Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure

Donald W. K. Andrews*
Cowles Foundation for Research in Economics
Yale University

Panle Jia
Department of Economics
MIT

June 2008
Revised: September 2008

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-0417911 and SES-0751517. The authors thank Steve Berry for numerous discussions and comments.

Abstract

This paper is concerned with tests and confidence intervals for partially-identified parameters that are defined by moment inequalities and equalities. In the literature, different test statistics, critical value methods, and implementation methods (i.e., asymptotic distribution versus the bootstrap) have been proposed. In this paper, we compare a wide variety of these methods. We provide a recommended test statistic, moment selection critical value method, and implementation method. In addition, we provide a data-dependent procedure for choosing the key moment selection tuning parameter κ and a data-dependent size-correction factor η.

Keywords: Asymptotic size, asymptotic power, confidence set, exact size, generalized moment selection, moment inequalities, partial identification, refined moment selection, test.

JEL Classification Numbers: C12, C15.

1 Introduction

This paper considers inference in moment inequality/equality models with parameters that need not be identified. Many models of this type have been considered recently in the literature. For example, such models arise from the necessary conditions for Nash equilibria, see Ciliberto and Tamer (2003), Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2004), and Bajari, Benkard, and Levin (2008). They also arise from the sufficient conditions for Nash equilibria, see Ciliberto and Tamer (2003) and Beresteanu, Molchanov, and Molinari (2008). In addition, moment inequality/equality models arise from data censoring, such as when a continuous variable is only observed to lie in an interval, see Manski and Tamer (2002), and in some macroeconomic models, see Moon and Schorfheide (2006).

In this paper we consider inference for the true parameter, as in Imbens and Manski (2004), rather than for the identified set. We believe that the former is of greater interest in most circumstances. This paper and many others in the literature construct confidence sets (CS's) by inverting Anderson-Rubin-type test statistics, following Chernozhukov, Hong, and Tamer (2007) (CHT). Several different test statistics have been proposed in the literature. Subsampling critical values have been employed by CHT, Andrews and Guggenberger (AG), and Romano and Shaikh (2005, 2008). Andrews and Soares (2007) (AS) use generalized moment selection (GMS) critical values. The critical value methods employed by Bugni (2007a,b), Canay (2007), and Fan and Park (2007) fall within the GMS class of critical values.

GMS and subsampling-based tests and CS's are the only methods in the literature that apply to arbitrary moment functions and have been shown to have correct asymptotic size in a uniform sense, see AS, Andrews and Guggenberger (2008) (AG), and Romano and Shaikh (2008). AS shows that GMS tests dominate subsampling tests in terms of asymptotic power. Bugni (2007a,b) shows that a particular GMS test has smaller errors in null rejection probabilities asymptotically than a corresponding (recentered) subsampling test. These power and size results imply that GMS critical values are preferred to subsampling critical values.

GMS tests and CS's depend on the specification of a test statistic function S, a critical value function φ, and a tuning parameter κ. Given the advantageous properties of GMS tests and CS's, it is desirable to compare different test statistic functions S and different critical value functions φ in terms of size and power and to find the combination that performs best and can be recommended for general use. In addition, it is very useful to determine (i) a data-dependent tuning parameter κ for the GMS critical value (because κ is a key parameter and the asymptotically optimal choice of κ depends on unknowns) and (ii) a data-dependent size-correction factor η (because asymptotic size-correction is necessary when one chooses the tuning parameter κ to maximize average asymptotic power). We call a GMS procedure that satisfies conditions (i) and (ii) a *refined moment selection* (RMS) procedure.

The present paper accomplishes the goal of determining a recommended RMS procedure. We find that the Gaussian quasi-likelihood ratio (QLR) test statistic combined with the "t-test moment selection" critical value performs very well in terms of average asymptotic power. We show that with i.i.d. observations the bootstrap implementation of this test out-performs the asymptotic-distribution implementation based on finite-sample size and power. We develop data-dependent methods of selecting κ and η and show that they yield very good asymptotic and finite-sample size and power. We provide a table that makes them easy to implement in practice. The results of the paper apply to i.i.d. and time series observations and to moment functions that are based on preliminary estimators of point-identified parameters.

To achieve the goals listed above, we consider asymptotics in which κ equals a *finite* constant plus $o_p(1)$, rather than asymptotics in which $\kappa \to \infty$ as $n \to \infty$. This differs from the asymptotics considered in other papers in this literature.

There are four reasons for using finite-κ asymptotics. First, they provide better approximations because κ is finite, not infinite, in any given application. Second, for any given (S, φ), they allow one to compute a best κ value in terms of average asymptotic power, which in turn allows one to compare different (S, φ) functions (each evaluated at its own best κ value) in terms of average asymptotic power. One cannot determine a best κ value in terms of average asymptotic power when $\kappa \to \infty$ because asymptotic power is always higher if κ is smaller, asymptotic size does not depend on κ, and finite-sample size is worse if κ smaller. Third, for the recommended (S, φ) functions, the finite-κ asymptotic formula for the best κ value lets one determine a data-dependent κ value that is approximately optimal in terms of average asymptotic power. Fourth, finite-κ asymptotics permit one to compute size-correction factors that depend on κ, which is a primary determinant of a test's finite-sample size. In contrast, if $\kappa \to \infty$ the asymptotic properties of tests under the null hypothesis do not depend on κ. Even the higher-order errors in null rejection probabilities do not depend on κ, see Bugni (2007a,b). Thus, with $\kappa \to \infty$ asymptotics, size-correction based on κ is not possible.

Using finite-κ asymptotics, we compare different choices of (S, φ) when each is evaluated at the infeasible asymptotically-optimal choice of κ according to an average power criterion. In particular, we consider (i) the modified method of moments (MMM) statistic S_1, which has been used in Pakes, Porter, Ishii, and Ho (2004), Romano and Shaikh (2005, 2008), AS, Bugni (2007a,b), CHT, Fan and Park (2007), and AG; (ii) the QLR statistic S_2, which has been considered in AG, AS, and Rosen (2008); and (iii) the Max and SumMax statistics S_3, which have been considered in AG and AS and by Azeem Shaikh.[1] We consider the $\varphi^{(1)}$ critical value function, which yields "t-test moment selection" critical values and has been considered in Soares (2005), AS, CHT, and Bugni (2007a,b); the $\varphi^{(3)}$ critical value function, which has been considered in AS and Canay (2007); the $\varphi^{(4)}$ critical value function, which has been considered in Fan and Park (2007); and the $\varphi^{(5)}$ critical value function, which yields modified moment selection criterion (MMSC) critical values and has been considered in Soares (2005) and AS.

The recommended (S, φ) functions are the QLR statistic and the t-test critical value functions $(S_2, \varphi^{(1)})$. This combination is found to have very good average asymptotic power. The $(S_2, \varphi^{(5)})$ and $(S_2, \varphi^{(4)})$ functions also have very good average asymptotic power, but they have computational drawbacks, especially the $(S_2, \varphi^{(5)})$ functions when the number of moment inequalities, p, is large. The $\varphi^{(1)}$ critical value function, on the other hand, is very attractive from a computational perspective.

The comparisons of the (S, φ) functions described above are based on infeasible values of κ. For our recommended choice $(S_2, \varphi^{(1)})$, we develop a feasible data-dependent method for choosing κ, denoted $\widehat{\kappa}$. The data-dependent method is based on an approximation to the function that maps the correlation matrix of the moment functions into an optimal value of κ. We show numerically that this approximation works extremely well in terms of average asymptotic power.

Finally, we compute a data-dependent size-correction factor $\widehat{\eta}$ for the recommended test based on $(S_2, \varphi^{(1)})$ and $\widehat{\kappa}$, and provide a table for easy determination of $\widehat{\kappa}$ and $\widehat{\eta}$.

The RMS test based on $(S_2, \varphi^{(1)})$, $\widehat{\kappa}$, and $\widehat{\eta}$ is our recommended RMS procedure. It can be implemented in finite samples using an "asymptotic normal" version of the moment selection critical value or a bootstrap version. Neither has superior asymptotic properties (because the tests are not asymptotically pivotal). Finite-sample simulations with i.i.d. observations show that the bootstrap performs better in terms of size and

[1] Personal communication.

power, especially for moment functions with skewed distributions. Furthermore, the power of the bootstrap-based test is very close to its asymptotic power across the range of cases considered. Hence, we recommend the bootstrap implementation.

We note that the finite-sample simulations carried out here are unusually general in their applicability. The finite-sample properties of GMS and RMS tests are shown to depend on the moment functions, $m(\cdot, \theta)$, and the observations, W_i, only through the distribution of $m(W_i, \theta_0)$, where θ_0 is the null parameter value. Hence, by considering a range of such distributions, one can cover any moment inequality model—the particular form of the moment functions does not need to be specified. From the asymptotic results, we know that the primary effect of the distribution is through its correlation matrix. Not surprisingly, secondary effects are found to be due to skewness and kurtosis of the distributions. Skewness effects are found to be more substantial than kurtosis effects for the "asymptotic normal" version of the test. The bootstrap version of the test has relatively little sensitivity to skewness or kurtosis.

The paper also compares the recommended RMS procedure to tests based on "plug-in asymptotic" (PA) critical values and to "pure" generalized empirical likelihood (GEL) tests in terms of average asymptotic power. PA critical values have been used widely in the statistical literature on multivariate one-sided tests, e.g., see Silvapulle and Sen (2005). PA critical values use a quantile from the least favorable null distribution given a consistent estimator of the correlation matrix of the moment functions. Pure GEL tests rely on a constant critical value that is least favorable with respect to both the null mean vectors and the correlation matrix of the moment functions. Pure GEL tests are shown in Otsu (2006) and Canay (2007) to have some optimal large-deviation asymptotic power properties. However, in our view, the large-deviation asymptotic optimality criterion is not appropriate when comparing tests with substantially different asymptotic properties under non-large deviations.

Our results show that the recommended RMS test dominates PA and pure GEL tests in terms of average asymptotic power and the power advantages are quite substantial in most cases, especially when the number of moment inequalities, p, is large. For example, when $p = 10$, the recommended RMS test is between three and six times more powerful than a pure GEL test (for alternatives where the asymptotic power envelope is .85).[2]

[2] GEL test statistics can be combined with the recommended RMS critical value. Such tests have the same asymptotic properties as the recommended RMS test. However, GEL test statistics are much more time consuming to compute than the QLR statistic. This is a distinct disadvantage because computation of the RMS critical value requires thousands of test statistic evaluations and construction

Related literature concerning inference for partially-identified parameters, not referenced above, includes Woutersen (2006), Bontemps, Magnac, and Maurin (2007), Moon and Schorfheide (2007), Stoye (2007), Beresteanu and Molinari (2008), Galichon and Henry (2008), Guggenberger, Hahn, and Kim (2008), and Andrews and Han (2009).

The remainder of this paper is organized as follows. Section 2 introduces the model and describes the preferred RMS confidence set and test. Section 3 defines the test statistics that are considered. Section 4 introduces RMS critical values. Section 5 provides asymptotic size and power results, discusses average asymptotic power, and introduces the asymptotic power envelope. Section 6 provides (i) numerical results comparing the average asymptotic power of tests based on different (S, φ) functions, (ii) a description of, and motivation for, how the recommended data-dependent tuning parameter $\widehat{\kappa}$ and size-correction factor $\widehat{\eta}$ are determined, and (iii) numerical results assessing the size and power properties of the recommended RMS test. Section 7 gives the finite-sample results. Appendix A given in the Supplement to this paper, i.e., Andrews and Jia (2008), provides proofs of the asymptotic results of the paper. Appendix B in the Supplement provides supplemental numerical results to those reported in Section 6. Appendix C in the Supplement contains details concerning the numerical results reported in Section 6.

We use the following notation. Let $R_+ = \{x \in R : x \geq 0\}$, $R_{++} = \{x \in R : x > 0\}$, $R_{+,\infty} = R_+ \cup \{+\infty\}$, $R_{[+\infty]} = R \cup \{+\infty\}$, $R_{[\pm\infty]} = R \cup \{\pm\infty\}$, $K^p = K \times ... \times K$ (with p copies) for any set K, $\infty^p = (+\infty, ..., +\infty)'$ (with p copies). All limits are as $n \to \infty$ unless specified otherwise. Let "pd" abbreviate "positive definite," $cl(\Psi)$ denote the closure of a set Ψ, and 0_v denote a v-vector of zeros.

2 Model and Recommended Confidence Set

2.1 Moment Inequality Model

The moment inequality/equality model is as follows. The true value θ_0 $(\in \Theta \subset R^d)$ is assumed to satisfy the moment conditions:

$$\begin{aligned} &E_{F_0} m_j(W_i, \theta_0) \geq 0 \text{ for } j = 1, ..., p \text{ and} \\ &E_{F_0} m_j(W_i, \theta_0) = 0 \text{ for } j = p+1, ..., p+v, \end{aligned} \tag{2.1}$$

of CS's requires many critical value calculations.

where $\{m_j(\cdot,\theta) : j = 1, ..., k\}$ are known real-valued moment functions, $k = p + v$, and $\{W_i : i \geq 1\}$ are i.i.d. or stationary random vectors with joint distribution F_0. Either p or v may be zero. The observed sample is $\{W_i : i \leq n\}$. The true value θ_0 is not necessarily identified.

We are interested in tests and confidence sets (CS's) for the true value θ_0.

Generic values of the parameters are denoted (θ, F). For the case of i.i.d. observations, the parameter space $\mathcal{F}$ for (θ, F) is the set of all (θ, F) that satisfy:

$$
\begin{aligned}
&\text{(i) } \theta \in \Theta, \text{ (ii) } E_F m_j(W_i,\theta) \geq 0 \text{ for } j = 1, ..., p, \text{ (iii) } E_F m_j(W_i,\theta) = 0 \\
&\text{for } j = p+1, ..., k, \text{ (iv) } \{W_i : i \geq 1\} \text{ are i.i.d. under } F, \\
&\text{(v) } \sigma^2_{F,j}(\theta) = Var_F(m_j(W_i,\theta)) > 0, \text{ (vi) } Corr_F(m(W_i,\theta)) \in \Psi, \text{ and} \\
&\text{(vii) } E_F|m_j(W_i,\theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M \text{ for } j = 1, ..., k, \qquad (2.2)
\end{aligned}
$$

where $Var_F(\cdot)$ and $Corr_F(\cdot)$ denote variance and correlation matrices, respectively, when F is the true distribution, Ψ is the parameter space for $k \times k$ correlation matrices specified at the end of Section 3, and $M < \infty$ and $\delta > 0$ are constants.

