Contributions to the Theory of Optimal Tests

Humberto Moreira and Marcelo J. Moreira
FGV/EPGE

This version: November 13, 2013

Abstract

This paper considers tests which maximize the weighted average power (WAP). The focus is on determining WAP tests subject to an uncountable number of equalities and/or inequalities. The unifying theory allows us to obtain tests with correct size, similar tests, and unbiased tests, among others.

A WAP test may be randomized and its characterization is not always possible. We show how to approximate the power of the optimal test by sequences of nonrandomized tests. Two alternative approximations are considered. The first approach considers a sequence of similar tests for an increasing number of boundary conditions. This discretization allows us to implement the WAP tests in practice. The second method finds a sequence of tests which approximate the WAP test uniformly. This approximation allows us to show that WAP similar tests are admissible.

The theoretical framework is readily applicable to several econometric models, including the important class of the curved-exponential family. In this paper, we consider the instrumental variable model with heteroskedastic and autocorrelated errors (HAC-IV) and the nearly integrated regressor model. In both models, we find WAP similar and (locally) unbiased tests which dominate other available tests.

1 Introduction

When making inference on parameters in econometric models, we rely on the classical hypothesis testing theory and specify a null hypothesis and alternative hypothesis. Following the Neyman-Pearson framework, we control size at some level α and seek to maximize power. Applied researchers often use the t-test, which can be motivated by asymptotic optimality. It is now understood that these asymptotic approximations may not be very reliable in practice. Two examples in which the existing theory fails are models in which parameters are weakly identified, e.g., Dufour (1997) and Staiger and Stock (1997); or when variables are highly persistent, e.g., Chan and Wei (1987) and Phillips (1987).

This paper aims to obtain finite-sample optimality and derive tests which maximize weighted average power (WAP). Consider a family of probability measures $\{P_v; v \in \mathbb{V}\}$ with densities f_v. For testing a null hypothesis $H_0 : v \in \mathbb{V}_0$ against an alternative hypothesis $H_1 : v \in \mathbb{V}_1$, we seek to decide which one is correct based on the available data. When the alternative hypothesis is composite, a commonly used device is to reduce the composite alternative to a simple one by choosing a weighting function Λ_1 and maximizing WAP. When the null hypothesis is also composite, we could proceed as above and determine a weight Λ_0 for the null. It follows from the Neyman-Pearson lemma that the optimal test rejects the null when $\int f_v \Lambda_1 (dv) / \int f_v \Lambda_0 (dv)$ is large. A particular choice of Λ_0, the least favorable distribution, yields a test with correct size α. Although this test is most powerful for the weight Λ_1, it can be biased and have bad power for many values of $v \in \mathbb{V}_1$.

An alternative strategy to the least-favorable-distribution approach is to obtain optimality results within a smaller class of procedures. For example, any unbiased test must be similar on $\overline{\mathbb{V}_0} \cap \overline{\mathbb{V}_1}$ by continuity of the power function. If the sufficient statistic for the submodel $v \in \mathbb{V}_0$ is complete, then all similar tests must be conditionally similar on the sufficient statistic. This is the theory behind the uniformly most powerful unbiased (UMPU) tests for multiparameter exponential models, e.g., Lehmann and Romano (2005). A caveat is that this theory does not hold true for many econometric models. Hence the need to develop a unifying optimality framework that encompasses tests with correct size, similar tests, and unbiased tests, among others. The theory is a simple generalization of the existing theory of WAP tests with correct size. This allows us to build on and connect to the existing literature

on WAP tests with correct size; e.g., Andrews, Moreira, and Stock (2008) and Müller and Watson (2013), among others.

We seek to find WAP tests subject to an uncountable number of equalities and/or inequalities. In practice, it may be difficult to implement the WAP test. We propose two different approximations. The first method finds a sequence of WAP tests for an increasing number of boundary conditions. We provide a pathological example which shows that the discrete approximation works even when the final test is randomized. The second approximation relaxes all equality constraints to inequalities. It allows us to show that WAP similar tests are admissible for an important class of econometric models (whether the sufficient statistic for the submodel $v \in V_0$ is complete or not). Both approximations are in finite samples only. In a companion paper, Moreira and Moreira (2011) extend the finite-sample theory to asymptotic approximations using limit of experiments. In the supplement, we also present an approximation in Hilbert spaces.

We apply our theory to find WAP tests in the weak instrumental variable (IV) model with heteroskedastic and autocorrelated (HAC) errors and the nearly integrated regressor model.

In the HAC-IV model, we obtain WAP unbiased tests based on two weighted average densities denoted the MM1 and MM2 statistics. We derive a locally unbiased (LU) condition from the power behavior near the null hypothesis. We implement both WAP-LU tests based on MM1 and MM2 using a nonlinear optimization package. Both WAP-LU tests are admissible and dominate both the Anderson and Rubin (1949) and score tests in numerical simulations. We derive a second condition for tests to be unbiased, the strongly unbiased (SU) condition. We implement the WAP-SU tests based on MM1 and MM2 using a conditional linear programming algorithm. The WAP-SU tests are easy to implement and have power close to the WAP-LU tests. We recommend the use of WAP-SU tests in empirical work.

In the nearly integrated regressor model, we find a WAP-LU (locally unbiased) test based on a two-sided, weighted average density (the MM-2S statistic). We show that the WAP-LU test must be similar (at the frontier between the null and alternative) and uncorrelated with another statistic. We approximate these two constraints to obtain the WAP-LU test using a linear programming algorithm. We compare the WAP-LU test with the similar L_2 test of Wright (2000) and a WAP test (with correct size) and a WAP similar test. The L_2 test is biased when the regressor is stationary, while the WAP size-corrected and WAP similar tests are biased when the

regressor is persistent. By construnction, the WAP-LU test does not suffer these difficulties. Hence, we recommend the WAP-LU test based on the MM-2S statistic for two-sided testing. In the supplement, we also propose a one-sided WAP test based on a one-sided (MM-1S) statistic.

The remainder of this paper is organized as follows. In Section 2, we discuss the power maximization problem. In Section 3, we present a version of the Generalized Neyman-Pearson (GNP) lemma when the number of boundary conditions is finite. By suitably increasing the number of boundary conditions, the tests based on discretization approximate the power function of the optimal similar test. We show how to implement these tests by a simulation method. In Section 4, we derive tests that are approximately similar in a uniform sense. We establish sufficient conditions for these tests to be nonrandomized. By decreasing the slackness condition, we approximate the power function of the optimal similar test. In Section 5, we present power plots for both HAC-IV and nearly integrated regressor models. In Section 6, we conclude. In Section 7, we provide proofs for all results. In the supplement to this paper, we provide an approximation in Hilbert spaces, all details for implementing WAP tests, and additional numerical simulations.

2 Weighted Average Power Tests

Consider a family of probability measures $\mathcal{P} = \{P_v; v \in \mathbb{V}\}$ on a measurable space $(Y, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-algebra. We assume that all probabilities P_v are dominated by a common σ-finite measure. By the Radon-Nikodym Theorem, these probability measures admit densities f_v.

Classical testing theory specifies a null hypothesis $H_0 : v \in \mathbb{V}_0$ against an alternative hypothesis $H_1 : v \in \mathbb{V}_1$ and seeks to determine which one is correct based on the available data. A test is defined to be a measurable function $\phi(y)$ that is bounded by 0 and 1 for all values of $y \in Y$. For a given outcome y, the test rejects the null with probability $\phi(y)$ and accepts the null with probability $1-\phi(y)$. The test is said to be nonrandomized if ϕ only takes values 0 and 1; otherwise, it is called a randomized test. The goal is to find a test that maximizes power for a given size α.

If both hypotheses are simple, $\mathbb{V}_0 = \{v_0\}$ and $\mathbb{V}_1 = \{v_1\}$, the Neyman-Pearson lemma establishes necessary and sufficient conditions for a test to be most powerful among all tests with null rejection probability no greater than α. This test rejects the null hypothesis when the likelihood ratio f_{v_1}/f_{v_0} is

sufficiently large.

When the alternative hypothesis is composite, the optimal test may or may not depend on the choice of $v \in \mathbb{V}_1$. If it does not, this test is called uniformly most powerful (UMP) at level α, e.g., testing one-sided alternatives $H_1 : v > v_0$ in a one-parameter exponential family. If it does depend on $v \in \mathbb{V}_1$, a commonly used device is to reduce the composite alternative to a simple one by choosing a weighting function Λ_1 and maximizing a weighted average density:

$$\sup_{0 \leq \phi \leq 1} \int \phi h, \text{ where } \int \phi f_{v_0} \leq \alpha,$$

where $h = \int_{\mathbb{V}_1} f_v \Lambda_1 (dv)$ for some probability measure Λ_1 that weights different alternatives in $\mathbb{V}_1$. If we seek to maximize power for a particular alternative $v_1 \in \mathbb{V}_1$, the weight function Λ_1 is given by

$$\Lambda_1 (dv) = \begin{cases} 1 & \text{if } v = v_1 \\ 0 & \text{otherwise} \end{cases} .$$

When the null hypothesis is also composite, we can proceed as above and determine a weight Λ_0 for the null. It follows from the Neyman-Pearson lemma that the optimal test rejects the null when $\int f_v \Lambda_1 (dv) / \int f_v \Lambda_0 (dv)$ is large. For an arbitrary choice of Λ_0, the test does not necessarily have null rejection probability smaller than the significance level α for all values $v \in \mathbb{V}_0$. Only a particular choice of Λ_0, the least favorable distribution, yields a test with correct size α. Although this test is most powerful for Λ_1, it can have undesirable properties; e.g., be highly biased.

An alternative strategy to the least-favorable-distribution approach is to obtain optimality within a smaller class of tests. For example, if a test is unbiased and the power curve is continuous, the test must be similar on the frontier between the null and alternative; that is, $\overline{\mathbb{V}_0} \cap \overline{\mathbb{V}_1}$. If the sufficient statistic for the submodel $v \in \overline{\mathbb{V}_0} \cap \overline{\mathbb{V}_1}$ is complete, then all similar tests must be conditionally similar on the sufficient statistic. These tests are said to have the so-called Neyman structure. This is the theory behind the uniformly most powerful unbiased (UMPU) tests for multiparameter exponential models, e.g., Lehmann and Romano (2005).

In this paper, we consider weighted average power maximization problems encompassing size-corrected tests, similar tests, and locally unbiased tests, among others. Therefore, we seek weighted average power (WAP) tests which

maximize power subject to several constraints:

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \gamma_v^1 \leq \int \phi g_v \leq \gamma_v^2, \forall v \in \mathbf{V}, \tag{2.1}$$

where $\mathbf{V} \subset \mathbb{V}$, g_v is a measurable function mapping Y onto $\mathbb{R}^m$ and γ_v^i are measurable functions mapping $\mathbf{V}$ onto $\mathbb{R}^m$ with h and g_v integrable for each $v \in \mathbf{V}$ and $i = 1, 2$. We use $\gamma_v^1 \leq \gamma_v^2$ to denote that each coordinate of the vector γ_v^1 is smaller than or equal to the corresponding coordinate of the vector γ_v^2. The functions γ_v^1 and γ_v^2 have no *a priori* restrictions and can be equal in an uncountable number of points. The problem (2.1) allows us to seek: WAP size-corrected tests for

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } 0 \leq \int \phi f_v \leq \alpha \text{ for } v \in \mathbb{V}_0; \tag{2.2}$$

WAP similar tests defined by

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \int \phi f_v = \alpha \text{ for } v \in \mathbf{V} \tag{2.3}$$

(typically with $\mathbf{V} = \overline{\mathbb{V}_0} \cap \overline{\mathbb{V}_1}$); WAP unbiased tests given by

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \int \phi f_{v_0} \leq \alpha \leq \int \phi f_{v_1} \text{ for } v_0 \in \mathbb{V}_0 \text{ and } v_1 \in \mathbb{V}_1; \tag{2.4}$$

among other constraints.

Our theoretical framework builds on and connects with many different applications. In this paper, we consider three econometric examples to illustrate the WAP maximization problem given in problem (2.1). Example 1 briefly discusses a simple moment inequality model in light of our theory. Example 2 presents the weak instrumental variable (WIV) model. We revisit the one-sided (Example 2.1) and two-sided (Example 2.2) testing problems with homoskedastic errors. We develop new WAP unbiased tests for heteroskedastic and autocorrelated errors (Example 2.3). Finally, Example 3 introduces novel WAP similar and WAP unbiased tests for the nearly integrated regressor model.

We use Examples 2.2 and 3 as the running examples as we present our theoretical findings.

Example 1: Moment Inequalities

Consider a simple model

$$Y \sim N(v, I_2),$$

where $v = (v_1, v_2)'$. We want to test $H_0 : v \geq 0$ against $H_1 : v \ngeq 0$. The boundary between the null and alternative is $\mathbf{V} = \{v \in \mathbb{R}^2; v \geq 0 \ \& \ v_1 = 0$ or $v_2 = 0\}$. The density of Y at y is given by

$$\begin{aligned} f_v(y) &= (2\pi)^{-1} \exp\left(-\frac{1}{2}\|y - v\|^2\right) \\ &= C(v) \exp(v'y)\, \eta(y), \end{aligned}$$

where $C(v) = (2\pi)^{-1} \exp\left(-\frac{\|v\|}{2}\right)$ and $\eta(y) = \exp\left(-\frac{\|y\|}{2}\right)$.

Andrews (2012) shows that similar tests have poor power for some alternatives. The power function $E_v\phi(Y)$ of any test is analytic in $v \in \mathbb{V}$; see Theorem 2.7.1 of Lehmann and Romano (2005, p. 49). The test is similar at the frontier between H_0 and H_1 if

$$E_v\phi(Y) = \alpha, \forall v \in \mathbf{V}.$$

That is, $E_{v_1,0}\phi(Y) = E_{0,v_2}\phi(Y) = \alpha$ for $v_1, v_2 \geq 0$. Because the power function is analytic then $E_{v_1,0}\phi(Y) = E_{0,v_2}\phi(Y) = \alpha$ for every v_1, v_2 for any similar test. Hence, similar tests have power equal to size for alternatives $(v_1, 0)$ or $(0, v_2)$ where $v_1, v_2 < 0$. Andrews (2012) also notes that similar tests may not have trivial power. He indeed provides a constructive proof of similar tests where $E_v\phi(Y) > \alpha$ for $v_1, v_2 > 0$.

Although similar tests have poor power for certain alternatives, we show in Section 4.1 that WAP similar tests which solve

$$\sup_{0 \leq \phi \leq 1} \int \phi h, \text{ where } \int \phi f_v = \alpha, \forall v \in \mathbf{V} \text{ and } \int \phi f_v \leq \alpha, \forall v \in \mathbb{V}_0$$

are still admissible. By the Complete Class Theorem, we can find a weight Λ_1 for $h = \int_{\mathbb{V}_1} f_v \Lambda_1(dv)$ so that the WAP test which solves

$$\sup_{0 \leq \phi \leq 1} \int \phi h, \text{ where } \int \phi f_v \leq \alpha, \forall v \in \mathbb{V}_0,$$

is approximately similar. This procedure is, however, not likely to be preferable to existing non-similar tests such as likelihood ratio or Bayes tests; see Sections 3.8 and 8.6 of Silvapulle and Sen (2005) and Chiburis (2009). Hence, choosing the weight Λ_1 requires some care in empirical practice.

Example 2: Weak Instrumental Variables (WIVs)

Consider the instrumental variable model

$$\begin{aligned} y_1 &= y_2\beta + u \\ y_2 &= Z\pi + w_2, \end{aligned}$$

where y_1 and y_2 are $n \times 1$ vectors of observations on two endogenous variables, Z is an $n \times k$ matrix of nonrandom exogenous variables having full column rank, and u and w_2 are $n \times 1$ unobserved disturbance vectors having mean zero. We are interested in the parameter β, treating π as a nuisance parameter. We look at the reduced-form model for $Y = [y_1, y_2]$:

$$Y = Z\pi a' + W, \tag{2.5}$$

the $n \times 2$ matrix of errors W is assumed to be iid across rows with each row having a mean zero bivariate normal distribution with nonsingular covariance matrix Ω.

