Quantitative Economics 8 (2017), 1–38 1759-7331/20170001

Inference for subvectors and other functions of partially identified parameters in moment inequality models

Federico A. Bugni
Department of Economics, Duke University

Ivan A. Canay
Department of Economics, Northwestern University

Xiaoxia Shi
Department of Economics, University of Wisconsin, Madison

This paper introduces a bootstrap-based inference method for functions of the parameter vector in a moment (in)equality model. These functions are restricted to be linear for two-sided testing problems, but may be nonlinear for one-sided testing problems. In the most common case, this function selects a subvector of the parameter, such as a single component. The new inference method we propose controls asymptotic size uniformly over a large class of data distributions and improves upon the two existing methods that deliver uniform size control for this type of problem: projection-based and subsampling inference. Relative to projection-based procedures, our method presents three advantages: (i) it weakly dominates in terms of finite sample power, (ii) it strictly dominates in terms of asymptotic power, and (iii) it is typically less computationally demanding. Relative to subsampling, our method presents two advantages: (i) it strictly dominates in terms of asymptotic power (for reasonable choices of subsample size), and (ii) it appears to be less sensitive to the choice of its tuning parameter than subsampling is to the choice of subsample size.

Keywords. Partial identification, moment inequalities, subvector inference, hypothesis testing.

JEL classification. C01, C12, C15.

Federico A. Bugni: federico.bugni@duke.edu
Ivan A. Canay: iacanay@northwestern.edu
Xiaoxia Shi: xshi@ssc.wisc.edu

We thank the editor and two anonymous referees whose valuable suggestions helped greatly to improve the paper. We also thank Francesca Molinari, Elie Tamer, Azeem Shaikh, the participants at the Workshop on Inference in Nonstandard Problems at Cornell University, the Workshop on Mathematical Statistics of Partially Identified at Oberwolfach, the Conference on Partial Identification, Weak Identification, and Related Econometric Problems at Cowles Fundation, the First Conference in Econometric Theory at Universidad de San Andrés, Argentina, the Workshop on Recent Developments in Partial Identification: Theory and Practice at Toulouse, the UCLA mini conference in Econometrics, and seminar participants at Brown, Northwestern, University of Maryland, Korea University, University of British Columbia, NYU, and Simon Fraser University for helpful comments. Bugni and Canay thank the National Science Foundation for research support via Grants SES-1123771 and SES-1123586, respectively. Takuya Ura and Zhonglin Li provided excellent research assistance. Any and all errors are our own.

DOI: 10.3982/QE490

1. Introduction

In recent years, substantial interest has been drawn to partially identified models defined by moment (in)equalities of the generic form

$$\begin{aligned} E_F\left[m_j(W_i,\theta)\right] &\geq 0 \quad \text{for } j=1,\dots,p, \\ E_F\left[m_j(W_i,\theta)\right] &= 0 \quad \text{for } j=p+1,\dots,k, \end{aligned} \tag{1.1}$$

where $\{W_i\}_{i=1}^n$ is an independent and identically distributed (i.i.d.) sequence of random variables with distribution F and $m=(m_1,\dots,m_k)':\mathbb{R}^d\times\Theta\to\mathbb{R}^k$ is a known measurable function of the finite dimensional parameter vector $\theta\in\Theta\subseteq\mathbb{R}^{d_\theta}$. Methods to conduct inference on θ have been proposed, for example, by Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010).[1] As a common feature, these papers construct *joint* confidence sets (CS's) for the vector θ by inverting hypothesis tests for $H_0:\theta=\theta_0$ versus $H_1:\theta\neq\theta_0$. However, in empirical work, researchers often report *marginal* confidence intervals for each coordinate of θ, either to follow the tradition of standard t-test-based inference or because only few individual coordinates of θ are of interest. The current practice appears to be reporting projections of the joint CS's for the vector θ, for example, Ciliberto and Tamer (2009) and Grieco (2014).

Although convenient, projecting joint CS's suffers from three problems. First, when interest lies in individual components of θ, projection methods are typically conservative (even asymptotically). This may lead to confidence intervals that are unnecessarily wide, a problem that gets exacerbated when the dimension of θ becomes reasonably large. Second, the projected confidence intervals do not necessarily inherit the good asymptotic power properties of the joint CS's. Yet, the available results in the literature are mostly limited to asymptotic properties of joint CS's. Finally, computing the projections of a joint CS is typically unnecessarily burdensome if the researcher is only interested in individual components. This is because one needs to compute the joint CS first, which itself requires searching over a potentially large dimensional space Θ for all the points not rejected by a hypothesis test.

In this paper, we address the practical need for marginal CS's by proposing a test to conduct inference directly on individual coordinates or, more generally, on a function $\lambda:\Theta\to\mathbb{R}^{d_\lambda}$ of the parameter vector θ. The function $\lambda(\cdot)$ is assumed to be linear in two-sided testing problems like (1.2) below, but may be quasiconvex in one-sided testing problems like those described in Remark 4.2. We then construct a CS for $\lambda(\theta)$ by inverting tests for the hypotheses

$$H_0:\lambda(\theta)=\lambda_0 \quad \text{vs.} \quad H_1:\lambda(\theta)\neq\lambda_0, \tag{1.2}$$

given a hypothetical value $\lambda_0\in\mathbb{R}^{d_\lambda}$. Our test controls asymptotic size uniformly over a large class of data distributions (Theorem 4.1) and has several attractive properties for

[1] Additional references include Imbens and Manski (2004), Beresteanu and Molinari (2008), Rosen (2008), Stoye (2009), Bugni (2010), Canay (2010), Romano and Shaikh (2010), Galichon and Henry (2011), Pakes, Porter, Ho, and Ishii (2015), Bontemps, Magnac, and Maurin (2012), Bugni, Canay, and Guggenberger (2012), and Romano, Shaikh, and Wolf (2014), among others.

practitioners: (i) it has finite sample power that weakly dominates that of projection-based tests for all alternative hypotheses (Theorem 4.2), (ii) it has asymptotic power that strictly dominates that of projection-based tests under reasonable assumptions (see Remark 4.6), and (iii) it is less computationally demanding than projection-based tests whenever the function $\lambda(\cdot)$ introduces dimension reduction (i.e., $d_\lambda \ll d_\theta$). In addition, one corollary of our analysis is that our marginal CS's are always a subset of those constructed by projecting joint CS's (see Remark 4.5).

The test we propose in this paper employs a profiled test statistic similar to the one suggested by Romano and Shaikh (2008) for testing the hypotheses in (1.2) via subsampling. However, our analysis of the testing problem in (1.2) and the properties of our test goes well beyond that in Romano and Shaikh (2008). First, one of our technical contributions is the derivation of the limit distribution of this profiled test statistic, which is an important step toward proving the validity of our bootstrap-based test. This is in contrast to Romano and Shaikh (2008, Theorem 3.4), as their result follows from a high-level condition regarding the relationship between the distribution of size n and that of size b_n (the subsample size), and thus avoids the need for a characterization of the limiting distribution of the profiled test statistic. Second, as opposed to Romano and Shaikh (2008), we present formal results on the properties of our test relative to projection-based inference. Third, we derive the following results that support our bootstrap-based inference over the subsampling inference in Romano and Shaikh (2008): (i) we show that our test is no less asymptotically powerful than the subsampling test under reasonable assumptions (see Theorem 4.3); (ii) we formalize the conditions under which our test has strictly higher asymptotic power (see Remark 4.9); and (iii) we note that our test appears to be less sensitive to the choice of its tuning parameter κ_n than subsampling is to the choice of subsample size (see Remark 4.10). All these results are in addition to the well known challenges behind subsampling inference that make some applied researchers reluctant to use it when other alternatives are available. In particular, subsampling inference is known to be very sensitive to the choice of subsample size, and even when the subsample size is chosen to minimize the error in the coverage probability, it is more imprecise than its bootstrap alternatives; see Politis and Romano (1994), Bugni (2010, 2016).

As previously mentioned, the asymptotic results in this paper hold uniformly over a large class of nuisance parameters. In particular, the test we propose controls asymptotic size over a large class of distributions F and can be inverted to construct uniformly valid CS's (see Remark 4.5). This represents an important difference with other methods that could also be used for inference on components of θ, such as Pakes, Porter, Ho, and Ishii (2015), Chen, Tamer, and Torgovitsky (2011), Kline and Tamer (2013), and Wan (2013). The test proposed by Pakes, Porter, Ho, and Ishii (2015) is, by construction, a test for each coordinate of the parameter θ. However, such test controls size over a much smaller class of distributions than the one we consider in this paper (cf. Andrews and Han (2009)). The approach recently introduced by Chen, Tamer, and Torgovitsky (2011) is especially useful for parametric models with unknown functions, which do not correspond exactly with the model in (1.1). In addition, the asymptotic results in that paper hold pointwise and so it is unclear whether it controls asymptotic size over the same class of distributions we consider. The method in Kline and Tamer (2013) is Bayesian in nature, requires either the function $m(W_i, \theta)$ to be separable (in W_i and θ) or the data to be discretely supported, and focuses on inference about the identified set as opposed to identifiable parameters. Finally, Wan (2013) introduces a computationally attractive inference method based on Markov chain Monte Carlo (MCMC) methods, but derives pointwise asymptotic results. Due to these reasons, we do not devote special attention to these papers.

We view our test as an attractive alternative for practitioners and so we start by presenting a step by step algorithm to implement our test in Section 2. We then present a simple example in Section 3 that illustrates why a straight application of the generalized moment selection (GMS) approach to the hypotheses in (1.2) does not deliver a valid test in general. The example also gives insight on why the test we propose does not suffer from similar problems. Section 4 presents all formal results on asymptotic size and power, while Section 5 presents Monte Carlo simulations that support all our theoretical findings. Proofs are provided in Appendix C and the Supplemental Material, available as a supplementary file on the journal website, http://qeconomics.org/supp/490/supplement.pdf. Replication files are posted as a supplementary file on the journal website, http://qeconomics.org/supp/490/code_and_data.zip.

2. Implementing the minimum resampling test

The minimum resampling test (Test MR) we propose in this paper rejects the null hypothesis in (1.2) for large values of a profiled test statistic, denoted by $T_n(\lambda_0)$. Specifically, it takes the form

$$\phi_n^{\mathrm{MR}}(\lambda_0) \equiv 1\{T_n(\lambda_0) > \hat{c}_n^{\mathrm{MR}}(\lambda_0, 1-\alpha)\}, \tag{2.1}$$

where $1\{\cdot\}$ denotes the indicator function, $\alpha \in (0,1)$ is the significance level, and $\hat{c}_n^{\mathrm{MR}}(\lambda_0, 1-\alpha)$ is the minimum resampling critical value that we formalize below. So as to describe how to implement this test, we need to introduce some notation. To this end, we define

$$\bar{m}_{n,j}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} m_j(W_i, \theta), \tag{2.2}$$

$$\hat{\sigma}_{n,j}^2(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\big(m_j(W_i, \theta) - \bar{m}_{n,j}(\theta)\big)^2 \tag{2.3}$$

for $j = 1, \dots, k$ to be the sample mean and sample variance of the moment functions in (1.1). Denote by

$$\Theta(\lambda_0) = \{\theta \in \Theta : \lambda(\theta) = \lambda_0\} \tag{2.4}$$

the subset of elements in Θ satisfying the null hypothesis in (1.2). Given this set, the profiled test statistic is

$$T_n(\lambda_0) = \inf_{\theta \in \Theta(\lambda_0)} Q_n(\theta), \tag{2.5}$$

where

$$Q_n(\theta)=\left\{\sum_{j=1}^{p}\left[\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}\right]_-^2+\sum_{j=p+1}^{k}\left(\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}\right)^2\right\} \tag{2.6}$$

and $[x]_-\equiv\min\{x,0\}$. The function $Q_n(\theta)$ in (2.6) is the so-called modified method of moments (MMM) test statistic and it is frequently used in the construction of joint CS's for θ. The results we present in Section 4 hold for a large class of possible test statistics, but to keep the exposition simple we use the MMM test statistic throughout this section and in all examples. See Section 4 for other test statistics.

We now describe the minimum resampling critical value, $\hat{c}_n^{\mathrm{MR}}(\lambda_0,1-\alpha)$. This critical value requires two approximations to the distribution of the profiled test statistic $T_n(\lambda_0)$ that share the common structure

$$\inf_{\theta\in\tilde{\Theta}}\left\{\sum_{j=1}^{p}\left[v_{n,j}^*(\theta)+\ell_j(\theta)\right]_-^2+\sum_{j=p+1}^{k}\left(v_{n,j}^*(\theta)+\ell_j(\theta)\right)^2\right\} \tag{2.7}$$

for a given set $\tilde{\Theta}$, stochastic process $v_{n,j}^*(\theta)$, and slackness function $\ell_j(\theta)$. Both approximations use the same stochastic process $v_{n,j}^*(\theta)$, which is defined as

$$v_{n,j}^*(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\left(m_j(W_i,\theta)-\bar{m}_{n,j}(\theta)\right)}{\hat{\sigma}_{n,j}(\theta)}\zeta_i \tag{2.8}$$

for $j=1,\dots,k$, where $\{\zeta_i\sim N(0,1)\}_{i=1}^n$ is i.i.d. and independent of the data. However, they differ in the set $\tilde{\Theta}$ and the slackness function $\ell_j(\theta)$ they use.

The first approximation to the distribution of $T_n(\lambda_0)$ is

$$T_n^{\mathrm{DR}}(\lambda_0)\equiv\inf_{\theta\in\hat{\Theta}_I(\lambda_0)}\left\{\sum_{j=1}^{p}\left[v_{n,j}^*(\theta)+\varphi_j(\theta)\right]_-^2+\sum_{j=p+1}^{k}\left(v_{n,j}^*(\theta)+\varphi_j(\theta)\right)^2\right\}, \tag{2.9}$$

where

$$\hat{\Theta}_I(\lambda_0)\equiv\left\{\theta\in\Theta(\lambda_0):Q_n(\theta)\le T_n(\lambda_0)\right\} \tag{2.10}$$

and

$$\varphi_j(\theta)=\begin{cases}\infty & \text{if } \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)>1 \text{ and } j\le p,\\ 0 & \text{if } \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)\le 1 \text{ or } j>p.\end{cases} \tag{2.11}$$

The set $\hat{\Theta}_I(\lambda_0)$ is the set of minimizers of the original test statistic $T_n(\lambda_0)$ in (2.5). For our method to work, it is enough for this set to be an approximation to the set of minimizers in the sense discussed in Remark 4.1. The function $\varphi_j(\theta)$ in (2.11) is one of the GMS functions in Andrews and Soares (2010). This function uses the information in the sequence

$$\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta) \tag{2.12}$$

for $j=1,\dots,k$ to determine whether the jth moment is binding or slack in the sample. Here κ_n is a tuning parameter that satisfies $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$ (e.g., $\kappa_n = \sqrt{\ln n}$). Although the results in Section 4 hold for a large class of GMS functions, we restrict our discussion here to the function in (2.11) for simplicity.

The second approximation to the distribution of $T_n(\lambda_0)$ is

$$T_n^{\mathrm{PR}}(\lambda_0) \equiv \inf_{\theta\in\Theta(\lambda_0)} \left\{ \sum_{j=1}^{p} \left[v_{n,j}^*(\theta) + \ell_j(\theta) \right]_-^2 + \sum_{j=p+1}^{k} \left(v_{n,j}^*(\theta) + \ell_j(\theta) \right)^2 \right\}, \tag{2.13}$$

with

$$\ell_j(\theta) = \kappa_n^{-1} \sqrt{n} \hat{\sigma}_{n,j}^{-1}(\theta) \bar{m}_{n,j}(\theta) \tag{2.14}$$

for $j=1,\dots,k$. This approximation employs the set $\Theta(\lambda_0)$ and a slackness function $\ell_j(\theta)$ that is not in the class of GMS functions. The reason why $\ell_j(\theta)$ is not a GMS function in Andrews and Soares (2010) is twofold: (i) it can take negative values (while $\varphi_j(\theta) \geq 0$) and (ii) it penalizes moment equalities (while $\varphi_j(\theta) = 0$ for $j=p+1,\dots,k$).

In the context of the common structure in (2.7), the first approximation sets $\tilde{\Theta} = \hat{\Theta}_I(\lambda_0)$ and $\ell_j(\theta) = \varphi_j(\theta)$ for $j=1,\dots,k$, while the second approximation sets $\tilde{\Theta} = \Theta(\lambda_0)$ and $\ell_j(\theta) = \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$ for $j=1,\dots,k$. Given these two approximations, the minimum resampling critical value $\hat{c}_n(\lambda_0, 1-\alpha)$ is defined to be the (conditional) $1-\alpha$ quantile of

$$T_n^{\mathrm{MR}}(\lambda_0) \equiv \min\left\{ T_n^{\mathrm{DR}}(\lambda_0), T_n^{\mathrm{PR}}(\lambda_0) \right\}, \tag{2.15}$$

where $T_n^{\mathrm{DR}}(\lambda_0)$ and $T_n^{\mathrm{PR}}(\lambda_0)$ are as in (2.9) and (2.13), respectively. Algorithm 1 below summarizes in a succinct way the steps required to implement Test MR, that is, $\phi_n^{\mathrm{MR}}(\lambda_0)$ in (2.1).

Remark 2.1. Two aspects about Algorithm 1 are worth emphasizing. First, note that in line 3 a matrix $n \times B$ of independent $N(0,1)$ is simulated and the same matrix is used to compute $T_n^{\mathrm{DR}}(\lambda_0)$ and $T_n^{\mathrm{PR}}(\lambda_0)$ (lines 23 and 24). Here B denotes the number of bootstrap replications. Second, the algorithm involves $2B+1$ optimization problems (lines 20, 23, and 24) that can be implemented via optimization packages available in standard computer programs. This is typically faster than projecting a joint confidence set for θ, which requires computing a test statistic and approximating a quantile for each $\theta \in \Theta$.

Remark 2.2. The leading application of our inference method is the construction of marginal CS's for coordinates of θ, which is done by setting $\lambda(\theta) = \theta_s$ for some $s \in \{1,\dots,d_\theta\}$ in (1.2) and collecting all values of λ_0 for which H_0 is not rejected. For this case, the set $\Theta(\lambda_0)$ in (2.4) becomes

$$\Theta(\lambda_0) = \{\theta \in \Theta : \theta_s = \lambda_0\}. \tag{2.16}$$

That is, optimizing over $\Theta(\lambda_0)$ is equivalent to optimizing over the $d_\theta - 1$ dimensional subspace of Θ that includes all except the sth coordinate.

Algorithm 1 Algorithm to implement the minimum resampling test.