The asymptotic results apply to the case of dependent observations. For brevity, we specify $\mathcal{F}$ for dependent observations in Appendix A in the Supplement. The asymptotic results also apply when the moment functions in (2.1) depend on a parameter τ, i.e., when they are of the form $\{m_j(W_i,\theta,\tau) : j \leq k\}$, and a preliminary consistent and asymptotically normal estimator $\widehat{\tau}_n(\theta_0)$ of τ exists (where θ_0 is the true value of θ). The existence of such an estimator requires that τ is identified given θ_0. In this case, the sample moment functions take the form $\overline{m}_{n,j}(\theta) = \overline{m}_{n,j}(\theta, \widehat{\tau}_n(\theta))$ $(= n^{-1}\sum_{i=1}^n m_j(W_i,\theta,\widehat{\tau}_n(\theta)))$. The asymptotic variance of $n^{1/2}\overline{m}_{n,j}(\theta)$ typically is affected by the estimation of τ and is defined accordingly. Nevertheless, all of the asymptotic results given below hold in this case using the definition of $\mathcal{F}$ given in Appendix A in the Supplement with the definitions of $m_j(W_i,\theta)$ and $\overline{m}_{n,j}(\theta)$ changed suitably, as described there.

2.2 Recommended Confidence Set

We consider a confidence set obtained by inverting a test. The test is based on a test statistic $T_n(\theta_0)$ for testing $H_0 : \theta = \theta_0$. The nominal level $1 - \alpha$ CS for θ is

$$
CS_n = \{\theta \in \Theta : T_n(\theta) \leq c_n(\theta)\}, \qquad (2.3)
$$

where $c_n(\theta)$ is a data-dependent critical value.[3] In other words, the confidence set includes all parameter values θ for which one does not reject the null hypothesis that θ is the true value.

We now describe the recommended test statistic and critical value. The justifications for these recommendations are given in the sections of the paper that follow. The recommended test statistic is a quasi-likelihood ratio (QLR) statistic, $T_{QLR,n}(\theta)$, that is a function of the sample moment conditions, $n^{1/2}\overline{m}_n(\theta)$, and an estimator, $\widehat{\Sigma}_n(\theta)$, of their asymptotic variance:

$$\begin{aligned} T_{QLR,n}(\theta) &= S_2(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)) \\ &= \inf_{t=(t_1,0_v):t_1\in R^p_{+,\infty}} (n^{1/2}\overline{m}_n(\theta) - t)'\widehat{\Sigma}_n^{-1}(\theta)(n^{1/2}\overline{m}_n(\theta) - t), \text{ where} \\ \overline{m}_n(\theta) &= (\overline{m}_{n,1}(\theta), ..., \overline{m}_{n,k}(\theta))' \text{ and} \\ \overline{m}_{n,j}(\theta) &= n^{-1}\sum_{i=1}^n m_j(W_i, \theta) \text{ for } j = 1, ..., k.^4 \end{aligned} \tag{2.4}$$

When the observations are i.i.d. and no parameter τ appears, we take

$$\begin{aligned} \widehat{\Sigma}_n(\theta) &= n^{-1}\sum_{i=1}^n (m(W_i, \theta) - \overline{m}_n(\theta))(m(W_i, \theta) - \overline{m}_n(\theta))', \text{ where} \\ m(W_i, \theta) &= (m_1(W_i, \theta), ..., m_k(W_i, \theta))'. \end{aligned} \tag{2.5}$$

With temporally dependent observations or when a preliminary estimator of a parameter τ appears, a different definition of $\widehat{\Sigma}_n(\theta)$ often is required. For example, with dependent observations, a heteroskedasticity and autocorrelation consistent (HAC) estimator may be required.

The correlation matrix $\widehat{\Omega}_n(\theta)$ that corresponds to $\widehat{\Sigma}_n(\theta)$ is defined by

$$\widehat{\Omega}_n(\theta) = \widehat{D}_n^{-1/2}(\theta)\widehat{\Sigma}_n(\theta)\widehat{D}_n^{-1/2}(\theta), \text{ where } \widehat{D}_n(\theta) = Diag(\widehat{\Sigma}_n(\theta)), \tag{2.6}$$

where $Diag(\Sigma)$ denotes the diagonal matrix based on the matrix Σ.

The test statistic $T_{QLR,n}(\theta)$ is computed using a quadratic programming algorithm. Such algorithms are built into GAUSS and Matlab. They are very fast even when p is large, although they are not as fast as computing a statistic that has a simple closed-

[3] When θ is in the interior of the identified set, it may be the case that $T_n(\theta) = 0$ and $c_n(\theta) = 0$. In consequence, it is important that the inequality in the definition of CS_n is $\leq$, not $<$.

form expression. For example, to compute the QLR test statistic 100,000 times takes 2.6, 2.9, 4.7, 10.7, 22.5, and 69.8 seconds when $p = 2$, 4, 10, 20, 30, and 50, respectively, using GAUSS on a PC with a 3.4 GHz processor.

The origin of the QLR statistic is as follows. Suppose one replaces $n^{1/2}\overline{m}_n(\theta)$ and $\widehat{\Sigma}_n(\theta)$ in (2.4) by a data vector $X \in R^k$ and a known $k \times k$ variance matrix Σ, respectively. Then, the QLR statistic is the likelihood ratio statistic for the model with $X \sim N(\mu, \Sigma)$, $\mu = (\mu_1', \mu_2')' \in R^p \times R^v = R^k$, the null hypothesis $H_0^* : \mu_1 \geq 0_p$ & $\mu_2 = 0_v$ and the alternative hypothesis $H_1^* : \mu_1 \ngeq 0_p$ &/or $\mu_2 \neq 0_v$. The QLR statistic has been considered in many papers on tests of inequality constraints, e.g., see Kudo (1963) and Silvapulle and Sen (2005, Sec. 3.8). In the moment inequality literature, it has been considered by AG, AS, and Rosen (2008).

The recommended RMS critical value is

$$c_n(\theta) = c_n(\theta, \widehat{\kappa}) + \widehat{\eta}, \tag{2.7}$$

where $c_n(\theta, \widehat{\kappa})$ is the $1 - \alpha$ quantile of a bootstrap (or "asymptotic normal") distribution of a moment selection form of $T_{QLR,n}(\theta)$ and $\widehat{\eta}$ is a data-dependent size-correction factor. For i.i.d. data, we recommend using a nonparametric bootstrap version of $c_n(\theta, \widehat{\kappa})$. For dependent data, either a block bootstrap or an asymptotic normal version can be applied. (To date, we have not determined which is preferable.)

We now describe the bootstrap version of $c_n(\theta, \widehat{\kappa})$. Let $\{W_{i,r}^* : i \leq n\}$ for $r = 1, ..., R$ denote R bootstrap samples of size n (i.i.d. across samples), such as nonparametric i.i.d. bootstrap samples in an i.i.d. scenario or block bootstrap samples in a time series scenario, where R is large. The k-vectors of re-centered and re-scaled bootstrap sample moments and bootstrap $k \times k$ correlation matrices for $r = 1, ..., R$ are defined by

$$\begin{aligned}
&M_{n,r}^*(\theta) = \left(\widehat{D}_n^*(\theta)\right)^{-1/2} n^{1/2}\left(\overline{m}_{n,r}^*(\theta) - \overline{m}_n(\theta)\right) \text{ and} \\
&\widehat{\Omega}_{n,r}^*(\theta) = \widehat{D}_{n,r}^*(\theta)^{-1/2}\widehat{\Sigma}_{n,r}^*(\theta)\widehat{D}_{n,r}^*(\theta)^{-1/2} \text{ for } r = 1, ..., R, \text{ where} \\
&\overline{m}_{n,r}^*(\theta) = n^{-1}\sum_{i=1}^{n} m(W_{i,r}^*, \theta),\ \widehat{D}_{n,r}^*(\theta) = Diag(\widehat{\Sigma}_{n,r}^*(\theta)),
\end{aligned} \tag{2.8}$$

and $\widehat{\Sigma}_{n,r}^*(\theta)$ is defined as $\widehat{\Sigma}_n(\theta)$ is defined (e.g., as in (2.5) in the i.i.d. case) with $\{W_{i,r}^* : i \leq n\}$ in place of $\{W_i : i \leq n\}$ throughout.[5]

[5] Note that when a preliminary consistent estimator of a parameter τ appears, the bootstrap moment

The idea behind the RMS critical value is to compute the critical value using only those moment inequalities that have a noticeable effect on the asymptotic null distribution of the test statistic. Note that moment inequalities that have large positive population means have little or no effect on the asymptotic null distribution. The preferred RMS procedure employs element-by-element t-tests of the null hypothesis that the mean of $\overline{m}_{n,j}(\theta)$ is zero versus the alternative that it is positive for $j = 1, ..., p$. The j-th moment inequality is selected if

$$\frac{n^{1/2}\overline{m}_{n,j}(\theta)}{\widehat{\sigma}_{n,j}(\theta)} \leq \widehat{\kappa}, \tag{2.9}$$

where $\widehat{\sigma}^2_{n,j}(\theta)$ is the (j,j) element of $\widehat{\Sigma}_n(\theta)$ for $j = 1, ..., p$ and $\widehat{\kappa}$ is a data-dependent tuning parameter (defined in (2.12) below) that plays the role of a critical value in selecting the moment inequalities. Let $\widehat{p}$ denote the number of selected moment inequalities.

For $r = 1, ..., R$, let $M^*_{n,r}(\theta, \widehat{\kappa})$ denote the $(\widehat{p} + v)$-sub-vector of $M^*_{n,r}(\theta)$ that includes the $\widehat{p}$ selected moment inequalities plus the v moment equalities. Analogously, let $\Omega^*_{n,r}(\theta, \widehat{\kappa})$ denote the $(\widehat{p} + v) \times (\widehat{p} + v)$-sub-matrix of $\Omega^*_{n,r}(\theta)$ that consists of the $\widehat{p}$ selected moment inequalities and the v moment equalities. The bootstrap critical value $c_n(\theta, \widehat{\kappa})$ is the $1 - \alpha$ sample quantile of

$$\{S_2(M^*_{n,r}(\theta, \widehat{\kappa}), \Omega^*_{n,r}(\theta, \widehat{\kappa})) : r = 1, ..., R\}, \tag{2.10}$$

where $S_2(\cdot, \cdot)$ is defined as in (2.4) but with p replaced by $\widehat{p}$.

An "asymptotic normal" version of the critical value is obtained by replacing the bootstrap quantities $M^*_{n,r}(\theta, \widehat{\kappa})$ and $\Omega^*_{n,r}(\theta, \widehat{\kappa})$ in (2.10) by $\widehat{\Omega}^{1/2}_n(\theta, \widehat{\kappa})Z^*_r$ and $\widehat{\Omega}_n(\theta, \widehat{\kappa})$, respectively, where $Z^*_r \sim i.i.d.\ N(0_{\widehat{p}+v}, I_{\widehat{p}+v})$ for $r = 1, ..., R$ (and $\{Z^*_r : r = 1, ..., R\}$ are independent of $\{W_i : i \leq n\}$ conditional on $\widehat{p}$).

The tuning parameter $\widehat{\kappa}$ in (2.9) and the size-correction factor $\widehat{\eta}$ in (2.7) depend on the estimator $\widehat{\Omega}_n(\theta)$ of the asymptotic correlation matrix $\Omega(\theta)$ of $n^{1/2}\overline{m}_n(\theta)$. In particular, they depend on $\widehat{\Omega}_n(\theta)$ through a $[-1, 1]$-valued function $\delta(\widehat{\Omega}_n(\theta))$ that is a measure of the amount of dependence in the correlation matrix $\widehat{\Omega}_n(\theta)$. We define

$$\delta(\Omega) = \text{smallest off-diagonal element in the upper } p \times p \text{ block of } \Omega, \tag{2.11}$$

conditions need to be based on a bootstrap estimator of this preliminary estimator. In such cases, the asymptotic normal version of the critical value may be much quicker to compute.

where Ω is a $k \times k$ correlation matrix. As defined, $\delta(\Omega)$ is a particular measure of the amount of negative correlation in Ω. Motivation for this choice of function $\delta(\Omega)$ is given in Section 6.3.1 below.

The moment selection tuning parameter $\widehat{\kappa}$ and the size-correction factor $\widehat{\eta}$ are defined by

$$\begin{aligned}&\widehat{\kappa} = \kappa(\widehat{\delta}_n(\theta)) \text{ and } \widehat{\eta} = \eta_1(\widehat{\delta}_n(\theta)) + \eta_2(p), \text{ where}\\&\widehat{\delta}_n(\theta) = \delta(\widehat{\Omega}_n(\theta)).\end{aligned} \tag{2.12}$$

Table I provides values of $\kappa(\delta)$, $\eta_1(\delta)$, and $\eta_2(p)$ for $\delta \in [-1, 1]$ and $p \in \{2, 3, ..., 50\}$ for use with tests with level $\alpha = .05$ and CS's with level $1 - \alpha = .95$. Table B-VIII in Appendix B in the Supplement provides simulated values of the mean and standard deviation of the asymptotic distribution of $c_n(\theta, \widehat{\kappa})$. These results, combined with the values of $\eta_1(\delta)$ and $\eta_2(p)$ in Table I, show that the size-correction factor $\widehat{\eta} = \eta_1(\widehat{\delta}_n(\theta)) + \eta_2(p)$ typically is small compared to $c_n(\theta, \widehat{\kappa})$, but not negligible.

In sum, the preferred RMS critical value, $c_n(\theta)$, and CS are computed using the following steps. One computes (i) $\widehat{\Omega}_n(\theta)$ defined in (2.6), (ii) $\widehat{\delta}_n(\theta)$ = smallest off-diagonal element in the upper $p \times p$ block of $\widehat{\Omega}_n(\theta)$, (iii) $\widehat{\kappa} = \kappa(\widehat{\delta}_n(\theta))$ using Table I, (iv) $\widehat{\eta} = \eta_1(\widehat{\delta}_n(\theta)) + \eta_2(p)$ using Table I, (v) the vector of selected moments using (2.9) with $\widehat{\kappa} = \kappa(\widehat{\delta}_n(\theta))$, (vi) the selected bootstrap sample moments and correlation matrices $\{(M^*_{n,r}(\theta, \widehat{\kappa}), \Omega^*_{n,r}(\theta, \widehat{\kappa})) : r = 1, ..., R\}$, defined in (2.8) with the non-selected moment inequalities omitted, (vii) $c_n(\theta, \widehat{\kappa})$, which is the .95 sample quantile of $\{S_2(M^*_{n,r}(\theta, \widehat{\kappa}), \Omega^*_{n,r}(\theta, \widehat{\kappa})) : r = 1, ..., R\}$ with $\widehat{\kappa} = \kappa(\widehat{\delta}_n(\theta))$ (for a test of level .05 and a CS of level .95) and (viii) $c_n(\theta) = c_n(\theta, \widehat{\kappa}) + \widehat{\eta}$. The preferred RMS confidence set is computed by determining all the values θ for which the null hypothesis that θ is the true value is not rejected. For the asymptotic normal version of the recommended RMS critical value, in step (vi) one computes the selected sub-vector and sub-matrix of $\widehat{\Omega}_n^{1/2}(\theta, \widehat{\kappa}) Z^*_r$ and $\widehat{\Omega}_n(\theta, \widehat{\kappa})$, defined in the paragraph following (2.10), and in step (vii) one computes the .95 sample quantile with these quantities in place of $M^*_{n,r}(\theta, \widehat{\kappa})$ and $\Omega^*_{n,r}(\theta, \widehat{\kappa})$, respectively.

