On the Computation of Size-Correct Power-Directed Tests with Null Hypotheses Characterized by Inequalities*

Adam McCloskey†

Brown University

November 2013; This Version: March 2015

Abstract

This paper presents theoretical results and a computational algorithm that allows a practitioner to conduct hypothesis tests in nonstandard contexts under which the null hypothesis is characterized by a finite number of inequalities on a vector of parameters. The algorithm allows one to obtain a test with uniformly correct asymptotic size, while directing power towards alternatives of interest, by maximizing a user-chosen local weighted average power criterion. Existing feasible methods for size control in this context do not allow the user to direct the power of the test toward alternatives of interest while controlling size. This is because presently available theoretical results require the user to search for a maximal empirical quantile over a potentially high-dimensional Euclidean space via repeated Monte Carlo simulation. The theoretical results I establish here reduce the space required for this search to a finite number of points for a large class of test statistics and data-dependent critical values, enabling power-direction to be computationally feasible. The results apply to a wide variety of testing contexts including tests on parameters in partially-identified moment inequality models and tests for the superior predictive ability of a benchmark forecasting model. I briefly analyze the asymptotic power properties of the new testing algorithm over existing feasible tests in a Monte Carlo study.

Keywords: composite hypothesis testing, weighted average power, moment inequalities, partial identification, forecast comparison, monotonicity testing

---

*I am grateful to Juan Carlos Escanciano, Kirill Evdokimov, Patrik Guggenberger, Marc Henry, Francesca Molinari, José Luis Montiel Olea, Joseph Romano and Jörg Stoye for helpful comments and to Simon Freyeldenhoven for excellent research assistance. Support from the NSF under grant SES-1357607 is gratefully acknowledged.

†Department of Economics, Brown University, Box B, 64 Waterman St., Providence, RI, 02912 (adam_mccloskey@brown.edu, http://www.econ.brown.edu/fac/adam_mccloskey/Home.html).

1 Introduction

Hypothesis tests characterized by a finite number of inequalities have appeared in numerous and varied forms throughout the econometrics literature. Though the contexts may at first glance appear distinct, examples of tests sharing this feature include tests of inequality constraints in regression models (Wolak, 1987, 1989, 1991), tests for the superior predictive ability of a forecasting model (White, 2000; Hansen, 2005; Romano and Wolf, 2005), tests of monotonicity in expected asset returns (Patton and Timmermann, 2010; Romano and Wolf, 2013)[1] and perhaps the most studied in the present econometrics literature: tests on parameters in partially identified moment inequality models (e.g., Manski and Tamer, 2002; Chernozhukov et al., 2007; Bajari et al., 2007; Cilberto and Tamer, 2009; Pakes et al., 2011; Beresteanu et al., 2011; Galichon and Henry, 2011; Bontemps et al., 2012). This class of testing problems is nonstandard in the sense that the null hypothesis is composite and the asymptotic distributions of test statistics used to conduct these tests are discontinuous in certain parameters allowed under the null.

To increase the power of asymptotically size-controlled tests, a handful of papers have suggested approaches that use the data to determine which point in the null hypothesis parameter space the critical values (CVs) ought to be constructed from. More specifically, these approaches require a tuning parameter that, along with the data, determines how far a given inequality under the null hypothesis should be set from binding in the construction of the CV. For example, see Andrews and Soares (2010) in the moment inequality context and Hansen (2005) in the context of testing for superior predictive ability. These approaches typically require the tuning parameter to diverge at an appropriate rate to result in a test with correct asymptotic size. To better account for the uncertainty involved in data-dependent CV construction, and consequently better control finite sample size, Andrews and Barwick (2012a) examine an alternative asymptotic approach that does not require such tuning parameter divergence. An appealing feature of this latter approach is that in theory, it should allow the user to choose the tuning parameter to maximize a local weighted average power (WAP) criterion. However, doing so is often computationally infeasible because it requires the user to search for a maximal empirical quantile over a potentially high-dimensional Euclidean space via repeated Monte Carlo simulation. On the other hand, Romano et al. (2014) advocate an approach based upon the Bonferroni inequality which also does not assume tun-

[1] Strictly speaking, the inequality testing framework of this paper applies to a particular form of null and alternative hypotheses for this problem. See Patton and Timmermann (2010) and Romano and Wolf (2013) for details.

ing parameter divergence (see also McCloskey, 2012). However, this latter approach can be conservative, with asymptotic size strictly less than its nominal level and an associated power loss.

The results presented in this paper simultaneously address three major issues of testing in these contexts: size-control, power-maximization and computational feasibility. Like Andrews and Barwick (2012a), I examine a class of test statistics and data-dependent CVs that allow for the construction of asymptotically size-correct tests that can be designed to maximize a WAP criterion. However, I establish theoretical results that substantially reduce the computational burden of test construction by reducing the required search space of maximal empirical quantiles to a finite set of points. More specifically, for a given tuning parameter, the construction of a valid CV requires the computation of a size-correction factor by simulating from asymptotic distributions that are determined by whether each inequality in the null hypothesis is binding or "far" from binding. The number of such distributions is equal to 2^p, where p is the number of inequalities present in the null hypothesis.[2] This contrasts with the existing technology, which would require simulation over an uncountably infinite number of such distributions in a potentially high-dimensional space. Once the relevant search space is determined for a given test statistic, these results naturally lead to a straightforward (WAP-maximizing) algorithm for CV construction.

The test statistics and CVs covered by this approach include many that have already been introduced in various contexts in the literature but also allow for potentially new constructions. Correspondingly, the results not only make WAP maximization feasible for existing testing procedures but also allow for power gains over existing approaches. In testing contexts with a large number of inequalities, even with these new computational simplifications, WAP maximization may not be feasible. However, the results presented here enable the feasible construction of tests that are guaranteed to be non-conservative in the sense of attaining asymptotic size *equal* to the nominal level for any given tuning parameter. I show how the theoretical results and corresponding computational testing algorithm can be applied to a leading example of inequality-based tests: tests of parameters in moment inequality models. Related approaches include Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009b), Fan and Park (2010), Andrews and Soares (2010), Canay (2010), Andrews and Barwick (2012a) and Romano et al. (2014). For a small number of inequalities (four), I compare the local asymptotic power of a test constructed from the

[2]Technically, the required number of such distributions is $2^p - 1$. See Theorem SC and the discussion following it.

WAP-maximizing algorithm presented in this paper with that of Andrews and Barwick (2012a), as the latter is generally considered to have the best power properties of those available in the literature for up to 10 inequalities (and tests of nominal level 5%). For a large number of inequalities (20), I compare the local asymptotic power of a different test with that of the new approach advanced by Romano et al. (2014). The power analysis reveals the potential for substantial power gains by using the new computational simplifications developed here.

The rest of this paper is organized as follows. Section 2 presents the general class of inequality testing problems we are interested in here, along with the classes of test statistics and CVs under study. In Section 3, I provide theoretical results that are instrumental in the feasible construction of uniformly size-controlled tests. I then provide details for power direction via the maximization of a WAP criterion in Section 4. A straightforward computational algorithm is presented there. Section 5 summarizes a brief simulation study comparing the local asymptotic power of two newly feasible testing procedures with the existing procedures of Andrews and Barwick (2012a) and Romano et al. (2014). Technical proofs are contained in a mathematical appendix and figures are collected at the end of the document. In what follows, ∞^p denotes $(\infty, \ldots, \infty)$ with p entries, $\mathbb{R}_{+,\infty} \equiv \mathbb{R}_+ \cup \{\infty\}$, $\mathbb{R}^p_{+,\infty} \equiv \mathbb{R}_{+,\infty} \times \ldots \times \mathbb{R}_{+,\infty}$ with the cross-product taken p times, $\mathbb{R}_{+\infty} \equiv \mathbb{R} \cup \{\infty\}$ and $\mathbb{R}^p_{+\infty} \equiv \mathbb{R}_{+\infty} \times \ldots \times \mathbb{R}_{+\infty}$. Finally, I use the convention that $Y + \infty = \infty$ with probability (wp) 1 for any random variable Y.

2 Class of Testing Problems

In this paper we are interested in testing null hypotheses for which a parameter vector satisfies a set of inequalities. Formally, the null hypothesis is defined as

$$H_0 : \gamma_1 \geq 0, \tag{1}$$

where $\gamma_1 \in \mathbb{R}^p$ and the inequality in (1) is meant to be taken element-by-element. In typical applications, γ_1 is (a normalized version of) a vector of moments of a function of an underlying random vector W for which we observe the realizations $\{W_i\}_{i=1}^n$. For example, γ_1 is a normalized version of $E_F[m(W_i)]$, where F denotes the probability measure generating the data and m is a vector-valued function that is measurable with respect to F.

