 0)=J_h(0)=1, \tag{C-14}$$

where we apply the argument in (A.12) of AG for the first equality and use Assumption A.3 for the second equality. Therefore, $\liminf_{n\to\infty}\Pr_{\gamma_{\omega_n}}(T_{\omega_n}(\theta_{\omega_n})\leq\widehat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n},1-\alpha))=1$. Note that when $h_1=\infty_p$, $J_h(c)=1$ for any $c\geq 0$. The last statement and Equations (C-8), (C-13), and (C-14) complete the proof of the lemma.

Subsampling critical value: Instead of $\{\gamma_{\omega_n,g_1,h}\}_{n\geq1} = \{\theta_{\omega_n,g_1,h}, F_{\omega_n,g_1,h}\}_{n\geq1}$ 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 $\gamma_3 = (\theta, F)$, $\gamma_1 = (\sigma^{-1}_{F,j}(\theta) E_F m_j(W_i, \theta))^k_{j=1} \in \mathbb{R}^k$, and $\gamma_2 = Corr_F(m(W_i, \theta)) \in \mathbb{R}^{k\times k}$ for (θ, F) introduced in the model defined in (2.14). Then, Assumption A0 in AG clearly holds. With $\{\gamma_{\omega_n,h}\}_{n\geq1}$ and H defined in definition A.1, Assumption B0 then holds by Lemma B.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 (9.4) in AG) which is the set of points $(g_1, h) \in GH$ such that for all sequences $\{\gamma_{w_n,g_1,h}\}_{n\geq1}$

$$\liminf_{n\to\infty} \Pr_{\gamma_{w_n,g_1,h}} (T_{w_n}(\theta_{w_n,g_1,h}) \leq c_{w_n,b_{w_n}}(\theta_{w_n,g_1,h}, 1-\alpha)) \geq J_h(c_{g_1}(h_2, 1-\alpha)). \tag{C-15}$$

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 (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 Lemma B.3. The FOC associated with the minimizers u_1 and u_2 in Equation (B-4) are

$$f'_{u_1} = -2(z_1 - u_1) + 2\rho(z_2 - u_2) \geq 0, \quad u_1 \times f'_{u_1} = 0, \quad u_1 \geq 0, \tag{C-16}$$
$$f'_{u_2} = -2(z_2 - u_2) + 2\rho(z_1 - u_1) \geq 0, \quad u_2 \times f'_{u_2} = 0, \quad u_2 \geq 0. \tag{C-17}$$

The SOC are immediately satisfied as the function on the RHS of Equation (B-4) is strictly convex for $\rho \in [-1+a, 1-a]$.

Consider Case 1. In this case, $u_1 = z_1$ and $u_2 = z_2$ satisfies Equations (C-16) and (C-17) and $f(z_1, z_2, \rho) = 0$ regardless of the value of ρ.

Now consider Case 2. First we note that $u_1 \geq 0$, $u_2 > 0$ is not a feasible solution as this results in $u_2 = z_2 < 0$ which is contradictory. The solution must then be of the form $u_1 \geq 0$ and $u_2 = 0$. Then, it follows from the conditions in Equation (C-16) that $u_1 \geq z_1 - \rho z_2$, so that $u_1 = \max\{z_1 - \rho z_2, 0\}$ and $u_2 = 0$. In addition, from the conditions in Equation (C-17) it follows that

$$z_2 - \rho z_1 + \rho u_1 \leq 0. \tag{C-18}$$

If $u_1 = 0$, then we need $z_2 - \rho z_1 \leq 0$ for Equation (C-18) to hold. However, $u_1 = 0$ is a valid solution provided $z_1 - \rho z_2 \leq 0$ which implies $\rho \leq z_1/z_2$. This is possible only if $z_1/z_2 \geq -1 + a$ which implies $z_2 + z_1 < 0$ and then $z_2 - \rho z_1 < z_2 + z_1 < 0$ satisfying Equation (C-18). Thus, if $\rho \leq z_1/z_2$, the unique solution is $(u_1, u_2) = (0, 0)$ and the objective function is given by Equation (B-5).

The additional possibility is that $\rho > z_1/z_2$ so that $u_1 = z_1 - \rho z_2 > 0$. In this case, Equation (C-18) holds immediately as $z_2 - \rho z_1 + \rho u_1 = (1-\rho^2)z_2 \leq 0$. Therefore, $(u_1, u_2) = (z_1 - \rho z_2, 0)$ is the unique solution and

$$f(z_1, z_2, \rho) = (1-\rho^2)^{-1} \left\{(z_1 - z_1 + \rho z_2)^2 + z_2^2 - 2\rho(z_1 - z_1 + \rho z_2)(z_2)\right\} = z_2^2. \tag{C-19}$$

Case 3 is exactly analogous to Case 2 by exchanging the subindices 1 and 2. Consider Case 4 then. First, we note again that $u_1 > 0$ and $u_2 > 0$ is not a feasible solution by the same arguments as before. Second, we note that $(u_1, u_2) = (0, 0)$ is a solution provided $\rho \leq \min\{z_1/z_2, z_2/z_1\}$, as this condition implies the correct sign of the derivatives in Equations (C-16) and (C-17). The remaining case is either $\rho > z_1/z_2$ or $\rho > z_2/z_1$. By similar steps as those used in Case 2 it follows that the solution for these cases are $(u_1, u_2) = (z_1 - \rho z_2, 0)$, $f(z_1, z_2, \rho) = z_2^2$ and $(u_1, u_2) = (0, z_2 - \rho z_1)$, $f(z_1, z_2, \rho) = z_1^2$ respectively. This completes the proof.