Inputs: λ_0, Θ, κ_n, B, $\lambda(\cdot)$, $\varphi(\cdot)$, $m(\cdot)$, α $\quad\triangleright$ $\kappa_n = \sqrt{\ln n}$ recommended by Andrews and Soares (2010)
$\Theta(\lambda_0) \leftarrow \{\theta \in \Theta : \lambda(\theta) = \lambda_0\}$
$\zeta \leftarrow n \times B$ matrix of independent $N(0, 1)$

function Qstat(type, θ, $\{W_i\}_{i=1}^n$, $\{\zeta_i\}_{i=1}^n$) $\quad\triangleright$ Computes criterion function for a given θ
  $\bar{m}_n(\theta) \leftarrow n^{-1}\sum_{i=1}^n m(W_i, \theta)$ $\quad\triangleright$ Moments for a given θ
  $\hat{D}_n(\theta) \leftarrow \mathrm{Diag}(\mathrm{var}(m(W_i, \theta)))$ $\quad\triangleright$ Variance matrix for a given θ
  if type $= 0$ **then** $\quad\triangleright$ Type 0 is for Test Statistic
    $v(\theta) \leftarrow \sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta)$
    $\ell(\theta) \leftarrow \mathbf{0}_{k\times 1}$ $\quad\triangleright$ Test Statistic does not involve ℓ
  else if type $= 1$ **then** $\quad\triangleright$ Type 1 is for $T_n^{\mathrm{DR}}(\lambda)$
    $v(\theta) \leftarrow n^{-1/2}\hat{D}_n^{-1/2}(\theta)\sum_{i=1}^n (m(W_i, \theta) - \bar{m}_n(\theta))\zeta_i$
    $\ell(\theta) \leftarrow \varphi(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta))$
  else if type $= 2$ **then** $\quad\triangleright$ Type 2 is for $T_n^{\mathrm{PR}}(\lambda)$
    $v(\theta) \leftarrow n^{-1/2}\hat{D}_n^{-1/2}(\theta)\sum_{i=1}^n (m(W_i, \theta) - \bar{m}_n(\theta))\zeta_i$
    $\ell(\theta) \leftarrow \kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta)$
  end if
  return $Q(\theta) \leftarrow \{\sum_{j=1}^p [v_j(\theta) + \ell_j(\theta)]_-^2 + \sum_{j=p+1}^k (v_j(\theta) + \ell_j(\theta))^2\}$
end function

function TestMR(B, $\{W_i\}_{i=1}^n$, ζ, $\Theta(\lambda_0)$, α) $\quad\triangleright$ Test MR
  $T_n \leftarrow \min_{\theta\in\Theta(\lambda_0)}$ Qstat$(0, \theta, \{W_i\}_{i=1}^n)$ $\quad\triangleright$ Compute test statistic
  $\hat{\Theta}_I(\lambda_0) \leftarrow \{\theta \in \Theta(\lambda_0) : \text{Qstat}(0, \theta, \{W_i\}_{i=1}^n) \le T_n\}$ $\quad\triangleright$ Estimated set of minimizers
  for $b = 1, \dots, B$ **do**
    $T^{\mathrm{DR}}[b] \leftarrow \min_{\theta\in\hat{\Theta}_I(\lambda_0)}$ Qstat$(1, \theta, \{W_i\}_{i=1}^n, \zeta[:, b])$ $\quad\triangleright$ type $= 1$. Uses bth column of ζ
    $T^{\mathrm{PR}}[b] \leftarrow \min_{\theta\in\Theta(\lambda_0)}$ Qstat$(2, \theta, \{W_i\}_{i=1}^n, \zeta[:, b])$ $\quad\triangleright$ type $= 2$. Uses bth column of ζ
    $T^{\mathrm{MR}}[b] \leftarrow \min\{T^{\mathrm{DR}}[b], T^{\mathrm{PR}}[b]\}$
  end for
  $\hat{c}_n^{\mathrm{MR}} \leftarrow$ quantile$(T^{\mathrm{MR}}, 1 - \alpha)$ $\quad\triangleright$ T^{MR} is $B \times 1$. Gets $1 - \alpha$ quantile
  return $\phi^{\mathrm{MR}} \leftarrow 1\{T_n > \hat{c}_n^{\mathrm{MR}}\}$
end function

3. Failure of naïve GMS and intuition for Test MR

Before we present the formal results on size and power for Test MR, we address two natural questions that may arise from Section 2. The first one is, “Why not simply use a straight GMS approximation to the distribution of $T_n(\lambda_0)$ in (2.5)?” We call this approach the naïve GMS approximation and denote it by

$$T_n^{\text{naive}}(\lambda_0) \equiv \inf_{\theta\in\Theta(\lambda_0)} \left\{ \sum_{j=1}^{p} \left[v_{n,j}^*(\theta) + \varphi_j(\theta) \right]_-^2 + \sum_{j=p+1}^{k} \left(v_{n,j}^*(\theta) + \varphi_j(\theta) \right)^2 \right\}, \tag{3.1}$$

where $v_{n,j}^*(\theta)$ is as in (2.8) and $\varphi_j(\theta)$ is as in (2.11). This approximation shares the common structure in (2.7) with $\tilde{\Theta} = \Theta(\lambda_0)$ and $\ell_j(\theta) = \varphi_j(\theta)$ for $j = 1, \dots, k$. After showing that this approximation does not deliver a valid test, the second question arises: "How is it that the two modifications in (2.9) and (2.13), which may look somewhat arbitrary ex ante, eliminate the problems associated with $T_n^{\text{naive}}(\lambda_0)$?" We answer these two questions in the context of the following simple example.

Let $\{W_i\}_{i=1}^n = \{(W_{1,i}, W_{2,i})\}_{i=1}^n$ be an i.i.d. sequence of random variables with distribution $F = N(\mathbf{0}_2, I_2)$, where $\mathbf{0}_2$ is a two dimensional vector of zeros and I_2 is the 2×2 identity matrix. Let $(\theta_1, \theta_2) \in \Theta = [-1, 1]^2$ and consider the moment inequality model

$$E_F\left[m_1(W_i, \theta)\right] = E_F[W_{1,i} - \theta_1 - \theta_2] \geq 0,$$

$$E_F\left[m_2(W_i, \theta)\right] = E_F[\theta_1 + \theta_2 - W_{2,i}] \geq 0.$$

If we denote by $\Theta_I(F)$ the so-called identified set, that is, the set of all parameters in Θ that satisfy the moment inequality model above, it follows that

$$\Theta_I(F) = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\}.$$

We are interested in testing the hypotheses

$$H_0 : \theta_1 = 0 \quad \text{vs.} \quad H_1 : \theta_1 \neq 0, \tag{3.2}$$

which correspond to choosing $\lambda(\theta) = \theta_1$ and $\lambda_0 = 0$ in (1.2). In this case, the set $\Theta(\lambda_0)$ is given by

$$\Theta(\lambda_0) = \left\{\theta \in \Theta : \theta_1 = 0, \theta_2 \in [-1, 1]\right\},$$

which is a special case of the one described in Remark 2.2. Since the point $\theta = (0, 0)$ belongs to $\Theta(\lambda_0)$ and $\Theta_I(F)$, the null hypothesis in (3.2) is true in this example.

Profiled test statistic

The profiled test statistic $T_n(\lambda_0)$ in (2.5) here takes the form

$$T_n(0) = \inf_{\theta_2\in[-1,1]} Q_n(0, \theta_2) = \inf_{\theta_2\in[-1,1]} \left\{ \left[\frac{\sqrt{n}(\bar{W}_{n,1} - \theta_2)}{\hat{\sigma}_{n,1}} \right]_-^2 + \left[\frac{\sqrt{n}(\theta_2 - \bar{W}_{n,2})}{\hat{\sigma}_{n,2}} \right]_-^2 \right\},$$

where we are implicitly using the fact that $\hat{\sigma}_{n,j}(\theta)$ does not depend on θ for $j = 1, 2$ in this example.

Simple algebra shows that the infimum is attained at

$$\theta_2^\star = \frac{\hat{\sigma}_{n,2}^2 \bar{W}_{n,1} + \hat{\sigma}_{n,1}^2 \bar{W}_{n,2}}{\hat{\sigma}_{n,2}^2 + \hat{\sigma}_{n,1}^2} \quad \text{w.p.a.1}, \tag{3.3}$$

and this immediately leads to

$$T_n(0) = Q_n\big(0, \theta_2^\star\big) = \frac{1}{\hat{\sigma}_{n,2}^2 + \hat{\sigma}_{n,1}^2}[\sqrt{n}\bar{W}_{n,1} - \sqrt{n}\bar{W}_{n,2}]_-^2 \xrightarrow{d} \frac{1}{2}[Z_1 - Z_2]_-^2, \tag{3.4}$$

where $(Z_1, Z_2) \sim N(\mathbf{0}_2, I_2)$. Thus, the profiled test statistic has a limiting distribution where both moments are binding and asymptotically correlated, something that arises from the common random element $\theta_2^\star$ appearing in both moments.

Naïve GMS

This approach approximates the limiting distribution in (3.4) using (3.1). To describe this approach, first note that $v_{n,j}^*(\theta)$ in (2.8) does not depend on θ in this example since

$$v_{n,1}^*(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\big[(W_{1,i} - \theta_1 - \theta_2) - (\bar{W}_{n,1} - \theta_1 - \theta_2)\big]\zeta_i}{\hat{\sigma}_{n,1}} = Z_{n,1}^*,$$

$$v_{n,2}^*(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\big[(\theta_1 + \theta_2 - W_{2,i}) - (\theta_1 + \theta_2 - \bar{W}_{n,2})\big]\zeta_i}{\hat{\sigma}_{n,2}} = -Z_{n,2}^*,$$

and $Z_{n,j}^* = \frac{1}{\sqrt{n}} \hat{\sigma}_{n,j}^{-1} \sum_{i=1}^n (W_{j,i} - \bar{W}_{n,j})\zeta_i$ for $j = 1, 2$. In addition,

$$\big\{Z_{n,1}^*, Z_{n,2}^* | \{W_i\}_{i=1}^n\big\} \xrightarrow{d} Z = (Z_1, Z_2) \sim N(\mathbf{0}_2, I_2) \quad \text{w.p.1.}$$

It follows that the naïve approximation in (3.1) takes the form

$$T_n^{\text{naive}}(0) = \inf_{\theta_2 \in [-1,1]} \big[Z_{n,1}^* + \varphi_1(0, \theta_2)\big]_-^2 + \big[-Z_{n,2}^* + \varphi_2(0, \theta_2)\big]_-^2,$$

where $(Z_{n,1}^*, Z_{n,2}^*)$ does not depend on θ and $\varphi_j(\theta)$ is defined as in (2.11). Some algebra shows that

$$\big\{T_n^{\text{naive}}(0) | \{W_i\}_{i=1}^n\big\} \xrightarrow{d} \min\big\{[Z_1]_-^2, [-Z_2]_-^2\big\} \quad \text{w.p.1.} \tag{3.5}$$

This result intuitively follows from the fact that the GMS functions depend on

$$\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,1}^{-1}\bar{m}_{n,1}(0, \theta_2) = \kappa_n^{-1}\sqrt{n}\frac{\bar{W}_{n,1}}{\hat{\sigma}_{n,1}} - \kappa_n^{-1}\sqrt{n}\frac{\theta_2}{\hat{\sigma}_{n,1}},$$

$$\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,2}^{-1}\bar{m}_{n,2}(0, \theta_2) = \kappa_n^{-1}\sqrt{n}\frac{\theta_2}{\hat{\sigma}_{n,2}} - \kappa_n^{-1}\sqrt{n}\frac{\bar{W}_{n,2}}{\hat{\sigma}_{n,2}}.$$

It thus follows that $(\varphi_1(0, \theta_2), \varphi_2(0, \theta_2)) \to^p (0, \infty)$ when $\theta_2 > 0$ and $(\varphi_1(0, \theta_2), \varphi_2(0, \theta_2)) \to^p (\infty, 0)$ when $\theta_2 < 0$. In other words, the naïve GMS approximation does not penalize large negative values of $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$ (due to the fact that $\varphi_j(\theta) \geq 0$) and thus can afford to treat an inequality as slack by making the remaining inequality very negative (and treat it as binding). When $\alpha = 10\%$, the $1 - \alpha$ quantile of the distribution in (3.4) is 1.64, while the $1 - \alpha$ quantile of the distribution in (3.5) is 0.23. This delivers a naïve GMS test with null rejection probability converging to 31%, which clearly exceeds 10%. It is worth noting that the results in Andrews and Soares (2010) do not cover profiled test statistics like the one in (2.5), so this example simply illustrates that their idea cannot be naïvely extended.

Test MR

Now consider the two approximations in (2.9) and (2.13) that lead to Test MR. The first approximation takes the form

$$T_n^{\mathrm{DR}}(0) = \inf_{\theta \in \hat{\Theta}_I(0)} \big[Z_{n,1}^* + \varphi_1(\theta_1, \theta_2)\big]_-^2 + \big[-Z_{n,2}^* + \varphi_2(\theta_1, \theta_2)\big]_-^2,$$

where, for $\theta_2^\star$ defined as in (3.3), it is possible to show that

$$\hat{\Theta}_I(0) = \big\{\theta \in \Theta : \theta_1 = 0 \text{ and } \theta_2 = \theta_2^\star \text{ if } \bar{W}_{n,1} \leq \bar{W}_{n,2} \text{ or}$$
$$\theta_2 \in [\bar{W}_{n,2}, \bar{W}_{n,1}] \text{ if } \bar{W}_{n,1} > \bar{W}_{n,2}\big\}.$$

We term this the *discard resampling* approximation for reasons explained below. Some algebra shows that

$$\big\{T_n^{\mathrm{DR}}(0) | \{W_i\}_{i=1}^n\big\} \xrightarrow{d} [Z_1]_-^2 + [-Z_2]_-^2 \quad \text{w.p.1}, \tag{3.6}$$

where $(Z_1, Z_2) \sim N(\mathbf{0}_2, I_2)$. Since $\frac{1}{2}[Z_1 - Z_2]_-^2 \leq [Z_1]_-^2 + [-Z_2]_-^2$, using the $1-\alpha$ quantile of $T_n^{\mathrm{DR}}(0)$ delivers an asymptotically valid (and possibly conservative) test. This approximation does not exhibit the problem we found in the naïve GMS approach because the set $\hat{\Theta}_I(0)$ does not allow the approximation to choose values of θ_2 far from zero to make one moment very negative and the other one slack. In other words, the set $\hat{\Theta}_I(0)$ *discards* the problematic points from $\Theta(\lambda_0)$ and this is precisely what leads to a valid approximation.

The second approximation takes the form

$$T_n^{\mathrm{PR}}(0) = \inf_{\theta_2 \in [-1,1]} \big[Z_{n,1}^* + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,1}^{-1}\bar{m}_{n,1}(0, \theta_2)\big]_-^2$$
$$+ \big[-Z_{n,2}^* + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,2}^{-1}\bar{m}_{n,2}(0, \theta_2)\big]_-^2.$$

We term this the *penalize resampling* approximation for reasons explained below. Some algebra shows that

$$\big\{T_n^{\mathrm{PR}}(0) | \{W_i\}_{i=1}^n\big\} \xrightarrow{d} \frac{1}{2}[Z_1 - Z_2]_-^2 \quad \text{w.p.1}, \tag{3.7}$$

and thus using the $1-\alpha$ quantile of $T_n^{\mathrm{PR}}(0)$ delivers an asymptotically valid (and exact in this case) test. This approximation does not exhibit the problem we found in the naïve GMS approach because the slackness function $\ell_j(\theta) = \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$, which may take negative values, *penalizes* the problematic points from $\Theta(\lambda_0)$. This feature implies that the infimum in $T_n^{\mathrm{PR}}(0)$ is attained at

$$\theta_2^{\dagger} = \theta_2^{\star} + \frac{(\kappa_n/\sqrt{n})\big(\hat{\sigma}_{n,2}^2 Z_{n,1}^* + \hat{\sigma}_{n,1}^2 Z_{n,2}^*\big)}{\hat{\sigma}_{n,2}^2 + \hat{\sigma}_{n,1}^2} \quad \text{w.p.a.1}, \tag{3.8}$$

where $\theta_2^{\star}$ is as in (3.3). Hence, using a slackness function that is not restricted to be nonnegative introduces a penalty when violating the inequalities that mimics the behavior of the profiled test statistic $T_n(\lambda_0)$.

Putting all these results together shows that

$$\big\{T_n^{\mathrm{MR}}(0) | \{W_i\}_{i=1}^n\big\} \xrightarrow{d} \frac{1}{2}[Z_1 - Z_2]_-^2 \quad \text{w.p.1}, \tag{3.9}$$

and thus Test MR, as defined in (2.1), has null rejection probability equal to α in this example, that is, it is an asymptotically valid test. We note that in this example the quantile of Test MR coincides with the one from the second resampling approximation. In general, the two resampling approximations leading to Test MR do not dominate each other; see Remark 4.11.

Remark 3.1. The example does not include moment equalities and thus does not illustrate why the penalize resampling approximation includes penalties for the moment equalities as well. However, the intuition behind it is analogous to that in the example. Using a slackness function that affects moment equalities enforces that the approximation cannot be determined by parameter values that are far from the identified set $\Theta_I(F)$, therefore mimicking the behavior of the profiled test statistic $T_n(\lambda_0)$.

Remark 3.2. In the next section we present formal results that show that tests that reject the null in (1.2) when the profiled test statistic in (2.5) exceeds the $1-\alpha$ quantile of either $T_n^{\mathrm{DR}}(\lambda_0)$, $T_n^{\mathrm{PR}}(\lambda_0)$, or $T_n^{\mathrm{MR}}(\lambda_0)$, control asymptotic size uniformly over a large class of distributions. We however recommend the use of the Test MR on the grounds that this test delivers the best power properties relative to tests based on $T_n^{\mathrm{DR}}(\lambda_0)$, $T_n^{\mathrm{PR}}(\lambda_0)$, projections, and subsampling.