Example 2.1: One-Sided IV

We want to test $H_0 : \beta \leq \beta_0$ against $H_1 : \beta > \beta_0$. A $2k$-dimensional sufficient statistic for β and π is given by

$$\begin{aligned} S &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2} \text{ and} \\ T &= (Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}, \text{ where} \\ b_0 &= (1, -\beta_0)' \text{ and } a_0 = (\beta_0, 1)'. \end{aligned} \tag{2.6}$$

Andrews, Moreira, and Stock (2006a) suggest to focus on tests which are invariant to orthogonal transformations on $[S, T]$. Invariant tests depend on the data only through

$$Q = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} = \begin{bmatrix} S'S & S'T \\ S'T & T'T \end{bmatrix}.$$

The density of Q at q for the parameters β and $\lambda = \pi'Z'Z\pi$ is

$$\begin{aligned} f_{\beta,\lambda}(q_S, q_{ST}, q_T) &= \kappa_0 \exp(-\lambda(c_\beta^2 + d_\beta^2)/2) \det(q)^{(k-3)/2} \\ &\quad \times \exp(-(q_S + q_T)/2)(\lambda\xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}(\sqrt{\lambda\xi_\beta(q)}), \end{aligned}$$

where $\kappa_0^{-1} = 2^{(k+2)/2} pi^{1/2} \Gamma_{(k-1)/2}$, $pi = 3.1415...$, $\Gamma_{(\cdot)}$ is the gamma function, $I_{(k-2)/2}(\cdot)$ denotes the modified Bessel function of the first kind, and

$$\begin{aligned} \xi_\beta(q) &= c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T, \\ c_\beta &= (\beta - \beta_0) \cdot (b_0' \Omega b_0)^{-1/2}, \text{ and} \\ d_\beta &= a' \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2}. \end{aligned} \tag{2.7}$$

Imposing similarity is not enough to yield tests with correct size. For example, Mills, Moreira, and Vilela (2013) show that POIS (Point Optimal Invariant Similar) tests do not have correct size. We can try to find choices of weights which yield a WAP similar test with correct size. However, this requires clever choices of weights. We can also find tests which are similar at $\beta = \beta_0$ and which have correct size for $\beta \leq \beta_0$. Alternatively, we can require the power function to be monotonic:

$$\sup_{0 \leq \phi \leq 1} \int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} = \alpha \text{ and } \int \phi f_{\beta_1,\lambda} \leq \int \phi f_{\beta_2,\lambda}, \forall \beta_1 < \beta_2, \lambda, \tag{2.8}$$

where the integrals are with respect to q. Problem (2.8) implies the test has correct size and is unbiased. If the power function is differentiable (as with normal errors), we can obtain

$$\sup_{0 \leq \phi \leq 1} \int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} = \alpha \text{ and } \int \phi \frac{\partial \ln f_{\beta,\lambda}}{\partial \beta} f_{\beta,\lambda} \geq 0, \forall \beta, \lambda. \tag{2.9}$$

There are two boundary conditions in (2.9). Some constraints may preclude admissibility whereas others not. In Section 4.1, we show that WAP similar (or unbiased) tests are admissible. On the other hand, the WAP test which solves (2.9) may not be admissible because the power function must be monotonic (although this does not seem a serious issue for one-sided testing for WIVs vis-à-vis the numerical findings of Mills, Moreira, and Vilela (2013)).

Example 2.2: Two-Sided IV

We want to test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$. An optimal WAP test which solves

$$\sup_{0 \leq \phi \leq 1} \int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} \leq \alpha, \forall \lambda \tag{2.10}$$

can be biased. We impose corrected size and unbiased conditions into the maximization problem:

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} \leq \alpha \leq \int \phi f_{\beta,\lambda}, \forall \beta \neq \beta_0, \lambda. \tag{2.11}$$

The first inequality implies that the test has correct size at level α. The second inequality implies that the test is unbiased. Because the power function is continuous, the test must be similar at $\beta = \beta_0$. The problem (2.11) is then equivalent to

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} = \alpha \leq \int \phi f_{\beta,\lambda}, \forall \beta \neq \beta_0, \lambda. \tag{2.12}$$

In practice, it is easier to require the test to be locally unbiased; that is, the power function derivative at $\beta = \beta_0$ equals zero:

$$\frac{\partial}{\partial\beta}\int \phi f_{\beta,\lambda}\bigg|_{\beta=\beta_0} = \int \phi \left.\frac{\partial \ln f_{\beta,\lambda}}{\partial\beta}\right|_{\beta=\beta_0} f_{\beta_0,\lambda} = 0.$$

The WAP locally unbiased test solves

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} = \alpha \text{ and } \int \phi \left.\frac{\partial \ln f_{\beta,\lambda}}{\partial\beta}\right|_{\beta=\beta_0} f_{\beta_0,\lambda} = 0, \forall\lambda. \tag{2.13}$$

Andrews, Moreira, and Stock (2006b) show that the optimization problem in (2.13) simplifies to

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} = \alpha \text{ and } \int \phi . q_{ST} f_{\beta_0,\lambda} = 0, \forall\lambda.$$

A clever choice of the WAP density $h(q)$ can yield a WAP similar test,

$$\sup_{0\leq\phi\leq 1}\int \phi h, \text{ where } \int \phi f_{\beta_0,\lambda} = \alpha, \tag{2.14}$$

which is automatically uncorrelated with the statistic Q_{ST}. Hence the WAP similar test is also locally unbiased. We could replace an arbitrary weight function $\Lambda_1(\beta,\lambda)$ in $h = \int f_{\beta_0,\lambda} d\Lambda_1(\beta,\lambda)$ by

$$\Lambda(\beta,\lambda) = \frac{\Lambda_1(\beta,\lambda) + \Lambda_1(\kappa\circ(\beta,\lambda))}{2},$$

for $\kappa \in \{-1, 1\}$. Define the action *sign* group at $\kappa = -1$ as

$$\kappa \circ (\beta, \lambda) = \left(\beta_0 - \frac{d_{\beta_0}(\beta - \beta_0)}{d_{\beta_0} + 2j_{\beta_0}(\beta - \beta_0)}, \lambda \frac{(d_{\beta_0} + 2j_{\beta_0}(\beta - \beta_0))^2}{d_{\beta_0}^2} \right), \text{ where}$$

$$d_{\beta_0} = (a_0' \Omega^{-1} a_0)^{1/2}, j_{\beta_0} = \frac{e_1' \Omega^{-1} a_0}{(a_0' \Omega^{-1} a_0)^{-1/2}}, \text{ and } e_1 = (1, 0)', \tag{2.15}$$

for $\beta \neq \beta_{AR}$ defined as

$$\beta_{AR} = \frac{\omega_{11} - \omega_{12}\beta_0}{\omega_{12} - \omega_{22}\beta_0}. \tag{2.16}$$

We note that

$$\int f_{\beta,\lambda}(q_S, q_{ST}, q_T) \, d\Lambda(\beta, \lambda) = \int \int f_{\beta,\lambda}(q_S, q_{ST}, q_T) \, d\Lambda_1(\kappa \circ (\beta, \lambda)) \, \nu(d\kappa),$$

where ν is the Haar probability measure on the group $\{-1, 1\}$: $\nu(\{1\}) = \nu(\{-1\}) = 1/2$. Because

$$\begin{aligned} \int f_{\beta,\lambda}(q_S, -q_{ST}, q_T) \, d\Lambda(\beta, \lambda) &= \int f_{(-1)\circ(\beta,\lambda)}(q_S, q_{ST}, q_T) \, d\Lambda(\beta, \lambda) \\ &= \int f_{\beta,\lambda}(q_S, q_{ST}, q_T) \, d\Lambda(\beta, \lambda), \end{aligned}$$

the WAP similar test only depends on $q_S, |q_{ST}|, q_T$, and the test is locally unbiased; see Corollary 1 of Andrews, Moreira, and Stock (2006b). Here, we are able to analytically replace Λ_1 by Λ because of the existence of a group structure in the WIV model. Yet, replacing Λ_1 by Λ does not necessarily solve (2.13) when the WAP density is $h = \int f_{\beta_0,\lambda} d\Lambda_1(\beta, \lambda)$. The question is: can we distort Λ_1 by a weight function so that the WAP similar test in (2.14), or even a WAP test in (2.10), is approximately the WAP locally unbiased test in (2.13)? In Section 4.1, we show that the answer is yes.

Example 2.3: Heteroskedastic Autocorrelated Errors (HAC-IV)

We now drop the assumption that W is iid across rows with each row having a mean zero bivariate normal distribution with nonsingular covariance matrix Ω. We allow the reduced-form errors W to have a more general covariance matrix.

For the instrument Z, define $P_1 = Z\left(Z'Z\right)^{-1/2}$ and choose $P = [P_1, P_2] \in \mathcal{O}_n$, the group of $n \times n$ orthogonal matrices. Pre-multiplying the reduced-form model (2.5) by P', we obtain

$$\begin{pmatrix} P_1'Y \\ P_2'Y \end{pmatrix} = \begin{pmatrix} \mu a' \\ 0 \end{pmatrix} + \begin{pmatrix} W_1 \\ W_2 \end{pmatrix},$$

where $\mu = \left(Z'Z\right)^{1/2}\pi$. The statistic $P_2'Y$ is ancillary and we do not have previous knowledge about the correlation structure on W. In consequence, we consider tests based on $P_1'Y$:

$$\left(Z'Z\right)^{-1/2} Z'Y = \mu a' + W_1.$$

If $W_1 \sim N\left(0, \Sigma\right)$, the sufficient statistic is given by the pair

$$\begin{aligned} S &= \left[\left(b_0' \otimes I_k\right)\Sigma\left(b_0 \otimes I_k\right)\right]^{-1/2}\left(Z'Z\right)^{-1/2} Z'Y b_0 \text{ and} \\ T &= \left[\left(a_0' \otimes I_k\right)\Sigma^{-1}\left(a \otimes I_k\right)\right]^{-1/2}\left(a_0' \otimes I_k\right)\Sigma^{-1} vec\left[\left(Z'Z\right)^{-1/2} Z'Y\right]. \end{aligned}$$

The statistic S is pivotal and the statistic T is complete and sufficient for μ under the null. By Basu's lemma, S and T are independent.

The joint density of (S, T) at (s, t) is given by

$$f_{\beta,\mu}\left(s, t\right) = \left(2pi\right)^{-k}\exp\left(-\frac{\left\|s - \left(\beta - \beta_0\right)C_{\beta_0}\mu\right\|^2 + \left\|t - D_\beta\mu\right\|^2}{2}\right),$$

where the population means of S and T depend on

$$\begin{aligned} C_{\beta_0} &= \left[\left(b_0' \otimes I_k\right)\Sigma\left(b_0 \otimes I_k\right)\right]^{-1/2} \text{ and} \\ D_\beta &= \left[\left(a_0' \otimes I_k\right)\Sigma^{-1}\left(a_0 \otimes I_k\right)\right]^{-1/2}\left(a_0' \otimes I_k\right)\Sigma^{-1}\left(a \otimes I_k\right). \end{aligned}$$

Examples of two-sided HAC-IV similar tests are the Anderson-Rubin and score tests. The Anderson-Rubin test rejects the null when

$$S'S > q\left(k\right)$$

where $q\left(k\right)$ is the $1 - \alpha$ quantile of a chi-square-k distribution. The score test rejects the null when

$$LM^2 > q\left(1\right),$$

where $q(1)$ is the $1-\alpha$ quantile of a chi-square-one distribution. In the supplement to this paper, we show that the score statistic is given by

$$LM = \frac{S' C_{\beta_0}^{-1/2} D_{\beta_0}^{-1/2} T}{\sqrt{T' D_{\beta_0}^{-1/2} C_{\beta_0}^{-1} D_{\beta_0}^{-1/2} T}}.$$

We now present novel WAP tests based on the weighted average density

$$h(s,t) = \int f_{\beta,\mu}(s,t) \; d\Lambda_1(\beta,\mu).$$

Proposition 2 in the supplement shows that there is no sign group structure which preserves the null and alternative. This makes the task of finding a weight function $h(s,t)$ which yields a two-sided WAP similar test more difficult.

Instead of seeking a weight function Λ_1 so that the WAP similar test is approximately unbiased, we can select an arbitrary weight and find the WAP locally unbiased test:

$$\sup_{0\leq\phi\leq 1} \int \phi h, \text{ where } \int \phi f_{\beta_0,\mu} = \alpha \text{ and } \int \phi \left. \frac{\partial \ln f_{\beta,\mu}}{\partial \beta} \right|_{\beta=\beta_0} f_{\beta_0,\mu} = 0, \; \forall \mu, \tag{2.17}$$

where the integrals are with respect to (s,t).

We now define two weighted average densities $h(s,t)$ based on different weights Λ_1. The weighting functions are chosen after approximating the covariance matrix Σ by the Kronecker product $\Omega \otimes \Phi$. Let $\|X\|_F = (tr(X'X))^{1/2}$ denote the Frobenius norm of a matrix X. For a positive-definite covariance matrix Σ, Van Loan and Ptsianis (1993, p. 14) find symmetric and positive definite matrices Ω and Φ with dimension 2×2 and $k\times k$ which minimize $\|\Sigma - \Omega_0 \otimes \Phi_0\|_F$.

For the MM1 statistic $h_1(s,t)$, we choose $\Lambda_1(\beta,\mu)$ to be $N(\beta_0,1) \times N(0,\Phi)$. For the MM2 statistic $h_2(s,t)$, we first make a change of variables from β to θ, where $\tan(\theta) = d_\beta / c_\beta$. We then choose $\Lambda_1(\beta,\mu)$ to be $Unif[-pi,pi] \times N(0, \left\|l_{\beta(\theta)}\right\|^{-2} \Phi)$, where $l_\beta = (c_\beta, d_\beta)'$.

In the supplement, we simplify algebraically both MM1 and MM2 test statistics. We also show there that if $\Sigma = \Omega \otimes \Phi$, then: (i) both MM1 and MM2 statistics are invariant to orthogonal transformations; and (ii) the MM2 statistic is invariant to sign transformations which preserve the two-sided hypothesis testing problem.

There are two boundary conditions in the maximization problem (2.17). The first one states that the test is similar. The second states that the test is locally unbiased. In the supplement, we use completeness of T to show that the locally unbiased (LU) condition simplifies to

$$E_{\beta_0,\mu}\phi\left(S,T\right)S'C_{\beta_0}\mu = 0, \forall\mu. \qquad \text{(LU condition)}$$

The LU condition states that the test is uncorrelated with linear combinations (which depend on the instruments' coefficient μ) of the pivotal statistic S. The LU condition holds if the test is uncorrelated with *any* linear combination of S; that is,

$$E_{\beta_0,\mu}\phi\left(S,T\right)S = 0, \forall\mu. \qquad \text{(SU condition)}$$

In the supplement, we show that this strongly unbiased (SU) condition is indeed stronger than the LU condition. In practice, numerical simulations indicate that there is little power gain (if any) in using LU instead of SU tests. We will show that strongly unbiased tests based on MM1 and MM2 statistics are easy to implement and have overall good power.

Example 3: Nearly Integrated Regressor

Consider a model with persistent regressors. There is a stochastic equation

$$y_{1,i} = \varphi + y_{2,i-1}\beta + \epsilon_{1,i},$$

where the variable $y_{1,i}$ and the regressor $y_{2,i}$ are observed, and $\epsilon_{1,i}$ is a disturbance variable, $i = 1, ..., n$. This equation is part of a larger model where the regressor has serial dependence and can be correlated with the unknown disturbance. More specifically, we have

$$y_{2,i} = y_{2,i-1}\pi + \epsilon_{2,i},$$

where the disturbance $\epsilon_{2,i}$ is unobserved and possibly correlated with $\epsilon_{1,i}$. We assume that $\epsilon_i = (\epsilon_{1,i}, \epsilon_{2,i}) \overset{iid}{\sim} N\left(0,\Omega\right)$ where

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix}$$

is a known positive definite matrix. The goal is to assess the predictive power of the past value of $y_{2,i}$ on the current value of $y_{1,i}$. For example, a variable observed at time $i-1$ can be used to forecast stock returns in period i.

Let $P = (P_1, P_2)$ be an orthogonal $N \times N$ matrix where the first column is given by $P_1 = 1_N/\sqrt{N}$ and 1_N is an N-dimensional vector of ones. Algebraic manipulations show that $P_2 P_2' = M_{1_N}$, where $M_{1_N} = I_N - 1_N \left(1_N' 1_N\right)^{-1} 1_N'$ is the projection matrix to the space orthogonal to 1_N. Define the $(N-1)$-dimensional vector $\widetilde{y}_1 = P_2' y_1$. The joint density of $\widetilde{y}_1 = P_2' y_1$ and y_2 does not depend on the nuisance parameter φ and is given by

$$f_{\beta,\pi}\left(\widetilde{y}_1, y_2\right) = \left(2\pi\omega_{22}\right)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\omega_{22}} \sum_{i=1}^{N} \left(y_{2,i} - y_{2,i-1}\pi\right)^2\right\} \tag{2.18}$$
$$\times \left(2\pi\omega_{11.2}\right)^{-\frac{N-1}{2}} \exp\left\{-\frac{1}{2\omega_{11.2}} \sum_{i=1}^{N} \left(\widetilde{y}_{1,i} - \widetilde{y}_{2,i}\frac{\omega_{12}}{\omega_{22}} - \widetilde{y}_{2,i-1}\left[\beta - \pi\frac{\omega_{12}}{\omega_{22}}\right]\right)^2\right\},$$

where $\omega_{11.2} = \omega_{11} - \omega_{12}^2/\omega_{22}$ is the variance of $\epsilon_{1,i}$ not explained by $\epsilon_{2,i}$.

We want to test the null hypothesis $H_0 : \beta = \beta_0$ against the two-sided alternative $H_1 : \beta \neq \beta_0$. We now introduce a novel WAP test. The optimal locally unbiased test solves

$$\max_{\phi \in \mathbb{K}} \int \phi h, \text{ where } \int \phi f_{\beta_0,\pi} = \alpha \text{ and } \int \phi \left.\frac{\partial \ln f_{\beta,\pi}}{\partial \beta}\right|_{\beta=\beta_0} f_{\beta_0,\pi} = 0, \forall \pi. \tag{2.19}$$

We now define the weighted average density $h(\widetilde{y}_1, y_2)$. For the two-sided MM-2S statistic

$$h(\widetilde{y}_1, y_2) = \int f_{\beta,\pi}\left(\widetilde{y}_1, y_2\right) \; d\Lambda_1\left(\beta, \pi\right),$$

we choose $\Lambda_1\left(\beta, \mu\right)$ to be the product of $N\left(\beta_0, 1\right)$ and $Unif[\underline{\pi}, \overline{\pi}]$. In the numerical results, we set $\underline{\pi} = 0.5$ and $\overline{\pi} = 1$.

As for the constraints in the maximization problem, there are two boundary conditions. The first one states that the test is similar. The second one asserts the power derivative is zero at the null $\beta_0 = 0$. In Section 4.1, we show that these tests are admissible. Hence, we can interpret the WAP locally unbiased test for (2.19) as being an optimal test with correct size where the weighted average density h is accordingly adjusted.