□

Proof of Lemma B.4. We begin by proving (1). Consider the set $A^a_{h_1,\rho}$. Note that we can always write $Z_{h_2,1} - \rho Z_{h_2,2} = \sqrt{1-\rho^2}W$ for $Z_{h_2,2} \perp W \sim N(0,1)$. Then, since $-h_{1,1} + \rho h_{1,2} \geq \beta > 0$ for $(h_1, h_2) \in \bar{H}_\beta$, it follows that

$$\Pr(Z_{h_2} \in A^a_{h_1,\rho}) = \Pr\left(Z_{h_2,2} \leq \min\{0, \frac{\sqrt{1-\rho^2}W}{-\rho}\}, W \leq \frac{-h_{1,1}+\rho h_{1,2}}{\sqrt{1-\rho^2}}\right) \to 1/2, \text{ as } \rho \to -1. \tag{C-20}$$

The same analysis applies for the set $A^b_{h_1,\rho}$ and the result immediately follows by continuity in ρ.

We now prove (2). Note that

$$S_2(z+h_1,h_2) = (1-\rho^2)^{-1} \min_{u \in \mathbb{R}^2_{+,\infty}} \left\{(z_1+h_{1,1}-u_1)^2 + (z_2+h_{1,2}-u_2)^2 - 2\rho(z_1+h_{1,1}-u_1)(z_2+h_{1,2}-u_2)\right\}, \tag{C-21}$$

is the same optimization problem as the one in Lemma B.3 by letting $\bar{z}_j = z_j + h_{1,j}$. It follows from Lemma B.3 that the solution of Equation (C-21) is,

$$S_2(z+h_1,h_2) = (1-\rho^2)^{-1}[(z_1+h_{1,1}-z_2-h_{1,2})^2 + 2(1-\rho)(z_1+h_{1,1})(z_2+h_{1,2})], \tag{C-22}$$

whenever $z \in A^a_{h_1,\rho}$ and $\rho < 0$. Consider first the following partition of the subset $A^a_{h_1,\rho}$,

$$A^{a_1}_{h_1,\rho} \equiv \{z \in \mathbb{R}^2 : z_1 \geq 0, z_2 < 0, 0 < z_1 - \rho z_2 \leq -h_{1,1} + \rho h_{1,2}\} \tag{C-23}$$

$$A^{a_2}_{h_1,\rho} \equiv \{z \in \mathbb{R}^2 : z_1 \geq 0, z_2 < 0, z_1 - \rho z_2 \leq 0\}. \tag{C-24}$$

Lemma B.3 implies that for $z \in A^{a_1}_{h_1,\rho}$, $S_2(z,h_2) = z_2^2$ and thus after some algebraic manipulations

$$S_2(z+h_1,h_2) = S_2(z,h_2) + \frac{1}{1-\rho^2}\tau_1(z,h_1,h_2), \tag{C-25}$$

where

$$\tau_1(z,h_1,h_2) = (z_1+h_{1,1}-\rho(z_2+h_{1,2}))^2 + (1-\rho^2)(h_{1,2}^2 + 2z_2h_{1,2}). \tag{C-26}$$

The term $\tau_1(z,h_1,h_2)$ is clearly positive for all $z \in A^{a_1}_{h_1,\rho}$ and $(h_1,h_2) \in \bar{H}_\beta$. We now show the statement for $z \in A^{a_2}_{h_1,\rho}$. Lemma B.3 implies that for $z \in A^{a_2}_{h_1,\rho}$, $S_2(z,h_2) = z_1^2 + (z_2-\rho z_1)^2/(1-\rho^2)$. Doing some algebraic manipulations it follows that

$$S_2(z+h_1,h_2) = S_2(z,h_2)\frac{1}{1-\rho^2}\tau_2(z,h_1,h_2), \tag{C-27}$$

where

$$\tau_2(z,h_1,h_2) = (h_{1,1}-h_{1,2})^2 + 2((z_1-\rho z_2)(h_{1,1}-\rho h_{1,2}) + h_{1,2}z_2(1-\rho^2) + (1-\rho)h_{1,1}h_{1,2}) \tag{C-28}$$

is positive for all $z \in A^{a_2}_{h_1,\rho}$ and $(h_1,h_2) \in \bar{H}_\beta$. This proves that (2) holds for $A^a_{h_1,\rho} = A^{a_1}_{h_1,\rho} \cup A^{a_2}_{h_1,\rho}$. A symmetric argument can be used for the subset $A^b_{h_1,\rho}$ and this completes the proof. □

Proof of Theorem 3.1. The proof makes use of the results in Lemma B.2. We first prove (1). Note that for $h \in H$ and $\kappa_n \to \infty$, there exists a subsequence $\{\omega_n\}_{n\geq1}$ and a sequence $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq1}$ for some $\pi_1 \in \mathbb{R}^k_\infty$ 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 $AsyCS_{PA} \geq AsyCS_{GMS}$. The result for subsampling CSs is verified analogously.

We now prove (2). Note that $AsyCS_{PA} = \inf_{h=(h_1,h_2)\in H} J_h(c_0(h_2, 1-\alpha)) \leq J_{h^*}(c_0(h_2^*, 1-\alpha)) < 1-\alpha$.

Finally, we prove (3). First, assume $(g_1, h) \in GH$. By Assumption A.1(a), it is enough to show 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 $\{\omega_n\}_{n\geq1}$ and a sequence

$\{\gamma_{\omega_n,g_1,h}\}_{n\geq 1}$. Because $\kappa_n^{-1}n^{1/2}/b^{1/2} \to \infty$ it follows that there exists a subsequence $\{v_n\}_{n\geq 1}$ of $\{\omega_n\}_{n\geq 1}$ such that under $\{\gamma_{v_n,g_1,h}\}_{n\geq 1}$

$$\kappa_{v_n}^{-1}v_n^{1/2}\sigma^{-1}_{F_{v_n,h},j}(\theta_{v_n,h})E_{F_{v_n,h}}m_j(W_i,\theta_{v_n,h}) \to \pi_{1,j}, \tag{C-29}$$

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 $j = p+1,\ldots,k$. We have just shown 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 therefore $AsyCS_{SS} \geq AsyCS_{GMS}$.

Second, assume $(\pi_1,h) \in \Pi H$ so that $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$ exists. 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 is 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, $AsyCS_{SS} \leq AsyCS_{GMS}$. This completes the proof.