Running Example: Tests on a Parameter in Moment Inequality Models

In a leading example, testing whether a parameter value satisfies the restrictions of a (partially-identified) moment inequality model, m can also be written as a function of an underlying (finite-dimensional) parameter θ so that the null hypothesis takes the form (1), where $\gamma_{1,j}$ is equal to $E_F[m_j(W_i;\theta_0)]/\sigma_j$ for $j = 1,\ldots,p$, $\{m_j(\cdot,\theta)\}$ are known real-valued moment functions, $\{W_i\}$ are i.i.d. or stationary random vectors with joint distribution F and $\sigma_j^2 = \mathrm{Var}_F(m_j(W_i;\theta_0))$ (see e.g., Andrews and Soares, 2010; Andrews and Barwick, 2012a; Romano et al., 2014). We have a sample of $i = 1,\ldots,n$ observations of W_i. ■

2.1 Test Statistics

Testing hypotheses of the form (1) typically proceeds by constructing a test statistic T_n and examining its asymptotic behavior under H_0. In this context, a complication arises from the fact that the asymptotic distribution of the test statistics used for this type of problem (see the following subsection for examples) is discontinuously nonpivotal, depending upon which of the elements of γ_1 are equal to zero and which are not. When coupled with a CV, in order to establish the asymptotic size of the resulting test, one approach is to examine the test statistic's behavior under appropriate drifting sequences of distributions (see e.g., Andrews and Guggenberger, 2009a, 2010; Andrews et al., 2011). More specifically, the relevant drifting sequences of distributions are characterized asymptotically by a vector of parameters $\gamma_n = (\gamma_{1,n}, \gamma_{2,n})$ such that $n^{1/2}\gamma_{1,n} \to h_1 \in H_1 \equiv \mathbb{R}^p_{+,\infty}$ and $\gamma_{2,n} \to h_2 \in H_2$, where $\gamma_{2,n}$ is a correlation matrix and H_2 is the closure of a set of correlation matrices. Let $\gamma_{n,h}$ denote a sequence with this characterization. Under sequences of distributions characterized by $\gamma_{n,h}$ and H_0, we then obtain weak convergence: $T_n \xrightarrow{d} W_h$, where W_h is a random variable that is completely characterized by the parameter $h \equiv (h_1, h_2) \in H_1 \times H_2 \equiv H$.

It is well known that under the $\gamma_{n,h}$ drifting sequences of distributions, h_1 cannot be consistently estimated. Nevertheless, by replacing population quantities with their finite sample counterparts, one can inconsistently "estimate" h_1 by some $\hat{h}_1$ such that $\hat{h}_1 \xrightarrow{d} h_1 + \mathcal{N}(0_p, h_2)$ and consistently estimate h_2 by some $\hat{h}_2$ such that $\hat{h}_2 \xrightarrow{p} h_2$ under any $\gamma_{n,h}$. The test statistics I focus on are functions of this natural localization parameter estimate, i.e., $T_n = S(\hat{h})$ for some non-negative-valued function S, where $\hat{h} = (\hat{h}_1, \hat{h}_2)$. Though not typically written this way, this is in fact true of the majority of test statistics for tests of (1) found in the literature. The test function corresponding to the modified method of moments (MMM) statistic considered by Chernozhukov et al. (2007), Romano and Shaikh (2008, 2010), Andrews and Guggenberger (2009b) and Andrews and Soares (2010) in moment-inequality testing contexts is one such example:

$$T_n = S(\hat{h}) = \sum_{j=1}^{p} [\hat{h}_{1,j}]_-^2, \text{ where } [x]_- \equiv x\mathbf{1}(x < 0). \tag{2}$$

Other leading examples of such test functions include weighted versions of the MMM test function and the test function corresponding to the minimum statistic (min-stat) considered by White (2000), Hansen (2005) and Romano et al. (2014):[3]

$$T_n = S(\hat{h}) = -\min\{\min_{j=1,\ldots,p} \hat{h}_{1,j}, 0\}. \tag{3}$$

In the typical application, we have a continuous $S(\cdot)$ and $\hat{h} \xrightarrow{d} \widetilde{h} = (h_1 + Z, h_2)$, where $Z \overset{d}{\sim} \mathcal{N}(0_p, h_2)$, so that $S(\hat{h}) \xrightarrow{d} S(\widetilde{h})$. This motivates the following assumption.

Assumption TeF. *For any $h \in H$, $W_h = S(\widetilde{h})$, for which the following holds:*

(i) $S : \mathbb{R}^p_{+\infty} \times H_2 \to \mathbb{R}_+$ is a continuous function that is non-increasing in its first p arguments.

(ii) For any $h_2 \in H_2$, $S(x, h_2) = 0$ if and only if $x \in \mathbb{R}^p_{+,\infty}$.

(iii) For any $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p \in \mathbb{R}_{+\infty}$ and $h_2 \in H_2$, $S(x, h_2)$ is constant in $x_i \in \mathbb{R}_{+,\infty}$.

The MMM test function, (weighted) variants of MMM (see e.g., Andrews and Soares, 2010) and the min-stat test functions clearly satisfy Assumption TeF. Similar assumptions have become standard in the moment-inequality literature. Though the quasi-likelihood ratio (QLR) test function, originally considered for tests of inequality constraints by Kudo (1963), Wolak (1987, 1989, 1991) and Sen and Silvapulle (2004), can be expressed as a function of $\hat{h}$, it violates Assumption TeF(iii) and is thus not covered by the results of this paper.

Running Example: Tests on a Parameter in Moment Inequality Models

In this problem, $\hat{h}_{1,j} = \sqrt{n}\bar{m}_j(\theta_0)/\hat{\sigma}_j(\theta_0)$ for $j = 1, \ldots, p$, where

$$\bar{m}_j(\theta) = n^{-1} \sum_{i=1}^{n} m_j(W_i; \theta) \quad \text{and} \quad \hat{\sigma}_j^2(\theta) = n^{-1} \sum_{i=1}^{n} (m_j(W_i; \theta) - \bar{m}_j(\theta))^2.$$

[3] These papers use a different variation of the statistic as written here. The equivalent formulation of the statistic is $T_n = S(\hat{h}) = \max\{\max_{j=1,\ldots,p} \hat{h}_{1,j}, 0\}$ when the null hypothesis is reversed, i.e., $H_0 : \gamma_1 \leq 0$.

The other parameter γ_2 is the correlation matrix of $m(W_i;\theta_0)$ and

$$\hat{h}_2 = \hat{D}^{-1/2}(\theta_0)\hat{\Sigma}(\theta_0)\hat{D}^{-1/2}(\theta_0), \quad \text{where}$$

$$\hat{\Sigma}(\theta) = n^{-1}\sum_{i=1}^{n}(m(W_i,\theta)-\bar{m}(\theta))(m(W_i,\theta)-\bar{m}(\theta))' \quad \text{and} \quad \hat{D}(\theta) = \text{Diag}(\hat{\Sigma}(\theta)) \quad \text{with}$$

$$m(W_i,\theta) = (m_1(W_i,\theta),\ldots,m_p(W_i,\theta))' \quad \text{and} \quad \bar{m}(\theta) = (\bar{m}_1(\theta),\ldots,\bar{m}_p(\theta))'.$$

Andrews and Barwick (2012b) provide the appropriate characterization of the parameter space H for this problem: it is the parameter space corresponding to the "standard problem" in which there are no restrictions on moment functions beyond the inequality restrictions and correlation matrices are "variation free". See (S4.14) of that paper. ■

2.2 Critical Values

The localization parameter h characterizes the local asymptotic behavior of the test statistic under H_0 and a given finite-sample distribution of T_n under H_0 is typically well-approximated by the distribution of W_h for some $h \in H$. Let $J_h(\cdot)$ denote the distribution function of W_h and $c_h(q) = c_{(h_1,h_2)}(q) = c_{((h_{1,1},\ldots,h_{1,p}),h_2)}(q)$ denote its q^{th} quantile. This motivates the use of size-corrected (SC) CVs that take the following form for tests with asymptotic size equal to α:

$$cv(\hat{h},\alpha) + \eta \equiv c_{((f_1(\hat{h}_{1,1},\hat{h}_2),\ldots,f_p(\hat{h}_{1,p},\hat{h}_2)),\hat{h}_2)}(1-\alpha) + \eta(\hat{h}_2), \tag{4}$$

where $\hat{h} \in \mathbb{R}^p_{+\infty} \times H_2$ is the same "estimate" of h used to construct the test statistic, $f_i : \mathbb{R}_{+\infty} \times H_2 \rightarrow \mathbb{R}_{+,\infty}$ for $i = 1,\ldots,p$ and $\eta : H_2 \rightarrow \mathbb{R}_+$ is a data-adaptive size-correction function (see e.g., Andrews and Barwick, 2012a and McCloskey, 2012). Though h_1 cannot be consistently estimated under data-generating process (DGP) sequences characterized by $\gamma_{n,h}$, the data can still provide information on the true value of h_1 since $\hat{h}_1 \xrightarrow{d} h_1 + \mathcal{N}(0,h_2)$. This motivates the use of CVs that are functions of $\hat{h}_1$. The "transition functions" f_i and size-correction factor $\eta(\hat{h}_2) \geq 0$ are used in the construction of the CV to account for the asymptotic uncertainty involved with using the inconsistent estimate $\hat{h}_1$. For example, to obtain a test with correct asymptotic size, one could use the simple choice of transition function $f_i(\hat{h}_{1,i},\hat{h}_2) = \max\{\hat{h}_{1,i},0\}$ as long as a large enough $\eta(\hat{h}_2)$ is used. However, such a construction often necessitates a large amount of size-correction (i.e., large $\eta(\hat{h}_2)$) for asymptotic size-control, leading to large CVs and low power. Thus, power considerations have led to more complex constructions in the literature, briefly discussed below. Since h_2 can be consistently estimated by $\hat{h}_2$ under $\gamma_{n,h}$, I examine "plug-in" CVs of the form (4).

