

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic
Theory in its Relation to Statistics and Mathematics*

<http://www.econometricsociety.org/>

Econometrica, Vol. 82, No. 5 (September, 2014), 1979–2002

A PRACTICAL TWO-STEP METHOD FOR TESTING MOMENT INEQUALITIES

JOSEPH P. ROMANO

Stanford University, Stanford, CA 94305-406, U.S.A.

AZEEM M. SHAIKH

University of Chicago, Chicago, IL 60637, U.S.A.

MICHAEL WOLF

University of Zurich, 8032 Zurich, Switzerland

The copyright to this Article is held by the Econometric Society. It may be downloaded, printed and reproduced only for educational or research purposes, including use in course packs. No downloading or copying may be done for any commercial purpose without the explicit permission of the Econometric Society. For such commercial purposes contact the Office of the Econometric Society (contact information may be found at the website <http://www.econometricsociety.org> or in the back cover of *Econometrica*). This statement must be included on all copies of this Article that are made available electronically or in any other format.

A PRACTICAL TWO-STEP METHOD FOR TESTING MOMENT INEQUALITIES

BY JOSEPH P. ROMANO, AZEEM M. SHAIKH, AND MICHAEL WOLF¹

This paper considers the problem of testing a finite number of moment inequalities. We propose a two-step approach. In the first step, a confidence region for the moments is constructed. In the second step, this set is used to provide information about which moments are “negative.” A Bonferroni-type correction is used to account for the fact that, with some probability, the moments may not lie in the confidence region. It is shown that the test controls size uniformly over a large class of distributions for the observed data. An important feature of the proposal is that it remains computationally feasible, even when the number of moments is large. The finite-sample properties of the procedure are examined via a simulation study, which demonstrates, among other things, that the proposal remains competitive with existing procedures while being computationally more attractive.

KEYWORDS: Bonferroni inequality, bootstrap, moment inequalities, partial identification, uniform validity.

1. INTRODUCTION

LET $W_i, i = 1, \dots, n$, BE AN INDEPENDENT AND IDENTICALLY DISTRIBUTED (i.i.d.) sequence of random variables with distribution $P \in \mathbf{P}$ on \mathbb{R}^k and consider the problem of testing

$$(1) \quad H_0: P \in \mathbf{P}_0 \quad \text{versus} \quad H_1: P \in \mathbf{P}_1,$$

where

$$(2) \quad \mathbf{P}_0 = \{P \in \mathbf{P}: \mathbb{E}_P[W_i] \leq 0\}$$

and $\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$. Here, the inequality in (2) is intended to be interpreted component-wise and \mathbf{P} is a “large” class of possible distributions for the observed data. Indeed, we will only impose below a mild (standardized) uniform integrability requirement on \mathbf{P} . Our goal is to construct tests $\phi_n = \phi_n(W_1, \dots, W_n)$ of (1) that are uniformly consistent in level, that is,

$$(3) \quad \limsup_{n \rightarrow \infty} \sup_{P \in \mathbf{P}_0} \mathbb{E}_P[\phi_n] \leq \alpha$$

for some prespecified value of $\alpha \in (0, 1)$.

¹We thank the co-editor, four anonymous referees, Ivan A. Canay, and Patrik Guggenberger for helpful comments that have improved the exposition of this paper. J. P. Romano’s research supported by NSF Grant DMS-0707085. A. M. Shaikh research supported by NSF Grant DMS-1227091 and the Alfred P. Sloan Foundation.

In the interest of constructing tests of (1) that not only satisfy (3), but also have good power properties, it may be desirable to incorporate information about which components of $\mathbb{E}_P[W_i]$ are “negative.” Examples of tests that incorporate such information implicitly using subsampling include Romano and Shaikh (2008) and Andrews and Guggenberger (2009), whereas examples of tests that incorporate such information more explicitly include the “generalized moment selection” procedures put forward by Andrews and Soares (2010), Canay (2010), and Bugni (2011). Andrews and Barwick (2012a) proposed a refinement of “generalized moment selection” termed “recommended moment selection” and discussed four reasons why such an approach is preferable. Therefore, our theoretical and numerical comparisons will be mainly restricted to the method of Andrews and Barwick (2012a); extensive comparisons with previous methods are already available in that paper.

Our two-step solution to this problem is similar in spirit to the recommended moment selection approach. In the first step, we construct a confidence region for $\mathbb{E}_P[W_i]$ at some “small” significance level $\beta \in [0, \alpha]$. In the second step, we then use this set to provide information about which components of $\mathbb{E}_P[W_i]$ are “negative” when constructing tests of (1). Importantly, similarly to the approach of Andrews and Barwick (2012a), we account in our asymptotic framework for the fact that, with some probability, $\mathbb{E}_P[W_i]$ may not lie in the confidence region, using a Bonferonni-type correction; see Remark 2.4 for further discussion.

Our testing procedure and those just cited are related to Hansen (2005), who uses a similar two-stage approach for the same problem, but does not account for the fact that with some probability, $\mathbb{E}_P[W_i]$ may not lie in the confidence region. He instead assumes that β tends to zero as n tends to infinity and only establishes that his test is pointwise consistent in level instead of the stronger requirement (3). The importance of the distinction between (3) and this weaker requirement has been emphasized in the recent literature on inference in partially identified models; for example, see Imbens and Manski (2004), Romano and Shaikh (2008), and Andrews and Guggenberger (2010). Another important feature of our approach stems from our choice of confidence region for $\mathbb{E}_P[W_i]$. Through an appropriate choice of confidence region for $\mathbb{E}_P[W_i]$, our approach remains computationally feasible even when the number of components of $\mathbb{E}_P[W_i]$, denoted by k , is large. In particular, unlike Hansen (2005), we are able to avoid having to optimize over the confidence region numerically.

As described in Remark 2.6, similar computational problems are also present in the approach put forward by Andrews and Barwick (2012a). As a result, they employ computational shortcuts whose validity is only justified using simulation. Even using these shortcuts, they must restrict attention to situations in which $k \leq 10$, which precludes many economic applications, including entry models, as in Ciliberto and Tamer (2009), where $k = 2^{m+1}$ when there are m firms, or dynamic models of imperfect competition, as in Bajari,

Benkard, and Levin (2007), where k may even be as large as 500. For situations in which $k \leq 10$ and $\alpha = 0.05$, both procedures are equally easy to implement; however, for situations in which $\alpha \neq 0.05$, our procedure is considerably easier to implement even when $k \leq 10$. This feature allows us, for example, to construct p -values more easily than Andrews and Barwick (2012a). On the other hand, in contrast to Andrews and Barwick (2012a), we are unable to establish that the left-hand side of (3) equals α and expect that it is strictly less than α , though we can argue it is not much less than α ; see Remark 2.2. Even so, for the situations when both procedures are available, we find in a simulation study that our procedure is nearly as powerful as the one proposed by Andrews and Barwick (2012a).

Other related literature that uses a two-stage approach like the one described in Hansen (2005) includes Loh (1985), who studies some parametric hypothesis testing problems, and Chernozhukov, Lee, and Rosen (2013), who study testing conditional moment inequalities. Importantly, these papers, like Hansen (2005), do not account for the fact that with some probability the nuisance parameter may not lie in the confidence region. Berger and Boos (1994) and Silvapulle (1996) improve upon Loh (1985) by introducing a Bonferroni-type correction similar to ours. This same idea has been used by Stock and Staiger (1997) to construct a confidence region for the parameters of a linear regression with possibly “weak” instrumental variables, by Romano and Wolf (2000) to construct a confidence interval for a univariate mean that has finite-sample validity and is “efficient” in a precise sense, and by Chernozhukov, Fernandez-Val, Hahn, and Newey (2013) to construct confidence intervals for marginal effects in nonlinear panel data models. Finally, this idea is introduced in a general setting by McCloskey (2012), though the assumptions there technically preclude moment inequality problems; see McCloskey (2012, Section 2.1.3) for further discussion. For a related construction in the context of testing conditional moment inequalities, see Chetverikov (2012).

The remainder of the paper is organized as follows. In Section S.1 of the Supplemental Material to this paper (Romano, Shaikh, and Wolf (2014)), we first consider the testing problem in the simplified setting where $\mathbf{P} = \{N(\mu, \Sigma) : \mu \in \mathbb{R}^k\}$ for a known covariance matrix Σ . Here, it is possible to illustrate the main idea behind our construction more clearly and also to obtain some exact results. In particular, we establish an upper bound on the power function of any level- α test of (1) by deriving the most powerful test against any fixed alternative. This result confirms the bound suggested by simulation in Andrews and Barwick (2012b, Section 7.3). We consider the more general, nonparametric setting in Section 2. We apply our main results to the problem of constructing confidence regions in partially identified models defined by a finite number of moment inequalities in Section 3. Section 4 sheds some light on the behavior of our procedures in finite samples via a simulation study, including an extensive comparison of our procedure with the one proposed recently by Andrews and Barwick (2012a). Proofs of all results can be found in the Appendix.