In the supplement, we also discuss testing $H_0 : \beta \leq \beta_0$ against the one-sided alternative $H_0 : \beta > \beta_0$ (the adjustment for $H_1 : \beta < \beta_0$ is straightforward by multiplying $y_{1,i}$ and ω_{12} by minus one).

2.1 The Maximization Problem

The problem given in equation (2.1) can be particularly difficult to solve. For example, consider the special case where we want to find WAP similar tests. For incomplete exponential families and under regularity conditions (on the densities and $\mathbb{V}_0$), Linnik (2000) proves the existence of a smooth α-similar test ϕ_ϵ such that

$$\int \phi_\epsilon h \geq \sup \int \phi h - \epsilon \tag{2.20}$$

for $\epsilon > 0$ among all α-similar tests on $\mathbf{V}$. If the test ϕ_ϵ satisfies (2.20), we say that it is an ϵ-optimal test. Here we show that the general problem (2.1) admits a maximum if we do not impose smoothness. Let $\mathcal{L}_1(Y)$ be the usual Banach space of integrable functions ϕ. We denote $\gamma = (\gamma_v^1, \gamma_v^2)$ and let $g = \{g_v \in \mathcal{L}_1(Y); v \in \mathbf{V}\}$.

Proposition 1. *Define*

$$M(h, g, \gamma) = \sup_\phi \int \phi h \text{ where } \phi \in \Gamma(g, \gamma) \tag{2.21}$$

for $\Gamma(g, \gamma) = \{\phi \in \mathbb{K}; \int \phi g_v \in [\gamma_v^1, \gamma_v^2], \forall v \in \mathbf{V}\}$ *and* $\mathbb{K} = \{\phi \in \mathcal{L}_\infty(Y); 0 \leq \phi \leq 1\}$. *If* $\Gamma(g, \gamma)$ *is not empty, then there exists* $\overline{\phi}$ *which solves (2.21).*

Comments: 1. The proof relies on the Banach-Alaoglu Theorem, which states that in a real normed linear space, the closed unit ball in its dual is weak*-compact. The $\mathcal{L}_\infty(Y)$ space is the dual of $\mathcal{L}_1(Y)$, that is, $\mathcal{L}_\infty(Y) = [\mathcal{L}_1(Y)]^*$. The functional $\phi \to \int \phi h$ is continuous in the weak* topology.

2. The optimal test may be randomized.

3. Consider the Banach space $\mathcal{C}(Y)$ of bounded continuous functions with supremum norm. The dual of $\mathcal{C}(Y)$ is $rba\,(Y)$, the space of regular bounded additive set functions defined on the algebra generated by closed sets whose norm is the total variation. However, the space $\mathcal{C}(Y)$ is not the dual of another Banach space $\mathcal{S}(Y)$. If there were such a space $\mathcal{S}(Y)$, then $\mathcal{S}(Y) \subset [\mathcal{C}(Y)]^* = rba\,(Y)$. Hence $[rba\,(Y)]^* \subset [\mathcal{S}(Y)]^* = \mathcal{C}(Y)$. Therefore, $\mathcal{C}(Y)$ would be a reflexive space which is not true; see Dunford and Schwartz (1988, p. 376).

4. Comment 3 shows that the proof would fail if we replaced $\mathcal{L}_\infty(Y)$ by $\mathcal{C}(Y)$. Indeed, an optimal $\phi \in \mathcal{C}(Y)$ does not exist even for testing a simple null against a simple alternative in one-parameter exponential families. The

failure to obtain an optimal continuous procedure justifies the search for an ϵ-optimal test given by Linnik (2000).

5. If g_v is the density f_v and $\alpha \in [\gamma_v^1, \gamma_v^2]$, for all $v \in \mathbf{V}$, then the set $\Gamma(g, \gamma)$ is non-empty because the trivial test $\phi = \alpha$ satisfies $\int \phi f_v = \alpha \in [\gamma_v^1, \gamma_v^2]$, $\forall v \in \mathbf{V}$.

Proposition 1 guarantees the existence of an optimal test $\overline{\phi}$. Lemma 1, stated in the Appendix, gives a characterization of $\overline{\phi}$ relying on properties of the epigraph of h under $\mathbb{K}$. It does not however present an explicit form of the optimal test.

For the remainder of the paper, we propose to approximate (2.21) by a sequence of problems. This simplification yields characterization of optimal tests. Continuity arguments guarantee that these tests nearly maximize the original problem given in (2.21).

3 Discrete Approximation

Implementing the optimal test $\overline{\phi}$ may be difficult with an uncountable number of boundary conditions. When $\mathbf{V}$ is finite, (2.21) simplifies to

$$\sup_{\phi \in \mathbb{K}} \int \phi h \text{ where } \int \phi g_{v_l} \in [\gamma_{v_l}^1, \gamma_{v_l}^2],\ l = 1, ..., n. \tag{3.22}$$

Corollary 1. *Suppose that* $\mathbf{V}$ *is finite and the constraint of (3.22) is not empty.*
(a) *There exists a test* $\overline{\phi}_n \in \mathbb{K}$ *that solves (3.22).*
(b) *A sufficient condition for* $\overline{\phi}_n$ *to solve (3.22) is the existence of a vector* $c_l = (c_l^1, ..., c_l^m) \in \mathbb{R}^m$, $l = 1, ..., n$, *such that* $c_l^j > 0$ *(resp.* < 0*) implies* $\int \overline{\phi}_n g_{v_l}^j = \gamma_{v_l}^{2,j}$ *(resp.* $\gamma_{v_l}^{1,j}$*) and*

$$\overline{\phi}_n(y) = \begin{cases} 1 & \text{if } h(y) > \sum_{l=1}^n c_l \cdot g_{v_l}(y) \\ 0 & \text{if } h(y) < \sum_{l=1}^n c_l \cdot g_{v_l}(y) \end{cases}. \tag{3.23}$$

(c) *If* ϕ *satisfies (3.23) with* $c_l \geq 0$, $l = 1, ..., n$, *then it solves*

$$\sup_{\phi \in \mathbb{K}} \int \phi h \text{ where } \int \phi g_{v_l} \leq \gamma_{v_l}^2,\ l = 1, ..., n.$$

(d) *If there exist tests* ϕ°, ϕ^1 *satisfying the constraints of problem (3.22) with*

strict slackness or $\int \phi^{\circ} h < \int \phi^{1} h$*, then there exist* c_l *and a test* ϕ *satisfying (3.23) such that* $c_l \cdot \left(\int \phi g_{v_l} - \gamma^1_{v_l}\right) \leq 0$ *and* $c_l \cdot \left(\int \phi g_{v_l} - \gamma^2_{v_l}\right) \leq 0$*,* $l = 1, ..., n$*. A necessary condition for* ϕ *to solve (3.22) is that (3.23) holds almost everywhere (a.e.).*

Corollary 1 provides a version of the Generalized Neyman-Pearson (GNP) lemma. In this paper, we are also interested in the special case in which g_v is the density f_v and the null rejection probabilities are all the same (and equal to $\alpha = \gamma^1_v = \gamma^2_v$) for $v \in \mathbf{V}$. This allows us to provide an easy-to-check condition for the characterization of the optimal procedure $\overline{\phi}_n$: if we find a (possibly non-optimal) test ϕ^1 whose power is strictly larger than $\phi^{\circ} = \alpha$, we can characterize the optimal procedure $\overline{\phi}_n$. This condition holds unless h is a linear combination of f_{v_l} a.e.; see Corollary 3.6.1 of Lehmann and Romano (2005).

The next lemma provides an approximation to $\overline{\phi}$ for the weak* topology.

Lemma 1. *Suppose that the correspondence* $\Gamma(g, \gamma)$ *has no empty value. Let the space of functions* (g, γ) *have the following topology: a net* $g^n \to g$ *when* $g^n_v \to g_v$ *in* $\mathcal{L}_1(Y)$ *for every* $v \in \mathbf{V}$ *and* $\gamma^n_v \to \gamma_v$ *a.e.* $v \in \mathbf{V}$*. We use the weak* topology on* $\mathbb{K}$*.*
(a) *The mapping* $\Gamma(g, \gamma)$ *is continuous in* (g, γ)*.*
(b) *The function* $M(h, g, \gamma)$ *is continuous and the mapping* Γ_M *defined by* $\Gamma_M(g, \gamma) = \left\{\phi \in \mathbb{K}; \phi \in \Gamma(g, \gamma) \text{ and } M(h, g, \gamma) = \int \phi h\right\}$ *is upper semicontinuous (u.s.c.).*

Comments: 1. The space of g functions is $\{g : \mathbf{V} \times Y \to \mathbb{R}^m; g(v, \cdot) \in \mathcal{L}_1(Y)$ and $g(\cdot, y) \in \mathcal{C}(\mathbf{V})$, for all $(v, y) \in \mathbf{V} \times Y$ $\}$ and the space of γ functions is $\{\gamma : \mathbf{V} \to \mathbb{R}^{2m}$ measurable function$\}$.

2. A net in a set $\mathbb{X}$ is a function $n : \mathbb{D} \to \mathbb{X}$, where $\mathbb{D}$ is a directed set. A directed set is any set $\mathbb{D}$ equipped with a direction which is a reflexive and transitive relation with the property that each pair has an upper bound.

Lemma 1 can be used to show convergence of the power function. Let $I(\cdot)$ be the indicator function.

Theorem 1. *Let* $\mathcal{P}_n = \{\mathbb{P}^n_l; l = 1, ..., m_n\}$ *be a partition of* $\mathbf{V}$ *and define for some* $v \in \mathbf{V}$

$$g^n_v(y) = \sum_{l=1}^{m_n} g_{v_l}(y) I\left(v \in \mathbb{P}^n_l\right),$$

where $v_l \in \mathbb{P}_l^n$, $l = 1, ..., m_n$. For this sequence the problem (2.21) becomes (3.22) with the optimal test $\overline{\phi}_n$ given in (3.23).
(a) *If the partition norm $|\mathcal{P}_n| \to 0$, then $g_v^n(y) \to g_v(y)$ for a.e. $y \in Y$ and $v \in \mathbf{V}$.*
(b) *If $g_v^n(y) \to g_v(y)$ for a.e. $y \in Y$ and $\int \sup_n |g_v^n| < \infty$ for every $v \in \mathbf{V}$, then $g^n \to g$.*
(c) *If $g^n \to g$, then $\int \overline{\phi}_n h \to \int \overline{\phi} h$. Furthermore, if $\overline{\phi}$ is the unique solution of (2.1), then $\int \overline{\phi}_n f \to \int \overline{\phi} f$, for any $f \in \mathcal{L}_1(Y)$.*

Comments: 1. If the sets $\mathbb{P}_l^n$ are intervals, we can choose the elements v_l to be the center of $\mathbb{P}_l^n$, $l = 1, ..., m_n$.

2. The norm $|\mathcal{P}_n|$ is defined as $\max_{l=1,...,m_n} \sup_{v_i, v_l \in \mathbb{P}_l^n} \|v_i - v_l\|$. We note that we can create a sequence of partitions $\mathcal{P}_n$ whose norm goes to zero if the set $\mathbf{V}$ is bounded.

3. This theorem is applicable to the nearly integrated regressor model in Example 3 where the regressors' coefficient is naturally bounded.

4. Finding a WAP similar or locally unbiased test for the HAC-IV model entails equality boundary constraints on an unbounded set $\mathbf{V}$. However, we can show that the power function is analytic in $v \in \mathbf{V}$. Hence, the WAP similar or locally unbiased test is the same if we replace $\mathbf{V}$ by a bounded set with non-empty interior $\mathbf{V}_2 \subset \mathbf{V}$ in the boundary conditions.

If the number of boundary conditions increases properly (i.e., $|\mathcal{P}_n| \to 0$ as $n \to \infty$), it is possible to approximate the power function of the optimal test $\overline{\phi}$. The approximation is given by a sequence of tests $\overline{\phi}_n$ for a finite number of boundary conditions. This is convenient as the tests $\overline{\phi}_n$ are given by Corollary 1 and are nonrandomized if g_v is analytic.

We can find the multipliers c_l, $l = 1, ..., n$, with nonlinear optimization algorithms. Alternatively, we can implement $\overline{\phi}_n$ numerically using a linear programming approach[1]. The method is simple and requires a sample drawn from only one law. Importance sampling can help to efficiently implement the numerical method.

Let $Y^{(j)}$, $j = 1, ..., J$, be i.i.d. random variables with positive density m.

[1]The connection between linear programming methods and (generalized) Neyman-Pearson tests is no surprise given the seminal contributions of George Dantzig to both fields; see Dantzig (1963, p. 23-24).

The discretized approximated problem (3.22) can be written as

$$\max_{0\leq\phi(Y^{(j)})\leq 1} \frac{1}{J}\sum_{j=1}^{J}\phi(Y^{(j)})\frac{h(Y^{(j)})}{m(Y^{(j)})}$$
$$\text{s.t.} \quad \frac{1}{J}\sum_{j=1}^{J}\phi(Y^{(j)})\frac{g_{v_l}(Y^{(j)})}{m(Y^{(j)})} \in [\gamma_{v_l}^1, \gamma_{v_l}^2],\ l = 1, ..., n.$$

We can rewrite the above problem as the following standard linear programming (primal) problem:

$$\max_{0\leq x_j\leq 1} \ r'x$$
$$\text{s.t.} \quad Ax \leq p,$$

where $x = (\phi(Y^{(1)}), ..., \phi(Y^{(J)}))'$, $r = \left(h(Y^{(1)})/m(Y^{(1)}), ..., h(Y^{(J)})/m(Y^{(J)})\right)'$ are vectors in $\mathbb{R}^J$. The $2n \times J$ matrix A and the $2n$-dimensional vector p are given by

$$A = \begin{bmatrix} -A_v \\ A_v \end{bmatrix} \text{ and } p = \begin{bmatrix} -\gamma_v^1 \\ \gamma_v^2 \end{bmatrix},$$

where the (l, j)-entry of the matrix A_v is $g_{v_l}(Y^{(j)})/m(Y^{(j)})$ and the l-entry of γ_v^i is $\gamma_{v_l}^i$ for $i = 1, 2$.

Its dual program is defined by

$$\min_{c\in\mathbb{R}_+^{2n}} \ p'c$$
$$\text{s.t.} \quad A'c \geq r.$$

Define the Lagrangian function by

$$L(x, c) = r'x + c'(p - Ax) = p'c + x'(r - A'c).$$

The optimal solutions of the primal and dual programs, $\overline{x}$ and $\overline{c}$, must satisfy the following saddle point condition:

$$L(x, \overline{c}) \leq L(\overline{x}, \overline{c}) \leq L(\overline{x}, c)$$

for all $x \in [0, 1]^J$ and $c \in \mathbb{R}_+^{2n}$. Krafft and Witting (1967) is the seminal reference of employing a linear programming method to characterize optimal tests. Chiburis (2009) uses an analogous approach to ours for the special case of approximating tests at level α in Example 1.

3.1 One-Parameter Exponential Family

We have shown that a suitably chosen sequence of tests can approximate the optimal WAP test. Linnik (2000) gives examples where similar tests are necessarily random. Our theoretical framework can be useful here as we can find a sequence of nonrandomized tests which approximate the optimal similar test — whether it is random or not. We now illustrate the importance of using the weak* topology in our theory with a knife-edge example. We consider a one-parameter exponential family model when a test is artificially required to be similar at level α in an interval.

Let the probability density function of Y belong to the exponential family

$$f_v(y) = C(v)e^{vR(y)}\eta(y),$$

where the parameter $v \in \mathbb{R}$ and $R : Y \to \mathbb{R}$. For testing $H_0 : v \leq v_0$ against $H_1 : v > v_0$, the UMP test rejects the null when $R(y) > c$, where the critical value c satisfies $P_{v_0}(R(Y) > c) = \alpha$; see Lehmann and Romano (2005, Corollary 3.4.1). The least favorable distribution is $\Lambda_0(v) = 1(v \geq v_0)$ and the test is similar at level α only at the boundary between the null and alternative hypotheses, $\mathbf{V} = \overline{\mathbb{V}}_0 \cap \overline{\mathbb{V}}_1 = \{v_0\}$.

Suppose instead that a test ϕ must be similar at level α for all values $v \leq v_0$; that is, $\mathbf{V} = (-\infty, v_0]$. Although there is no reason to impose this requirement, this pathological example highlights the power convergence and weak*-convergence of Theorem 1. Here, the optimal test is randomized and known. Because the sufficient statistic $R = R(Y)$ is complete, any similar test ϕ equals α (up to a set of measure zero).

For some fixed alternative $\overline{v} > v_0$, let $\left(\overline{\phi}_n\right)_{n\in\mathbb{N}}$ be the sequence of uniformly most powerful tests $\overline{\phi}_n$ similar at values v_0 and $v_l = v_0 - 2^{-(n-2)}l$, $l = 0, 1, ..., 2^{2n-4}$ for $n \geq 2$. As n increases, we augment the interval $[v_0 - 2^{n-2}, v_0]$ to be covered and provide a finer grid by the rate $2^{-(n-2)}$.