4. Main results on size and power

4.1 *Minimum resampling test*

We now describe the minimum resampling test in (2.1) for a generic test statistic and a generic GMS slackness function. So as to do this, we introduce some notation. Let $\bar{m}_n(\theta) \equiv (\bar{m}_{n,1}(\theta), \ldots, \bar{m}_{n,k}(\theta))$, where $\bar{m}_{n,j}(\theta)$ is as in (2.2) for $j = 1, \ldots, k$. Denote by

$$\hat{D}_n(\theta) \equiv \operatorname{diag}\big\{\hat{\sigma}_{n,1}^2(\theta), \ldots, \hat{\sigma}_{n,k}^2(\theta)\big\}$$

the diagonal matrix of variances, where $\hat{\sigma}_{n,j}^2(\theta)$ is as in (2.3), and let $\hat{\Omega}_n(\theta)$ be the sample correlation matrix of the vector $m(W_i, \theta)$. For a given $\lambda \in \Lambda$, the profiled test statistic is

$$T_n(\lambda) \equiv \inf_{\theta \in \Theta(\lambda)} Q_n(\theta), \tag{4.1}$$

where

$$Q_n(\theta) = S\big(\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\big) \tag{4.2}$$

and $S(\cdot)$ is a test function satisfying Assumptions M.1–M.9. In the context of the moment (in)equality model in (1.1), it is convenient to consider functions $Q_n(\theta)$ that take the form in (4.2) (see, e.g., Andrews and Guggenberger (2009), Andrews and Soares (2010), Bugni, Canay, and Guggenberger (2012)). Some common examples of test functions satisfying all of the required conditions are the MMM function in (2.6), the maximum test statistic in Romano, Shaikh, and Wolf (2014), and the adjusted quasi-likelihood ratio statistic in Andrews and Barwick (2012).

The critical value of Test MR requires two resampling approximations to the distribution of $T_n(\lambda)$. The discard resampling approximation uses the statistic

$$T_n^{\mathrm{DR}}(\lambda) \equiv \inf_{\theta\in\hat{\Theta}_I(\lambda)} S\big(v_n^*(\theta) + \varphi\big(\kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta)\big), \hat{\Omega}_n(\theta)\big), \tag{4.3}$$

where $\hat{\Theta}_I(\lambda)$ is as in (2.10), $\varphi = (\varphi_1, \dots, \varphi_k)$, and φ_j for $j = 1, \dots, k$ is a GMS function satisfying Assumption A.1. Examples of functions φ_j satisfying our assumptions include the one in (2.11), $\varphi_j(x_j) = \max\{x_j, 0\}$, and several others; see Remark B.1. We note that the previous sections treated φ_j as a function of θ when in fact these are mappings from $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$ to $\mathbb{R}_{+,\infty}$. We did this to keep the exposition as simple as possible in those sections, but in what follows we properly view φ_j as a function of $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$.

The use of $T_n^{\mathrm{DR}}(\lambda)$ to approximate the quantiles of the distribution of $T_n(\lambda)$ is based on an approximation that forces θ to be *close* to the identified set

$$\begin{aligned} \Theta_I(F) \equiv \big\{\theta \in \Theta : {} & E_F\big[m_j(W_i, \theta)\big] \geq 0 \text{ for } j = 1, \dots, p \text{ and} \\ & E_F\big[m_j(W_i, \theta)\big] = 0 \text{ for } j = p+1, \dots, k\big\}. \end{aligned} \tag{4.4}$$

This is achieved by using the approximation $\hat{\Theta}_I(\lambda)$ for the intersection of $\Theta(\lambda)$ and $\Theta_I(F)$, that is, for

$$\Theta(\lambda) \cap \Theta_I(F) = \big\{\theta \in \Theta_I(F) : \lambda(\theta) = \lambda\big\}.$$

The approximation therefore discards the points in $\Theta(\lambda)$ that are far from $\Theta_I(F)$. Note that replacing $\hat{\Theta}_I(\lambda)$ with $\Theta(\lambda)$ while keeping the function $\varphi(\cdot)$ in (4.3) leads to the naïve GMS approach. As illustrated in Section 3, such an approach does not deliver a valid approximation.

Remark 4.1. The set $\hat{\Theta}_I(\lambda)$ could be defined as $\hat{\Theta}_I(\lambda) \equiv \{\theta \in \Theta(\lambda) : Q_n(\theta) \leq T_n(\lambda) + \delta_n\}$, with $\delta_n \geq 0$ and $\delta_n = o_p(1)$, without affecting our results. This is relevant for situations where the optimization is only guaranteed to approximate exact minimizers. In addition, the set $\hat{\Theta}_I(\lambda)$ is not required to contain *all* the minimizers of $Q_n(\theta)$, in the sense that our results hold as long as $\hat{\Theta}_I(\lambda)$ approximates *at least one* of the possible minimizers. More specifically, all we need is that

$$P_F\big(\hat{\Theta}_I(\lambda) \subseteq \Theta(\lambda) \cap \Theta_I^{\ln \kappa_n}(F)\big) \to 1 \tag{4.5}$$

uniformly over the parameter space defined in the next section, where $\Theta_I^{\ln \kappa_n}(F)$ is a nonrandom expansion of $\Theta_I(F)$ defined in Table 1. It follows from Bugni, Canay, and Shi (2015, Lemma D.13) that all the variants of $\hat{\Theta}_I(\lambda)$ just discussed satisfy the above property.

The penalize resampling approximation uses the statistic

$$T_n^{\mathrm{PR}}(\lambda) \equiv \inf_{\theta \in \Theta(\lambda)} S\big(v_n^*(\theta) + \kappa_n^{-1}\sqrt{n}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\big). \tag{4.6}$$

This second approximation does not require the set $\hat{\Theta}_I(\lambda)$ and it uses a slackness function that does not belong to the class of GMS functions. This is so because GMS functions are assumed to satisfy $\varphi_j(\cdot) \geq 0$ for $j = 1, \dots, p$ and $\varphi_j(\cdot) = 0$ for $j = p+1, \dots, k$ to ensure that GMS tests have good power properties; see Andrews and Soares (2010, Assumption GMS6 and Theorem 3). As illustrated in Section 3, the fact that $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$ may be negative for $j = 1, \dots, p$ and $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$ may be nonzero for $j = p+1, \dots, k$ is fundamental for how this approximation works. This is because using this slackness function penalizes θ values away from the identified set (for equality and inequality restrictions) and thus automatically restricts the effective infimum range to a neighborhood of the identified set.

Definition 4.1 ((Minimum Resampling Test)). Let $T_n^{\mathrm{DR}}(\lambda)$ and $T_n^{\mathrm{PR}}(\lambda)$ be defined as in (4.3) and (4.6), respectively, where $v_n^*(\theta)$ is defined as in (2.8) and is common to both resampling statistics. Let the critical value $\hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha)$ be the (conditional) $1-\alpha$ quantile of

$$T_n^{\mathrm{MR}}(\lambda) \equiv \min\big\{T_n^{\mathrm{DR}}(\lambda), T_n^{\mathrm{PR}}(\lambda)\big\}.$$

The minimum resampling test (or Test MR) is

$$\phi_n^{\mathrm{MR}}(\lambda) \equiv 1\big\{T_n(\lambda) > \hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha)\big\}.$$

The profiled test statistic $T_n(\lambda)$ is standard in point identified models. It has been considered in the context of partially identified models for a subsampling test by Romano and Shaikh (2008), although Romano and Shaikh (2008, Theorem 3.4) did not derive asymptotic properties of $T_n(\lambda)$ and proved the validity of their test under high-level conditions. The novelty in Test MR lies in the critical value $\hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha)$. This is because each of the two basic resampling approximations we combine—embedded in $T_n^{\mathrm{DR}}(\lambda)$ and $T_n^{\mathrm{PR}}(\lambda)$—has good power properties in particular directions and neither dominates the other in terms of asymptotic power; see Example 4.1. By combining the two approximations into the resampling statistic $T_n^{\mathrm{MR}}(\lambda)$, the test $\phi_n^{\mathrm{MR}}(\lambda)$ not only dominates each of these basic approximations, but also dominates projection-based tests and subsampling tests. We formalize these properties in the following sections.

Remark 4.2. Test MR and all our results can be extended to one-sided testing problems where

$$H_0 : \lambda(\theta) \leq \lambda_0 \quad \text{vs.} \quad H_1 : \lambda(\theta) > \lambda_0.$$

The only modification lies in the definition of $\Theta(\lambda)$, which should now be $\{\theta \in \Theta : \lambda(\theta) \leq \lambda\}$. This change affects the profiled test statistic and the two approximations, $T_n^{\mathrm{DR}}(\lambda)$ and $T_n^{\mathrm{PR}}(\lambda)$, leading to Test MR.

4.2 *Asymptotic size*

In this section we show that Test MR controls asymptotic size uniformly over an appropriately defined parameter space. We define the parameter space after introducing some additional notation. First, we assume that F, the distribution of the observed data, belongs to a *baseline distribution space* denoted by $\mathcal{P}$.

Definition 4.2 ((Baseline Distribution Space)). The baseline space of distributions $\mathcal{P}$ is the set of distributions F satisfying the following properties:

(i) The data $\{W_i\}_{i=1}^n$ are i.i.d. under F.

(ii) We have $\sigma_{F,j}^2(\theta) = \mathrm{Var}_F(m_j(W_i, \theta)) \in (0, \infty)$ for $j = 1, \dots, k$ and all $\theta \in \Theta$.

(iii) For all $j = 1, \dots, k$, $\{\sigma_{F,j}^{-1}(\theta) m_j(\cdot, \theta) : \mathcal{W} \to \mathbb{R}\}$ is a *measurable* class of functions indexed by $\theta \in \Theta$.

(iv) The empirical process $v_n(\theta)$ with jth component as in Table 1 is asymptotically ρ_F-equicontinuous uniformly in $F \in \mathcal{P}$ in the sense of van der Vaart and Wellner (1996, p. 169). That is, for any $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{F \in \mathcal{P}} P_F^*\Big(\sup_{\rho_F(\theta, \theta') < \delta} \big\| v_n(\theta) - v_n(\theta') \big\| > \varepsilon \Big) = 0,$$

where P_F^* denotes outer probability and ρ_F is the coordinate-wise intrinsic variance semimetric in (A.1).

(v) For some constant $a > 0$ and all $j = 1, \dots, k$.

$$\sup_{F \in \mathcal{P}} E_F \left[\sup_{\theta \in \Theta} \left| \frac{m_j(W, \theta)}{\sigma_{F,j}(\theta)} \right|^{2+a} \right] < \infty.$$

(vi) For $\Omega_F(\theta, \theta')$ being the $k \times k$ correlation matrix with $[j_1, j_2]$ component as defined in Table 1,

$$\lim_{\delta \downarrow 0} \sup_{\|(\theta_1, \theta_1') - (\theta_2, \theta_2')\| < \delta} \sup_{F \in \mathcal{P}} \big\| \Omega_F(\theta_1, \theta_1') - \Omega_F(\theta_2, \theta_2') \big\| = 0.$$

Parts (i)–(iii) in Definition 4.2 are mild conditions. In fact, the kind of uniform laws of large numbers we need for our analysis would not hold without part (iii) (see van der

Vaart and Wellner (1996, p. 110)). Part (iv) is a uniform stochastic equicontinuity assumption that, in combination with the other requirements, is used to show that the class of functions $\{\sigma_{F,j}^{-1}(\theta)m_j(\cdot,\theta):\mathcal{W}\to\mathbb{R}\}$ is Donsker and pre-Gaussian uniformly in $F\in\mathcal{P}$ (see Lemma S.3.1). Part (v) provides a uniform (in F and θ) envelope function that satisfies a uniform integrability condition. This is essential to obtaining uniform versions of the laws of large numbers and central limit theorems. Finally, part (vi) requires the correlation matrices to be uniformly equicontinuous, which is used to show pre-Gaussianity.

Second, we introduce a parameter space for the tuple (λ, F). Note that inference for the entire parameter θ requires a parameter space for the tuple (θ, F); see, for example, Andrews and Soares (2010). Here the hypotheses in (1.2) are determined by the function $\lambda(\cdot):\Theta\to\Lambda$, and so the relevant tuple becomes (λ, F).

Definition 4.3 ((Parameter Space for (λ, F))). The parameter space for (λ, F) is given by

$$\mathcal{L}\equiv\big\{(\lambda,F):F\in\mathcal{P},\lambda\in\Lambda\big\}.$$

The subset of $\mathcal{L}$ that is consistent with the null hypothesis, referred to as the *null parameter space*, is

$$\mathcal{L}_0\equiv\big\{(\lambda,F):F\in\mathcal{P},\lambda\in\Lambda,\Theta(\lambda)\cap\Theta_I(F)\neq\emptyset\big\}.$$

The following theorem states that Test MR controls asymptotic size uniformly over parameters in $\mathcal{L}_0$.

Theorem 4.1. *Let Assumptions A.1–A.3 hold and let $\phi_n^{\mathrm{MR}}(\lambda)$ be the test in Definition 4.1. Then, for $\alpha\in(0,\frac{1}{2})$,*

$$\limsup_{n\to\infty}\sup_{(\lambda,F)\in\mathcal{L}_0}E_F\big[\phi_n^{\mathrm{MR}}(\lambda)\big]\leq\alpha.$$

All the assumptions we use throughout the paper can be found in Appendix B. Assumption A.1 restricts the class of GMS functions we allow for; see Remark B.1. Assumption A.2 is a continuity assumption on the limit distribution of $T_n(\lambda)$; see Remark B.2. Finally, Assumption A.3 is a key sufficient condition for the asymptotic validity of our test that requires the population version of $Q_n(\theta)$ to satisfy a minorant-type condition as in Chernozhukov, Hong, and Tamer (2007) and the normalized population moments to be sufficiently smooth. This assumption also requires $\Theta(\lambda)$ to be convex, which is satisfied for linear $\lambda(\cdot)$ in two-sided testing problems, and quasi-convex $\lambda(\cdot)$ for one-sided testing problems. See Remark B.3 for a detailed discussion. We verified that all these assumptions hold in the examples we use throughout the paper.

Remark 4.3. We can construct examples where Assumption A.3 is violated and Test MR overrejects. Interestingly enough, in those examples the subsampling-based test proposed by Romano and Shaikh (2008), and discussed in Section 4.4, also exhibits overrejection. We conjecture that Assumption A.3 is part of the primitive conditions that may be required to satisfy the high-level conditions stated in Romano and Shaikh (2008). This is, however, beyond the scope of this paper as here we recommend Test MR.

Remark 4.4. The proof of Theorem 4.1 relies on Theorem S.2.4 in the Supplemental Material, which derives the limiting distribution of $T_n(\lambda)$ along sequences of parameters $(\lambda_n, F_n) \in \mathcal{L}_0$. The expression of this limit distribution is not particularly insightful, so we refer the reader to the Supplemental Material for it. We do emphasize that the result in Theorem S.2.4 is new, represents an important milestone into Theorem 4.1, and is part of the technical contributions of this paper.

Remark 4.5. By exploiting the well known duality between tests and confidence sets, Test MR may be inverted to construct confidence sets for the parameter λ. That is, if we let

$$\mathrm{CS}_n^\lambda(1-\alpha) \equiv \big\{\lambda \in \Lambda : T_n^{\mathrm{MR}}(\lambda) \leq \hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha)\big\},$$

it follows from Theorem 4.1 that

$$\liminf_{n\to\infty} \inf_{(\lambda,F)\in\mathcal{L}_0} P_F\big(\lambda \in \mathrm{CS}_n^\lambda(1-\alpha)\big) \geq 1-\alpha. \tag{4.7}$$

In particular, when $\lambda(\theta) = \theta_s$ for some $s \in \{1, \dots, d_\theta\}$, $\mathrm{CS}_n^\lambda(1-\alpha)$ constitutes a confidence interval for the component θ_s.

4.3 *Power advantage over projection tests*

To test the hypotheses in (1.2), a common practice in applied work involves projecting joint CS's for the entire parameter θ into the image of the function $\lambda(\cdot)$. This practice requires one to first compute

$$\mathrm{CS}_n^\theta(1-\alpha) \equiv \big\{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\big\}, \tag{4.8}$$

where $Q_n(\theta)$ is as in (4.2) and $\hat{c}_n(\theta, 1-\alpha)$ is such that $\mathrm{CS}_n^\theta(1-\alpha)$ has the correct asymptotic coverage. CS's that have the structure in (4.8) and control asymptotic coverage have been proposed by Romano and Shaikh (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Canay (2010), and Bugni (2010, 2016), among others. The projection test then rejects the null hypothesis in (1.2) when the image of $\mathrm{CS}_n^\theta(1-\alpha)$ under $\lambda(\cdot)$ does not include the value λ_0. Formally,

$$\phi_n^{\mathrm{BP}}(\lambda) \equiv 1\big\{\mathrm{CS}_n^\theta(1-\alpha) \cap \Theta(\lambda) = \emptyset\big\}. \tag{4.9}$$

We refer to this test as a projection test, or Test BP (by-product projection), to emphasize the fact that this test comes as a by-product of constructing CS's for the entire parameter θ. Applied papers using this test include Ciliberto and Tamer (2009), Grieco (2014), Dickstein and Morales (2015), and Wollmann (2015), among others.

Test BP inherits its size and power properties from the properties of $\mathrm{CS}_n^\theta(1-\alpha)$. These properties depend on the particular choice of test statistic and critical value entering

$\text{CS}_n^\theta(1-\alpha)$ in (4.8). All the tests we consider in this paper are functions of the same $Q_n(\theta)$ and thus their relative power properties do not depend on the choice of test function $S(\cdot)$. However, the performance of Test BP tightly depends on the critical value used in $\text{CS}_n^\theta(1-\alpha)$. Bugni (2016) shows that GMS tests have more accurate asymptotic size than subsampling tests. Andrews and Soares (2010) show that GMS tests are more powerful than plug-in asymptotics or subsampling tests. This means that, asymptotically, Test BP implemented with a GMS CS will be less conservative and more powerful than the analogous test implemented with plug-in asymptotics or subsampling. We therefore adopt the GMS version of Test BP as the "benchmark version." This is stated formally in the maintained Assumption M.4 in Appendix B.

The next theorem formalizes the power advantage of Test MR over Test BP.

Theorem 4.2. *Let $\phi_n^{\text{MR}}(\lambda)$ and $\phi_n^{\text{BP}}(\lambda)$ be implemented with the same sequence $\{\kappa_n\}_{n\geq 1}$ and GMS function $\varphi(\cdot)$. Then, for any $(\lambda, F)\in\mathcal{L}$ and all $n\in\mathbb{N}$, it follows that $\phi_n^{\text{MR}}(\lambda)\geq \phi_n^{\text{BP}}(\lambda)$.*

Corollary 4.1. *For any sequence $\{(\lambda_n, F_n)\in\mathcal{L}\}_{n\geq 1}$, $\liminf_{n\to\infty}(E_{F_n}[\phi_n^{\text{MR}}(\lambda_n)] - E_{F_n}[\phi_n^{\text{BP}}(\lambda_n)])\geq 0$.*

Theorem 4.2 is a statement for all $n\in\mathbb{N}$ and $(\lambda,F)\in\mathcal{L}$, and thus it is a result about finite sample power *and* size. This theorem also implies that the CS for λ defined in Remark 4.5 is always a subset of the one produced by projecting the joint CS in (4.8).