To compute the recommended bootstrap RMS test using 10,000 simulation repetitions takes 1.3, 1.7, 3.2, 8.4, 17.2, and 52.0 seconds when $p = 2$, 4, 10, 20, 30, and 50, respectively, and $n = 250$ using GAUSS on a PC with a 3.4 GHz processor. For the "asymptotic normal" version, the times are .25, .31, .71, 2.4, 6.1, and 21.8 seconds, respectively.[6]

Table I. Moment Selection Tuning Parameters $\kappa(\delta)$ and Size-Correction Factors $\eta_1(\delta)$ and $\eta_2(p)$ for $\alpha = .05$

δ	$\kappa(\delta)$	$\eta_1(\delta)$	δ	$\kappa(\delta)$	$\eta_1(\delta)$	δ	$\kappa(\delta)$	$\eta_1(\delta)$
$[-1, -.975)$	2.9	.000	$[-.30, -.25)$	1.9	.113	$[.45, .50)$	0.8	.072
$[-.975, -.95)$	2.9	.001	$[-.25, -.20)$	1.9	.151	$[.50, .55)$	0.8	.043
$[-.95, -.90)$	2.9	.002	$[-.20, -.15)$	1.9	.144	$[.55, .60)$	0.6	.067
$[-.90, -.85)$	2.9	.013	$[-.15, -.10)$	1.9	.122	$[.60, .65)$	0.6	.041
$[-.85, -.80)$	2.8	.043	$[-.10, -.05)$	1.8	.112	$[.65, .70)$	0.4	.021
$[-.80, -.75)$	2.7	.076	$[-.05, .00)$	1.7	.094	$[.70, .75)$	0.4	.023
$[-.75, -.70)$	2.7	.077	$[.00, .05)$	1.5	.131	$[.75, .80)$	0.001	.030
$[-.70, -.65)$	2.7	.075	$[.05, .10)$	1.5	.103	$[.80, .85)$	0.001	.011
$[-.65, -.60)$	2.6	.086	$[.10, .15)$	1.4	.108	$[.85, .90)$	0.001	.002
$[-.60, -.55)$	2.4	.139	$[.15, .20)$	1.3	.093	$[.90, .95)$	0.001	.000
$[-.55, -.50)$	2.4	.113	$[.20, .25)$	1.3	.102	$[.95, .975)$	0.001	.000
$[-.50, -.45)$	2.4	.106	$[.25, .30)$	1.2	.099	$[.975, .99)$	0.001	.000
$[-.45, -.40)$	2.4	.094	$[.30, .35)$	1.1	.089	$[.99, 1.0]$	0.001	.000
$[-.40, -.35)$	2.2	.131	$[.35, .40)$	0.8	.113			
$[-.35, -.30)$	2.1	.131	$[.40, .45)$	0.8	.091			

p	2	3	4	5	6	7	8	9	10
$\eta_2(p)$	.00	.05	.09	.14	.18	.23	.27	.31	.35

$p \in [11, 50]$: $\eta_2(p) = .04743\ (p-2) - .00040\ (p-2)^2$

[6] When constructing a CS, if the computation time is burdensome (because one needs to carry out many tests with different values of θ as the null value), then a useful approach is to map out the general features of the CS using the asymptotic normal version of the MMM/t-Test/κ=2.35 test, defined below, which is very fast to compute, see Section B of the Supplement, and then switch to the bootstrap version of the recommended RMS test to find the boundaries of the CS more precisely.

3 Test Statistics

We now start the justification for the recommended RMS test. In this section, we define the test statistics $T_n(\theta)$ that we consider. The statistic $T_n(\theta)$ is of the form

$$T_n(\theta) = S(n^{1/2}\overline{m}_n(\theta), \widehat{\Sigma}_n(\theta)), \tag{3.1}$$

where S is a real function on $(R^p_{[+\infty]} \times R^v) \times \mathcal{V}_{k\times k}$ and $\mathcal{V}_{k\times k}$ is the space of $k \times k$ variance matrices. (The set $R^p_{[+\infty]} \times R^v$ contains k-vectors whose first p elements are either real or $+\infty$ and whose last v elements are real.)

We now give the leading examples of the test statistic function S. The first is the modified method of moments (MMM) test function S_1 defined by

$$S_1(m, \Sigma) = \sum_{j=1}^{p} [m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v} (m_j/\sigma_j)^2, \text{ where}$$
$$[x]_- = \begin{cases} x & \text{if } x < 0 \\ 0 & \text{if } x \geq 0, \end{cases} \quad m = (m_1, ..., m_k)', \tag{3.2}$$

and σ_j^2 is the jth diagonal element of Σ. The Introduction lists papers in the literature that consider this test statistic and the other test statistics below.[7]

The second function S is the QLR test function S_2 that is defined in (2.4).

Note that under the null and local alternative hypotheses, GEL test statistics behave asymptotically (to the first order) the same as the statistic $T_n(\theta)$ based on S_2 (see Sections 8.1 and 10.3 of AG and Section 10.1 of AS). Although GEL statistics are not of the form given in (3.1), the results of the present paper, viz., Theorems 1 and 2 below, hold for such statistics under the assumptions given in AG.

The third function is a test function, S_3, that directs power against alternatives with p_1 $(< p)$ moment inequalities violated. The test function S_3 is defined by

$$S_3(m, \Sigma) = \sum_{j=1}^{p_1} [m_{(j)}/\sigma_{(j)}]_-^2 + \sum_{j=p+1}^{p+v} (m_j/\sigma_j)^2, \tag{3.3}$$

where $[m_{(j)}/\sigma_{(j)}]_-^2$ denotes the jth largest value among $\{[m_\ell/\sigma_\ell]_-^2 : \ell = 1, ..., p\}$ and

[7] Several papers in the literature use a variant of S_1 that is not invariant to rescaling of the moment functions (i.e., with $\sigma_j = 1$ for all j), which is not desirable in terms of the power of the resulting test.

$p_1 < p$ is some specified integer.[8,9]

The asymptotic results given in Section 5 below hold for all functions S that satisfy the following assumption.

Assumption S. (a) $S(m, \Sigma) = S(Dm, D\Sigma D)$ for all $m \in R^k$, $\Sigma \in R^{k \times k}$, and pd diagonal $D \in R^{k \times k}$.

(b) $S(m, \Omega) \geq 0$ for all $m \in R^k$ and $\Omega \in \Psi$.

(c) $S(m, \Omega)$ is continuous at all $m \in R^p_{[+\infty]} \times R^v$ and $\Omega \in \Psi$.[10]

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

(e) For all $\ell \in R^p_{[+\infty]} \times R^v$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the df of $S(Z + \ell, \Omega)$ at x is (i) continuous for $x > 0$ and (ii) unless $v = 0$ and $\ell = \infty^p$, strictly increasing for $x > 0$.

In Assumption S, the set Ψ is as in condition (vi) of (2.2) when the observations are i.i.d. and no preliminary estimator of a parameter τ appears. Otherwise, Ψ is the parameter space for the correlation matrix of the asymptotic distribution of $n^{1/2}\overline{m}_n(\theta)$ under (θ, F), denoted $AsyCorr_F(n^{1/2}\overline{m}_n(\theta))$.[11]

The functions S_1, S_2, and S_3 satisfy Assumption S.[12]

4 Refined Moment Selection

This section is concerned with critical values for use with the test statistics introduced in Section 3. We proceed in four steps. First, we explain the idea behind moment selection critical values and discuss a tuning parameter $\widehat{\kappa}$ that determines the extent of the moment selection. Second, we introduce a function φ that helps one to select "relevant" moment inequalities. Third, we define the RMS critical value. Lastly, we

[8] When constructing a CS, a natural choice for p_1 is the dimension d of θ, see Section 5.3 below.

[9] With the functions S_1 and S_3, the parameter space Ψ for the correlation matrices in Assumption S and in condition (vi) of (2.2) can be any non-empty subset of the set Ψ_1 of all $k \times k$ correlation matrices. With the function S_2, the asymptotic results below require that the correlation matrices in Ψ have determinants bounded away from zero because Σ^{-1} appears in the definition of S_2. It may be possible to extend the results to allow Ψ to equal Ψ_1 by replacing Σ^{-1} by the Moore-Penrose inverse Σ^+ in the definition of S_2.

[10] Let $B \subset R^w$. We say that a real function G on $R^p_{[+\infty]} \times B$ is continuous at $x \in R^p_{[+\infty]} \times B$ if $y \to x$ for $y \in R^p_{[+\infty]} \times B$ implies that $G(y) \to G(x)$. In Assumption S(c), $S(m, \Omega)$ is viewed as a function with domain Ψ_1.

[11] More specifically, for dependent observations or when a preliminary estimator of a parameter τ appears, Ψ is as in condition (v) of (1.3) in Section A of the Supplement.

[12] See Lemma 1 of AG for a proof for Assumptions S(a)-S(d) and AS for a proof for Assumption S(e).

specify a size-correction factor $\widehat{\eta}$ that delivers correct asymptotic size even when $\widehat{\kappa}$ does not diverge to infinity. Because the CS's defined in (2.3) are obtained by inverting tests, we discuss both tests and CS's below.

4.1 Basic Idea and Tuning Parameter $\widehat{\kappa}$

The idea behind *generalized moment selection* and *refined moment selection* is to use the data to determine whether a given moment inequality is satisfied and is far from being an equality. If so, one takes the critical value to be smaller than it would be if all moment inequalities were binding—both under the null and under the alternative.

Under a suitable sequence of null distributions $\{F_n : n \geq 1\}$, the asymptotic null distribution of $T_n(\theta)$ is the distribution of

$$S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0), \text{ where } Z^* \sim N(0_k, I_k), \tag{4.1}$$

$h_1 \in R^p_{+,\infty}$, Ω_0 is a $k \times k$ correlation matrix, and both h_1 and Ω_0 typically depend on the true value of θ. The correlation matrix Ω_0 can be consistently estimated. But the "$1/n^{1/2}$-local asymptotic mean parameter h_1 cannot be (uniformly) consistently estimated.[13]

A moment selection critical value is the $1-\alpha$ quantile of a data-dependent version of the asymptotic null distribution, $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$, that replaces Ω_0 by a consistent estimator and replaces h_1 with a p-vector in $R^p_{+,\infty}$ whose value depends on a measure of the slackness of the moment inequalities. The measure of slackness is

$$\xi_n(\theta) = \widehat{\kappa}^{-1} n^{1/2} \widehat{D}_n^{-1/2}(\theta) \overline{m}_n(\theta) \in R^k, \tag{4.2}$$

where $\widehat{\kappa}$ is a tuning parameter. For a GMS critical value, $\{\widehat{\kappa} = \kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \to \infty$, such as $\kappa_n = (\ln n)^{1/2}$ or $\kappa_n = (2 \ln \ln n)^{1/2}$. In contrast, for an RMS critical value, $\widehat{\kappa}$ does not go to infinity as $n \to \infty$ and is data-dependent.

[13] The asymptotic distribution of the test statistic $T_n(\theta)$ is a discontinuous function of the expected values of the moment inequality functions. This is not a feature of its finite sample distribution. For this reason, sequences of distributions $\{F_n : n \geq 1\}$ in which these expected values may drift to zero—rather than a fixed distribution F—need to be considered. See Andrews and Guggenberger (2008) for details.

The local parameter h_1 cannot be estimated consistently because doing so requires an estimator of the mean $h_1/n^{1/2}$ that is consistent at rate $o_p(n^{-1/2})$, which is not possible.

Data-dependence of $\widehat{\kappa}$ is obtained by taking $\widehat{\kappa}$ to depend on $\widehat{\Omega}_n(\theta)$:

$$\widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta)), \tag{4.3}$$

where $\kappa(\cdot)$ is a function from Ψ to R_{++}. A suitable choice of function $\kappa(\cdot)$ improves the power properties of the RMS procedure because the asymptotic power of the test depends on the probability limit of $\widehat{\kappa}$ through $\Omega(\theta)$.

We assume that $\kappa(\Omega)$ satisfies:

Assumption κ. (a) $\kappa(\Omega)$ is continuous at all $\Omega \in \Psi$. (b) $\kappa(\Omega) > 0$ for all $\Omega \in \Psi$.[14]

4.2 Moment Selection Function φ

Next, we discuss the moment selection function φ that determines how non-binding moment inequalities are detected. Let $\xi_{n,j}(\theta)$, $h_{1,j}$, and $[\Omega_0^{1/2} Z^*]_j$ denote the jth elements of $\xi_n(\theta)$, h_1, and $\Omega_0^{1/2} Z^*$, respectively, for $j = 1, ..., p$. When $\xi_{n,j}(\theta)$ is zero or close to zero, this indicates that $h_{1,j}$ is zero or fairly close to zero and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is 0. On the other hand, when $\xi_{n,j}(\theta)$ is large, this indicates $h_{1,j}$ is large and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is ∞ or some large value.

We replace $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ by $\varphi_j(\xi_n(\theta), \widehat{\Omega}_n(\theta))$ for $j = 1, ..., p$, where $\varphi_j : (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi \to R_{[\pm\infty]}$ is a function that is chosen to deliver the properties described above. The leading choices for the function φ_j are

$$\varphi_j^{(1)}(\xi, \Omega) = \begin{cases} 0 & \text{if } \xi_j \leq 1 \\ \infty & \text{if } \xi_j > 1, \end{cases} \quad \varphi_j^{(2)}(\xi, \Omega) = \psi(\xi_j),$$
$$\varphi_j^{(3)}(\xi, \Omega) = [\xi_j]_+, \text{ and } \varphi_j^{(4)}(\xi, \Omega) = \begin{cases} 0 & \text{if } \xi_j \leq 1 \\ \kappa(\Omega)\xi_j & \text{if } \xi_j > 1 \end{cases} \tag{4.4}$$

for $j = 1, ..., p$, where ψ is defined below and $\kappa(\Omega)$ in $\varphi_j^{(4)}$ is the same tuning parameter

[14] For simplicity, the recommended function $\kappa(\Omega) = \kappa(\delta(\Omega))$ given in Section 2.2 is constant on intervals of $\delta(\Omega)$ values and has jumps from one interval to the next. Hence, it does not satisfy Assumption κ. However, the function $\kappa(\delta)$ in Table I can be replaced by a continuous linearly-interpolated function whose value at the left-hand point in each interval of δ equals the value given in Table I. Such a function satisfies Assumption κ. Numerical calculations show that the grid of δ values in Table I is sufficiently fine that the finite-sample and asymptotic properties of the recommended RMS test are not sensitive to whether the $\kappa(\delta)$ function is linearly interpolated or not.

function that appears in (4.3). Let $\varphi^{(r)}(\xi, \Omega) = (\varphi_1^{(r)}(\xi, \Omega), ..., \varphi_p^{(r)}(\xi, \Omega),\ 0, ..., 0)' \in R^p_{[\pm\infty]} \times \{0\}^v$ for $r = 1, ..., 4$. CHT, AS, and Bugni (2007a,b) consider the function $\varphi^{(1)}$; AS considers $\varphi^{(2)}$; AS and Canay (2007) consider $\varphi^{(3)}$; and Fan and Park (2007) consider $\varphi^{(4)}$.[15]

The function $\varphi^{(1)}$ generates a "moment selection t-test" procedure, which is the recommended φ function. Note that $\xi_{n,j}(\theta_0) \leq 1$ is equivalent to the condition in (2.9).