The test $\overline{\phi}_n$ accepts the alternative when

$$e^{\overline{v}R(y)} > \sum_{l=0}^{2^{2n-4}} c_l e^{v_l R(Y)},$$

where the multipliers c_l are determined by

$$P_{v_l}\left(e^{\overline{v}R(Y)} > \sum_{l=0}^{2^{2n-4}} c_l e^{v_l R(Y)}\right) = \alpha,\ l = 0, ..., 2^{2n-4}.$$

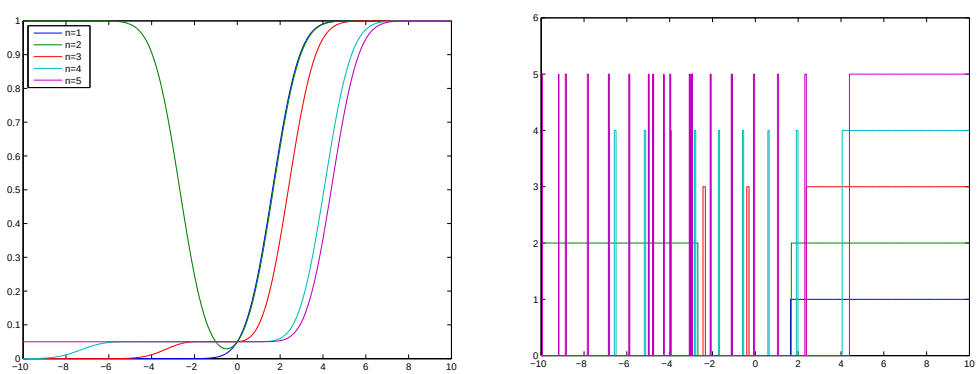
There are two interesting features for this sequence of tests. First, imposing similarity on a finite number of points gives some slack in terms of power. By Lehmann and Romano (2005, Corollary 3.6.1), $E_{\overline{v}}\overline{\phi}_n(Y) > \alpha \equiv E_{\overline{v}}\overline{\phi}(Y)$. Second, the spurious power vanishes as the number of boundary conditions increases. Define the collection $\mathcal{P}_n = \{\mathbb{P}_l^n; l = 0, ..., 2^{2n-4}\}$, where

$$\mathbb{P}_l^n = \left(v_0 - 2^{-(n-2)}(l+1), v_0 - 2^{-(n-2)}l\right]$$

for $l = 0, ..., 2^{2n-4} - 1$ and $\mathbb{P}_{2^{2n-4}}^n = (-\infty, v_0 - 2^{n-2}]$. As n increases, $E_v\overline{\phi}_n(Y) \to \alpha$ for $v > 0$.

To illustrate this convergence, let $Y \sim N(v, 1)$. We take the alternative $\overline{v} = 1$ and consider $v_0 = 0$. Figure 1 presents the power function for $v \in [-10, 10]$ of $\overline{\phi}_n$, $n = 1, ..., 5$. The total number of boundary conditions is respectively 1, 2, 5, 17, and 65. The power curve for $\overline{\phi}$ is trivially equal to $\alpha = 0.05$. As n increases, the null rejection probability approaches α. For the alternative $\overline{v} = 1$, the rejection probability with 17 boundary conditions is also close to α. This behavior is also true for any alternative v, although the convergence is not uniform in $v > 0$.

Figure 1: Knife-Edge Example



Power Curves

Rejection Regions

Because each test $\overline{\phi}_n$ is nonrandom, $\left(\overline{\phi}_n\right)_{n\in\mathbb{N}}$ does *not* converge to $\overline{\phi}(y) \equiv \alpha$ for any value of y. This example shows that establishing almost sure (a.s.) or $\mathcal{L}_\infty(Y)$ convergence in general is hopeless. However, $\int \overline{\phi}_n g \to \int \overline{\phi} g$ for any $g \in \mathcal{L}_1(Y)$. In particular, take $g = (\kappa_2 - \kappa_1)^{-1} I(\kappa_1 \leq y \leq \kappa_2)$. Then

the integral

$$\int \overline{\phi}_n(y)\, g(y)\ dy = \frac{1}{\kappa_2 - \kappa_1} \int_{\kappa_1}^{\kappa_2} \overline{\phi}_n(y)\ dy$$

converges to α. This implies that, for any interval $[\kappa_1, \kappa_2]$, the rejection and acceptance regions need to alternate more often as n increases. Figure 1 illustrates the oscillation between the acceptance regions and rejection regions for $y \in [-10, 10]$ of $\overline{\phi}_n$, $n = 1, ..., 5$. The x-axis shows the value of y which ranges from -10 to 10. The y-axis represents the rejection region for $n = 1, ..., 5$. For example, for the test with two boundary conditions, we reject the null when y is smaller than -2.7 and larger than 1.7.

4 Uniform Approximation

If $\mathbf{V}$ is finite, we can characterize the optimal test $\overline{\phi}$ from Lagrangian multipliers in a Euclidean space. Another possibility is to relax the constraint $\Gamma(g, \gamma) = \{\phi \in \mathbb{K}; \int \phi g_v \in [\gamma_v^1, \gamma_v^2], \forall v \in \mathbf{V}\}$. Consider the following problem

$$M(h, g, \gamma, \delta) = \sup_{\phi \in \Gamma(g,\gamma,\delta)} \int \phi h, \tag{4.24}$$

where $\Gamma(g, \gamma, \delta) = \{\phi \in \mathbb{K}; \int \phi g_v \in [\gamma_v^1 - \delta, \gamma_v^2 + \delta], \forall v \in \mathbf{V}\}$.

Lemma 2. *If $\Gamma(g, \gamma) \neq \emptyset$, then for sufficiently small $\delta > 0$ the following hold:*
(a) *There exists a test $\overline{\phi}_\delta \in \Gamma(g, \gamma, \delta)$ which solves (4.24).*
(b) *There are vector positive regular counting additive (*rca*) measures Λ_δ^+ and Λ_δ^- on compact $\mathbf{V}$ which are Lagrangian multipliers for problem (4.24):*

$$\overline{\phi}_\delta \in \ \underset{\phi \in \mathbb{K}}{arg\ max} \int \phi h + \int \phi \int_{\mathbf{V}} g_v \cdot \left(\Lambda_\delta^+(dv) - \Lambda_\delta^-(dv)\right),$$

where $\int \int_{\mathbf{V}} \left(\overline{\phi}_\delta g_v - \gamma_v^2 - \delta\right) \cdot \Lambda_\delta^+(dv) = 0$ and $\int \int_{\mathbf{V}} \left(\overline{\phi}_\delta g_v - \gamma_v^1 + \delta\right) \cdot \Lambda_\delta^-(dv) = 0$ are the usual slackness conditions.

Comments: 1. Finding Λ_δ^+ and Λ_δ^- is similar to the problem of seeking a least-favorable distribution associated with max-min optimal tests; see Krafft and Witting (1967). Polak (1997) develops implementation algorithms for related problems.

2. If $\mathbf{V}$ is not compact and $\sup_{v\in\mathbf{V}} \int |g_v| < \infty$, then Λ_δ^+ and Λ_δ^- are regular bounded additive (*rba*) set functions. See Dunford and Schwartz (1988, p. 261) for details.

From Lemma 2 and using Fubini's Theorem (see Dunford and Schwartz (1988, p. 190)) the optimal test is given by

$$\overline{\phi}_\delta(y) = \begin{cases} 1, & \text{if } h(y) > c_\delta(y) \\ 0, & \text{if } h(y) < c_\delta(y) \end{cases}$$

where

$$c_\delta(y) \equiv \int_{\mathbf{V}} g_v(y)\Lambda_\delta(dv) \tag{4.25}$$

and $\Lambda_\delta = \Lambda_\delta^+ - \Lambda_\delta^-$ for positive *rca* measures Λ_δ^+ and Λ_δ^- on $\mathbf{V}$.

The next theorem shows that $\overline{\phi}_\delta$ provides an approximation of the optimal test $\overline{\phi}$. We again consider the weak* topology on $\mathcal{L}_\infty(Y)$. Because the objective function $\int \phi h$ is continuous in the weak* topology, we are able to prove the following lemma.

Lemma 3. *Suppose that the correspondence $\Gamma(g,\gamma,\delta)$ has no empty value. The following holds under the weak* topology on $\mathcal{L}_\infty(Y)$.*
(a) *The correspondence $\Gamma(g,\gamma,\delta)$ is continuous in δ.*
(b) *The function $M(h,g,\gamma,\delta)$ is continuous and the mapping Γ_M defined by $\Gamma_M(g,\gamma,\delta) = \left\{\phi; \phi \in \Gamma(g,\gamma,\delta) \text{ and } M(h,g,\gamma,\delta) = \int \phi h\right\}$ is u.s.c.*

Lemma 3 can be used to show convergence of the power function.

Theorem 2. *Let $\overline{\phi}_\delta$ and $\overline{\phi}$ be respectively the solutions for (4.24) and (2.21). Then $\int \overline{\phi}_\delta h \to \int \overline{\phi} h$ when $\delta \to 0$. Furthermore, if $\overline{\phi}$ is the unique solution of (2.21), then $\int \overline{\phi}_\delta f \to \int \overline{\phi} f$ when $\delta \to 0$ for any $f \in \mathcal{L}_1(Y)$.*

4.1 WAP Similar Test and Admissibility

In this section, let us consider the case of similar tests, i.e., when $g_v = f_v$ is a density and $\gamma_v^1 = \gamma_v^2 = \alpha$, for all $v \in \mathbf{V}$. Hence, we drop the notation dependence on γ.

By construction, the rejection probability of $\overline{\phi}_\delta$ is uniformly bounded by $\alpha + \delta$ for any $v \in \mathbf{V}$. We say that the optimal test $\overline{\phi}_\delta$ is trivial if $\overline{\phi}_\delta = \alpha + \delta$ almost everywhere. If the optimal test has power greater than size, then

it cannot be trivial for sufficiently small $\delta > 0$. The following assumption provides a sufficient condition for $\overline{\phi}_\delta$ to be nonrandomized.

Assumption U-BD. $\{f_v, v \in \mathbf{V}\}$ is a family of uniformly bounded analytic functions.

Assumption U-BD states that each $f_v(y)$ is a restriction to Y of a holomorphic function defined on a domain $\mathcal{D}$ such that for any given compact $\mathbb{D} \subset \mathcal{D}$

$$\sup_{v \in \mathbf{V}} |f_v(z)| < \infty$$

holds for every $z \in \mathbb{D}$, where domain means an open set in $\mathbb{C}^m$. The joint densities $f_{\beta,\mu}(s,t)$ of Example 2.3 and $f_{\beta,\pi}(\widetilde{y}_1, y_2)$ of Example 3 satisfy Assumption U-BD. On the other hand, the density of the maximal invariant to affine data transformations in the Behrens-Fisher problem is non-analytic and does not satisfy Assumption U-BD.

Theorem 3. *Suppose that $h(y)$ is an analytic function on $\mathbb{R}^m$ and the optimal test $\overline{\phi}_\delta$ is not trivial. If Assumption U-BD holds, then the optimal test $\overline{\phi}_\delta$ is nonrandomized.*

For distributions with a unidimensional sufficient statistic, Besicovitch (1961) shows there exist approximately similar regions. Let P_v, $v \in \mathbf{V}$, with density f_v satisfying $|f_v'(y)| \leq \kappa$. Then for $\alpha \in (0,1)$ and $\delta > 0$, there exists a set $\mathbb{A}_\delta \in \mathcal{B}$ such that $|P_v(\mathbb{A}_\delta) - \alpha| < \delta$ for $v \in \mathbf{V}$. This method also yields a δ-similar nonrandomized test $\phi_\delta(y) = I(y \in \mathbb{A}_\delta)$. A caveat is that $\phi_\delta(y)$ is not based on optimality considerations; see also the discussion on similar tests by Perlman and Wu (1999). Indeed, for most distributions in which $|f_v'(y)| \leq \kappa$, $v \in \mathbf{V}$, $f_v'(y)$ is also bounded for $v \in \mathbb{V}_1$ compact. Hence, the rejection probability of $\phi_\delta(y)$ is approximately equal to α even for alternatives $v \in \mathbb{V}_1$; see the discussion by Linnik (2000). By construction, the test $\overline{\phi}_\delta$ instead has desirable optimality properties.

An important property of the WAP similar test is admissibility. The following theorem shows that for a relevant class of problems, the optimal similar test is admissible.

Theorem 4. *Let $\mathbb{B}$ and $\mathbb{P}$ be Borel sets in $\mathbb{R}^k$ and $\mathbb{R}^m$ such that $\mathbb{V} = \mathbb{B} \times \mathbb{P}$, $\mathbb{V}_0 = \mathbf{V} = \{\beta_0\} \times \mathbb{P}$, $\mathbb{V}_1 = \mathbb{V} - \mathbb{V}_0$ and $h = \int_{\mathbb{V}_1} f_v \Lambda_1(dv)$ for some probability measure Λ_1. Let β_0 be a cumulative point in $\mathbb{B}$, the set $\mathbb{V}$ be compact, and Λ_1*

be a rca measure with full support on $\mathbb{V}_1$ and $f_v > 0$, for all $v \in \mathbb{V}_0$. Then there exists a sequence of tests with Neyman structure which weakly converges to a WAP similar test. In particular, the WAP similar test is admissible.

Comment: 1. If the power function is continuous, an unbiased test ϕ satisfies $\int \phi f_{v_0} = \alpha \leq \int \phi f_{v_1}$ for $v_0 \in \mathbb{V}_0$ and $v_1 \in \mathbb{V}_1$. Because the multiplier associated to the inequality $\alpha \leq \int \phi f_{v_1}$ is non-positive, we can extend this theorem to show that a WAP unbiased test is admissible as well.

By imposing inequality constraints, the choice of Λ_1 does not matter. In some sense, the equality conditions adjust the arbitrary choice of Λ_1 to yield a WAP test that is approximately similar.

We prefer not to give a firm recommendation on which constraints are reasonable for WAP tests to have. Example 1 on moment inequalities shows that we should not try to require the test to be similar indiscriminately. On the other hand, take Example 2.2 on WIVs with homoskedastic errors. Moreira's (2003) conditional likelihood ratio (CLR) test is by construction similar, whereas the likelihood ratio (LR) test is severely biased. Chernozhukov, Hansen, and Jansson (2009) and Anderson (2011) respectively show that the CLR and LR tests are admissible. However, Andrews, Moreira, and Stock (2006a) demonstrate that the CLR test dominates all invariant tests (including the LR test) for all practical purposes. In Section 5, we show that WAP similar or unbiased tests have overall good power also for Example 2.3 on the HAC-IV model and Example 3 on the nearly integrated regressor.

5 Numerical Simulations

In this section, we provide numerical results for the two running examples in this paper. Section 5.1 presents power curves for the AR, LM, and the novel WAP tests for the HAC-IV model. Section 5.2 provides power plots for the L_2 test of Wright (2000) as well as the novel WAP tests for the nearly integrated regressor model.

5.1 HAC-IV

We can write

$$\Omega = \begin{bmatrix} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{bmatrix} P \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix} P' \begin{bmatrix} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{bmatrix},$$

where P is an orthogonal matrix and $\rho = \omega_{12}/\omega_{11}^{1/2}\omega_{22}^{1/2}$. For the numerical simulations, we specify $\omega_{11} = \omega_{22} = 1$.

We use the decomposition of Ω to perform numerical simulations for a class of covariance matrices:

$$\Sigma = P \begin{bmatrix} 1+\rho & 0 \\ 0 & 0 \end{bmatrix} P' \otimes diag\left(\varsigma_1\right) + P \begin{bmatrix} 0 & 0 \\ 0 & 1-\rho \end{bmatrix} P' \otimes diag\left(\varsigma_2\right),$$

where ς_1 and ς_2 are k-dimensional vectors.

We consider two possible choices for ς_1 and ς_2. For the first design, we set $\varsigma_1 = \varsigma_2 = (1/\varepsilon - 1, 1, ..., 1)'$. The covariance matrix then simplifies to a Kronecker product: $\Sigma = \Omega \otimes diag\left(\varsigma_1\right)$. For the non-Kronecker design, we set $\varsigma_1 = (1/\varepsilon - 1, 1, ..., 1)'$ and $\varsigma_2 = (1, ..., 1, 1/\varepsilon - 1)'$. This setup captures the data asymmetry in extracting information about the parameter β from each instrument. For small ε, the angle between ς_1 and ς_2 is nearly zero. We report numerical simulations for $\varepsilon = (k+1)^{-1}$. As k increases, the vector ς_1 becomes orthogonal to ς_2 in the non-Kronecker design.

