SUPPLEMENT TO "A PRACTICAL TWO-STEP METHOD FOR TESTING MOMENT INEQUALITIES"
(*Econometrica*, Vol. 82, No. 5, September 2014, 1979–2002)

By Joseph P. Romano, Azeem M. Shaikh, and Michael Wolf

This document provides additional results for the authors' paper "A Practical Two-Step Method for Testing Moment Inequalities."

S.1. THE GAUSSIAN PROBLEM

IN THIS SECTION, we assume that $W = (W_1, \ldots, W_k)' \sim P \in \mathbf{P} = \{N(\mu, \Sigma) : \mu \in \mathbb{R}^k\}$ for a known covariance matrix Σ. In this setting, we may equivalently describe the problem of testing (1) as the problem of testing

$$(\text{S.1}) \qquad H_0 : \mu \in \Omega_0 \quad \text{versus} \quad H_1 : \mu \in \Omega_1,$$

where

$$(\text{S.2}) \qquad \Omega_0 = \{\mu : \mu_j \leq 0 \text{ for } 1 \leq j \leq k\}$$

and $\Omega_1 = \mathbb{R}^k \setminus \Omega_0$. Here, it is possible to obtain some exact results, so we focus on tests $\phi_n = \phi_n(W_1, \ldots, W_n)$ of (S.1) that satisfy

$$(\text{S.3}) \qquad \sup_{\mu \in \Omega_0} \mathbb{E}_P[\phi_n] \leq \alpha$$

for some prespecified value of $\alpha \in (0, 1)$ rather than (3). In Section S.1.1 below, we first establish an upper bound on the power function of any test of (S.1) that satisfies (S.3) by deriving the most powerful test against any fixed alternative. We then describe our two-step procedure for testing (S.1) in Section S.1.2. Proofs of all results can be found in the Supplement Appendix.

Before proceeding, note that, by sufficiency, we may assume without loss of generality that $n = 1$. Hence, the data consist of a single random variable W distributed according to the multivariate Gaussian distribution with unknown mean vector $\mu \in \mathbb{R}^k$ and known covariance matrix Σ. For $1 \leq j \leq k$, we denote by W_j the jth component of W and by μ_j the jth component of μ. Note further that, because Σ is assumed known, we may assume without loss of generality that its diagonal consists of ones; otherwise, we can simply replace W_j by W_j divided by its standard deviation.

S.1.1. *Power Envelope*

In this subsection only, we assume further that Σ is invertible.

Below, we calculate the most powerful (MP) test of $\mu \in \Omega_0$ satisfying (S.3) against a fixed alternative $\mu = a$, where $a \in \Omega_1$. The power of such a test, as

a function of a, provides an upper bound on the power function of any test of (S.1) satisfying (S.3) and is, therefore, referred to as the power envelope function. In Andrews and Barwick (2012a, 2012b), numerical evidence was given to justify their conjecture of how to calculate the MP test of $\mu \in \Omega_0$ satisfying (S.3) against $\mu = a$ and hence how to calculate the power envelope function. Theorem S.1 below verifies the claim made by Andrews and Barwick (2012a). Note that the power of the MP test of $\mu \in \Omega_0$ satisfying (S.3) against $\mu = a$ depends on a through its "distance" from Ω_0 in terms of the Mahanolobis metric $d(x, y) = \sqrt{(x - y)'\Sigma^{-1}(x - y)}$, that is,

$$(\text{S.4}) \qquad \inf_{\mu \in \Omega_0} \sqrt{\{(\mu - a)'\Sigma^{-1}(\mu - a)\}}.$$

THEOREM S.1: *Let W be multivariate normal with unknown mean vector μ and known covariance matrix Σ. For testing $\mu \in \Omega_0$ against the fixed alternative $\mu = a$, where $a \in \Omega_1$, the MP test satisfying (S.3) rejects for large values of $T = W'\Sigma^{-1}(a - \bar{\mu})$, where*

$$\bar{\mu} = \underset{\mu \in \Omega_0}{\arg\min}(\mu - a)'\Sigma^{-1}(\mu - a).$$