CVs based on the transition function defined as

$$f_i(\hat{h}_{1,i}, \hat{h}_2) = \begin{cases} 0, & \text{if } \hat{h}_{1,i} \leq K_{1-\beta}(\hat{h}_2) \\ \hat{h}_{1,i} - K_{1-\beta}(\hat{h}_2), & \text{if } \hat{h}_{1,i} > K_{1-\beta}(\hat{h}_2), \end{cases} \tag{5}$$

where $K_{1-\beta}(\hat{h}_2)$ is a CV used to construct a rectangular one-sided confidence set for h_1, were suggested by Romano et al. (2014) in a moment-inequality testing framework. Certain constructions of the Bonferroni and Adjusted-Bonferroni CVs suggested by McCloskey (2012) also fit this context. Replacing "$K_{1-\beta}(\hat{h}_2)$" by a tuning parameter "κ" in the previous display leads to CVs examined by Canay (2010) and Andrews and Barwick (2012a). Allowing κ to be a continuous function of h_2 allows for data-dependent tuning parameter construction that fits the context of (4) above. This type of transition function is continuous. Additional continuous formulations have been suggested by Hansen (2005), Andrews and Soares (2010) and Andrews and Barwick (2012a), and the Type II Robust CV of Andrews and Cheng (2012) is quite similar in spirit. The general form of CV (4) also offers more flexibility in the choice of transition function than what has been previously considered: here we may allow f_i to vary across $i \in \{1, \ldots, p\}$.

These transition functions motivate the following assumption.

Assumption TrF. *The following conditions hold for all $i = 1, \ldots, p$:*

(i) $f_i : \mathbb{R}_{+\infty} \times H_2 \to \mathbb{R}_{+,\infty}$ is continuous.

(ii) f_i is non-decreasing in its first argument.

(iii) For each $h_2 \in H_2$, there is some $K_i \in [0, \infty)$ such that $f_i(z_i, h_2)$ is constant and finite for $z_i \in [-\infty, K_i]$.

(iv) For each $h_2 \in H_2$, $f_i(\infty, h_2) = \infty$.

Though the continuous transition functions mentioned above satisfy Assumption TrF, CVs based upon binary decision rules have also been examined in the literature. In contrast to the examples mentioned above, these essentially involve transition functions that are discontinuous in the localization parameter estimate, in violation of Assumption TrF(i). Perhaps the most popular of these is the "moment selection" CV, used by e.g., Chernozhukov et al. (2007), Andrews and Soares (2010), Bugni (2010) and Andrews and Barwick (2012a). Though not typically formulated this way, these are based upon the following discontinuous transition function:

$$f_i(\hat{h}_{1,i}, \hat{h}_2) = f(\hat{h}_{1,i}) = \begin{cases} 0, & \text{if } \hat{h}_{1,i} \leq \kappa \\ \infty, & \text{if } \hat{h}_{1,i} > \kappa. \end{cases}$$

Other examples of "abrupt transition" CVs include those of Hansen (2005) and Fan and Park (2010).

Finally, least-favorable CVs (see e.g., Andrews and Guggenberger, 2009a, 2009b), corresponding here to $f_i(z_i) = 0$ for all $z_i \in \mathbb{R}_\infty$ and $i = 1, \ldots, p$, violate Assumption TrF(iv). These CVs do not adapt to the data through a transition function and lead to conservative inference when some elements of γ_1 are large and positive. Nevertheless, it is straightforward to show that the results of Theorem SC below still hold for these CVs but $\mathcal{H}_1$ (defined in the theorem) need only contain the single point 0_p.

3 Theoretical Results for Test Implementation

The asymptotic size of a test based on the statistic T_n and CV of the form (4) is defined as follows:

$$\text{AsySz}(T_n, cv(\hat{h}, \alpha) + \eta(\hat{h}_2)) = \limsup_{n \to \infty} \sup_{F \in \mathcal{F}} P_F(T_n > cv(\hat{h}, \alpha) + \eta(\hat{h}_2)),$$

where

$$cv(\hat{h}, \alpha) \equiv c_{((f_1(\hat{h}_{1,1}, \hat{h}_2), \ldots, f_p(\hat{h}_{1,p}, \hat{h}_2)), \hat{h}_2)}(1 - \alpha),$$

P_F is the probability under measure F and $\mathcal{F}$ is the set of probability measures specified by the null hypothesis. The function $\eta : H_2 \to \mathbb{R}_+$ is a data-adaptive size-correction function (see e.g., Andrews and Barwick, 2012a and McCloskey, 2012). The size-correction function (SCF) $\eta(\cdot)$ must be constructed carefully to control the asymptotic size of the test by the nominal level α. This is detailed in the algorithm of the following section.

To use these critical values, I impose a weak continuity condition on the asymptotic versions of the test statistic and localization parameter estimate.

Assumption C. *(i)* $\widetilde{h} = (Z + h_1, h_2)$, *where* $Z \overset{d}{\sim} \mathcal{N}(0_p, h_2)$ *and* h_2 *is positive definite.*

(ii) For each $h \in H$, $J_h(x)$ *is continuous for* $x > 0$.

(iii) Unless $h_1 = \infty^p$, $J_h(x)$ *is strictly increasing for* $x > 0$.

This condition holds in the typical application. Parts (ii) and (iii) combine to form a slightly modified version of Assumption S(e) in Andrews and Barwick (2012a).

By Assumption C(i), under H_0 and $\gamma_{n,h}$, $\hat{h} \xrightarrow{d} \widetilde{h} = (Z + h_1, h_2)$. Hence, by Assumption TeF(i), $T_n = S(\hat{h}) \xrightarrow{d} S(\widetilde{h}) = W_h$. Similarly, by Assumptions TeF(i), TrF(i) and C, if $\eta(\cdot)$ is continuous for all $i = 1, \ldots, p$, then $cv(\hat{h}, \alpha) + \eta(\hat{h}_2)$ is continuous in $\hat{h}$ almost everywhere so that $cv(\hat{h}, \alpha) + \eta(\hat{h}_2) \xrightarrow{d} cv(\widetilde{h}, \alpha) + \eta(h_2)$. Since they are functions of the same underlying random vector $\hat{h}$, joint convergence of T_n and $cv(\hat{h},\alpha)+\eta(\hat{h}_2)$ follows. Hence under H_0 and $\gamma_{n,h}$, the asymptotic probability of rejecting H_0 is $P(S(\widetilde{h})>cv(\widetilde{h},\alpha)+\eta(h_2))$. Using arguments found in, inter alia, Andrews and Guggenberger (2010) and Andrews et al. (2011), this fact allows us to simplify the problem of controlling the asymptotic size of the test to controlling the asymptotic null rejection probability $P(W_h>cv(\widetilde{h},\alpha)+\eta(h_2))$ for all $h\in H$, as described by the following assumption.

Assumption SC. $\text{AsySz}(T_n,cv(\hat{h},\alpha)+\eta(\hat{h}_2))=\sup_{h\in H}P(W_h>cv(\widetilde{h},\alpha)+\eta(h_2))$

Assumption SC can be verified in specific applications via the following more primitive condition that is ensured to hold by proper construction of the SCF. It is similar to Assumptions η1 and η3 of Andrews and Barwick (2012a) and η(i)-(ii) of McCloskey (2012).

Assumption η. *(i) $\eta(\cdot)$ is continuous and*
(ii) $\sup_{h\in H}P(W_h>cv(\widetilde{h},\alpha)+\eta(h_2))=\sup_{h\in H}\lim_{x\downarrow 0}P(W_h>cv(\widetilde{h},\alpha)+\eta(h_2)-x)$.