To describe the mechanics behind Theorem 4.2, let $\hat{c}_n^{\text{DR}}(\lambda, 1-\alpha)$ be the (conditional) $1-\alpha$ quantile of $T_n^{\text{DR}}(\lambda)$ in (4.3) and let

$$\phi_n^{\text{DR}}(\lambda)\equiv 1\big\{T_n(\lambda)>\hat{c}_n^{\text{DR}}(\lambda, 1-\alpha)\big\} \tag{4.10}$$

be the test associated with the discard resampling approximation leading to Test MR. To prove the theorem, we first modify the arguments in Bugni, Canay, and Shi (2015) to show that $\phi_n^{\text{DR}}(\lambda)\geq\phi_n^{\text{BP}}(\lambda)$, provided these tests are implemented with the same sequence $\{\kappa_n\}_{n\geq 1}$ and GMS function $\varphi(\cdot)$. We then extend the result to $\phi_n^{\text{MR}}(\lambda)$ by using

$$\phi_n^{\text{MR}}(\lambda)\geq\phi_n^{\text{DR}}(\lambda) \tag{4.11}$$

for all $(\lambda,F)\in\mathcal{L}$ and $n\in\mathbb{N}$, which in turn follows from $\hat{c}_n^{\text{MR}}(\lambda, 1-\alpha)\leq\hat{c}_n^{\text{DR}}(\lambda,1-\alpha)$.

Remark 4.6. Under a condition similar to Bugni, Canay, and Shi (2015, Assumption A.9), $\phi_n^{\text{DR}}(\lambda)$ has asymptotic power that is *strictly* higher than that of $\phi_n^{\text{BP}}(\lambda)$ for certain local alternative hypotheses. The proof is similar to that in Bugni, Canay, and Shi (2015, Theorem 6.2), so we omit it here. We do illustrate this in Example 4.1.

Remark 4.7. The test $\phi_n^{\text{DR}}(\lambda)$ in (4.10) corresponds to one of the tests introduced by Bugni, Canay, and Shi (2015) to test the correct specification of the model in (1.1). By (4.11), this test controls asymptotic size for the null hypothesis in (1.2). However, $\phi_n^{\text{DR}}(\lambda)$ presents two disadvantages relative to $\phi_n^{\text{MR}}(\lambda)$. First, the power results we present in the next section for $\phi_n^{\mathrm{MR}}(\lambda)$ do not necessarily hold for $\phi_n^{\mathrm{DR}}(\lambda)$; that is, $\phi_n^{\mathrm{DR}}(\lambda)$ may not have better power than the subsampling test proposed by Romano and Shaikh (2008). Second, $\phi_n^{\mathrm{MR}}(\lambda)$ has strictly higher asymptotic power than $\phi_n^{\mathrm{DR}}(\lambda)$ in some cases; see Example 4.1 for an illustration.

We conclude this section with two aspects that go beyond Theorem 4.2. First, when the function $\lambda(\cdot)$ selects one of several elements of Θ and so $\dim(\Theta) > \dim(\Lambda)$, the implementation of Test MR is computationally attractive as it involves inverting a test over a smaller dimension. In those cases, Test MR has power *and* computational advantages over Test BP. Second, Test BP requires fewer assumptions to control asymptotic size relative to Test MR. It is fair to say then that Test BP is more "robust" than Test MR, in the sense that if some of the Assumptions A.1–A.3 fail, Test BP may still control asymptotic size.

4.4 *Power advantage over subsampling tests*

In this section we show that Test MR dominates subsampling-based tests by exploiting its connection to the second resampling approximation $T_n^{\mathrm{PR}}(\lambda)$ in (4.6). We follow a proof approach analogous to the one in the previous section, first deriving results for the test associated with $T_n^{\mathrm{PR}}(\lambda)$ and then extending these results to $\phi_n^{\mathrm{MR}}(\lambda)$ by exploiting the finite sample inequality in (4.15) below.

We start by describing subsampling-based tests. Romano and Shaikh (2008, Section 3.4) propose to test the hypothesis in (1.2) using $T_n(\lambda)$ in (4.1) with a subsampling critical value. Concretely, the test they propose, which we denote by Test SS, is

$$\phi_n^{\mathrm{SS}}(\lambda) \equiv 1\{T_n(\lambda) > \hat{c}_n^{\mathrm{SS}}(\lambda, 1-\alpha)\}, \tag{4.12}$$

where $\hat{c}_n^{\mathrm{SS}}(\lambda, 1-\alpha)$ is the (conditional) $1-\alpha$ quantile of the distribution of $T_n^{\mathrm{SS}}(\lambda)$, which is identical to $T_n(\lambda)$ but computed using a random sample of size b_n without replacement from $\{W_i\}_{i=1}^n$. We assume the subsample size satisfies $b_n \to \infty$ and $b_n/n \to 0$. Romano and Shaikh (2008, Remark 3.11) note that projection-based tests may lead to conservative inference, and they use this as a motivation for introducing Test SS. However, they provide neither formal comparisons between their test and projection-based tests nor primitive conditions for their test to control asymptotic size; see Remark 4.3.

To compare Test MR and Test SS, we define a class of distributions in the alternative hypotheses that are local to the null hypothesis. After noticing that the null hypothesis in (1.2) can be written as $\Theta(\lambda_0) \cap \Theta_I(F) \neq \emptyset$, we do this by defining sequences of distributions F_n for which $\Theta(\lambda_0) \cap \Theta_I(F_n) = \emptyset$ for all $n \in \mathbb{N}$, but where $\Theta(\lambda_n) \cap \Theta_I(F_n) \neq \emptyset$ for a sequence $\{\lambda_n\}_{n\geq 1}$ that approaches the value λ_0 in (1.2). These alternatives are conceptually similar to those in Andrews and Soares (2010), but the proof of our result involves additional challenges that are specific to the infimum present in the definition of our test statistic. The following definition formalizes the class of local alternative distributions we consider.

Definition 4.4 ((Local Alternatives)). Let $\lambda_0 \in \Lambda$ be the value in (1.2). The sequence $\{F_n\}_{n\geq 1}$ is a sequence of local alternatives if there is a $\{\lambda_n \in \Lambda\}_{n\geq 1}$ such that $\{(\lambda_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ and the following statements hold:

(a) For all $n \in \mathbb{N}$, $\Theta_I(F_n) \cap \Theta(\lambda_0) = \emptyset$.

(b) We have $d_H(\Theta(\lambda_n), \Theta(\lambda_0)) = O(n^{-1/2})$.

(c) For any $\theta \in \Theta$, $G_{F_n}(\theta) = O(1)$, where $G_F(\theta) \equiv \partial D_F^{-1/2}(\theta) E_F[m(W, \theta)]/\partial\theta'$.

Under the assumption that F_n is a local alternative (Assumption A.5), a restriction on κ_n and b_n (Assumption A.4), and smoothness conditions (Assumptions A.3 and A.6), we show the following result.

Theorem 4.3. *Let Assumptions* A.1–A.6 *hold. Then*

$$\liminf_{n\to\infty}\big(E_{F_n}\big[\phi_n^{\mathrm{MR}}(\lambda_0)\big] - E_{F_n}\big[\phi_n^{\mathrm{SS}}(\lambda_0)\big]\big) \geq 0. \tag{4.13}$$

To describe the mechanics behind Theorem 4.3, let $\hat{c}_n^{\mathrm{PR}}(\lambda, 1-\alpha)$ be the (conditional) $1-\alpha$ quantile of $T_n^{\mathrm{PR}}(\lambda)$ and let

$$\phi_n^{\mathrm{PR}}(\lambda) \equiv 1\big\{T_n(\lambda) > \hat{c}_n^{\mathrm{PR}}(\lambda, 1-\alpha)\big\} \tag{4.14}$$

be the test associated with the penalize resampling approximation leading to Test MR. The test in (4.14) is not part of the tests discussed in Bugni, Canay, and Shi (2015) but has recently been used for a different testing problem in Gandhi, Lu, and Shi (2013). By construction, $\hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) \leq \hat{c}_n^{\mathrm{PR}}(\lambda, 1-\alpha)$, and thus

$$\phi_n^{\mathrm{MR}}(\lambda) \geq \phi_n^{\mathrm{PR}}(\lambda), \tag{4.15}$$

for all $(\lambda, F) \in \mathcal{L}$ and $n \in \mathbb{N}$. To prove the theorem we first show that (4.13) holds with $\phi_n^{\mathrm{PR}}(\lambda_0)$ in place of $\phi_n^{\mathrm{MR}}(\lambda_0)$, and then use (4.15) to complete the argument.

Remark 4.8. To show that the asymptotic power of Test MR weakly dominates that of Test SS, Theorem 4.3 relies on Assumption A.4, which requires

$$\limsup_{n\to\infty} \kappa_n\sqrt{b_n/n} \leq 1. \tag{4.16}$$

For the problem of inference on the entire parameter θ, Andrews and Soares (2010) show the analogous result that the asymptotic power of the GMS test weakly dominates that of subsampling tests, based on the stronger condition that $\lim_{n\to\infty} \kappa_n\sqrt{b_n/n} = 0$. Given that Theorem 4.3 allows for $\limsup_{n\to\infty} \kappa_n\sqrt{b_n/n} = K \in (0, 1]$, we view our result as relatively more robust to the choice of κ_n and b_n.[2] We notice that Theorem 4.3 is consistent with the possibility of a failure of (4.13) whenever Assumption A.4 is violated, that is, when $\limsup_{n\to\infty} \kappa_n\sqrt{b_n/n} > 1$. Remark 4.13 provides a concrete example in which this possibility occurs. In any case, for the recommended choice of $\kappa_n = \sqrt{\ln n}$ in Andrews

[2]We would like to thank a referee for suggesting this generalization.

and Soares (2010, p. 131), a violation of this assumption implies a b_n larger than $O(n^c)$ for all $c \in (0, 1)$, which can result in Test SS having poor finite sample power properties, as discussed in Andrews and Soares (2010, p. 137).

Remark 4.9. The inequality in (4.13) can be strict for certain sequences of local alternatives. Lemma S.3.10 proves this result under the conditions in Assumption A.7. Intuitively, we require a sequence of alternative hypotheses in which one or more moment (in)equality is slack by magnitude that is $o(b_n^{-1/2})$ and larger than $O(\kappa_n n^{-1/2})$. We provide an illustration of Assumption A.7 in Example 4.2.

Remark 4.10. There are reasons to support Test MR over Test SS that go beyond asymptotic power. First, we find in our simulations that Test SS is significantly more sensitive to the choice of b_n than Test MR is to the choice of κ_n. Second, in the context of inference on the entire parameter θ, subsampling tests have been shown to have an error in rejection probability (ERP) of $O(b_n/n + b_n^{-1/2}) \geq O(n^{-1/3})$, while GMS-type tests have an ERP of $O(n^{-1/2})$ (cf. Bugni (2016)). We expect an analogous result to hold for the problem of inference on $\lambda(\theta)$, but a formal proof is well beyond the scope of this paper.

4.5 *Understanding the power results*

Theorems 4.2 and 4.3 follow by proving weak inequalities for $\phi_n^{\mathrm{DR}}(\lambda)$ and $\phi_n^{\mathrm{PR}}(\lambda)$, and then using the weak inequalities in (4.11) and (4.15) to extend the results to $\phi_n^{\mathrm{MR}}(\lambda)$. In this section we present two examples that illustrate how each of these weak inequalities may become strict in some cases. Example 4.1 illustrates a case where $\phi_n^{\mathrm{MR}}(\lambda)$ has strictly better asymptotic power than both $\phi_n^{\mathrm{DR}}(\lambda)$ and $\phi_n^{\mathrm{PR}}(\lambda)$. Example 4.2 illustrates a case where $\phi_n^{\mathrm{PR}}(\lambda)$—and so $\phi_n^{\mathrm{MR}}(\lambda)$—has strictly better asymptotic power than $\phi_n^{\mathrm{SS}}(\lambda)$.

Example 4.1. Let $\{W_i\}_{i=1}^n = \{(W_{1,i}, W_{2,i}, W_{3,i})\}_{i=1}^n$ be an i.i.d. sequence of random variables with distribution F_n, $V_{F_n}[W] = I_3$, $E_{F_n}[W_1] = \mu_1\kappa_n/\sqrt{n}$, $E_{F_n}[W_2] = \mu_2\kappa_n/\sqrt{n}$, and $E_{F_n}[W_3] = \mu_3/\sqrt{n}$ for some $\mu_1 > 1$, $\mu_2 \in (0, 1)$, and $\mu_3 \in \mathbb{R}$. Consider the following model with $\Theta = [-1, 1]^3$:

$$\begin{aligned} E_{F_n}\big[m_1(W_i, \theta)\big] &= E_{F_n}[W_{1,i} - \theta_1] \geq 0, \\ E_{F_n}\big[m_2(W_i, \theta)\big] &= E_{F_n}[W_{2,i} - \theta_2] \geq 0, \\ E_{F_n}\big[m_3(W_i, \theta)\big] &= E_{F_n}[W_{3,i} - \theta_3] = 0. \end{aligned} \tag{4.17}$$

We are interested in testing the hypotheses

$$H_0 : \theta = (0, 0, 0) \quad \text{vs.} \quad H_1 : \theta \neq (0, 0, 0),$$

which implies that $\lambda(\theta) = \theta$, $\Theta(\lambda) = \{(0, 0, 0)\}$, and $\hat{\Theta}_I(\lambda) = \{(0, 0, 0)\}$.[3] Note that H_0 is true if and only if $\mu_3 = 0$. The model in (4.17) is linear in θ and so several expressions

[3]In this example we use $\lambda(\theta) = \theta$ for simplicity, as it makes the infimum over $Q_n(\theta)$ trivial. We could generate the same conclusions using a different function by adding some complexity to the structure of the example.

do not depend on θ. These include $\hat{\sigma}_{n,j}(\theta) = \hat{\sigma}_{n,j}$ and $v^*_{n,j}(\theta) = v^*_{n,j}$ for $j = 1, 2, 3$, where $v^*_{n,j}(\theta)$ is defined in (2.8). As in Section 3, we use the MMM test statistic in (2.6) and the GMS function in (2.11). Below we also use $Z = (Z_1, Z_2, Z_3) \sim N(\mathbf{0}_3, I_3)$.

Simple algebra shows that the test statistic satisfies

$$\begin{aligned} T_n(\lambda) &= \inf_{\theta \in \Theta(\lambda)} Q_n(\theta) \\ &= \big[\sqrt{n}\hat{\sigma}^{-1}_{n,1}\bar{W}_{n,1}\big]^2_- + \big[\sqrt{n}\hat{\sigma}^{-1}_{n,2}\bar{W}_{n,2}\big]^2_- + \big(\sqrt{n}\hat{\sigma}^{-1}_{n,3}\bar{W}_{n,3}\big)^2 \xrightarrow{d} (Z_3 + \mu_3)^2. \end{aligned}$$

Test MR Consider the approximations leading to Test MR. The discard approximation takes the form

$$\begin{aligned} T^{\mathrm{DR}}_n(\lambda) &= \big[v^*_{n,1} + \varphi_1\big(\kappa^{-1}_n\sqrt{n}\hat{\sigma}^{-1}_{n,1}\bar{W}_{n,1}\big)\big]^2_- \\ &\quad + \big[v^*_{n,2} + \varphi_2\big(\kappa^{-1}_n\sqrt{n}\hat{\sigma}^{-1}_{n,2}\bar{W}_{n,2}\big)\big]^2_- + \big(v^*_{n,3}\big)^2, \end{aligned}$$

since $\hat{\Theta}_I(\lambda) = \{(0, 0, 0)\}$. Using that $\mu_1 > 1$ and $\mu_2 < 1$ (which imply $\varphi_1 \to \infty$ and $\varphi_2 \to 0$), it follows that

$$\big\{T^{\mathrm{DR}}_n(\lambda) | \{W_i\}^n_{i=1}\big\} \xrightarrow{d} [Z_2]^2_- + (Z_3)^2 \quad \text{w.p.1.}$$

The penalize approximation takes the form

$$\begin{aligned} T^{\mathrm{PR}}_n(\lambda) &= \big[v^*_{n,1} + \kappa^{-1}_n\sqrt{n}\hat{\sigma}^{-1}_{n,1}\bar{W}_{n,1}\big]^2_- \\ &\quad + \big[v^*_{n,2} + \kappa^{-1}_n\sqrt{n}\hat{\sigma}^{-1}_{n,2}\bar{W}_{n,2}\big]^2_- + \big(v^*_{n,3} + \kappa^{-1}_n\sqrt{n}\hat{\sigma}^{-1}_{n,3}\bar{W}_{n,3}\big)^2, \end{aligned}$$

since $\Theta(\lambda) = \{(0, 0, 0)\}$. Simple algebra shows that

$$\big\{T^{\mathrm{PR}}_n(\lambda) | \{W_i\}^n_{i=1}\big\} \xrightarrow{d} [Z_1 + \mu_1]^2_- + [Z_2 + \mu_2]^2_- + (Z_3)^2 \quad \text{w.p.1.}$$

Putting all these results together shows that

$$\big\{T^{\mathrm{MR}}_n(\lambda) | \{W_i\}^n_{i=1}\big\} \xrightarrow{d} \min\big\{[Z_1 + \mu_1]^2_- + [Z_2 + \mu_2]^2_-, [Z_2]^2_-\big\} + (Z_3)^2 \quad \text{w.p.1.}$$

The example provides important lessons about the relative power of all these tests. To see this, note that

$$\begin{aligned} &P\big([Z_1 + \mu_1]^2_- + [Z_2 + \mu_2]^2_- < [Z_2]^2_-\big) \ge P(Z_1 + \mu_1 \ge 0)P(Z_2 < 0) > 0, \\ &P\big([Z_1 + \mu_1]^2_- + [Z_2 + \mu_2]^2_- > [Z_2]^2_-\big) \ge P(Z_1 + \mu_1 < 0)P(Z_2 \ge 0) > 0, \end{aligned} \tag{4.18}$$

which implies that whether $T^{\mathrm{MR}}_n(\lambda)$ equals $T^{\mathrm{DR}}_n(\lambda)$ or $T^{\mathrm{PR}}_n(\lambda)$ is random, conditionally on $\{W_i\}^n_{i=1}$. This means that using Test MR is not equivalent to using either $\phi^{\mathrm{DR}}_n(\lambda)$ in (4.10) or $\phi^{\mathrm{PR}}_n(\lambda)$ in (4.14).