2. THE NONPARAMETRIC MULTI-SIDED TESTING PROBLEM

Let $W_i, i = 1, \dots, n$, be an i.i.d. sequence of random vectors with distribution $P \in \mathbf{P}$ on \mathbb{R}^k and consider the problem of testing (1). The unknown family of distributions \mathbf{P} will be a nonparametric class of distributions defined by a mild (standardized) uniform integrability condition, as described in the main results below. Before proceeding, we introduce some useful notation. Below, \hat{P}_n denotes the empirical distribution of the $W_i, i = 1, \dots, n$. The notation $\mu(P)$ denotes the mean of P and $\mu_j(P)$ denotes the j th component of $\mu(P)$. Let $\bar{W}_n = \mu(\hat{P}_n)$ and $\bar{W}_{j,n} = \mu_j(\hat{P}_n)$. The notation $\Sigma(P)$ denotes the covariance matrix of P and $\sigma_j^2(P)$ denotes the variance of the j th component of P . The notation $\Omega(P)$ denotes the correlation matrix of P . Let $\hat{\Omega}_n = \Omega(\hat{P}_n)$ and $S_{j,n}^2 = \sigma_j^2(\hat{P}_n)$. Finally, let $S_n^2 = \text{diag}(S_{1,n}^2, \dots, S_{k,n}^2)$.

Our methodology incorporates information about which components of $\mu(P)$ are “negative” by first constructing a (nonparametric) upper confidence rectangle for μ at nominal level $1 - \beta$. Our bootstrap confidence region for this purpose is given by

$$(4) \quad M_n(1 - \beta) = \left\{ \mu \in \mathbb{R}^k : \max_{1 \leq j \leq k} \frac{\sqrt{n}(\mu_j - \bar{W}_{j,n})}{S_{j,n}} \leq K_n^{-1}(1 - \beta, \hat{P}_n) \right\},$$

where

$$(5) \quad K_n(x, P) = P \left\{ \max_{1 \leq j \leq k} \frac{\sqrt{n}(\mu_j(P) - \bar{W}_{j,n})}{S_{j,n}} \leq x \right\}.$$

Next, a test statistic T_n is required such that large values of T_n provide evidence against H_0 . For simplicity, below we consider several different test statistics of the form

$$T_n = T(S_n^{-1} \sqrt{n} \bar{W}_n, \hat{\Omega}_n)$$

for some function $T : \mathbb{R}^k \times (\mathbb{R}^k)^2 \rightarrow \mathbb{R}$ that is continuous in both arguments and weakly increasing in each component of its first argument. As in [Andrews and Barwick \(2012a\)](#), other test statistics may be considered as well. In particular, we consider

$$(6) \quad T_n^{\max} = \max_{1 \leq j \leq k} \frac{\sqrt{n} \bar{W}_{j,n}}{S_{j,n}},$$

$$(7) \quad T_n^{\text{qlr}} = \inf_{t \in \mathbb{R}^k : t < 0} Z_n(t)' \hat{\Omega}_n^{-1} Z_n(t),$$

where

$$Z_n(t) = \left(\frac{\sqrt{n}(\bar{W}_{1,n} - t)}{S_{1,n}}, \dots, \frac{\sqrt{n}(\bar{W}_{k,n} - t)}{S_{k,n}} \right)$$

and the inequality in the infimum is interpreted component-wise. Following Andrews and Barwick (2012a), we also consider an “adjusted” version of T_n^{qlr} in which $\hat{\Omega}_n$ is replaced with

$$\tilde{\Omega}_n = \max\{\varepsilon - \det(\hat{\Omega}_n), 0\} \cdot I_k + \hat{\Omega}_n$$

for some fixed $\varepsilon > 0$, with I_k denoting the k -dimensional identity matrix, that is,

$$(8) \quad T_n^{\text{qlr,ad}} = \inf_{t \in \mathbb{R}^k: t < 0} Z_n(t)' \tilde{\Omega}_n^{-1} Z_n(t).$$

This modification accommodates situations in which $\Omega(P)$ may be singular. Finally, we also consider the “modified method of moments” test statistic of Andrews and Soares (2010) defined as

$$(9) \quad T_n^{\text{mmm}} = \sum_{j=1}^k \left(\frac{\sqrt{n} \bar{W}_{j,n}}{S_{j,n}} \right)^2 \cdot \mathbb{1}\{\bar{W}_{j,n} > 0\}.$$

We also require a critical value with which to compare T_n . For $x \in \mathbb{R}$ and $\lambda \in \mathbb{R}^k$, let

$$(10) \quad J_n(x, \lambda, P) = P\{T(S_n^{-1}(\sqrt{n}(\bar{W}_n - \mu(P))) + S_n^{-1}\sqrt{n}\lambda, \hat{\Omega}_n) \leq x\}.$$

Note that

$$(11) \quad P\{T_n \leq x\} = J_n(x, \mu(P), P).$$

Importantly, for any x and P , $J_n(x, \lambda, P)$ is nonincreasing in each component of λ . It is natural to replace P in the right-hand side of (11) with \hat{P}_n , but this approximation to the distribution of T_n fails when P is on the “boundary” of the null hypothesis; for example, see Andrews (2000). On the other hand, if $\mu(P)$ were known exactly, then one could plug in this value for $\mu(P)$ and replace the final P in the right-hand side of (11) with \hat{P}_n . Obviously, $\mu(P)$ is not known exactly, but we may use the confidence region for $\mu(P)$ defined in (4) to limit the possible values for $\mu(P)$. This idea leads us to consider the critical value defined by

$$(12) \quad \hat{c}_n(1 - \alpha + \beta) = \sup_{\lambda \in M_n(1-\beta) \cap \mathbb{R}_-^k} J_n^{-1}(1 - \alpha + \beta, \lambda, \hat{P}_n),$$

where $\mathbb{R}_- = (-\infty, 0]$. The addition of β to the quantile is necessary to account for the possibility that $\mu(P)$ may not lie in $M_n(1 - \beta)$. It may be removed by allowing β to tend to zero with the sample size. However, the spirit of this

paper, as well as Andrews and Barwick (2012a), is to account for the selection of moments in order to achieve better finite-sample size performance; see Remark 2.4 below for further discussion.

The calculation of $\hat{c}_n(\cdot)$ in (12) is straightforward because $J_n^{-1}(1 - \alpha + \beta, \lambda, \hat{P}_n)$ is nondecreasing in each component of λ . It follows that the supremum in (12) is attained when $\lambda = \lambda^*$ has j th component equal to the minimum of zero and the upper confidence bound for the μ_j , that is,

$$(13) \quad \lambda_j^* = \min \left\{ \bar{W}_{j,n} + \frac{S_{j,n} K_n^{-1}(1 - \beta, \hat{P}_n)}{\sqrt{n}}, 0 \right\}.$$

Then,

$$(14) \quad \hat{c}_n(1 - \alpha + \beta) = J_n^{-1}(1 - \alpha + \beta, \lambda^*, \hat{P}_n).$$

Since $\beta \in (0, \alpha)$, we define our test so that it fails to reject the null hypothesis not only whenever T_n is less than or equal to the critical value defined above, but also whenever $M_n(1 - \beta) \subseteq \mathbb{R}_-^k$. Formally, our test is, therefore, given by

$$(15) \quad \phi_n = \phi_n(\alpha, \beta) = 1 - \mathbb{1} \left\{ \left\{ M_n(1 - \beta) \subseteq \mathbb{R}_-^k \right\} \cup \left\{ T_n \leq \hat{c}_n(1 - \alpha + \beta) \right\} \right\},$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. The following theorem shows that this test controls the probability of a Type I error uniformly over \mathbf{P} in the sense that (3) holds, as long as \mathbf{P} satisfies a mild (standardized) uniform integrability condition.

THEOREM 2.1: *Let $W_i, i = 1, \dots, n$, be an i.i.d. sequence of random vectors with distribution $P \in \mathbf{P}$ on \mathbb{R}^k . Suppose \mathbf{P} is such that, for all $1 \leq j \leq k$,*

$$(16) \quad \limsup_{\lambda \rightarrow \infty} \sup_{P \in \mathbf{P}} \mathbb{E}_P \left[\left(\frac{W_{j,1} - \mu_j(P)}{\sigma_j(P)} \right)^2 \mathbb{1} \left\{ \left| \frac{W_{j,1} - \mu_j(P)}{\sigma_j(P)} \right| > \lambda \right\} \right] = 0.$$

Fix $0 \leq \beta \leq \alpha$. The test ϕ_n of (1) defined by (15) with T_n given by (6), (8), or (9) satisfies (3).

REMARK 2.1: If, in addition to satisfying the requirements of Theorem 2.1, \mathbf{P} is required to satisfy $\inf_{P \in \mathbf{P}} \det(\Omega(P)) > 0$, then the conclusion of Theorem 2.1 holds when T_n is given by (7).