The function $\varphi^{(2)}$ in (4.4) depends on a non-decreasing function $\psi(x)$ that satisfies $\psi(x) = 0$ if $x \leq a_L$, $\psi(x) \in [0, \infty]$ if $a_L < x < a_U$, and $\psi(x) = \infty$ if $x > a_U$, for some $0 < a_L \leq a_U \leq \infty$. A key condition is that $a_L > 0$. The function $\varphi^{(2)}$ is a continuous version of $\varphi^{(1)}$ when ψ is taken to be continuous on R (where continuity at a_U means that $\lim_{x \to a_U} \psi(x) = \infty$).

The functions $\varphi^{(3)}$ and $\varphi^{(4)}$ exhibit less steep rates of increase than $\varphi^{(1)}$ as functions of ξ_j for $j = 1, ..., p$.

For the asymptotic results given below, the only condition needed on the φ_j functions is that they are continuous on a set that has probability one under a certain distribution:

Assumption φ. For all $j = 1, ..., p$, all $\beta \in R^p_{[+\infty]} \times R^v$, and all $\Omega \in \Psi$, $\varphi_j(\xi, \Omega)$ is continuous at (ξ, Ω) for all ξ in a set $\Xi(\beta, \Omega) \subset (R^p_{[+\infty]} \times R^v) \times \Psi$ for which $P(\kappa^{-1}(\Omega)[\Omega^{1/2} Z^* + \beta \in \Xi(\beta, \Omega)) = 1$, where $Z^* \sim N(0_k, I_k)$.

The functions φ_j in (4.4) all satisfy Assumption φ.

The functions $\varphi^{(r)}$ for $r = 1, ..., 4$ all exhibit "element by element" determination of which moments to "select" because $\varphi_j^{(r)}(\xi, \Omega)$ only depends on (ξ, Ω) through ξ_j. This has significant computational advantages because $\varphi_j^{(r)}(\xi_n(\theta), \widehat{\Omega}_n(\theta))$ is very easy to compute. On the other hand, when $\widehat{\Omega}_n(\theta)$ is non-diagonal, the whole vector $\xi_n(\theta)$ contains information about the magnitude of the mean of $\overline{m}_n(\theta)$. The function $\varphi^{(5)}$ that is introduced in AS and defined below exploits this information. It is related to the information criterion-based moment selection criteria (MSC) considered in Andrews (1999) for a different moment selection problem. We refer to $\varphi^{(5)}$ as the modified MSC (MMSC) φ function. It is computationally more expensive than the functions $\varphi^{(1)}$-$\varphi^{(4)}$ considered above.

Define $c = (c_1, ..., c_k)'$ to be a selection k-vector of $0's$ and $1's$. If $c_j = 1$, the jth

[15] The function used by Fan and Park (2007) is not exactly equal to $\varphi_j^{(4)}$. Let $\widehat{\sigma}_{n,j}(\theta)$ denote the (j, j) element of $\widehat{\Sigma}_n(\theta)$. The function Fan and Park (2007) use is $\varphi_j^{(4)}(\xi, \Omega)$ with "if $\xi_j \leq 1$" replaced by "if $\widehat{\sigma}_{n,j}(\theta)\xi_j \leq 1$," and likewise for $>$ in place of $<$. This yields a non-scale-invariant φ_j function, which is not desirable, so we define $\varphi_j^{(4)}$ as is.

moment condition is selected; if $c_j = 0$, it is not selected. The moment equality functions are always selected, so $c_j = 1$ for $j = p+1, ..., k$. Let $|c| = \sum_{j=1}^{k} c_j$. For $\xi \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, define $c \cdot \xi = (c_1\xi_1, ..., c_k\xi_k)' \in R^p_{[+\infty]} \times R^v_{[\pm\infty]}$, where $c_j\xi_j = 0$ if $c_j = 0$ and $\xi_j = \infty$. Let $\mathcal{C}$ denote the parameter space for the selection vectors, e.g., $\mathcal{C} = \{0,1\}^p \times \{1\}^v$. Let $\zeta(\cdot)$ be a strictly increasing real function on R_+. Given $(\xi, \Omega) \in (R^p_{[+\infty]} \times R^v_{[\pm\infty]}) \times \Psi$, the selection vector $c(\xi, \Omega) \in \mathcal{C}$ that is chosen is the vector in $\mathcal{C}$ that minimizes the MMSC defined by

$$S(-c \cdot \xi, \Omega) - \zeta(|c|). \tag{4.5}$$

The minus sign that appears in the first argument of the S function ensures that a large *positive* value of ξ_j yields a large value of $S(-c \cdot \xi, \Omega)$ when $c_j = 1$, as desired. Since $\zeta(\cdot)$ is increasing, $-\zeta(|c|)$ is a bonus term that rewards inclusion of more moments. For $j = 1, ..., p$, define

$$\varphi_j^{(5)}(\xi, \Omega) = \begin{cases} 0 & \text{if } c_j(\xi, \Omega) = 1 \\ \infty & \text{if } c_j(\xi, \Omega) = 0. \end{cases} \tag{4.6}$$

The MMSC is analogous to the Bayesian information criterion (BIC) and the Hannan-Quinn information criterion (HQIC) when $\zeta(x) = x$, $\kappa_n = (\log n)^{1/2}$ for BIC, and $\kappa_n = (Q \ln\ln n)^{1/2}$ for some $Q \geq 2$ for HQIC, see AS. Some calculations show that when $\widehat{\Omega}_n(\theta)$ is diagonal, $S = S_1$ or S_2, and $\zeta(x) = x$, the function $\varphi^{(5)}$ reduces to $\varphi^{(1)}$.

4.3 RMS Critical Value $c_n(\theta)$

The (asymptotic normal) RMS critical value is equal to the $1 - \alpha$ quantile of $S\left(\Omega^{1/2} Z^* + \beta, \Omega\right)$ evaluated at $\beta = \varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta))$ and $\Omega = \widehat{\Omega}_n(\theta)$ plus a size-correction factor $\widehat{\eta}$. More specifically, given a choice of function

$$\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), ..., \varphi_p(\xi, \Omega), 0, ..., 0)' \in R^p_{[+\infty]} \times \{0\}^v, \tag{4.7}$$

the replacement for the k-vector $(h_1, 0_v)$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is given by

$$\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)). \tag{4.8}$$

For $Z^* \sim N(0_k, I_k)$ (independent of $\{W_i : i \geq 1\}$) and $\beta \in R^k_{[+\infty]}$, let $q_S(\beta, \Omega)$ denote the $1 - \alpha$ quantile of

$$S\left(\Omega^{1/2} Z^* + \beta, \Omega\right). \tag{4.9}$$

One can compute $q_S(\beta,\Omega)$ by simulating R i.i.d. random variables $\{Z_r^* : r = 1, ..., R\}$ with $Z_r^* \sim N(0_k, I_k)$ and taking $q_S(\beta,\Omega)$ to be the $1-\alpha$ sample quantile of $\{S(\Omega^{1/2}Z_r^* + \beta,\Omega) : r = 1, ..., R\}$, where R is large.

The nominal $1-\alpha$ (asymptotic normal) RMS critical value is

$$c_n(\theta) = q_S\left(\varphi\Big(\xi_n(\theta), \widehat{\Omega}_n(\theta)\Big), \widehat{\Omega}_n(\theta)\right) + \eta(\widehat{\Omega}_n(\theta)), \tag{4.10}$$

where $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ is a size-correction factor that is specified in Section 4.4 below.

The bootstrap RMS critical value is obtained by replacing $q_S(\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)), \widehat{\Omega}_n(\theta))$ in (4.10) by $q_S^*(\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)))$, where $q_S^*(\beta)$ is the $1-\alpha$ quantile of $S(M_{n,r}^*(\theta) + \beta, \widehat{\Omega}_{n,r}^*(\theta))$ for $\beta \in R_{[+\infty]}^k$ and $M_{n,r}^*(\theta)$ and $\widehat{\Omega}_{n,r}^*(\theta)$ are defined in (2.8). The quantity $q_S^*(\varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)))$ can be computed by taking the $1-\alpha$ sample quantile of $\{S(M_{n,r}^*(\theta) + \varphi(\xi_n(\theta), \widehat{\Omega}_n(\theta)), \widehat{\Omega}_{n,r}^*(\theta)) : r = 1, ..., R\}$.

For our preferred RMS critical value discussed in Section 2.2, the asymptotic normal critical value is of the form in (4.10) with $S = S_2$, $\varphi = \varphi^{(1)}$, and $\eta(\Omega) = \eta_1(\delta(\Omega)) + \eta_2(p)$. The bootstrap critical value uses $q_{S_2}^*(\cdot)$ in place of $q_{S_2}(\cdot, \widehat{\Omega}_n(\theta))$.

4.4 Size-Correction Factor $\widehat{\eta}$

We now discuss the size-correction factor $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$. Such a factor is necessary to deliver correct asymptotic size under asymptotics in which $\widehat{\kappa}$ does not diverge to infinity. This factor can viewed as giving an asymptotic size refinement to a GMS critical value.

As noted above, we show in the proofs (see Appendix A in the Supplement) that under a suitable sequence of true parameters and distributions $\{(\theta_n, F_n) : n \geq 1\}$, $T_n(\theta_n) \to_d S(\Omega^{1/2}Z^* + (h_1, 0_v), \Omega)$ for some $(h_1, \Omega) \in R_{+,\infty}^p \times \Psi$. Furthermore, we show that under such a sequence the asymptotic coverage probability of an RMS CS based on a data-dependent tuning parameter $\widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta))$ and a fixed size-correction constant η is

$$\begin{aligned} CP(h_1, \Omega, \eta) = P\big(& S\left(\Omega^{1/2}Z^* + (h_1, 0_v), \Omega\right) \leq \\ & q_S\left(\varphi\left(\kappa^{-1}(\Omega)[\Omega^{1/2}Z^* + (h_1, 0_v)], \Omega\right), \Omega\right) + \eta\big), \end{aligned} \tag{4.11}$$

where $Z^* \sim N(0_k, I_k)$. (Correspondingly, the null rejection probability of an RMS test with fixed η for testing $H_0 : \theta = \theta_0$ is $1 - CP(h_1, \Omega, \eta)$.)

We let $\Delta \subset R_{+,\infty}^p \times cl(\Psi)$ denote the set of all (h_1, Ω) values that can arise given the model specification $\mathcal{F}$.[16] Our primary focus is on the standard case in which

$$\Delta = R^p_{+,\infty} \times cl(\Psi). \tag{4.12}$$

This arises when there are no restrictions on the moment functions beyond the inequality/equality restrictions and h_1 and Ω are variation free. Our asymptotic results cover the general case in which Δ may be restricted, as well as the standard case in (4.12).

To determine the asymptotic size of an RMS test or CS, it suffices to have $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ satisfy:

Assumption η1. $\eta(\Omega)$ is continuous at all $\Omega \in \Psi$.[17]

However, for an RMS CS to have asymptotic size greater than or equal to $1-\alpha$, $\eta(\cdot)$ must be chosen to satisfy the first condition that follows. If it also satisfies the second, stronger, condition, then its asymptotic size equals $1-\alpha$. Let $CP(h_1, \Omega, \eta(\Omega)-) = \lim_{x \downarrow 0} CP(h_1, \Omega, \eta(\Omega) - x)$.

Assumption η2. $\inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega)-) \geq 1-\alpha$.

Assumption η3. (a) $\inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega)) = 1-\alpha$.

(b) $\inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega)-) = \inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega))$.

Assumption η3(b) is a continuity condition that is not restrictive. The left-hand side (lhs) quantity inside the probability in (4.11) has a df that is continuous and strictly increasing for positive values. The corresponding right-hand side (rhs) quantity is positive. These two quantities are quite different nonlinear functions of the same underlying normal random vector. Hence, they are equal with probability zero, which implies that Assumption η3(b) holds.

The function $\eta(\Omega)$ depends on S, φ, and the tuning parameter function $\kappa(\Omega)$. For notational simplicity, we suppress this dependence. Functions $\eta(\cdot)$ that satisfy Assumptions η2 and/or η3 are not uniquely defined. The smallest function that satisfies Assumption η3(a), denoted $\eta^*(\Omega)$, exists and is defined as follows. For each $\Omega \in \Psi$, define $\eta^*(\Omega)$ to be the smallest value η for which

$$\inf_{h_1:(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta) = 1-\alpha.^{18} \tag{4.13}$$

[16] A more precise/detailed definition of Δ is given in Section A of the Supplement.

[17] An analogous comment to that in footnote 14 also applies to the recommended function $\eta(\cdot)$ given in Section 2.2 and Assumption η1.

[18] A smallest value exists because $CP(h_1, \Omega, \eta)$ is right continuous in η.

When Δ satisfies (4.12), the infimum is over $h_1 \in R^p_{+,\infty}$. For purposes of minimizing the probability of false coverage of the CS (or equivalently, maximizing the power of the tests upon which the CS is based), it is desirable to take $\eta(\Omega)$ as close to $\eta^*(\Omega)$ as possible subject to $\eta(\Omega) \geq \eta^*(\Omega)$. For computational tractability and storability, however, it is convenient to use a function $\eta(\cdot)$ that is simpler than $\eta^*(\Omega)$, e.g., a function that depends on Ω only through a scalar function of Ω, as with the recommended RMS critical value described in Section 2.2.[19]

4.5 Plug-in Asymptotic Critical Values

We now discuss CS's based on a plug-in asymptotic (PA) critical value. The least-favorable asymptotic null distributions of the statistic $T_n(\theta)$ are those for which the moment inequalities hold as equalities. These distributions depend on the correlation matrix Ω of the moment functions. PA critical values are determined by the least-favorable asymptotic null distribution for given Ω evaluated at a consistent estimator of Ω. Such critical values have been considered in the literature on multivariate one-sided tests, see Silvapulle and Sen (2005) for references. CHT, AG, and AS consider them in the context of the moment inequality literature. Rosen (2008) considers variations of PA critical values that make adjustments in the case where it is known that if one moment inequality holds as an equality then another cannot.[20]

The PA critical value is

$$q_S(0_k, \widehat{\Omega}_n(\theta)). \tag{4.14}$$

The PA critical value can be viewed as a special case of an RMS critical value with $\varphi_j(\xi, \Omega) = 0$ for all $j = 1, ..., k$ and $\eta(\widehat{\Omega}_n(\theta)) = 0$. This implies that the asymptotic results stated below for RMS CS's and tests also apply to PA CS's and tests.

