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

Federico A. Bugni †
Department of Economics
Duke University

Ivan A. Canay ‡
Department of Economics
Northwestern University

Patrik Guggenberger §
Department of Economics
University of California, San Diego

October 19, 2010

Abstract

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

---

*This paper was previously circulated under the title "Asymptotic Distortions in Locally Misspecified Moment Inequality Models". We would like to thank seminar participants at Caltech, Harvard/MIT, Penn State University, UC Riverside, UCLA, University of Pennsylvania, Xiamen University, the 2010 Econometric Society World Congress, the Cemmap/Cowles "Advancing Applied Microeconometrics" conference, and the Econometrics Jamboree at Duke for helpful comments.

†Email: federico.bugni@duke.edu.

‡Email: iacanay@northwestern.edu.

§E-mail: pguggenberger@ucsd.edu. Guggenberger would like to thank the Alfred P. Sloan Foundation for a 2009-2011 fellowship and the NSF for research support through grant SES-1021101.

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), Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009b)(AG from now on), Fan and Park (2009), Stoye (2009), Andrews and Soares (2010)(AS from now on), Bugni (2010), and Canay (2010)).[1] As a consequence, there are currently several different testing procedures and methods to construct confidence sets (CSs) based on test inversion that have been compared in terms of asymptotic confidence size and asymptotic power 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.[2] We show that this intuition holds for CSs based on plug-in asymptotic (PA) critical values compared to subsampling and generalized moment selection (GMS, see AS), as well as CSs based on the modified methods of moments (MMM) test statistic compared to the quasi likelihood ratio (QLR) test statistic. However, the main contribution of this paper are two results that go beyond this 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 can lead to CSs with smaller expected volume under correct model specification (see AS). Second, we show that under certain conditions the CSs based on the QLR test statistic have asymptotic confidence size that can be an arbitrary fraction of the asymptotic confidence size of CSs obtained by using the MMM test statistic.

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 infer-

[1] There is a related literature about partially identified models that focuses on inference on the identified set rather than the identifiable parameters. This includes Pakes, Porter, Ho, and Ishii (2005), Beresteanu and Molinari (2008), Bontemps, Magnac, and Maurin (2008), Galichon and Henry (2009a,b), Moon and Schorfheide (2009), and Romano and Shaikh (2010) among others.

[2] In the context of hypothesis tests, local power is the limit of the rejection probability under a sequence of parameters that belong to the alternative hypothesis and approach the null hypothesis. Tests with high local power reject these sequences relatively often. In the context of local misspecification, some of the local sequences are part of the parameter space that determine the asymptotic size. Consequently, test with high local power might result in relatively high asymptotic size distortion. However, it is worth noting that the analysis of robustness conducted in this paper is relatively more complex than the study of local power, as here an essential part of the analysis is to consider all possible sequences of local parameters and search for the one that leads to the highest limiting rejection probability.

ence (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). However, the researcher has to make a stand on a second group of less fundamental assumptions (typically related to parametric functional forms). For example, in a standard simultaneous entry game where firms have profit functions given by $\pi_l = (u_l - \theta_l W_{-l})I(W_l = 1)$, where W_l denotes the entry decision of firm l, W_{-l} denotes the entry decision of the other firm, θ_l is the parameter of interest, $I(\cdot)$ is the indicator function, and u_l the monopoly profits of firm l; 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_l are part of the maintained assumptions. One justification for this asymmetry in the way assumptions are treated lies behind the idea that there are certain assumptions that directly restrict 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,[3] 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 now.

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.[4] 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 (2010) discuss the impact of global misspecification on the set of identifiable parameters.

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, the analysis under local misspecification provides guidance in situations where the true model is just a small perturbation away from the model proposed by the researcher. Newey (1985) applies this type of analysis in the context of over-identified moment equality models. More recently Guggenberger (2009) studies the size properties of hypothesis tests in the linear IV model under local violations of instrument exogeneity conditions, while Kitamura, Otsu, and Evdokimov (2009) consider local deviations within

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

[4] For example, in the case of maximum likelihood estimation the pseudo-true value minimizes the Kullback-Leibler discrepancy between the true model and the incorrect parametric model.

shrinking topological neighborhoods of point identified moment equality models and propose an estimator that achieves optimal minimax robust properties. Since the limit of locally misspecified models equals the correctly specified model, the parameter of interest under local misspecification and correct specification coincides. This facilitates the interpretation relative to pseudo-true values in globally misspecified models. Therefore, if the probability law generating the observations is a small perturbation of the true law, then it is of interest to seek for an inference procedure whose size is robust against such perturbations. This is the motivation behind the approach we propose in this paper.

The paper is organized as follows. Section 2 introduces the model, the notation, and provides two examples that illustrate the nature of misspecification that we capture with our framework. 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 and the Appendix includes the assumptions, proofs of the results, and verification of some of the assumptions for the two examples.

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

2 Locally Misspecified Moment Inequality/Equality 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 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 $(1-\alpha)$ level CS for θ_0 is

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

A CS is determined by the choice of test statistic and critical value. Several CSs have been suggested in the literature whose asymptotic confidence size is at least 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.[5]

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 are as powerful as subsampling tests for all sequences of local alternatives and strictly more powerful along certain sequences of local alternatives. In turn, subsampling tests are as powerful as PA tests for all sequences of local alternatives and strictly more powerful along some sequences of local alternatives. Andrews and Jia (2008) compare different combinations of tests statistics and critical values and provide a recommended test based on the 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). In this paper we are interested in ranking the resulting CSs in terms of asymptotic confidence size distortion when the moment (in)equalities in Equation (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 Equation (2.4), the true θ_0 satisfies the following inequalities for $l=1,\dots,d_x$,

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

Now suppose that in fact the data come from a local perturbation F_n of the hypothesized model F_0 such that

$$E_{F_n}\left(Y|X=x_l\right)=G_n\left(x_l,\theta_0\right),\quad \forall l=1,\dots,d_x, \tag{2.6}$$

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

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

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

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

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

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

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

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

when entering market $i \in \{1, \ldots, n\}$. Here $W_{l,i} = 1$ or 0 denotes "entering" or "not entering" market i by firm l, respectively, the subscript $-l$ denotes the decision of the other firm, the continuous random variable $u_{l,i}$ denotes the monopoly profits of firm l in market i, and $\theta_l \in [0, 1]$ is the profit reduction incurred by firm l if $W_{-l,i} = 1$. If a firm does not enter a market, it gets zero profits in that market. Therefore, entering is always profitable for at least one firm.

Define $W_i = (W_{1,i}, W_{2,i})$ and $\theta_0 = (\theta_1, \theta_2)$. There are four possible outcomes: (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. Therefore, under the assumption $u \sim G$, for some joint distribution G, the model implies that

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

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

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

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

To do inference on θ_0, the researcher assumes G is the joint distribution of the unobserved random vector u_i.[6] Now suppose that the data comes from a local perturbation of the hypothesized model. More specifically, suppose for example that for some $r = (r_1, r_2, r_3)' \in \mathbb{R}^3_+$

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

where G_n denotes the true distribution of u_i for sample size n and $G_{n,j}(\theta_0)$ is defined as $G_j(\theta_0)$ above when $u_i \sim G_n$ rather than $u_i \sim G$. Denote by F_n the true distribution of W_i that 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.[7] ■

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, as demonstrated, if the researcher assumes an incorrect parametric structure, the moment (in)equalities are potentially violated for every given sample size n at the true θ_0. The assumption of correct specification by the researcher of the distribution of u_i is very strong - it is therefore of critical importance to assess how robust in terms of size distortion competing inference procedures are when the assumption fails.

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

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

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

(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) $\{W_i\}_{i=1}^n$ 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) $Corr_F(m(W_i, \theta)) \in \Psi$, and,

(vii) $E_F|m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M, \quad j = 1, \ldots, k,$ (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 (in)equality violation r. Conditions (ii)-(iii) are modifications of (3.3) in AG (or (2.2) in AS) to account for local model misspecification. Finally, we use

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

to measure the amount of misspecification. Notice that the definition of the parameter space captures the framework in Examples 2.1 and 2.2.

Remark 2.2. 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 (in)equality is strictly locally violated. For example, adding the restriction

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

would be one way of doing this.

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 inferior of the magnitude one aims to control in finite samples, i.e., the exact confidence size of the CS. The existing literature on inference in partially identified moment (in)equality models shows that several inference procedures achieve $AsyCS \geq 1-\alpha$ when $r^* = 0$. In this paper we are interested in comparing these inference procedures when there is local misspecification (i.e, $r^* > 0$). In particular, we are interested in ranking the procedures according to their level of confidence size distortion, defined as

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

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

Remark 2.3. We could alternatively focus on the asymptotic size distortion of the tests for the null $H_0 : \theta_0 = \theta$. The asymptotic size in that case would be

$$AsySz(\theta) = \limsup_{n\to\infty} \sup_{F:(\theta,F)\in\mathcal{F}_n} \Pr_{\theta,F}(T_n(\theta) > c_n(\theta, 1-\alpha)), \tag{2.19}$$

where the supremum is only with respect to F and θ is fixed. Analytically, studying $AsySz(\theta)$ is less complex than studying $AsyCS$ as in the former case θ is fixed at a particular value while in the latter case θ may depend on n.

2.1 Test Statistics and Critical Values

We now present several test statistics $T_n(\theta)$ and corresponding critical values $c_n(\theta, 1-\alpha)$ to test (2.2) or, equivalently, to construct a CS as in (2.3). Define the sample moment functions $\bar{m}_n(\theta) = (\bar{m}_{n,1}(\theta), \dots, \bar{m}_{n,k}(\theta))$, where

$$\bar{m}_{n,j}(\theta) = n^{-1}\sum_{i=1}^{n} m_j(W_i, \theta) \text{ for } j = 1, \dots, k. \tag{2.20}$$

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

$$\hat{\Sigma}_n(\theta) = n^{-1}\sum_{i=1}^{n}(m(W_i,\theta) - \bar{m}_n(\theta))(m(W_i,\theta) - \bar{m}_n(\theta))'. \tag{2.21}$$

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

$$T_n(\theta) = S(n^{1/2}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)), \tag{2.22}$$

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

We now describe two popular choices of test functions. The first test function S is the Modified Method of Moments (MMM) given by

$$S_1(m, \Sigma) = \sum_{j=1}^{p}[m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{k}(m_j/\sigma_j)^2, \tag{2.23}$$

where $[x]_- = xI(x<0)$, $m = (m_1, \dots, m_k)$, and $\sigma_j^2 = \Sigma_{[j,j]}$. For this function, the parameter space Ψ for the correlation matrices in condition (v) of Equation (2.14) is not restricted. That is, the space in (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

$$T_{1,n}(\theta) = n \sum_{j=1}^{p} [\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta)]_-^2 + n \sum_{j=p+1}^{k} (\bar{m}_{n,j}(\theta)/\hat{\sigma}_{n,j}(\theta))^2, \tag{2.24}$$

where $\hat{\sigma}_{n,j}^2(\theta) = \hat{\Sigma}_n(\theta)_{[j,j]}$.

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

$$S_2(m, \Sigma) = \inf_{t=(t_1,0_v):t_1 \in \mathbb{R}^p_{+,+\infty}} (m-t)'\Sigma^{-1}(m-t). \tag{2.25}$$

This function requires Σ to be non-singular so we take $\Psi = \Psi_{2,\varepsilon}$, where

$$\Psi_{2,\varepsilon} = \{\Sigma \in \Psi_1 : \det(\Sigma) \geq \varepsilon\}, \tag{2.26}$$

for some $\varepsilon > 0$. The function S_2 leads to the test statistic

$$T_{2,n}(\theta) = \inf_{t=(t_1,0_v):t_1 \in \mathbb{R}^p_{+,+\infty}} (n^{1/2}\bar{m}_n(\theta) - t)'\hat{\Sigma}_n(\theta)^{-1}(n^{1/2}\bar{m}_n(\theta) - t). \tag{2.27}$$

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.[8]

We next describe three main choices of critical values. Assuming the limiting correlation matrix of $m(W_i, \theta)$ is given by Ω and that $r^* = 0$ in Equation (2.14), it follows from Lemma B.1 that

$$T_n(\theta) \to_d S(\Omega^{1/2}Z + h_1, \Omega), \tag{2.28}$$

where $Z \sim N(0_k, I_k)$, h_1 is a k-vector with $h_{1,j} = 0$ for $j > p$ and $h_{1,j} \in [0, \infty]$ for $j \leq p$ (see Lemma B.1), and $\Omega^{1/2}$ denotes a lower triangular matrix such that $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$. Therefore, ideally one would 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 (see AS and AG), and so some approximation to $c_{h_1}(\Omega, 1-\alpha)$ is necessary.

Under the assumptions in the Appendix, the asymptotic distribution in Equation (2.28) 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 $S(\Omega^{1/2}Z, \Omega)$ (i.e., $h_1 = 0_k$).[9] PA critical values are based on this "worst case" and are defined as consistent estimators of $c_0(\Omega, 1-\alpha)$.