Example 4.1 and (4.18) also show that the conditional distribution of $T^{\mathrm{MR}}_n(\lambda)$ is (asymptotically) strictly first order stochastically dominated by the conditional distributions of $T^{\mathrm{DR}}_n(\lambda)$ or $T^{\mathrm{PR}}_n(\lambda)$. Since all these tests reject for large values of $T_n(\lambda)$, their relative asymptotic power depends on the limit of their respective critical values. In the example above, we can numerically compute the $1-\alpha$ quantiles of the limit distributions of $T_n^{\mathrm{DR}}(\lambda)$, $T_n^{\mathrm{PR}}(\lambda)$, and $T_n^{\mathrm{MR}}(\lambda)$ after fixing some values for μ_1 and μ_2. Setting both of these parameters close to 1 results in asymptotic 95% quantiles of $T_n^{\mathrm{DR}}(\lambda)$, $T_n^{\mathrm{PR}}(\lambda)$, and $T_n^{\mathrm{MR}}(\lambda)$ equal to 5.15, 4.18, and 4.04, respectively.

Remark 4.11. Example 4.1 illustrates that the two resampling approximations leading to Test MR do not dominate each other in terms of asymptotic power. For example, if we consider the model in (4.17) with the second inequality removed, it follows that

$$T_n^{\mathrm{DR}}(\lambda) \xrightarrow{d} (Z_3)^2 \quad \text{and} \quad T_n^{\mathrm{PR}}(\lambda) \xrightarrow{d} [Z_1+\mu_1]_-^2 + (Z_3)^2.$$

In this case $\phi_n^{\mathrm{DR}}(\lambda)$ has strictly better asymptotic power than $\phi_n^{\mathrm{PR}}(\lambda)$: taking μ_1 close to 1 gives asymptotic 95% quantiles of $\phi_n^{\mathrm{DR}}(\lambda)$ and $\phi_n^{\mathrm{PR}}(\lambda)$ equal to 3.84 and 4.00, respectively. On the other hand, if we consider the model in (4.17) with the first inequality removed, it follows that

$$T_n^{\mathrm{DR}}(\lambda) \xrightarrow{d} [Z_2]_-^2 + (Z_3)^2 \quad \text{and} \quad T_n^{\mathrm{PR}}(\lambda) \xrightarrow{d} [Z_2+\mu_2]_-^2 + (Z_3)^2.$$

Since $[Z_2+\mu_2]_-^2 \le [Z_2]_-^2$ (with strict inequality when $Z_2<0$), this case represents a situation where $\phi_n^{\mathrm{DR}}(\lambda)$ has strictly worse asymptotic power than $\phi_n^{\mathrm{PR}}(\lambda)$: taking μ_2 close to 1 results in asymptotic 95% quantiles of $\phi_n^{\mathrm{DR}}(\lambda)$ and $\phi_n^{\mathrm{PR}}(\lambda)$ equal to 5.13 and 4.00, respectively.

Example 4.2. Let $\{W_i\}_{i=1}^n = \{(W_{1,i}, W_{2,i}, W_{3,i})\}_{i=1}^n$ be an i.i.d. sequence of random variables with distribution F_n, $V_{F_n}[W]=I_3$, $E_{F_n}[W_1]=\mu_1\kappa_n/\sqrt{n}$, $E_{F_n}[W_2]=\mu_2/\sqrt{n}$, and $E_{F_n}[W_3]=0$ for some $\mu_1\ge 0$ and $\mu_2\le 0$. Consider the model in (4.17) with $\Theta=[-1,1]^3$ and the hypotheses

$$H_0:\lambda(\theta)=(\theta_1,\theta_2)=(0,0) \quad \text{vs.} \quad H_1:\lambda(\theta)=(\theta_1,\theta_2)\neq(0,0).$$

In this case $\Theta(\lambda)=\{(0,0,\theta_3):\theta_3\in[-1,1]\}$ and H_0 is true if and only if $\mu_2=0$. The model in (4.17) is linear in θ and so several expressions do not depend on θ. These include $\hat\sigma_{n,j}(\theta)=\hat\sigma_{n,j}$ and $v_{n,j}^*(\theta)=v_{n,j}^*$ for $j=1,2,3$, where $v_{n,j}^*(\theta)$ is defined in (2.8). As in Section 3, we use the MMM test statistic in (2.6) and the GMS function in (2.11). Below we also use $Z=(Z_1,Z_2,Z_3)\sim N(\mathbf{0}_3,I_3)$.

Simple algebra shows that the test statistic satisfies

$$\begin{aligned}
T_n(\lambda) &= \inf_{\theta\in\Theta(\lambda)} Q_n(\theta) \\
&= \inf_{\theta_3\in[-1,1]} \big[\sqrt{n}\hat\sigma_{n,1}^{-1}\bar W_{n,1}\big]_-^2 + \big[\sqrt{n}\hat\sigma_{n,2}^{-1}\bar W_{n,2}\big]_-^2 + \big(\sqrt{n}\hat\sigma_{n,3}^{-1}(\bar W_{n,3}-\theta_3)\big)^2 \\
&\xrightarrow{d} [Z_1]_-^2 1\{\mu_1=0\} + [Z_2+\mu_2]_-^2,
\end{aligned}$$

where we used $\Theta(\lambda)=\{(0,0,\theta_3):\theta_3\in[-1,1]\}$.

Penalize resampling test This test uses the (conditional) $(1-\alpha)$ quantile of

$$T_n^{\mathrm{PR}}(\lambda) = \inf_{\theta_3 \in [-1,1]} \big\{\big[v_{n,1}^* + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,1}^{-1}\bar{W}_{n,1}\big]_-^2 + \big[v_{n,2}^* + \kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,2}^{-1}\bar{W}_{n,2}\big]_-^2 + \big(\sqrt{n}\hat{\sigma}_{n,3}^{-1}(\bar{W}_{n,3} - \theta_3)\big)^2\big\},$$

where we used $\Theta(\lambda) = \{(0,0,\theta_3) : \theta_3 \in [-1,1]\}$. Simple arguments show that

$$\big\{T_n^{\mathrm{PR}}(\lambda) | \{W_i\}_{i=1}^n\big\} \xrightarrow{d} [Z_1 + \mu_1]_-^2 + [Z_2]_-^2 \quad \text{w.p.1.}$$

Test SS This test draws $\{W_i^{\mathrm{SS}}\}_{i=1}^{b_n}$ i.i.d. with replacement from $\{W_i\}_{i=1}^n$ and computes $v_n^{\mathrm{SS}}(\theta) \equiv \frac{1}{\sqrt{b_n}}\sum_{i=1}^{b_n} \hat{D}_n^{\mathrm{SS},-1/2}(\theta)m(W_i^{\mathrm{SS}},\theta)$, where $\hat{D}_n^{\mathrm{SS}}(\theta)$ is as $\hat{D}_n(\theta)$ but based on $\{W_i^{\mathrm{SS}}\}_{i=1}^{b_n}$. Letting

$$\tilde{v}_n^{\mathrm{SS}}(\theta) \equiv \frac{1}{\sqrt{b_n}}\sum_{i=1}^{b_n} \hat{D}_n^{\mathrm{SS},-1/2}(\theta)\big\{m\big(W_i^{\mathrm{SS}},\theta\big) - E_{F_n}\big[m(W_i,\theta)\big]\big\}$$

and noting that $\tilde{v}_n^{\mathrm{SS}}(\theta) = \tilde{v}_n^{\mathrm{SS}}$, Politis, Romano, and Wolf (1999, Theorem 2.2.1) implies that $\{\tilde{v}_n^{\mathrm{SS}} | \{W_i\}_{i=1}^n\} \xrightarrow{d} N(0,1)$ almost surely (a.s.).

Test SS uses the conditional $(1-\alpha)$ quantile of the random variable

$$T_n^{\mathrm{SS}}(\lambda) = \inf_{\theta_3 \in [-1,1]} \big\{\big[\sqrt{b_n}\hat{\sigma}_{n,1}^{\mathrm{SS},-1}\bar{W}_{n,1}^{\mathrm{SS}}\big]_-^2 + \big[\sqrt{b_n}\hat{\sigma}_{n,2}^{\mathrm{SS},-1}\bar{W}_{n,2}^{\mathrm{SS}}\big]_-^2 + \big(\sqrt{b_n}\hat{\sigma}_{n,3}^{\mathrm{SS},-1}\big(\bar{W}_{n,3}^{\mathrm{SS}} - \theta_3\big)\big)^2\big\},$$

where we used $\Theta(\lambda) = \{(0,0,\theta_3) : \theta_3 \in [-1,1]\}$. Simple arguments show that

$$\big\{T_n^{\mathrm{SS}}(\lambda) | \{W_i\}_{i=1}^n\big\} \xrightarrow{d} [Z_1 + K\mu_1]_-^2 + [Z_2]_-^2 \quad \text{w.p.1,}$$

where, for simplicity, we assume that $\kappa_n\sqrt{b_n/n} \to K$.

Remark 4.12. In Example 4.2, $T_n^{\mathrm{PR}}(\lambda)$ and $T_n^{\mathrm{SS}}(\lambda)$ have the same asymptotic distribution, conditionally on $\{W_i\}_{i=1}^n$, when $\mu_1 = 0$ or $K = 1$. However, if $\mu_1 > 0$ and $K < 1$, it follows that $T_n^{\mathrm{PR}}(\lambda)$ is (asymptotically) strictly first order stochastically dominated by $T_n^{\mathrm{SS}}(\lambda)$, conditionally on $\{W_i\}_{i=1}^n$. Specifically,

$$P\big([Z_2 + \mu_2]_-^2 > q_{1-\alpha}\big([Z_1 + \mu_1]_-^2 + [Z_2]_-^2\big)\big) > P\big([Z_2 + \mu_2]_-^2 > q_{1-\alpha}\big([Z_1 + K\mu_1]_-^2 + [Z_2]_-^2\big)\big),$$

where $q_{1-\alpha}(X)$ denotes the $1-\alpha$ quantile of X. Thus, Test MR is *strictly* less conservative under H_0 (i.e., when $\mu_2 = 0$) and *strictly* more powerful under H_1 (i.e., when $\mu_2 < 0$).

Remark 4.13. Example 4.2 shows that Test SS could deliver higher power than $\phi_n^{\mathrm{PR}}(\lambda)$ if $\mu_1 > 0$ and $K > 1$, that is, if Assumption A.4 is violated. However, for the recommended choice of $\kappa_n = \sqrt{\ln n}$ in Andrews and Soares (2010, p. 131), a violation of this assumption can result in Test SS having poor finite sample power properties, as already discussed in Remark 4.8.

5. Monte Carlo simulations

In this section we consider an entry game model similar to that in Canay (2010) with the addition of market-type fixed effects. Consider a firm $j \in \{1, 2\}$ deciding whether to enter $(A_{j,i} = 1)$ a market $i \in \{1, \dots, n\}$ or not $(A_{j,i} = 0)$ based on its profit function

$$\pi_{j,i} = \left(\varepsilon_{j,i} - \theta_j A_{-j,i} + \sum_{q=0}^{d_X} \beta_q X_{q,i}\right) 1\{A_{j,i} = 1\},$$

where $\varepsilon_{j,i}$ is firm j's benefit of entry in market i, $A_{-j,i}$ is the decision of the rival firm, and $X_{q,i}$, $q \in \{0, \dots, d_X\}$, are observed market type indicators with distribution $P(X_{q,i} = 1) = p_q$ (assumed to be known for simplicity). We normalize (p_0, β_0) to $(1, 0)$ and let $\varepsilon_{j,i} \sim \text{Uniform}(0, 1)$ conditional on all market characteristics. We also assume that the parameter space for the vector $\theta = (\theta_1, \theta_2, \beta_1, \dots, \beta_{d_X})$ is

$$\begin{aligned} \Theta = \big\{ &\theta \in \mathbb{R}^{d_X + 2} : (\theta_1, \theta_2) \in (0, 1)^2 \text{ and} \\ &\beta_q \in \big[0, \min\{\theta_1, \theta_2\}\big] \text{ for all } q \in \{1, \dots, d_X\}\big\}. \end{aligned}$$

This space guarantees that there are three pure strategy Nash equilibria (NE), conditional on a given market type. To be clear, the four possible outcomes in market q are (i) $A_i \equiv (A_{1,i}, A_{2,i}) = (1, 1)$ is the unique NE if $\varepsilon_{j,i} > \theta_j - \beta_q$ for all j, (ii) $A_i = (1, 0)$ is the unique NE if $\varepsilon_{1,i} > \theta_1 - \beta_q$ and $\varepsilon_{2,i} < \theta_2 - \beta_q$, (iii) $A_i = (0, 1)$ is the unique NE if $\varepsilon_{1,i} < \theta_1 - \beta_q$ and $\varepsilon_{2,i} > \theta_2 - \beta_q$, and (iv) there are multiple equilibria if $\varepsilon_{j,i} < \theta_j - \beta_q$ for all j as both $A_i = (1, 0)$ and $A_i = (0, 1)$ are NE. Without further assumptions, this model implies

$$\begin{aligned} E_F\big[m_{1,q}(W_i, \theta)\big] &= E_F\big[A_{1,i}A_{2,i}X_{q,1}/p_q - (1 - \theta_1 + \beta_q)(1 - \theta_2 + \beta_q)\big] = 0, \\ E_F\big[m_{2,q}(W_i, \theta)\big] &= E_F\big[A_{1,i}(1 - A_{2,i})X_{q,1}/p_q - (\theta_2 - \beta_q)(1 - \theta_1 + \beta_q)\big] \\ &\geq 0, \\ E_F\big[m_{3,q}(W_i, \theta)\big] &= E_F\big[\theta_2 - \beta_q - A_{1,i}(1 - A_{2,i})X_{q,1}/p_q\big] \geq 0, \end{aligned} \tag{5.1}$$

where $W_i \equiv (A_i, X_i)$, $X_i \equiv (X_{0,i}, \dots, X_{d_X,i})$, and $q \in \{0, \dots, d_X\}$. This model delivers $d_X + 1$ unconditional moment equalities and $2(d_X + 1)$ unconditional moment inequalities. The identified set for (θ_1, θ_2), which are the parameters of interest, is a curve in $\mathbb{R}^2$ and the shape of this curve depends on the values of the nuisance parameters $(\beta_1, \dots, \beta_{d_X})$.

We generate data using $d_X = 3$, $(\theta_1, \theta_2, \beta_1, \beta_2, \beta_3) = (0.3, 0.5, 0.05, 0, 0)$, and $\delta = 0.6$, where δ is the probability of selecting $A_i = (1, 0)$ in the region of multiple equilibria. The identified set for each coordinate of $(\theta_1, \theta_2, \beta_1, \beta_2, \beta_3)$ is given by

$$\begin{aligned} &\theta_1 \in [0.230, 0.360], \qquad \theta_2 \in [0.455, 0.555], \\ &\beta_1 \in [0.0491, 0.0505], \quad \text{and} \quad \beta_2 = \beta_3 = 0. \end{aligned}$$

Having a five dimensional parameter θ already presents challenges for projection-based tests and represents a case of empirical relevance (e.g., see Dickstein and Morales (2015) and Wollmann (2015)). For example, a grid with 100 points in the $(0, 1)$ interval for each element in θ (imposing the restrictions in Θ for $(\beta_1, \beta_2, \beta_3)$) involves 1,025 million evaluations of test statistics and critical values. This is costly for doing Monte Carlo simulations so we do not include Test BP in this exercise. However, in simulations not reported for the case where $d_X = 0$, Test BP is always dominated by Test MR in terms of size control and power.

We set $n = 1{,}000$ and $\alpha = 0.10$, and simulate the data by taking independent draws of $\varepsilon_{j,i} \sim \text{Uniform}(0, 1)$ for $j \in \{1, 2\}$ and computing the equilibrium according to the region in which $\varepsilon_i \equiv (\varepsilon_{1,i}, \varepsilon_{2,i})$ falls. We consider subvector inference for this model, with

$$H_0 : \lambda(\theta) = \theta_1 = \lambda_0 \quad \text{vs.} \quad H_1 : \lambda(\theta) = \theta_1 \neq \lambda_0,$$

and perform MC = 2,000 Monte Carlo replications. We report results for Test MR1 (Test MR with $\kappa_n = \sqrt{\ln n} = 2.63$ as recommended by Andrews and Soares (2010)), Test MR2 (Test MR with $\kappa_n = 0.8\sqrt{\ln n} = 2.10$), Test SS1 (Test SS with $b_n = n^{2/3} = 100$ as considered in Bugni (2010, 2016)), and Test SS2 (Test SS with $b_n = n/4 = 250$ as considered in Ciliberto and Tamer (2009)).[4]

Figure 1 shows the rejection probabilities under the null and alternative hypotheses for the first coordinate, that is, $\lambda(\theta) = \theta_1$. Since there are sharp differences in the behavior of the tests to the right and left of the identified set (due to the asymmetric nature of the model), we comment on each direction separately. Let us first focus on values of θ_1

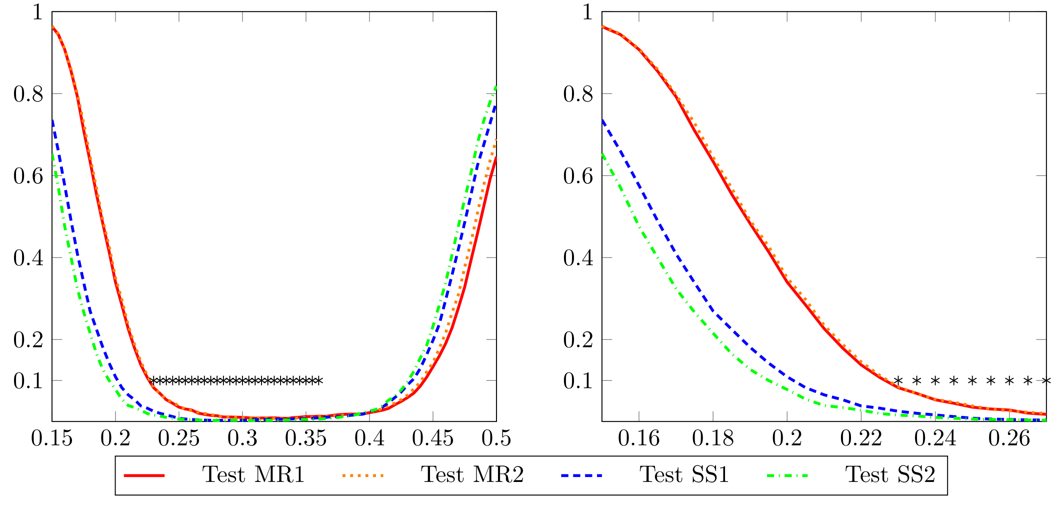

FIGURE 1. Rejection probabilities under the null and alternative hypotheses when $\lambda(\theta) = \theta_1$. Tests considered are Test MR1 (solid line), Test MR2 (dotted line), Test SS1 (dashed line), and Test SS2 (dashed–dotted line). Asterisks indicate values of θ_1 in the identified set at the nominal level. The left panel shows rejection rates to the left and right of the identified set. The right panel zooms-in the power differences to the left. In all cases $n = 1{,}000$, $\alpha = 0.10$, and MC = 2,000.