[19] Note that even if $\eta(\Omega) \neq \eta^*(\Omega)$, Assumption η3(a) still can hold.

[20] This method delivers corrrect asymptotic size in a uniform sense only if when one moment inequality holds as an equality then the other is strictly bounded away from zero.

5 Asymptotic Results

5.1 Asymptotic Size

The exact and asymptotic confidence sizes of an RMS CS are

$$ExCS_n = \inf_{(\theta,F)\in\mathcal{F}} P_F(T_n(\theta) \leq c_n(\theta)) \text{ and } AsyCS = \liminf_{n\to\infty} ExCS_n, \tag{5.1}$$

respectively. The definition of $AsyCS$ takes the "inf" before the "lim." This builds uniformity over (θ, F) into the definition of $AsyCS$. Uniformity is required for the asymptotic size to give a good approximation to the finite-sample size of a CS.

Theorems 1 and 2 below apply to i.i.d. observations, in which case $\mathcal{F}$ is defined in (2.2). They also apply to stationary temporally-dependent observations and to cases in which the moment functions depend on a preliminary consistent estimator of a parameter τ, in which cases for brevity $\mathcal{F}$ is defined in Appendix A in the Supplement.

Theorem 1 *Suppose Assumptions* S, κ, φ, *and* $\eta 1$ *hold and* $0 < \alpha < 1$. *Then, the nominal level* $1-\alpha$ RMS CS *based on* S, φ, $\widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta))$, *and* $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ *satisfies*

(a) $AsyCS \in [\inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega)-), \inf_{(h_1,\Omega)\in\Delta} CP(h_1, \Omega, \eta(\Omega))]$,

(b) $AsyCS \geq 1-\alpha$ *provided Assumption* $\eta 2$ *holds, and*

(c) $AsyCS = 1-\alpha$ *provided Assumption* $\eta 3$ *holds.*

Comments. 1. Theorem 1(b) shows that an RMS CS based on a size-correction factor $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ that satisfies Assumption $\eta 2$ is asymptotically valid in a uniform sense under asymptotics that do not require $\widehat{\kappa} \to \infty$ as $n \to \infty$. In contrast, the GMS CS introduced in AS requires $\widehat{\kappa} \to \infty$ as $n \to \infty$.

2. Theorem 1 holds even if there are restrictions such that one moment inequality cannot hold as an equality if another moment inequality does. Rosen (2008) discusses models in which restrictions of this sort arise.

3. Theorem 1 applies to moment conditions based on weak instruments (because the tests considered are of an Anderson-Rubin form.)

4. Define the asymptotic size of an RMS test of $H_0 : \theta = \theta_0$ by

$$AsySz(\theta_0) = \limsup_{n\to\infty} \sup_{(\theta,F)\in\mathcal{F}:\theta=\theta_0} P_F(T_n(\theta_0) > c_n(\theta_0)). \tag{5.2}$$

The proof of Theorem 1 shows that under the assumptions in Theorem 1, (a) $AsySz(\theta_0) \in$

$[1 - \inf_{(h_1,\Omega)\in\Delta_0} CP(h_1, \Omega, \eta(\Omega)), 1 - \inf_{(h_1,\Omega)\in\Delta_0} CP(h_1, \Omega, \eta(\Omega)-)]$, where Δ_0 is defined as Δ is defined in (4.12) or in a more general case Δ is defined as in (1.2) of Appendix A in the Supplement but with the sequence $\{\theta_{w_n} : n \geq 1\}$ replaced by the constant θ_0, (b) $AsySz(\theta_0) \leq \alpha$ provided Assumption $\eta 2$ holds, and (c) $AsySz(\theta_0) = \alpha$ provided Assumption $\eta 3$ holds, where Δ in Assumptions $\eta 2$ and $\eta 3$ is replaced by Δ_0. The primary case of interest is when $\Delta_0 = R^p_{+,\infty} \times cl(\Psi)$, which occurs when there are no restrictions on the moment functions beyond the inequality/equality restrictions and h_1 and Ω are variation free.

5. The proofs of Theorem 1 and all other results in the paper are provided in Appendix A in the Supplement.

5.2 Asymptotic Power

In this section, we compute the asymptotic power of RMS tests against $1/n^{1/2}$-local alternatives. These results have immediate consequences for the length or volume of a CS based on these tests because the power of a test for a point that is not the true value is the probability that the CS does not include that point. (See Pratt (1961) for an equation that links CS volume and probabilities of false coverage.) We use these results to define tuning parameters $\kappa = \kappa(\Omega)$ and size-correction factors $\eta = \eta(\Omega)$ that maximize average power for a selected set of alternative parameter values. We also use the results to compare different choices of test function S and moment selection function φ in terms of average asymptotic power.

For given θ_0, we consider tests of

$$\begin{aligned} H_0 : E_F m_j(W_i, \theta_0) &\geq 0 \text{ for } j = 1, ..., p \text{ and} \\ E_F m_j(W_i, \theta_0) &= 0 \text{ for } j = p+1, ..., k, \end{aligned} \tag{5.3}$$

where F denotes the true distribution of the data. (More precisely, by this we mean H_0: the true $(\theta, F) \in \mathcal{F}$ satisfies $\theta = \theta_0$.) The alternative is $H_1 : H_0$ does not hold.

Let

$$\begin{aligned} \sigma^2_{F,j}(\theta) &= AsyVar_F(n^{1/2}\overline{m}_{n,j}(\theta)) \text{ for } j = 1, ..., p, \\ D(\theta, F) &= Diag\{\sigma^2_{F,1}(\theta), ..., \sigma^2_{F,k}(\theta)\}, \text{ and} \\ \Omega(\theta, F) &= AsyCorr_F(n^{1/2}\overline{m}_n(\theta)). \end{aligned} \tag{5.4}$$

Note that this definition of $\sigma^2_{F,j}(\theta)$ reduces to that given in (2.2) when the observations are i.i.d.

We now introduce the $1/n^{1/2}$-local alternatives. The first two assumptions are the same as in AS. The third assumption is a high-level assumption that allows for dependent observations and sample moment functions that may depend on a preliminary estimator $\widehat{\tau}_n(\theta)$. It is shown to hold automatically with i.i.d. observations when there is no preliminary estimator of a parameter τ.

Assumption LA1. The true parameters $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ satisfy:

(a) $\theta_n = \theta_0 - \lambda n^{-1/2}(1 + o(1))$ for some $\lambda \in R^d$ and $F_n \to F_0$ for some $(\theta_0, F_0) \in \mathcal{F}$,

(b) $n^{1/2} E_{F_n} m_j(W_i, \theta_n)/\sigma_{F_n,j}(\theta_n) \to h_{1,j}$ for some $h_{1,j} \in R_{+,\infty}$ for $j = 1, ..., p$, and

(c) $\sup_{n \geq 1} E_{F_n} |m_j(W_i, \theta_0)/\sigma_{F_n,j}(\theta_0)|^{2+\delta} < \infty$ for $j = 1, ..., k$ for some $\delta > 0$.

Assumption LA2. The $k \times d$ matrix $\Pi(\theta, F) = (\partial/\partial\theta')[D^{-1/2}(\theta, F) E_F m(W_i, \theta)]$ exists and is continuous in (θ, F) for all (θ, F) in a neighborhood of (θ_0, F_0).[21]

Assumption LA3. The true parameters $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ satisfy:

(a) $A^0_n = (A^0_{n,1}, ..., A^0_{n,k})' \to_d Z \sim N(0_k, \Omega_0)$ as $n \to \infty$, where $A^0_{n,j} = n^{1/2}(\overline{m}_{n,j}(\theta_0) - E_{F_n} m_j(W_i, \theta_0))/\sigma_{F_n,j}(\theta_0)$,

(b) $\widehat{\sigma}_{n,j}(\theta_0)/\sigma_{F_n,j}(\theta_0) \to_p 1$ as $n \to \infty$ for $j = 1, ..., k$, and

(c) $\widehat{D}_n^{-1/2}(\theta_0)\widehat{\Sigma}_n(\theta_0)\widehat{D}_n^{-1/2}(\theta_0) \to_p \Omega_0$ as $n \to \infty$.

When the observations are i.i.d. for each $(\theta, \Omega) \in \mathcal{F}$, Assumption LA3 holds automatically as shown in the following Lemma.

Assumption LA3*. (a) For each $n \geq 1$, the observations $\{W_i : i \leq n\}$ are i.i.d. under $(\theta_n, F_n) \in \mathcal{F}$, (b) $\widehat{\Sigma}_n(\theta)$ is defined by (2.5), and (c) no preliminary estimator of a parameter τ appears in the sample moment functions.

Lemma 1 *Assumptions* LA1 *and* LA3* *imply Assumption* LA3.

The asymptotic distribution of $T_n(\theta_0)$ under local alternatives depends on the limit of the normalized moment inequality functions when evaluated at the null value θ_0. Under Assumptions LA1 and LA2, it can be shown that

$$\lim_{n\to\infty} n^{1/2} D^{-1/2}(\theta_0, F_n) E_{F_n} m(W_i, \theta_0) = \mu = (h_1, 0_v) + \Pi_0 \lambda \in R^k, \text{ where}$$

[21] When a preliminary estimator of a parameter τ appears in the sample moment functions, then in Assumptions LA1 and LA2 and (5.5) below, $m_j(W_i, \theta)$ and $m(W_i, \theta)$ are defined to be $m_j(W_i, \theta, \tau_0)$ and $m(W_i, \theta, \tau_0)$, respectively, where τ_0 denotes the true value of the parameter τ under the true distribution F.

$$h_1 = (h_{1,1}, ..., h_{1,p})' \text{ and } \Pi_0 = \Pi(\theta_0, F_0). \tag{5.5}$$

By definition, if $h_{1,j} = \infty$, then $h_{1,j} + x = \infty$ for any $x \in R$. Let $\Pi_{0,j}$ denote the jth row of Π_0 written as a column d-vector for $j = 1, ..., k$. Note that $(h_1, 0_v) + \Pi_0\lambda \in R^p_{[+\infty]} \times R^v$. Let $\mu = (\mu_1, ..., \mu_k)'$. The true distribution F_n is in the alternative, not the null (for n large) when $\mu_j = h_{1,j} + \Pi'_{0,j}\lambda < 0$ for some $j = 1, ..., p$ or $\Pi'_{0,j}\lambda \neq 0$ for some $j = p+1, ..., k$.

For constants $\kappa > 0$ and $\eta \geq 0$, define

$$\begin{aligned} & AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta) \\ &= P\left(S(\Omega^{1/2}Z^* + \mu, \Omega) > q_S\left(\varphi(\kappa^{-1}[\Omega^{1/2}Z^* + \mu], \Omega), \Omega\right) + \eta\right) \text{ and} \\ & AsyPow^-(\mu, \Omega_0, S, \varphi, \kappa, \eta) = \lim_{x \downarrow 0} AsyPow(\mu, \Omega_0, S, \varphi, \kappa, \eta - x), \end{aligned} \tag{5.6}$$

where $Z^* \sim N(0_k, I_k)$, $\mu \in R^k$, $\Omega \in \Psi$, $\kappa \in R_{++}$, the functions S, φ, and q_S are as defined in Section 3, (4.4) or (4.6), and (4.9), respectively.[22] Typically, $AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta) = AsyPow^-(\mu, \Omega, S, \varphi, \kappa, \eta)$ because the lhs quantity in the probability in (5.6) is a nonlinear function of a normal random vector that has a continuous and strictly increasing df (unless $v = 0$ and $\mu = \infty^p$, which cannot hold under the alternative hypothesis) and the rhs quantity in the probability in (5.6) is a quite different nonlinear function of the same normal random vector.

For a sequence of constants $\{\zeta_n : n \geq 1\}$, let $\zeta_n \to [\zeta_{1,\infty}, \zeta_{2,\infty}]$ denote that $\zeta_{1,\infty} \leq \liminf_{n\to\infty} \zeta_n \leq \limsup_{n\to\infty} \zeta_n \leq \zeta_{2,\infty}$.

Theorem 2 *Under Assumptions* S, κ, φ, η1, *and* LA1-LA3, *the* RMS *test based on* $S, \varphi, \widehat{\kappa} = \kappa(\widehat{\Omega}_n(\theta))$, *and* $\widehat{\eta} = \eta(\widehat{\Omega}_n(\theta))$ *satisfies*

$$\begin{aligned} & P_{F_n}(T_n(\theta_0) > c_n(\theta_0)) \\ & \to [AsyPow(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0)), AsyPow^-(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0))], \end{aligned}$$

where $\mu = (h_1, 0_v) + \Pi_0\lambda$.

Comments. 1. Theorem 2 provides the $1/n^{1/2}$-local alternative power function of RMS and PA tests. Typically, $AsyPow(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0)) = AsyPow^-(\mu, \Omega_0, S, \varphi, \kappa(\Omega_0), \eta(\Omega_0))$ and the asymptotic local power function is unique for any given (μ, Ω_0).

[22] For some functions φ, such as $\varphi^{(1)}$ and $\varphi^{(4)}$, $\kappa = 0$ is permissible because $\lim_{\kappa\downarrow 0} \varphi(\kappa^{-1}[\Omega^{1/2}Z + \mu], \Omega)$ is well-defined. For example, for $\varphi^{(1)}$ and $x \in R$, $\lim_{\kappa\downarrow 0} \varphi(\kappa^{-1}x, \Omega) = 0$ if $x \leq 0$ and $\lim_{\kappa\downarrow 0} \varphi(\kappa^{-1}x, \Omega) = \infty$ if $x > 0$.

2. The results of Theorem 2 hold under the null and alternative hypotheses.

3. For moment conditions based on weak instruments, the results of Theorem 2 still hold. But, with weak instruments, RMS and PA tests have power less than or equal to α against $1/n^{1/2}$-local alternatives because $\Pi_{0,j}'\lambda = 0$ for all $j = 1, ..., k$.