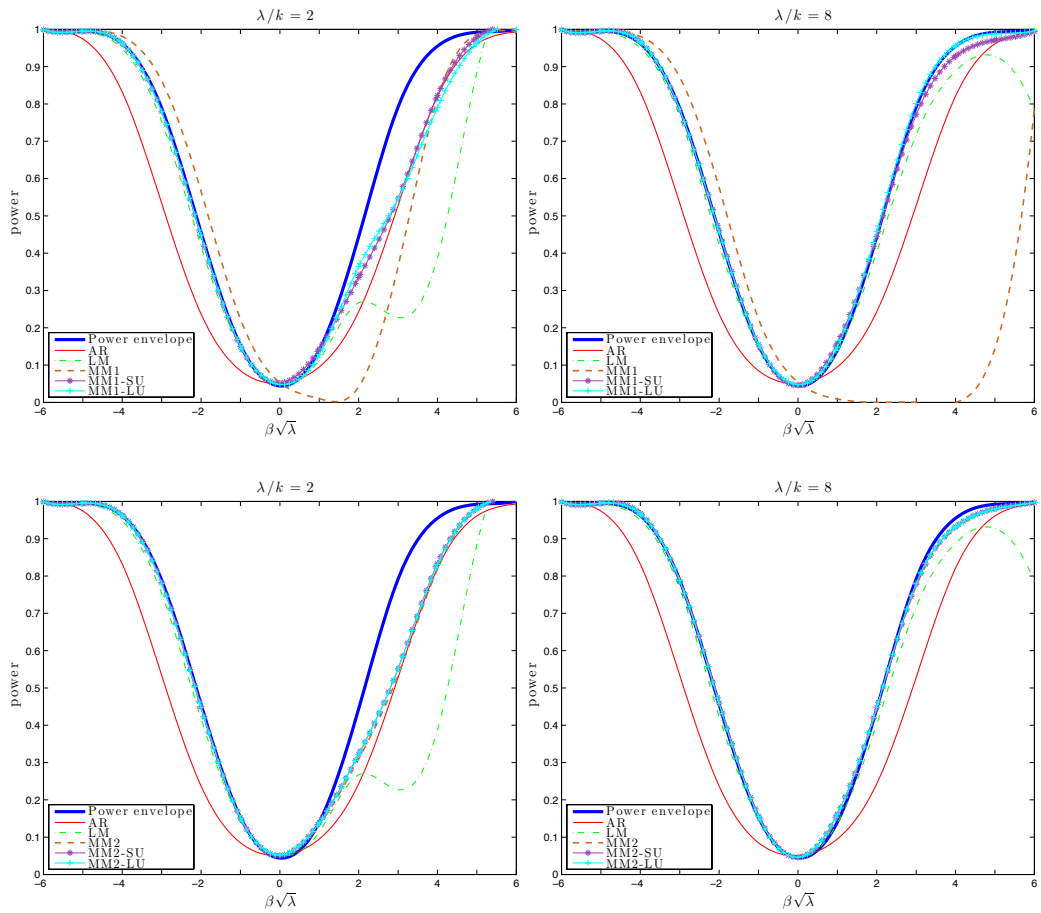
We set the parameter $\mu = \left(\lambda^{1/2}/\sqrt{k}\right) 1_k$ for $k = 2, 5, 10, 20$ and $\rho = -0.5, 0.2, 0.5, 0.9$. We choose $\lambda/k = 0.5, 1, 2, 4, 8, 16$, which span the range from weak to strong instruments. We focus on tests with significance level 5% for testing $\beta_0 = 0$. To conserve space, we report here only power plots for $k = 5$, $\rho = 0.9$, and $\lambda/k = 2, 8$. The full set of simulations is available in the supplement.

We present plots for the power envelope and power functions against various alternative values of β and λ. All results reported here are based on 1,000 Monte Carlo simulations. We plot power as a function of the rescaled alternative $(\beta - \beta_0)\lambda^{1/2}$, which reflects the difficulty in making inference on β for different instruments' strength.

Figure 2 reports numerical results for the Kronecker product design. All four pictures present a power envelope (as defined in the supplement to this paper) and power curves for two existing tests, the Anderson-Rubin (AR) and score (LM) tests.

The first two graphs plot the power curves for three different similar tests based on the MM1 statistic. The MM1 test is a WAP similar test based on $h_1(s,t)$ as defined in Section 2. The MM1-SU and MM1-LU tests also satisfy respectively the strongly unbiased and locally unbiased conditions. All three tests reject the null when the $h_1(s,t)$ statistic is larger than an adjusted critical value function. In practice, we approximate these critical

Figure 2: Power Comparison (Kronecker Covariance)



value functions with 10,000 replications. The MM1 test sets the critical value function to be the 95% empirical quantile of $h_1(S,t)$. The MM1-SU test uses a conditional linear programming algorithm to find its critical value function. The MM1-LU test uses a nonlinear optimization package. The supplement provides more details for each numerical algorithm.

The AR test has power considerably lower than the power envelope when instruments are both weak ($\lambda/k = 2$) and strong ($\lambda/k = 8$). The LM test does not perform well when instruments are weak, and its power function is not monotonic even when instruments are strong. These two facts about the AR and LM tests are well documented in the literature; see Moreira

(2003) and Andrews, Moreira, and Stock (2006a). The figure also reveals some salient findings for the tests based on the MM1 statistic. First, all MM1-based tests have correct size. Second, the bias of the MM1 similar test increases as the instruments get stronger. Hence, a naive choice for the density can yield a WAP test which can have overall poor power. We can remove this problem by imposing an unbiased condition when obtaining an optimal test. The MM1-SU test is easy to implement and has power closer to the power upper bound. When instruments are weak, its power lies moderately below the reported power envelope. This is expected as the number of parameters is too large[2]. When instruments are strong, its power is virtually the same as the power envelope.

To support the use of the MM1-SU test we also consider the MM1-LU test, which imposes a weaker unbiased condition. Close inspection of the graphs show that the derivative of the power function of the MM1 test is different from zero at $\beta = \beta_0$. This observation suggests that the power curve of the WAP test would change considerably if we were to force the power derivative to be zero at $\beta = \beta_0$. Indeed, we implement the MM1-LU test where the locally unbiased condition is true at only one point, the true parameter μ. This parameter is of course unknown to the researcher and this test is not feasible. However, by considering the locally unbiased condition for other values of the instruments' coefficients, the WAP test would be smaller –not larger. The power curves of MM1-LU and MM1-SU tests are very close, which shows that there is not much gain in relaxing the strongly unbiased condition.

The last two graphs plot the power curves for the three similar tests based on the MM2 statistic $h_2(s,t)$ as defined in Section 2. By using the density $h_2(s,t)$, we avoid the pitfalls for the MM1 test. In the supplement, we show that $h_2(s,t)$ is invariant to data transformations which preserve the two-sided hypothesis testing problem; see Andrews, Moreira, and Stock (2006a) for details on the sign-transformation group. Hence, the MM2 similar test is unbiased and has overall good power without imposing any additional unbiased conditions. The graphs illustrate this theoretical finding, as the MM2, MM2-SU, and MM2-LU tests have numerically the same power curves. This conclusion changes dramatically when the covariance matrix is no longer

[2]The MM1-SU power is nevertheless close to the two-sided power envelope for orthogonally invariant tests as in Andrews, Moreira, and Stock (2006a) (which is applicable to this design, but not reported here).

a Kronecker product.

Figure 3: Power Comparison (Non-Kronecker Covariance)

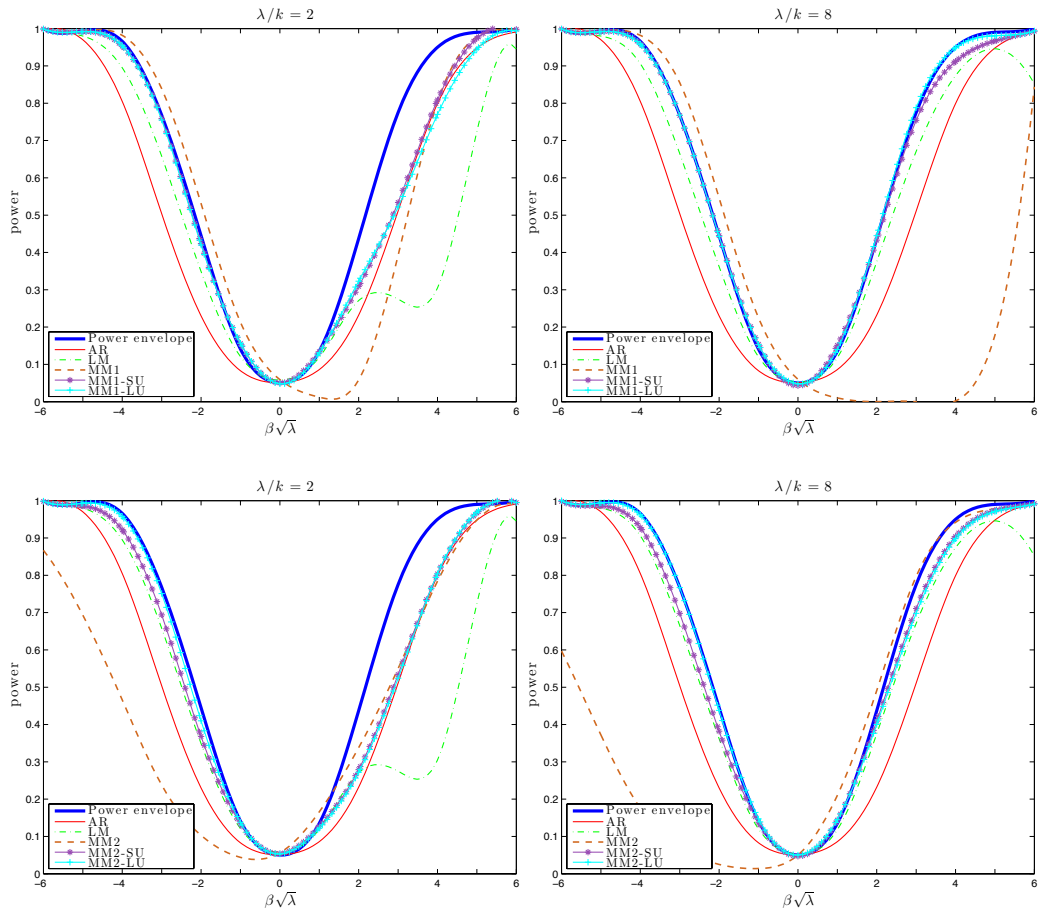


Figure 3 presents the power curves for all reported tests for the non-Kronecker design. Both MM1 and MM2 tests are severely biased and have overall bad power when instruments are strong. In the supplement, we show that in this design we cannot find a group of data transformations which preserve the two-sided testing problem. Hence, a choice for the density for the WAP test based on symmetry considerations is not obvious. The correct density choice can be particularly difficult due to the large parameter-dimension (the coefficients μ and covariance Σ). Instead, we can endogenize the weight choice so that the WAP test will be automatically unbiased. This is done

by the MM1-LU and MM2-LU tests. These two tests perform as well as the MM1-SU and MM2-SU tests. Because the latter two tests are easy to implement, we recommend their use in empirical practice.

5.2 Nearly Integrated Regressor

To evaluate rejection probabilities, we perform 1,000 Monte Carlo simulations following the design of Jansson and Moreira (2006). The disturbances ε_t^y and ε_t^x are serially iid, with variance one and correlation $\rho = \omega_{12}/\omega_{11}^{1/2}\omega_{22}^{1/2}$. We use 1,000 replications to find the Lagrange multipliers using linear programming (LP). The number of replications for LP is considerably smaller than what is recommended for empirical work. However, the Monte Carlo experiment attenuates the randomness for power comparisons. We refer the reader to MacKinnon (2006, p. S8) for a similar argument on the bootstrap.

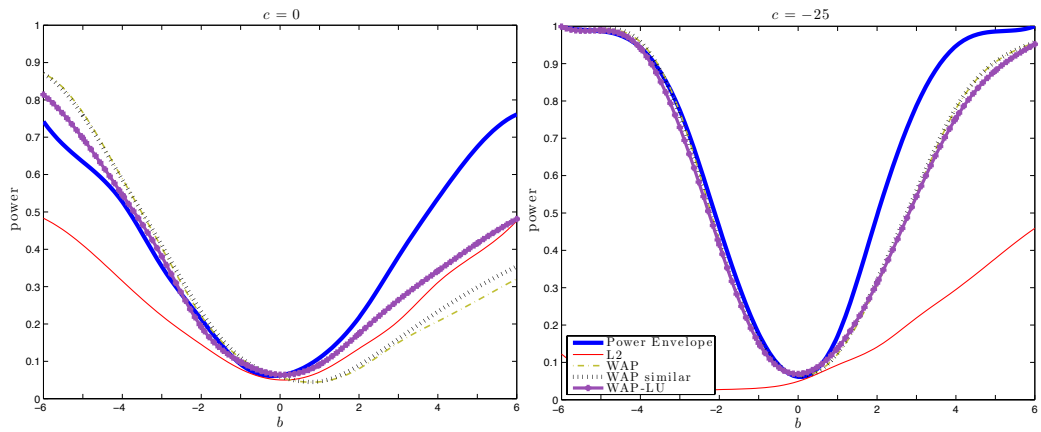
We consider three WAP tests based on the two-sided weighted average density MM-2S statistic. We present power plots for the WAP (size-corrected) test, the WAP similar test, and the WAP locally unbiased test (whose power derivative is zero at the null $\beta_0 = 0$). We choose 15 evenly-spaced boundary constraints for $\pi \in [0.5, 1]$. We compare the WAP tests with the L_2 test of Wright (2000) and a power envelope. The envelope is the power curve for the unfeasible UMPU test for the parameter β when the regressor's coefficient π *is known*.

The numerical simulations are done for $\rho = -0.5, 0.5$, $\gamma_N = 1 + c/N$ for $c = 0, -5, -10, -15, -25, -40$, and $\beta = b \cdot \omega_{11.2} g(\gamma_N)$ for $b = -6, -5, ..., 6$. The scaling function $g(\gamma_N) = \left(\sum_{i=1}^{N-1}\sum_{l=0}^{i-1}{\gamma_N}^{2l}\right)^{-1/2}$ allows us to look at the relevant power plots as γ_N changes. The value $b = 0$ corresponds to the null hypothesis $H_0 : \beta = 0$.

Figure 4 plots power curves for $\rho = 0.5$ and $c = 0, -25$. All other numerical results are available in the supplement to this paper.

When $c = 0$ (integrated regressor), the power curve of the L_2 test is considerably lower than the power envelope. The WAP size-corrected test has correct size but is highly biased. For negative b, its power is above the two-sided power envelope. For positive b, the WAP test has power considerably lower than the power upper bound. The WAP similar test decreases the bias and performs slightly better than the WAP size-corrected test. The power curve behavior of both tests near the null explains why those two WAP tests do not perform so well.

Figure 4: Power Comparison



The WAP-LU (locally unbiased) test removes the bias of the other two WAP test considerably and has very good power. We did not remove the bias completely because we implemented the WAP-LU test with only 15 points (its power is even slightly above the power envelope for unbiased tests for some negative values of b when $c = 0$). By increasing the number of boundary conditions, the power curve for $c = 0$ would be slightly smaller for negative values of b with power gains for positive values of b.

The WAP-LU test seems to dominate the L_2 test for most alternatives and has power closer to the power envelope. As c goes away from zero, all three WAP tests behave more similarly. When $c = -25$, their power is the same for all purposes. On the other hand, the bias of the L_2 test increases with power being close to zero for some alternatives far from the null. Overall, the WAP-LU test is the only test which is well-behaved regardless of whether or not the regressor is integrated. Hence, we recommend the WAP-LU test based on the MM-2S statistic for empirical work.

6 Conclusion

This paper considers tests which maximize the weighted average power (WAP). The focus is on determining WAP tests subject to an uncountable number of equalities and/or inequalities. The unifying theory allows us to obtain tests with correct size, similar tests, and unbiased tests, among others. Character-

ization of WAP tests is, however, a non-trivial task in our general framework. This problem is considerably more difficult to solve than the standard problem of maximizing power subject to size constraints.

We propose relaxing the original maximization problem and using continuity arguments to approximate the power function. The results obtained here follow from the Maximum Theorem of Berge (1997). Two main approaches are considered: discretization and uniform approximation. The first method considers a sequence of tests for an increasing number of boundary conditions. This approximation constitutes a natural and easy method of approximating WAP tests. The second approach builds a sequence of tests that approximate the WAP tests uniformly. Approximating equalities by inequalities implies that the resulting tests are weighted averages of the densities using regular additive measures. The problem is then analogous to finding least favorable distributions when maximizing power subject to size constraints.

We prefer not to give a firm recommendation on which constraints are reasonable for WAP tests to have (such as correct size, similarity on the boundary, unbiasedness, local unbiasedness, and monotonic power). However, our theory allows us to show that WAP similar tests are admissible for an important class of testing problems. Hence, we provide a theoretical justification for a researcher to seek a weighted average density so that the WAP test is approximately similar. Better yet, we do not need to blindly search for a correct weighted average density. A standard numerical algorithm can automatically adjust it for the researcher.

Finally, we apply our theory to the weak instrumental variable (IV) model with heteroskedastic and autocorrelated (HAC) errors and to the nearly integrated regressor model. In both models, we find WAP-LU (locally unbiased) tests which have correct size and overall good power.

7 Proofs

Proof of Proposition 1. Let $\phi_n \in \Gamma(g,\gamma)$ such that $\int \phi_n h \to \sup_{\phi\in\Gamma(g,\gamma)} \int \phi h$. We note that $\Gamma(g,\gamma)$ is contained in the unit ball on $\mathcal{L}_\infty(Y)$. By the Banach-Alaoglu Theorem, there exist a subsequence $\left(\phi_{n_k}\right)$ and a function $\overline{\phi}$ such that $\int \phi_{n_k} f \to \int \overline{\phi} f$ for every $f \in \mathcal{L}_1(Y)$. Trivially, we have $\int \overline{\phi} g_v \in [\gamma_v^1, \gamma_v^2]$, $\forall v \in \mathbf{V}$. Hence, $\overline{\phi} \in \Gamma(g,\gamma)$ solves $\sup_{\phi\in\Gamma(g,\gamma)} \int \phi h$. $\square$

Let $F(\phi) = \int \phi g_v$ be an operator defined on $\mathcal{L}_\infty(Y)$ into a space of real continuous functions defined on $\mathbf{V}$ and $\mathcal{G} = \{F(\phi); \phi \in \mathcal{L}_\infty(Y)\}$ its image. By the Dominated Convergence Theorem, $\mathcal{G}$ is a subspace (not necessarily topological) of $\mathcal{C}(\mathbf{V})$. Characterization of $\overline{\phi}$ relies on properties of the epigraph of h under $\mathbb{K}$:

$$[h, \mathbb{K}] = \left\{ (a, b)\,; b = F(\phi), \phi \in \mathbb{K}, a < \int \phi h \right\}.$$

Lemma A.1. *Suppose that there exists $\phi^o \in \Gamma(g, \gamma)$ such that $\int \phi^o h < \int \overline{\phi} h$. The following hold:*
(a) *The set $[h, \mathbb{K}]$ is also convex.*
(b) *The element $\left(\int \phi^o h, F(\overline{\phi})\right)$ is an internal point of $[h, \mathbb{K}]$.*
(c) *There exists a linear functional G^* defined on $\mathcal{G}$ such that*

$$\int \phi h + G^* \left(\int \phi g_v \right) \leq \int \overline{\phi} h + G^* \left(\int \overline{\phi} g_v \right)$$

for all $\phi \in \mathbb{K}$.