Part (i) holds by proper construction of the SCF and part (ii) is an unrestrictive continuity condition. Since $W_h=S(\widetilde{h})$, the left-hand and right-hand side quantities inside the probabilities of Assumption η(ii) are very different continuous nonlinear functions of the Gaussian random vector $\widetilde{h}$ under Assumption C(i). This implies that, with the exception of the degenerate case which occurs when $h_1=\infty^p$, the left-hand and right-hand side quantities are equal with probability zero. So long as η is constructed appropriately, these probabilities will never be maximized at an h for which $h_1=\infty^p$, implying that part (ii) holds.[4]

Running Example: Tests on a Parameter in Moment Inequality Models

Under the parameter space (S9.2)-(S9.3) of Andrews and Barwick (2012b), H_0 and $\gamma_{n,h}$, $\hat{h}\xrightarrow{d}\widetilde{h}=(Z+h_1,h_2)$, where $Z\overset{d}{\sim}\mathcal{N}(0,h_2)$ and h_2 is the asymptotic correlation matrix of $m(W_i;\theta_0)$. Lemma 5 of Andrews and Barwick (2012b) provides that $c_h(1-\alpha)$ is continuous in h. The following proposition verifies that, by properly constructing the test statistic and CV, Assumption SC holds.

Proposition MI 1. *In the above moment inequality testing context satisfying the parameter space definitions given by (S9.2)-(S9.3) and (S4.14) of Andrews and Barwick (2012b), under Assumptions TeF, TrF and C, Assumption η implies Assumption SC holds.* ■

[4]See the proof of Theorem SC in the appendix.

We are now prepared to state the main theoretical result of the paper, which is useful for implementing tests with asymptotically correct size based upon the test statistics and CVs described in Section 2.

Theorem SC. *Let* $\mathcal{H}_1 = \{(a_1, \ldots, a_p) : a_i = 0 \text{ or } \infty \text{ for } i = 1, \ldots, p\} \setminus \{\infty^p\}$. *Under Assumptions TeF, TrF, C and SC,*

$$\text{AsySz}(T_n, cv(\hat{h}_n, \alpha) + \eta(\hat{h}_2)) = \sup_{(h_1,h_2)\in\mathcal{H}_1\times H_2} P(W_h > cv(\widetilde{h}, \alpha) + \eta(h_2)).$$

This theorem tells us that, for the tests studied in this paper, only the extreme points of the parameter space H_1 are relevant to establishing asymptotic size control. Note that, although it simplifies the problem of computing asymptotic size, Assumption SC still requires one to search over all null rejection probabilities in a potentially high-dimensional uncountably infinite parameter space $H_1 = \mathbb{R}^p_{+,\infty}$ to find the asymptotic size of a test. Theorem SC reduces the space over which this must be established to a finite number of points, $|\mathcal{H}_1| = 2^p - 1$, making computation of size-correct CVs feasible in practice.

Remark 1. *Theorem SC can be slightly modified if one does not wish to impose part (iv) of Assumption TrF. In this case, the theorem holds with* $\mathcal{H}_1 = \{(a_1, \ldots, a_p) : a_i = 0 \text{ or } \infty \text{ for } i = 1, \ldots, p\}$ *(see the proof of Theorem SC in the Mathematical Appendix).*

4 Implementation of Power-Directed Tests

If we know the local asymptotic power function, we can construct the transition functions f_i to maximize a local WAP criterion while controlling size. For implementation, I restrict focus to transition functions relying on a tuning parameter that depends on the data through $\hat{h}_2$ (see Section 2.2 for examples). More specifically, let $f_i(h_{1,i}, h_2) = g_i(h_{1,i}, \kappa(h_2))$, where $\kappa : H_2 \to \mathbb{R}_+$ is continuous.[5] We want to find the function κ that maximizes a WAP criterion using a SCF to control asymptotic size.

Suppose the null hypothesis is given by (1), where γ_1 is a normalized version of $E_F[m(W_i)]$. Then, under conditions permitting a central limit theorem to hold for the sample mean $n^{-1}\sum_{i=1}^n m(W_i)$ (with $\sqrt{n}$ convergence rate), the contiguous local alternatives take the form $\gamma_{1,n}$ such that $n^{1/2}\gamma_{1,n} \to \mu$ for some $\mu \in \mathbb{R}^p_{+\infty}$ with $\mu_j < 0$ for some $j \in \{1, \ldots, p\}$. The following is a high-level assumption similar to Assumption SC regarding the power of a SC test under a given sequence of local alternatives characterized in the limit by $\mu \in \mathbb{R}^p_{+\infty}$.

[5] The theoretical results of this paper actually allow for a different κ function for each i, i.e., $f_i(h_{1,i}, h_2) = g_i(h_{1,i}, \kappa_i(h_2))$. The algorithm that follows can be easily modified to allow for this. This may result in greater power performance, however at the price of increased computational burden.

Assumption AsyPow SC. *For some* $\mu \in \mathbb{R}^p_{+\infty}$ *and* $h_2 \in H_2$,

$$P_{F_n}(T_n > cv(\hat{h}_n, \alpha) + \eta(\hat{h}_2)) \to P(S(Z+\mu, h_2) > cv((Z+\mu, h_2), \alpha) + \eta(h_2)).$$

Verification of this assumption is quite similar to verification of Assumption SC. It typically follows from continuity conditions and distributional convergence results. As is the case for Assumption SC, Assumption AsyPow SC can be verified in applications via more primitive conditions that hold with proper test construction. The following assumption is the counterpart to Assumption $\eta(ii)$ for local alternative vectors μ of interest.

Assumption AsyPow η**.** *For some* $\mu \in \mathbb{R}^p_{+\infty}$ *and* $h_2 \in H_2$,

$$P(S(Z+\mu, h_2) > cv((Z+\mu, h_2), \alpha)+\eta(h_2)) = \lim_{x \downarrow 0} P(S(Z+\mu, h_2) > cv((Z+\mu, h_2), \alpha)+\eta(h_2)-x).$$

Running Example: Tests on a Parameter in Moment Inequality Models

The contiguous local alternatives in this context are characterized in Andrews and Barwick (2012b). They are given by $\theta_n = \theta_0 - \lambda n^{-1/2}(1+o(1))$ for some λ. Delta-method-based arguments provide that this can be equivalently expressed as $n^{1/2}\gamma_{1,n} \to \mu = h_1 + \Pi\lambda$, where Π is a matrix of partial derivatives. Similarly to Proposition MI 1, the following proposition provides the conditions under which Assumption AsyPow SC holds for this example.

Proposition MI 2. *In the above moment inequality testing context satisfying the parameter space definitions given by (S9.2)-(S9.3) and (S4.14) of Andrews and Barwick (2012b), under Assumptions TeF, TrF, C and LA1-LA3 of Andrews and Barwick (2012b), Assumptions* $\eta(i)$ *and AsyPow* η *imply Assumption AsyPow SC holds.* ■

4.1 Algorithm for Weighted Average Power Maximization

For some $a > 0$, let $\{\mu^1, \ldots, \mu^a\}$ denote the set of relevant local alternative parameter vectors with corresponding weights $\{w_1, \ldots, w_a\}$. For given h_2 and $\{g_i\}_{i=1}^p$, the goal is to then choose $\kappa(h_2)$ to maximize the WAP criterion

$$\sum_{i=1}^{a} w_i P(S(Z+\mu^i, h_2) > cv((Z+\mu^i, h_2), \alpha) + \eta(h_2)). \tag{6}$$

Note that $cv((Z+\mu^i, h_2), \alpha)$ implicitly depends upon $\kappa(h_2)$ via the transition functions $f_i(h_{1,i}, h_2) = g_i(h_{1,i}, \kappa(h_2))$. Presumably the value of h_2 the practitioner is most interested in is $\hat{h}_2$ so that construction of the entire function $\kappa(\cdot)$ is unnecessary in a single given application.

Choosing a tuning parameter function κ to maximize (6) proceeds similarly to the methods used to determine the "recommended" moment selection procedure of Andrews and Barwick (2012a). A key difference here is that Andrews and Barwick (2012a) advocate a particular data-adaptive tuning parameter function that maximizes *a particular* WAP criterion (with equal weights), restricting attention to a particular type of transition function and test function. In contrast, the goal here is to (i) allow for the construction of tests of size different from 5% and for more than 10 inequalities and (ii) allow the user to specify the WAP criterion of interest in order to direct power toward alternatives he considers to be the most relevant to his application. As outlined in the algorithm below, the theoretical results established here make (i) and (ii) computationally tractable for the first time. Moreover, the user may choose both the test statistic and transition functions (provided that they satisfy the relevant assumptions) based upon power and computational tradeoffs, within the context of his testing problem.

Using the results of Theorem SC, for a given test function S and set of transition functions $\{g_i\}_{i=1}^p$ satisfying the above conditions, the following algorithm yields a SC CV producing a test that maximizes the WAP criterion (6) subject to correct asymptotic size.

Algorithm WAP Max.