5.3 Average Power

RMS tests depend on S, φ, $\kappa(\Omega)$, and $\eta(\Omega)$. We compare the power of RMS tests by comparing their average asymptotic power for a chosen set $\mathcal{M}_k(\Omega)$ of alternative parameter vectors $\mu \in R^k$ for $\Omega \in \Psi$.[23] Let $|\mathcal{M}_k(\Omega)|$ denote the number of elements in $\mathcal{M}_k(\Omega)$. The average asymptotic power of the RMS test based on $(S, \varphi, \kappa, \eta)$ for constants $\kappa > 0$ and $\eta \geq 0$ is

$$|\mathcal{M}_k(\Omega)|^{-1} \sum_{\mu \in \mathcal{M}_k(\Omega)} AsyPow(\mu, \Omega, S, \varphi, \kappa, \eta). \tag{5.7}$$

We are interested in comparing the (S, φ) functions defined in (2.4), (3.2), (3.3), (4.4), and (4.6) in terms of $\mathcal{M}_k(\Omega)$-average asymptotic power. To do so requires choices of functions $(\kappa(\cdot), \eta(\cdot))$ for each (S, φ). We use the tuning and size-correction functions $\kappa^*(\Omega)$ and $\eta^*(\Omega)$ that are optimal in terms of $\mathcal{M}_k(\Omega)$-average asymptotic power. They are defined as follows. Given Ω and $\kappa > 0$, let $\eta^*(\Omega, \kappa)$ be defined as in (4.13) with $\Delta = R^p_{+,\infty} \times cl(\Omega)$ and tuning parameter $\kappa > 0$. The optimal tuning parameter $\kappa^*(\Omega)$ maximizes (5.7) with η replaced by $\eta^*(\Omega, \kappa)$ over $\kappa > 0$. The optimal size-correction factor then is $\eta^*(\Omega) = \eta^*(\Omega, \kappa^*(\Omega))$ and the test based on $(\kappa^*(\Omega), \eta^*(\Omega))$ has asymptotic size α. (Obviously, $\kappa^*(\cdot)$ and $\eta^*(\cdot)$ depend on (S, φ).)

Given $\eta^*(\Omega)$ and $\kappa^*(\Omega)$, we compare (S, φ) functions by comparing their values of

$$|\mathcal{M}_k(\Omega)|^{-1} \sum_{\mu \in \mathcal{M}_k(\Omega)} AsyPow(\mu, \Omega, S, \varphi, \kappa^*(\Omega), \eta^*(\Omega)), \tag{5.8}$$

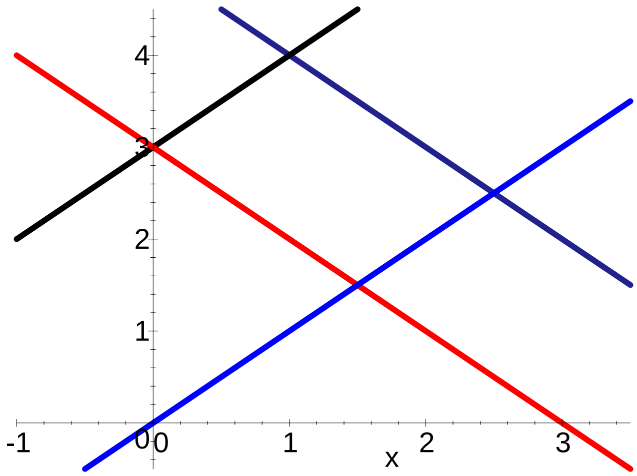
which depend on Ω.

We are interested in constructing tests that yield CS's that are as small as possible. The boundary of a CS, like the boundary of the identified set, is determined at any given point by the moment inequalities that are binding at that point. The number of

[23] As indicated, we allow this set to depend on Ω. The reason is that the power of any test and the asymptotic power envelope depend on Ω. Hence, it is natural to vary the magnitude of $||\mu||$ for $\mu \in \mathcal{M}_k(\Omega)$ as Ω varies.

binding moment inequalities at a point depends on the dimension, d, of the parameter θ. Typically, the boundary of a confidence set is determined by d (or fewer) moment inequalities. That is, at most d moment inequalities are binding and at least $p - d$ are slack, see Figure 1. In consequence, we specify the sets $\mathcal{M}_k(\Omega)$ considered below to be ones for which most vectors μ have half or more elements positive (since positive elements correspond to non-binding inequalities), which is suitable for the typical case in which $p \geq 2d$.

Figure 1. Confidence Set for a Parameter $\theta \in R^d$ for $d = 2$ Based on $p = 4$ Moment Inequalities



5.4 Asymptotic Power Envelope

To assess the power performance of RMS tests in an absolute sense, it is of interest to compare their asymptotic power to the asymptotic power envelope. For details on the determination and computation of the latter, see Appendix C in the Supplement.[24]

We note that the asymptotic power envelope is a "uni-directional" envelope. One does not expect a test that is designed to perform well for multi-directional alternatives to be on, or close to, the uni-directional envelope. This is analogous to the fact that the power of a standard F-test for a p-dimensional restriction with an unrestricted alternative hypothesis in a normal linear regression model is not close to the uni-dimensional power envelope. For example, for $p = 2, 4, 10$, when the asymptotic power envelope

[24] Methods similar to those used in Andrews, Moreira, and Stock (2008) and Müller and Watson (2008) are employed.

is $.75, .80, .85$, respectively, the F test has power $.65, .60, .49$, respectively.[25] Clearly, the larger is p the greater is the difference between the power of a test designed for p-directional alternatives and the uni-directional power envelope.

6 Numerical Results

6.1 Introduction

In the numerical work, we focus on results for $p = 2, 4$, and 10 and $v = 0$, which represent small, medium, and large numbers of moment inequalities respectively. Results for $p = 2$ are of special interest because the correlation matrix Ω is very simple in this case. It just depends on a scalar $\rho \in [-1, 1]$. Hence, it is easy to see how the magnitude of ρ affects key quantities, such as asymptotic null rejection probabilities of tests, size-corrected asymptotic power of tests, and the asymptotic power envelope.

For each value of p, we consider three representative correlation matrices Ω: Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}. The matrix Ω_{Zero} equals I_p for $p = 2, 4$, and 10. The matrices Ω_{Neg} and Ω_{Pos} are Toeplitz matrices with correlations on the diagonals given by the following: For $p = 2$: $\rho = -.9$ for Ω_{Neg} and $\rho = .5$ for Ω_{Pos}. For $p = 4$: $\rho = (-.9, .7, -.5)$ for Ω_{Neg} and $\rho = (.9, .7, .5)$ for Ω_{Pos}. For $p = 10$: $\rho = (-.9, .8, -.7, .6, -.5, .4, -.3, .2, -.1)$ for Ω_{Neg} and $\rho = (.9, .8, .7, .6, .5, ..., .5)$ for Ω_{Pos}.

For $p = 2$, the set of μ vectors $\mathcal{M}_2(\Omega)$ for which average asymptotic power is computed includes seven elements:

$$\begin{aligned}\mathcal{M}_2(\Omega) = \{&(-\mu_1, 0), (-\mu_2, 1), (-\mu_3, 2), (-\mu_4, 3),\\ &(-\mu_5, 4), (-\mu_6, 7), (-\mu_7, -\mu_7)\},\end{aligned} \tag{6.1}$$

where μ_j depends on Ω and is such that the power envelope is .75 at each element of $\mathcal{M}_2(\Omega)$. Consistent with the discussion in Section 5.3, most elements of $\mathcal{M}_2(\Omega)$ have less than p negative elements. The positive elements of the μ vectors are chosen to cover a reasonable range of the parameter space. The simulations used to compute the values μ_j for $\Omega = \Omega_{Neg}, \Omega_{Zero}, \Omega_{Pos}$ are based on 40,000 simulation repetitions to determine the critical value of the simple-versus-simple LR tests that yield the power envelope and

[25] These asymptotic power results are obtained by some simple calculations based on the distribution function of the noncentral χ^2 distribution with $p = 1, 2, 4, 10$ degrees of freedom, where the noncentral χ^2 distribution with $p = 1$ degrees of freedom is used for the power envelope calculations.

40,000 repetitions to determine the power of these tests. (The same is true for the cases where $p = 4, 10$ discussed below.) For brevity, the values of μ_j in (6.1) are given in Appendix C in the Supplement.

For $p = 4$, $\mathcal{M}_4(\Omega)$ includes 24 elements:

$$\begin{aligned}
&\mathcal{M}_4(\Omega)\\
&= \{(-\mu_1, -\mu_1, 1, 1), (-\mu_2, -\mu_2, 2, 2), (-\mu_3, -\mu_3, 3, 3), (-\mu_4, -\mu_4, 4, 4), (-\mu_5, -\mu_5, 7, 7),\\
&\quad (-\mu_6, -\mu_6, 1, 7), (-\mu_7, -\mu_7, 2, 7), (-\mu_8, -\mu_8, 3, 7), (-\mu_9, -\mu_9, 4, 7),\\
&\quad (-\mu_{10}, 1, 1, 1), (-\mu_{11}, 2, 2, 2), (-\mu_{12}, 3, 3, 3), (-\mu_{13}, 4, 4, 4), (-\mu_{14}, 7, 7, 7),\\
&\quad (-\mu_{15}, 1, 1, 7), (-\mu_{16}, 2, 2, 7), (-\mu_{17}, 3, 3, 7), (-\mu_{18}, 4, 4, 7), (-\mu_{19}, -\mu_{19}, 0, 0),\\
&\quad (-\mu_{20}, 0, 0, 0), (-\mu_{21}, 25, 25, 25), (-\mu_{22}, -\mu_{22}, 25, 25), (-\mu_{23}, -\mu_{23}, -\mu_{23}, 25),\\
&\quad (-\mu_{24}, -\mu_{24}, -\mu_{24}, -\mu_{24})\}, \qquad (6.2)
\end{aligned}$$

where μ_j depends on Ω and is such that the power envelope is .80 at each element of $\mathcal{M}_4(\Omega)$.

For $p = 10$, $\mathcal{M}_{10}(\Omega)$ includes 40 vectors. For brevity, they are specified in Appendix C in the Supplement. They include 10 vectors with 2 negative components and with the other components taking a variety of positive values, 10 vectors with 4 negative components, 10 vectors with 1 negative component, and 10 vectors with 1-10 negative components and with the other elements positive and large.

In addition to the main results based on (i) the correlation matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos}, we also provide results based on (ii) a grid of 19 different Ω matrices, each with a different "amounts" of correlation, and (iii) 500 Ω matrices for $p = 2, 4$ and 250 for $p = 10$ obtained by simulation. Details concerning these Ω matrices are given in Appendix C in the Supplement.

6.2 Comparison of (S, φ) Functions

In this section, we compare tests based on different (S, φ) functions. We consider the following combinations: (S, φ) = (MMM, PA), (MMM, t-Test), (Max, PA), (Max, t-Test), (SumMax, PA), (SumMax, t-Test), (QLR, PA), (QLR, t-Test), (QLR, $\varphi^{(3)}$), (QLR, $\varphi^{(4)}$), and (QLR, MMSC).[26] We also consider "pure" GEL tests, which combine

[26] The statistics MMM, QLR, Max, and SumMax are based on the functions S_1, S_2, S_3 with $p_1 = 1$, and S_3 with $p_1 = 2$, respectively. The t-Test and MMSC critical values corresponds to the functions

GEL statistics with a critical value that is the same for all Ω. GEL statistics behave the same as the QLR statistic asymptotically.[27]

For each RMS test, we report the average asymptotic power for the κ value that maximizes average asymptotic power, denoted κ=Best. We do this because we are interested in determining first which test has the highest power when κ is chosen optimally. Then we determine a suitable data-dependent choice of κ.

In this section we report results for the three matrices $(\Omega_{Neg}, \Omega_{Zero}, \Omega_{Pos})$. In Appendix B in the Supplement we report additional results based on 19 Ω matrices that cover a grid of $\delta(\Omega)$ values from $-.99$ to $.99$. The qualitative results reported here are found to apply as well to the broader range of 19 Ω matrices.

The best κ values for the RMS tests are determined numerically using grid search, see Appendix C in the Supplement for details. The best κ values are specified in Table B-I in Appendix B in the Supplement. The table shows that for all tests and $p = 2, 4, 10$, the best κ values are decreasing from Ω_{Neg} to Ω_{Zero} to Ω_{Pos}. For the QLR/t-Test test, the best κ values for $(\Omega_{Neg}, \Omega_{Zero}, \Omega_{Pos})$ are $(2.50, 1.75, .00)$ for $p = 10$, $(2.75, 1.50, .25)$ for $p = 4$, and $(2.75, 1.50, .75)$ for $p = 2$. The best κ values for the other tests that use the t-Test and $\varphi^{(4)}$ critical values are roughly similar. The best κ values for the tests that use the $\varphi^{(3)}$ and MMSC critical values are noticeably larger, at least for Ω_{Neg}.

Table II provides asymptotic average power results for $p = 2, 4, 10$ and $\Omega = \Omega_{Neg}$, $\Omega_{Zero}, \Omega_{Pos}$. The asymptotic power results are size-corrected.[28] Except where stated otherwise, the size-correction factors are calculated using $40,000$ simulation repetitions and the power results are obtained using $40,000$ repetitions, which yields a simulation standard error of .0011.

Now we discuss the asymptotic power results given in Table II. Table II shows that the MMM/PA test has very low asymptotic power compared to the QLR/t-Test/κBest test (which is shown in boldface) especially for $p = 4, 10$. Similarly, the Max/PA and SumMax/PA tests have low power. The QLR/PA test has better power than the other PA tests, but it is still very low compared to the QLR/t-Test/κBest test.

$\varphi^{(1)}$ and $\varphi^{(5)}$, respectively.

[27] The level .05 pure GEL asymptotic critical values are determined numerically by calculating the constant for which the maximum null rejection probability of the QLR statistic over all mean vectors in the null hypothesis and over all positive definite correlation matrices Ω is .05. The critical values are found to be 5.07, 7.94, and 16.2 for $p = 2$, 4, and 10, respectively. These critical values yield null rejection rates of .05 when Ω contains elements that are close to -1.0.

[28] Size-correction here is done for the fixed known value of Ω. It is not based on the least-favorable Ω matrix because the results are asymptotic and Ω can be estimated consistently.

The "pure GEL" test has very poor power properties. For example, for $p = 10$, its power is between 1/3 and 1/6 that of the QLR/t-Test/κBest test (and of the feasible QLR/t-Test/κAuto test, which is the recommended test of Section 2.2).

Table II shows that the MMM/t-Test/κBest test has equal average asymptotic power to the QLR/t-Test/κBest test for Ω_{Zero} and only slightly lower power for Ω_{Pos}. But, it has substantially lower power for Ω_{Neg}. For example, for $p = 10$, the comparison is .19 versus .59. The Max/t-Test/κBest test has noticeably lower average power than the QLR/t-Test/κBest test for Ω_{Neg} and Ω_{Zero} and essentially equal power for Ω_{Pos}. It is strongly dominated in terms of average power. Results for individual μ vectors show that the Max/t-Test/κBest and QLR/t-Test/κBest tests have similar average power over μ vectors that have only one negative element, but the Max/t-Test/κBest has substantially lower average power over μ vectors that have more than one negative element. For example, for $p = 4$ and Ω_{Neg}, the Max/t-Test/κBest and QLR/t-Test/κBest tests have average asymptotic powers of .62 and .63, respectively, for μ vectors that have one element negative, but .13 and .61 for μ vectors with two or more negative elements. The SumMax/t-Test/κBest test also is strongly dominated by the QLR/t-Test/κBest test in terms of average asymptotic power. The power differences between these two tests are especially large for Ω_{Neg}. For example, for $p = 10$ and Ω_{Neg}, their powers are .14 and .59, respectively.