REMARK 2.2: By arguing as in Remark S.2 in the Supplemental Material, it is in fact possible to show that the left-hand side of (3) is at least $\alpha - \beta$, so that, for small β , the test is not overly conservative.

REMARK 2.3: In some cases, the null hypothesis may be such that some components of $\mathbb{E}_P[W_i]$ are equal to zero rather than less than or equal to zero. That is, rather than testing that P belongs in \mathbf{P}_0 given by (2), the problem is to test that P belongs to $\tilde{\mathbf{P}}_0$ given by

$$\tilde{\mathbf{P}}_0 = \{P \in \mathbf{P} : \mathbb{E}_P[W_{j,1}] = 0 \text{ for } j \in J_1, \mathbb{E}_P[W_{j,1}] \leq 0 \text{ for } j \in J_2\},$$

where J_1 and J_2 form a partition of $\{1, \dots, k\}$. Such a situation may be accommodated in the framework described above by writing $\mathbb{E}_P[W_{j,1}] = 0$ as two inequalities $\mathbb{E}_P[W_{j,1}] \leq 0$ and $-\mathbb{E}_P[W_{j,1}] \leq 0$. Note that it may be possible to improve upon this approach by exploiting the additional structure of the null hypotheses, as is done in Remark S.4 in the Supplemental Material.

REMARK 2.4: For $\beta = \beta_n$ tending to zero, it follows from our analysis that the test $\phi_n^*(\beta_n)$, where

$$\phi_n^*(\beta) = 1 - \mathbb{1}\{\{M_n(1 - \beta) \subseteq \mathbb{R}_-^k\} \cup \{T_n \leq \hat{c}_n(1 - \alpha)\}\},$$

satisfies

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathbf{P}_0} \mathbb{E}_P[\phi_n^*(\beta_n)] \leq \alpha$$

under the assumptions of Theorem 2.1. To see this, suppose that the assumptions of Theorem 2.1 hold. Let $\phi_n = \phi_n(\alpha, \beta)$ be defined as in (15). Fix any $\varepsilon > 0$. By monotonicity, we have, for all large enough n , that $M_n(1 - \beta_n) \subseteq M_n(1 - \varepsilon)$. Hence, for all such n , we have that $\phi_n^*(\beta_n) \leq \phi_n^*(\varepsilon)$. Moreover, $\phi_n(\alpha + \varepsilon, \varepsilon) = \phi_n^*(\varepsilon)$. It, therefore, follows from Theorem 2.1 that

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathbf{P}_0} \mathbb{E}_P[\phi_n^*(\beta_n)] \leq \limsup_{n \rightarrow \infty} \sup_{P \in \mathbf{P}_0} \mathbb{E}_P[\phi_n(\alpha + \varepsilon, \varepsilon)] \leq \alpha + \varepsilon.$$

Since the choice of $\varepsilon > 0$ was arbitrary, the desired result follows. The test $\phi_n^*(\beta_n)$ defined in this way is similar to the “generalized moment selection” procedures of Andrews and Soares (2010), Canay (2010), and Bugni (2011). On the other hand, the test ϕ_n defined by (15), which accounts for the impact of the choice of β on the finite-sample behavior of the testing procedure, is more similar to the procedure of Andrews and Barwick (2012a).

REMARK 2.5: An “optimal” approach to choosing β is described in Remark S.6. We have found that a reasonable simple choice is $\beta = \alpha/10$. Further discussion is given in Section 4.

REMARK 2.6: For the hypothesis testing problem considered in this section, Andrews and Barwick (2012a) considered an alternative testing procedure that they termed “recommended moment selection.” In order to describe a version

of their method based on the bootstrap, fix $\kappa < 0$. Let $\hat{\lambda}_n$ be the k -dimensional vector whose j th component equals zero if $\sqrt{n}\bar{W}_{j,n}/S_{j,n} > \kappa$ and $-\infty$ otherwise (or, for practical purposes, some very large negative number). Define the “size correction factor”

$$(17) \quad \hat{\eta}_n = \inf \left\{ \eta > 0 : \sup_{\lambda \in \mathbb{R}^k : \lambda \leq 0} J_n(J_n^{-1}(1 - \alpha, \hat{\lambda}_n, \hat{P}_n) + \eta, \lambda, \hat{P}_n) \geq \alpha \right\}.$$

The proposed test is then given by

$$\phi_n(\alpha) = \mathbb{1}\{T_n > J_n^{-1}(1 - \alpha, \hat{\lambda}_n, \hat{P}_n) + \hat{\eta}_n\},$$

where T_n is given by T_n^{qlr} or $T_n^{\text{qlr,ad}}$; see (7) and (8). The addition of $\hat{\eta}_n$ is required because, in order to allow the asymptotic framework to better reflect the finite-sample situation, the authors did not allow κ to tend to zero with the sample size n . As explained in Remark S.5 in the Supplemental Material, determination of $\hat{\eta}_n$ defined in (17) is computationally prohibitive, even in a parametric setting. This remains true here, so the authors resorted to an approximation to the supremum in (17) analogous to the one described in Remark S.5. The authors provided an extensive simulation study, but no proof, in favor of this approximation and restricted attention to situations in which $k \leq 10$ and $\alpha = 0.05$. The authors also provided simulation-based evidence to support a further approximation to $\hat{\eta}_n$ that only depends on k and the smallest off-diagonal element of $\hat{\Omega}_n$. A data-dependent way of choosing κ similar to the way of choosing β described in Remark S.6 was described as well.

3. CONFIDENCE REGIONS FOR PARTIALLY IDENTIFIED MODELS

In this section, we consider the related problem of constructing a confidence region for identifiable parameters that is uniformly consistent in level. Concretely, let $X_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on some general sample space \mathcal{S} , where \mathbf{P} is again a nonparametric class of distributions defined by a mild (standardized) uniform integrability requirement on \mathbf{P} . We consider the class of partially identified models in which the *identified set*, $\Theta_0(P)$, is given by

$$(18) \quad \Theta_0(P) = \{ \theta \in \Theta : \mathbb{E}_P[g(X_i, \theta)] \leq 0 \},$$

where Θ is some parameter space (usually some subset of Euclidean space) and $g : \mathcal{S} \times \Theta \rightarrow \mathbb{R}^k$. Here, for each θ , $g(\cdot, \theta)$ is a vector of k real-valued functions, and the inequality in (18) is intended to be interpreted component-wise. We wish to construct random sets $\mathcal{C}_n = \mathcal{C}_n(X_1, \dots, X_n)$ satisfying

$$(19) \quad \liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} \inf_{\theta \in \Theta_0(P)} P\{\theta \in \mathcal{C}_n\} \geq 1 - \alpha$$

for some prespecified $\alpha \in (0, 1)$. As in Romano and Shaikh (2008), we refer to such sets as confidence regions for identifiable parameters that are uniformly consistent in level. Note that, in this paper, we will not consider the construction of confidence regions for the identified set itself; see Chernozhukov, Hong, and Tamer (2007), Bugni (2010), and Romano and Shaikh (2010) for further discussion of such confidence regions.

As in Romano and Shaikh (2008), our construction will be based upon the duality between constructing confidence regions and hypothesis tests. Specifically, we will consider tests of the null hypotheses

$$(20) \quad H_\theta : \mathbb{E}_P[g(X_i, \theta)] \leq 0$$

for each $\theta \in \Theta$ that control the usual probability of a Type I error at level α . To this end, for each $\theta \in \Theta$, let $\phi_n(\theta)$ be the test of (20) given by the following algorithm.

ALGORITHM 3.1:

- (a) Set $W_i = g(X_i, \theta)$.
- (b) Compute the bootstrap quantile $K_n^{-1}(1 - \beta, \hat{P}_n)$, where $K_n(x, P)$ is given by (5).
- (c) Using $K_n^{-1}(1 - \beta, \hat{P}_n)$ from (b), compute $M_n(1 - \beta)$ via equation (4).
- (d) Using $K_n^{-1}(1 - \beta, \hat{P}_n)$ from (b), compute λ^* via equation (13).
- (e) Compute the bootstrap quantile $\hat{c}_n(1 - \alpha + \beta) = J_n^{-1}(1 - \alpha + \beta, \lambda^*, \hat{P}_n)$, where $J_n(x, \lambda, P)$ is given by (10).
- (f) Compute $\phi_n(\theta) = \phi_n$, where ϕ_n is given by (15).

Consider

$$(21) \quad C_n = \{\theta \in \Theta : \phi_n(\theta) = 0\}.$$

The following theorem shows that C_n satisfies (19). In the statement of the theorem, we denote by $\mu_j(\theta, P)$ and $\sigma_j^2(\theta, P)$ the mean and variance, respectively, of $g_j(X_i, \theta)$ under P .