[4]The choice $b_n = n^{2/3}$ corresponds to the optimal rate for the subsample size to minimize ERP; see Bugni (2010, 2016). The choice $b_n = n/4$ is the subsample size rule used by Ciliberto and Tamer (2009).

below the boundary point 0.23. The null rejection probabilities at this boundary point are close to the nominal level of $\alpha = 0.10$ for Test MR (regardless of the choice of κ_n), but below 0.025 for Tests SS1 and SS2 (note that the simulation standard error with 2,000 replications is 0.007). In addition, the power of Test MR in this direction is higher than that of Tests SS1 and SS2, with an average difference of 0.27 and a maximum difference of 0.39 with respect to Test SS1, which exhibits higher power than Test SS2. The right panel illustrates the power differences more clearly. Now focus on values of θ_1 above the boundary point 0.36. In this case all tests have null rejection probabilities below the nominal level, with rejection probabilities equal to 0.013, 0.012, 0.009, and 0.006 for Tests MR1, MR2, SS1, and SS2, respectively. When we look at power in this direction, Test SS2 delivers the highest power, with an average difference of 0.05 and a maximum difference of 0.17 with respect to Test MR2. All in all, the results from this simulation exercise illustrate that there are cases where Test SS may deliver higher power than Test MR (possibly due to the sufficient conditions in Theorem 4.3 not holding). At the same time, the results also suggest that the power gains delivered by Test MR could be significantly higher.

6. Concluding remarks

This paper introduces a test for the null hypothesis $H_0 : \lambda(\theta) = \lambda_0$, where $\lambda(\cdot)$ is a known function, λ_0 is a known constant, and θ is a parameter that is partially identified by a moment (in)equality model. The test can be inverted to construct CS's for $\lambda(\theta)$. The leading application of our inference method is the construction of marginal CS's for individual coordinates of a parameter vector θ, which is implemented by setting $\lambda(\theta) = \theta_s$ for $s \in \{1, \dots, d_\theta\}$ and collecting all values of $\lambda_0 \in \Lambda$ for which H_0 is not rejected.

We show that our inference method controls asymptotic size uniformly over a large class of distributions of the data. The current literature describes only two other procedures that deliver uniform size control for these types of problems: projection-based and subsampling inference. Relative to the projection-based procedure, our method presents three advantages: (i) it weakly dominates in terms of finite sample power, (ii) it strictly dominates in terms of asymptotic power, and (iii) it may be less computationally demanding. Relative to a subsampling, our method presents two advantages: (i) it strictly dominates in terms of asymptotic power under certain conditions and (ii) it appears to be less sensitive to the choice of its tuning parameter than subsampling is to the choice of subsample size.

Moving forward, there are some open questions that are worth highlighting. First, our paper does not cover conditional moment restrictions (cf. Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Armstrong (2014), and Chetverikov (2013)). Second, alternative asymptotic frameworks like those in Andrews and Barwick (2012) and Romano, Shaikh, and Wolf (2014) may provide a better asymptotic approximation for the type of problems we study in this paper. Finally, although our method eases the computational burden for some problems, further improvements in this dimension may be available. Recent work by Kaido, Molinari, and Stoye (2016), who propose a novel projection-based method to build CS's for single components of θ, explores innovative ideas to increase computational tractability that may also prove useful for profiling-based tests.

Appendix A: Notation

Throughout the appendix, the lemmas, equations, and theorems labeled with the letter "S" belong to the Supplemental Material. We also employ the notation in Table 1, which was not necessarily introduced in the text.

For any $u \in \mathbb{N}$, $\mathbf{0}_u$ is a column vector of zeros of size u, $\mathbf{1}_u$ is a column vector of 1s of size u, and I_u is the $u \times u$ identity matrix. We use $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$, $\mathbb{R}_+ = \mathbb{R}_{++} \cup \{0\}$, $\mathbb{R}_{+,\infty} = \mathbb{R}_+ \cup \{+\infty\}$, $\mathbb{R}_{[+\infty]} = \mathbb{R} \cup \{+\infty\}$, and $\mathbb{R}_{[\pm\infty]} = \mathbb{R} \cup \{\pm\infty\}$. We equip $\mathbb{R}^u_{[\pm\infty]}$ with the following metric d. For any $x_1, x_2 \in \mathbb{R}^u_{[\pm\infty]}$, $d(x_1, x_2) = (\sum_{i=1}^u (G(x_{1,i}) - G(x_{2,i}))^2)^{1/2}$, where $G : \mathbb{R}_{[\pm\infty]} \to [0, 1]$ is such that $G(-\infty) = 0$, $G(\infty) = 1$, and $G(y) = \Phi(y)$ for $y \in \mathbb{R}$, where Φ is the standard normal cumulative distribution function (CDF). Also, $1\{\cdot\}$ denotes the indicator function.

Let $\mathcal{C}(\Theta^2)$ denote the space of continuous functions that map Θ^2 to Ψ, where Ψ is the space of $k \times k$ correlation matrices, and $\mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})$ denotes the space of compact subsets of the metric space $(\Theta \times \mathbb{R}^k_{[\pm\infty]}, d)$. Let d_H denote the Hausdorff metric associated with d. We use $\overset{H}{\to}$ to denote convergence in the Hausdorff metric, that is, $A_n \overset{H}{\to} B \iff d_H(A_n, B) \to 0$. For nonstochastic functions of $\theta \in \Theta$, we use $\overset{u}{\to}$ to denote uniform in θ convergence, for example, $\Omega_{F_n} \overset{u}{\to} \Omega \iff \sup_{\theta,\theta' \in \Theta} d(\Omega_{F_n}(\theta, \theta'), \Omega(\theta, \theta')) \to 0$. Finally, we use $\Omega(\theta)$ and $\Omega(\theta, \theta)$ equivalently.

We denote by $l^\infty(\Theta)$ the set of all uniformly bounded functions that map $\Theta \to \mathbb{R}^u$, equipped with the supremum norm. The sequence of distributions $\{F_n \in \mathcal{P}\}_{n\geq 1}$ determines a sequence of probability spaces $\{(\mathcal{W}, \mathcal{A}, F_n)\}_{n\geq 1}$. Stochastic processes are then random maps $X : \mathcal{W} \to l^\infty(\Theta)$. In this context, we use $\overset{d}{\to}$, $\overset{p}{\to}$, and $\overset{\text{a.s.}}{\to}$ to denote weak

Table 1. Important notation.

$\mathcal{P}_0$	$\{F \in \mathcal{P} : \Theta_I(F) \neq \emptyset\}$
$\Sigma_F(\theta)$	$\mathrm{Var}_F(m(W, \theta))$
$D_F(\theta)$	$\mathrm{diag}(\Sigma_F(\theta))$
$Q_F(\theta)$	$S(E_F[m(W, \theta)], \Sigma_F(\theta))$
$\Theta_I^{\ln \kappa_n}(F)$	$\{\theta \in \Theta : S(\sqrt{n}E_F[m(W, \theta)], \Sigma_F(\theta)) \leq \ln \kappa_n\}$
$\Theta_I(F, \lambda)$	$\Theta(\lambda) \cap \Theta_I(F)$
$\Gamma_{n,F}(\lambda)$	$\{(\theta, \ell) \in \Theta(\lambda) \times \mathbb{R}^k : \ell = \sqrt{n}D_F^{-1/2}(\theta)E_F[m(W_i, \theta)]\}$
$\Gamma^{\mathrm{SS}}_{b_n,F}(\lambda)$	$\{(\theta, \ell) \in \Theta(\lambda) \times \mathbb{R}^k : \ell = \sqrt{b_n}D_F^{-1/2}(\theta)E_F[m(W, \theta)]\}$
$\Gamma^{\mathrm{PR}}_{n,F}(\lambda)$	$\{(\theta, \ell) \in \Theta(\lambda) \times \mathbb{R}^k : \ell = \kappa_n^{-1}\sqrt{n}D_F^{-1/2}(\theta)E_F[m(W_i, \theta)]\}$
$\Gamma^{\mathrm{DR}}_{n,F}(\lambda)$	$\{(\theta, \ell) \in \Theta(\lambda) \cap \Theta_I^{\ln \kappa_n}(F) \times \mathbb{R}^k : \ell = \kappa_n^{-1}\sqrt{n}D_F^{-1/2}(\theta)E_F[m(W, \theta)]\}$
$v_{n,j}(\theta)$	$n^{-1/2}\sigma_{F,j}^{-1}(\theta)\sum_{i=1}^n (m_j(W_i, \theta) - E_F[m_j(W_i, \theta)])$, $j = 1, \ldots, k$
$\Omega_F(\theta, \theta')_{[j_1,j_2]}$	$E_F[(\frac{m_{j_1}(W,\theta) - E_F[m_{j_1}(W,\theta)]}{\sigma_{F,j_1}(\theta)})(\frac{m_{j_2}(W,\theta') - E_F[m_{j_2}(W,\theta')]}{\sigma_{F,j_2}(\theta')})]$

convergence, convergence in probability, and convergence almost surely in the $l^\infty(\Theta)$ metric, respectively, in the sense of van der Vaart and Wellner (1996). In addition, for every $F \in \mathcal{P}$, we use $\mathcal{M}(F) \equiv \{D_F^{-1/2}(\theta)m(\cdot,\theta) : \mathcal{W} \to \mathbb{R}^k\}$ and denote by ρ_F the coordinate-wise version of the "intrinsic" variance semimetric, that is,

$$\rho_F(\theta,\theta') \equiv \left\| \left\{ V_F\left[\sigma_{F,j}^{-1}(\theta)m_j(W,\theta) - \sigma_{F,j}^{-1}(\theta')m_j(W,\theta')\right]^{1/2}\right\}_{j=1}^k \right\|. \tag{A.1}$$

Appendix B: Assumptions

B.1 *Assumptions for asymptotic size control*

Assumption A.1. *Given the GMS function $\varphi(\cdot)$, there is a function $\varphi^* : \mathbb{R}^k_{[\pm\infty]} \to \mathbb{R}^k_{[+\infty]}$ that takes the form $\varphi^*(\xi) = (\varphi_1^*(\xi_1), \dots, \varphi_p^*(\xi_p), \mathbf{0}_{k-p})$ and, for all $j = 1, \dots, p$,*

(a) *$\varphi_j^*(\xi_j) \geq \varphi_j(\xi_j)$ for all $\xi_j \in \mathbb{R}_{[+\infty]}$,*

(b) *$\varphi_j^*(\cdot)$ is continuous,*

(c) *$\varphi_j^*(\xi_j) = 0$ for all $\xi_j \leq 0$ and $\varphi_j^*(\infty) = \infty$.*

Remark B.1. Assumption A.1 is satisfied when φ is any of the functions $\varphi^{(1)}$–$\varphi^{(4)}$ described in Andrews and Soares (2010) or Andrews and Barwick (2012). This follows from Bugni, Canay, and Shi (2015, Lemma D.8).

Assumption A.2. *For any $\{(\lambda_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$, let (Γ, Ω) be such that $\Omega_{F_n} \xrightarrow{u} \Omega$ and $\Gamma_{n,F_n}(\lambda_n) \xrightarrow{H} \Gamma$ with $(\Gamma, \Omega) \in \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]}) \times \mathcal{C}(\Theta^2)$ and $\Gamma_{n,F_n}(\lambda_n)$ as in Table 1. Let $c_{(1-\alpha)}(\Gamma, \Omega)$ be the $(1-\alpha)$ quantile of*

$$J(\Gamma,\Omega) \equiv \inf_{(\theta,\ell)\in\Gamma} S\big(v_\Omega(\theta) + \ell, \Omega(\theta)\big). \tag{B.1}$$

Then the following statements hold:

(a) *If $c_{(1-\alpha)}(\Gamma,\Omega) > 0$, the distribution of $J(\Gamma,\Omega)$ is continuous at $c_{(1-\alpha)}(\Gamma,\Omega)$.*

(b) *If $c_{(1-\alpha)}(\Gamma,\Omega) = 0$, $\liminf_{n\to\infty} P_{F_n}(T_n(\lambda_n) = 0) \geq 1-\alpha$, where $T_n(\lambda_n)$ is as in (4.1).*

Remark B.2. Without Assumption A.2 the asymptotic distribution of the test statistic could be discontinuous at the probability limit of the critical value, resulting in asymptotic overrejection under the null hypothesis. One could add an infinitesimal constant to the critical value and avoid introducing such assumption, but this introduces an additional tuning parameter that needs to be chosen by the researcher. Note that this assumption holds in Examples 4.1 and 4.2 where $J(\cdot)$ is continuous at $x \in \mathbb{R}$. Also notice that $c_{(1-\alpha)}(\Gamma,\Omega) = 0$ implies $P(J(\Gamma,\Omega) = 0) \geq 1-\alpha$. Thus, the requirement for $c_{(1-\alpha)}(\Gamma,\Omega) = 0$ is automatically satisfied whenever $P_{F_n}(T_n(\lambda_n) = 0) \to P(J(\Gamma,\Omega) = 0)$.

Assumption A.3. *The following conditions hold.*

(a) *For all $(\lambda, F) \in \mathcal{L}_0$ and $\theta \in \Theta(\lambda)$, $Q_F(\theta) \geq c\min\{\delta, \inf_{\tilde\theta\in\Theta(\lambda)\cap\Theta_I(F)} \|\theta - \tilde\theta\|\}^\chi$ for constants $c, \delta > 0$ and for χ as in Assumption M.1.*

(b) *We have that* $\Theta(\lambda)$ *is convex.*

(c) *The function* $g_F(\theta) \equiv D_F^{-1/2}(\theta)E_F[m(W,\theta)]$ *is differentiable in* θ *for any* $F \in \mathcal{P}_0$, *and the class of functions* $\{G_F(\theta) \equiv \partial g_F(\theta)/\partial\theta' : F \in \mathcal{P}_0\}$ *is equicontinuous, that is,*

$$\lim_{\delta\to 0} \sup_{F\in\mathcal{P}_0,(\theta,\theta'):\|\theta-\theta'\|\leq\delta} \|G_F(\theta) - G_F(\theta')\| = 0.$$

Remark B.3. Assumption A.3(a) states that $Q_F(\theta)$ can be bounded below in a neighborhood of the null identified set $\Theta(\lambda) \cap \Theta_I(F)$ and so it is analogous to the polynomial minorant condition in Chernozhukov, Hong, and Tamer (2007, Eqs. (4.1) and (4.5)). The convexity in Assumption A.3(b) follows from a convex parameter space Θ and a linear function $\lambda(\cdot)$ in the case of the null in (1.2). However, in one-sided testing problems like those described in Remark 4.2, this assumption holds for quasi-convex functions. Finally, Assumption A.3(c) is a smoothness condition that would be implied by the class of functions $\{G_F(\theta) \equiv \partial g_F(\theta)/\partial\theta' : F \in \mathcal{P}_0\}$ being Lipschitz. These three parts are sufficient conditions for our test to be asymptotically valid (see Lemmas S.3.7 and S.3.8).

B.2 *Assumptions for asymptotic power*

Assumption A.4. *The sequences* $\{\kappa_n\}_{n\geq 1}$ *and* $\{b_n\}_{n\geq 1}$ *in Assumption* M.2 *satisfy* $\limsup_{n\to\infty} \kappa_n\sqrt{b_n/n} \leq 1$.

Remark B.4. Assumption A.4 is a weaker version of Andrews and Soares (2010, Assumption GMS5) and it holds for all typical choices of κ_n and b_n. For example, it holds if we use the recommended choice of $\kappa_n = \sqrt{\ln n}$ in Andrews and Soares (2010, p. 131) and $b_n = n^c$ for any $c \in (0,1)$. Note that the latter includes as a special case $b_n = n^{2/3}$, which has been shown to be optimal according to the rate of convergence of the error in the coverage probability (see Politis and Romano (1994), Bugni (2010, 2016)).

Assumption A.5. *For* $\lambda_0 \in \Lambda$, *there is* $\{\lambda_n \in \Lambda\}_{n\geq 1}$ *such that* $\{(\lambda_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ *satisfies the following statements*:

(a) *For all* $n \in \mathbb{N}$, $\Theta_I(F_n) \cap \Theta(\lambda_0) = \emptyset$ (*i.e.*, $(\lambda_0, F_n) \notin \mathcal{L}_0$).

(b) *We have* $d_H(\Theta(\lambda_n), \Theta(\lambda_0)) = O(n^{-1/2})$.

(c) *For any* $\theta \in \Theta$, $G_{F_n}(\theta) = O(1)$.

Assumption A.6. *For* $\lambda_0 \in \Lambda$ *and* $\{\lambda_n \in \Lambda\}_{n\geq 1}$ *as in Assumption* A.5, *let* $(\Gamma, \Gamma^{\mathrm{SS}}, \Gamma^{\mathrm{PR}}, \Omega) \in \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})^3 \times \mathcal{C}(\Theta^2)$ *be such that* $\Omega_{F_n} \xrightarrow{u} \Omega$, $\Gamma_{n,F_n}(\lambda_0) \xrightarrow{H} \Gamma$, $\Gamma^{\mathrm{PR}}_{n,F_n}(\lambda_0) \xrightarrow{H} \Gamma^{\mathrm{PR}}$, $\Gamma^{\mathrm{SS}}_{b_n,F_n}(\lambda_0) \xrightarrow{H} \Gamma^{\mathrm{SS}}$ *for* $\Gamma_{n,F_n}(\lambda_0)$, $\Gamma^{\mathrm{PR}}_{n,F_n}(\lambda_0)$, *and* $\Gamma^{\mathrm{SS}}_{b_n,F_n}(\lambda_0)$ *as in Table* 1. *Then the following statements hold*:

(a) *The distribution of* $J(\Gamma, \Omega)$ *is continuous at* $c_{1-\alpha}(\Gamma^{\mathrm{SS}}, \Omega)$.

(b) *The distributions of* $J(\Gamma, \Omega)$, $J(\Gamma^{\mathrm{SS}}, \Omega)$, *and* $J(\Gamma^{\mathrm{PR}}, \Omega)$ *are strictly increasing at* $x > 0$.