1. *For a given $\kappa \in \mathbb{R}_+$, compute the $(1-\alpha)^{th}$ empirical quantile of $W_h - cv(\widetilde{h}, \alpha)$ at $h = (h_1, \hat{h}_2)$ for each $h_1 \in \mathcal{H}_1$ ($2^p - 1$ points) via simulation. For each $h_1 \in \mathcal{H}_1$, label this quantity $\eta_{h_1}(\hat{h}_2)$.*[6]

2. *For the given $\kappa \in \mathbb{R}_+$ from step 1., compute $\eta(\hat{h}_2) \equiv \max_{h_1 \in \mathcal{H}_1} \eta_{h_1}(\hat{h}_2)$.*

3. *For the given $\kappa \in \mathbb{R}_+$ from steps 1. and 2. and corresponding $\eta(\hat{h}_2)$ from step 2., compute WAP* (6).

4. *Repeat steps 1.-3. over a grid of $\kappa \in \mathbb{R}_+$, choosing κ to maximize WAP* (6). *Form the SC CV using this κ and corresponding $\eta(\hat{h}_2)$. Reject H_0 if $T_n = S(\hat{h})$ exceeds this CV.*

[6] In practice, large positive numbers may be used in places of the ∞'s in $\mathcal{H}_1$. In the following Monte Carlo analysis, I substitute the number 25 for ∞, following Andrews and Barwick (2012a). Unreported simulation results show that this makes no detectable difference.

Steps 1. and 2. allow one to compute a size-correction $\eta(\hat{h}_2)$ that uniformly controls the null rejection probability of the test at $\hat{h}_2$ for given S, $\{g_i\}_{i=1}^p$ and κ.[7] Existing methods of computation in this context would require one to search over a fine grid of the uncountably infinite and potentially high-dimensional $H_1 = \mathbb{R}^p_{+,\infty}$ parameter space to find maximal null rejection probabilities. More specifically, this would require one to find the smallest $\eta(\hat{h}_2) \geq 0$ such that $\sup_{h_1 \in H_1} P(W_{(h_1,\hat{h}_2)} > cv((\widetilde{h}_1, \hat{h}_2), \alpha) + \eta(\hat{h}_2)) \leq \alpha$. Steps 3.-4. enables one to choose $\hat{\kappa} = \kappa(\hat{h}_2)$ to maximize (6) while simultaneously controlling the asymptotic size of the test.

Remark 2. *If p is too large, even with the results of this paper, WAP maximization may become very computationally burdensome since step 1. needs to be repeatedly computed over $2^p - 1$ points. Nevertheless, steps 1. and 2. can be used to construct size-correct tests for a given $\kappa \in \mathbb{R}_+$, making construction of tests with asymptotic size equal to their nominal level feasible in many previously infeasible cases. Though the approach of Romano et al. (2014) allows one to construct asymptotically size-controlled tests when p is large, these tests do not necessarily have asymptotic size equal to their nominal level. This difference allows the tests presented here to have power gains over those of Romano et al. (2014) in the large p context.*

Remark 3. *It is interesting to note that one may also use the results presented in this paper to compute a non-positive SCF for the CVs used by Romano et al. (2014), provided that the test function used satisfies Assumption TeF. Adding this non-positive SCF to their CVs would allow their tests to have size equal to the nominal level, rather than being bounded above by it.*[8]

5 Local Asymptotic Power Analysis

I now briefly analyze the asymptotic power properties of a test constructed from Algorithm WAP Max. First, it is instructive to examine how the choice of tuning parameter affects CV formation and the subsequent power of the test. To fix ideas, let us consider the MMM test function (2) and CVs constructed using transition function (5) for each $i = 1, \ldots, p$, replacing $K_{1-\beta}$ simply with $\kappa \geq 0$. We now graphically analyze SC CVs computed for $p = 2$ and $\alpha = 0.05$ at the given value $h_2 = I_2$. Figure 1 plots the transition function for two different κ values, 0.1 and 2, as a function of μ_2, where μ_2 takes the place of $\hat{h}_{1,i}$ in (5).

[7] As mentioned above, the CVs are computed in step 2. only at the point $h_2 = \hat{h}_2$, rather than computing the entire function $\eta(\cdot)$, since this is the most relevant point in a given application and $\hat{h}_2$ is consistent under drifting $\gamma_{n,h}$ DGPs.

[8] I thank an anonymous referee for pointing this out.

Figure 2 plots the corresponding SC CVs under the local alternative $\mu = (-1, \mu_2)$, as a function of μ_2. That is, Figure 2 graphs $cv(((-1, \mu_2), I_2), \alpha) + \eta(I_2)$ as a function of μ_2, where $\eta(I_2)$ is determined by steps 1. and 2. of Algorithm WAP for each $\kappa = 0.1$ and 2. The underlying localized quantile function that the CVs are constructed from, $c_{((0,\mu_2),I_2)}(1-\alpha)$, is also included in the figure for comparison. Though the μ values are nonrandom, the graph provides a heuristic illustration of how the CVs behave under the alternative hypothesis.

We can see that the CV constructed with $\kappa = 2$ outperforms that constructed with $\kappa = 0.1$ at either small or large values of μ_2 since a smaller CV corresponds to higher power in the resulting test. Conversely, the CV constructed with $\kappa = 0.1$ performs best over an intermediate range of μ_2 values. The features of this specific example generalize to all of the problems considered in this paper: the power properties of tests based upon SC CVs are sensitive to the choice of tuning parameter used. Different choices of κ direct power toward different regions of the alternative hypothesis.

Turning now to local asymptotic power comparison with the test advocated by Andrews and Barwick (2012a), Figures 3 and 4 graph the local asymptotic power curves for $p = 4$ and $\alpha = 0.05$ of (i) SC tests constructed from Algorithm WAP Max using the min-stat test function (3) and transition function (5), now replacing $K_{1-\beta}(\hat{h}_2)$ by the "optimal" κ determined by step 4. of the algorithm; and (ii) Andrews and Barwick's (2012a) SC test based upon the QLR test function, moment selection transition functions and a tuning parameter κ selected to maximize their particular WAP criterion (which was chosen for the test to have good power over a wide range of alternatives). The local alternatives under study are characterized by $\mu = [\mu_1, 1, 1, 1]$ with $\mu_1 < 0$ and local asymptotic power is graphed as a function of μ_1. Figure 3 corresponds to the correlation matrix $h_2 = I_4$ and Figure 4 corresponds to h_2 equal to a Toeplitz matrix with correlations $(-0.9, 0.7, -0.5)$, as was examined by Andrews and Barwick (2012a). Direct comparison of the power curves in the figures is not entirely fair because the κ used for power curves (i) are chosen in Algorithm WAP Max to maximize average power for local alternatives $\mu = [\mu_1, 1, 1, 1]$ with $\mu_1 \in \{-0.1, -0.2, \ldots, -3.5\}$, precisely the type of alternatives under study, while Andrews and Barwick's (2012a) SC test directs power toward a diffuse set of alternatives. Nevertheless, some interesting results emerge.

First, we can see that gains are possible when the practitioner has a specific type of alternative in mind. Second, for the correlation matrix used in Figure 4, it is interesting to note that the relative power performance of tests (i) and (ii) depends on how "far" the local alternative is from H_0, with the min-stat test seeming to perform better for "more local" alternatives. I also made the analogous power comparison for the case of very large positive correlations with the off-diagonal elements of h_2 all equal to 0.9. The results are omitted for brevity but the power of the directed min-stat test and Andrews and Barwick's (2012a) test were very close, with the former tending to be about 0.01 above the latter over the alternatives under study. In summary, for tests of size $\alpha = 0.05$ and $p = 4$ inequalities, Andrews and Barwick's (2012a) test performs well, even against directed alternatives, though significant power gains can be realized.

I conclude with a similar local asymptotic power comparison with the test advocated by Romano et al. (2014), but now for a large number of inequalities: 20. For this large number of inequalities, choosing κ to maximize a WAP criterion is computationally expensive. So for this experiment, I simply set $\kappa = 2$ and computed the appropriate SCF via steps 1. and 2. of Algorithm WAP Max (see Remark 2). Figure 5 graphs the local asymptotic power curves for $p = 20$, $h_2 = I_{20}$ and $\alpha = 0.05$ of (i) SC tests using the MMM test function (2) and transition function (5), replacing $K_{1-\beta}(\hat{h}_2)$ with the number two and (ii) Romano et al.'s (2014) test based upon the QLR test function (which is asymptotically equivalent to MMM in this context) and $K_{1-\alpha/10}(\hat{h}_2)$. The local alternatives under study are characterized by $\mu = [\mu_1, 3, \ldots, 3]$ with $\mu_1 < 0$ and local asymptotic power is graphed as a function of μ_1. Here we can see that there is potential for very large power gains over the potentially conservative approach of Romano et al. (2014), which is based upon the Bonferroni inequality. Though the tuning parameter κ used in the SC test was not chosen to maximize a WAP criterion, by construction, the SC test has asymptotic size equal to the nominal level of the test. The test of Romano et al. (2014) may have asymptotic size strictly less than its nominal level, allowing for potential gains in power by tests with exact asymptotic size.