[8]Note $S_1(m, \Sigma)$ is increasing in $|m_j|$ for $j = p+1, \ldots, k$, while $S_2(m, \Sigma)$ is not. To see this take $p = 0$, $k = 2$, and Σ with ones in the diagonal and $\sigma_{12} = 1/2$ off-diagonal. Then $S_2(m, \Sigma) = (4/3)(m_1^2 + m_2^2 - m_1 m_2)$ which is not necessarily increasing in $|m_1|$ or $|m_2|$.

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

Define

$$\hat{\Omega}_n(\theta) = \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta), \tag{2.29}$$

where $\hat{D}_n(\theta) = Diag(\hat{\Sigma}_n(\theta))$ and $\hat{\Sigma}_n(\theta)$ is defined in Equation (2.21). The PA test rejects H_0 if $T_n(\theta) > c_0(\hat{\Omega}_n(\theta), 1-\alpha)$, where the PA critical value is

$$c_0(\hat{\Omega}_n(\theta), 1-\alpha) \equiv \inf\{x \in \mathbb{R} : \Pr(S(\hat{\Omega}_n(\theta)^{1/2}Z, \hat{\Omega}_n(\theta)) \leq x) \geq 1-\alpha\}, \tag{2.30}$$

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

$$\xi_n(\theta) = \kappa_n^{-1}\hat{D}_n^{-1/2}(\theta)n^{1/2}\bar{m}_n(\theta), \tag{2.31}$$

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

$$\varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \tag{2.32}$$

where $\varphi = (\varphi_1, \ldots, \varphi_p, 0_v) \in \mathbb{R}^k_{[+\infty]}$ is a function chosen by the researcher that is assumed to satisfy Assumption A.4 in the Appendix. Examples include $\varphi_j^{(1)}(\xi, \Omega) = \infty I(\xi_j > 1)$, where we use the convention $\infty 0 = 0$, $\varphi_j^{(2)}(\xi, \Omega) = \psi(\xi_j)$, $\varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+$, 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 $0 < a_L \leq a_U \leq \infty$. 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

$$\hat{c}_{n,\kappa_n}(\theta, 1-\alpha) \equiv \inf\{x \in \mathbb{R} : \Pr(S(\hat{\Omega}_n^{1/2}(\theta)Z + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta)) \leq x) \geq 1-\alpha\}, \tag{2.33}$$

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_n denote the subsample size when the sample size is n. Throughout the paper we assume $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$. The number of different subsamples of size b_n is q_n (with i.i.d. observations, $q_n = n!/((n-b_n)!b_n!)$). The subsample statistics used to construct the subsampling critical value are $\{T_{n,b,s}(\theta)\}_{s=1}^{q_n}$, 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_n rather than the full sample. The empirical distribution function of $\{T_{n,b,s}(\theta)\}_{s=1}^{q_n}$ is

$$U_{n,b}(\theta, x) = q_n^{-1} \sum_{s=1}^{q_n} I(T_{n,b,s}(\theta) \leq x) \text{ for } x \in \mathbb{R}. \tag{2.34}$$

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

$$\hat{c}_{n,b}(\theta, 1-\alpha) \equiv \inf\{x \in \mathbb{R} : U_{n,b}(\theta, x) \geq 1-\alpha\}. \tag{2.35}$$

Having introduced the different test statistics and critical values typically used in the literature, we devote the next section to the analysis of the asymptotic properties of the different CSs under the locally misspecified models introduced 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 the resulting CSs based on PA, GMS, and subsampling critical values perform under local misspecification. In this case we write $AsyCS_{PA}$, $AsyCS_{GMS}$, and $AsyCS_{SS}$ for PA, GMS, and subsampling CSs, respectively, to make explicit the choice of critical value. Second, we take the critical value as given and compare how CSs based on the test functions S_1 and S_2 perform under local misspecification. In this case we write $AsyCS_l^{(1)}$ and $AsyCS_l^{(2)}$, for $l \in \{PA, GMS, SS\}$, to denote the asymptotic confidence size of the CSs based on test functions S_1 and S_2, respectively.

3.1 Comparison across Critical Values

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

Theorem 3.1. *Assume the parameter space is given by $\mathcal{F}_n$ in (2.14), $0 < \alpha < 1/2$, and that S satisfies Assumptions A.1-A.3. For GMS CSs assume that $\varphi(\xi, \Omega)$ satisfies Assumption A.4, and that $\kappa_n \to \infty$ and $\kappa_n^{-1} n^{1/2} \to \infty$. For subsampling CSs suppose $b_n \to \infty$ and $b_n/n \to 0$.*

1. *It follows that*

$$AsyCS_{PA} \geq AsyCS_{SS} \text{ and } AsyCS_{PA} \geq AsyCS_{GMS}. \tag{3.1}$$

Therefore, PA CSs are at least as robust as GMS and subsampling CSs under local violations of the moment (in)equalities.

2. *Suppose that Assumption A.6 holds. Then*

$$AsyCS_{PA} < 1 - \alpha. \tag{3.2}$$

By Equation (3.1) *it follows that* $AsyCS_{SS} < 1 - \alpha$ *and* $AsyCS_{GMS} < 1 - \alpha$.

3. *Suppose that Assumption A.5 holds and* $\kappa_n^{-1} n^{1/2} / b_n^{1/2} \to \infty$. *Then*

$$AsyCS_{SS} = AsyCS_{GMS}. \tag{3.3}$$

Therefore, subsampling CSs and GMS CSs are equally robust under local violations of the moment (in)equalities.

Assumptions A.1-A.4 are slight modifications of the corresponding assumptions in AS and AG. Assumptions A.5 and A.6 are introduced in this paper and ensure that the parameter space is rich enough. These two new assumptions are mild and we verify them for the two lead examples in Section D of the Appendix under the mild primitive conditions of Assumptions D.1 and D.2. Under a reasonable set of assumptions the theorem implies

$$AsyCS_{GMS} = AsyCS_{SS} \leq AsyCS_{PA} < 1 - \alpha. \tag{3.4}$$

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, CSs under-cover the true parameter. Second, the equation reveals that PA CSs suffer the least amount of asymptotic confidence size distortion. This is expected as this CS uses a conservative critical value, 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 suggesting that the expected volume of the corresponding GMS CSs are no larger than that of subsampling CSs. 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 result 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 CSs differ in their asymptotic behavior along certain sequences of locally misspecified models, these sequences turn out not to be the relevant ones for the computation of the asymptotic confidence sizes, i.e., the ones that attain the infimum in Equation (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 (and their implication for the expected volume of the corresponding CSs), our results indicate that GMS CSs are preferable to subsampling CSs: there can be a reduction in expected volume under correct specification without worsening the asymptotic confidence size distortion when the model is locally misspecified.

According to Equation (3.4), PA CSs are the most robust CSs among the procedures 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 in terms of expected volume 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 distortion of CSs based on the test functions S_1 and S_2 defined in Equations (2.23) and (2.25), respectively. The main result of this section has two parts. First, we show that the $AsyCS$ of CSs based on 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. *Assume the parameter space is given by $\mathcal{F}_n$ in (2.14) and $0 < \alpha < 1/2$. For GMS CSs assume that $\varphi(\xi, \Omega)$ satisfies Assumption A.4 and that $\kappa_n \to \infty$ and $\kappa_n^{-1} n^{1/2} \to \infty$. For subsampling CSs suppose $b_n \to \infty$ and $b_n/n \to 0$.*

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

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

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 when ε in $\Psi_{2,\varepsilon}$ is small enough. This is, the asymptotic confidence size of CSs based on the test function S_2 is severely affected by the smallest amount of misspecification while CSs based on the test function S_1 have positive asymptotic confidence size. Combining these two results we derive the following corollary.

Corollary 3.1. *Suppose that all the assumptions in Theorems 3.1 and 3.2 hold. Then, there exists $B > 0$ and $\varepsilon > 0$ in $\Psi_{2,\varepsilon}$ such that whenever $r^* \in (0, B]$,*

$$AsyCS^{(2)}_l < AsyCS^{(1)}_l < 1 - \alpha, \quad l \in \{PA, GMS, SS\}. \tag{3.7}$$

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 that Equation (3.7) should hold. However, Theorem 3.2 quantifies this relationship by showing that the cost of having a smaller expected volume under correct specification for CSs based on S_2 is an arbitrarily low asymptotic confidence size under local misspecification. The quantitative differences between CSs based on S_1 and S_2 can be dramatic, as we illustrate below in Table 1.

Remark 3.1. The generalized empirical likelihood (GEL) test statistics are asymptotically equivalent to $T_{2,n}(\theta)$ up to first order (see AG and Canay (2010)), and so the asymptotic confidence size distortion of CSs based on GEL is equal to $AsyCS^{(2)}_{GMS}$ in Theorem 3.2.

To understand the intuition behind Theorem 3.2 it is enough to consider the case with two moment inequalities, $p = k = 2$, 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

$$AsyCS^{(1)}_{PA} \leq \Pr([Z^\star_1 - r_1]^2_- + [-Z^\star_1]^2_- \leq c_0(\Omega, 1-\alpha)), \tag{3.8}$$

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 $AsyCS^{(1)}_{PA}$ 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 $A \equiv [Z^\star_1 - r_1]^2_-$ and $B \equiv [-Z^\star_1]^2_-$ so that the right hand side of Equation (3.8) is $\Pr(A + B \leq c_0(\Omega, 1-\alpha))$. On the one hand, if $Z^\star_1 \notin [0, r_1]$ 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 $Z^\star_1 \in [0, r_1]$, $A + B = (Z^\star_1 - r_1)^2 + Z^{\star 2}_1 \geq r_1^2/2 > c_0(\Omega, 1-\alpha)$. We can then conclude that

$$\Pr([Z^\star_1 - r_1]^2_- + [-Z^\star_1]^2_- \leq c_0(\Omega, 1-\alpha)) = 0, \tag{3.9}$$

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

In addition, Theorem 3.2 shows that $AsyCS^{(2)}_{PA}$ can be arbitrarily close to zero when ε in the space $\Psi_{2,\varepsilon}$ is small. What drives this result is the possibility that at least two inequalities are violated (or one is violated and the other one is binding) and strongly negatively correlated. To illustrate this, consider again 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

$$AsyCS^{(2)}_{PA} \leq \Pr(S_2(Z^\star, r_1, \Omega_\varepsilon) \leq c_0(\Omega_\varepsilon, 1-\alpha)), \tag{3.10}$$

where $Z^\star \sim N(0, \Omega_\varepsilon)$, Ω_ε is a matrix with $\rho = -(1-\varepsilon)^{1/2}$, and

$$S_2(Z^\star, r_1, \Omega_\varepsilon) = \frac{1}{\varepsilon} \inf_{t \in \mathbb{R}^2_{+,+\infty}} \{\sum_{j=1}^{2} (Z_j^\star - r_1 - t_j)^2 + 2(1-\varepsilon)^{1/2}(Z_1^\star - r_1 - t_1)(Z_2^\star - r_1 - t_2)\}. \quad (3.11)$$

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 $(Z_1^\star, Z_2^\star)$. However, there exists a set $A \subset \mathbb{R}^2$ such that for all $(z_1, z_2) \in A$

$$S_2(z, r_1, \Omega_\varepsilon) \geq S_2(z, 0, \Omega_\varepsilon) + \frac{2}{\varepsilon}[r_1^2 - z_1 - z_2], \quad (3.12)$$

with $[r_1^2 - z_1 - z_2] > 0$, and $\Pr(Z^\star \in A) \to 1$ as $\varepsilon \to 0$. It is immediate from Equation (3.12) that small distortions $r_1 > 0$ can produce a value of $S_2(Z^\star, r_1, \Omega_\varepsilon)$ that is arbitrarily high on the set A by allowing ε to be arbitrarily close to 0 (i.e., correlation close to -1). Since $c_0(\Omega_\varepsilon, 1-\alpha)$ can be shown to be bounded in $\Psi_{2,\varepsilon}$, it follows that

$$\Pr(S_2(Z^\star, r_1, \Omega_\varepsilon) \leq c_0(\Omega_\varepsilon, 1-\alpha) | A) \to 0, \quad (3.13)$$

as $\varepsilon \to 0$. Therefore, Equation (3.10) implies that CSs based on S_2 have asymptotic confidence size arbitrarily close to zero when ε is small.

		$AsyCS_{PA}^{(1)}$	$AsyCS_{PA}^{(2)}$	
p	r^*		$\varepsilon = 0.10$	$\varepsilon = 0.05$
	0.25	0.888	0.637	0.351
2	0.50	0.800	0.101	0.003
	1.00	0.502	0.000	0.000
	0.25	0.866	0.588	0.314
4	0.50	0.739	0.071	0.001
	1.00	0.256	0.000	0.000
	0.25	0.847	0.631	0.347
6	0.50	0.674	0.091	0.002
	1.00	0.153	0.000	0.000
	0.25	0.830	0.713	0.441
8	0.50	0.617	0.134	0.009
	1.00	0.082	0.000	0.000
	0.25	0.804	0.720	0.461
10	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. The numbers above were computed using the explicit formula for $AsyCS$ provided in Equation B-2 and the infimum with respect to Ω for S_1 and S_2 was carried out by minimizing over 15000 random correlation matrices in Ψ_1 and $\Psi_{2,\varepsilon}$, respectively.