□

Proof of Theorem 3.2. **Part 1**. By Lemma B.2

$$AsyCS^{(1)}_{GMS} \geq \inf_{(\pi_1,h)\in\Pi H} \Pr\left(S_1(h_2^{1/2}Z + h_1, h_2) \leq c_{\pi_1^*}(h_2, 1-\alpha)\right), \tag{C-30}$$

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.2. Recall that

$$S_1(h_2^{1/2}Z + h_1, h_2) = \sum_{j=1}^{p}[h_2^{1/2}(j)Z + h_{1,j}]_-^2 + \sum_{j=p+1}^{k}(h_2^{1/2}(j)Z + h_{1,j})^2, \tag{C-31}$$

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 $j \geq 1$

$$\sum_{s=1}^{k}(h_2^{1/2}(j,s))^2 = 1, \quad h_2^{1/2}(j,s) = 0,\ \forall s > j, \quad |h_2^{1/2}(j,s)| \leq 1,\ \forall s \geq 1. \tag{C-32}$$

The properties in Equation (C-32) follow by h_2 having ones in the main diagonal and $h_2^{1/2}$ being lower triangular. We use Equation (C-32) and the Cauchy-Schwarz inequality to derive the following three useful inequalities. For any $z \in \mathbb{R}^k$ and $j = 1,\ldots,k$,

$$\begin{aligned}(h_2^{1/2}(j)z + h_{1,j})^2 &= \left(\sum_{s=1}^{j}h_2^{1/2}(j,s)z_s + \sum_{s=1}^{j}(h_2^{1/2}(j,s))^2h_{1,j}\right)^2 \\ &\leq \sum_{m=1}^{j}(h_2^{1/2}(j,m))^2\sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2 = \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \quad \text{(C-33)}\end{aligned}$$

$$[h_2^{1/2}(j)z + h_{1,j}]_-^2 \leq \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \tag{C-34}$$

$$[h_2^{1/2}(j)z + h_{1,j}]_-^2 \leq [h_2^{1/2}(j)z]_-^2 \leq \sum_{s=1}^{j}z_s^2, \quad \text{provided } h_{1,j} \in (0,\infty). \tag{C-35}$$

Therefore, for every $z \in \mathbb{R}^k$ and $h \in H$ define

$$\tilde{S}_1(z,h) = \sum_{j=1}^{p}\sum_{s=1}^{j} z_s^2 I(h_{1,j} \in (0,\infty)) + \sum_{j=1}^{p}\sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2 I(h_{1,j} \leq 0)$$
$$+ \sum_{j=p+1}^{k}\sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \qquad \text{(C-36)}$$

and it follows from Equations (C-33), (C-34), and (C-35) that $\tilde{S}_1(z,h) \geq S_1(h_2^{1/2}z + h_1, h_2)$ for all $z \in \mathbb{R}^k$. From the above, we conclude that the function $\tilde{S}_1(Z,h)$ in Equation (C-36) is such that

$$\Pr(S_1(h_2^{1/2}Z + h_1, h_2) \leq x) \geq \Pr(\tilde{S}_1(Z,h) \leq x), \qquad \text{(C-37)}$$

for all $h \in H$ and $x \in \mathbb{R}$. To this end,

Let $B > 0$ and

$$A_B \equiv \{z_s \in \mathbb{R} : |z_s| \leq B\} \quad \text{and} \quad A_B^k = A_B \times \cdots \times A_B \text{ (with } k \text{ copies).} \qquad \text{(C-38)}$$

Since A_B has positive length on $\mathbb{R}$, it follows that for $Z \sim N(0_k, I_k)$,

$$\Pr(Z \in A_B^k) = \prod_{s=1}^{k} \Pr(Z_s \in A_B) > 0. \qquad \text{(C-39)}$$

Let $\{\pi_{1,l}, h_l\}_{l\geq 1}$ be a sequence such that $h_l = (h_{1,l}, h_{2,l})$, $(\pi_{1,l}, h_l) \in \Pi H$ for all $l \in \mathbb{N}$ and

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

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

We now consider two cases. In the first case $\liminf_{l\to\infty} c_{\pi_{1,l}^*}(h_{2,l}, 1-\alpha) > 0$ and in the second case $\liminf_{l\to\infty} c_{\pi_{1,l}^*}(h_{2,l}, 1-\alpha) = 0$. To deal with the first case, let $B = \liminf_{l\to\infty} B_l > 0$ and 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, when $|h_{1,j}| \leq r_j$ it follows that for all $z_s \in A_{B_{\omega_l}}$ and $j = 1, \ldots, k$

$$(z_s + h_2^{1/2}(j,s)h_{1,j})^2 \leq B_{\omega_l}^2 + r^{*2} + 2B_{\omega_l}r^*, \qquad \text{(C-40)}$$

and then for all $z \in A_{B_{\omega_l}}^k$

$$\tilde{S}_1(z,h_l) \leq \sum_{j=1}^{k}\sum_{s=1}^{j} 4B_{\omega_l}^2 = 2k(k+1)B_{\omega_l}^2 = c_{\pi_{1,\omega_l}^*}(h_{2,\omega_l}, 1-\alpha). \qquad \text{(C-41)}$$

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

$$\Pr(\tilde{S}_1(Z, h_{\omega_l}) \leq c_{\pi_{1,\omega_l}^*}(h_{2,\omega_l}, 1-\alpha)) \geq \Pr(Z \in A_{B_{\omega_l}}^k) > 0. \qquad \text{(C-42)}$$

It follows from Equations (C-37), (C-39), and (C-42) that,

$$\begin{aligned} AsyCS^{(1)}_{GMS} &\geq \inf_{(\pi,h)\in\Pi H} \Pr(S_1(h_2^{1/2}Z + h_1, h_2) \leq c_{\pi_1^*}(h_2, 1-\alpha)) \\ &= \lim_{l\to\infty} \Pr(S_1(h_{2,l}^{1/2}Z + h_{1,l}, h_{2,l}) \leq c_{\pi_{1,l}^*}(h_{2,l}, 1-\alpha)). \\ &\geq \lim_{\omega_l\to\infty} \Pr(\tilde{S}_1(Z, h_{\omega_l}) \leq c_{\pi_{1,\omega_l}^*}(h_{2,\omega_l}, 1-\alpha)). \\ &\geq \liminf_{\omega_l\to\infty} \Pr(Z \in A^k_{B_{\omega_l}}) > 0. \end{aligned} \tag{C-43}$$