6 Mathematical Appendix

For notational simplicity, the dependence of $f_i(z_i, h_2)$ on h_2 is suppressed in this appendix. That is, for a given h_2, $f_i(z_i, h_2)$ will be denoted $f_i(z_i)$. I begin by presenting an auxiliary lemma used to prove Theorem SC.

Lemma SC. *Under Assumptions TeF, TrF and C, for $i = 1, \ldots, p$, the function $S((z, h_2)) - cv((z, h_2), \alpha)$, where $cv((z, h_2), \alpha) \equiv c_{((f_1(z_1),\ldots,f_p(z_p)),h_2)}(1-\alpha)$, is (i) continuously non-increasing in $z_i \in (-\infty, K_i]$ and (ii) continuously non-decreasing in $z_i \in [K_i, \infty]$.*

Proof: The following arguments apply to any given $h_2 \in H_2$. By Assumption TrF(iii), $f_i(z_i)$ is constant when $z_i \leq K_i \in [0, \infty)$ so that for $z \in \mathbb{R}^p_{+\infty}$ and any $i \in \{1, \ldots, p\}$, $cv((z, h_2), \alpha)$ is constant in $z_i \in (-\infty, K_i]$. On the other hand, $S((z, h_2))$ is non-increasing in z_i by Assumption TeF(i), which implies that $S((z, h_2)) - cv((z, h_2), \alpha)$ is non-increasing in $z_i \in (-\infty, K_i]$. Assumption TrF(ii) provides that $f_i(\cdot)$ is non-decreasing. Thus, $cv((z, h_2), \alpha)$ is non-increasing in z_i since Assumptions TeF(i) and C(i) provide that $c_h(1 - \alpha)$ is non-increasing in h_1. On the other hand, $S((z, h_2))$ is constant in z_i when $z_i > K_i$ by Assumption TeF(iii) so that $S((z, h_2)) - cv((z, h_2), \alpha)$ is non-decreasing in $z_i \in (K_i, \infty]$. Finally, by Lemma 5 of Andrews and Barwick (2012b), Assumptions TeF(i) and C imply that $c_h(q)$ is continuous in h for any $q \in (0, 1)$. Hence, Assumptions TeF(i) and TrF(i) imply that $S((z, h_2)) - cv((z, h_2), \alpha)$ is continuous in $z \in \mathbb{R}^p_{+\infty}$. ■

Proof of Theorem SC: Given Assumptions C(i) and SC, the goal is to find which values of $h_1 \in \mathbb{R}^p_{+,\infty}$ maximize the quantity

$$P(S(X, h_2) > cv((X, h_2), \alpha) + \eta(h_2)), \tag{7}$$

for a given $h_2 \in H_2$, where $X \overset{d}{\sim} \mathcal{N}(h_1, h_2)$. We will proceed by maximizing (7) in $h_{1,1}$ for any given $h_{1,2}, \ldots, h_{1,p} \in \mathbb{R}^{p-1}_{+,\infty}$. Without loss of generality, suppose $h_{1,2}, \ldots, h_{1,k} < \infty$ and $h_{1,k+1}, \ldots, h_{1,p} = \infty$ for some $k \in \{1, \ldots, p\}$. For a given $h_2 \in H_2$, then let $g : \mathbb{R}^k_{+\infty} \to \mathbb{R}$ such that

$$g(x) \equiv S((x_1, \ldots, x_k, \infty, \ldots, \infty), h_2) - cv((x_1, \ldots, x_k, \infty, \ldots, \infty), \alpha)$$

so that (7) is equal to

$$P(g(X^k) > 0) = \int \mathbf{1}(g(x) > \eta(h_2)) f_h(x) dx,$$

where

$$f_h(x) \equiv (2\pi)^{-k/2} |h_2^k|^{-1/2} \exp(-\frac{1}{2}(x - h_1^k)'(h_2^k)^{-1}(x - h_1^k)),$$

with $h_1^k \equiv (h_{1,1}, \ldots, h_{1,k})$, h_2^k being the upper $k \times k$ block of correlation matrix h_2 and $X^k \overset{d}{\sim} \mathcal{N}(h_1^k, h_2^k)$. Using invertibility implied by Assumption C(i), to simplify notation, let $h_2^k = \Omega$ and partition Ω and Ω^{-1} conformably so that

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \quad \text{and} \quad \Omega^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} \\ \Omega^{21} & \Omega^{22} \end{pmatrix},$$

where Ω_{11} and Ω^{11} are scalar and Ω_{22} and Ω^{22} are $(k-1)\times(k-1)$ submatrices. Also, let abs$(\cdot)$ be an operator such that for an arbitrary matrix A, abs(A) is equal to the matrix composed of the absolute values of the entries of A.

Since $f_h(x)$ is continuously differentiable in $h_{1,1}$ and

$$\left|\frac{\partial f_h(x)}{\partial h_{1,1}}\right| \leq \Omega^{11}\{|x_1-h_{1,1}|+\text{abs}(\Omega_{12})\,\text{abs}(\Omega_{22}^{-1})(|x_2-h_{1,2}|,\ldots,|x_k-h_{1,k}|)'\}f_h(x),$$

which is (Lebesgue) integrable due to the integrability of $|x_i-h_{1,i}|f_h(x)$ for all $i=1,\ldots,k$, application of the dominated convergence and mean value theorems provides that

$$\begin{aligned}\frac{\partial P(g(X^k)>\eta(h_2))}{\partial h_{1,1}} &= \int \mathbf{1}(g(x)>\eta(h_2))\frac{\partial f_h(x)}{\partial h_{1,1}}dx \\ &= \Omega^{11}\int \mathbf{1}(g(x)>\eta(h_2)) \\ &\qquad \times\{(x_1-h_{1,1})-\Omega_{12}\Omega_{22}^{-1}(x_2-h_{1,2},\ldots,x_k-h_{1,k})'\}f_h(x)dx \\ &= \Omega^{11}E[\mathbf{1}(g(Z+h_1^k)>\eta(h_2))(Z_1-\Omega_{12}\Omega_{22}^{-1}(Z_2,\ldots,Z_k)')] \\ &= \Omega^{11}E[\mathbf{1}(g((\widetilde{Z}+\Omega_{12}\Omega_{22}^{-1}\bar{Z}+h_{1,1},\bar{Z}+\bar{h}_1))>\eta(h_2))\widetilde{Z}], \end{aligned} \tag{8}$$

where $Z\overset{d}{\sim}\mathcal{N}(0_k,h_2^k)$, $\bar{Z}\equiv(Z_2,\ldots,Z_k)'$, $\bar{h}_1\equiv(h_{1,2},\ldots,h_{1,k})'$, $\widetilde{Z}\overset{d}{\sim}\mathcal{N}(0,1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21})$ and $\widetilde{Z}$ is independent of $\bar{Z}$. Since $f_h(x)$ is twice continuously differentiable in $h_{1,1}$ and

$$\left|\frac{\partial^2 f_h(x)}{\partial h_{1,1}^2}\right| \leq \{\Omega^{11}+(\Omega^{11})^2[(x_1-h_{1,1})-\Omega_{12}\Omega_{22}^{-1}(x_2-h_{1,2},\ldots,x_k-h_{1,k})']^2\}f_h(x),$$

which is integrable due to the integrability of $f_h(x)$ and $(x_i-h_{1,i})^2f_h(x)$ for all $i=1,\ldots,k$, another application of the dominated convergence and mean value theorems implies that for any given $\bar{h}_1\in\mathbb{R}_+^{k-1}$, (8) is differentiable and thus continuous in $h_{1,1}$.