For a topological vector space $\mathcal{X}$, consider the dual space $\mathcal{X}^*$ of continuous linear functionals $\langle \phi, \phi^* \rangle$. The following result is useful to prove Lemma A.1.

Lemma A.2. *Let $\mathcal{X}$ be a topological vector space and $\mathbb{K} \subset \mathcal{X}$ be a convex set. Let $\phi^* \in \mathcal{X}^*$, $F : \mathcal{X} \to \mathcal{G}$ be a linear operator such that $\mathcal{G} = F(\mathcal{X})$ and $C \subset \mathcal{G}$ be a convex set. Consider the problem*

$$\sup_{\phi \in \mathbb{K}} \ \langle \phi, \phi^* \rangle \text{ where } F(\phi) \in C. \tag{7.26}$$

Suppose that there exist $\overline{\phi} \in \mathbb{K}$ which solves (7.26) and $\phi^\circ \in \mathbb{K}^\circ$ (i.e. an interior point of $\mathbb{K}$) such that $F(\phi^\circ) \in C$ and $\langle \phi^\circ, \phi^* \rangle < \left\langle \overline{\phi}, \phi^* \right\rangle$. Then, there exists a linear functional G^* defined in $\mathcal{G}$ such that

$$\langle \phi, \phi^* \rangle + \langle F(\phi), G^* \rangle \leq \left\langle \overline{\phi}, \phi^* \right\rangle + \left\langle F(\overline{\phi}), G^* \right\rangle,$$

for all $\phi \in \mathbb{K}$.

Proof of Lemma A.2. Define $\mathbb{F} = \{(a, b) \in \mathbb{R} \times \mathcal{G}; a < \langle \phi, \phi^* \rangle$ and $b = F(\phi)$ for some $\phi \in \mathbb{K}\}$. The set $\mathbb{F}$ is trivially a convex set.

Since $\phi^\circ \in \mathbb{K}^\circ$ and $\phi^* \in \mathcal{X}^*$, for every $\epsilon > 0$, there is an open set $\mathbb{U} \subset \mathbb{K}$ such that $\phi^\circ \in \mathbb{U}$ and $\langle \phi, \phi^* \rangle > \langle \phi^\circ, \phi^* \rangle - \epsilon$, for all $\phi \in \mathbb{U}$. Let $\epsilon > 0$ be sufficiently small such that $\left\langle \overline{\phi}, \phi^* \right\rangle - \epsilon > \langle \phi^\circ, \phi^* \rangle$. For fixed $t \in (1/2, 1)$, let $\mathbb{B} = t\overline{\phi} + (1-t)\mathbb{U}$. The set $\mathbb{B}$ is an open subset of $\mathbb{K}$, since $\mathbb{K}$ is convex and $\overline{\phi} \in \mathbb{K}$. Since F is linear, $tF(\overline{\phi}) + (1-t)F(\phi^\circ) = G_t \in C$. We claim that G_t is an internal point of $F(\mathbb{B})$. Indeed, given $y \in \mathcal{G}$, there exists $x \in \mathcal{X}$ such that $y = F(x)$. Since $\mathbb{U}$ is an open set, there exists $\lambda > 0$ such that $\phi^\circ + \lambda x \in \mathbb{U}$ and, consequently, $t\overline{\phi} + (1-t)(\phi^\circ + \lambda x) \in \mathbb{B}$. Hence, the linearity of F implies that

$$G_t + (1-t)\,\lambda y = F(t\overline{\phi} + (1-t)(\phi^\circ + \lambda x)) \in F(\mathbb{B}).$$

Moreover,

$$\begin{aligned}
\left\langle t\overline{\phi} + (1-t)\phi, \phi^* \right\rangle &= t\left\langle \overline{\phi}, \phi^* \right\rangle + (1-t)\left\langle \phi, \phi^* \right\rangle \\
&> t(\langle \phi^\circ, \phi^* \rangle + \epsilon) + (1-t)(\langle \phi^\circ, \phi^* \rangle - \epsilon) \\
&= \langle \phi^\circ, \phi^* \rangle + (2t-1)\epsilon \\
&> \langle \phi^\circ, \phi^* \rangle,
\end{aligned}$$

for all $\phi \in \mathbb{U}$. These trivially imply that $(\langle \phi^\circ, \phi^* \rangle, G_t)$ is an internal point of $\mathbb{F}$.

The set $\left\{ (\left\langle \overline{\phi}, \phi^* \right\rangle, G_t); t \in (1/2, 1] \right\}$ is convex and does not intercept $\mathbb{F}$ by the definition of $\overline{\phi}$. By the basic separation theorem (see Dunford and Schwartz (1988, p. 412)), there exist $\kappa \in \mathbb{R}$ and a linear functional G^* defined in $\mathcal{G}$ such that for all $(a, b) \in \mathbb{F}$ and $t \in (1/2, 1]$

$$\kappa a + \langle b, G^* \rangle \leq \kappa \left\langle \overline{\phi}, \phi^* \right\rangle + \langle G_t, G^* \rangle.$$

We claim that $\kappa > 0$. First, $\kappa < 0$ would lead to a contradiction with the previous inequality given that a can be arbitrarily negative. If $\kappa = 0$, then the previous inequality becomes

$$\langle F(\phi), G^* \rangle \leq \langle G_t, G^* \rangle,$$

for all $\phi \in \mathbb{K}$. Since G_t is an internal point of $F(\mathbb{K})$ for some $t \in (1/2, 1)$, then the previous inequality would imply that the linear functional G^* would be null, which contradicts the basic separation theorem.

Normalizing $\kappa = 1$ and taking $a = \langle \phi, \phi^* \rangle$, $b = F(\phi)$ and taking $t = 1$ we obtain the desired result, once $G_1 = F(\overline{\phi})$. $\square$

Proof of Lemma A.1. Part (a) follows from Proposition 2 of Section 7.8 of Luenberger (1969) because $\int \phi h$ is linear, hence concave, in ϕ. Defining $\mathcal{X} = \mathcal{L}_\infty(Y)$, $F(\phi) = \int \phi g_v$, $\mathbb{K} = \{\phi \in \mathcal{X}; 0 \leq \phi \leq 1\}$ and $C = \{f; f_v \in [\gamma_v^1, \gamma_v^2], v \in \mathbf{V}\}$. Parts (b) and (c) follow from Lemma A.2. $\square$

Lemma A.1 uses the fact that the set of bounded measurable functions is a vector space. However, it does not require any topology for the vector space $\mathbb{R} \times \mathcal{G}$ in which $[h, \mathbb{K}]$ is contained. The difficulty arises in transforming the internal point $\left(\int \phi^o h, G_t\right)$ into an interior point of $[h, \mathbb{K}]$. Even if we were able to find such a topology, characterization of the linear functional G^* (and consequently of $\overline{\phi}$) may not be trivial.

Proof of Corollary 1. Parts (a)-(c) follow directly from Theorem 3.6.1 of Lehmann and Romano (2005). Part (d) follows from Lemma 2 for $\mathcal{G} = \mathbb{R}^m$. The result now follows trivially. $\square$

Proof of Lemma 1. For part (a), we need to show that Γ is both upper semi-continuous (u.s.c.) and lower semi-continuous (l.s.c.).

Since $\mathbb{K}$ is compact in the weak* topology, upper semi-continuity of Γ is equivalent to the closed graph property; see Berge (1997, p. 112). With a slight abuse of notation, we use n to index nets. Let (g^n, γ^n, ϕ_n) be a net such that $\phi_n \in \Gamma(g^n, \gamma^n)$ and $(g^n, \gamma^n, \phi_n) \to (g, \gamma, \phi)$, where $\phi_n \to \phi$ in the weak* topology sense. Notice that

$$\begin{aligned} \left| \int \phi g_v - \int \phi_n g_v^n \right| &\leq \left| \int (\phi - \phi_n) g_v \right| + \left| \int \phi_n (g_v - g_v^n) \right| \\ &\leq \left| \int (\phi - \phi_n) g_v \right| + \int |g_v - g_v^n|. \end{aligned} \tag{7.27}$$

Since $\phi_n \to \phi$ in the weak* topology, $\phi \in \mathbb{K}$ and $\left|\int (\phi - \phi_n) g_v\right| \to 0$ and since $g_v^n \to g_v$ in the $\mathcal{L}_1$, $\int |g_v - g_v^n| \to 0$, for every $v \in \mathbf{V}$. Since $\gamma_v^n \to \gamma_v$ and $\int \phi_n g_v^n \in [\gamma_v^{1,n}, \gamma_v^{2,n}]$ for all $v \in \mathbf{V}$ and n, we have that $\int \phi g_v \in [\gamma_v^1, \gamma_v^2]$, for all $v \in \mathbf{V}$, i.e., $\phi \in \Gamma(g, \gamma)$ which proves the closed graph property.

It remains to show that Γ is l.s.c. Let G be a weak* open set such that $G \cap \Gamma(g, \gamma) \neq \varnothing$. We have to show that there exists a neighborhood $\mathbb{U}(g, \gamma)$ of (g, γ) such that $G \cap \Gamma(\widetilde{g}, \widetilde{\gamma}) \neq \varnothing$, for all $(\widetilde{g}, \widetilde{\gamma}) \in \mathbb{U}(g, \gamma)$. Suppose that this is not the case. Then there exists a net $g_v^n \to g_v$ in the $\mathcal{L}_1$ sense and $\gamma_v^n \to \gamma_v$ pointwise a.e. $v \in \mathbf{V}$ such that $G \cap \Gamma(g^n, \gamma^n) = \varnothing$, for all n. Take

$\phi \in G \cap \Gamma(g,\gamma)$. Now define $\phi_n \in \Gamma(g^n,\gamma^n)$ a point of minimum distance from ϕ in $\Gamma(g^n,\gamma^n)$ according to a given metric on $\mathbb{K}$ equivalent to the weak* topology (notice that the weak* topology is metrizable on $\mathbb{K}$). There exists a subnet of (ϕ_n) which converges in weak* sense to $\widetilde{\phi} \in \mathbb{K}$ (because $\mathbb{K}$ is weak* compact). Passing to this subnet, for a.e. $v \in \mathbf{V}$,

$$[\gamma_v^{1,n}, \gamma_v^{2,n}] \ni \int \phi_n g_v^n = \int \phi_n g_v + \int \phi_n (g_v - g_v^n) \to \int \widetilde{\phi} g_v$$

because $g_v^n \to g_v$ in the $\mathcal{L}^1$ sense for every $v \in \mathbf{V}$ and (ϕ_n) is bounded. Since $\gamma_v^n \to \gamma_v$, we have that $\widetilde{\phi} \in \Gamma(g,\gamma)$. By construction, ϕ_n must converge (in the weak* sense) to $\phi \in \Gamma(g,\gamma)$, i.e., $\widetilde{\phi} = \phi$. Thus, for n sufficiently large, $\phi_n \in G \cap \Gamma(g^n,\gamma^n)$. However, this contradicts the hypothesis that $G \cap \Gamma(g^n,\gamma^n) = \varnothing$, for all n.

For part (b), by hypothesis $\Gamma(g,\gamma) \neq \emptyset$ for all (g,γ). The functional $\phi \to \int \phi h$ is continuous in the weak* topology. The result now follows from the Maximum Theorem of Berge (1997, p. 116). □

Proof of Theorem 1. For part (a), the result follows from continuity in v.

For part (b), fix $v \in \mathbf{V}$ in which $g_v^n(y) \to g_v(y)$ for a.e. $y \in Y$ and $\int \sup_n |g_v^n| < \infty$. As $n \to \infty$,

$$\int |g_v^n - g_v| \to 0$$

by the Dominated Convergence Theorem.

For part (c), convergence of $\int \overline{\phi}_n h$ to $\int \overline{\phi} h$ follows directly from Lemma 1. Convergence of the power function $\int \overline{\phi}_n g$ also follows from Lemma 1 if every convergent subsequence of (ϕ_n) converges to $\overline{\phi}$. By Lemma 1 (b), this subsequence should converge to a point in $\Gamma_M(g,\gamma)$ and, by hypothesis, $\Gamma_M(g,\gamma) = \{\overline{\phi}\}$ which implies the claim. Suppose, by contradiction, that the sequence (ϕ_n) does not converge to $\overline{\phi}$. Hence, there exists a neighborhood $\mathbb{U}$ of $\overline{\phi}$ in the weak* topology and subsequence (ϕ_{n_k}) in the complement of $\mathbb{U}$. Since this subsequence is bounded, we can find a convergent subsequence of it. However, this limit point is different from $\overline{\phi}$ and the resulting subsequence is a subsequence of (ϕ_n). This, however, contradicts the initial claim. □

Proof of Lemma 2. Part (a) is an immediate consequence of the Banach-Alaoglu Theorem.

Part (b) follows from Theorem 1 in Luenberger (1969, p. 217). Define in Luenberger's (1969) notation: $X = \mathcal{L}_\infty(Y)$, $\Omega = \mathbb{K}$, $Z = \mathcal{C}(\mathbf{V}) \times \mathcal{C}(\mathbf{V})$ (the space of continuous and bounded real functions on $\mathbf{V}$), P is the set of nonnegative functions of Z, $f(\phi) = -\int \phi h$ and $G(\phi) = (\gamma_v^1 - \delta - \int \phi g_v, \int \phi g_v - \gamma_v^2 - \delta)$. Let $\phi^o \in \Gamma(g, \gamma)$. We only need to observe that $G(\phi^o) < 0$ and the dual of Z is $rca(\mathbf{V}) \times rca(\mathbf{V})$; see Dunford and Schwartz (1988, p. 376). □

Proof of Lemma 3. Let us first prove that Γ is u.s.c. Since $\mathbb{K}$ is compact in the weak* topology, u.s.c. of Γ is equivalent to the closed graph property; see Berge (1997, p. 112).

Let (δ_n, ϕ_n) be a net such that $\phi_n \in \Gamma(g, \gamma, \delta_n)$ and $(\delta_n, \phi_n) \to (\delta, \phi)$, where $\phi_n \to \phi$ in the weak* topology sense. Since $\phi_n \to \phi$ in the weak* topology, $\phi \in \mathbb{K}$ and $\left|\int (\phi - \phi_n) g_v\right| \to 0$, a.e. Therefore, $\int \phi_n g_v \in [\gamma_v^1 - \delta_n, \gamma_v^2 + \delta_n]$ for all $v \in \mathbf{V}$ (because $g_v(y)$ is a continuous function in v for each y) and n implies that $\int \phi g_v \in [\gamma_v^1 - \delta, \gamma_v^2 + \delta]$, i.e., $\phi \in \Gamma(g, \gamma, \delta)$ which proves the closed graph property.

Let us now prove that Γ is l.s.c. at δ. Let G be a weak* open set such that $G \cap \Gamma(g, \gamma, \delta) \neq \varnothing$. We have to show that there exists a neighborhood $\mathbb{U}(\delta)$ of δ such that $G \cap \Gamma(g, \gamma, \widetilde{\delta}) \neq \varnothing$, for all $\widetilde{\delta} \in \mathbb{U}(\delta)$. Suppose that this is not the case. Then, there exists a sequence $\delta_n \to \delta$ such that $G \cap \Gamma(g, \gamma, \delta_n) = \varnothing$, for all n. Take $\phi \in G \cap \Gamma(g, \gamma, \delta)$. Now define $\phi_n \in \Gamma(g, \gamma, \delta_n)$ as a point of minimum distance from ϕ in $\Gamma(g, \gamma, \delta_n)$ according to a given metric on $\mathbb{K}$ equivalent to the weak* topology (notice that the weak* topology is metrizable on $\mathbb{K}$). There exists a subsequence of (ϕ_n) that converges in weak* sense to $\widetilde{\phi} \in \mathbb{K}$ (because $\mathbb{K}$ is weak* compact). Since for a.e. $v \in$ and all n

$$\int \phi_n g_v \in [\gamma_v^1 - \delta_n, \gamma_v^2 + \delta_n],$$

taking the limit to this subsequence, we get $\widetilde{\phi} \in \Gamma(g, \gamma, \delta)$. However, by construction ϕ_n must then converge (in the weak* sense) to $\phi \in \Gamma(g, \gamma, \delta)$, i.e., $\widetilde{\phi} = \phi$ a.e. Thus, for a sufficiently large n, $\phi_n \in G \cap \Gamma(g, \gamma, \delta_n)$. However, this contradicts the hypothesis that $G \cap \Gamma(g, \gamma, \delta_n) = \varnothing$, for all n.

For part (b), since $\Gamma(g, \gamma, \delta) \neq \varnothing$, this theorem is an immediate consequence of the Maximum Theorem; see Berge (1997, p. 116). □

Proof of Theorem 2. The proof is similar to that of Theorem 1 (c). □

Proof of Theorem 3. Under assumption U-BD, for each finite regular counting additive measure Λ,

$$\int f_v(y)\Lambda(dv)$$

is a pointwise limit of a sequence of uniformly bounded analytic functions a.e. on $\mathbb{R}^m$. By the Generalized Vitali Theorem, it is an analytic function as well; see Dunford and Schwartz (1988, p. 228) and Gunning and Rossi (1965, p. 11).