Now consider the second case. It follows that there exists a subsequence $\{\omega_l\}_{l\geq 1}$ of $\mathbb{N}$ such that $\lim_{l\to\infty} c_{\pi_{1,\omega_l}^*}(h_{2,\omega_l}, 1-\alpha) = 0$. 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 $\{0,\infty\}$. Since $\lim_{l\to\infty} c_{\bar{\pi}_1^*}(h_{2,\tilde{\omega}_l}, 1-\alpha) = 0$, Assumption A.2 implies $\bar{\pi}_1^* = \infty_p$. It follows immediately that $h_{1,\tilde{\omega}_l} = \infty_p$ and $S_1(h_{2,\tilde{\omega}_l}^{1/2}Z + h_{1,\tilde{\omega}_l}, h_{2,\tilde{\omega}_l}) = 0$ a.s. along the subsequence. This completes the proof.

Part 2. By Lemma B.2, the PA asymptotic confidence size of the test function S_2 is

$$AsyCS^{(2)}_{PA} = \inf_{h\in H} \Pr\left(S_2(h_2^{1/2}Z + h_1, h_2) \leq c_0(h_2, 1-\alpha)\right), \tag{C-44}$$

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. The function S_2 is defined in Equation (2.24).

Consider the following correlation matrix h_2

$$h_2 = \begin{bmatrix} A & B \\ B' & D \end{bmatrix}, \quad A = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad \rho = corr(Z_{h_2,1}, Z_{h_2,2}). \tag{C-45}$$

Since $h_2 \in \Psi_2$, then $\rho \in [-\sqrt{1-\varepsilon}, \sqrt{1-\varepsilon}]$ and $\det(A) = 1-\rho^2 \geq \varepsilon > 0$. Thus, $\det(h_2) = \det(A)\det(D - B'A^{-1}B)$. Furthermore, consider the case where $B = 0_{2\times(k-2)}$ and $D = I_{k-2}$, so that $\det(h_2) = \det(A) = 1-\rho^2$. The inverse of h_2 is

$$h_2^{-1} = \begin{bmatrix} A^{-1} & 0_{2\times(k-2)} \\ 0_{(k-2)\times 2} & I_{k-2} \end{bmatrix}, \quad \text{where} \quad A^{-1} = (1-\rho^2)^{-1}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}. \tag{C-46}$$

In what follows we let $h_2^\star$ be the matrix defined in Equation (C-45) with $B = 0_{2\times(k-2)}$ and $D = I_{k-2}$. By assumption A.7, there exists h_1 such that $(h_1, h_2^\star) \in H$. To simplify notation let $Z^\star \sim N(0_k, h_2^\star)$ so that

$$\begin{aligned} S_2(Z^\star + h_1, h_2^\star) = \inf_{t\in\mathbb{R}^p_{+,+\infty}} &\Big\{(1-\rho^2)^{-1}[(Z_1^\star + h_{1,1} - t_1)^2 + (Z_2^\star + h_{1,2} - t_2)^2 \\ &-2\rho(Z_1^\star + h_{1,1} - t_1)(Z_2^\star + h_{1,2} - t_2)] + \sum_{j=3}^{p}(Z_j^\star + h_{1,j} - t_j)^2\Big\} + \sum_{j=p+1}^{k}(Z_j^\star + h_{1,j})^2. \end{aligned} \tag{C-47}$$

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

$$\begin{aligned} S_2(Z^\star + h_1, h_2^\star) = \inf_{t\in\mathbb{R}^2_{+,+\infty}} &\big\{(1-\rho^2)^{-1}[(Z_1^\star + h_{1,1} - t_1)^2 + (Z_2^\star + h_{1,2} - t_2)^2 \\ &-2\rho(Z_1^\star + h_{1,1} - t_1)(Z_2^\star + h_{1,2} - t_2)]\big\} + \sum_{j=3}^{p}[Z_j^\star + h_{1,j}]_-^2 + \sum_{j=p+1}^{k}(Z_j^\star + h_{1,j})^2. \end{aligned} \tag{C-48}$$

The optimization problem in the RHS of Equation (C-48) is the same as the one in Lemma B.3 and, by that lemma, the solution can be divided in four cases depending on whether $Z_1^\star$ and $Z_2^\star$ are positive or negative. We focus on the set of $Z^\star$ that yield the solution in Equation (B-5) in Lemma B.3 and show that such solution holds with probability one. To do this let $r^* > 0$ be given. By assumption A.7 there are at least two $j = 1, \dots, k$ such that $h_{1,j} < 0$, so wlog assume $h_{1,1} < 0$ and $h_{1,2} < 0$. Now, for some $\beta > 0$, let

$$\bar{H}_\beta \equiv \{(h_1, h_2^\star) \in H : h_{1,1} \le -\beta,\ h_{1,2} \le -\beta,\ \rho \in [-\sqrt{1-\varepsilon}, -\beta]\}.$$

$\bar{H}_\beta$ does not restrict $h_{1,j}$ for $j = 3, \dots, k$. Next define the event $A_{h_1,\rho} = A^a_{h_1,\rho} \cup A^b_{h_1,\rho}$, where

$$A^a_{h_1,\rho} \equiv \{z \in \mathbb{R}^k : z_1 \ge 0, z_2 < 0, z_1 - \rho z_2 \le -h_{1,1} + \rho h_{1,2}\}, \tag{C-49}$$

$$A^b_{h_1,\rho} \equiv \{z \in \mathbb{R}^k : z_2 \ge 0, z_1 < 0, z_2 - \rho z_1 \le -h_{1,2} + \rho h_{1,1}\}. \tag{C-50}$$