In fact, the distribution that puts mass 1 at the point $\bar{\mu}$ is least favorable, and the critical value at level α can be determined so that

$$P_{\bar{\mu}}\{T > c_{1-\alpha}\} = \alpha.$$

Under $\mu = \bar{\mu}$,

$$\mathbb{E}[T] = \bar{\mu}'\Sigma^{-1}(a - \bar{\mu}),$$
$$\text{Var}[T] = (\bar{\mu} - a)'\Sigma^{-1}(\bar{\mu} - a),$$

so

$$c_{1-\alpha} = \bar{\mu}'\Sigma^{-1}(a - \bar{\mu}) + z_{1-\alpha}\sqrt{(\bar{\mu} - a)'\Sigma^{-1}(\bar{\mu} - a)},$$

where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution. Moreover, the power of this test is given by

$$1 - \Phi\big(z_{1-\alpha} - \sqrt{(\bar{\mu} - a)'\Sigma^{-1}(\bar{\mu} - a)}\big),$$

where $\Phi(\cdot)$ denotes the standard normal c.d.f.

Since the most powerful tests vary as a function of the vector a, it follows that there is no uniformly most powerful test. Furthermore, as argued in Lehmann

(1952), the only unbiased test is the trivial test whose power function is constant and equal to α. Invariance considerations do not appear to lead to any useful simplification of the problem either; also see Andrews (2012) for some negative results concerning similarity.

REMARK S.1: Note that $T = W'\Sigma^{-1}(a - \bar{\mu})$ in Theorem S.1 is a linear combination $\sum_{1 \le j \le k} c_j W_j$ of the $W_1, \ldots, W_k$. Even if all components of a are positive, depending on Σ, $\bar{\mu}$ may not equal zero. One might, therefore, suspect that the test described in Theorem S.1 does not satisfy (S.3). However, the proof of the theorem shows that if $\bar{\mu}$ has any components that are negative, then the corresponding coefficient of W_j in T must be zero; components of $\bar{\mu}$ that are zero have corresponding coefficient of W_j in T that are nonnegative.

S.1.2. *A Two-Step Procedure*

There are, of course, many ways in which to construct a test of (S.1) that controls size at level α. For instance, given any test statistic $T = T(W_1, \ldots, W_k)$ that is nondecreasing in each of its arguments, we may consider a test that rejects H_0 for large values of T. Note that, for any given fixed critical value c, $P_\mu\{T(W_1, \ldots, W_k) > c\}$ is a nondecreasing function of μ. Therefore, if $c = c_{1-\alpha}$ is chosen to satisfy

$$P_0\big\{T(W_1, \ldots, W_k) > c_{1-\alpha}\big\} \le \alpha,$$

then the test that rejects H_0 when $T > c_{1-\alpha}$ is a level α test. A reasonable choice of test statistic T is the likelihood ratio statistic, which is given by

$$(S.5) \qquad T = \inf_{\mu \in \Omega_0} \big\{(W - \mu)'\Sigma^{-1}(W - \mu)\big\}.$$

By analogy with (S.4) and Theorem S.1, rejecting for large values of the "distance" of W to Ω_0 is intuitively appealing. It is easy to see that such a test statistic T is nondecreasing in each of its arguments.

A second choice of monotone test statistic is the "modified method of moments" test statistic

$$T = \sum_{j=1}^{k} W_j^2 \cdot \mathbb{1}\{W_j > 0\}.$$

A further choice of monotone test statistic is the maximal order statistic $T = \max\{W_1, \ldots, W_k\}$. For any given choice of monotone test statistic, a critical value $c_{1-\alpha}$ may be determined as the $1 - \alpha$ quantile of the distribution of T when $(W_1, \ldots, W_k)'$ is multivariate normal with mean 0 and covariance matrix Σ. Unfortunately, as k increases, so does the critical value, which can make it difficult to have any reasonable power against alternatives. The main

idea of our procedure, as well as that of Andrews and Barwick (2012a), is to essentially remove from consideration those μ_j that are "negative." If we can eliminate such μ_j from consideration, then we may use a smaller critical value with the hopes of increased power against alternatives.