Theorem 3.2 presents a theoretical result regarding the relative amount of confidence size distortion of different test functions. We now quantify these results by numerically computing the asymptotic confidence size of the CSs based on S_1 and S_2 using the formulas provided in Lemma B.2. Table 1 reports the cases where $p \in \{2, 4, 6, 8, 10\}$, $k = p$, $\varepsilon \in \{0.10, 0.05\}$, and $r^* \in \{0.25, 0.50, 1.00\}$. Table 1 shows that S_2 is significantly distorted even for relatively high values of ε (i.e., $\varepsilon = 0.1$). For example, when $p = 2$ and $r^* = 0.5$, the asymptotic confidence size of the test function S_1 is 0.80 while the asymptotic confidence size of S_2 is 0.10 at best. The asymptotic confidence size for both test functions decreases as p grows,[10] and as predicted by Theorem 3.2, the asymptotic confidence size of S_2 is always significantly below than that of S_1 and very close to zero for relatively large values of r^*.

Two aspects related to the second part of Theorem 3.2 are worth mentioning. The first aspect is that the result still holds if we modify the test function S_2 in order to admit any matrix in the space of all correlation matrices Ψ_1 (even singular ones). This is, suppose that for $\varepsilon > 0$ we define the test function

$$\tilde{S}_2(m, \Sigma) = \inf_{t=(t_1,0_v):t_1 \in \mathbb{R}^p_{+,+\infty}} (m - t)' \tilde{\Sigma}_\varepsilon^{-1} (m - t), \tag{3.14}$$

where

$$\tilde{\Sigma}_\varepsilon = \Sigma + 1\{\det(\Omega) < \varepsilon\} \varepsilon D, \quad D = diag(\Sigma), \quad \Omega = D^{-1/2} \Sigma D^{-1/2}. \tag{3.15}$$

The function $\tilde{S}_2$ is well defined on Ψ_1 and leads to the test statistic

$$\tilde{T}_{2,n}(\theta) = \inf_{t=(t_1,0_v):t_1 \in \mathbb{R}^p_{+,+\infty}} (n^{1/2} \bar{m}_n(\theta) - t)' \tilde{\Sigma}_{\varepsilon,n}(\theta)^{-1} (n^{1/2} \bar{m}_n(\theta) - t). \tag{3.16}$$

where $\tilde{\Sigma}_{\varepsilon,n}(\theta)$ is a sample analog of $\tilde{\Sigma}_\varepsilon$ based on Equations (2.21), (2.29), and (3.15). This new test function is numerically equal to S_2 when the determinant of the correlation matrix is at least ε, but it changes the weighting matrix when Ω is singular or close to singular. If we let $AsyCS^{(\tilde{2})}$ denote the asymptotic confidence size of CSs based on $\tilde{S}_2$, we have the following corollary to Theorem 3.2.

Corollary 3.2. *Suppose all the assumptions in Theorem 3.2 hold and that r in the definition of $\mathcal{F}_n$ is such that $r^* > 0$. Then, for every $\eta > 0$ there exists an $\varepsilon > 0$ in the definition of $\tilde{S}_2$ such that*

$$AsyCS^{(\tilde{2})}_{PA} \leq \eta. \tag{3.17}$$

The second aspect is related to the requirement in Assumption A.7 that the parameter space is sufficiently rich in the following sense. Assumption A.7 requires that at least one inequality in Equation (2.1) is violated and strongly negatively correlated with another inequality that is either violated or equal to zero. This ensures that the relationship illustrated in Equation (3.12) holds and so the difference between $S_2(z, r_1, \Omega_\varepsilon)$ and $S_2(z, 0, \Omega_\varepsilon)$ increases as $\varepsilon \to 0$. When $p = 2$ it can be shown that this is actually a necessary condition to obtain

[10] Table 1 shows this clearly for S_1, but less clearly 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 reported are relatively optimistic for S_2. The random correlation matrices were generated with the method of Marsaglia and Olkin (1984).

the second part in Theorem 3.2.[11] In the general case, there are alternative ways to make the parameter space rich enough, but Assumption A.7 has the additional advantage of making the optimization problem in Equation (2.25) tractable. However, the second part of Theorem 3.2 should be interpreted as a warning message. Unless the researcher is certain that it is impossible for inequalities that are violated to be negatively correlated with each other or with other inequalities that are binding, inference based on S_2 could be extremely distorted.

4 Numerical Simulations

In this section we perform a small simulation study to analyze whether the result in Theorem 3.1 are relevant in finite samples. We consider the model in Example 2.1 where

$$E_{F_0}(Y|X=x) = G(x,\theta_0), \forall x \in S_X, \tag{4.1}$$

$\theta_0 \in \mathbb{R}^2$ is the true parameter value, and $S_X = \{x_l\}_{l=1}^{d_x} \subset \mathbb{R}^2$ is the support of X. The data is simulated according to the following parametrization. $Y \in \{0,1\}$ is a binary random variable, $S_X = \{(1,0),(0,1)\}$, $\theta_0 = (0.1,-0.5)$, $G(x,\theta_0) = \Phi(\theta_0' x)$ (where $\Phi(\cdot)$ denotes the standard normal cdf so that the model under F_0 is a Probit model) and

$$\pi(x) = \Pr(Z=1|X) = I(X=(1,0)) + 1/2 I(X=(0,1)). \tag{4.2}$$

The observed data is $W_i = (Y_i Z_i, Z_i, X_i)$. Note that when $X=(1,0)$, Y is always observed. The model then results in one equality and two inequalities. We compare PA, GMS, and subsampling critical values together with the test function S_1 from Equation (2.23). The rest of the parameters are as follows: sample size is $n=1000$, number of subsampling/GMS replications is $B=200$, number of simulations is $MC=500$, subsampling block sizes are $b=\{n^{1/3},(n^{1/3}+n^{1/2})/2,n^{1/2}\}$, and GMS tuning parameters are $\kappa_n = \{\ln\ln n, (\ln\ln n + \ln n)/2, \ln n\}$.

For simplicity, instead of focusing on the asymptotic confidence sizes of the CSs we look at the asymptotic size distortion of the tests for the null hypothesis $H_0 : \theta_0 = (0.1,-0.5)$. This simplifies the computations significantly and provides an analogous analysis by Remark 2.3. Under the null hypothesis it follows that

$$E_{F_0}(Y|X=(1,0)) = 0.54 \text{ and } E_{F_0}(Y|X=(0,1)) = 0.31. \tag{4.3}$$

A perturbation F' of F_0 results in different values of the expectations $E_{F'}(Y|X=x)$.[12] For the simulation exercise we consider the set of all distributions F' such that $E_{F'}(Y|X=x)$ is

[11] For instance, in Examples 2.1 and 2.2 there are only two inequalities and the models are restricted in a way than when one inequality is negative, the other one is necessarily positive. However, in Example 2.2 this relationship is not present when there are more than two firms and the model includes additional covariates.

[12] For example, if the model is Logit then $E_{F'}(Y|x=(1,0)) = \Lambda(0.1) = 0.53$ and $E_{F'}(Y|x=(0,1)) = \Lambda(-0.5) = 0.38$, where Λ denotes the logistic cumulative distribution function.

r^*	PA-non-adj	PA	SS1	SS2	SS3	GMS1	GMS2	GMS3
0.00	0.05	0.10	0.10	0.10	0.10	0.10	0.10	0.10
0.30	0.06	0.12	0.13	0.11	0.11	0.12	0.12	0.12
0.60	0.10	0.17	0.15	0.15	0.14	0.17	0.17	0.17
0.90	0.19	0.27	0.24	0.28	0.23	0.27	0.27	0.27
1.20	0.28	0.41	0.34	0.37	0.34	0.39	0.39	0.39
1.50	0.61	0.70	0.66	0.67	0.66	0.69	0.69	0.69
1.80	0.70	0.79	0.77	0.76	0.76	0.77	0.77	0.77
2.10	0.79	0.84	0.81	0.83	0.81	0.84	0.84	0.84
2.40	0.94	0.97	0.95	0.96	0.96	0.96	0.96	0.96
2.70	0.94	0.97	0.95	0.96	0.96	0.96	0.96	0.96
3.00	0.99	1.00	0.99	0.99	1.00	1.00	1.00	1.00
3.30	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: Adjusted maximum rejection probability over models that are a distance r^* or less away from F_0 under H_0. Simulation parameters: $n = 1000$, $\alpha = 0.10$, $B = 200$, $b = \{n^{1/3}, (n^{1/3} + n^{1/2})/2, n^{1/2}\}$, $\kappa_n = \{\ln \ln n, (\ln \ln n + \ln n)/2, \ln n\}$, and $MC = 500$.

in a neighborhood of $E_{F_0}(Y|X = x) = (0.54, 0.31)$. In particular, the set of models that are a distance $d = r^* n^{-1/2} \geq 0$ from F_0 are defined as

$$\mathcal{F}_{r^*} \equiv \{F' : \max\{d_{F',1}(\theta_0), d_{F',2}(\theta_0), d_{F',3}(\theta_0)\} \leq r^* n^{-1/2}\} \tag{4.4}$$

where

$$d_{F',1}(\theta_0) = |E_{F'} m_1(W_i, \theta_0)/\sigma_{F',1}(\theta_0)| \tag{4.5}$$

$$d_{F',2}(\theta_0) = |[E_{F'} m_2(W_i, \theta_0)/\sigma_{F',2}(\theta_0)]_-| \tag{4.6}$$

$$d_{F',3}(\theta_0) = |[E_{F'} m_3(W_i, \theta_0)/\sigma_{F',3}(\theta_0)]_-|. \tag{4.7}$$

Given $r^* \geq 0$ we explore *all* models that are in the ball and compare the maximum rejection probabilities across inferential methods. This is, we report

$$\sup_{F \in \mathcal{F}_{r^*}} \Pr_F(T_n(\theta_0) > c_n(\theta_0, 1 - \alpha)), \tag{4.8}$$

for each choice of critical value, which involves simulating data from all $F' \in \mathcal{F}_{r^*}$.

The results are reported in Table 2. The table shows size corrected maximum rejection probabilities (for $\alpha = 0.10$). From the table we see that the maximum rejection probability of subsampling and GMS are extremely close. Results for subsampling are particularly sensitive to the choice of the block size. Overall, the finite sample rejection probabilities of GMS and subsampling are very similar and the differences are not statistically significant given the $MC = 500$ simulations. Finally, the table also shows the non-adjusted maximum rejection probabilities for PA, as PA is actually conservative for testing H_0 in this model. This illustrates that the robustness of PA is related to the fact that under correct specification ($r^* = 0$) the method is conservative. All these results are in line with Theorem 3.1.

Appendices

Appendix A Additional Definitions and Assumptions

To determine the asymptotic confidence size in Equation (2.17) we calculate the limiting coverage probability along a sequence of "worst case parameters" $\{\theta_n, F_n\}_{n\geq 1}$ with $(\theta_n, F_n) \in \mathcal{F}_n, \forall n \in \mathbb{N}$. See also Andrews and Guggenberger (2009a,b,2010a,b). We start with the following definition.

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

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

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

Denote by GH the set of points $(g_1, h) \in \mathbb{R}^k_{+\infty} \times H$ such that there is a subsequence $\{\omega_n\}_{n\geq 1}$ of $\mathbb{N}$ and a sequence $\{\gamma_{\omega_n,h}\}_{n\geq 1}$ that satisfies[13]

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

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

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

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

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

Our assumptions imply that elements of H satisfy certain properties. For example, for any $h \in H$, h_1 is constrained to satisfy $h_{1,j} \geq -r_j$ for $j = 1,\ldots,p$ and $|h_{1,j}| \leq r_j$ for $j = p+1,\ldots,k$, and h_2 is a correlation matrix. Note that the set H depends on the choice of S through Ψ. Note that $b/n \to 0$ implies that if $(g_1, h) \in GH$ and $h_{1,j}$ is finite $(j = 1,\ldots,k)$, then $g_{1,j} = 0$. In particular, $g_{1,j} = 0$ for $j > p$ by Equation (2.14)(iii). Analogous statements hold for ΠH. Finally, the spaces H, GH, and ΠH for the case of hypothesis testing (see Remark 2.3) are defined analogously for a sequence $\gamma_{\omega_n,h} = \{\theta, F_{\omega_n,h}\}_{n\geq 1}$ where θ is fixed.

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 slightly 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). Assumptions A.5-A.7 are new.

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

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

(a) $S\left((m_1, m_1^*), \Sigma\right)$ *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 matrix* $\Delta \in \mathbb{R}^{k\times k}$,

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

(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* $h_1 \in [-r_j, \infty]_{j=1}^p \times [-r_j, r_j]_{j=p+1}^k$, *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* $h_1 = \infty_p$, *and*

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

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

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

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

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

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

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

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

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

Let $\Xi_{l,l'}(\varepsilon) \in \mathbb{R}^{k\times k}$ be an identity matrix except for the (l, l') and (l', l) components that are equal to $-\sqrt{1-\varepsilon}$ for some $l, l' \in \{1, \ldots, p\}$.