Suppose now that there exists a positive Lebesgue measurable set D in $\mathbb{R}^m$ such that for all $y \in D$

$$h(y) = c_\delta(y), \tag{7.28}$$

where $c_\delta(y)$ is defined in expression (4.25). Since the functions h and c_δ are analytic, $h - c_\delta = 0$ in $\mathbb{R}^m$.

Indeed, the case $m = 1$ is straightforward since D has at least one cumulative point (in fact, there are infinitely many such points) which immediately implies the result. Suppose that $m = 2$. For each $y_1 \in \mathbb{R}$, define $D_{y_1} = \{y_2; (y_1, y_2) \in D\}$. The set D has a positive Lebesgue measure in $\mathbb{R}^2$. Hence the set of y_1 such that D_{y_1} has a positive Lebesgue measure also has a positive Lebesgue measure. For each such y_1, we know that $(h - c_\delta)(y_1, y_2)$ is an analytic function of y_2 and is identical to zero in the positive measure set D_{y_1}. Therefore, $(h - c_\delta)(y_1, z_2) = 0$ in the domain of the holomorphic extension of the second complex variable when the first is fixed at y_1, which has a positive Lebesgue measure in $\mathbb{C}$ (or $\mathbb{R}^2$). Interchanging the places of y_1 and y_2 and making the same argument, we are able to build a positive measure set in $\mathbb{C}^2$ such that $h - c_\delta$ is null. From Theorem 3.7 of Range (1986) this equality must hold for all $y \in \mathbb{R}^m$. The proof is analogous for all $m > 2$.

By the necessary conditions of Lemma 2, $supp(\Lambda^+) \subset \mathbb{V}_-$ and $supp(\Lambda^-) \subset \mathbb{V}_+$, where

$$\mathbb{V}_- = \{v \in \mathbf{V}; \int \overline{\phi} f_v = \alpha - \delta\} \text{ and } \mathbb{V}_+ = \{v \in \mathbf{V}; \int \overline{\phi} f_v = \alpha + \delta\}$$

are disjoint sets. The optimal test is not trivial and cannot be identical to $\alpha - \delta$. Hence, the sets $\mathbb{V}_-$ and $\mathbb{V}_+$ cannot both be of zero measure. Indeed, suppose that $\mathbb{V}_-$ has positive measure (the case in which $\mathbb{V}_+$ has positive measure is analogous). Since the optimal test is not trivial,

$$\int (\alpha + \delta)\, h < \int \overline{\phi} h.$$

Substituting (7.28) into the previous expression:

$$(\alpha+\delta)\int\int_{\mathbb{V}} f_v(y)\Lambda(dv) < \int \overline{\phi}\int_{\mathbb{V}} f_v(y)\Lambda(dv).$$

Since $\int \overline{\phi} f_v = \alpha-\delta$ on $\mathbb{V}_-$ and $\int \overline{\phi} f_v = \alpha+\delta$ on $\mathbb{V}_+$, using Fubini's Theorem

$$(\alpha+\delta)\int_{\mathbb{V}} \Lambda(dv) < \int_{\mathbb{V}} \left[\int \overline{\phi} f_v(y)\right] \Lambda(dv) = (\alpha-\delta)\int_{\mathbb{V}_-} \Lambda(dv) + (\alpha+\delta)\int_{\mathbb{V}_+} \Lambda(dv)$$

which is a contradiction. □

Proof of Theorem 4. Farrell (1968a) and Farrell (1968b) consider a more concrete version of the Stein's (1955) necessary and sufficient condition for admissibility. We follow Farrell's approach here to prove our admissibility result. For more details on this topic, see Subsection 8.9 of Berger (1985). First, we need the following lemma for the proof of Theorem 4.

Lemma A.3. *For each $\delta > 0$, there exists a sequence of Bayes tests $\left(\overline{\phi}_{\delta,n}\right)$ which converges pointwisely (and therefore weakly) to $\overline{\phi}_\delta$.*

Proof of Lemma A.3. Let $\delta > 0$. From Lemma 2 for $\gamma_v^1 = \alpha-\delta$ and $\gamma_v^2 = \alpha+\delta$, there are positive *rca* measures Λ_δ^- and Λ_δ^+ on the compact $\mathbb{V}_0$ such that

$$\overline{\phi}_\delta(y) = \begin{cases} 1, \text{ if } \int_{\mathbb{V}_1} f_v(y)\Lambda_1(dv) + \int_{\mathbb{V}_0} f_v(y)\Lambda_\delta^-(dv) > \int_{\mathbb{V}_0} f_v(y)\Lambda_\delta^+(dv) \\ 0, \text{ if } \int_{\mathbb{V}_1} f_v(y)\Lambda_1(dv) + \int_{\mathbb{V}_0} f_v(y)\Lambda_\delta^-(dv) < \int_{\mathbb{V}_0} f_v(y)\Lambda_\delta^+(dv) \end{cases}$$

is the optimal test for problem (2.21). Notice that Λ_δ^+ cannot be zero, otherwise we have a contradiction with the optimality of $\overline{\phi}_\delta$. From Theorem 2, $(\overline{\phi}_\delta)$ weakly converges to $\overline{\phi}$ when $\delta \to 0$.

Let (β_n) be a sequence in $\mathbb{B}-\{\beta_0\}$ converging to β_0. Define the following sequence of measures $(\Lambda_{\delta,n}^-)$ with support on $\mathbb{V}_1$. For each $n \in \mathbb{N}$ and $\mathbb{B}_1$ a Borel set in $\mathbb{B}\times\mathbb{P}$, define

$$\Lambda_{\delta,n}^-(\mathbb{B}_1) = \begin{cases} \Lambda_\delta^-(\{\beta_0\}\times\mathbb{P}), \text{ if } \mathbb{B}_1 = \{\beta_n\}\times\mathbb{P} \\ 0, \text{ if otherwise.} \end{cases}$$

It is easy to see that the sequence $(\Lambda^-_{\delta,n})$ has support in $\mathbb{V}_1$ and weakly converges to Λ^-_δ. Define $\overline{\phi}_{\delta,n}$ as $\overline{\phi}_\delta$ by substituting Λ^-_δ for $\Lambda^-_{\delta,n}$ in the above expression of $\overline{\phi}_\delta$. Normalizing these measures, for each $\delta > 0$ and $n \in \mathbb{N}$, there exist a positive constant $\kappa_{\delta,n}$ and probability distribution measures $\Lambda^-_{\delta,n}$ and Λ^+_δ with support in $\mathbb{V}_1$ and $\mathbb{V}_0$ such that

$$\overline{\phi}_{\delta,n}(y) = \begin{cases} 1, \text{ if } h_1^{\delta,n}(y) > \kappa_{\delta,n} h_0^\delta(y) \\ 0, \text{ if } h_1^{\delta,n}(y) < \kappa_{\delta,n} h_0^\delta(y) \end{cases},$$

where $h_1^{\delta,n}(y) = \int_{\mathbb{V}_1} f_v(y)\Lambda^-_{\delta,n}(dv)$ and $h_0^\delta(y) = \int_{\mathbb{V}_0} f_v(y)\Lambda^+_\delta(dv)$.

We define now the functions

$$\begin{aligned} c_n(y) &= \int_{\mathbb{V}_1} f_v(y)\Lambda_1(dv) + \int_{\mathbb{V}_0} f_v(y)\Lambda^-_{\delta,n}(dv), \\ c(y) &= \int_{\mathbb{V}_1} f_v(y)\Lambda_1(dv) + \int_{\mathbb{V}_0} f_v(y)\Lambda^-_{\delta}(dv), \text{ and} \\ d(y) &= \int_{\mathbb{V}_0} f_v(y)\Lambda^+_{\delta}(dv). \end{aligned}$$

From Theorem 3 $\overline{\phi}_\delta$ is nonrandomized, hence $\overline{\phi}_\delta = I\,(c > d)$. Since $\Lambda^-_{\delta,n}$ weakly converges to Λ^-_δ, we have that $c_n(y)$ pointwisely converges to $c(y)$. Therefore, for any compact $\mathbb{A} \subset Y$ such that $c(y) > d(y)$ for all $y \in \mathbb{A}$, we have that $c_n(y) > d(y)$, for all $y \in \mathbb{A}$ and n sufficiently large (because $c(\cdot)$ and $c_n(\cdot)$ are continuous functions). Analogously, for any compact $\mathbb{A} \subset Y$ such that $c(y) < d(y)$ for all $y \in \mathbb{A}$, we have that $c_n(y) < d(y)$, for all $y \in \mathbb{A}$ and n sufficiently large. Since in the set $\{c = d\}$ we only have isolated points, $\overline{\phi}_{\delta,n} = I\,(c_n > d)$ pointwisely converge to $\overline{\phi}_\delta = I\,(c > d)$. $\square$

Proof of Theorem 4 (cont.). We now show that there exists a sequence of tests with Neyman structure that weakly converge to an optimal test for problem (2.1).

For each m, the previous lemma implies that the sequence $\left(\overline{\phi}_{1/m,n}\right)$ pointwisely converges to $\overline{\phi}_{1/m}$ when $n \to \infty$. Hence, from the definition of $\overline{\phi}_{1/m}$ we can find $n\,(m)$ sufficiently large such that the sequence of tests $\left(\overline{\phi}_{1/m,n(m)}\right)$ has a Neyman structure and satisfies

$$\left| \int \left(\overline{\phi}_{1/m,n(m)} - \overline{\phi}\right) h \right| < \frac{1}{m} \text{ and}$$

$$\int \overline{\phi}_{1/m,n(m)} f_v \in \left[\alpha - \frac{2}{m}, \alpha + \frac{2}{m}\right], \text{ for all } v \in \mathbb{V}_0.$$

Since this sequence is in $\mathbb{K}$, by the Banach-Alaoglu theorem, there exists a subsequence weakly converging to a point in $\mathbb{K}$. Considering a convergent subsequence, without loss of generality, we can assume that the sequence is convergent to a test $\phi^\circ \in \mathbb{K}$. We claim that this limit point must be an optimal test for problem (2.1). Indeed, taking the limit when $m \to \infty$, we have that

$$\int \phi^\circ f_v = \alpha, \text{ for all } v \in \mathbb{V}_0 \text{ and } \int \phi^\circ h = \int \overline{\phi} h,$$

that is, ϕ° is a feasible test for problem (2.1) and provides the same maximum value.

Suppose by contradiction that $\overline{\phi}$ is inadmissible. Let $\phi \in \mathbb{K}$ be a test such that $\int \phi f_v \leq \alpha = \int \overline{\phi} f_v$, for all $v \in \mathbb{V}_0$ and

$$\int \overline{\phi} f_v \leq \int \phi f_v$$

for all $v \in \mathbb{V}_1$ with strict inequality for some $v \in \mathbb{V}_1$. For a given $\eta > 0$, define $\widetilde{\phi} = [\phi - \eta]^+ \equiv \max\{\phi - \eta, 0\} \in \mathbb{K}$. Since $f_v > 0$ and continuous at $v \in \mathbb{V}_0$, for each $\eta > 0$, there exists $\epsilon_0 > 0$ such that $\int \widetilde{\phi} f_v \leq \alpha - \epsilon_0$, for all $v \in \mathbb{V}_0$. There exist $\widetilde{\epsilon} > 0$ and a closed neighborhood $\overline{\mathbb{U}} \subset \mathbb{V}_1$ such that $\overline{\mathbb{U}} \cap \mathbb{V}_0 = \emptyset$ and

$$\int \overline{\phi} f_v + \widetilde{\epsilon} < \int \phi f_v, \forall v \in \overline{\mathbb{U}}.$$

Taking n sufficiently large we have that $\overline{\mathbb{U}} \cap (\{\beta_n\} \times \mathbb{P}) = \emptyset$ and integrating with respect to the probability measure $\Lambda^-_{\delta,n}$ we get

$$\int \overline{\phi} h_1^{\delta,n} + \widetilde{\epsilon} \Lambda_1(\overline{\mathbb{U}}) < \int \phi h_1^{\delta,n}.$$

Since Λ_1 has full support on $\mathbb{V}_1$, for each $0 < \epsilon_1 < \widetilde{\epsilon} \Lambda_1(\overline{\mathbb{U}})$, we can find $\eta > 0$ sufficiently small such that

$$\int \overline{\phi} h_1^{\delta,n} + \epsilon_1 < \int \widetilde{\phi} h_1^{\delta,n}.$$

For each $\epsilon_1 > 0$, for sufficiently large m we have that

$$\int \overline{\phi}_{1/m,n(m)} h_1^{1/m,n(m)} < \int \overline{\phi} h_1^{1/m,n(m)} + \epsilon_1.$$

For each $\epsilon_0 > 0$, for sufficiently large $n\,(m)$ we have that

$$\int \overline{\phi}_{1/m,n(m)} h_0^{1/m} > \alpha - \epsilon_0.$$

Take $m \in \mathbb{N}$ sufficiently large and choose $\eta > 0$ such that

$$\int \overline{\phi} h_1^{1/m,n(m)} + \epsilon_1 < \int \widetilde{\phi} h_1^{1/m,n(m)}.$$

Choosing an even bigger m, we have

$$\int \widetilde{\phi} h_0^{1/m} \leq \alpha - \epsilon_0 < \int \overline{\phi}_{1/m,n(m)} h_0^{1/m} \text{ and}$$
$$\int \overline{\phi}_{1/m,n(m)} h_1^{1/m,n(m)} < \int \overline{\phi} h_1^{1/m,n(m)} + \epsilon_1 < \int \widetilde{\phi} h_1^{1/m,n(m)}.$$

However, the test $\overline{\phi}_{1/m,n(m)}$ is a Neyman-Pearson test for the null pdf $h_0^{1/m}$ against the pdf $h_1^{1/m,n(m)}$. Hence, we obtain a contradiction. □

References

Anderson, T. W. (2011): "Optimal Significance Tests in Simultaneous Equations Models," *Unpublished Manuscript, Stanford University.*

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K. (2012): "Similar-On-The-Boundary Tests for Moment Inequalities Exist, But Have Poor Power," Discussion Paper 1815R, Yale University.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006a): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752.

——— (2006b): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752, Supplement.

——— (2008): "Optimal Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 146, 241–254.

Berge, C. (1997): *Topological Spaces*. Dover Books on Mathematics.

Berger, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*. Second Edition. Springer Series in Statistics.

Besicovitch, A. S. (1961): "On Diagonal Values of Probability Vectors of Infinitely Many Components," *Proceedings of Cambridge Philosophical Society*, 57, 759–766.

Chan, N. H., and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics*, 15, 1050–1063.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Admissible Invariant Similar Tests for Instrumental Variables Regression," *Econometric Theory*, 25, 806–818.

Chiburis, R. C. (2009): "Approximately Most Powerful Tests for Moment Inequalities," Working paper, University of Texas.

Dantzig, G. B. (1963): *Linear Programming and Extensions*. New Jersey: Princeton University Press.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

Dunford, N., and J. T. Schwartz (1988): *Linear Operators Part I: General Theory*. Wiley Classics Library. John Wiley and Sons.

Farrell, R. H. (1968a): "On Necessary and Sufficient Condition for Admissibility," *The Annals of Mathematical Statistics*, 38, 23–28.

——— (1968b): "Towards a Theory of Generalized Bayes Tests," *The Annals of Mathematical Statistics*, 38, 1–22.

Gunning, R., and H. Rossi (1965): *Analytic Functions of Several Complex Variables*. Englewood Cliffs, NJ: Prentice-Hall.

Jansson, M., and M. J. Moreira (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74, 681–715.

Krafft, O., and H. Witting (1967): "Optimale Tests and ungünstigsten Verteilungen," *Zeitschrift für Wahrscheinlinchkeitstheorie und verwandte Gebiete*, 7, 289–302.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Third edn., Springer.

Linnik, J. V. (2000): *Statistical Problems with Nuisance Parameters*. Translations of Mathematical Monographs, vol. 20.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*. John Wiley and Sons, Inc.

MacKinnon, J. G. (2006): "Bootstrap Methods in Econometrics," *The Economic Record (Special Issue)*, 82, S2–S18.

Mills, B., M. J. Moreira, and L. P. Vilela (2013): "Tests Based on t-Statistics for IV Regression with Weak Instruments," Unpublished Manuscript, FGV/EPGE.

Moreira, H., and M. J. Moreira (2011): "Inference with Persistent Regressors," *Unpublished Manuscript*, Columbia University.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Müller, U. K., and M. Watson (2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81.

Perlman, M. D., and L. Wu (1999): "The Emperor's New Tests," *Statistical Science*, 14, 355–369.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

Polak, E. (1997): *Optimization: Algorithms and Consistent Approximations*. New York: Springer Verlag.

Range, R. M. (1986): *Holomorphic Functions and Integral Representations in Several Complex Variables*. Graduate Texts in Mathematics.

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

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stein, C. (1955): "A Necessary and Sufficient Condition for Admissibility," *The Annals of Mathematical Statistics*, 26, 518–522.

Van Loan, C., and N. Ptsianis (1993): "Approximation with Kronecker Products," in *Linear Algebra for Large Scale and Real-Time Applications*, ed. by M. Moonen, and G. Golub, pp. 293–314. Springer.

Wright, J. H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business and Economic Statistics*, 18, 211–222.

Supplement to "Contributions to the Theory of Optimal Tests"

Humberto Moreira and Marcelo J. Moreira
FGV/EPGE

This version: September 10, 2013

1 Introduction

This paper contains supplemental material to Moreira and Moreira (2013), hereafter MM.