Using this reasoning as a motivation, we may use a confidence region to help determine which μ_j are "negative." To this end, let $M(1 - \beta)$ denote an upper confidence rectangle for all the μ_j simultaneously at level $1 - \beta$. Specifically, let

$$(S.6) \qquad M(1 - \beta) = \left\{ \mu \in \mathbb{R}^k : \max_{1 \leq j \leq k}(\mu_j - W_j) \leq K^{-1}(1 - \beta) \right\}$$

$$= \left\{ \mu \in \mathbb{R}^k : \mu_j \leq W_j + K^{-1}(1 - \beta) \text{ for all } 1 \leq j \leq k \right\},$$

where $K^{-1}(1 - \beta)$ is the $1 - \beta$ quantile of the distribution

$$K(x) = P_\mu \left\{ \max_{1 \leq j \leq k}(\mu_j - W_j) \leq x \right\}.$$

Note that $K(\cdot)$ depends only on the dimension k and the underlying covariance matrix Σ. In particular, it does not depend on the μ_j, so it can be computed under the assumption that all $\mu_j = 0$. By construction, we have, for any $\mu \in \mathbb{R}^k$, that

$$P_\mu \{ \mu \in M(1 - \beta) \} = 1 - \beta.$$

The idea is that, with probability at least $1 - \beta$, we may assume that, under the null hypothesis, μ in fact will lie in $\Omega_0 \cap M(1 - \beta)$ rather than just Ω_0. Instead of computing the critical value under $\mu = 0$, the largest value of μ in Ω_0, we may, therefore, compute the critical value under $\tilde{\mu}$, the "largest" value of μ in the (data-dependent) set $\Omega_0 \cap M(1 - \beta)$. It is straightforward to determine $\tilde{\mu}$ explicitly. In particular, $\tilde{\mu}$ has jth component equal to

$$(S.7) \qquad \tilde{\mu}_j = \min\{W_j + K^{-1}(1 - \beta), 0\}.$$

But, to account for the fact that μ may not lie in $M(1 - \beta)$ with probability at most β, we reject H_0 when $T(W_1, \ldots, W_k)$ exceeds the $1 - \alpha + \beta$ quantile of the distribution of T under $\tilde{\mu}$ rather than the $1 - \alpha$ quantile of the distribution of T under $\tilde{\mu}$. Such an adjustment is in the same spirit as the "size correction factor" in Andrews and Barwick (2012a), but requires no computation to determine; see Remark S.5 for further discussion. The following theorem establishes that this test of (S.1) satisfies (S.3).

THEOREM S.2: *Let $T(W_1, \ldots, W_k)$ denote any test statistic that is nondecreasing in each of its arguments. For $\mu \in \mathbb{R}^k$ and $\gamma \in (0, 1)$, define*

$$b(\gamma, \mu) = \inf\{x \in \mathbb{R} : P_\mu\{T(W_1, \ldots, W_k) \leq x\} \geq \gamma\}.$$

Fix $0 \leq \beta \leq \alpha$. The test of (S.1) *that rejects H_0 if $T > b(1 - \alpha + \beta, \tilde{\mu})$ satisfies* (S.3).

REMARK S.2: Although we are unable to establish that the left-hand side of (S.3) equals α, we are able to establish that the left-hand side of (S.3) is at least $\alpha - \beta$. To see this, simply note that $b(1 - \alpha + \beta, \tilde{\mu}) \leq b(1 - \alpha + \beta, 0)$, so

$$\sup_{\mu \in \Omega_0} P_\mu \{T > b(1 - \alpha + \beta, \tilde{\mu})\} \geq P_0 \{T > b(1 - \alpha + \beta, 0)\} = \alpha - \beta.$$

REMARK S.3: As emphasized above, an attractive feature of our procedure is that the "largest" value of μ in $\Omega_0 \cap M(1 - \beta)$ may be determined explicitly. This follows from our particular choice of initial confidence region for μ, namely, from its rectangular shape. If, for example, we had instead chosen $M(1 - \beta)$ to be the usual confidence ellipsoid, then there might not even be a "largest" value of μ in $\Omega_0 \cap M(1 - \beta)$, and one would have to compute

$$\sup_{\mu \in \Omega_0 \cap M(1-\beta)} b(1 - \alpha + \beta, \mu).$$

This problem persists even if the initial confidence region is chosen by inverting tests based on the likelihood ratio statistic (S.5), despite the resulting confidence region being monotone decreasing in the sense that, if x lies in the region, then so does y whenever $y_j \leq x_j$ for all $1 \leq j \leq k$.