Note that $z \in A_{h_1,\rho}$ does not restrict z_j for $j = 3, \dots, k$. To further simplify notation let $\Pr(A_{h_1,\rho}) \equiv \Pr(Z^\star \in A_{h_1,\rho})$ and $\underline{h}_2^\star$ denote the matrix $h_2^\star$ for the value $\rho = \underline{\rho} \equiv -\sqrt{1-\varepsilon}$. It follows from Lemma B.4 that $\forall \eta > 0$, $\exists \varepsilon > 0$ such that

$$\inf_{(h_1, h_2^\star = \underline{h}_2^\star) \in \bar{H}_\beta} \Pr(A_{h_1,\underline{\rho}}) \ge 1 - \eta. \tag{C-51}$$

For the next step define the function

$$\tilde{S}_2(Z^\star, \rho) = \inf_{t \in \mathbb{R}^2_{+,+\infty}} \left\{(1-\rho^2)^{-1}[(Z_1^\star - t_1)^2 + (Z_2^\star - t_2)^2 - 2\rho(Z_1^\star - t_1)(Z_2^\star - t_2)]\right\}, \tag{C-52}$$

and note that Lemma B.4 implies that there exists a function $\tau(z, h_1, h_2) > 0$ such that

$$S_2(z + h_1, h_2^\star) \ge \tilde{S}_2(z, \rho) + \frac{1}{1-\rho^2}\tau(z, h_1, h_2^\star), \text{ for all } z \in A_{h_1,\rho}. \tag{C-53}$$

We wish to show that $\forall \eta > 0$, $\exists \varepsilon > 0$ such that

$$\inf_{\rho \in [-\sqrt{1-\varepsilon}, -\beta]} (c_0(h_2, 1-\alpha) - \frac{1}{1-\rho^2}\tau(z, h_1, h_2^\star)) \le -\eta. \tag{C-54}$$

To this end, note that by Lemma B.3 it follows that with probability one

$$S_2(Z^\star, h_2^\star) = \sum_{j=3}^{p}[Z_j^\star]_-^2 + \sum_{j=p+1}^{k}(Z_j^\star)^2 + f(Z_1^\star, Z_2^\star, \rho) \le \sum_{j=3}^{p}[Z_j^\star]_-^2 + \sum_{j=p+1}^{k}(Z_j^\star)^2 + (Z_1^\star)^2 + W^2, \tag{C-55}$$

where $f(\cdot)$ is defined in Lemma B.3 (Equation (B-5)) and satisfies $f(Z_1^\star, Z_2^\star, \rho) \le (Z_1^\star)^2 + W^2$ with probability one for all $\rho \in [-\sqrt{1-\varepsilon}, -\beta]$ and $\varepsilon > 0$, and $Z_1^\star \perp W \sim N(0,1)$. As a result, the $1-\alpha$ quantile of $S_2(Z^\star, h_2^\star)$, $c_0(h_2, 1-\alpha)$, is bounded above by the $1-\alpha$ quantile of the RHS of Equation (C-55), denoted by $\tilde{c}_0(1-\alpha)$. Note that $\tilde{c}_0(1-\alpha)$ does not depend on ρ. It then follows that $c_0(h_2, 1-\alpha) \le \tilde{c}_0(1-\alpha) < \infty$ and Equation (C-54) follows immediately from

$$\inf_{\rho \in [-\sqrt{1-\varepsilon}, -\beta]} (\tilde{c}_0(1-\alpha) - \frac{1}{1-\rho^2}\tau(z, h_1, h_2^\star)) < 0, \tag{C-56}$$

for $\varepsilon > 0$ small enough, as $\inf_{(h_1, h_2^\star) \in \bar{H}_\beta} \tau(z, h_1, h_2^\star) > 0$. Finally, to complete the proof we note that for every $\eta > 0$, $\exists \varepsilon > 0$ such that

$$
\begin{aligned}
AsyCS_{PA}^{(2)} &\leq \inf_{h \in \bar{H}_\beta} \Pr\left(S_2(Z^\star + h_1, h_2^\star) \leq c_0(h_2^\star, 1-\alpha)\right), \\
&= 1 - \sup_{h \in \bar{H}_\beta} \Pr\left(S_2(Z^\star + h_1, h_2^\star) > c_0(h_2^\star, 1-\alpha)\right), \\
&\leq 1 - \sup_{h \in \bar{H}_\beta} \Pr\left(S_2(Z^\star + h_1, h_2^\star) > c_0(h_2^\star, 1-\alpha) | A_{h_1,\rho}\right) \Pr(A_{h_1,\rho}), \\
&\leq 1 - \sup_{h \in \bar{H}_\beta} \Pr\left(\tilde{S}_2(Z^\star, \rho) > c_0(h_2^\star, 1-\alpha) - \frac{1}{1-\rho^2}\tau(Z^\star, h_1, h_2^\star) | A_{h_1,\rho}\right) \Pr(A_{h_1,\rho}), \\
&\leq \eta, \qquad \text{(C-57)}
\end{aligned}
$$

where the first inequality follows from $\bar{H}_\beta \subseteq H$, the second inequality from $A_{h_1,\rho} \subseteq \mathbb{R}^k$, the third one from Equation (C-53) and the last one from Equations (C-51), (C-54) and $\tilde{S}_2(z, \rho) \geq 0$, $\forall z \in \mathbb{R}^k$. □

Proof of Corollary 3.1. By Theorem 3.2(1) there exists $B > 0$ such that for all $r^* \leq B$, $AsyCS_{GMS}^{(1)} > 0$. Pick $\eta = AsyCS_{GMS}^{(1)}/2 > 0$. By Theorem 3.2(2) there exists ε such that $AsyCS_{PA}^{(2)} \leq \eta = AsyCS_{GMS}^{(1)}/2$. Therefore, by Theorem 3.1

$$
AsyCS_{GMS}^{(2)} = AsyCS_{SS}^{(2)} \leq AsyCS_{PA}^{(2)} < AsyCS_{GMS}^{(1)} = AsyCS_{SS}^{(1)} \leq AsyCS_{PA}^{(1)}. \qquad \text{(C-58)}
$$