Assumption A.7. *There exists* $h \in H$ *such that* $h_{1,l} \leq 0$, $h_{1,l'} \leq 0$, $\min\{h_{1,l}, h_{1,l'}\} < 0$, *and* $h_2 = \Xi_{l,l'}(\varepsilon)$ *for some* $l, l' \in \{1, \ldots, p\}$ *with* $l \neq l'$.

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

Note that S_1 and S_2 satisfy Assumption A.2 which is a strengthened version of Assumption 2 from AG using the same proof as in AG. Assumption A.3 implies that $S(\infty_p, \Omega) = 0$ when $v = 0$. Assumption A.5 makes sure the parameter space is sufficiently rich. Assumption A.6 holds by Assumption A.2(a) if there exists $h^* \in H$ such that $J_{h^*}(c_0(h_2^*, 1-\alpha)) < J_{(0,h_2^*)}(c_0(h_2^*, 1-\alpha))$. Also note that by Assumption A.1(a), a $h^* \in H$ as in Assumption A.6 needs to have $h_{1,j}^* < 0$ for some $j \le p$ or $h_{1,j}^* \ne 0$ for some $j > p$. Assumptions A.5 and A.6 are verified for the two lead example in Appendix D. Assumption A.7 guarantees two things. First, it guarantees that at least two inequalities in Equation (2.1) are violated (or at least, one is violated and the other one is binding) and negatively correlated. Second, it guarantees that there are correlation matrices with zeros outside the diagonal except at two spots. This part of the assumption simplifies the proof significantly but it could be replaced with alternative forms of correlation matrices.

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 $\{\gamma_{\omega_n,h}\}_{n\ge1} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\ge1}$ defined in definition A.1 for a subsequence $\{\omega_n\}_{n\ge1}$ 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 $Z = (Z_1, \dots, Z_k) \sim N(0_k, I_k)$.

Lemma B.2. *Consider CSs with nominal confidence size $1-\alpha$ for $0 < \alpha < 1/2$. Assume the nonempty parameter space is given by $\mathcal{F}_n$ in Equation (2.14) for some $r \in \mathbb{R}_+^k$, $\delta > 0$, and $M < \infty$. Assume S satisfies Assumptions A.1-A.3. For GMS CSs assume that $\varphi(\xi, \Omega)$ satisfies Assumption A.4 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*

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

where $J_h(x) = P(J_h \le x)$ and π_1^*, $\pi_1^{**} \in \mathbb{R}_{+\infty}^k$ with jth element defined by

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

Lemma B.3. *For any $a \in (0,1)$ and $\rho \in [-1+a, 1-a]$ define*

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

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

1. *Let $z_1 \ge 0$ and $z_2 \ge 0$. Then, $f(z_1, z_2, \rho) = 0$.*
2. *Let $z_1 \ge 0$ and $z_2 < 0$. If $\rho \le z_1/z_2$, 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. *Let $z_1 < 0$ and $z_2 \ge 0$. If $\rho \le z_2/z_1$, then Equation (B-5) holds. Otherwise $f(z_1, z_2, \rho) = z_1^2$.*
4. *Let $z_1 < 0$ and $z_2 < 0$. If $\rho \le \min\{z_1/z_2, z_2/z_1\}$, then Equation (B-5) holds. Otherwise $f(z_1, z_2, \rho) = \max\{z_1^2, z_2^2\}$.*

Lemma B.4. *Suppose that $k = p = 2$. For $\beta > 0$ define*

$$\bar{H}_\beta \equiv \left\{(h_1, h_2) \in \mathbb{R}^2 \times \Psi_1 : h_{1,1} \le -\beta,\ h_{1,2} \le 0,\ h_2 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},\ \rho \le -\beta\right\}. \tag{B-6}$$

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

$$A^a_{h_1,\rho} \equiv \{z \in \mathbb{R}^2 : z_1 \ge 0, z_2 < 0, 0 < z_1 - \rho z_2 \le -h_{1,1} + \rho h_{1,2}\}, \tag{B-7}$$
$$A^b_{h_1,\rho} \equiv \{z \in \mathbb{R}^2 : z_1 < 0, z_2 \ge 0, 0 < z_2 - \rho z_1 \le -h_{1,2} + \rho h_{1,1}\}, \tag{B-8}$$
$$A^{c1}_{h_1,\rho} \equiv \{z \in \mathbb{R}^2 : z_1 \ge 0, z_2 < 0, z_1 - \rho z_2 \le 0\}, \tag{B-9}$$
$$A^{c2}_{h_1,\rho} \equiv \{z \in \mathbb{R}^2 : z_1 < 0, z_2 \ge 0, z_2 - \rho z_1 \le 0\}, \tag{B-10}$$

and $A^c_{h_1,\rho} \equiv A^{c1}_{h_1,\rho} \cup A^{c2}_{h_1,\rho}$. Letting $Z_{h_2} \sim N(0, h_2)$ we have

1. *$\forall \eta > 0$, $\exists \bar{\rho} > -1$ such that $\inf_{(h_1, h_2 = \bar{h}_2) \in \bar{H}_\beta} \Pr(Z_{\bar{h}_2} \in A_{h_1,\bar{\rho}}) \ge 1 - \eta$, where $\bar{h}_2 = \begin{bmatrix} 1 & \bar{\rho} \\ \bar{\rho} & 1 \end{bmatrix}$.*
2. *There exists a real valued function $\tau(z, h_1, h_2)$ such that*

$$\Pr(\inf_{(h_1,h_2) \in \bar{H}_\beta} \tau(Z_{h_2}, h_1, h_2) > 0 | Z_{h_2} \in A_{h_1,\rho}) = 1 \tag{B-11}$$

and, for the function S_2 defined in Equation (2.25),

$$S_2(z + h_1, h_2) = S_2(z, h_2) + \frac{1}{1 - \rho^2}\tau(z, h_1, h_2), \quad \forall z \in A_{h_1,\rho},\ \forall (h_1, h_2) \in \bar{H}_\beta. \tag{B-12}$$

Appendix C Proof of 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 $j = 1, \ldots, k$, 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 Assumption (A.3)(x) in that paper, we have that

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

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 by 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 $\{\theta_s, F_s\}_{s \ge 1}$ 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 Equation (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 Equation (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 $j = 1, \ldots, k$, we have

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

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

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

by Equations (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 $j = 1, \ldots, p$), then

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

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

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

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 A.1(d) $S^*(y, \Omega)$ is continuous at all (y, Ω) for $y \in (0,1]^p \times (0,1)^v$ and $\Omega \in \Psi$. We now have

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

where the convergence holds by Equations (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

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

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

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

and $\hat{\Omega}_s(\theta) = (\hat{D}_s(\theta))^{-1/2}\hat{\Sigma}_s(\theta)(\hat{D}_s(\theta))^{-1/2}$. From Equation (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 $S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n})$ for $Z \sim N(0_k, I_k)$. We first show the existence of random variables $c^*_{\omega_n}$ and $c^{**}_{\omega_n}$ such that under $\{\gamma_{\omega_n}\}$

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

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 jth component given by

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

Define $c^*_{\omega_n}$ as the $1-\alpha$ quantile of $S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi^*(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n})$. As $\varphi^*_j \geq \varphi_j$ it follows from Assumption A.1(a) that $c^*_{\omega_n} \leq \hat{c}_{\omega_n,\kappa_{\omega_n}}(\theta_{\omega_n}, 1-\alpha)$ a.s. $[Z]$ under $\{\gamma_{\omega_n}\}_{n\geq 1}$. Furthermore, by Lemma 2(a) in the 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 $S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi^{**}(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n})$. Note that the definition of $\varphi^{**}_j(\xi, \Omega)$ implies that $\varphi^{**}_j \leq \varphi_j$. The same steps as in the proof of 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, $0 \leq S(\hat{h}_{2,\omega_n}^{1/2} Z + \varphi(\xi_{\omega_n}(\theta_{\omega_n}, \hat{h}_{2,\omega_n})), \hat{h}_{2,\omega_n}) \to_p 0$. This implies the second line in 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 Equation (C-10) and

Lemma 5 in AG that

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

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

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 $\{\gamma_{\omega_n,\pi_1,h}\}_{n\ge1}$ with $h = (\infty_p, h_2)$ we have

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

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

Subsampling critical value: Instead of $\{\gamma_{\omega_n,g_1,h}\}_{n\ge1} = \{\theta_{\omega_n,g_1,h}, F_{\omega_n,g_1,h}\}_{n\ge1}$ 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 $\gamma_3 = (\theta, F)$, $\gamma_1 = (\sigma^{-1}_{F,j}(\theta) E_F m_j(W_i, \theta))_{j=1}^k \in \mathbb{R}^k$, and $\gamma_2 = Corr_F(m(W_i, \theta)) \in \mathbb{R}^{k\times k}$ for (θ, F) introduced in the model defined in (2.14). Then, Assumption A0 in AG clearly holds. With $\{\gamma_{\omega_n,h}\}_{n\ge1}$ 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 $\{\gamma_{w_n,g_1,h}\}_{n\ge1}$

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

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

$$-(z_1 - u_1) + \rho(z_2 - u_2) \ge 0, \quad u_1[-(z_1 - u_1) + \rho(z_2 - u_2)] = 0, \quad u_1 \ge 0, \tag{C-16}$$
$$-(z_2 - u_2) + \rho(z_1 - u_1) \ge 0, \quad u_2[-(z_2 - u_2) + \rho(z_1 - u_1)] = 0, \quad u_2 \ge 0. \tag{C-17}$$

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 $u_1 \ge 0$, $u_2 > 0$ 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 $u_1 \ge 0$ and $u_2 = 0$. Then, it follows from the conditions in Equation (C-16) that $u_1 \ge z_1 - \rho z_2$, so that $u_1 = \max\{z_1 - \rho z_2, 0\}$ and $u_2 = 0$ is the solution. This is a strictly convex optimization problem and so the solution exists and is unique. Then, if $\rho \le z_1/z_2$, the unique solution is $(u_1, u_2) = (0, 0)$ and the objective function is given by Equation (B-5). On the other hand, if $\rho < z_1/z_2$, $(u_1, u_2) = (z_1 - \rho z_2, 0)$ is the unique solution and

$$f(z_1, z_2, \rho) = (1-\rho^2)^{-1}\left\{(z_1 - z_1 + \rho z_2)^2 + z_2^2 - 2\rho(z_1 - z_1 + \rho z_2)(z_2)\right\} = z_2^2. \tag{C-18}$$

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). Define the set $A^{ac1}_{h_1,\rho} \equiv A^a_{h_1,\rho} \cup A^{c1}_{h_1,\rho}$. 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

$$\Pr(Z_{h_2} \in A^{ac1}_{h_1,\rho}) = \Pr\left(Z_{h_2,2} \leq \min\{0, \frac{\sqrt{1-\rho^2}W}{-\rho}\}, W \leq \frac{-h_{1,1} + \rho h_{1,2}}{\sqrt{1-\rho^2}}\right) \to 1/2, \text{ as } \rho \to -1. \tag{C-19}$$

The same applies for the set $A^{bc2}_{h_1,\rho} \equiv A^b_{h_1,\rho} \cup A^{c2}_{h_1,\rho}$ and the result follows by continuity in ρ.