THEOREM 3.1: *Let $X_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. Suppose \mathbf{P} is such that, for all $1 \leq j \leq k$,*

$$\limsup_{\lambda \rightarrow \infty} \sup_{P \in \mathbf{P}} \sup_{\theta \in \Theta_0(P)} \mathbb{E}_P \left[\left(\frac{g_j(X_i, \theta) - \mu_j(\theta, P)}{\sigma_j(\theta, P)} \right)^2 \right] \times \mathbb{1} \left[\left| \frac{g_j(X_i, \theta) - \mu_j(\theta, P)}{\sigma_j(\theta, P)} \right| > \lambda \right] = 0.$$

Then, C_n defined by (21) with T_n given by (6), (8), or (9) satisfies (19).

4. SIMULATION STUDY

The goal of this section is to study the finite-sample performance of our two-step procedure. For the reasons mentioned in the [Introduction](#), the comparison with other procedures is reserved to the newly recommended procedure of [Andrews and Barwick \(2012a\)](#) (henceforth abbreviated as AB) with certain details provided in [Andrews and Barwick \(2012b\)](#). In their notation, the preferred procedure is the “recommended moment selection” (RMS) test based on $(S_2, \varphi^{(1)})$ with data-dependent tuning parameters $\hat{\kappa}$ and $\hat{\eta}$ and it is termed “qlr, ad/*t*-Test/ κ Auto.”

We compare finite-sample performance in terms of both maximum null rejection probability (MNRP) and average power for a nominal level of $\alpha = 0.05$. The design of the simulation study is equal to the one used by AB for their Table III. We focus on results for $k = 2, 4$, and 10. For each value of k , we consider three correlation matrices: Ω_{Neg} , Ω_{Zero} , and Ω_{Pos} . The matrix Ω_{Zero} equals I_k (i.e., the identity matrix). The matrices Ω_{Neg} and Ω_{Pos} are Toeplitz matrices with correlations on the diagonals (as they go away from the main diagonal) given by the following: for $k = 2$: $\rho = -0.9$ for Ω_{Neg} and $\rho = 0.5$ for Ω_{Pos} ; for $k = 4$: $\rho = (-0.9, 0.7, -0.5)$ for Ω_{Neg} and $\rho = (0.9, 0.7, 0.5)$ for Ω_{Pos} ; for $k = 10$: $\rho = (-0.9, 0.8, -0.7, 0.6, -0.5, 0.4, -0.3, 0.2, -0.1)$ for Ω_{Neg} and $\rho = (0.9, 0.8, 0.7, 0.6, 0.5, \dots, 0.5)$ for Ω_{Pos} .

For $k = 2$, the set of μ vectors $\mathcal{M}_2(\Omega)$ for which asymptotic average power is computed includes seven elements: $\mathcal{M}_2(\Omega) = \{(\mu_1, 0), (\mu_2, 1), (\mu_3, 2), (\mu_4, 3), (\mu_5, 4), (\mu_6, 7), (\mu_7, \mu_7)\}$, where μ_j depends on Ω . For brevity, the values of μ_j in $\mathcal{M}_2(\Omega)$ and the sets $\mathcal{M}_k(\Omega)$ for $k = 4, 10$ are given in Section 7.1 of [Andrews and Barwick \(2012b\)](#). We point out, however, that we reverse the signs of the mean vectors used by AB, since in our framework the inequality signs are reversed in the null and alternative hypotheses.

To showcase the value, in terms of power properties, of incorporating information about which components of $\mathbb{E}_p[W_i]$ are “negative,” we also include a one-step procedure which ignores such information. This one-step procedure simply uses $J_n^{-1}(1 - \alpha, \lambda, \hat{P}_n)$ with the “least favorable” value of λ , that is, $\lambda = 0$, as the critical value for the test statistic. Equivalently, it can be described as our two-step procedure using $\beta = 0$. Such an approach is expected to have higher power when all non-positive moments are equal to zero (or at least very close to zero), but is expected to have reduced power when some non-positive moments are far away from zero.

AB found that a bootstrap version of their test has better finite-sample size properties than a version based on asymptotic (normal) critical values. Therefore, we only implement bootstrap versions, both for the qlr, ad/*t*-Test/ κ Auto test and our two-step and one-step procedures. All bootstraps use $B = 499$ resamples; this is also the case for the first step of our two-step procedure.

The two-step procedure uses $\beta = 0.005$ for the construction of the confidence region in the first step. Using larger values of β leads to somewhat reduced average power in general. Lower values of β do not make a noticeable

difference in terms of average power, but require a larger number of bootstrap resamples in the first step. (The reason is that the number of bootstrap samples needed to accurately estimate a β quantile is inversely related to β , for small values of β .)

Unlike Andrews and Barwick (2012b), we do not consider any singular covariance matrices Ω . Therefore, the qlr, ad/*t*-Test/ κ Auto test as well as our two-step and one-step procedures use, for simplicity and reduced computational burden, the “unadjusted” quasi-likelihood ratio test statistic (7) rather than the “adjusted” version (8). For the scenarios that we consider, this does not make any difference.

4.1. Maximum Null Rejection Probabilities

Following AB, to ensure computational feasibility, empirical MNRPs are simulated as the maximum rejection probability over all μ vectors that are composed only of zero and $-\infty$ entries, containing at least one zero entry. So for dimension k , there are a total of $2^k - 1$ null vectors to consider. It is worth emphasizing, however, that it has not been proven that the maximum over these $2^k - 1$ null vectors equals the maximum over all μ vectors satisfying the null.

For each scenario, we use 10,000 repetitions to compute empirical MNRPs. The results are presented in the upper half of Table I and can be summarized as follows; from here on, we use the term AB-Rec to denote the recommended procedure of AB, that is, the qlr, ad/*t*-Test/ κ Auto test.

- All procedures achieve a satisfactory performance.
- The empirical MNRP of the AB-Rec procedure is generally somewhat higher compared to the two-step and one-step procedures.
- The empirical MNRPs are somewhat higher when the distribution of the elements is heavy-tailed (i.e., t_3) or skewed (i.e., χ_3^2) versus standard normal.

4.2. Average Powers

Empirical average powers are computed over a set of m different alternative μ vectors, with $m = 7$ when $k = 2$, $m = 24$ when $k = 4$, and $m = 40$ when $k = 10$. For a fixed k , the specific set of μ vectors depends on the correlation matrix $\Omega \in \{\Omega_{\text{Neg}}, \Omega_{\text{Zero}}, \Omega_{\text{Pos}}\}$; see Andrews and Barwick (2012b, Section 7.2) for the details. For each scenario, we use 10,000 repetitions to compute empirical average powers when $k = 2$ and $k = 4$, and 5,000 repetitions to compute empirical average powers when $k = 10$. Unlike AB, we first report “raw” empirical average powers instead of size-corrected empirical average powers. If anything, this slightly favors the recommended procedure of AB, since our two-step and one-step procedures were seen to have (somewhat) lower empirical MNRPs in general. The results are presented in the lower half of Table I and can be summarized as follows.

TABLE I
 EMPIRICAL MAXIMUM NULL REJECTION PROBABILITIES (MNRPs), UPPER HALF,
 AND EMPIRICAL AVERAGE POWERS, LOWER HALF, OF THE AB-RECOMMENDED
 PROCEDURE, THE TWO-STEP PROCEDURE, AND THE ONE-STEP PROCEDURE^a

Test	Distribution	H_0/H_1	$k = 2$			$k = 4$			$k = 10$		
			Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
AB-Rec	$N(0, 1)$	H_0	5.3	5.1	4.9	5.3	5.0	5.1	5.8	5.9	5.6
Two-Step	$N(0, 1)$	H_0	5.0	4.8	4.5	5.1	4.9	5.0	5.3	5.2	5.4
One-Step	$N(0, 1)$	H_0	5.2	5.1	4.9	4.9	5.0	5.1	5.2	4.9	5.3
AB-Rec	t_3	H_0	6.2	6.2	5.9	5.7	5.9	5.7	5.4	5.5	5.3
Two-Step	t_3	H_0	5.6	5.7	5.6	5.3	5.7	5.4	5.7	5.6	5.6
One-Step	t_3	H_0	5.2	6.1	5.7	4.7	5.3	5.7	5.3	5.2	5.7
AB-Rec	χ_3^2	H_0	5.2	4.9	5.1	5.3	4.8	4.9	5.8	5.9	6.0
Two-Step	χ_3^2	H_0	4.8	4.4	4.8	5.1	4.7	4.8	5.6	5.3	5.7
One-Step	χ_3^2	H_0	4.6	4.9	5.1	4.9	5.0	5.0	5.3	4.9	5.5
AB-Rec	$N(0, 1)$	H_1	64.1	68.1	71.4	59.1	66.6	77.5	54.7	63.6	78.9
Two-Step	$N(0, 1)$	H_1	62.0	65.1	66.4	56.1	60.6	74.4	51.0	54.8	75.6
One-Step	$N(0, 1)$	H_1	52.7	61.1	64.2	41.3	50.4	72.6	23.9	32.6	68.4
AB-Rec	t_3	H_1	68.1	72.4	75.2	63.9	71.5	79.5	58.9	68.2	80.4
Two-Step	t_3	H_1	66.0	69.1	71.0	61.1	66.1	76.6	54.9	58.9	77.4
One-Step	t_3	H_1	61.7	66.2	68.8	46.7	57.2	74.9	27.6	37.7	71.5
AB-Rec	χ_3^2	H_1	69.3	76.4	77.9	63.1	74.5	82.4	57.8	69.8	82.6
Two-Step	χ_3^2	H_1	67.6	73.7	74.3	61.0	70.8	80.1	55.5	63.7	80.7
One-Step	χ_3^2	H_1	63.7	70.1	71.7	46.9	59.5	77.9	26.1	37.2	73.5

^aThe nominal level is $\alpha = 5\%$ and the sample size is $n = 100$. All results are based on 10,000 repetitions when $k = 2, 4$ and on 5,000 repetitions when $k = 10$.