□

Appendix D Verification of the Assumptions in Examples

D.1 Example 2.1

We start by writing the example using the notation in Definition 2.1. For simplicity assume $Y_L(x_l)$ and $Y_H(x_l)$ are both finite for all $l = 1, \ldots, d_x$. Without loss of generality, assume that $Y_L(x_l) = 0$ and $Y_H(x_l) = 1$ and let P_n denote the probability with respect to the distribution F_n so that

$$
\begin{aligned}
\gamma_{1,l,1,n} &\equiv \sigma_{F_n,l,1}^{-1} E_{F_n} m_{l,1}(W_i, \theta_n) \\
&= \sigma_{F_n,l,1}^{-1} E_{F_n}[(YZ - G(x_l, \theta_n) + 1 - Z) I(X = x_l)] \geq -r_{l,1} n^{-1/2}, \\
\gamma_{1,l,2,n} &\equiv \sigma_{F_n,l,2}^{-1} E_{F_n} m_{l,2}(W_i, \theta_n) \\
&= \sigma_{F_n,l,2}^{-1} E_{F_n}[(G(x_l, \theta_n) - YZ) I(X = x_l)] \geq -r_{l,2} n^{-1/2}. \qquad \text{(D-1)}
\end{aligned}
$$

for $l = 1, \ldots, d_x$. This model satisfies the following relationship

$$
m_{l,1}(W_i, \theta_n) = (1 - Z) I(X = x_l) - m_{l,2}(W_i, \theta_n), \qquad \text{(D-2)}
$$

for $l = 1, \ldots, d_x$, so that

$$
\gamma_{1,l,1,n} = \sigma_{F_n,l,1}^{-1}(1 - \pi_{l,n}) p_{l,n} - \sigma_{F_n,l,1}^{-1} \sigma_{F_n,l,2} \gamma_{1,l,2,n}, \qquad \text{(D-3)}
$$

where $p_{l,n} = P_n(X = x_l)$ and $\pi_{l,n} = P_n(Z = 1 | X = x_l)$. Also, assume $\min_{l \leq d_x} p_{l,n} \geq c$ for all $n \geq 1$ and some $c > 0$ and that $\{Y | X = x_l\}$ is bounded and non-degenerate for all $l = 1, \ldots, d_x$.

The are two fixed sequences that serve as benchmark. If $\pi_{l,n} = 1$ (i.e., there is no missing data) then $\rho_{1,2,l,n} = Corr_{F_n}(m_{l,2}, m_{1,1}) = -1$, and $\sigma_{F_n,l,1}^2 = \sigma_{F_n,l,2}^2 = Var_{F_n}(Y | X = x_l) \in (0, \infty)$ for $l = 1, \ldots, d_x$. Also, if $\pi_{l,n} = 0$ (i.e., all data are missing) then $\rho_{1,2,l,n} = 0$ and $\sigma_{F_n,l,1}^2 = \sigma_{F_n,l,2}^2 = 0$. Any other value of $\pi_{l,n}$ results in $\rho_{1,2,l,n} \in (0, 1)$ and $\sigma_{F_n,l,s}^2 \in (0, \infty)$ as $\{Y | X = x_l\}$ is bounded and non-degenerate.

Let $n^{1/2}\gamma_{1,l,1,n} \to h_{1,l,1}$, $n^{1/2}\gamma_{1,l,2,n} \to h_{1,l,2}$, $\lambda_l = \lim n^{1/2} \sigma_{F_n,l,1}^{-1}(1 - \pi_{l,n}) p_{l,n} \geq 0$ and $\delta_l =$

$\lim \sigma^{-1}_{F_n,l,1}\sigma_{F_n,l,2} \in (0,\infty)$ for $l = 1,\ldots,d_x$. The last result follows by getting the expression of $\frac{\sigma^2_{F_n,l,2}}{\sigma^2_{F_n,l,1}}$ and doing some algebraic manipulations. By (D-3)

$$h_{1,l,1} + \delta_l h_{1,l,2} = \lambda_l \in \mathbb{R}_{+,\infty}. \tag{D-4}$$

Equation (D-4) shows that $h_{1,l,2} < 0$ implies $h_{1,l,1} > 0$ and that it is impossible to have both strictly negative. The values λ_l can take depend on the sequence $\{\pi_{l,n}\}_{n\geq 1}$. If $\pi_{l,n} \to 1$, λ_l can take any value on $[0,\infty]$ depending on the rate at which $\pi_{l,n}$ converges to 1. If $\pi_{l,n} \to \pi \in [0,1)$, then $\lambda_l = \infty$ as $\sigma_{F_n,l,1}$ is bounded.

D.1.1 On Assumption A.5

Since the analysis is the same for all $l = 1,\ldots,d_x$, we focus on one l at a time. Assume $h_{1,l,2} < 0$ and $h_{1,l,1} = \infty$.[6] In this case $g_{1,l,2} = \lim b_n^{1/2}\gamma_{1,l,2,n} = 0$ and

$$\begin{aligned} g_{1,l,1} &= \lim b_n^{1/2}\gamma_{1,l,1,n} = \lim\{b_n^{1/2}\sigma^{-1}_{F_n,l,1}[(1-\pi_{l,n})p_{l,n} - \sigma_{F_n,l,2}\gamma_{1,l,2,n}]\} \\ &= \lim b_n^{1/2}\sigma^{-1}_{F_n,l,1}(1-\pi_{l,n})p_{l,n} - \lim b_n^{1/2}\sigma^{-1}_{F_n,l,1}\sigma_{F_n,l,2}\gamma_{1,l,2,n} \\ &= \lim b_n^{1/2}\sigma^{-1}_{F_n,l,1}(1-\pi_{l,n})p_{l,n}. \end{aligned} \tag{D-5}$$

Suppose that $\lim b_n^{1/2}(1-\pi_{l,n}) = \infty$. Then $g_{1,l,1} = \infty$ and Assumption A.5 holds. Now suppose $\lim b_n^{1/2}(1-\pi_{l,n}) < \infty$. In such case we can find another sequence $\{\theta_n', F_n'\}_{n\geq 1}$ such that $\theta_n' = \theta_n$ and F_n' induces $1-\pi'_{l,n} = b_n^{-1/2}\log b_n = o(1)$. This new sequence is such that $g_{1,l,1} = \lim b_n^{1/2}(1-\pi'_{l,n}) = \infty$. Therefore, Assumption A.5 holds in this example.