We now prove (2). Note that

$$S_2(z + h_1, h_2) = (1-\rho^2)^{-1} \min_{t \in \mathbb{R}^2_{+,\infty}} \left\{(\bar{z}_1 - t_1)^2 + (\bar{z}_2 - t_2)^2 - 2\rho(\bar{z}_1 - t_1)(\bar{z}_2 - t_2)\right\} \tag{C-20}$$

is the same optimization problem as the one in Lemma B.3 by letting $\bar{z}_j = z_j + h_{1,j}$, $j = 1, 2$. It follows from Lemma B.3 that the solution of Equation (C-20) for $z \in A_{h_1,\rho}$ and $(h_1, h_2) \in \bar{H}_\beta$ is

$$S_2(z + h_1, h_2) = (1-\rho^2)^{-1}[(z_1 + h_{1,1} - z_2 - h_{1,2})^2 + 2(1-\rho)(z_1 + h_{1,1})(z_2 + h_{1,2})]. \tag{C-21}$$

In addition, it follows from Lemma B.3 that the solution when $h_{1,1} = h_{1,2} = 0$ is given by $S_2(z, h_2) = z_2^2$ for $z \in A^a_{h_1,\rho}$, $S_2(z, h_2) = z_1^2$ for $z \in A^b_{h_1,\rho}$, and $S_2(z, h_2) = z_1^2 + (z_2 - \rho z_1)^2/(1-\rho^2)$ for $z \in A^c_{h_1,\rho} \equiv A^{c1}_{h_1,\rho} \cup A^{c2}_{h_1,\rho}$. After doing some algebraic manipulations it follows that

$$S_2(z + h_1, h_2) = S_2(z, h_2) + \frac{1}{1-\rho^2}\tau_l(z, h_1, h_2) \quad \forall z \in A^l_{h_1,\rho},\ l \in \{a, b, c\}, \tag{C-22}$$

where

$$\tau_a(z, h_1, h_2) = (z_1 + h_{1,1} - \rho(z_2 + h_{1,2}))^2 + (1-\rho^2)(h_{1,2}^2 + 2z_2 h_{1,2}), \tag{C-23}$$

$$\tau_b(z, h_1, h_2) = (z_2 + h_{1,2} - \rho(z_1 + h_{1,1}))^2 + (1-\rho^2)(h_{1,1}^2 + 2z_1 h_{1,1}), \tag{C-24}$$

$$\tau_c(z, h_1, h_2) = (h_{1,1} - h_{1,2})^2 + 2((z_2 - \rho z_1)(h_{1,2} - \rho h_{1,1}) + h_{1,1}z_1(1-\rho^2) + (1-\rho)h_{1,1}h_{1,2}). \tag{C-25}$$

Note that $\tau_a(z, h_1, h_2) = 0$ on $A^a_{h_1,\rho} \times \bar{H}_\beta$ if and only if $h_{1,2} = 0$ and $h_{1,1} = \rho z_2 - z_1$, while $\tau_b(z, h_1, h_2) > 0$ on $A^b_{h_1,\rho} \times \bar{H}_\beta$ and $\tau_c(z, h_1, h_2) > 0$ on $A^c_{h_1,\rho} \times \bar{H}_\beta$. Thus, letting

$$\tau(z, h_1, h_2) \equiv \sum_{l \in \{a,b,c\}} \tau_l(z, h_1, h_2) I(z \in A^l_{h_1,\rho}), \tag{C-26}$$

it follows immediately that

$$\Pr(\inf_{(h_1,h_2)\in\bar{H}_\beta} \tau(Z_{h_2}, h_1, h_2) > 0 | Z_{h_2} \in A_{h_1,\rho}) = 1. \tag{C-27}$$

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 $\{\omega_n\}_{n\geq 1}$ and a sequence $\{\gamma_{\omega_n,\pi_1,h}\}_{n\geq 1}$ 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 $(g_1, h) \in GH$. To show $AsyCS_{SS} \geq AsyCS_{GMS}$, 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 $\{\omega_n\}_{n\geq 1}$ and a sequence $\{\gamma_{\omega_n,g_1,h}\}_{n\geq 1}$. Because $\kappa_n^{-1} n^{1/2}/b_n^{1/2} \to \infty$ it follows that there exists a subsequence $\{v_n\}_{n\geq 1}$ of $\{\omega_n\}_{n\geq 1}$ such that under $\{\gamma_{v_n,g_1,h}\}_{n\geq 1}$

$$\kappa_{v_n}^{-1} v_n^{1/2} \sigma^{-1}_{F_{v_n,h},j}(\theta_{v_n,h}) E_{F_{v_n,h}} m_j(W_i, \theta_{v_n,h}) \to \pi_{1,j}, \tag{C-28}$$

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

□

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

$$AsyCS_{GMS}^{(1)} \geq \inf_{(\pi_1,h)\in\Pi H} \Pr\left(S_1(h_2^{1/2} Z + h_1, h_2) \leq c_{\pi_1^*}(h_2, 1-\alpha)\right), \tag{C-29}$$

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

$$S_1(h_2^{1/2} Z + h_1, h_2) = \sum_{j=1}^{p} [h_2^{1/2}(j) Z + h_{1,j}]_-^2 + \sum_{j=p+1}^{k} (h_2^{1/2}(j) Z + h_{1,j})^2, \tag{C-30}$$

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 $j \geq 1$

$$\sum_{s=1}^{k} (h_2^{1/2}(j,s))^2 = 1, \quad h_2^{1/2}(j,s) = 0,\ \forall s > j, \quad |h_2^{1/2}(j,s)| \leq 1,\ \forall s \geq 1. \tag{C-31}$$

The properties in Equation (C-31) follow by h_2 having ones in the main diagonal and $h_2^{1/2}$ being lower triangular. We use Equation (C-31) and the Cauchy-Schwarz inequality to derive the following three useful inequalities. For any $z \in \mathbb{R}^k$ and $j = 1, \ldots, k$,

$$(h_2^{1/2}(j)z + h_{1,j})^2 = \left(\sum_{s=1}^{j} h_2^{1/2}(j,s)z_s + \sum_{s=1}^{j}(h_2^{1/2}(j,s))^2 h_{1,j}\right)^2$$

$$\leq \sum_{m=1}^{j}(h_2^{1/2}(j,m))^2 \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2 = \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \quad \text{(C-32)}$$

$$[h_2^{1/2}(j)z + h_{1,j}]_-^2 \leq \sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \; \textit{and} \quad \text{(C-33)}$$

$$[h_2^{1/2}(j)z + h_{1,j}]_-^2 \leq [h_2^{1/2}(j)z]_-^2 \leq \sum_{s=1}^{j} z_s^2, \quad \text{provided } h_{1,j} \in (0,\infty). \quad \text{(C-34)}$$

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

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

$$+ \sum_{j=p+1}^{k}\sum_{s=1}^{j}(z_s + h_2^{1/2}(j,s)h_{1,j})^2, \quad \text{(C-35)}$$

and it follows from Equations (C-32), (C-33), and (C-34) that $\tilde{S}_1(z,h) \geq S_1(h_2^{1/2}z + h_1, h_2)$ for all $z \in \mathbb{R}^k$. Therefore, for all $h \in H$ and $x \in \mathbb{R}$

$$\Pr(S_1(h_2^{1/2}Z + h_1, h_2) \leq x) \geq \Pr(\tilde{S}_1(Z,h) \leq x). \quad \text{(C-36)}$$

Let $B > 0$ and

$$A_B \equiv \{z \in \mathbb{R} : |z| \leq B\} \quad \text{and} \quad A_B^k = A_B \times \cdots \times A_B \text{ (with } k \text{ copies)}. \quad \text{(C-37)}$$

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

$$\Pr(Z \in A_B^k) = \prod_{s=1}^{k}\Pr(Z_s \in A_B) > 0. \quad \text{(C-38)}$$

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

$$\inf_{(\pi,h)\in\Pi H}\Pr(S_1(h_2^{1/2}Z + h_1, h_2) \leq c_{\pi_1^*}(h_2, 1-\alpha)) = \lim_{l\to\infty}\Pr(S_1(h_{2,l}^{1/2}Z + h_{1,l}, h_{2,l}) \leq c_{\pi_{1,l}^*}(h_{2,l}, 1-\alpha)), \quad \text{(C-39)}$$

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, it follows that for all $z_s \in A_{B_{\omega_l}}$ and $j = 1, \ldots, k$

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

when $|h_{1,j}| \leq r_j$. Then, for all $z \in A_{B_{\omega_l}}^k$

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

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

$$\Pr(\tilde{S}_1(Z, h_{\omega_l}) \leq c_{\pi^*_{1,\omega_l}}(h_{2,\omega_l}, 1-\alpha)) \geq \Pr(Z \in A^k_{B_{\omega_l}}) > 0. \tag{C-42}$$

It follows from Equations (C-29), (C-36), (C-38), (C-39), and (C-42) that

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

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

Part 2. By Lemma B.2

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

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. Let $h^*_{2,\varepsilon} = \Xi_{1,2}(\varepsilon)$, where $\Xi_{1,2}(\varepsilon) \in \mathbb{R}^{k\times k}$ is an identity matrix except for the $(1,2)$ and $(2,1)$ components that are equal to $-\sqrt{1-\varepsilon}$. By Assumption A.7 and without loss of generality, there exists $h_1 \in \mathbb{R}^k$ with $h_{1,1} \leq 0$, $h_{1,2} \leq 0$, and $\min\{h_{1,1}, h_{1,2}\} < 0$ such that $(h_1, h^*_{2,\varepsilon}) \in H$. It follows that $\det(h^*_{2,\varepsilon}) = \varepsilon$ and

$$(h^*_{2,\varepsilon})^{-1} = \begin{bmatrix} A^{-1} & 0_{2\times(k-2)} \\ 0_{(k-2)\times 2} & I_{k-2} \end{bmatrix}, \quad \text{where} \quad A^{-1} = \frac{1}{1-\underline{\rho}^2}\begin{bmatrix} 1 & -\underline{\rho} \\ -\underline{\rho} & 1 \end{bmatrix}, \tag{C-45}$$

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

$$\begin{aligned} S_2(Z^\star + h_1, h^\star_{2,\varepsilon}) = \inf_{t\in\mathbb{R}^p_{+,+\infty}} &\big\{(1-\underline{\rho}^2)^{-1}[(Z^\star_1 + h_{1,1} - t_1)^2 + (Z^\star_2 + h_{1,2} - t_2)^2 \\ &- 2\underline{\rho}(Z^\star_1 + h_{1,1} - t_1)(Z^\star_2 + h_{1,2} - t_2)] + \sum_{j=3}^{p}(Z^\star_j + h_{1,j} - t_j)^2\big\} + \sum_{j=p+1}^{k}(Z^\star_j + h_{1,j})^2. \end{aligned} \tag{C-46}$$

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

$$\begin{aligned} S_2(Z^\star + h_1, h^\star_{2,\varepsilon}) = \inf_{t\in\mathbb{R}^2_{+,+\infty}} &\big\{(1-\underline{\rho}^2)^{-1}[(Z^\star_1 + h_{1,1} - t_1)^2 + (Z^\star_2 + h_{1,2} - t_2)^2 \\ &-2\underline{\rho}(Z^\star_1 + h_{1,1} - t_1)(Z^\star_2 + h_{1,2} - t_2)]\big\} + \sum_{j=3}^{p}[Z^\star_j + h_{1,j}]^2_- + \sum_{j=p+1}^{k}(Z^\star_j + h_{1,j})^2. \end{aligned} \tag{C-47}$$

The optimization problem in the RHS of Equation (C-47) is the same as the one in Lemma B.3 and, by that lemma, the solution can be divided into four cases depending on whether $Z^\star_1$ and $Z^\star_2$ are positive or negative. We will show that the solution described in Equation (B-5) in Lemma B.3 holds with probability approaching one as $\varepsilon \to 0$. To do this let $r^* > 0$ be given. By Assumption A.7 $\min\{h_{1,1}, h_{1,2}\} < 0$ so without loss of generality let $h_{1,1} < 0$. For small $\beta > 0$ let

$$\bar{H}_{\beta,\varepsilon} \equiv \{h_1 \in \mathbb{R}^k : (h_1, h^\star_{2,\varepsilon}) \in H,\ h_{1,1} \le -\beta,\ h_{1,2} \le 0\}. \tag{C-48}$$

Define subsets of $\mathbb{R}^k$ by letting $\tilde{A}_{h_1,\rho} \equiv A_{h_1,\rho} \times \mathbb{R}^{k-2}$ for $A_{h_1,\rho}$ defined in Lemma B.4. The set $A_{h_1,\rho}$ only depends on $(h_{1,1}, h_{1,2})$ but we use $A_{h_1,\rho}$ instead of $A_{(h_{1,1},h_{1,2}),\rho}$ to simplify the notation. Note that $\tilde{A}_{h_1,\rho}$ does not restrict z_j for $j = 3, \dots, k$. It follows from Lemma B.4 that $\forall \eta > 0$, $\exists \varepsilon > 0$ such that

$$\inf_{h_1 \in \bar{H}_{\beta,\varepsilon}} \Pr(Z^\star \in \tilde{A}_{h_1,\underline{\rho}}) = \inf_{h_1 \in \bar{H}_{\beta,\varepsilon}} \Pr((Z^\star_1, Z^\star_2) \in A_{h_1,\underline{\rho}}) \ge 1 - \eta. \tag{C-49}$$

For the next step define the function

$$\underline{S_2}(z, \underline{\rho}) = \inf_{t \in \mathbb{R}^2_{+,+\infty}} \left\{(1-\underline{\rho}^2)^{-1}[(z_1 - t_1)^2 + (z_2 - t_2)^2 - 2\underline{\rho}(z_1 - t_1)(z_2 - t_2)]\right\}, \tag{C-50}$$

and note that by Equation (C-47) and Lemma B.4, there exists a function $\tau(z, h_1, h^*_{2,\varepsilon})$ that is positive with probability 1 such that

$$S_2(z + h_1, h^\star_{2,\varepsilon}) \ge \underline{S_2}(z, \underline{\rho}) + \frac{1}{1-\underline{\rho}^2}\tau(z, h_1, h^\star_{2,\varepsilon}),\ \text{for all } z \in \tilde{A}_{h_1,\underline{\rho}}. \tag{C-51}$$

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

$$\Pr\left(\inf_{h_1 \in \bar{H}_{\beta,\varepsilon}} [c_0(h^\star_{2,\varepsilon}, 1-\alpha) - \frac{1}{1-\underline{\rho}^2}\tau(Z^\star, h_1, h^\star_{2,\varepsilon})] \le -\eta | Z^\star \in \tilde{A}_{h_1,\underline{\rho}}\right) = 1. \tag{C-52}$$