- For every scenario, the AB-Rec procedure has the highest empirical average power and the one-step procedure has the lowest empirical average power. However, this does not mean that the AB-Rec procedure is uniformly more powerful than the other two procedures. For individual alternative μ vectors, even the one-step procedure can have higher empirical power than the AB-Rec procedure; for example, this happens when all non-positive moments are equal to zero.

- The two-step procedure generally picks up most of the difference in empirical average powers between AB-Rec and the one-step procedure; across the 27 scenarios, the average pickup is 74.1% and the median pickup is 76.4%. In particular, the relative improvement of the two-step procedure over the one-step procedure tends to be largest when it is needed most, that is, when the differences between AB-Rec and the one-step procedure are the largest. Such cases correspond to Ω_{Neg} ; across these 9 scenarios, the average pickup of the two-step procedure is 82.2% and the median pickup is 83.7%.

TABLE II
 EMPIRICAL AVERAGE POWERS OF THE AB-RECOMMENDED PROCEDURE
 AND THE TWO-STEP PROCEDURE AND EMPIRICAL “SIZE-CORRECTED”
 AVERAGE POWERS OF THE TWO-STEP PROCEDURE^a

Test	Distribution	H_0/H_1	$k = 2$			$k = 4$		
			Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
AB-Rec	$N(0, 1)$	H_1	64.1	68.1	71.4	59.1	66.6	77.5
Two-Step _{sc}	$N(0, 1)$	H_1	63.3	66.3	67.8	56.7	62.1	75.2
Two-Step	$N(0, 1)$	H_1	62.0	65.1	66.4	56.1	60.6	74.4
AB-Rec	t_3	H_1	68.1	72.4	75.2	63.9	71.5	79.5
Two-Step _{sc}	t_3	H_1	67.5	70.2	72.4	61.7	67.0	77.3
Two-Step	t_3	H_1	66.0	69.1	71.0	61.1	66.1	76.6
AB-Rec	χ_3^2	H_1	69.3	76.4	77.9	63.1	74.5	82.4
Two-Step _{sc}	χ_3^2	H_1	69.0	74.8	75.6	61.8	71.8	80.6
Two-Step	χ_3^2	H_1	67.6	73.7	74.3	61.0	70.8	80.1

^aThe nominal level is $\alpha = 5\%$ and the sample size is $n = 100$. Empirical (size-corrected) average powers are based on 10,000 repetitions.

As mentioned before, reporting “raw” empirical average powers slightly favors the recommended procedure of AB, so we also compute “size-corrected” average powers for the two-step procedure. Because of the extremely high computational burden when $k = 10$, we are only able to do this for $k = 2$ and $k = 4$, however, as follows. For a given combination of $k \in \{2, 4\}$, $\Omega \in \{\Omega_{\text{Neg}}, \Omega_{\text{Zero}}, \Omega_{\text{Pos}}\}$, and $\text{Dist} \in \{N(0, 1), t_3, \chi_3^2\}$, we vary the nominal level α for the two-step procedure, keeping $\beta = 0.005$ fixed, until the resulting MNRP matches that of the AB-Rec procedure with $\alpha = 0.05$. Denote the corresponding nominal level α for the two-step procedure by α_{sc} ; for the 18 different combinations of (k, Ω, Dist) considered, we find that $\alpha_{\text{sc}} \in [0.051, 0.055]$. We then use α_{sc} to compute the “size-corrected” average empirical power for the given combination of (k, Ω, Dist) . The results are presented in Table II. The “fair” comparison is the one between AB-Rec and Two-Step_{sc}. It can be seen that the difference is always smaller than for the “unfair” comparison between AB-Rec and Two-Step.

4.3. Maximum Null Rejection Probabilities for a Large Number of Moment Inequalities

We finally turn attention to a case with a large number of inequalities, that is, a case with $k > 10$, for which the procedures of AB are no longer available.

We feel that it is most informative to compute MNRPs. Since a comparison to AB-Rec (or any other of the procedures suggested by AB) is no longer possible, it is not clear what useful information could be taken away from computing empirical average powers.

TABLE III
 EMPIRICAL MAXIMUM NULL REJECTION PROBABILITIES OF THE TWO-STEP PROCEDURE
 AND THE ONE-STEP PROCEDURE BASED ON VARIOUS TEST STATISTICS^a

Test	Distribution	H_0/H_1	$k = 50, n = 100$			$k = 50, n = 500$		
			T_n^{qlr}	T_n^{MMM}	T_n^{max}	T_n^{qlr}	T_n^{MMM}	T_n^{max}
Two-Step	$N(0, 1)$	H_0	NA	4.9	5.1	4.9	4.8	5.1
One-Step	$N(0, 1)$	H_0	NA	4.5	4.9	5.2	5.1	5.2
Two-Step	t_3	H_0	NA	4.3	4.4	4.4	4.7	4.9
One-Step	t_3	H_0	NA	2.9	2.1	4.7	4.5	4.0
Two-Step	χ_3^2	H_0	NA	4.5	4.7	5.2	5.2	5.1
One-Step	χ_3^2	H_0	NA	3.0	4.3	4.9	5.0	5.2

^aThe nominal level is $\alpha = 5\%$ and the covariance matrix is Ω_{Zero} . All results are based on 5,000 repetitions.

As discussed before, computing MNRPs, in principle, involves the evaluation of $2^k - 1$ NRPs. Given current computational powers, this is infeasible for any value of k much larger than 10. However, for the special case of $\Omega = \Omega_{\text{Zero}}$, the problem is reduced to the evaluation of k NRPs only. This is because, under the identity covariance matrix, for a given number of zero entries, the position of these entries does not matter. So if there are p zero entries, say, one only has to evaluate a single NRP rather than $\binom{k}{p}$ NRPs; and without loss of generality, the corresponding single null vector can be chosen as $(0, \dots, 0, -\infty, \dots, -\infty)'$.

We use $k = 50$, which corresponds to roughly the limit of our computational capabilities. The sample sizes considered are $n = 100, 500$. It turns out that for $n = 100$, in many instances, the qlr test statistic cannot be computed because of numerical difficulties. We suspect that the reason is that, for $(k = 50, n = 100)$, the sample covariance matrix is ill-conditioned; this problem is exacerbated in the bootstrap world where, in a given data set, there are always some repeated observations.

Therefore, in addition to the qlr test statistic, we also consider the following two alternative test statistics: first, the “modified method of moments” (MMM) test statistic T_n^{MMM} defined in (9), and second, the maximum test statistic T_n^{max} defined in (6).

For each scenario, we use 5,000 repetitions to compute empirical MNRPs. The results are presented in Table III and can be summarized as follows.

- For $n = 100$, the results for the qlr test statistic are not available due to the numerical difficulties described above. The other two test statistics yield satisfactory performance throughout, though the one-step procedure is somewhat conservative when the distribution of the elements is heavy-tailed (i.e., t_3) or skewed (i.e., χ_3^2).

- For $n = 500$, both the two-step method and the one-step procedure yield satisfactory performance for all test statistics and all distributions of the elements considered.

APPENDIX

In Section A.1, we establish a series of results that will be used in the proof of Theorem 2.1 in Section A.2. The proof of Theorem 3.1 is then provided in Section A.3.