Assumption A.7. *For $\lambda_0 \in \Lambda$, there is $\{\lambda_n \in \Lambda\}_{n\geq 1}$ such that $\{(\lambda_n, F_n) \in \mathcal{L}_0\}_{n\geq 1}$ satisfies the following statements*:

(a) *The conditions in Assumption A.5 hold.*

(b) *There is a* (*possibly random*) *sequence $\{\hat{\theta}_n \in \Theta(\lambda_0)\}_{n\geq 1}$ such that* (*s.t.*) *the the following statements hold*:

(i) *We have $T_n(\lambda_0) - S(\sqrt{n}\bar{m}_n(\hat{\theta}_n), \hat{\Sigma}_n(\hat{\theta}_n)) = o_p(1)$.*

(ii) *We have $\sqrt{n}D_{F_n}^{-1/2}(\hat{\theta}_n)E_{F_n}[m(W, \hat{\theta}_n)] = \lambda + o_p(1)$, where $\lambda_j \in \mathbb{R}$ for some $j = 1, \dots, k$,*

(iii) *We have $\hat{\theta}_n = \theta + o_p(1)$ for some $\theta \in \Theta$.*

(c) *There are* (*possibly random*) *sequences $\{\hat{\theta}_n^{\mathrm{SS}} \in \Theta(\lambda_0)\}_{n\geq 1}$ and $\{\tilde{\theta}_n^{\mathrm{SS}} \in \Theta_I(F_n)\}_{n\geq 1}$ s.t., conditionally on $\{W_i\}_{i=1}^n$, the following statements hold*:

(i) *We have $T_n^{\mathrm{SS}}(\lambda_0) - S(\sqrt{b_n}\bar{m}_n^{\mathrm{SS}}(\hat{\theta}_n^{\mathrm{SS}}), \hat{\Sigma}_n(\hat{\theta}_n^{\mathrm{SS}})) = o_p(1)$ a.s.*

(ii) *We have $\sqrt{n}(D_{F_n}^{-1/2}(\hat{\theta}_n^{\mathrm{SS}})E_{F_n}[m(W, \hat{\theta}_n^{\mathrm{SS}})] - D_{F_n}^{-1/2}(\tilde{\theta}_n^{\mathrm{SS}})E_{F_n}[m(W, \tilde{\theta}_n^{\mathrm{SS}})]) = O_p(1)$ a.s.*

(iii) *We have $\sqrt{b_n}D_{F_n}^{-1/2}(\tilde{\theta}_n^{\mathrm{SS}})E_{F_n}[m(W, \tilde{\theta}_n^{\mathrm{SS}})] = (g, \mathbf{0}_{k-p}) + o_p(1)$ a.s. with $g \in \mathbb{R}^p_{[+\infty]}$ and $g_j \in (0, \infty)$ for some $j = 1, \dots, p$. In addition, either $k > p$ or $k = p$, where $g_l = 0$ for some $l = 1, \dots, p$.*

(iv) *We have $\hat{\theta}_n^{\mathrm{SS}} = \theta^* + o_p(1)$ a.s. for some $\theta^* \in \Theta$.*

(d) *Assumption A.4 holds with strict inequality, that is,* $\limsup_{n\to\infty} \kappa_n\sqrt{b_n/n} < 1$.

B.3 *Maintained assumptions*

The literature routinely assumes that the function $S(\cdot)$ entering $Q_n(\theta)$ in (4.2) satisfies the following assumptions (see, e.g., Andrews and Soares (2010), Andrews and Guggenberger (2009), and Bugni, Canay, and Guggenberger (2012)). We therefore treat the assumptions below as maintained (denoted by the letter “M” in the labeling). We note in particular that the constant χ in Assumption M.1 equals 2 when the function $S(\cdot)$ is either the modified methods of moments in (2.6) or the quasi-likelihood ratio.

Assumption M.1. *For some $\chi > 0$, $S(am, \Omega) = a^\chi S(m, \Omega)$ for all scalars $a > 0$, $m \in \mathbb{R}^k$, and $\Omega \in \Psi$.*

Assumption M.2. *The sequence $\{\kappa_n\}_{n\geq 1}$ satisfies $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$. The sequence $\{b_n\}_{n\geq 1}$ satisfies $b_n \to \infty$ and $b_n/n \to 0$.*

Assumption M.3. *For each $\lambda \in \Lambda$, $\Theta(\lambda)$ is a nonempty and compact subset of $\mathbb{R}^{d_\theta}$ ($d_\theta < \infty$).*

Assumption M.4. *Test BP is computed using the GMS approach in Andrews and Soares (2010); that is, $\phi_n^{\mathrm{BP}}(\cdot)$ in (4.9) is based on $\mathrm{CS}_n(1-\alpha) = \{\theta \in \Theta : Q_n(\theta) \leq \hat{c}_n(\theta, 1-\alpha)\}$, where $\hat{c}_n(\theta, 1-\alpha)$ is the GMS critical value constructed using the GMS function $\varphi(\cdot)$ and threshholding sequence $\{\kappa_n\}_{n\geq 1}$ satisfying Assumption M.2.*

Assumption M.5. *The function $S(\cdot)$ satisfies the following conditions.*

(a) *The term $S((m_1, m_2), \Sigma)$ is nonincreasing in m_1 for all $(m_1, m_2) \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$ and all variance matrices $\Sigma \in \mathbb{R}^{k\times k}$.*

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

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

(d) *The term $S(m, \Omega)$ is continuous at all $m \in \mathbb{R}^k_{[\pm\infty]}$ and $\Omega \in \Psi$.*

Assumption M.6. *For all $h_1 \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$, all $\Omega \in \Psi$, and $Z \sim N(\mathbf{0}_k, \Omega)$, the distribution function of $S(Z + h_1, \Omega)$ at $x \in \mathbb{R}$ has the following properties*:

(a) *It is continuous for $x > 0$.*

(b) *It is strictly increasing for $x > 0$ unless $p = k$ and $h_1 = \infty^p$.*

(c) *It is less than or equal to $1/2$ at $x = 0$ when $k > p$ or when $k = p$ and $h_{1,j} = 0$ for some $j = 1, \dots, p$.*

(d) *It is degenerate at $x = 0$ when $p = k$ and $h_1 = \infty^p$.*

(e) *It satisfies $P(S(Z + (m_1, \mathbf{0}_{k-p}), \Omega) \leq x) < P(S(Z + (m_1^*, \mathbf{0}_{k-p}), \Omega) \leq x)$ for all $x > 0$ and all $m_1, m_1^* \in \mathbb{R}^p_{[+\infty]}$ with $m_{1,j} \leq m^*_{1,j}$ for all $j = 1, \dots, p$ and $m_{1,j} < m^*_{1,j}$ for some $j = 1, \dots, p$.*

Assumption M.7. *The function $S(\cdot)$ satisfies the following conditions.*

(a) *The distribution function of $S(Z, \Omega)$ is continuous at its $(1-\alpha)$ quantile, denoted $c_{(1-\alpha)}(\Omega)$, for all $\Omega \in \Psi$, where $Z \sim N(\mathbf{0}_k, \Omega)$ and $\alpha \in (0, 0.5)$.*

(b) *The term $c_{(1-\alpha)}(\Omega)$ is continuous in Ω uniformly for $\Omega \in \Psi$.*

Assumption M.8. *We have $S(m, \Omega) > 0$ if and only if $m_j < 0$ for some $j = 1, \dots, p$ or $m_j \neq 0$ for some $j = p+1, \dots, k$, where $m = (m_1, \dots, m_k)'$ and $\Omega \in \Psi$. Equivalently, $S(m, \Omega) = 0$ if and only if $m_j \geq 0$ for all $j = 1, \dots, p$ and $m_j = 0$ for all $j = p+1, \dots, k$, where $m = (m_1, \dots, m_k)'$ and $\Omega \in \Psi$.*

Assumption M.9. *For all $n \geq 1$, $S(\sqrt{n}\bar{m}_n(\theta), \hat{\Sigma}(\theta))$ is a lower semicontinuous function of $\theta \in \Theta$.*

Appendix C: Proofs of the main theorems

Proof of Theorem 4.1. We divide the proof into six steps and show that for $\eta \geq 0$,

$$\limsup_{n\to\infty} \sup_{(\lambda,F)\in\mathcal{L}_0} P_F\big(T_n(\lambda) > \hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) + \eta\big) \leq \alpha.$$

Steps 1–4 hold for $\eta \geq 0$, Step 5 needs $\eta > 0$, and Step 6 holds for $\eta = 0$ under Assumption A.2.

Step 1. For any $(\lambda, F) \in \mathcal{L}_0$, let $\tilde{T}_n^{\mathrm{DR}}(\lambda)$ be as in S.1 of the Supplemental Material and let $\tilde{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha)$ be the conditional $(1-\alpha)$ quantile of $\min\{\tilde{T}_n^{\mathrm{DR}}(\lambda), T_n^{\mathrm{PR}}(\lambda)\}$. Consider the derivation

$$\begin{aligned} P_F\big(T_n(\lambda) > \hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) + \eta\big) &\leq P_F\big(T_n(\lambda) > \tilde{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) + \eta\big) \\ &\quad + P_F\big(\hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) < \tilde{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha)\big) \\ &\leq P_F\big(T_n(\lambda) > \tilde{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) + \eta\big) \\ &\quad + P_F\big(\hat{\Theta}_I(\lambda) \nsubseteq \Theta(\lambda) \cap \Theta_I^{\ln \kappa_n}(F)\big), \end{aligned}$$

where the second inequality follows from the fact that Assumption A.1 and $\hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) < \tilde{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha)$ imply that $\hat{\Theta}_I(\lambda) \nsubseteq \Theta(\lambda) \cap \Theta_I^{\ln \kappa_n}(F)$. By this and Lemma D.13 in Bugni, Canay, and Shi (2015) (with a redefined parameter space equal to $\Theta(\lambda)$), it follows that

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{(\lambda,F)\in\mathcal{L}_0} P_F\big(T_n(\lambda) > \hat{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) + \eta\big) \\ &\quad \leq \limsup_{n\to\infty} \sup_{(\lambda,F)\in\mathcal{L}_0} P_F\big(T_n(\lambda) > \tilde{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) + \eta\big). \end{aligned}$$

Step 2. By definition, there exists a subsequence $\{a_n\}_{n\geq 1}$ of $\{n\}_{n\geq 1}$ and a subsequence $\{(\lambda_{a_n}, F_{a_n})\}_{n\geq 1}$ s.t.

$$\begin{aligned} &\limsup_{n\to\infty} \sup_{(\lambda,F)\in\mathcal{L}_0} P_F\big(T_n(\lambda) > \tilde{c}_n^{\mathrm{MR}}(\lambda, 1-\alpha) + \eta\big) \\ &\quad = \lim_{n\to\infty} P_{F_{a_n}}\big(T_{a_n}(\lambda_{a_n}) > \tilde{c}_{a_n}^{\mathrm{MR}}(\lambda_{a_n}, 1-\alpha) + \eta\big). \end{aligned} \tag{C.1}$$

By Lemma S.3.2, there is a further sequence $\{u_n\}_{n\geq 1}$ of $\{a_n\}_{n\geq 1}$ s.t. $\Omega_{F_{u_n}} \xrightarrow{u} \Omega$, $\Gamma_{u_n,F_{u_n}}(\lambda_{u_n}) \xrightarrow{H} \Gamma$, $\Gamma_{u_n,F_{u_n}}^{\mathrm{DR}}(\lambda_{u_n}) \xrightarrow{H} \Gamma^{\mathrm{DR}}$, and $\Gamma_{u_n,F_{u_n}}^{\mathrm{PR}}(\lambda_{u_n}) \xrightarrow{H} \Gamma^{\mathrm{PR}}$ for some $(\Gamma, \Gamma^{\mathrm{DR}}, \Gamma^{\mathrm{PR}}, \Omega) \in \mathcal{S}(\Theta \times \mathbb{R}_{[\pm\infty]}^k)^3 \times \mathcal{C}(\Theta^2)$. Since $\Omega_{F_{u_n}} \xrightarrow{u} \Omega$ and $\Gamma_{u_n,F_{u_n}}(\lambda_{u_n}) \xrightarrow{H} \Gamma$, Theorem S.2.4 implies that $T_{u_n}(\lambda_{u_n}) \xrightarrow{d} J(\Gamma, \Omega) \equiv \inf_{(\theta,\ell)\in\Gamma} S(v_\Omega(\theta) + \ell, \Omega(\theta))$. Similarly, Theorem S.2.2 implies that $\{\min\{\tilde{T}_{u_n}^{\mathrm{DR}}(\lambda_{u_n}), T_{u_n}^{\mathrm{PR}}(\lambda_{u_n})\} | \{W_i\}_{i=1}^{u_n}\} \xrightarrow{d} J(\Gamma^{\mathrm{MR}}, \Omega)$ a.s.

Step 3. We show that $J(\Gamma^{\mathrm{MR}}, \Omega) \geq J(\Gamma, \Omega)$. Suppose not, that is, $\exists (\theta, \ell) \in \Gamma_*^{\mathrm{DR}} \cup \Gamma^{\mathrm{PR}}$ s.t. $S(v_\Omega(\theta) + \ell, \Omega(\theta)) < J(\Gamma, \Omega)$. If $(\theta, \ell) \in \Gamma_*^{\mathrm{DR}}$ then by definition $\exists (\theta, \ell') \in \Gamma^{\mathrm{DR}}$ s.t. $\varphi^*(\ell') = \ell$ and $S(v_\Omega(\theta) + \varphi^*(\ell'), \Omega(\theta)) < J(\Gamma, \Omega)$. By Lemma S.3.7, $\exists (\theta, \tilde{\ell}) \in \Gamma$, where $\tilde{\ell}_j \geq \varphi_j^*(\ell_j')$ for $j \leq p$ and $\tilde{\ell}_j = 0$ for $j > p$. Thus

$$\begin{aligned} S\big(v_\Omega(\theta) + \tilde{\ell}, \Omega(\theta)\big) &\leq S\big(v_\Omega(\theta) + \varphi^*(\ell'), \Omega(\theta)\big) \\ &< J(\Gamma, \Omega) \equiv \inf_{(\theta,\ell)\in\Gamma} S\big(v_\Omega(\theta) + \ell, \Omega(\theta)\big), \end{aligned}$$

which is a contradiction to $(\theta, \tilde{\ell}) \in \Gamma$. If $(\theta, \ell) \in \Gamma^{\mathrm{PR}}$, we first need to show that $\ell \in \mathbb{R}_{[+\infty]}^p \times \mathbb{R}^{k-p}$. Suppose not, that is, suppose that $\ell_j = -\infty$ for some $j \leq p$ or $|\ell_j| = \infty$ for some $j > p$. Since $v_\Omega : \Theta \to \mathbb{R}^k$ is a tight Gaussian process, it follows that $v_{\Omega,j}(\theta) +$

$\ell_j = -\infty$ for some $j \le p$ or $|v_{\Omega,j}(\theta) + \ell_j| = \infty$ for some $j > p$. By Lemma S.3.6, we have $S(v_\Omega(\theta) + \ell, \Omega(\theta)) = \infty$, which contradicts $S(v_\Omega(\theta) + \ell, \Omega(\theta)) < J(\Gamma, \Omega)$. Since $\ell \in \mathbb{R}^p_{[+\infty]} \times \mathbb{R}^{k-p}$, Lemma S.3.8 implies that $\exists(\theta, \tilde{\ell}) \in \Gamma$, where $\tilde{\ell}_j \ge \ell_j$ for $j \le p$ and $\tilde{\ell}_j = \ell_j$ for $j > p$. We conclude that

$$S\big(v_\Omega(\theta) + \tilde{\ell}, \Omega(\theta)\big) \le S\big(v_\Omega(\theta) + \ell, \Omega(\theta)\big) < J(\Gamma, \Omega) \equiv \inf_{(\theta,\ell)\in\Gamma} S\big(v_\Omega(\theta) + \ell, \Omega(\theta)\big),$$

which is a contradiction to $(\theta, \tilde{\ell}) \in \Gamma$.

Step 4. We now show that for $c_{(1-\alpha)}(\Gamma, \Omega)$ being the $(1-\alpha)$ quantile of $J(\Gamma, \Omega)$ and any $\varepsilon > 0$,

$$\lim_{n\to\infty} P_{F_{u_n}}\big(\tilde{c}^{\mathrm{MR}}_{u_n}(\lambda_{u_n}, 1-\alpha) \le c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon\big) = 0. \tag{C.2}$$

Let $\varepsilon > 0$ be s.t. $c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon$ is a continuity point of the CDF of $J(\Gamma, \Omega)$. Then

$$\begin{aligned}
&\lim_{n\to\infty} P_{F_{u_n}}\big(\min\big\{\tilde{T}^{\mathrm{DR}}_{u_n}(\lambda_{u_n}), T^{\mathrm{PR}}_{u_n}(\lambda_{u_n})\big\} \le c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon \big| \{W_i\}_{i=1}^{u_n}\big) \\
&\quad = P\big(J\big(\Gamma^{\mathrm{MR}}, \Omega\big) \le c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon\big) \\
&\quad \le P\big(J(\Gamma, \Omega) \le c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon\big) < 1 - \alpha,
\end{aligned}$$

where the first equality holds because $\{\min\{\tilde{T}^{\mathrm{DR}}_{u_n}(\lambda_{u_n}), T^{\mathrm{PR}}_{u_n}(\lambda_{u_n})\} | \{W_i\}_{i=1}^{u_n}\} \xrightarrow{d} J(\Gamma^{\mathrm{MR}}, \Omega)$ a.s., the second weak inequality is a consequence of $J(\Gamma^{\mathrm{MR}}, \Omega) \ge J(\Gamma, \Omega)$, and the final strict inequality follows from $c_{(1-\alpha)}(\Gamma, \Omega)$ being the $(1-\alpha)$ quantile of $J(\Gamma, \Omega)$. Next, notice that

$$\begin{aligned}
&\Big\{\lim_{n\to\infty} P_{F_{u_n}}\big(\min\big\{\tilde{T}^{\mathrm{DR}}_{u_n}(\lambda_{u_n}), T^{\mathrm{PR}}_{u_n}(\lambda_{u_n})\big\} \le c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon \big| \{W_i\}_{i=1}^{u_n}\big) < 1 - \alpha\Big\} \\
&\quad \subseteq \Big\{\liminf_{n\to\infty}\big\{\tilde{c}^{\mathrm{MR}}_{u_n}(1-\alpha) > c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon\big\}\Big\}.
\end{aligned}$$

Since the right-hand side (RHS) occurs a.s., the left-hand side (LHS) must also occur a.s. Then (C.2) is a consequence of this and Fatou's lemma.