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

$$S_2(Z^\star, h^\star_{2,\varepsilon}) = \sum_{j=3}^{p}[Z^\star_j]^2_- + \sum_{j=p+1}^{k}(Z^\star_j)^2 + f(Z^\star_1, Z^\star_2, \underline{\rho}) \le \sum_{j=3}^{p}[Z^\star_j]^2_- + \sum_{j=p+1}^{k}(Z^\star_j)^2 + (Z^\star_1)^2 + W^2, \tag{C-53}$$

where $f(\cdot)$ is defined in Lemma B.3 (Equation (B-5)) and satisfies $f(Z^\star_1, Z^\star_2, \underline{\rho}) \le (Z^\star_1)^2 + W^2$ with probability one for all $\varepsilon > 0$, and $Z^\star_1 \perp W \sim N(0,1)$. As a result, the $1-\alpha$ quantile of $S_2(Z^\star, h^\star_{2,\varepsilon})$, denoted by $c_0(h^*_{2,\varepsilon}, 1-\alpha)$, is bounded above by the $1-\alpha$ quantile of the RHS of Equation (C-53), denoted by $\tilde{c}_0(1-\alpha)$. Note that $\tilde{c}_0(1-\alpha)$ does not depend on ε. It then follows that $c_0(h^*_{2,\varepsilon}, 1-\alpha) \le \tilde{c}_0(1-\alpha) < \infty$ and Equation (C-52) follows immediately from

$$\Pr\left(\inf_{h_1 \in \bar{H}_{\beta,\varepsilon}} [\tilde{c}_0(1-\alpha) - \frac{1}{1-\underline{\rho}^2}\tau(Z^\star, h_1, h^\star_{2,\varepsilon})] < 0 | Z^\star \in \tilde{A}_{h_1,\underline{\rho}}\right) = 1 \tag{C-54}$$

for $\varepsilon > 0$ small enough by Lemma B.4. Finally, to further simplify the notation below let $\Pr(\tilde{A}_{h_1,\rho}) \equiv \Pr(Z^\star \in \tilde{A}_{h_1,\rho})$ and note that for every $\eta > 0$, $\exists \varepsilon > 0$ such that

$$\begin{aligned}
AsyCS^{(2)}_{PA} &\le \inf_{h_1 \in \bar{H}_{\beta,\varepsilon}} \Pr\left(S_2(Z^\star + h_1, h^\star_{2,\varepsilon}) \le c_0(h^\star_{2,\varepsilon}, 1-\alpha)\right) \\
&= 1 - \sup_{h_1 \in \bar{H}_{\beta,\varepsilon}} \Pr\left(S_2(Z^\star + h_1, h^\star_{2,\varepsilon}) > c_0(h^\star_{2,\varepsilon}, 1-\alpha)\right) \\
&\le 1 - \sup_{h_1 \in \bar{H}_{\beta,\varepsilon}} \Pr\left(S_2(Z^\star + h_1, h^\star_{2,\varepsilon}) > c_0(h^\star_{2,\varepsilon}, 1-\alpha) | \tilde{A}_{h_1,\underline{\rho}}\right) \Pr(\tilde{A}_{h_1,\underline{\rho}}) \\
&\le 1 - \sup_{h_1 \in \bar{H}_{\beta,\varepsilon}} \Pr\left(\underline{S_2}(Z^\star, \underline{\rho}) > c_0(h^\star_{2,\varepsilon}, 1-\alpha) - \frac{1}{1-\underline{\rho}^2}\tau(Z^\star, h_1, h^\star_{2,\varepsilon}) | \tilde{A}_{h_1,\underline{\rho}}\right) \Pr(\tilde{A}_{h_1,\underline{\rho}}) \\
&\le \eta,
\end{aligned} \tag{C-55}$$

where the first inequality follows from $\bar{H}_{\beta,\varepsilon} \times h^{\star}_{2,\varepsilon} \subseteq H$, the second inequality from $\tilde{A}_{h_1,\underline{\rho}} \subseteq \mathbb{R}^k$, the third one from Equation (C-51), and the last one from Equations (C-49), (C-52) and $\underline{S_2}(z,\underline{\rho}) \geq 0$, $\forall z \in \mathbb{R}^k$. □

Proof of Corollary 3.1. By Theorem 3.2.1 there exists $B > 0$ such that for all $r^* \leq B$, $AsyCS^{(1)}_{GMS} > 0$. Pick $\eta = AsyCS^{(1)}_{GMS}/2 > 0$. By Theorem 3.2.2 there exists ε such that $AsyCS^{(2)}_{PA} \leq \eta = AsyCS^{(1)}_{GMS}/2$. Therefore, by Theorem 3.1

$$AsyCS^{(2)}_{GMS} = AsyCS^{(2)}_{SS} \leq AsyCS^{(2)}_{PA} < AsyCS^{(1)}_{GMS} = AsyCS^{(1)}_{SS} \leq AsyCS^{(1)}_{PA}. \tag{C-56}$$

□

Appendix D Verification of Assumptions in the Examples

D.1 Example 2.1

We start by writing the example using the notation in Definition 2.1 and using the following primitive assumption. For the assumption we use the following notation. $\Pr_n$ denotes the probability with respect to the distribution F_n, $I_l \equiv I(X = x_l)$, $p_{l,n} = \Pr_n(X = x_l)$, $\pi_{l,n} = \Pr_n(Z = 1|X = x_l)$, $E_{l,n} = E_{F_n}(Y|Z = 1, X = x_l)$, $H_{l,n} = E_{F_n}\left(Y^2|Z = 1, X = x_l\right)$, and $G_{l,n} = G(x_l, \theta_n)$.

Assumption D.1. *Assume that for $c_1, c_2, c_3, c_4 \in \mathbb{R}$: (i) $p_{l,n} \geq c_1 > 0$, (ii) $H_{n,l} \leq c_2 < \infty$, $H_{n,l} - E^2_{n,l} \geq c_3 > 0$, and $\pi_{l,n} \geq c_4 > 0$ for all $n \geq 1$ and $l = 1, \ldots, d_x$.*

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 $Y_H(x_l) = 1$ so that

$$\gamma_{1,l,1,n} \equiv \sigma^{-1}_{F_n,l,1} E_{F_n} m_{1,l}(W_i, \theta_n) = \sigma^{-1}_{F_n,l,1} E_{F_n}[(YZ - G(x_l, \theta_n) + 1 - Z)I_l], \tag{D-1}$$

$$= \sigma^{-1}_{F_n,l,1}\left(\pi_{l,n} E_{l,n} - G_{l,n} + (1 - \pi_{l,n})\right) p_{l,n} \geq -r_{l,1} n^{-1/2}, \tag{D-2}$$

$$\gamma_{1,l,2,n} \equiv \sigma^{-1}_{F_n,l,2} E_{F_n} m_{2,l}(W_i, \theta_n) = \sigma^{-1}_{F_n,l,2} E_{F_n}[(G(x_l, \theta_n) - YZ)I_l], \tag{D-3}$$

$$= \sigma^{-1}_{F_n,l,2}\left(G_{l,n} - E_{l,n}\pi_{l,n}\right) p_{l,n} \geq -r_{l,2} n^{-1/2}, \tag{D-4}$$

where $\sigma^2_{F_n,l,j} \equiv V_{F_n}(m_{j,l}(W_i, \theta_n))$ for $j = 1, 2$ and $l = 1, \ldots, d_x$ is given by

$$\sigma^2_{F_n,l,1} = p_{l,n}\pi_{l,n}\left[\left(H_{l,n} - E^2_{l,n}\right) + (1 - \pi_{l,n})(1 - E_{l,n})^2\right], \tag{D-5}$$

$$\sigma^2_{F_n,l,2} = p_{l,n}\pi_{l,n}\left[\left(H_{l,n} - E^2_{l,n}\right) + (1 - \pi_{l,n})(E_{l,n} - 2G_{l,n})^2\right]. \tag{D-6}$$

Also, for $l = 1, \ldots, d_x$

$$\rho_{12,l,n} \equiv E_{F_n}\left(m_{1,l}(W_i, \theta_n)\, m_{2,l}(W_i, \theta_n)\right) = E_{F_n}[(YZ - G_{l,n} + 1 - Z)(G_{l,n} - YZ)I_l] \tag{D-7}$$

$$= (1 - \pi_{l,n})p_{l,n}[G_{l,n}(1 - p_{l,n}) + E_{l,n}\pi_{l,n}p_{l,n}] - \sigma^2_{F_n,l,2}. \tag{D-8}$$

This model satisfies the following relationship

$$m_{l,1}(W_i, \theta_n) + m_{l,2}(W_i, \theta_n) = (1 - Z)I_l, \tag{D-9}$$

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

$$\gamma_{1,l,1,n} = \sigma^{-1}_{F_n,l,1}(1 - \pi_{l,n})p_{l,n} - \sigma^{-1}_{F_n,l,1}\sigma_{F_n,l,2}\gamma_{1,l,2,n}. \tag{D-10}$$

D.1.1 On Assumption A.5

We begin with the case $d_x = 1$ and cover the case $d_x > 1$ afterwards. By Definition A.1, $\gamma_{\omega_n,g_1,h}$ denotes a sequence of parameter vectors θ_{ω_n} and distributions F_{ω_n} for W_i such that $\omega_n^{1/2}\gamma_{1,j,\omega_n} \to h_{1,j}$ and $b_{\omega_n}^{1/2}\gamma_{1,j,\omega_n} \to g_{1,j}$ for $j \in \{1,2\}$.

For a given $\gamma_{\omega_n,g_1,h}$ denote by J the set of $j \in \{1,2\}$ that satisfy $h_{1,j} = \infty$ and $g_{1,j} < \infty$. By Assumption D.1, there are constants $0 < B_1 < B_2 < \infty$ such that $\sigma_{F_n,j} \in [B_1, B_2]$ for all $j \in \{1,2\}$ and $n \in \mathbb{N}$, which implies that $E_{F_n} m_j(W_i,\theta_n) = o(1)$ for all $j \in J$. When J is empty, there is nothing to show. We are therefore left with cases (I) $J = \{1\}$, (II) $J = \{2\}$, and (III) $J = \{1,2\}$. We start with the case $J = \{1\}$ and consider two subcases. In Case (a) $h_{1,2} < \infty$ while in Case (b) we have $h_{1,2} = \infty$ and $g_{1,2} = \infty$. To simplify notation we write n rather than $\tilde{\omega}_n$ and b instead of b_n.

Case (I)(a): Since $h_{1,2} < \infty$, it follows by previous arguments that $E_{F_n} m_2(W_i,\theta_n) = o(1)$. By Equation (D-9), $(1-\pi_n) = o(1)$ and $E_n = G_n + o(1)$. It then follows that $\rho_{12,n} \to -1$. Consider an alternative sequence of parameters $\{\theta_n', F_n'\}_{n\geq 1}$ such that $\theta_n' = \theta_n$ (so $G_n' = G_n$), $H_n' = H_n$, π_n' and E_n' given by

$$\pi_n' = \left(1 - (\ln b)b^{-1/2}\right) \to 1, \tag{D-11}$$

$$E_n' = \left(G_n + h_{1,2}\sigma_{F_n,2}n^{-1/2} - (\ln b)b^{-1/2}\right)\left(1-(\ln b)b^{-1/2}\right)^{-1} = E_n + o(1). \tag{D-12}$$

This implies

$$(1-\pi_n') = (\ln b)b^{-1/2} = o(1), \tag{D-13}$$

$$(G_n' - \pi_n' E_n') = -h_{1,2}\sigma_{F_n,2}n^{-1/2} + (\ln b)b^{-1/2} = o(1), \tag{D-14}$$

and $\sigma_{F_n',j}\sigma_{F_n,j}^{-1} = 1 + o(1)$ for $j = 1,2$. As a result

$$b^{1/2}\sigma_{F_n',1}^{-1}E_{F_n'}m_1(W_i,\theta_n') = \sigma_{F_n',1}^{-1}\left(-b^{1/2}(h_{1,2}\sigma_{F_n,1})n^{-1/2} + \ln b\right) \to \infty, \tag{D-15}$$

$$n^{1/2}\sigma_{F_n',2}^{-1}E_{F_n'}m_2(W_i,\theta_n') = \sigma_{F_n',2}^{-1}(h_{1,2}\sigma_{F_n,1}) \to h_{1,2}. \tag{D-16}$$

Finally, by $\pi_n' \to 1$ and Assumption D.1, $\rho_{12,n}' \equiv Corr_{F_n'}(m_1(W_i,\theta_n'), m_2(W_i,\theta_n')) \to -1$.