A.1. Auxiliary Results

LEMMA A.1: Suppose μ_n is a sequence in \mathbb{R}_-^k such that $\mu_n \rightarrow \mu$ with $\mu \in \bar{\mathbb{R}}_-^k = (\mathbb{R}_- \cup \{-\infty\})^k$. For $\tau \in \mathbb{R}^k$ and Γ a positive definite $k \times k$ real matrix, define

$$f_n(\tau, \Gamma) = \inf_{t \in \mathbb{R}^k: t < -\mu_n} \|\tau - t\|_\Gamma,$$

where $\|x\|_\Gamma = (x' \Gamma x)^{1/2}$ for $x \in \mathbb{R}^k$. (Below, we may simply write $\|x\|$ for $\|x\|_{I_k}$.) Suppose $(\tau_n, \Gamma_n) \rightarrow (\tau, \Gamma)$, where Γ is positive definite. Then, $f_n(\tau_n, \Gamma_n) \rightarrow f(\tau, \Gamma)$, where

$$f(\tau, \Gamma) = \inf_{t \in \mathbb{R}^k: t < -\mu} \|\tau - t\|_\Gamma.$$

PROOF: We first argue that $f_n(\tau_n, \Gamma_n) - f_n(\tau, \Gamma) \rightarrow 0$. To see this, first note, by strict convexity and continuity of $\|\Gamma^{1/2}(\tau - t)\|$ as a function of $t \in \mathbb{R}^k$, that there exists $t_n^* \leq -\mu_n$ such that

$$\inf_{t \in \mathbb{R}^k: t < -\mu_n} \|\Gamma^{1/2}(\tau - t)\| = \min_{t \in \mathbb{R}^k: t \leq -\mu_n} \|\Gamma^{1/2}(\tau - t)\| = \|\Gamma^{1/2}(\tau - t_n^*)\|.$$

Next, since $0 \leq -\mu_n$, note that

$$(22) \quad \|\Gamma^{1/2}(\tau - t_n^*)\| \leq \|\Gamma^{1/2}\tau\|.$$

Finally, observe that

$$\begin{aligned} & f_n(\tau_n, \Gamma_n) - f_n(\tau, \Gamma) \\ &= \inf_{t \in \mathbb{R}^k: t < -\mu_n} \|\Gamma_n^{1/2}(\tau_n - t)\| - \inf_{t \in \mathbb{R}^k: t < -\mu_n} \|\Gamma^{1/2}(\tau - t)\| \\ &= \min_{t \in \mathbb{R}^k: t \leq -\mu_n} \|\Gamma_n^{1/2}(\tau_n - t)\| - \min_{t \in \mathbb{R}^k: t \leq -\mu_n} \|\Gamma^{1/2}(\tau - t)\| \\ &\leq \|\Gamma_n^{1/2}(\tau_n - t_n^*)\| - \|\Gamma^{1/2}(\tau - t_n^*)\| \\ &\leq \|\Gamma_n^{1/2}(\tau_n - t_n^*) - \Gamma^{1/2}(\tau - t_n^*)\| \end{aligned}$$

$$\begin{aligned}
 &= \left\| \Gamma_n^{1/2}(\tau_n - \tau) + \Gamma_n^{1/2}(\tau - t_n^*) - \Gamma^{1/2}(\tau - t_n^*) \right\| \\
 &= \left\| \Gamma_n^{1/2}(\tau_n - \tau) + \Gamma_n^{1/2}\Gamma^{-1/2}\Gamma^{1/2}(\tau - t_n^*) - \Gamma^{1/2}(\tau - t_n^*) \right\| \\
 &\leq \left\| \Gamma_n^{1/2}(\tau_n - \tau) \right\| + \left\| \Gamma_n^{1/2}\Gamma^{-1/2} - I_k \right\|_{\text{op}} \left\| \Gamma^{1/2}(\tau - t_n^*) \right\| \\
 &\leq \left\| \Gamma_n^{1/2}(\tau_n - \tau) \right\| + \left\| \Gamma_n^{1/2}\Gamma^{-1/2} - I_k \right\|_{\text{op}} \left\| \Gamma^{1/2}\tau \right\| \\
 &\rightarrow 0,
 \end{aligned}$$

where the first equality follows from the definition of the relevant norms, the second equality follows from strict convexity and continuity, the first inequality follows from the definition of t_n^* and the fact that $t_n^* \leq -\mu_n$, the second inequality follows from the reverse triangle inequality, the third and fourth equalities follow by inspection, the third inequality follows from the triangle inequality and the definition of the operator norm, the fourth inequality follows from (22), and the convergence to zero follows from the assumed convergences of τ_n and Γ_n .

Next, we argue that $f_n(\tau, \Gamma) \rightarrow f(\tau, \Gamma)$. For this purpose, it is useful to assume, without loss of generality, that $\mu_n = (\mu_n^{(1)}, \mu_n^{(2)})$ and $\mu = (\mu^{(1)}, \mu^{(2)})$, where all components of $\mu^{(1)}$ are finite and all components of $\mu^{(2)}$ are infinite. Define $\iota^{(1)}$ to be a vector of ones with the same length as $\mu^{(1)}$; define $\iota^{(2)}$ similarly. First note, for $0 < \varepsilon_n \rightarrow 0$ sufficiently slowly and n sufficiently large, that

$$\begin{aligned}
 \inf_{t \in \mathbb{R}^k: t < -\mu_n} \|\tau - t\|_G &\geq \inf_{t \in \mathbb{R}^k: t < -(\mu^{(1)}, \mu_n^{(2)}) + (\varepsilon_n \iota^{(1)}, 0 \iota^{(2)})} \|\tau - t\|_G \\
 &= \inf_{t \in \mathbb{R}^k: t < -(\mu^{(1)}, \mu_n^{(2)})} \|\tau - (\varepsilon_n \iota^{(1)}, 0 \iota^{(2)}) - t\|_G.
 \end{aligned}$$

But, by identifying τ_n in the preceding paragraph with $\tau - (\varepsilon_n \iota^{(1)}, 0 \iota^{(2)})$ here, we see that the final expression equals

$$(23) \quad \inf_{t \in \mathbb{R}^k: t < -(\mu^{(1)}, \mu_n^{(2)})} \|\tau - t\|_G + o(1).$$

The same argument with $\varepsilon < 0$ establishes that $\inf_{t \in \mathbb{R}^k: t < -\mu_n} \|\tau - t\|_G$ in fact equals (23). To complete the argument, we argue that

$$(24) \quad \inf_{t \in \mathbb{R}^k: t < -(\mu^{(1)}, \mu_n^{(2)})} \|\tau - t\|_G \rightarrow \inf_{t \in \mathbb{R}^k: t < -\mu} \|\tau - t\|_G.$$

To establish this fact, given any subsequence n_k , consider a further subsequence n_{k_ℓ} such that $\mu_{n_{k_\ell}}^{(2)}$ is strictly increasing. By the monotone convergence theorem, we see that

$$\inf_{t \in \mathbb{R}^k: t < -(\mu^{(1)}, \mu_{n_{k_\ell}}^{(2)})} \|\tau - t\|_G \rightarrow \inf_{t \in \mathbb{R}^k: t < -\mu} \|\tau - t\|_G.$$

Hence, (24) holds.

Q.E.D.

LEMMA A.2: Let $W_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on \mathbb{R}^k , where \mathbf{P} satisfies (16). Then, $M_n(1 - \beta)$ defined by (4) satisfies

$$(25) \quad \liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} P\{\mu(P) \in M_n(1 - \beta)\} \geq 1 - \beta.$$

PROOF: Follows immediately from Theorem 3.7 in Romano and Shaikh (2012). Q.E.D.

LEMMA A.3: Consider a sequence $\{P_n \in \mathbf{P}: n \geq 1\}$ where \mathbf{P} is a set of distributions on \mathbb{R}^k satisfying (16). Let $W_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P_n . Suppose

$$\frac{\sqrt{n}\mu_j(P_n)}{\sigma_j(P_n)} \rightarrow -\infty$$

for all $1 \leq j \leq k$. Then,

$$P_n\{M_n(1 - \beta) \subseteq \mathbb{R}_-^k\} \rightarrow 1.$$

PROOF: Note that we may write $M_n(1 - \beta)$ as the set of all $\mu \in \mathbb{R}^k$ such that

$$\mu_j \leq \frac{\sigma_j(P_n)}{\sqrt{n}} \left[\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P_n))}{\sigma_j(P_n)} + \frac{\sqrt{n}\mu_j(P_n)}{\sigma_j(P_n)} + \frac{K_n^{-1}(1 - \beta, \hat{P}_n)}{\sigma_j(P_n)} S_{j,n} \right]$$

for all $1 \leq j \leq k$. From Lemma 11.4.1 of Lehmann and Romano (2005), we see that

$$\frac{\sqrt{n}(\bar{W}_{j,n} - \mu_j(P_n))}{\sigma_j(P_n)} = O_{P_n}(1).$$

By assumption,

$$\frac{\sqrt{n}\mu_j(P_n)}{\sigma_j(P_n)} \rightarrow -\infty.$$

From Lemma 4.8 in Romano and Shaikh (2012), we see that

$$\frac{S_{j,n}}{\sigma_j(P_n)} \xrightarrow{P_n} 1.$$

Finally, note that

$$K_n^{-1}(1 - \beta, \hat{P}_n) = O_{P_n}(1)$$

because, using the Bonferroni inequality, it is asymptotically bounded above by $\Phi^{-1}(1 - \beta/k)$, from which the desired result follows. Q.E.D.