REMARK S.4: In some cases, it may be desired to test the null hypothesis that $\mu \in \tilde{\Omega}_0$, where

$$\tilde{\Omega}_0 = \{\mu : \mu_j = 0 \text{ for } j \in J_1, \mu_j \leq 0 \text{ for } j \in J_2\}$$

and J_1 and J_2 form a partition of $\{1, \ldots, k\}$. Such a situation may be accommodated in the framework described above simply by writing $\mu_j = 0$ as $\mu_j \leq 0$ and $-\mu_j \leq 0$, but the resulting procedure may be improved upon by exploiting the additional structure of the null hypothesis. In particular, Theorem S.2 remains valid if T is only required to be nondecreasing in its $|J_2|$ arguments with $j \in J_2$ and $\tilde{\mu}$ is replaced by the vector whose jth component is equal to 0 for $j \in J_1$ and $\min\{W_j + \tilde{K}^{-1}(1 - \beta), 0\}$ for $j \in J_2$, where $\tilde{K}^{-1}(1 - \beta)$ is the $1 - \beta$ quantile of the distribution

$$\tilde{K}(x) = P_\mu \left\{ \max_{j \in J_2} (\mu_j - W_j) \leq x \right\}.$$

REMARK S.5: In the context of the Gaussian model considered in this section, it is instructive for comparison purposes to consider a parametric coun-

terpart to the nonparametric method of Andrews and Barwick (2012a). To describe their approach, fix $\kappa < 0$. Let $\hat{\mu}$ be the k-dimensional vector whose jth component equals zero if $W_j > \kappa$ and $-\infty$ otherwise (or, for practical purposes, some very large negative number). Define the "size correction factor"

$$(S.8) \qquad \hat{\eta} = \inf\Big\{\eta > 0 : \sup_{\mu \in \Omega_0} P_\mu\big\{T > b(1 - \alpha, \hat{\mu}) + \eta\big\} \le \alpha\Big\}.$$

The proposed test of (S.1) then rejects H_0 if $T > b(1 - \alpha, \hat{\mu}) + \hat{\eta}$. The addition of $\hat{\eta}$ 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. Note that the computation of $\hat{\eta}$ as defined in (S.8) is complicated by the fact that there is no explicit solution to the supremum in (S.8). One must, therefore, resort to approximating the supremum in (S.8) in some fashion. Andrews and Barwick (2012a) proposed to approximate $\sup_{\mu \in \Omega_0} P_\mu\{T > b(1 - \alpha, \hat{\mu}) + \eta\}$ with $\sup_{\mu \in \tilde{\Omega}_0} P_\mu\{T > b(1 - \alpha, \hat{\mu}) + \eta\}$, where $\tilde{\Omega}_0 = \{-\infty, 0\}^k$. Andrews and Barwick (2012a) provided an extensive simulation study, but no proof, in favor of this approximation. Even so, the problem remains computationally demanding and, as a result, the authors only considered situations in which $k \le 10$ and $\alpha = 0.05$. In contrast, our two-step procedure is simple to implement even when k is large, as it does not require optimization over Ω_0, and has proven size control for any value of α (thereby allowing, among other things, one to compute a p-value as the smallest value of α for which the null hypothesis is rejected). In the nonparametric setting considered below, where the underlying covariance matrix is also unknown, further approximations are required to implement the method of Andrews and Barwick (2012a). See Remark 2.6 for related discussion.

REMARK S.6: Let $\phi_{\alpha,\beta}$ be the test as described in Theorem S.2. Similarly to the approach of Andrews and Barwick (2012a), one can determine β to maximize (weighted) average power. In the parametric context considered in this section, one can achieve this exactly modulo simulation error. To describe how, let $\mu_1, \ldots, \mu_d$ be alternative values in Ω_1, and let $w_1, \ldots, w_d$ be nonnegative weights that add up to 1. Then, β can be chosen to maximize

$$\sum_{i=1}^d w_i \mathbb{E}_{\mu_i}[\phi_{\alpha,\beta}].$$

This can be accomplished by standard simulation from $N(\mu_i, \Sigma)$ and discretizing β between 0 and α. The drawback here is the specification of the μ_i and w_i. In our simulations, we have found that a reasonable choice is simply $\beta = \alpha/10$.

SUPPLEMENT APPENDIX

PROOF OF THEOREM S.1: For $1 \leq j \leq k$, let e_j be the jth unit basis vector having a 1 in the jth coordinate. To determine $\bar{\mu}$ for the given a, we must minimize

$$f(\mu) = (\mu - a)' \Sigma^{-1} (\mu - a)$$

over $\mu \in \Omega_0$. Note that

$$\frac{\partial f(\mu)}{\partial \mu_j} = 2(\mu - a)' \Sigma^{-1} e_j.$$

First of all, we claim that the minimizing $\bar{\mu}$ cannot have all of its components negative. This follows because, if it did, the line joining the claimed solution and a itself would intersect the boundary of Ω_0 at a point with a smaller value of $f(\mu)$. Therefore, the solution $\bar{\mu}$ must have at least one zero entry.