Case (I)(b): Since $\sigma_{F_n,2} \in [B_1,B_2]$, it follows that $\lim E_{F_n} m_2(W_i,\theta_n) \in [0,\infty]$. In this case $\lim(1-\pi_n) \in [0,1]$. Consider an alternative sequence of parameters $\{\theta_n', F_n'\}_{n\geq 1}$ such that $\theta_n' = \theta_n$ (so $G_n' = G_n$), $H_n' = H_n$, π_n' and E_n' given by

$$\pi_n' = \pi_n - 2(\ln b)b^{-1/2} = \pi_n + o(1), \tag{D-17}$$

$$E_n' = \left(\pi_n E_n - (\ln b)b^{-1/2}\right)\left(\pi_n - 2(\ln b)b^{-1/2}\right)^{-1} = E_n + o(1), \tag{D-18}$$

where we used $\pi_n \geq c_4 > 0$. This implies $\sigma_{F_n',j}\sigma_{F_n,j}^{-1} = 1 + o(1)$ for $j = 1,2$. It then follows that

$$b^{1/2}\sigma_{F_n',1}^{-1}E_{F_n'}m_1(W_i,\theta_n') = b^{1/2}\sigma_{F_n,1}^{-1}E_{F_n}m_1(W_i,\theta_n) + \sigma_{F_n',1}^{-1}\ln b + o(1) \to \infty, \tag{D-19}$$

$$b^{1/2}\sigma_{F_n',2}^{-1}E_{F_n'}m_2(W_i,\theta_n') = b^{1/2}\sigma_{F_n,2}^{-1}E_{F_n}m_2(W_i,\theta_n) + \sigma_{F_n',2}^{-1}\ln b + o(1) \to \infty. \tag{D-20}$$

Finally, Assumption D.1 and Equation (D-8) imply

$$\rho_{12,n}' \equiv Corr_{F_n'}(m_1(W_i,\theta_n'), m_2(W_i,\theta_n')) = \rho_{12,n} + o(1). \tag{D-21}$$

Case (II): This case is analogous to Case (I) and is therefore omitted.

Case (III): By Equation (D-9) $(1-\pi_n) = o(1)$ and $E_n = G_n + o(1)$. As a consequence of this and Assumption D.1, it follows that $\rho_{12,n} \to -1$. Consider an alternative sequence of parameters $\{\theta_n', F_n'\}_{n\geq 1}$ such that $\theta_n' = \theta_n$ (so $G_n' = G_n$), $H_n' = H_n$, π_n' and E_n' given by Equations (D-17) and (D-18). Then, Equations (D-19)-(D-21) follow and this concludes the proof for the case $d_x = 1$.

Now consider the case $d_x > 1$. Notice that in the case with $d_x = 1$, we considered a sequence of parameters $\{\theta_n', F_n'\}_{n\geq 1}$ such that $\theta_n' = \theta_n$ and

$$\sigma_{F_n',1}^{-1} E_{F_n'} m_1 (W_i, \theta_n') = \sigma_{F_n,1}^{-1} E_{F_n} m_1 (W_i, \theta_n) + o(1), \tag{D-22}$$

$$\sigma_{F_n',2}^{-1} E_{F_n'} m_2 (W_i, \theta_n') = \sigma_{F_n,2}^{-1} E_{F_n} m_2 (W_i, \theta_n) + o(1), \tag{D-23}$$

$$\lim Corr_{F_n'} (m_1 (W_i, \theta_n'), m_2 (W_i, \theta_n')) = \lim Corr_{F_n} (m_1 (W_i, \theta_n), m_2 (W_i, \theta_n)). \tag{D-24}$$

When $d_x > 1$, we consider an alternative sequence of parameters $\{\theta_n', F_n'\}_{n\geq 1}$ such that for each $l = 1, \ldots, d_x$, we set $p_{l,n}' = p_{l,n}$, $\theta_n' = \theta_n$ (so $G_{l,n}' = G_{l,n}$), and the rest of the choices of the alternative distribution would be chosen according to the corresponding case in the previous part. According to this, it follows that for every $l = 1, \ldots, d_x$,

$$\sigma_{F_n',1,l}^{-1} E_{F_n'} m_{1,l} (W_i, \theta_n') = \sigma_{F_n,1,l}^{-1} E_{F_n} m_{1,l} (W_i, \theta_n) + o(1), \tag{D-25}$$

$$\sigma_{F_n',2,l}^{-1} E_{F_n'} m_{2,l} (W_i, \theta_n') = \sigma_{F_n,2,l}^{-1} E_{F_n} m_{2,l} (W_i, \theta_n) + o(1), \tag{D-26}$$

$$\lim Corr_{F_n'} (m_{1,l} (W_i, \theta_n'), m_{2,l} (W_i, \theta_n')) = \lim Corr_{F_n} (m_{1,l} (W_i, \theta_n), m_{2,l} (W_i, \theta_n)) \tag{D-27}$$

To conclude the proof, we notice that for $l, j = 1, \ldots, d_x$ with $l \neq j$ and $a, b \in \{1, 2\}$

$$\begin{aligned} Corr_{F_n'} (m_{a,l} (W_i, \theta_n'), m_{b,k} (W_i, \theta_n')) &= -\sigma_{F_n',a,l}^{-1} E_{F_n'} m_{a,l} (W_i, \theta_n') \sigma_{F_n',b,k}^{-1} E_{F_n'} m_{b,k} (W_i, \theta_n') + o(1), \\ &= -\sigma_{F_n,a,l}^{-1} E_{F_n} m_{a,l} (W_i, \theta_n) \sigma_{F_n,b,k}^{-1} E_{F_n} m_{b,k} (W_i, \theta_n) + o(1), \\ &= Corr_{F_n} (m_{a,l} (W_i, \theta_n), m_{b,k} (W_i, \theta_n)) + o(1). \end{aligned} \tag{D-28}$$

This verifies all the desired properties and concludes the proof.

D.1.2 On Assumption A.6

We verify Assumption A.6 for $r^* > 0$. For simplicity consider the case $d_x = 1$. Choose a sequence of parameters $\{\theta_n, F_n\}_{n\geq 1}$ with $1 - \pi_n = o(1)$ and limiting parameter $h_{1,1}^* < 0$. By Equation (D-10), $h_{1,2}^* = -h_{1,1}^* > 0$ and h_2^* is a 2×2 matrix equal to $[1, -1; -1, 1]$.

First, consider the test function S_1. Let $c_0(h_2^*, 1 - \alpha)$ be the $1 - \alpha$ quantile of

$$S_1(Z_{h_2^*}, h_2^*) = [Z_1]_-^2 + [-Z_1]_-^2 = Z_1^2, \quad Z_{h_2^*} = (Z_1, Z_2) \sim N(0, h_2^*). \tag{D-29}$$

Note that

$$S_1(Z_{h_2^*} + h_1^*, h_2^*) = [Z_1 + h_{1,1}^*]_-^2 + [-Z_1 - h_{1,1}^*]_-^2 = (Z_1 + h_{1,1}^*)^2, \tag{D-30}$$

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.

Second, consider the test function S_2. We consider the version of S_2 in Equation (3.14) as here the limit correlation matrix are singular and so we need a test function defined on Ψ_1. Using the definition of $\tilde{\Sigma}_\varepsilon$ in Equation (3.15), it follows that

$$\tilde{\Omega}_\varepsilon^* = \begin{bmatrix} 1+\epsilon & -1 \\ -1 & 1+\epsilon \end{bmatrix} \quad \text{and} \quad \tilde{\Omega}_\varepsilon^{*,-1} = a(\varepsilon) \begin{bmatrix} 1+\epsilon & 1 \\ 1 & 1+\epsilon \end{bmatrix}, \tag{D-31}$$

where $a(\varepsilon) = [(1+\varepsilon)^2 - 1]^{-1}$. As a result

$$\begin{aligned}\tilde{S}_2(Z_{h_2^*}, \tilde{\Omega}_\varepsilon^*) &= (1+\varepsilon)a(\varepsilon) \inf_{t_1 \geq 0, t_2 \geq 0} \{(Z_1 - t_1)^2 + (Z_2 - t_2)^2 + 2(1+\epsilon)^{-1}(Z_1 - t_1)(Z_2 - t_2)\} \\ &= (1+\varepsilon)a(\varepsilon) \inf_{t_1 \geq 0, t_2 \geq 0} \{(Z_1 - t_1)^2 + (Z_1 + t_2)^2 - 2(1+\epsilon)^{-1}(Z_1 - t_1)(Z_1 + t_2)\} \\ &= \frac{1}{1+\varepsilon} Z_1^2, \end{aligned} \tag{D-32}$$

where the first equality holds by definition, the second equality holds because $Z_{h_2^*}$ is such that $Z_2 = -Z_1$, and the third equality by solving the optimization. Since $Z_2 + h_{1,2}^* = -Z_1 - h_{1,1}^*$ we also have

$$\tilde{S}_2(Z_{h_2^*} + h_1^*, \tilde{\Omega}_\varepsilon^*) = \frac{1}{1+\varepsilon}(Z_1 + h_{1,1}^*)^2, \tag{D-33}$$

where $h_{1,1}^* < 0$. Let $c_0(h_2^*, 1-\alpha)$ be the $1-\alpha$ quantile of $\tilde{S}_2(Z, \tilde{\Omega}_\varepsilon^*)$ in Equation (D-32). Since $\Pr((Z_1 + h_{1,1}^*)^2 \leq x) < \Pr(Z_1^2 \leq x)$ for $h_{1,1}^* < 0$, $\Pr((1+\varepsilon)^{-1}(Z_1 + h_{1,1}^*)^2 \leq c_0(h_2^*, 1-\alpha)) < 1-\alpha$. 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. For simplicity of the argument we assume that the distribution G is uniform, as stated below.

Assumption D.2. *Under the distribution G, $u_i = (u_{1,i}, u_{2,i})$ is uniformly distributed on $[0,1]^2$.*

Let $\theta_n = (\theta_{1,n}, \theta_{2,n})$ be the true parameter vector, $\Pr_n(\cdot)$ the probability with respect to the distribution F_n of W_i, $p_{rs,n} \equiv \Pr_n(W_{1,i} = r, W_{2,i} = s)$ for $r, s \in \{0,1\}$, and $\rho_{jj',n} \equiv Corr_{F_n}[m_j(W_i, \theta_n), m_{j'}(W_i, \theta_n)]$ for $j, j' \in \{1,2,3\}$. As defined in the text, G_n denotes the true distribution of u_i for sample size n. Under Assumption D.2 we have

$$\begin{aligned} \gamma_{1,1,n} &\equiv \sigma_{F_n,1}^{-1} E_{F_n}[G_1(\theta_n) - W_{1,i}(1 - W_{2,i})] = \sigma_{F_n,1}^{-1}(\theta_{2,n} - p_{10,n}), \\ \gamma_{1,2,n} &\equiv \sigma_{F_n,2}^{-1} E_{F_n}[W_{1,i}(1 - W_{2,i}) - G_2(\theta_n)] = \sigma_{F_n,2}^{-1}(p_{10,n} - (1 - \theta_{1,n})\theta_{2,n}), \\ \gamma_{1,3,n} &\equiv \sigma_{F_n,3}^{-1} E_{F_n}[W_{1,i}W_{2,i} - G_3(\theta_n)] = \sigma_{F_n,3}^{-1}(p_{11,n} - (1 - \theta_{1,n})(1 - \theta_{2,n})). \end{aligned} \tag{D-34}$$

By simple calculations we have

$$\begin{aligned} &\sigma_{F_n,1}^2 = \sigma_{F_n,2}^2 = Var_{F_n}[m_1(W_i, \theta_n)] = p_{10,n}(1 - p_{10,n}) \in (0, 1/4], \\ &\sigma_{F_n,3}^2 = Var_{F_n}[m_3(W_i, \theta_n)] = p_{11,n}(1 - p_{11,n}) \in (0, 1/4], \\ &\rho_{12,n} = -1, \ \rho_{13,n} = \frac{p_{10,n}p_{11,n}}{\sigma_{F_n,1}\sigma_{F_n,3}}, \text{ and } \rho_{23,n} = -\rho_{13,n}, \end{aligned} \tag{D-35}$$

where zero variances have been ruled out by Definition 2.1(iv). By Definition A.1, $\gamma_{\omega_n, g_1, h}$ denotes a sequence of parameter vectors θ_{ω_n} and distributions F_{ω_n} for W_i such that $\omega_n^{1/2}\gamma_{1,j,\omega_n} \to h_{1,j}$ and $b_{\omega_n}^{1/2}\gamma_{1,j,\omega_n} \to g_{1,j}$ for $j \in \{1,2,3\}$. Recall that $\gamma_{\omega_n, g_1, h}$ defines $\theta_{\omega_n} = (\theta_{1,\omega_n}, \theta_{2,\omega_n})$ and thus defines $G_1(\theta_{\omega_n}), G_2(\theta_{\omega_n})$, and $G_3(\theta_{\omega_n})$.