Step 5. For $\eta > 0$, we can define $\varepsilon > 0$ in Step 4 so that $\eta - \varepsilon > 0$ and $c_{(1-\alpha)}(\Gamma, \Omega) + \eta - \varepsilon$ is a continuity point of the CDF of $J(\Gamma, \Omega)$. It then follows that

$$\begin{aligned}
&P_{F_{u_n}}\big(T_{u_n}(\lambda_{u_n}) > \tilde{c}^{\mathrm{MR}}_{u_n}(\lambda_{u_n}, 1-\alpha) + \eta\big) \\
&\quad \le P_{F_{u_n}}\big(\tilde{c}^{\mathrm{MR}}_{u_n}(\lambda_{u_n}, 1-\alpha) \le c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon\big) \\
&\qquad + 1 - P_{F_{u_n}}\big(T_{u_n}(\lambda_{u_n}) \le c_{(1-\alpha)}(\Gamma, \Omega) + \eta - \varepsilon\big).
\end{aligned} \tag{C.3}$$

Taking the limit supremum on both sides, using Steps 2 and 4, and that $\eta - \varepsilon > 0$, yields

$$\begin{aligned}
&\limsup_{n\to\infty} P_{F_{u_n}}\big(T_{u_n}(\lambda_{u_n}) > \tilde{c}^{\mathrm{MR}}_{u_n}(\lambda_{u_n}, 1-\alpha) + \eta\big) \\
&\quad \le 1 - P\big(J(\Gamma, \Omega) \le c_{(1-\alpha)}(\Gamma, \Omega) + \eta - \varepsilon\big) \le \alpha.
\end{aligned}$$

This combined with Steps 1 and 2 completes the proof under $\eta > 0$.

Step 6. For $\eta = 0$, there are two cases to consider. First, suppose $c_{(1-\alpha)}(\Gamma, \Omega) = 0$. Then, by Assumption A.2,

$$\limsup_{n\to\infty} P_{F_{u_n}}\big(T_{u_n}(\lambda_{u_n}) > \tilde{c}^{\mathrm{MR}}_{u_n}(\lambda_{u_n}, 1-\alpha)\big) \le \limsup_{n\to\infty} P_{F_{u_n}}\big(T_{u_n}(\lambda_{u_n}) \ne 0\big) \le \alpha.$$

The proof is completed by combining the previous equation with Steps 1 and 2. Second, suppose $c_{(1-\alpha)}(\Gamma, \Omega) > 0$. Consider a sequence $\{\varepsilon_m\}_{m\ge1}$ s.t. $\varepsilon_m \downarrow 0$ and $c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon_m$ is a continuity point of the CDF of $J(\Gamma, \Omega)$ for all $m \in \mathbb{N}$. For any $m \in \mathbb{N}$, it follows from (C.3) and Steps 2 and 3 that

$$\limsup_{n\to\infty} P_{F_{u_n}}\big(T_{u_n}(\lambda_{u_n}) > \tilde{c}^{\mathrm{MR}}_{u_n}(\lambda_{u_n}, 1-\alpha)\big) \le 1 - P\big(J(\Gamma, \Omega) \le c_{(1-\alpha)}(\Gamma, \Omega) - \varepsilon_m\big).$$

Taking $\varepsilon_m \downarrow 0$ and using continuity gives the RHS equal to α. This, with Steps 1 and 2, completes the proof. □

The proof of Theorem 4.2 follows identical steps to those in the proof of Bugni, Canay, and Shi (2015, Theorem 6.1) and is therefore omitted.

Proof of Theorem 4.3. Suppose not, that is, suppose that $\liminf(E_{F_n}[\phi^{\mathrm{PR}}_n(\lambda_0)] - E_{F_n}[\phi^{\mathrm{SS}}_n(\lambda_0)]) \equiv -\delta < 0$. Consider a subsequence $\{k_n\}_{n\ge1}$ of $\{n\}_{n\ge1}$ such that

$$\begin{aligned} P_{F_{k_n}}\big(T_{k_n}(\lambda_0) > c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha)\big) &= E_{F_{k_n}}\big[\phi^{\mathrm{PR}}_{k_n}(\lambda_0)\big] < E_{F_{k_n}}\big[\phi^{\mathrm{SS}}_{k_n}(\lambda_0)\big] - \delta/2 \\ &= P_{F_{k_n}}\big(T_{k_n}(\lambda_0) > c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha)\big) - \delta/2 \end{aligned}$$

or, equivalently,

$$P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \le c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha)\big) + \delta/2 < P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \le c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha)\big). \tag{C.4}$$

Lemma S.3.2 implies that for some $(\Gamma, \Gamma^{\mathrm{PR}}, \Gamma^{\mathrm{SS}}, \Gamma^{\mathrm{PR}}_A, \Gamma^{\mathrm{SS}}_A, \Omega) \in \mathcal{S}(\Theta \times \mathbb{R}^k_{[\pm\infty]})^5 \times \mathcal{C}(\Theta^2)$, $\Omega_{F_{k_n}} \xrightarrow{u} \Omega$, $\Gamma^{\mathrm{PR}}_{k_n,F_{k_n}}(\lambda_0) \xrightarrow{H} \Gamma^{\mathrm{PR}}$, $\Gamma^{\mathrm{SS}}_{b_{k_n},F_{k_n}}(\lambda_0) \xrightarrow{H} \Gamma^{\mathrm{SS}}$, $\Gamma^{\mathrm{PR}}_{k_n,F_{k_n}}(\lambda_{k_n}) \xrightarrow{H} \Gamma^{\mathrm{PR}}_A$, and $\Gamma^{\mathrm{SS}}_{b_{k_n},F_{k_n}}(\lambda_{k_n}) \xrightarrow{H} \Gamma^{\mathrm{SS}}_A$. Then Theorems S.2.1, S.2.3, and S.2.4 imply that $T_{k_n}(\lambda_0) \xrightarrow{d} J(\Gamma, \Omega)$, $\{T^{\mathrm{PR}}_{k_n}(\lambda_0) | \{W_i\}_{i=1}^{k_n}\} \xrightarrow{d} J(\Gamma^{\mathrm{PR}}, \Omega)$ a.s., and $\{T^{\mathrm{SS}}_{k_n}(\lambda_0) | \{W_i\}_{i=1}^{k_n}\} \xrightarrow{d} J(\Gamma^{\mathrm{SS}}, \Omega)$ a.s.

We next show that $c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha) \xrightarrow{\text{a.s.}} c_{1-\alpha}(\Gamma^{\mathrm{PR}}, \Omega)$. Let $\varepsilon > 0$ be arbitrary and pick $\tilde{\varepsilon} \in (0, \varepsilon)$ s.t. $c_{(1-\alpha)}(\Gamma^{\mathrm{PR}}, \Omega) + \tilde{\varepsilon}$ and $c_{(1-\alpha)}(\Gamma^{\mathrm{PR}}, \Omega) - \tilde{\varepsilon}$ are both a continuity points of the CDF of $J(\Gamma^{\mathrm{PR}}, \Omega)$. Then

$$\begin{aligned} &\lim_{n\to\infty} P_{F_{k_n}}\big(T^{\mathrm{PR}}_{k_n}(\lambda_0) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) + \tilde{\varepsilon} \big| \{W_i\}_{i=1}^{n}\big) \\ &\quad = P\big(J\big(\Gamma^{\mathrm{PR}}, \Omega\big) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) + \tilde{\varepsilon}\big) > 1 - \alpha \quad \text{a.s.}, \end{aligned} \tag{C.5}$$

where the first equality holds because of $\{T^{\mathrm{PR}}_{k_n}(\lambda_0) | \{W_i\}_{i=1}^{k_n}\} \xrightarrow{d} J(\Gamma^{\mathrm{PR}}, \Omega)$ a.s., and the strict inequality is due to $\tilde{\varepsilon} > 0$ and $c_{(1-\alpha)}(\Gamma^{\mathrm{PR}}, \Omega) + \tilde{\varepsilon}$ being a continuity point of the

CDF of $J(\Gamma^{\mathrm{PR}}, \Omega)$. Similarly,

$$\begin{aligned}&\lim_{n\to\infty} P_{F_{k_n}}\big(T^{\mathrm{PR}}_{k_n}(\lambda_0) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) - \tilde{\varepsilon}\big|\{W_i\}_{i=1}^n\big) \\ &\quad = P\big(J\big(\Gamma^{\mathrm{PR}}, \Omega\big) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) - \tilde{\varepsilon}\big) < 1-\alpha.\end{aligned} \tag{C.6}$$

Next, notice that

$$\begin{aligned}&\Big\{\lim_{n\to\infty} P_{F_{k_n}}\big(T^{\mathrm{PR}}_{k_n}(\lambda_0) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) + \tilde{\varepsilon}\big|\{W_i\}_{i=1}^n\big) > 1-\alpha\Big\} \\ &\quad \subseteq \Big\{\liminf_{n\to\infty}\big\{c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha) < c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) + \tilde{\varepsilon}\big\}\Big\},\end{aligned} \tag{C.7}$$

with the same result holding with $-\tilde{\varepsilon}$ replacing $\tilde{\varepsilon}$. From (C.5), (C.6), and (C.7), we conclude that

$$P_{Fn}\Big(\liminf_{n\to\infty}\big\{\big|c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha) - c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big)\big| \le \varepsilon\big\}\Big) = 1,$$

which is equivalent to $c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha) \overset{\text{a.s.}}{\to} c_{(1-\alpha)}(\Gamma^{\mathrm{PR}}, \Omega)$. By similar arguments, $c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha) \overset{\text{a.s.}}{\to} c_{(1-\alpha)}(\Gamma^{\mathrm{SS}}, \Omega)$.

Let $\varepsilon > 0$ be s.t. $c_{(1-\alpha)}(\Gamma^{\mathrm{SS}}, \Omega) - \varepsilon$ is a continuity point of the CDF of $J(\Gamma, \Omega)$ and note that

$$\begin{aligned}P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \le c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha)\big) \ge{}& P_{F_{k_n}}\Big(\big\{T_{k_n}(\lambda_0) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}}, \Omega\big) - \varepsilon\big\} \\ &\cap \big\{c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha) \ge c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}}, \Omega\big) - \varepsilon\big\}\Big) \\ &+ P_{F_{k_n}}\Big(\big\{T_{k_n}(\lambda_0) \le c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha)\big\} \\ &\cap \big\{c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha) < c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}}, \Omega\big) - \varepsilon\big\}\Big).\end{aligned}$$

Taking lim inf and using that $T_{k_n}(\lambda_0) \overset{d}{\to} J(\Gamma, \Omega)$ and $c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha) \overset{\text{a.s.}}{\to} c_{(1-\alpha)}(\Gamma^{\mathrm{SS}}, \Omega)$, we deduce that

$$\liminf_{n\to\infty} P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \le c^{\mathrm{SS}}_{k_n}(\lambda_0, 1-\alpha)\big) \ge P\big(J(\Gamma, \Omega) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}}, \Omega\big) - \varepsilon\big). \tag{C.8}$$

Fix $\varepsilon > 0$ arbitrarily and pick $\tilde{\varepsilon} \in (0, \varepsilon)$ s.t. $c_{(1-\alpha)}(\Gamma^{\mathrm{PR}}, \Omega) + \tilde{\varepsilon}$ is a continuity point of the CDF of $J(\Gamma, \Omega)$. Then

$$\begin{aligned}P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \le c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha)\big) \le{}& P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) + \tilde{\varepsilon}\big) \\ &+ P_{F_{k_n}}\big(c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha) > c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) + \tilde{\varepsilon}\big).\end{aligned}$$

Taking lim sup on both sides, and using that $T_{k_n}(\lambda_0) \overset{d}{\to} J(\Gamma, \Omega)$, $c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha) \overset{\text{a.s.}}{\to} c_{(1-\alpha)}(\Gamma^{\mathrm{PR}}, \Omega)$, and $\tilde{\varepsilon} \in (0, \varepsilon)$, yields

$$\limsup_{n\to\infty} P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \le c^{\mathrm{PR}}_{k_n}(\lambda_0, 1-\alpha)\big) \le P\big(J(\Gamma, \Omega) \le c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}}, \Omega\big) + \tilde{\varepsilon}\big). \tag{C.9}$$

Next consider the derivation

$$
\begin{aligned}
&P\big(J(\Gamma,\Omega) \leq c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}},\Omega\big)-\varepsilon\big)+\delta/2 \\
&\quad \leq \liminf P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \leq c^{\mathrm{SS}}_{k_n}(\lambda_0,1-\alpha)\big)+\delta/2 \\
&\quad \leq \limsup P_{F_{k_n}}\big(T_{k_n}(\lambda_0) \leq c^{\mathrm{PR}}_{k_n}(\lambda_0,1-\alpha)\big) \\
&\quad \leq P\big(J(\Gamma,\Omega) \leq c_{(1-\alpha)}\big(\Gamma^{\mathrm{PR}},\Omega\big)+\varepsilon\big) \\
&\quad \leq P\big(J(\Gamma,\Omega) \leq c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}},\Omega\big)+\varepsilon\big),
\end{aligned}
$$

where the first inequality follows from (C.8), the second inequality follows from (C.4), the third inequality follows from (C.9), and the fourth inequality follows from $c_{(1-\alpha)}(\Gamma^{\mathrm{PR}},\Omega) \leq c_{(1-\alpha)}(\Gamma^{\mathrm{SS}},\Omega)$ by Lemma S.3.9. We conclude that

$$
P\big(J(\Gamma,\Omega) \leq c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}},\Omega\big)+\varepsilon\big) - P\big(J(\Gamma,\Omega) \leq c_{(1-\alpha)}\big(\Gamma^{\mathrm{SS}},\Omega\big)-\varepsilon\big) \geq \delta/2 > 0.
$$

Taking $\varepsilon \downarrow 0$ and using Assumption A.6, the LHS converges to zero, which is a contradiction. □

References

Andrews, D. W. K. and P. J. Barwick (2012), "Inference for parameters defined by moment inequalities: A recommended moment selection procedure." *Econometrica*, 80, 2805–2826. [12, 26, 28]

Andrews, D. W. K. and P. Guggenberger (2009), "Validity of subsampling and 'plug-in asymptotic' inference for parameters defined by moment inequalities." *Econometric Theory*, 25, 669–709. [2, 12, 16, 30]

Andrews, D. W. K. and S. Han (2009), "Invalidity of the bootstrap and the *m* out of *n* bootstrap for confidence interval endpoints defined by moment inequalities." *Econometrics Journal*, 12, S172–S199. [3]

Andrews, D. W. K. and X. Shi (2013), "Inference based on conditional moment inequalities." *Econometrica*, 81, 609–666. [26]

Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78, 119–157. [2, 5, 6, 7, 10, 12, 13, 15, 16, 17, 18, 19, 20, 23, 25, 28, 29, 30]

Armstrong, T. B. (2014), "Weighted KS statistics for inference on conditional moment inequalities." *Journal of Econometrics*, 181, 92–116. [26]

Beresteanu, A. and F. Molinari (2008), "Asymptotic properties for a class of partially identified models." *Econometrica*, 76, 763–814. [2]

Bontemps, C., T. Magnac, and E. Maurin (2012), "Set identified linear models." *Econometrica*, 80, 1129–1155. [2]

Bugni, F. A. (2010), "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set." *Econometrica*, 78, 735–753. [2, 3, 16, 25, 29]

Bugni, F. A. (2016), "Comparison of inferential methods in partially identified models in terms of error in coverage probability." *Econometric Theory*, 32, 187–242. Formerly circulated as "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the elements of the identified set." [3, 16, 17, 20, 25, 29]

Bugni, F. A., I. A. Canay, and P. Guggenberger (2012), "Distortions of asymptotic confidence size in locally misspecified moment inequality models." *Econometrica*, 80, 1741–1768. [2, 12, 30]

Bugni, F. A., I. A. Canay, and X. Shi (2015), "Specification tests for partially identified models defined by moment inequalities." *Journal of Econometrics*, 185, 259–282. [13, 17, 19, 28, 32, 34]

Canay, I. A. (2010), "EL inference for partially identified models: Large deviations optimality and bootstrap validity." *Journal of Econometrics*, 156, 408–425. [2, 16, 24]

Chen, X., E. Tamer, and A. Torgovitsky (2011), "Sensitivity analysis in partially identified semiparametric models." Manuscript, Northwestern University. [3]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models." *Econometrica*, 75, 1243–1284. [2, 15, 29]

Chernozhukov, V., S. Lee, and A. M. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81, 667–737. [26]

Chetverikov, D. (2013), "Adaptive test of conditional moment inequalities." Available at arXiv:1201.0167. [26]

Ciliberto, F. and E. Tamer (2009), "Market structure and multiple equilibria in airline industry." *Econometrica*, 77, 1791–1828. [2, 16, 25]

Dickstein, M. and E. Morales (2015), "What do exporters know?" Technical report, Princeton University. [16, 24]

Galichon, A. and M. Henry (2011), "Set identification in models with multiple equilibria." *Review of Economic Studies*, 78, 1264–1298. [2]

Gandhi, A., Z. Lu, and X. Shi (2013), "Estimating demand for differentiated products with error in market shares: A moment inequalities approach." Manuscript, University of Wisconsin—Madison. [19]

Grieco, P. (2014), "Discrete games with flexible information structures: An application to local grocery markets." *RAND Journal of Economics*, 45, 303–340. [2, 16]

Imbens, G. and C. F. Manski (2004), "Confidence intervals for partially identified parameters." *Econometrica*, 72, 1845–1857. [2]

Kaido, H., F. Molinari, and J. Stoye (2016), "Inference for projections of identified sets." Manuscript. Cornell Univerisity. [26]

Kline, B. and E. Tamer (2013), "Default Bayesian inference in a class of partially identified models." Manuscript, Northwestern University. [3]

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015), "Moment inequalities and their application." *Econometrica*, 83, 315–334. [2, 3]

Politis, D. N. and J. P. Romano (1994), "Large sample confidence regions based on subsamples under minimal assumptions." *The Annals of Statistics*, 22, 2031–2050. [3, 29]

Politis, D. N., J. P. Romano, and M. Wolf (1999), *Subsampling*. Springer, New York. [23]

Romano, J. P. and A. M. Shaikh (2008), "Inference for identifiable parameters in partially identified econometric models." *Journal of Statistical Planning and Inference*, 138, 2786–2807. [2, 3, 13, 15, 16, 18]

Romano, J. P. and A. M. Shaikh (2010), "Inference for the identified set in partially identified econometric models." *Econometrica*, 78, 169–211. [2]

Romano, J. P., A. M. Shaikh, and M. Wolf (2014), "A practical two-step method for testing moment inequalities." *Econometrica*, 82, 1979–2002. [2, 12, 26]

Rosen, A. M. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146, 107–117. [2]

Stoye, J. (2009), "More on confidence intervals for partially identified parameters." *Econometrica*, 77, 1299–1315. [2]

van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*. Springer-Verlag, New York. [14, 15, 28]

Wan, Y. Y. (2013), "An integration-based approach to moment inequality models." Manuscript, University of Toronto. [3, 4]

Wollmann, T. (2015), "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles." Manuscript. [16, 25]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 11 September, 2014; final version accepted 10 February, 2016; available online 6 April, 2016.