D.1.2 On Assumption A.6

For simplicity consider the case $d_x = 1$ and the function S_1. Let $c_0(h_2^*, 1-\alpha)$ be the $1-\alpha$ quantile of

$$S_1(Z_{h_2^*}, h_2^*) = [Z_1]_-^2 + [-Z_1]_-^2 = Z_1^2, \quad Z_{h_2^*} = (Z_1, Z_2) \sim N(0, h_2^*), \tag{D-6}$$

where h_2^* has ones in the diagonal and $\rho_{2,1} = -1$ (i.e., $Z_2 = -Z_1$) off-diagonal. Choose a sequence of parameters with $\lambda^* = 0$ and $h_{1,1}^* = -h_{1,2}^* < 0$, following Equation (D-4). It follows that

$$S_1(Z_{h_2^*} + h_1^*, h_2^*) = [Z_1 + h_{1,1}^*]_-^2 + [-Z_1 - h_{1,1}^*]_-^2 = (Z_1 + h_{1,1}^*)^2, \tag{D-7}$$

and since $\Pr((Z_1 + h_{1,1}^*)^2 \leq x) < \Pr(Z_1^2 \leq x)$ for $h_{1,1}^* < 0$, $\Pr((Z_1 + h_{1,1}^*)^2 \leq c_0(h_2^*, 1-\alpha)) < 1-\alpha$. Assumption A.6 then holds. The general case where $d_x > 1$ follows by applying the previous argument to each pair of moment inequalities.

D.2 Example 2.2

We start again by writing the example using the notation in Definition 2.1. Let $\theta_n = (\theta_{1,n}, \theta_{2,n})$ and P_n denote the probability with respect to the distribution F_n. Define

$$\begin{aligned} \gamma_{1,1,n} &\equiv \sigma^{-1}_{F_n,1}E_{F_n}m_1(W_i,\theta_n) = \sigma^{-1}_{F_n,1}E_{F_n}[G_1(\theta_n) - W_{1,i}(1-W_{2,i})], \\ \gamma_{1,2,n} &\equiv \sigma^{-1}_{F_n,2}E_{F_n}m_2(W_i,\theta_n) = \sigma^{-1}_{F_n,2}E_{F_n}[W_{1,i}(1-W_{2,i}) - G_2(\theta_n)], \\ \gamma_{1,3,n} &\equiv \sigma^{-1}_{F_n,3}E_{F_n}m_3(W_i,\theta_n) = \sigma^{-1}_{F_n,3}E_{F_n}[W_{1,i}W_{2,i} - G_3(\theta_n)]. \end{aligned} \tag{D-8}$$

[6] The case where $h_{1,l,2} = \infty$ and $h_{1,l,1} < 0$ or $h_{1,l,1} = h_{1,l,2} = \infty$ is similar.

Define $p_{jk,n} \equiv P_n(W_1 = j, W_2 = k)$ so that

$$\begin{aligned}
\sigma^2_{F_n,2} = \sigma^2_{F_n,1} &\equiv Var_{F_n}[m_1(W_i,\theta_n)] = p_{10,n}(1-p_{10,n}) \in [0,1], \\
\sigma^2_{F_n,3} &\equiv Var_{F_n}[m_3(W_i,\theta_n)] = p_{11,n}(1-p_{11,n}) \in [0,1], \\
\rho_{12,n} &\equiv Corr_{F_n}[m_1, m_2] = -1, \\
\rho_{13,n} &\equiv Corr_{F_n}[m_1, m_3] = \frac{p_{11,n}p_{10,n}}{\sigma_{F_n,1}\sigma_{F_n,3}}, \\
\rho_{23,n} &\equiv Corr_{F_n}[m_2, m_3] = -\frac{p_{11,n}p_{10,n}}{\sigma_{F_n,1}\sigma_{F_n,3}}.
\end{aligned} \tag{D-9}$$

This model satisfies the following relationship

$$m_2(W,\theta_n) = \Delta G(\theta_n) - m_1(W,\theta_n), \tag{D-10}$$

where $\Delta G(\theta) = G_1(\theta) - G_2(\theta) = \Pr(u_{1,i} < \theta_1, u_{2,i} < \theta_2)$. This results in

$$\gamma_{1,2,n} = \sigma^{-1}_{F_n,2}\Delta G(\theta_n) - \gamma_{1,1,n}. \tag{D-11}$$

The parameter space for the misspecified model imposes for $r = (r_1, r_2, r_3) \in \mathbb{R}^3_+$

$$n^{1/2}\gamma_{1,1,n} \geq -r_1, \quad n^{1/2}\gamma_{1,2,n} \geq -r_2, \quad |n^{1/2}\gamma_{1,3,n}| \leq r_3. \tag{D-12}$$

Let $n^{1/2}\gamma_{1,j,n} \to h_{1,j}$ for $j = 1,2,3$ and $\lambda = \lim n^{1/2}\sigma^{-1}_{F_n,2}\Delta G(\theta_n) \geq 0$. By (D-11)

$$h_{1,1} + h_{1,2} = \lambda \in \mathbb{R}_{+,\infty}. \tag{D-13}$$

Equation (D-13) shows that $h_{1,1} < 0$ implies $h_{1,2} > 0$, and $h_{1,2} < 0$ implies $h_{1,1} > 0$. Also note that $\sigma_{F_n,2} \to 0$ iff $p_{10,n} \to 0$ or $p_{10,n} \to 1$. In addition, given that $u_i = (u_{1,i}, u_{2,i})$ is supported on $[0,1]^2$ under the distribution G

1. $G_1(\theta) = 0$ iff $\theta_2 = 0$ and $G_1(\theta) = 1$ iff $\theta_2 = 1$.
2. $G_2(\theta) = 0$ iff $\theta_1 = 1$ or $\theta_2 = 0$, and $G_2(\theta) = 1$ iff $(\theta_1, \theta_2) = (0, 1)$.
3. $\Delta G(\theta) = 0$ iff $\theta_1 = 0$ or $\theta_2 = 0$, and $\Delta G(\theta) = 1$ iff $\theta = (1, 1)$.