D.2.1 On Assumption A.5

For a given $\gamma_{\omega_n, g_1, h}$ denote by J the set of $j \in \{1,2\}$ that satisfy $h_{1,j} = \infty$ and $g_{1,j} < \infty$. When J is empty, there is nothing to show. We are therefore left with the cases $J = \{1\}$, $J = \{2\}$, and $J = \{1,2\}$. We start with the case $J = \{1\}$ and consider two subcases. In Case (I) $h_{1,2} < \infty$ while in Case (II) we have $h_{1,2} = \infty$ and $g_{1,2} = \infty$. For each subcase we consider two further subcases: In Case (a) $\rho_{13,n} \to 0$ while in Case (b) $\rho_{13,n} \to \rho_{13} \in (0,1]$. To simplify notation we write n rather than $\tilde{\omega}_n$ and b instead of b_n.

Remark D.1. Note that for any positive numbers a_{10}, a_{01}, a_{11} whose sum equals 1 and $\theta = (\theta_1, \theta_2) \in (0,1)^2$, there exists a random variable $u_i = (u_{1,i}, u_{2,i})$ on $[0,1]^2$ with continuous distribution such that $\Pr(u_{1,i} > \theta_1 \ \& \ u_{2,i} < \theta_2) = a_{10}$, $\Pr(u_{1,i} < \theta_1 \ \& \ u_{2,i} > \theta_2) = a_{01}$, and $\Pr(u_{1,i} > \theta_1 \ \& \ u_{2,i} > \theta_2) = a_{11}$ (and consequently $\Pr(u_{1,i} < \theta_1 \ \& \ u_{2,i} < \theta_2) = 0$).

Letting a_{10}, a_{01}, a_{11} play the role of $p_{10,n}$, $p_{01,n}$, $p_{11,n}$ the Remark D.1 implies that for a given vector $\theta_n = (\theta_{1,n}, \theta_{2,n})$ any desired outcome probabilities $p_{10,n}$, $p_{01,n}$, $p_{11,n}$ can be generated by a random variable $u_i = (u_{1,i}, u_{2,i})$ that has a continuous distribution G_n.

Case (I)(a): We have to produce a sequence $\gamma_{\tilde{\omega}_n, \tilde{g}_1, h}$ for which $\tilde{g}_{1,1} = \infty$, $h_{1,2}$ is a specific finite number, and the upper right element of h_2 equals 0. Define

$$p'_{10,n} = b^{-3/7}. \quad \text{(D-36)}$$

Let $\theta'_n = (\theta'_{1,n}, \theta'_{2,n})$ for $\theta'_{1,n}$ and $\theta'_{2,n}$ defined next. Pick $\theta'_{2,n} \in (0,1)$ such that

$$G_1(\theta'_n) = \theta'_{2,n} = p'_{10,n} + b^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} b^{2/7} \quad \text{(D-37)}$$

and pick $\theta'_{1,n} \in (0,1)$ such that

$$G_2(\theta'_n) = (1 - \theta'_{1,n})\theta'_{2,n} = p'_{10,n} - n^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} h_{1,2}. \quad \text{(D-38)}$$

This is clearly possible because $p'_{10,n} \to 0$, $|n^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} h_{1,2}| < p'_{10,n}$, and $b^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} b^{2/7} \to 0$. We have $G_1(\theta'_n) = \theta'_{2,n} = 2b^{-3/7}(1 + o(1))$, $G_2(\theta'_n) = b^{-3/7}(1 + o(1))$. Now

$$b^{-3/7}(1 + o(1)) = G_2(\theta'_n) = (1 - \theta'_{1,n})\theta'_{2,n} = 2(1 - \theta'_{1,n}) b^{-3/7}(1 + o(1)) \quad \text{(D-39)}$$

which implies that $\theta'_{1,n}$ cannot converge to 1. Without loss of generality we can therefore assume that $\theta'_{1,n} \to \theta'_1$ for some $\theta'_1 \in [0,1)$. We then have

$$G_3(\theta'_n) = (1 - \theta'_{1,n})(1 - \theta'_{2,n}) \to (1 - \theta'_1). \quad \text{(D-40)}$$

Consider the function

$$f(x) \equiv x - h_{1,3} n^{-1/2}(x(1 - x))^{1/2} \quad \text{(D-41)}$$

for $x \in [0,1]$. The function f is continuous and satisfies $f(0) = 0$ and $f(1) = 1$. Therefore, for given $G_3(\theta'_n)$, the intermediate value theorem implies there exists a value $p'_{11,n}$ such that

$$G_3(\theta'_n) = p'_{11,n} - h_{1,3} n^{-1/2}(p'_{11,n}(1 - p'_{11,n}))^{1/2}. \quad \text{(D-42)}$$

Define $p'_{11,n}$ to be any value in $(0,1)$ that satisfies Equation (D-42). It cannot be the case that $p'_{11,n} \to 1$ as otherwise we would have $G_3(\theta'_n) \to 1$ contradicting Equation (D-40). Therefore, without loss of generality $p'_{11,n} \to p'_{11}$ for some $p'_{11} \in [0,1)$. Note that $p'_{10,n} \to 0$ and $p'_{11} \in [0,1)$ imply that

$$\rho'_{13,n} = \frac{p'_{10,n} p'_{11,n}}{\sigma'_{F_n,1} \sigma'_{F_n,3}} \to 0. \quad \text{(D-43)}$$

For these given choices of $p'_{10,n}$, $p'_{11,n}$ and $\theta'_n = (\theta'_{1,n}, \theta'_{2,n})$, and by Remark D.1, there exists a continuous distribution G'_n for the random variable $u_i = (u_{1,i}, u_{2,i})$ such that $\Pr(u_{1,i} > \theta'_{1,n} \ \& \ u_{2,i} < \theta'_{2,n}) = p'_{10,n}$, $\Pr(u_{1,i} < \theta'_{1,n} \ \& \ u_{2,i} > \theta'_{2,n}) = 1 - p'_{10,n} - p'_{11,n}$, and $\Pr(u_{1,i} > \theta'_{1,n} \ \& \ u_{2,i} > \theta'_{2,n}) = p'_{11,n}$. By construction all requirements are fulfilled under the sequence θ'_n and G'_n.

Case (I)(b): We have to produce a sequence $\gamma_{\tilde{\omega}_n, \tilde{g}_1, h}$ for which $\tilde{g}_{1,1} = \infty$, $h_{1,2} < \infty$ and the upper right element of h_2 equals $\rho_{13} \in (0,1]$. Assume first that $\rho_{13} \in (0,1)$. Define

$$p'_{10,n} = cb^{-1/7} \quad \text{(D-44)}$$

for some constant $c > 0$, and define $(\theta'_{1,n}, \theta'_{2,n}) \in (0,1)^2$ as in Equations (D-37) and (D-38). We then have $\theta'_{2,n} = cb^{-1/7} + c^{1/2}b^{-2/7}(1 + o(1))$ and $(1 - \theta'_{1,n})\theta'_{2,n} = cb^{-1/7} + o(b^{-2/7})$ and thus

$$\theta'_{1,n} = 1 - \frac{cb^{-1/7} + o(b^{-2/7})}{cb^{-1/7} + c^{1/2}b^{-2/7}(1 + o(1))} = c^{-1/2}b^{-1/7}(1 + o(1)). \tag{D-45}$$

Next

$$\begin{aligned} G_3(\theta'_n) &= (1 - \theta'_{1,n})(1 - \theta'_{2,n}) = (1 - c^{-1/2}b^{-1/7}(1 + o(1)))(1 - cb^{-1/7}(1 + o(1))) \\ &= 1 - (c^{-1/2} + c)b^{-1/7}(1 + o(1)). \end{aligned} \tag{D-46}$$

Arguing as in Case (I)(a) there is a value $p'_{11,n} \in (0,1)$ such that

$$G_3(\theta'_n) = p'_{11,n} - h_{1,3}n^{-1/2}(p'_{11,n}(1 - p'_{11,n}))^{1/2}. \tag{D-47}$$

As $G_3(\theta'_n) \to 1$ we have $p'_{11,n} \to 1$. More precisely,

$$p'_{11,n} = G_3(\theta'_n) + h_{1,3}n^{-1/2}(p'_{11,n}(1 - p'_{11,n}))^{1/2} = 1 - (c^{-1/2} + c)b^{-1/7}(1 + o(1)). \tag{D-48}$$

Therefore,

$$\begin{aligned} \rho'_{13,n} \equiv \left(\frac{p'_{10,n}p'_{11,n}}{(1 - p'_{10,n})(1 - p'_{11,n})} \right)^{1/2} &= \left(\frac{cb^{-1/7}(1 - (c^{-1/2} + c)b^{-1/7})}{(1 - cb^{-1/7})((c^{-1/2} + c)b^{-1/7})} \right)^{1/2} (1 + o(1)) \\ \to (c/(c^{-1/2} + c))^{1/2}. \end{aligned} \tag{D-49}$$

The function $(c/(c^{-1/2} + c))^{1/2}$ is continuous for $c > 0$ and converges to 1 as $c \to \infty$ and to 0 as $c \to 0$. There is therefore $c > 0$ such that $(c/(c^{-1/2} + c))^{1/2} = \rho_{13}$. The proof is then concluded as in Case (I)(a). If $\rho_{13} = 1$, the same proof applies once the constant c in Equation (D-44) is replaced by the sequence $c_n = \ln b$ that slowly converges to infinity.

Case (II)(a): We have to produce a sequence $\gamma_{\tilde{\omega}_n, \tilde{g}_1, h}$ for which $\tilde{g}_{1,1} = \tilde{g}_{1,2} = \infty$ and the upper right element of h_2 equals zero. Define

$$p'_{10,n} = b^{-3/7}. \tag{D-50}$$

Let $\theta'_n = (\theta'_{1,n}, \theta'_{2,n}) \in (0,1)^2$ be defined as follows. Let $\theta'_{2,n} \in (0,1)$ be such that

$$G_1(\theta'_n) = \theta'_{2,n} = p'_{10,n} + b^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2}b^{2/7} \tag{D-51}$$

and pick $\theta'_{1,n} \in (0,1)$ such that

$$G_2(\theta'_n) = (1 - \theta'_{1,n})\theta'_{2,n} = p'_{10,n} - b^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2}b^{1/7}. \tag{D-52}$$

As in Case (I)(a) we have $G_1(\theta'_n) = 2b^{-3/7}(1+o(1))$ and $G_2(\theta'_n) = b^{-3/7}(1+o(1))$. Using the same steps as in Case (I)(a) we have $\theta'_{1,n} \to \theta'_1$ for some $\theta'_1 \in [0,1)$ and thus that $G_3(\theta'_n)$ converges to a number smaller than 1. Then again, there exists $p'_{11,n}$ such that $G_3(\theta'_n) = p'_{11,n} - h_{1,3}n^{-1/2}(p'_{11,n}(1-p'_{11,n}))^{1/2}$ and $p'_{11,n} \to p'_{11}$ for some $p'_{11} \in [0,1)$. Therefore, we have again that $\rho'_{13,n} \to 0$ and the proof concludes as in Case (I)(i).

Case (II)(b): We have to produce a sequence $\gamma_{\tilde{\omega}_n, \tilde{g}_1, h}$ for which $\tilde{g}_{1,1} = \tilde{g}_{1,2} = \infty$ and the upper right element of h_2 equals $\rho_{13} \in (0,1]$. The proof follows along the same lines as Case (I)(b) with the one difference that $G_2(\theta'_n)$ is defined as in Equation (D-52).

That concludes the verification of the assumption for the case $J = \{1\}$. Regarding the other cases, note that the case $J = \{1, 2\}$ is covered by Cases (II)(a) and (II)(b) above. The verification of the assumption in case $J = \{2\}$ is also partially covered by Cases (II)(a) and (II)(b) and the remaining cases for $J = \{2\}$ are similar to Cases (I)(a) and (I)(b) above for $J = \{1\}$ and therefore omitted.

D.2.2 On Assumption A.6

The verification of Assumption A.6 follows almost identical steps to those used in Section D.1.2 and is therefore omitted.

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(3), 669–709.

——— (2010a): "Applications of Hybrid and Size-Corrected Subsampling Methods," *Journal of Econometrics*, 158(2), 285–305.

——— (2010b): "Asymptotic Size and a Problem with Subsampling and with the m Out of n Bootstrap," *Econometric Theory*, 26(2), 426–468.

ANDREWS, D. W. K., AND P. JIA (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," manuscript, Yale University.

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–158.

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2008): "Set Identified Linear Models," manuscript, University of Toulouse.

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(2), 408–425.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

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

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

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

——— (2009b): "Inference in Models with Multiple Equilibria," manuscript, University of Montreal.

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.

Marsaglia, G., and I. Olkin (1984): "Generating Correlation Matrices," *SIAM Journal on Scientific Computing*, 5(2), 470–475.

Moon, H. R., and F. Schorfheide (2009): "Bayesian and Frequentist Inference in Partially Identified Models," manuscript, University of Pennsylvania.

Newey, W. K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29(3), 229–256.

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

Politis, D. N., and J. P. Romano (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22(4), 2031–2050.

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

Ponomareva, M., and E. Tamer (2010): "Misspecification in Moment Inequality Models: Back to Moment Equalities?," forthcoming in *The Econometrics Journal*.

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.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–212.

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

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77(4), 1299–1315.

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