Letting $\bar{f}(\cdot)$ denote the multivariate normal pdf of $\bar{Z}$ and $\phi(\cdot)$ denote the standard normal pdf, note that for a given $\bar{h}_1\in\mathbb{R}_+^{k-1}$, (8) is equal to

$$\begin{aligned}&\frac{\Omega^{11}}{\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}}}\int\int_0^\infty \mathbf{1}(g(\widetilde{z}+\Omega_{12}\Omega_{22}^{-1}\bar{z}+h_{1,1},\bar{z}+\bar{h}_1)>\eta(h_2)) \\ &\qquad\qquad \times\widetilde{z}\phi(\widetilde{z}/\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}})\bar{f}(\bar{z})d\widetilde{z}d\bar{z} \\ &\quad+\frac{\Omega^{11}}{\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}}}\int\int_{-\infty}^0 \mathbf{1}(g(\widetilde{z}+\Omega_{12}\Omega_{22}^{-1}\bar{z}+h_{1,1},\bar{z}+\bar{h}_1)>\eta(h_2)) \\ &\qquad\qquad \times\widetilde{z}\phi(\widetilde{z}/\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}})\bar{f}(\bar{z})d\widetilde{z}d\bar{z} \\ &=\frac{\Omega^{11}}{\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}}}\int\bar{f}(\bar{z})\left(\int_{S^+_{\bar{z},\bar{h}_1}(h_{1,1})}\widetilde{z}\phi(\widetilde{z}/\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}})d\widetilde{z}\right)d\bar{z}\end{aligned}$$

$$+\frac{\Omega^{11}}{\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}}}\int \bar{f}(\bar{z})\left(\int_{S^-_{\bar{z},\bar{h}_1}(h_{1,1})}\widetilde{z}\phi(\widetilde{z}/\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}})d\widetilde{z}\right)d\bar{z}$$
$$= A_{\bar{h}_1}(h_{1,1}) - B_{\bar{h}_1}(h_{1,1}),$$

say, where

$$S^+_{\bar{z},\bar{h}_1}(h_{1,1}) \equiv \{\widetilde{z} \geq 0 : g(\widetilde{z} + \Omega_{12}\Omega_{22}^{-1}\bar{z} + h_{1,1}, \bar{z} + \bar{h}_1) > \eta(h_2)\},$$
$$S^-_{\bar{z},\bar{h}_1}(h_{1,1}) \equiv \{\widetilde{z} < 0 : g(\widetilde{z} + \Omega_{12}\Omega_{22}^{-1}\bar{z} + h_{1,1}, \bar{z} + \bar{h}_1) > \eta(h_2)\}.$$

Note that for any $h_1^k \in \mathbb{R}^k_+$, $\bar{z} \in \mathbb{R}^{k-1}$ and $\varepsilon \geq 0$, Lemma SC implies that (i) $S^+_{\bar{z},\bar{h}_1}(h_{1,1}) \subseteq S^+_{\bar{z},\bar{h}_1}(h_{1,1}+\varepsilon)$ and (ii) $S^-_{\bar{z},\bar{h}_1}(h_{1,1}+\varepsilon) \subseteq S^-_{\bar{z},\bar{h}_1}(h_{1,1})$.

Case I: for $\bar{h}_1 \in \mathbb{R}^{k-1}_+$ given, there is some $h^*_{1,1} > 0$ such that $E[\mathbf{1}(g(\widetilde{Z} + \Omega_{12}\Omega_{22}^{-1}\bar{Z} + h^*_{1,1}, \bar{Z} + \bar{h}_1) > \eta(h_2))\widetilde{Z}] = 0$, i.e., $A_{\bar{h}_1}(h^*_{1,1}) = B_{\bar{h}_1}(h^*_{1,1})$. Then, for any $0 \leq h_{1,1} < h^*_{1,1}$, $A_{\bar{h}_1}(h_{1,1}) \leq A_{\bar{h}_1}(h^*_{1,1})$ by property (i) since

$$\int_{S^+_{\bar{z},\bar{h}_1}(h_{1,1})}\widetilde{z}\phi(\widetilde{z}/\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}})d\widetilde{z} \leq \int_{S^+_{\bar{z},\bar{h}_1}(h^*_{1,1})}\widetilde{z}\phi(\widetilde{z}/\sqrt{1-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}})d\widetilde{z}$$

for all $\bar{z} \in \mathbb{R}^{k-1}$ and $\Omega^{11} \geq 0$ by the positive semi-definiteness of Ω. Similarly, $B_{\bar{h}_1}(h_{1,1}) \geq B_{\bar{h}_1}(h^*_{1,1})$ by property (ii). Hence, $E[\mathbf{1}(g(\widetilde{Z} + \Omega_{12}\Omega_{22}^{-1}\bar{Z} + h_{1,1}, \bar{Z} + \bar{h}_1) > \eta(h_2))\widetilde{Z}] \leq 0$ for any $0 \leq h_{1,1} < h^*_{1,1}$. An exactly symmetric argument shows that $E[\mathbf{1}(g(\widetilde{Z} + \Omega_{12}\Omega_{22}^{-1}\bar{Z} + h_{1,1}, \bar{Z} + \bar{h}_1) > \eta(h_2))\widetilde{Z}] \geq 0$ for any $h^*_{1,1} < h_{1,1} < \infty$. Hence for any $\bar{h}_1 \in \mathbb{R}^{k-1}_+$, (8) implies that (7) is continuously non-increasing in $h_{1,1} \in [0, h^*_{1,1}]$ and continuously non-decreasing in $h_{1,1} \in [h^*_{1,1}, \infty)$ so that it attains a supremum at either $h_{1,1} = 0$ or $h_{1,1} = \infty$.

Case II: for $\bar{h}_1 \in \mathbb{R}^{k-1}_+$ given, there is no $h^*_{1,1} > 0$ such that $E[\mathbf{1}(g(\widetilde{Z}+\Omega_{12}\Omega_{22}^{-1}\bar{Z}+h^*_{1,1}, \bar{Z}+\bar{h}_1) > \eta(h_2))\widetilde{Z}] = 0$. Then, by the continuity of $E[\mathbf{1}(g(Z + h_1^k) > \eta(h_2))\widetilde{Z}] = E[\mathbf{1}(g(\widetilde{Z} + \Omega_{12}\Omega_{22}^{-1}\bar{Z} + h_{1,1}, \bar{Z} + \bar{h}_1) > \eta(h_2))\widetilde{Z}]$ in $h_{1,1}$, we have either $E[\mathbf{1}(g(Z + h_1^k) > \eta(h_2))\widetilde{Z}] \leq 0$ for all $h_{1,1} \geq 0$ or $E[\mathbf{1}(g(Z + h_1) > \eta(h_2))\widetilde{Z}] \geq 0$ for all $h_{1,1} \geq 0$. In either case (8) implies that (7) is continuous and (weakly) monotone in $h_{1,1} \geq 0$, so that it attains a supremum at either $h_{1,1} = 0$ or $h_{1,1} = \infty$.

The exactly analogous argument can be made for any $h_{1,i}$ with $i = 1, \ldots, p$, implying that (7) is maximized at $h_1 \in \mathbb{R}^p_{+,\infty}$ such that either $h_{1,i} = 0$ or $h_{1,i} = \infty$ for $i = 1, \ldots, p$. Finally, for $h_1 = \infty^p$, $S(\widetilde{h}) = S((\infty^p, h_2)) = 0$ wp 1 by Assumption TeF(ii) and

$$cv(\widetilde{h}, \alpha) + \widetilde{\eta}(h_2) = cv((\infty^p, \alpha)) + \widetilde{\eta}(h_2) = c_{((\infty,\ldots,\infty),h_2)}(1-\alpha) + \widetilde{\eta}(h_2) = \widetilde{\eta}(h_2) \geq 0$$

wp 1 by Asumptions TeF(ii) and TrF(iv) so that $P(S(\widetilde{h}) > cv(\widetilde{h}, \alpha) + \widetilde{\eta}(h_2)) = 0$ at $h_1 = \infty^p$. Thus, using Assumptions C(i) and SC,

$$\text{AsySz}(T_n, cv(\hat{h}_n, \alpha) + \widetilde{\eta}(\hat{h}_2)) = \sup_{h \in H} P(S(\widetilde{h}) > cv(\widetilde{h}, \alpha) + \widetilde{\eta}(h_2))$$

$$= \sup_{h \in H} P(S(\widetilde{h}) - cv(\widetilde{h}, \alpha) - \widetilde{\eta}(h_2) > 0)$$

$$= \sup_{(h_1,h_2) \in \mathcal{H}_1 \times H_2} P(W_h > cv(\widetilde{h}, \alpha) + \widetilde{\eta}(h_2)). \blacksquare$$

Proof of Proposition MI 1: The proof is very similar to the proof of Theorem 1(c) in Andrews and Barwick (2012b) and therefore mostly omitted. Let $\gamma = (\gamma_1, \gamma_2)$. Note that under any subsequence $\{t_n\}$ of $\{n\}$ for which $t_n^{1/2}\gamma_{1,t_n} \to h_1$ for some $h_1 \in \mathbb{R}^p_{+,\infty}$ and $\gamma_{2,t_n} \to h_2$ for some $h_2 \in H_2$, where $\{\gamma_{1,t_n}\}$ is a sequence in $\mathbb{R}^p_+$ and $\{\gamma_{2,t_n}\}$ is a sequence in H_2,

$$\begin{pmatrix} T_{t_n} \\ cv(\hat{h}_{t_n}, \bar{a}(\hat{h}_{2,t_n})) \end{pmatrix} = \begin{pmatrix} S(\hat{h}_{t_n}) \\ cv(\hat{h}_{t_n}, \bar{a}(\hat{h}_{2,t_n})) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S(\widetilde{h}) \\ cv(\widetilde{h}, \bar{a}(h_2)) \end{pmatrix} = \begin{pmatrix} W_h \\ cv(\widetilde{h}, \bar{a}(h_2)) \end{pmatrix},$$

where $\widetilde{h} = (h_1 + Z, h_2)$ with $Z \sim \mathcal{N}(0_p, h_2)$ by the following facts. The parameter space restriction (S9.3) of Andrews and Barwick (2012b), provides that $\hat{h}_{t_n} \xrightarrow{d} \widetilde{h}$, under a DGP sequence characterized by $\{\gamma_{t_n}\}$. Lemma 5 of Andrews and Barwick (2012b) provides that $c_h(1-\alpha)$ is continuous in h for all $\alpha \in (0,1)$. Hence, $cv(h,\alpha) + \eta(h_2)$ is continuous in h almost everywhere by Assumptions TrF(i) and $\eta(i)$. Finally, Assumption TeF(i) ensures continuity of $S(\cdot)$. $\blacksquare$

Proof of Proposition MI 2: The proof is very similar to the proof of Theorem 3 in Andrews and Barwick (2012b), making use of similar reasoning to that used in the proof of Proposition MI 1. $\blacksquare$

References

Andrews, D. W. K., Barwick, P. J., 2012a. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. Econometrica 80, 2805–2826.