We assume the distribution G assumed by the econometrician is continuous on $\mathbb{R}^2$ and that for a $r = (r_1, r_2, r_3)' \in \mathbb{R}^3_+$

$$\sup_{\theta\in\Theta} |G_j(\theta) - G_{nj}(\theta)| \leq r_j n^{-1/2}, \quad j = 1,2,3, \tag{D-14}$$

where G_n denotes the true distribution of u_i for sample size n and $G_{nj}(\theta_0)$ is defined as $G_j(\theta_0)$ above when $u_i \sim G_n$ rather than $u_i \sim G_n$. Thus, F_n is the distribution of W_i compatible with the joint distribution G_n. Since $G_{n2}(\theta_n) \leq p_{10,n} \leq G_{n1}(\theta_n)$, it follows from (D-14) and the continuity of G that $p_{10,n} \in (0,1)$ for n large enough whenever $\theta_{k,n} \to \theta_k \in (0,1)$ for all $k = 1,2$. Also, $p_{10,n} \to 0$ iff $\theta_{2,n} \to 0$ and $p_{10,n} \to 1$ iff $(\theta_{1,n}, \theta_{2,n}) \to (0,1)$.

D.2.1 On Assumption A.5

Consider a sequence $\{\theta_n, F_n\}_{n\geq 1}$ with limits $\lambda = \infty$, $h_{1,3} \in [-r_3, r_3]$ and $h_{1,1} < 0$.[7] Equation (D-13) implies that $h_{1,2} = \infty$. In this case $g_{1,1} = \lim b_n^{1/2}\gamma_{1,1,n} = 0$, and

$$g_{1,2} = \lim b_n^{1/2}\gamma_{1,2,n} = \lim b_n^{1/2}\sigma^{-1}_{F_n,2}\Delta G(\theta_n) \in [0, \infty]. \tag{D-15}$$

[7] The case in which $h_{1,2} < 0$ is symmetric.

If $\theta_{k,n} \to \theta_k \in (0,1)$ for $k = 1,2$ then $g_{1,2} = \infty$ and Assumption A.5 holds. Suppose instead that the sequence $\{\theta_n, F_n\}_{n\geq 1}$ is such that $\theta_{k,n} \to \theta_k$, where $\theta_k = 0$ for at least one $k = 1,2$, and that

$$\sigma_{F_n,2}^{-1} \Delta G(\theta_n) = O(b_n^{-1/2}). \tag{D-16}$$

In such case, we can find another sequence $\{\theta'_n, F'_n\}_{n\geq 1}$ such that $\theta'_{k,n} \to \theta_k$ at a much slower rate and therefore implying $g_{1,2} = \infty$. Choosing F'_n appropriately guarantees that γ'_n converges to the same limit h. Thus, Assumption A.5 holds in this case.

Finally consider a sequence $\{\theta_n, F_n\}$ with $\lambda < \infty$, $h_{1,3} = r_3$ and $h_{1,1} < 0$. Since $\lambda < \infty$ it follows that $h_{1,2} = \lambda - h_{1,1} \in (0, \infty)$ and Assumption A.5 holds.

D.2.2 On Assumption A.6

Consider the function S_1 and let $c_0(h_2^*, 1-\alpha)$ denote the $1-\alpha$ quantile of

$$S_1(Z_{h_2^*}, h_2^*) = [Z_1]_-^2 + [-Z_1]_-^2 + Z_3^2 = Z_1^2 + Z_3^2, \quad Z_{h_2^*} = (Z_1, Z_2, Z_3) \sim N(0, h_2^*), \tag{D-17}$$

where h_2^* is a correlation matrix with $\rho_{12} = -1$. Choose a sequence of parameters with $\lambda^* = 0$, $h_{1,1}^* = h_{1,2}^* < 0$ and $h_{1,3}^* = 0$. It follows that

$$S_1(Z_{h_2^*} + h_1^*, h_2^*) = [Z_1 + h_{1,1}^*]_-^2 + [-Z_1 - h_{1,1}^*]_-^2 + Z_3^2 = (Z_1 + h_{1,1}^*)^2 + Z_3^2 \tag{D-18}$$

and since $\Pr((Z_1 + h_{1,1}^*)^2 + Z_3^2 \leq x) < \Pr(Z_1^2 + Z_3^2 \leq x)$ for $h_{1,1}^* < 0$,

$$\Pr((Z_1 + h_{1,1}^*)^2 + Z_3^2 \leq c_0(h_2^*, 1-\alpha)) < 1 - \alpha. \tag{D-19}$$

Then, Assumption A.6 holds.

References

Andrews, D. W. K., and P. Guggenberger (2009a): "Hybrid and size-corrected subsample methods," *Econometrica*, 77(3), 721–762.

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

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

——— (2010b): "Applications of Hybrid and Size-Corrected Subsampling Methods," *Journal of Econometrics,* forthcoming.

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(1), 119–158.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

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

Ciliberto, F., and E. Tamer (2010): "Market Structure and Multiple Equilibria in Airline Industry," *Econometrica*, 77(6), 1791–1828.

Grieco, P. (2009): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," manuscript, Northwestern University.

Guggenberger, P. (2009): "On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Assumption," Manuscript, UCSD.

Hall, A. R., and A. Inoue (2003): "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models," *Journal of Econometrics*, 114(2), 361–394.

Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 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.

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

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 (2009): "Misspecification in Moment Inequality Models: Back to Moment Equalities?," manuscript, Northwestern University.

Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

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

Tamer, E. (2009): "Partial Identification in Econometrics," manuscript, Northwestern University.