Next we compare tests that use the QLR test statistic but different critical values—due to the use of different moment selection functions φ. The QLR/$\varphi^{(3)}$/κBest test has noticeably lower average asymptotic power than the QLR/t-Test/κBest test for Ω_{Neg}, somewhat lower power for Ω_{Zero}, and equal power for Ω_{Pos}. The differences increase with p.

The QLR/$\varphi^{(4)}$/κBest test has the same average asymptotic power as the QLR/t-Test/ κBest test in all cases considered. This is because the $\varphi^{(4)}$ and $\varphi^{(1)}$ functions are similar. The QLR/MMSC/κBest test has the same average asymptotic power as the QLR/t-Test/κBest test for $p = 2$, for $p = 4$ with Ω_{Zero} and Ω_{Pos}, and for $p = 10$ with Ω_{Zero}. For $p = 4$ and Ω_{Neg}, its power is higher by .03 and for $p = 10$ and Ω_{Neg}, its power is higher by .06, but for $p = 10$ and Ω_{Pos}, its power is lower by .04. Hence, these two tests have similar power but, if anything, that of the QLR/MMSC/κBest test is slightly superior. Nevertheless, this test is not the recommended test for reasons given below.

We experimented with several smooth versions of the $\varphi^{(1)}$ critical value function, viz. functions of the form $\varphi^{(2)}$, in conjunction with the QLR statistic. We were not able to

Table II. Asymptotic Power Comparisons (Size-Corrected): MMM, Max, SumMax, & QLR Statistics, & PA, t-Test, $\varphi^{(3)}$, $\varphi^{(4)}$, & MMSC Critical Values with κ=Best[1]

	Crit.	Tuning	$p=10$			$p=4$			$p=2$		
Stat.	Val.	Par. κ	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
MMM	PA	-	.04	.36	.36	.20	.52	.46	.48	.62	.59
MMM	t-Test	Best	.19	.67	.79	.32	.69	.77	.51	.69	.71
Max	PA	-	.18	.44	.72	.30	.55	.71	.48	.63	.66
Max	t-Test	Best	.25	.59	.82	.35	.66	.79	.51	.69	.72
SumMax	PA	-	.10	.43	.64	.20	.54	.60	.48	.62	.59
SumMax	t-Test	Best	.14	.55	.71	.24	.64	.65	.51	.69	.71
GEL	Const.	-	.19	.18	.12	.44	.42	.39	.52	.54	.54
QLR	PA	-	.28	.36	.70	.44	.52	.71	.58	.62	.65
QLR	***t*-Test**	**Best**	**.59**	**.67**	**.82**	**.62**	**.69**	**.78**	**.65**	**.69**	**.72**
QLR	t-Test	Auto	.58	.67	.82	.62	.69	.78	.65	.69	.72
QLR	$\varphi^{(3)}$	Best	.49†	.62*	.83†	.54*	.67*	.78*	.60*	.67*	.72*
QLR	$\varphi^{(4)}$	Best	.59†	.67*	.82†	.62*	.69*	.78*	.65*	.69*	.72*
QLR	MMSC	Best	.65	.67	.78	.65	.69	.78	.65	.69	.72
Power	Envelope	-	.85	.85	.85	.80	.80	.80	.75	.75	.75

[1] κ=Best denotes the κ value that maximizes average asymptotic power.

*Results are based on (5000, 5000) size-correction and power repetitions.

†Results are based on (2000, 2000) size-correction and power repetitions.

find any that improved upon the average asymptotic power of the QLR/t-Test/κBest test. Some were inferior. All such tests have substantial disadvantages relative to the QLR/t-test in terms of the computational ease of determining suitable data-dependent κ and η values, as explained below.

In conclusion, we find that the best (S, φ) choices in terms of average asymptotic power (based on κ=Best) are, in order: QLR/MMSC, QLR/t-Test, and QLR/$\varphi^{(4)}$. Each of these tests out-performs the PA tests and "pure GEL" tests by a wide margin in terms of asymptotic power. Although the QLR/MMSC test is slightly better than the QLR/t-Test in terms of average asymptotic power, it has the following drawbacks: (i) its computation time is very high when p is large, such as $p = 10$, and is prohibitive for $p \geq 15$, because the QLR test statistic must be computed for all possible combinations of selected moment vectors, (ii) the best κ value varies widely with Ω and p, which makes it quite difficult to specify a data-dependent κ value that performs well, and (iii) the power differences between the QLR/MMSC and QLR/t-Test tests are relatively small and the latter test does not suffer from the aforementioned drawbacks.

Similarly, the QLR/$\varphi^{(2)}$ and QLR/$\varphi^{(4)}$ tests have a substantial drawback relative to the QLR/MMSC and QLR/t-Test tests. The latter two tests are pure moment selection tests and have the feature that a moment condition is either included or not included when computing the critical value. In consequence, for any given p and Ω combination, only a finite number of different critical values are possible—each one corresponding to a different combination of selected moments. This allows one to simulate these critical values initially once and then simulate the size or power of the test using these critical values in each size/power simulation repetition. If R repetitions are used for both critical values and size/power, then $2R$ simulations are required for these tests. On the other hand, the QLR/$\varphi^{(2)}$ and QLR/$\varphi^{(4)}$ tests are not pure moment selection tests. One has to simulate the critical value separately for each repetition in a size or power calculation, which requires R^2 simulation repetitions.

When developing a data-dependent method of selecting κ and computing asymptotic size-correction values η, one needs to simulate asymptotic size and power for a very large number of cases and, hence, computational speed is very important. To obtain accurate results (especially accurate size results), a large number of simulation repetitions is desirable. This is possible with pure moment selection tests, but not with the QLR/$\varphi^{(2)}$ and QLR/$\varphi^{(4)}$ tests.

Based primarily on the power results discussed above and secondarily on the computational factors, we take the QLR/t-Test to be the recommended test and we develop data-dependent $\widehat{\kappa}$ and $\widehat{\eta}$ for this test in Section 6.3.

We conclude this section by discussing the asymptotic power envelope and the asymptotic size of the RMS tests in Table II. The last row of Table II gives values of the asymptotic power envelope. The table shows that the QLR/t-Test/κBest test is quite close to the power envelope when $\Omega = \Omega_{Pos}$. This is remarkable because the testing problem is one in which the alternative hypothesis is multi-directional. In general, with multi-directional alternatives, one does not expect a test that is designed to have power in all directions of interest to be close to the power envelope (which is determined by a uni-directional test). For $\Omega = \Omega_{Neg}, \Omega_{Zero}$, the difference between the power of the QLR/t-Test/κBest test and the power envelope is fairly substantial, especially for Ω_{Neg}, and the amount is increasing in p. Note that for all Ω matrices, the power differences are noticeably smaller than the differences between asymptotic power of the F test and the asymptotic power envelope (for its testing problem) reported in Section 5.4 above.

Asymptotic size results for the RMS tests in Table II are given in Table B-II in Appendix B in the Supplement. The size results are for the case where κ=Best and $\eta = 0$. The size results show that the QLR/t-Test and QLR/MMSC tests with κ=Best have size close to the nominal level .05. For example, the QLR/t-Test/κ=Best test has size between .051 and .057 for all values of p and Ω considered. Thus, if one uses the optimal value of κ in terms of power, then the amount of asymptotic size-correction that is needed is small for these two tests. On the other hand, the sizes of the SumMax/t-Test and QLR/$\varphi^{(3)}$ tests are quite poor when κ=Best for Ω_{Neg} and Ω_{Pos}. For example, for $p = 10$ and $\Omega = \Omega_{Neg}$, these tests have size .17 and .10, respectively.

6.3 Approximately Optimal $\kappa(\Omega)$ and $\eta(\Omega)$ Functions

6.3.1 Definitions of $\kappa(\Omega)$ and $\eta(\Omega)$

In this section, we describe how the recommended $\kappa(\Omega)$ and $\eta(\Omega)$ functions defined in Section 2.2 are determined. These functions are for use with the QLR/t-Test test.

First, for $p = 2$ and given $\rho \in (-1, 1)$, where ρ denotes the correlation that appears in Ω, we compute numerically the values of κ that maximizes the average asymptotic (size-corrected) power of the nominal .05 QLR/t-Test test over a fine grid of 31 κ values.[29] We do this for each ρ in a fine grid of 43 values. Because the power results are size-corrected, a by-product of determining the best κ value for each ρ value is the size-correction value

[29] The grid of 31 κ values is $\{0, .2, .4, .6, .8, 1.0, 1.1, 1.2, ...,2.9, 3.0, 3.2, ..., 3.8, 4.0\}$. The grid of 43 δ values is $\{.99, .975, .95, .90, .85, ..., -.90, -.95, -.975, -.99\}$. The results are based on 40,000 critical value repetitions and 40,000 size and power repetitions. Size-corrrection is done for the given value of ρ, not uniformly over $\rho \in [-1, 1]$, because ρ can be consistently estimated and hence is known asymptotically

η that yields asymptotically correct size for each ρ.[30]

Second, by a combination of intuition and the analysis of numerical results, we postulate that for $p \geq 3$ the optimal function $\kappa^*(\Omega)$ defined in Section 5.3 is well approximated by a function that depends on Ω only through the $[-1, 1]$-valued function $\delta(\Omega)$ defined in (2.11).

The explanation for this is as follows: (i) Given Ω, the value $\kappa^*(\Omega)$ that yields maximum average asymptotic power is such that the size-correction value $\eta^*(\Omega)$ is not very large. (This is established numerically for a variety of p and Ω.) The reason is that the larger is $\eta^*(\Omega)$, the closer is the test to the PA test and the lower is the power of the test for μ vectors that have less than p elements negative. (ii) The size-correction value $\eta^*(\Omega)$ is small if the rejection probability at the least-favorable null vector μ is close to α when using the size-correction factor $\eta(\Omega) = 0$. (This is self-evident.) (iii) We postulate that null vectors μ that have two elements equal to zero and the rest equal to infinity are nearly least-favorable null vectors. If true, then the size of the QLR/t-Test test depends on the two-dimensional sub-matrices of Ω that are the correlation matrices that correspond to the cases where only two moment conditions appear. (iv) The size of a test for given κ and $p = 2$ is decreasing in the correlation ρ. In consequence, the least-favorable two-dimensional sub-matrix of Ω is the one with the smallest correlation. Hence, the value of κ that makes the size of the test equal to α for a small value of η is (approximately) a function of Ω through $\delta(\Omega)$ defined in (2.11). Note that this is just a heuristic explanation. It is not intended to be a proof.

Next, because $\delta(\Omega)$ corresponds to a particular 2 by 2 submatrix of Ω with correlation δ $(= \delta(\Omega))$, we take $\kappa(\Omega)$ to be the value that maximizes average asymptotic power when $p = 2$ and $\rho = \delta$, as specified in Table I and described in the second paragraph of this section.[31] We take $\eta(\Omega)$ to be the value determined by $p = 2$ and δ, i.e., $\eta_1(\delta)$ in (2.12)

[30] The asymptotic size of the QLR/tTest for given κ is found numerically to be decreasing in ρ for $\rho \in [-1, 1]$. Hence, for $\rho \in [a_1, a_2)$, we take η to be the size-correction value that yields correct asymptotic size for $\rho = a_1$.

[31] For $\rho \in [-.8, 1.0]$, we use the κ values that maximize average asymptotic power for $p = 2$ as the automatic κ values. For $\rho \in [-1.0, -.8)$, however, we use somewhat larger κ values than the ones that maximize average power. The reason is as follows. Numerical results show that the best κ values (in terms of power) for $\rho \in [-1.0, -.85]$ (and $p = 2$) are somewhat smaller than for $\rho = -.80$. Thus, there is a small deviation from the feature that the best κ value is monotone decreasing in ρ. When using the κ values for $p = 2$ with $p = 4, 10$, numerical results show that imposing monotonicity of κ in ρ yields better results for $p = 4$ in the sense that a smaller value $\eta_2(p)$ is needed for size-correction (which leads to higher power over the entire range of δ values). For this reason, we define $\kappa(\delta)$ in Table I to take values for $\delta \in [-1.0, -.80)$ that are slightly larger than the power maximizing values. The resultant loss in power for $p = 2$ is small, being around .01 for $\delta \in [-1.0, -.80)$.

and Table I, but allow for an adjustment that depends on p, viz., $\eta_2(p)$, that is defined to guarantee that the test has correct asymptotic significance level (up to numerical error).[32] In particular, $\eta_1(\delta) \in R$ is defined to be such that

$$\inf_{h_1 \in R^2_{+,\infty}} CP(h_1, \Omega_\delta, \eta_1(\delta)) = 1 - \alpha, \tag{6.3}$$

where Ω_δ is the 2 by 2 correlation matrix with correlation δ (and $\kappa(\Omega)$ that appears in the definition of $CP(h_1, \Omega, \eta)$ in (4.11) is as just defined). The numerical calculation of $\eta_1(\delta)$ is described above in the second paragraph of this section. Next, $\eta_2(p) \in R$ is defined to be such that

$$\inf_{h_1 \in R^p_{+,\infty}, \Omega \in \Psi} CP(h_1, \Omega, \eta_1(\delta(\Omega)) + \eta_2(p)) = 1 - \alpha, \tag{6.4}$$

where $\kappa(\Omega)$ and $\eta_1(\delta(\Omega))$ are defined as described above. The numerical calculation of $\eta_2(p)$ is described in Appendix C in the Supplement.

6.3.2 Automatic κ Power Assessment

We now examine numerically how well the proposed method does in approximating the best κ, viz., $\kappa^*(\Omega)$. We provide three groups of results and consider $p = 2, 4, 10$ for each group. The first group consists of the three Ω matrices considered in Table II. The second group consists of a fixed set of 19 Ω matrices chosen such that $\delta(\Omega)$ takes values on a grid in $[-.99, .99]$, see Appendix C in the Supplement for details. The third group considers 500 randomly generated Ω matrices for $p = 2, 4$ and 250 randomly generated Ω matrices for $p = 10$, see Appendix C in the Supplement for details. The asymptotic power results are size-corrected, are based on (40000, 40000) size-correction and power simulation repetitions for $p = 2, 4$ and (3000, 3000) simulations for $p = 10$. Average power is computed for μ vectors that consist of linear combinations of the μ vectors defined above in (6.1)-(6.2) of Section 6 and Appendix C in the Supplement, see Appendix C in the Supplement for details. In all three groups, we assess the proposed method of selecting κ, referred to as the κAuto method, by comparing the average asymptotic power of the κAuto test with the corresponding κBest test, whose κ value

[32] One could define $\eta(\Omega)$ to depend separately on $\delta(\Omega)$ and p, say $\eta(\Omega) = \overline{\eta}(\delta(\Omega), p)$ for some function $\overline{\eta}$. This would yield a much more complicated function $\eta(\Omega)$ than the function $\eta(\Omega) = \eta_1(\delta(\Omega)) + \eta_2(p)$ that we use. Numerical results indicate that more complicated functions $\overline{\eta}$ are not needed. The simple function that we use works quite well.

is determined numerically to maximize average asymptotic power.