Suppose that $\bar{\mu}$ is the solution and that $\bar{\mu}_j = 0$ for $j \in J$, where J is some nonempty subset of $\{1, \ldots, k\}$. Let $f_J(\mu) = f(\mu)$ viewed as a function of μ_j with $j \notin J$ and with $\mu_j = 0$ for $j \in J$. Then, the solution to the components $\bar{\mu}_j$ with $j \notin J$ (if there are any) must be obtained by setting partial derivatives equal to zero, leading to the solution of the equations

$$(\mu - a)' \Sigma^{-1} e_j = 0 \quad \forall j \notin J$$

with μ_j fixed at 0 for $j \in J$. Now, the MP test for testing $\bar{u}$ against a rejects for large values of $W' \Sigma^{-1} (a - \bar{u})$, which is a linear combination of $W_1, \ldots, W_k$. The coefficient multiplying W_j is $e_j' \Sigma^{-1} (a - \bar{u})$. But for $j \notin J$, this coefficient is zero by the gradient calculation above.

Next we claim that, for $j = 1, \ldots, k$, the coefficient of W_j is nonnegative. Fix j. Consider $f(\mu)$ as a function of μ_j alone with the other components fixed at the claimed solution for $\bar{\mu}$. If the derivative with respect to μ_j at 0 were positive, that is,

$$(\bar{\mu} - a)' \Sigma^{-1} e_j > 0,$$

then the value of μ_j could decrease and result in a smaller minimizing value for $f(\mu)$. Therefore, it must be the case that

$$(a - \bar{\mu})' \Sigma^{-1} e_j \geq 0;$$

the left-hand side is precisely the coefficient of W_j.

Thus, the solution $\bar{\mu}$ has the property that, for testing $\bar{\mu}$ against a, the MP test rejects for large $\sum_{1 \leq j \leq k} c_j W_j$ such that $\bar{\mu}_j = 0$ implies $c_j \geq 0$ and $\bar{\mu}_j < 0$ implies $c_j = 0$. This property allows us to prove that $\bar{\mu}$ is least favorable. Indeed,

if the critical value c is determined so that the test is level α under $\bar{\mu}$, then for $\mu \in \Omega_0$,

$$P_\mu \left\{ \sum_{j \in J} c_j W_j > c \right\} \le P_0 \left\{ \sum_{j \in J} c_j W_j > c \right\} = P_{\bar{\mu}} \left\{ \sum_{j \in J} c_j W_j > c \right\}.$$

The first inequality follows by monotonicity and the second one by the fact that $\bar{\mu}_j = 0$ for $j \in J$. The least favorable property now follows by Lehmann and Romano (2005, Theorem 3.8.1).

The remainder of the proof is obvious.   *Q.E.D.*

PROOF OF THEOREM S.2: First note that $b(\gamma, \mu)$ is nondecreasing in μ, since T is nondecreasing in its arguments. Fix any μ with $\mu_i \le 0$. Let E be the event that $\mu \in M(1 - \beta)$. Then, the Type I error satisfies

$$P_\mu\{\text{reject } H_0\} \le P_\mu\{E^c\} + P_\mu\{E \cap \{\text{reject } H_0\}\}$$
$$= \beta + P_\mu\{E \cap \{\text{reject } H_0\}\}.$$

But when the event E occurs and H_0 is rejected—so that $T > b(1 - \alpha + \beta, \tilde{\mu})$, then the event $T > b(1 - \alpha + \beta, \mu)$ must occur, since $b(1 - \alpha + \beta, \mu)$ is nondecreasing in μ and $\mu \le \tilde{\mu}$ when E occurs. Hence, the Type I error is bounded above by

$$\beta + P_\mu\{T > b(1 - \alpha + \beta, \mu)\} \le \beta + (1 - (1 - \alpha + \beta)) = \alpha. \quad \textit{Q.E.D.}$$

REFERENCES

ANDREWS, D. (2012): "Similar-on-the-Boundary Tests for Moment Inequalities Exist, But Have Poor Power," Technical Report, Department of Economics, Yale University. [3]

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. [2,4,6]

———— (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. [2]

LEHMANN, E. L. (1952): "Testing Multiparameter Hypotheses," *The Annals of Mathematical Statistics*, 23, 541–552. [2]

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer. [8]

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.