LEMMA A.4: Let \mathbf{P}' be the set of all distributions on \mathbb{R}^k and let \mathbf{P} be a set of distributions on \mathbb{R}^k satisfying (16). For $(P, Q) \in \mathbf{P}' \times \mathbf{P}$, define

$$\rho(Q, P) = \max \left\{ \max_{1 \leq j \leq k} \left\{ \int_0^\infty |r_j(\lambda, Q) - r_j(\lambda, P)| \exp(-\lambda) d\lambda \right\}, \max_{1 \leq j \leq k} \left| \frac{\sigma_j(P)}{\sigma_j(Q)} - 1 \right|, \|\Omega(Q) - \Omega(P)\| \right\},$$

where

$$(26) \quad r_j(\lambda, P) = \mathbb{E}_P \left[\left(\frac{X_j - \mu_j(P)}{\sigma_j(P)} \right)^2 \mathbb{1} \left\{ \left| \frac{X_j - \mu_j(P)}{\sigma_j(P)} \right| > \lambda \right\} \right],$$

and the norm $\|\cdot\|$ is the component-wise maximum of the absolute value of all elements. Let $\{Q_n \in \mathbf{P}' : n \geq 1\}$ and $\{P_n \in \mathbf{P} : n \geq 1\}$ be such that $\rho(P_n, Q_n) \rightarrow 0$ and, for some $\emptyset \neq I \subseteq \{1, \dots, k\}$,

$$\frac{\sqrt{n}\lambda_{j,n}}{\sigma_j(P_n)} \rightarrow -\delta_j \quad \text{for all } j \in I \text{ and some } \delta_j \geq 0$$

and

$$\frac{\sqrt{n}\lambda_{j,n}}{\sigma_j(P_n)} \rightarrow -\infty \quad \text{for all } j \notin I.$$

Then, for T_n given by (6), (8), or (9), we have

$$(27) \quad \limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^k} |J_n(x, \lambda_n, P_n) - J_n(x, \lambda_n, Q_n)| = 0.$$

PROOF: Consider first the case where T_n is given by (6). Note that

$$\frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} = \frac{\sigma_j(P_n)}{S_{j,n}} \frac{\sqrt{n}\lambda_{j,n}}{\sigma_j(P_n)}.$$

From Lemma 4.8 in Romano and Shaikh (2012), we see that

$$\frac{S_{j,n}}{\sigma_j(P_n)} \xrightarrow{P_n} 1.$$

Hence,

$$(28) \quad \frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} \xrightarrow{P_n} -\delta_j \quad \text{for all } j \in I$$

and

$$(29) \quad \frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} \xrightarrow{P_n} -\infty \quad \text{for all } j \notin I.$$

It follows that

$$(30) \quad \begin{aligned} & \max_{1 \leq j \leq k} \left(\frac{\sqrt{n}(\bar{W}_{j,n} - \mu_j(P_n))}{S_{j,n}} + \frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} \right) \\ &= \max_{j \in I} \left(\frac{\sqrt{n}(\bar{W}_{j,n} - \mu_j(P_n))}{S_{j,n}} + \frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} \right) + o_{P_n}(1). \end{aligned}$$

Next, we argue that

$$(31) \quad \begin{aligned} & \max_{1 \leq j \leq k} \left(\frac{\sqrt{n}(\bar{W}_{j,n} - \mu_j(Q_n))}{S_{j,n}} + \frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} \right) \\ &= \max_{j \in I} \left(\frac{\sqrt{n}(\bar{W}_{j,n} - \mu_j(Q_n))}{S_{j,n}} + \frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} \right) + o_{Q_n}(1). \end{aligned}$$

For this purpose, it suffices to show that the convergences in (28) and (29) also hold with P_n replaced by Q_n . To see this, first note that by arguing as in the proof of Lemma 4.11 in Romano and Shaikh (2012), we have that

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} r_j(\lambda, Q_n) = 0.$$

The convergence $\rho(P_n, Q_n) \rightarrow 0$ implies further that

$$\frac{\sigma_j(P_n)}{\sigma_j(Q_n)} \rightarrow 1 \quad \text{for all } 1 \leq j \leq k.$$

Since

$$\frac{\sqrt{n}\lambda_{j,n}}{S_{j,n}} = \frac{\sigma_j(Q_n)}{S_{j,n}} \frac{\sigma_j(P_n)}{\sigma_j(Q_n)} \frac{\sqrt{n}\lambda_{j,n}}{\sigma_j(P_n)},$$

the desired conclusion follows. Finally, (27) now follows from (30) and (31) and by arguing as in the proof of Lemma 4.11 in Romano and Shaikh (2012).

Now consider the case where T_n is given by (8). Note that

$$T_n^{\text{qlr,ad}} = \inf_{t \in \mathbb{R}^k : t < -\sqrt{n}D^{-1}(P_n)\lambda_n} \tilde{Z}_n(t)' \tilde{\Omega}_n D^2(P_n) S_n^{-2} \tilde{Z}_n(t),$$

where

$$\tilde{Z}_n(t) = \left(\frac{\sqrt{n}(\bar{W}_{1,n} - \mu_1(P_n))}{\sigma_1(P_n)} - t_1, \dots, \frac{\sqrt{n}(\bar{W}_{k,n} - \mu_k(P_n))}{\sigma_k(P_n)} - t_k \right)',$$

$$D^2(P_n) = \text{diag}(\sigma_1^2(P_n), \dots, \sigma_k^2(P_n)).$$

Now suppose by way of contradiction that (27) fails. It follows that there exists a subsequence n_k along which the left-hand side of (27) converges to a nonzero constant and

$$(32) \quad \Omega(P_{n_k}) \rightarrow \Omega^*, \quad \text{as well as}$$

$$(33) \quad \left(\frac{\bar{W}_{1,n_k} - \mu_1(P_{n_k})}{\sigma_1(P_{n_k})}, \dots, \frac{\bar{W}_{k,n_k} - \mu_k(P_{n_k})}{\sigma_k(P_{n_k})} \right)'$$

$$\xrightarrow{d} Z \sim N(0, \Omega^*) \quad \text{under } P_{n_k}.$$

Since

$$D^2(P_{n_k})S_{n_k}^{-2} \rightarrow I_k,$$

we have further that

$$(34) \quad \tilde{\Omega}_{n_k} D^2(P_{n_k})S_{n_k}^{-2} \xrightarrow{P_{n_k}} \max\{\varepsilon - \det(\Omega^*), 0\}I_k + \Omega^* = \bar{\Omega}.$$

Note that, along such a subsequence n_k , we also have that

$$(35) \quad \frac{\sqrt{n_k}\lambda_{j,n_k}}{\sigma_j(P_{n_k})} \rightarrow -\delta_j \quad \text{for all } j \in I$$

and

$$(36) \quad \frac{\sqrt{n_k}\lambda_{j,n_k}}{\sigma_j(P_{n_k})} \rightarrow -\infty \quad \text{for all } j \notin I.$$

Hence, by Lemma A.1 and the extended continuous mapping theorem (van der Vaart and Wellner (1996, Theorem 1.11.1)), we have that

$$(37) \quad T_{n_k}^{\text{qlr,ad}} \xrightarrow{d} \inf_{t \in \mathbb{R}^k: t < -\delta} (Z - t)' \bar{\Omega}^{-1} (Z - t) \quad \text{under } P_{n_k}.$$

Note that a similar result under slightly stronger assumptions could be established using, for example, Lemma S.1 in Bugni, Canay, and Guggenberger (2012). Moreover, by Chow and Teicher (1978, Lemma 3, p. 260), we have that

$$\sup_{x \in \mathbb{R}} \left| P_{n_k} \{ T_{n_k}^{\text{qlr,ad}} \leq x \} - P \left\{ \inf_{t \in \mathbb{R}^k: t < -\delta} (Z - t)' \bar{\Omega}^{-1} (Z - t) \leq x \right\} \right| \rightarrow 0,$$

since the distribution of $\inf_{t \in \mathbb{R}^k: t < -\delta} (Z-t)' \bar{\Omega}^{-1} (Z-t)$ is continuous everywhere except possibly at zero and

$$P_{n_k} \{ T_{n_k}^{\text{qtr,ad}} \leq 0 \} \rightarrow P\{Z \leq -\delta\} = P\left\{ \inf_{t \in \mathbb{R}^k: t < -\delta} (Z-t)' \bar{\Omega}^{-1} (Z-t) \leq 0 \right\}.$$

Next, note that by arguing as above, it follows from the assumed convergence $\rho(P_{n_k}, Q_{n_k}) \rightarrow 0$ that (32)–(36) all hold when P_{n_k} is replaced by Q_{n_k} . Hence, by the triangle inequality, we see that along n_k , the left-hand side of (27) must converge to zero, from which the desired result follows.