Andrews, D. W. K., Barwick, P. J., 2012b. Supplement to 'inference for parameters defined by moment inequalities: A recommended moment selection procedure'. Econometrica Supplementary Material.

Andrews, D. W. K., Cheng, X., 2012. Estimation and inference with weak, semi-strong, and strong identification. Econometrica 80, 2153–2211.

Andrews, D. W. K., Cheng, X., Guggenberger, P., 2011. Generic results for establishing the asymptotic size of confidence sets and tests, Cowles Foundation Discussion Paper No. 1813.

Andrews, D. W. K., Guggenberger, P., 2009a. Hybrid and size-corrected subsampling methods. Econometrica 77, 721–762.

Andrews, D. W. K., Guggenberger, P., 2009b. Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities. Econometric Theory 25, 669–709.

Andrews, D. W. K., Guggenberger, P., 2010. Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Econometric Theory 26, 426–468.

Andrews, D. W. K., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–157.

Bajari, P., Benkard, C. L., Levin, J., 2007. Estimating dynamic models of imperfect competition. Econometrica 75, 1331–1370.

Beresteanu, A., Molchanov, I., Molinari, F., 2011. Sharp identification regions in models with convex moment predictions. Econometrica 79, 1785–1821.

Bontemps, C., Magnac, T., Maurin, E., 2012. Set identified linear models. Econometrica 80, 1129–1155.

Bugni, F. A., 2010. Bootstrap inference in partially identified models defined by moment inequalities: coverage of the identified set. Econometrica 78, 735–753.

Canay, I. A., 2010. EL inference for partially identified models: large deviations optimality and bootstrap validity. Journal of Econometrics 156, 408–425.

Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75, 1243–1284.

Cilberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline markets. Econometrica 77, 1791–1828.

Fan, Y., Park, S. S., 2010. Confidence sets for some partially identified parameters. Economics, Managemnt, and Financial Market 5, 37–87.

Galichon, A., Henry, M., 2011. Set identification in models with multiple equilibria. Review of Economic Studies 78, 1264–1298.

Hansen, P. R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365–380.

Kudo, A., 1963. A multivariate analog of a one-sided test. Biometrika 59, 403–418.

Manski, C. F., Tamer, E., 2002. Inference on regressions with interval data on a regressor or outcome. Econometrica 70, 519–546.

McCloskey, A., 2012. Bonferroni-based size-correction for nonstandard testing problems, Working Paper, Department of Economics, Brown University.

Pakes, A., Porter, J., Ho, K., Ishii, J., 2011. Moment inequalities and their application, Working Paper, Department of Economics, Harvard University.

Patton, A. J., Timmermann, A., 2010. Monotonicity in asset returns: new tests with applications to the term structure, the CAPM and portfolio sorts. Journal of Financial Economics 98, 605–625.

Romano, J. P., Shaikh, A. M., 2008. Inference for identifiable parameters in partially identified econometric models. Journal of Statistical Planning and Inference 138, 2786–2807.

Romano, J. P., Shaikh, A. M., 2010. Inference for the identified set in partially identified econometric models. Econometrica 78, 169–211.

Romano, J. P., Shaikh, A. M., Wolf, M., 2014. A practical two-step method for testing moment inequalities. Econometrica 82, 1979–2002.

Romano, J. P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica 73, 1237–1282.

Romano, J. P., Wolf, M., 2013. Testing for monotonicity in expected asset returns. Journal of Empirical Finance 23, 93–116.

Rosen, A. M., 2008. Confidence sets for partially identified parameters that satisfy moment inequalities. Journal of Econometrics 146, 107–117.

Sen, P. K., Silvapulle, M. J., 2004. Constrained Statistical Inference: Inequality, Order, and Shape Restrictions. Wiley-Interscience, New York.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.

Wolak, F. A., 1987. An exact test for multiple inequality and equality constraints in the linear regression model. Journal of the American Statistical Association 82, 782–793.

Wolak, F. A., 1989. Testing inequality constraints in linear econometric models. Journal of Econometrics 41, 205–235.

Wolak, F. A., 1991. The local nature of hypothesis tests involving inequality constraints in nonlinear models. Econometrica 59, 981–995.

Figure 1: Transition Functions

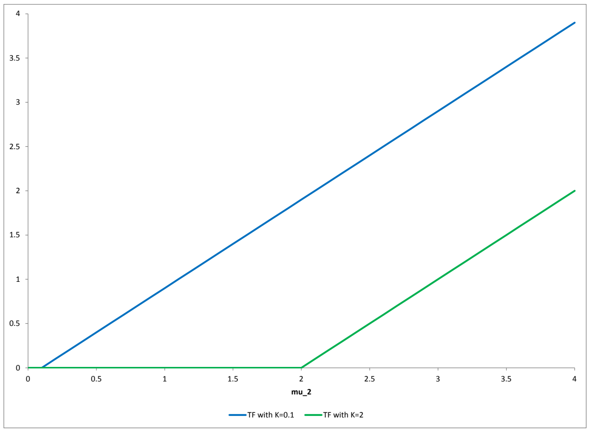

Figure 2: Size-Corrected Critical Values

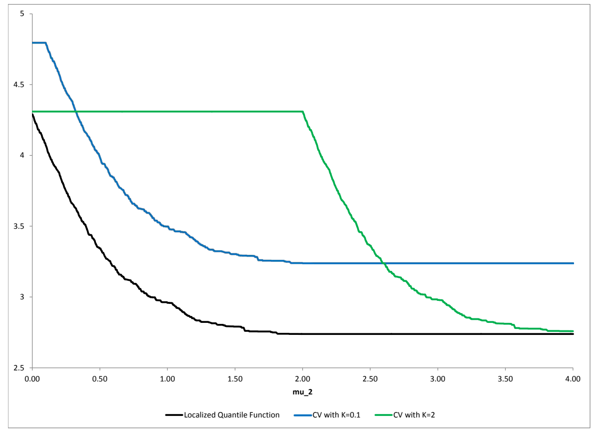

Figure 3: Power $p = 4$, Zero Correlation

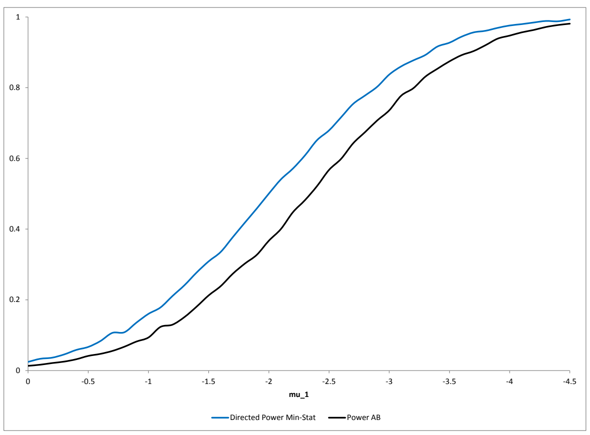

Figure 4: Power $p = 4$, Negative Correlation

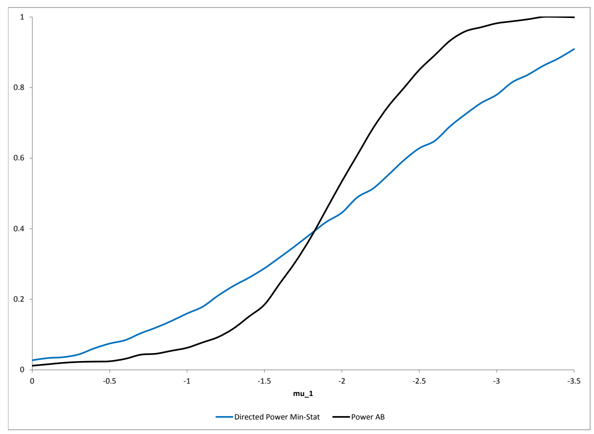

Figure 5: Power $p = 20$, Zero Correlation