The rows of Table III for the QLR/t-Test/κBest and QLR/t-Test/κAuto tests show that the κAuto method works very well. It has the same average asymptotic power as the QLR/t-Test/κBest test for all p and Ω values except one and in this one case the difference is just .01.

The results for the 19 Ω matrices are given in Table III. These results also show that the κAuto method works very well. There is very little difference between the average asymptotic power of the QLR/t-Test/κAuto and QLR/t-Test/κBest tests. Only in a few scenarios is a difference of .01 or more detected.

Table III. Asymptotic Power Differences Between QLR/t-Test/κAuto and QLR/t-Test/κBest Tests for Nominal Level .05 Size-Corrected Tests

δ	-.99	-.975	-.95	-.9	-.8	-.7	-.6	-.5	-.4	-.2
p=2	.022	.016	.009	.003	.000	.000	.000	.000	.000	.000
p=4	.007	.004	.003	.003	.002	.003	.000	.003	.003	.000
p=10	.003	.006	.006	.008	.004	.009	.001	.005	.002	.002

δ	.0	.2	.4	.6	.8	.9	.95	.975	.99
p=2	.000	.000	.000	.000	.000	.000	.000	.000	.000
p=4	.001	.001	.000	.000	.000	.000	.000	.000	.000
p=10	.003	.004	.002	.003	.000	.000	.000	.000	.000

The results for the randomly generated Ω matrices are similarly good for the κAuto method. For $p = 2$, across the 500 Ω matrices, the average power differences have average equal to .0023, standard deviation equal to .0059, and range equal to $[.000, .026]$. For $p = 4$, across the 500 Ω matrices, the average power difference is .0018, the standard deviation is .0022, and the range is $[.000, .012]$. For $p = 10$, across the 250 Ω matrices, the average power differences have average equal to .0148, standard deviation equal to .0060, and range equal to $[.000, .036]$.

In conclusion, the κAuto method performs very well in terms of selecting κ values that maximize the average asymptotic power.

7 Finite Sample Results

The recommended RMS test, QLR/t-Test/κAuto, can be implemented in finite samples via the "asymptotic normal" and the bootstrap versions of the t-Test/κAuto critical value. In this section we determine which of these two methods performs better in finite samples. We also assess how well these tests perform in finite samples in an absolute sense. In short, we find that the bootstrap version (denoted Boot) performs better than the asymptotic normal version (denoted Norm) in terms of the closeness of its null rejection probabilities to its nominal level and in terms of its power. The Boot test is found to perform quite well in that its null rejection probabilities are close to its nominal level and the difference between its finite-sample and asymptotic power is relatively small.

We provide results for sample size $n = 100$. We consider the same correlation matrices Ω_{Neg}, Ω_{Zero}, and Ω_{Pos} as above and the same numbers of moment inequalities $p = 2$, 4, and 10. We take the mean zero variance I_p random vector $Z^\dagger = Var^{-1/2}(m(W_i,\theta))(m(W_i,\theta) - Em(W_i,\theta))$ to be i.i.d. across elements and consider six distributions for the elements: standard normal (i.e., N(0, 1)), t_5, t_3, t_2, uniform, and chi-squared with three degrees of freedom χ^2_3. All of these distributions are centered and scaled to have mean zero and variance one except the t_2, whose variance is infinite. The t distributions have thick tails, the uniform has thin tails, and the χ^2_3 is skewed. (The t_3 and t_2 distributions may not be of much practical interest because their tails are extremely thick, but they are included as extreme cases.) We use (5000, 5000) critical value and rejection probability repetitions.

We note that the finite-sample testing problem for *any* moment inequality model fits into the framework above for some correlation matrix Ω and some distribution of $Z^\dagger$. Hence, the finite-sample results given here provide a level of generality that usually is lacking with finite-sample simulation results.

Table IV provides the finite-sample maximum null rejection probabilities (MNRPs) of the nominal .05 Norm and Boot versions of the recommended RMS test based on the QLR statistic. The MNRP is the maximum rejection probability over mean vectors μ in the null hypothesis for a given correlation matrix Ω and a given distribution of $Z^\dagger$. Table V provides MNRP-corrected finite-sample average power for the same two tests. The average power results are for the same mean vectors μ in the alternative hypothesis as considered above for asymptotic power.

Table IV shows that for the normal, t_5, uniform, and χ^2_3 distributions, the Boot test

Table IV. Finite-Sample Maximum Null Rejection Probabilities (MNRPs) of the Nominal .05 QLR/t-Test/κAuto Test Based on Normal and Bootstrap-Based Critical Values

			$p=10$			$p=4$			$p=2$		
Test	Dist	n	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
Norm	N(0,1)	100	.071	.066	.045	.058	.058	.045	.044	.049	.052
Boot	N(0,1)	100	.044	.048	.043	.058	.055	.047	.050	.046	.051
Norm	t_5	100	.073	.069	.046	.053	.050	.047	.048	.050	.049
Boot	t_5	100	.050	.051	.050	.052	.051	.051	.053	.053	.052
Norm	t_3	100	.071	.069	.047	.057	.052	.051	.048	.055	.054
Boot	t_3	100	.052	.056	.053	.064	.060	.063	.066	.063	.065
Norm	t_2	100	.056	.057	.037	.045	.044	.042	.040	.041	.043
Boot	t_2	100	.056	.055	.058	.072	.067	.072	.073	.066	.072
Norm	Uniform	100	.075	.069	.045	.055	.049	.045	.048	.049	.046
Boot	Uniform	100	.046	.048	.041	.047	.047	.045	.047	.046	.044
Norm	χ^2_3	100	.143	.146	.067	.091	.096	.065	.074	.083	.078
Boot	χ^2_3	100	.052	.054	.045	.054	.055	.046	.053	.052	.053

performs very well with MNRPs in the range of $[.041, .058]$. For the t_3 and t_2 distributions, its MNRPs are in the ranges of $[.052, .066]$ and $[.055, .077]$, respectively, which is quite good considering how thick the tails are of these distributions. (Note that the asymptotic results given above do not hold for the t_2 distribution because its variance is infinite.)

In contrast, the Norm test over-rejects somewhat in some cases even for the normal distribution for which its MNRPs are in the range of $[.044, .071]$. For the thick- and thin-tailed distributions (t_5, t_3, t_2, and uniform), the MNRPs of the Norm test are in the range $[.037, .075]$, which is similar to those for the normal distribution. However, with the skewed distribution, χ^2_3, the Norm test over-rejects the null hypothesis substantially, especially with the Ω_{Neg} and Ω_{Zero} matrices. Its MNRPs are in the range $[.067, .147]$ for the χ^2_3 distribution. It should not be too surprising that skewed distributions cause the most problems for the Norm test because the first term in the Edgeworth expansion of a sample average is a skewness term and the statistics considered here are simple functions of sample averages.

The results show that the bootstrap is able to detect skewness of the underlying distributions and hence the Boot test does not over-reject in the presence of skewness. Note that this occurs even though the statistics considered are not asymptotically pivotal (which implies that the bootstrap does not provide higher-order asymptotic improvements over standard asymptotic approximations).

We conclude that the Boot version of the recommended test noticeably out-performs the Norm version in terms of its properties under the null hypothesis.

Table V shows that the Boot test has superior finite-sample average power compared to the Norm test for the N(0, 1), t_5, uniform, and χ^2_3 distributions, especially for $p = 10$ with Ω_{Neg} and Ω_{Zero}. The differences are largest with the uniform and χ^2_3 distributions. The superior performance of the Boot test occurs in the cases in which the Norm test over-rejects under the null hypothesis. The reason is that over-rejection leads to an increase in the critical value for the Norm test given that the power results are MNRP-corrected. With the t_3 and t_2 distributions, the Norm test has slightly higher power than the Boot test, but this result is mitigated by (i) the fact that both distributions are quite extreme in terms of tail thickness and (ii) the power of both tests for the t_2 distribution is very low.

For comparative purposes, Table V also provides finite-sample results for the QLR /PA test. These results indicate that the asymptotic dominance of moment selection-based critical values over PA-based critical values also is apparent in finite samples.

Recall that GEL statistics have the same asymptotic distribution as the QLR statistic. Hence, the recommended RMS test, QLR/t-Test/κAuto, also can be implemented in finite samples by combining GEL statistics with Norm and Boot versions of the t-Test/κAuto critical value. We do not report any results for such tests here for several reasons. First, with normally-distributed moment functions, the difference between the finite-sample and asymptotic properties of the tests is due solely to the estimation of the variance matrix. Hence, the only way in which the GEL statistic can out-perform the QLR statistic is by providing a better estimator of the variance matrix. However, we find that the results for the QLR-based Boot test vary very little between the

Table V. Finite-Sample ("Size-Corrected") Power of the Nominal .05 QLR/PA and QLR/t-Test/κAuto Tests Based on Normal and Bootstrap-Based Critical Values

			$p=10$			$p=4$			$p=2$		
Test	Dist	n	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}	Ω_{Neg}	Ω_{Zero}	Ω_{Pos}
QLR/PA	N(0,1)	100	.31	.39	.69	.45	.53	.69	.57	.63	.66
κAuto/Norm			.51	.61	.81	.58	.66	.77	.65	.69	.71
κAuto/Boot			.56	.67	.82	.59	.67	.77	.65	.71	.72
Power Envel.			.85	.85	.84	.79	.78	.77	.75	.74	.74
QLR/PA	t_5	100	.32	.40	.69	.45	.53	.69	.57	.62	.65
κAuto/Norm			.50	.61	.80	.61	.69	.77	.64	.68	.71
κAuto/Boot			.54	.65	.78	.60	.68	.76	.64	.68	.71
QLR/PA	t_3	100	.42	.50	.77	.54	.61	.76	.64	.68	.70
κAuto/Norm			.61	.72	.85	.67	.75	.81	.71	.73	.75
κAuto/Boot			.60	.71	.81	.63	.71	.77	.66	.69	.72
QLR/PA	t_2	100	.05	.07	.19	.08	.12	.20	.15	.18	.19
κAuto/Norm			.09	.14	.26	.14	.20	.25	.19	.22	.23
κAuto/Boot			.06	.13	.23	.09	.18	.23	.16	.21	.23
QLR/PA	Uniform	100	.30	.39	.70	.45	.51	.68	.55	.61	.64
κAuto/Norm			.49	.60	.73	.59	.68	.78	.62	.67	.71
κAuto/Boot			.55	.67	.82	.60	.69	.78	.63	.69	.73
QLR/PA	χ^2_3	100	.40	.48	.66	.49	.56	.69	.59	.63	.65
κAuto/Norm			.38	.44	.70	.50	.55	.70	.58	.58	.61
κAuto/Boot			.49	.56	.71	.54	.60	.71	.58	.60	.64

case of known and unknown variance matrix. In consequence, there is little room for GEL-based tests to provide improvements in terms of MNRP or average power. Second, GEL-based tests have an enormous disadvantage in terms of computation compared to QLR-based tests. To compute a confidence set using an RMS procedure one needs to compute the test statistic hundreds of thousands of times. For example, to determine whether a single point is in the confidence set one needs to simulate the critical value once which requires, say, 10,000 statistic evaluations. For the QLR statistic it is fast to do so because the QLR statistic is the solution to a quadratic programming problem which is very well behaved. On the other hand, GEL statistics require the solution to a general nonlinear optimization problem which is much slower. Third, Canay (2007) provides some finite-sample simulation results for GEL statistics and does not find any power advantages for them.

References

Andrews, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–564.

Andrews, D. W. K., S. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," unpublished manuscript, Cowles Foundation, Yale University.

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, forthcoming.

Andrews, D. W. K. and S. Han (2009): "Invalidity of the Bootstrap and the m Out of n Bootstrap for Interval Endpoints Defined by Moment Inequalities," *Econometrics Journal,* forthcoming. Also available as Cowles Foundation Discussion Paper No. 1671, Yale University.

Andrews, D. W. K. and P. Jia (2008): "Supplement to 'Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure'," unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2008): "Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics,* forthcoming.

Andrews, D. W. K. and G. Soares (2007): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," Cowles Foundation Discussion Paper No. 1631, Yale University.

Beresteanu, A., I. Molchanov, and F. Molinari (2008): "Sharp Identification Regions in Games," CEMMAP Working Paper CWP15/08, Institute for Fiscal Studies, UCL.

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica,* 76, 763-814.

Bontemps, C., T. Magnac, and E. Maurin (2007): "Set Identified Linear Models," unpublished manuscript, Toulouse School of Economics.

Bugni, F. (2007a): "Bootstrap Inference in Partially Identified Models," unpublished manuscript, Department of Economics, Northwestern University.

——— (2007b): "Bootstrap Inference in Partially Identified Models: Pointwise Construction," unpublished manuscript, Department of Economics, Northwestern University.

Canay, I. A. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," unpublished manuscript, Department of Economics, University of Wisconsin.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica,* 75, 1243-1284.

Ciliberto, F. and E. Tamer (2003): "Market Structure and Market Equilibrium in Airline Markets," unpublished manuscript, Department of Economics, Princeton University.

Galichon, A. and M. Henry (2008): "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics,* forthcoming.

Guggenberger, P., J. Hahn, and K. Kim (2008): "Specification Testing Under Moment Inequalities," *Economics Letters,* 99, 375-378.

Imbens, G. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica,* 72, 1845-1857.

Kudo, A. (1963): "A Multivariate Analog of a One-Sided Test," *Biometrika,* 50, 403-418.

Manski, C. F. and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica,* 70, 519-546.

Moon, H. R. and F. Schorfheide (2006): "Boosting Your Instruments: Estimation with Overidentifying Inequality Moment Conditions," unpublished working paper, Department of Economics, University of Southern California.

——— (2007): "A Bayesian Look at Partially–Identified Models," unpublished manuscript, Department of Economics, University of Pennsylvania.

Müller, U. and M. Watson (2007): "Low-Frequency Robust Cointegration Testing," *Econometrica,* 76, 979-1016.

Otsu, T. (2006): "Large Deviation Optimal Inference for Set Identified Moment Inequality Models," unpublished manuscript, Cowles Foundation, Yale University.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2004): "Applications of Moment Inequalities," unpublished working paper, Department of Economics, Harvard University.

Pratt, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association,* 56, 541-567.

Romano, J. P. and A. M. Shaikh (2005): "Inference for the Identified Set in Partially Identified Econometric Models," unpublished working paper, Department of Economics, University of Chicago.

——— (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Inference and Planning (Special Issue in Honor of T. W. Anderson),* forthcoming.

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

Silvapulle, M. J. and P. K. Sen (2005): *Constrained Statistical Inference.* New York: Wiley.

Soares, G. (2005): "Inference with Inequality Moment Constraints," unpublished working paper, Department of Economics, Yale University.

Stoye, J. (2007): "More on Confidence Intervals for Partially Identified Parameters," unpublished manuscript, Department of Economics, New York University.

Woutersen, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters," unpublished manuscript, Department of Economics, Johns Hopkins University.