Finally, consider the test statistic (9), for which the argument is easier. For example, the above argument for (8) can be used with $\tilde{\Omega}_n$ replaced by the identity, so that the convergence (37) holds with $\bar{\Omega}$ replaced by the identity. *Q.E.D.*

LEMMA A.5: Consider a sequence $\{P_n \in \mathbf{P} : n \geq 1\}$ where \mathbf{P} is a set of distributions on \mathbb{R}^k satisfying (16). Let $W_{n,i}, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P_n . Suppose that, for some $\emptyset \neq I \subseteq \{1, \dots, k\}$,

$$\frac{\sqrt{n} \mu_j(P_n)}{\sigma_j(P_n)} \rightarrow -\delta_j \quad \text{for all } j \in I \text{ and some } \delta_j \geq 0$$

and

$$\frac{\sqrt{n} \mu_j(P_n)}{\sigma_j(P_n)} \rightarrow -\infty \quad \text{for all } j \notin I.$$

Then,

$$P_n \{ T_n > J_n^{-1}(1 - \alpha + \beta, \mu(P_n), \hat{P}_n) \} \rightarrow \alpha - \beta.$$

PROOF: Let \mathbf{P}' and $\rho(P, Q)$ be defined as in Lemma A.4. Trivially,

$$P_n \{ \hat{P}_n \in \mathbf{P}' \} \rightarrow 1.$$

From Lemma 4.8 in Romano and Shaikh (2012), we see that

$$\max_{1 \leq j \leq k} \left| \frac{S_{j,n}}{\sigma_j(P_n)} - 1 \right| \xrightarrow{P_n} 0.$$

From Lemma 4.9 in Romano and Shaikh (2012), we see that

$$\| \Omega(\hat{P}_n) - \Omega(P_n) \| \xrightarrow{P_n} 0.$$

It follows from Lemma 4.12 in Romano and Shaikh (2012) that

$$\rho(\hat{P}_n, P_n) \xrightarrow{P_n} 0.$$

The desired result now follows by applying Lemma A.4 with $\lambda_n = \mu(P_n)$ and Theorem 2.4 in Romano and Shaikh (2012). Q.E.D.

A.2. Proof of Theorem 2.1

Suppose by way of contradiction that (3) fails. It follows that there exist a subsequence n_k and $\eta > \alpha$ such that

$$(38) \quad \mathbb{E}_{P_{n_k}}[\phi_{n_k}] \rightarrow \eta.$$

There are two cases to consider.

First, consider the case where there exists a further subsequence (which, by an abuse of notation, we continue to denote by n_k) such that

$$\frac{\sqrt{n_k} \mu_j(P_{n_k})}{\sigma_j(P_{n_k})} \rightarrow -\infty$$

for all $1 \leq j \leq k$. Then, by Lemma A.3, we see that

$$P_{n_k} \{M_{n_k}(1 - \beta) \subseteq \mathbb{R}_-^k\} \rightarrow 1.$$

Hence,

$$\mathbb{E}_{P_{n_k}}[\phi_{n_k}] \rightarrow 0,$$

contradicting (38).

Second, consider the case where there exists a further subsequence (which, by an abuse of notation, we continue to denote by n_k) and $\emptyset \neq I \subseteq \{1, \dots, k\}$ such that

$$\frac{\sqrt{n_k} \mu_j(P_{n_k})}{\sigma_j(P_{n_k})} \rightarrow -\delta_j \quad \text{for all } j \in I \text{ and some } \delta_j \geq 0$$

and

$$\frac{\sqrt{n_k} \mu_j(P_{n_k})}{\sigma_j(P_{n_k})} \rightarrow -\infty \quad \text{for all } j \notin I.$$

Next, recall the definition of $\hat{c}_n(1 - \alpha + \beta)$ in (14) and note that

$$\begin{aligned} \mathbb{E}_{P_{n_k}}[\phi_{n_k}] &\leq P_{n_k} \{T_{n_k} > \hat{c}_{n_k}(1 - \alpha + \beta)\} \\ &\leq P_{n_k} \{T_{n_k} > J_{n_k}^{-1}(1 - \alpha + \beta, \mu(P_{n_k}), \hat{P}_{n_k})\} \\ &\quad + P_{n_k} \{\mu(P_{n_k}) \notin M_{n_k}(1 - \beta)\}. \end{aligned}$$

Then, by Lemmas A.2 and A.5, we have that

$$\limsup_{k \rightarrow \infty} \mathbb{E}_{P_{n_k}} [\phi_{n_k}] \leq \alpha,$$

contradicting (38).

A.3. Proof of Theorem 3.1

Follows immediately from Theorem 2.1 by identifying the distribution of $g(X_i, \theta)$ under $P \in \mathbf{P}$ and $\theta \in \Theta_0(P)$ in the present context with the distribution of W_i under P in Theorem 2.1.

REFERENCES

- ANDREWS, D. W. K. (2000): “Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space,” *Econometrica*, 68, 399–405. [1983]
- ANDREWS, D. W. K., AND P. J. BARWICK (2012a): “Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure,” *Econometrica*, 80 (6), 2805–2826. [1980-1985,1988]
- (2012b): “Supplement to ‘Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure,’” *Econometrica Supplemental Material*, 80 (6), http://www.econometricsociety.org/ecta/supmat/8166_miscellaneous.pdf; http://www.econometricsociety.org/ecta/supmat/8166_data_and_programs-1.zip; http://www.econometricsociety.org/ecta/supmat/8166_data_and_programs-2.zip. [1981,1988,1989]
- 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 (3), 669–709. [1980]
- (2010): “Asymptotic Size and a Problem With Subsampling and With the m out of n Bootstrap,” *Econometric Theory*, 26 (2), 426–468. [1980]
- ANDREWS, D. W. K., AND G. SOARES (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78 (1), 119–157. [1980,1983, 1985]
- BAJARI, P., C. BENKARD, AND J. LEVIN (2007): “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 75 (5), 1331–1370. [1980]
- BERGER, R. L., AND D. D. BOOS (1994): “ P Values Maximized Over a Confidence Set for the Nuisance Parameter,” *Journal of the American Statistical Association*, 89 (427), 1012–1016. [1981]
- BUGNI, F. A. (2010): “Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set,” *Econometrica*, 78 (2), 735–753. [1987]
- (2011): “A Comparison of Inferential Methods in Partially Identified Models in Terms of Error in the Coverage Probability,” Working Paper, Department of Economics, Duke University. [1980,1985]
- BUGNI, F. A., I. A. CANAY, AND P. GUGGENBERGER (2012): “Distortions of Asymptotic Confidence Size in Locally Misspecified Moment Inequality Models,” *Econometrica*, 80 (4), 1741–1768. [1998]
- CANAY, I. A. (2010): “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” *Journal of Econometrics*, 156 (2), 408–425. [1980,1985]
- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, J. HAHN, AND W. NEWEY (2013): “Average and Quantile Effects in Nonseparable Panel Models,” *Econometrica*, 81 (2), 535–580. [1981]
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 75 (5), 1243–1284. [1987]

- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81 (2), 667–737. [1981]
- CHETVERIKOV, D. (2012): "Adaptive Tests of Conditional Moment Inequalities," Working Paper, Department of Economics, University of California-Los Angeles. [1981]
- CHOW, Y. S., AND H. TEICHER (1978): *Probability Theory: Independence, Interchangeability and Martingales*. New York: Springer. [1998]
- CILIBERTO, F., AND E. TAMER (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77 (6), 1791–1828. [1980]
- HANSEN, P. R. (2005): "A Test for Superior Predictive Ability," *Journal of Business & Economic Statistics*, 23, 365–380. [1980,1981]
- IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72 (6), 1845–1857. [1980]
- LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer. [1995]
- LOH, W. (1985): "A New Method for Testing Separate Families of Hypotheses," *Journal of the American Statistical Association*, 80, 362–368. [1981]
- MCCLOSKEY, A. (2012): "Bonferroni-Based Size Correction for Nonstandard Testing Problems," Working Paper, Department of Economics, Brown University. [1981]
- ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138 (9), 2786–2807. [1980,1987]
- (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78 (1), 169–211. [1987]
- (2012): "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap," *The Annals of Statistics*, 40 (6), 2798–2822. [1995-1997,1999,2000]
- ROMANO, J. P., AND M. WOLF (2000): "Finite Sample Nonparametric Inference and Large Sample Efficiency," *The Annals of Statistics*, 28 (3), 756–778. [1981]
- ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2014): "Supplement to 'A Practical Two-Step Method for Testing Moment Inequalities'," *Econometrica Supplemental Material*, 82, http://www.econometricsociety.org/ecta/supmat/11011_miscellaneous.pdf; http://www.econometricsociety.org/ecta/supmat/11011_data_and_programs.zip. [1981]
- SILVAPULLE, M. J. (1996): "A Test in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, 91 (436), 1690–1693. [1981]
- STAIGER, D., AND J. STOCK (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65 (3), 557–586. [1981]
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer. [1998]

Dept. of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305-406, U.S.A.; romano@stanford.edu,

Dept. of Economics, University of Chicago, 1126 e. 59th street, Chicago, IL 60637, U.S.A.; amshaikh@uchicago.edu,

and

Dept. of Economics, University of Zurich, Wilfriedstrasse 6, 8032 Zurich, Switzerland; michael.wolf@econ.uzh.ch.

Manuscript received August, 2012; final revision received April, 2014.