Section 2 provides details for the tests for the HAC-IV model. We derive both WAP (weighted average power) MM1 and MM2 statistics presented in the paper. We discuss different requirements for tests to be unbiased. We show that the locally unbiased (LU) condition is less restrictive than the strongly unbiased (SU) condition. We implement numerically the WAP tests using approximation (MM similar tests), non-linear optimization (MM-LU tests), and conditional linear programming (MM-SU tests) methods. Appendix A contains all numerical simulations for the Anderson and Rubin (1949), score, and MM tests. Based on power comparisons, we recommend the MM1-SU and MM2-SU tests in empirical practice.

Section 3 derives one-sided and two-sided WAP for the nearly integrated model. We show how to carry out these tests using linear programming algorithms. Moreira and Moreira (2011) compare one-sided tests, including a similar t-test, the UMPCU test of Jansson and Moreira (2006), and the refined Bonferroni test of Campbell and Yogo (2006). Appendix B presents power curves for two-sided tests, including the L_2 test of Wright (2000) and three WAP (similar, correct size, and locally unbiased) tests based on the two-sided MM-2S statistic. We recommend the WAP-LU (locally unbiased) test based on the MM-2S statistic.

Section 4 approximates the WAP test by a sequence of tests in a Hilbert space. We can fully characterize the approximating tests since they are equivalent to distance minimization for closed and convex sets. The power function of this sequence of optimal tests converges uniformly to the WAP test. The implementation method for a smaller class of tests is readily available.

Section 5 derives the score test in the HAC-IV model and provides proofs for all results presented in this supplement.

2 HAC-IV

The statistics S and T are independent and have distribution

$$\begin{aligned} S &\sim N\left(\left(\beta-\beta_0\right)C_{\beta_0}\mu, I_k\right) \text{ and } T \sim N\left(D_\beta\mu, I_k\right), \text{ where} \\ C_{\beta_0} &= \left[\left(b_0' \otimes I_k\right)\Sigma\left(b_0 \otimes I_k\right)\right]^{-1/2} \text{ and} \\ D_\beta &= \left[\left(a_0' \otimes I_k\right)\Sigma^{-1}\left(a_0 \otimes I_k\right)\right]^{-1/2}\left(a_0' \otimes I_k\right)\Sigma^{-1}\left(a \otimes I_k\right). \end{aligned} \tag{2.1}$$

The density $f_{\beta,\mu}(s,t)$ is given by

$$\begin{aligned} f_{\beta,\mu}(s,t) &= (2pi)^{-k}\exp\left(-\frac{\left\|s-\left(\beta-\beta_0\right)C_{\beta_0}\mu\right\|^2+\left\|t-D_\beta\mu\right\|^2}{2}\right) \\ &= (2pi)^{-k/2}\exp\left(-\frac{\left\|s-\left(\beta-\beta_0\right)C_{\beta_0}\mu\right\|^2}{2}\right)\times(2pi)^{-k/2}\exp\left(-\frac{\left\|t-D_\beta\mu\right\|^2}{2}\right) \\ &= f^S_{\beta,\mu}(s)\times f^T_{\beta,\mu}(t). \end{aligned}$$

Under the null hypothesis,

$$\begin{aligned} f_{\beta_0,\mu}(s,t) &= (2pi)^{-k/2}\exp\left(-\frac{\left\|s\right\|^2}{2}\right)\times(2pi)^{-k/2}\exp\left(-\frac{\left\|t-D_{\beta_0}\mu\right\|^2}{2}\right) \\ &= f^S_{\beta_0}(s)\times f^T_{\beta_0,\mu}(t), \end{aligned}$$

where the mean of T is given by

$$D_{\beta_0}\mu = \left[\left(a_0' \otimes I_k\right)\Sigma^{-1}\left(a_0 \otimes I_k\right)\right]^{1/2}\mu.$$

2.1 Weighted-Average Power (WAP)

The weighting function is chosen after approximating the covariance matrix Σ by the Kronecker product $\Omega \otimes \Phi$. Let $\|X\|_F = \left(tr\left(X'X\right)\right)^{1/2}$ denote the Frobenius norm of a matrix X. For a positive-definite covariance matrix Σ, Van Loan and Ptsianis (1993, p. 14) find symmetric and positive definite matrices Ω and Φ with dimension 2×2 and $k\times k$ which minimize $\|\Sigma-\Omega_0\otimes\Phi_0\|_F$.

We now integrate out the distribution given in (2.1) with respect to a prior for μ and β. For the prior $\mu \sim N\left(0,\sigma^2\Phi\right)$, the integrated likelihood is

$$(2pi)^{-k} \left|\Psi_\beta\right|^{-1/2} \exp\left(-\frac{(s',t')\,\Psi_\beta^{-1}\,(s',t')'}{2}\right)$$

where the $2k \times 2k$ covariance matrix is given by

$$\Psi_\beta = I_2 \otimes I_k + \sigma^2 \begin{bmatrix} (\beta-\beta_0)^2\, C_{\beta_0}\Phi C_{\beta_0} & (\beta-\beta_0)\, C_{\beta_0}\Phi D'_\beta \\ (\beta-\beta_0)\, D_\beta \Phi C_{\beta_0} & D_\beta \Phi D'_\beta \end{bmatrix}.$$

We now pick the prior $\beta \sim N(\beta_0, 1)$. The integrated likelihood for S and T using the prior $N(0, \sigma^2\Phi) \times N(\beta_0, 1)$ on μ and β yields

$$h_1(s,t) = (2pi)^{-k-1/2} \int \left|\Psi_\beta\right|^{-1/2} \exp\left(-\frac{(s',t')\,\Psi_\beta^{-1}\,(s',t')'}{2}\right) \exp\left(-\frac{(\beta-\beta_0)^2}{2}\right) d\beta.$$

We will set $\sigma^2 = 1$ for the simulations.

2.1.1 A Sign Invariant WAP Test

We can adjust the weights for β so that the WAP similar test is unbiased when $\Sigma = \Omega \otimes \Phi$.

We choose the (conditional on β) prior $\mu \sim N\left(0, \|l_\beta\|^{-2}\,\zeta \cdot \Phi\right)$ for a scalar ζ and the two-dimensional vector

$$l_\beta = \begin{bmatrix} (\beta-\beta_0)\cdot(b_0'\Omega b_0)^{-1/2} \\ a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \end{bmatrix} = \begin{bmatrix} (\beta-\beta_0)\cdot(b_0'\Omega b_0)^{-1/2} \\ b'\Omega b_0 \cdot (b_0'\Omega b_0)^{-1/2}\,|\Omega|^{-1/2} \end{bmatrix}.$$

The integrated density is

$$(2pi)^{-k} \left|\Psi_{\beta,\zeta}\right|^{-1/2} \exp\left(-\frac{(s',t')\,\Psi_{\beta,\zeta}^{-1}\,(s',t')'}{2}\right),$$

where the $2k \times 2k$ covariance matrix is given by

$$\Psi_{\beta,\zeta} = I_2 \otimes I_k + \frac{\zeta}{\|l_\beta\|^2} \begin{bmatrix} (\beta-\beta_0)^2\, C_{\beta_0}\Phi C_{\beta_0} & (\beta-\beta_0)\, C_{\beta_0}\Phi D'_\beta \\ (\beta-\beta_0)\, D_\beta \Phi C_{\beta_0} & D_\beta \Phi D'_\beta \end{bmatrix}.$$

It is now convenient to change variables:

$$(\cos(\theta), \sin(\theta))' = l_\beta / \|l_\beta\|.$$

The one-to-one mapping $\beta(\theta)$ is then

$$\beta = \beta_0 + \frac{b_0' \Omega b_0}{e_2' \Omega b_0 + \tan(\theta) \cdot |\Omega|^{1/2}}.$$

We choose the prior for θ to be uniform on $[-pi, pi]$. The integrated likelihood for S and T using the prior $N(0, \left\| l_{\beta(\theta)} \right\|^{-2} \zeta \cdot \Phi) \times Unif[-pi, pi]$ on μ and θ yields

$$h_2(s,t) = (2pi)^{-(k+1)} \int_{-pi}^{pi} \left| \Psi_{\beta(\theta),\zeta} \right|^{-1/2} \exp\left(-\frac{(s',t')\, \Psi_{\beta(\theta),\zeta}^{-1} (s',t')'}{2} \right) d\theta.$$

We will set $\zeta = 1$ for the simulations.

The following proposition shows that the WAP densities $h_1(s,t)$ and $h_2(s,t)$ enjoy invariance properties when the covariance matrix is a Kronecker product.

Proposition 1 *The following holds when $\Sigma = \Omega \otimes \Phi$:*
(i) The weighted average density $h_1(s,t)$ is invariant to orthogonal transformations. That is, it depends on the data only through

$$Q = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} = \begin{bmatrix} S'S & S'T \\ S'T & T'T \end{bmatrix}.$$

(ii) The weighted average density $h_2(s,t)$ is invariant to orthogonal sign transformations. That is, it depends on the data only through Q_S, $|Q_{ST}|$, and Q_T.

Tests which depend on the data only through Q_S, $|Q_{ST}|$, and Q_T are locally unbiased; see Corollary 1 of Andrews, Moreira, and Stock (2006). Hence, tests based on the WAP $h_2(s,t)$ are naturally two-sided tests for the null $H_0 : \beta = \beta_0$ against the alternative $H_1 : \beta \neq \beta_0$ when $\Sigma = \Omega \otimes \Phi$.

2.2 Two-Sided Boundary Conditions

Tests depending on the data only through Q_S, $|Q_{ST}|$, and Q_T are locally unbiased; see Corollary 1 of Andrews, Moreira, and Stock (2006). Hence,

a WAP similar test based on $h_2(s,t)$ is naturally a two-sided test for the null $H_0 : \beta = \beta_0$ against the alternative $H_1 : \beta \neq \beta_0$ when $\Sigma = \Omega \otimes \Phi$. When errors are autocorrelated and heteroskedastic, the covariance Σ will typically not have a Kronecker product structure. In this case, the WAP similar test based on $h_2(s,t)$ may not have good power. Indeed, this test is truly a two-sided test exactly because the sign-group of transformations preserves the two-sided testing problem when $\Sigma = \Omega \otimes \Phi$. When there is no Kronecker product structure, there is actually no sign invariance argument to accommodate two-sided testing.

Proposition 2 *Assume that we cannot write $\Sigma = \Omega \otimes \Phi$ for a 2×2 matrix Ω and a $k \times k$ matrix Φ, both symmetric and positive definite. Then for the data group of transformations $[S,T] \to [\pm S, T]$, there exists no group of transformations in the parameter space which preserves the testing problem.*

Proposition 2 asserts that we cannot simplify the two-sided hypothesis testing problem using sign invariance arguments. An unbiasedness condition instead adjusts the bias automatically (whether Σ has a Kronecker product or not). Hence, we seek approximately optimal unbiased tests.

2.2.1 Locally Unbiased (LU) condition

The next proposition provides necessary conditions for a test to be unbiased.

Proposition 3 *A test is said to be locally unbiased (LU) if*

$$E_{\beta_0,\mu}\phi(S,T)\, S' C_{\beta_0}\mu = 0, \forall \mu. \qquad \text{(LU condition)}$$

If a test is unbiased, then it is similar and locally unbiased.

Following Proposition 3, we would like to find WAP locally unbiased tests:

$$\max_{\phi \in \mathbb{K}} \int \phi h, \text{ where } \int \phi f_{\beta_0,\mu} = \alpha \text{ and } \int \phi s' C_{\beta_0} \mu f_{\beta_0,\mu} = 0, \forall \mu. \qquad (2.2)$$

The optimal tests based on $h_1(s,t)$ and $h_2(s,t)$ are denoted respectively MM1-LU and MM2-LU tests. Relaxing both constraints in (2.2) will assure

us the existence of multipliers. We solve the approximated maximization problem:

$$\begin{aligned} \max_{\phi\in\mathbb{K}} \quad & \int \quad \phi h, \text{ where } \int \phi f_{\beta_0,\mu} \in [\alpha - \ \epsilon, \alpha+\epsilon], \forall \mu \\ \text{and} \quad & \int \quad \phi s' C_{\beta_0} \mu_l f_{\beta_0,\mu_l} = 0, \text{ for } l = 1, ..., n, \end{aligned} \tag{2.3}$$

when ϵ is small and the number of discretizations n is large.

The optimal test rejects the null hypothesis when

$$h(s,t) - s' C_{\beta_0} \sum_{l=1}^{n} c_l^{\epsilon} \mu_l f_{\beta_0,\mu_l}(s,t) > \int f_{\beta_0,\mu}(s,t) \Lambda_{\epsilon}(d\mu),$$

where the multipliers c_l^{ϵ}, $l = 1, ..., n$, and Λ_{ϵ} satisfy the constraints in the maximization problem (2.3). We can write

$$\frac{h(s,t)}{f_{\beta_0}^S(s)} - s' C_{\beta_0} \sum_{l=1}^{n} c_l^{\epsilon} \mu_l f_{\beta_0,\mu_l}^T(t) > \int f_{\beta_0,\mu}^T(t) \Lambda_{\epsilon}(d\mu).$$

Letting $\epsilon \downarrow 0$, the optimal test rejects the null hypothesis when

$$\frac{h(s,t)}{f_{\beta_0}^S(s)} - s' C_{\beta_0} \sum_{l=1}^{n} c_l \mu_l f_{\beta_0,\mu_l}^T(t) > q(t),$$

where $q(t)$ is the conditional $1-\alpha$ quantile of

$$\frac{h(S,t)}{f_{\beta_0}^S(S)} - S' C_{\beta_0} \sum_{l=1}^{n} c_l \mu_l f_{\beta_0,\mu_l}^T(t).$$

This representation is very convenient as we can find

$$q(t) = \lim_{\epsilon\downarrow 0} \int f_{\beta_0,\mu}^T(t) \Lambda_{\epsilon}(d\mu)$$

by numerical approximations of the conditional distribution instead of searching for an infinite-dimensional multiplier Λ_{ϵ}. In the second step, we search for the values c_l so that

$$E_{\beta_0,\mu_l} \phi(S,T) S' C_{\beta_0} \mu_l = \int \phi(s,t) s' C_{\beta_0} \mu_l f_{\beta_0}^S(s) f_{\beta_0,\mu_l}^T(t) = 0,$$

by taking into consideration that $q(t)$ depends on c_l, $l = 1, ..., n$. We use a nonlinear numerical algorithm to find c_l, $l = 1, ..., n$.

As an alternative procedure, we consider a condition stronger than the LU condition which is simpler to implement numerically. This strategy turns out to be useful because it gives a simple way to implement tests with overall good power. We provide details for this alternate condition next.

2.2.2 Strongly Unbiased (SU) condition

The LU condition asserts that the test ϕ is uncorrelated with a linear combination indexed by the instruments' coefficients μ and the pivotal statistic S. We note that the LU condition trivially holds if

$$E_{\beta_0,\mu}\phi(S,T)\,S = 0, \forall\mu. \qquad \text{(SU condition)}$$

That is, the test ϕ is uncorrelated with the k-dimensional statistic S itself under the null. The strongly unbiased (SU) condition above states that the test $\phi(S,T)$ is uncorrelated with S for all instruments' coefficients μ. The following lemma shows that there are tests which satisfy the LU condition, but not the SU condition. Hence, finding WAP similar tests that satisfy the SU instead of the LU condition in theory may entail unnecessary power losses (in Appendix A, we show that those power losses in practice are numerically quite small).

Lemma 1 *Define the mapping*

$$G_\phi(s,t,z_1,z_2) = \phi(s,t)s'C_{\beta_0}z_1 \cdot exp\left(-s's/2\right) \cdot exp\left(-(t-z_2)'(t-z_2)/2\right)$$

and the integral

$$F_\phi(z_1,z_2) = \int G_\phi(s,t,z_1,z_2)d(s,t).$$

Then there exists $\phi \in \mathbb{K} \subset \mathcal{L}_\infty(\mathbb{R}^{2k})$ such that $F_\phi(z_1,z_1) = 0$, for all z_1, and $F_\phi(z_1,z_2) \neq 0$, for some z_1 and z_2.

The WAP strongly unbiased (SU) test solves

$$\max_{\phi\in\mathbb{K}} \int \phi h, \text{ where } \int \phi f_{\beta_0,\mu} = \alpha \text{ and } \int \phi s f_{\beta_0,\mu} = 0, \forall\mu.$$

Because the statistic T is complete, we can carry on power maximization for each level of $T = t$:

$$\max_{\phi \in \mathbb{K}} \int \phi h, \text{ where } \int \phi f_{\beta_0}^S = \alpha \text{ and } \int \phi s f_{\beta_0}^S = 0, \forall t, \tag{2.4}$$

where the integrals are taken with respect to s only. The optimal test rejects the null when

$$\frac{h(s,t)}{f_{\beta_0}^S(s)} > c(s,t),$$

where the function $c(s,t) = c_0(t) + s'c_1(t)$ satisfies the boundary conditions in (2.4).

In practice, we can find $c_0(t)$ and $c_1(t)$ using linear programming based on simulations for the statistic S. Consider the approximated problem

$$\begin{aligned} \max_{0 \le x^{(j)} \le 1} \quad & J^{-1} \sum_{j=1}^{J} x^{(j)} h\left(s^{(j)}, t\right) \exp\left(s^{(j)\prime} s^{(j)}/2\right) (2pi)^{k/2} \\ \text{s.t.} \quad & J^{-1} \sum_{j=1}^{J} x^{(j)} = \alpha \text{ and} \\ & J^{-1} \sum_{j=1}^{J} x^{(j)} s_l^{(j)} = 0, \text{ for } l = 1, ..., n. \end{aligned}$$

Each j-th draw of S is iid standard-normal:

$$S^{(j)} = \begin{bmatrix} S_1^{(j)} \\ \vdots \\ S_k^{(j)} \end{bmatrix} \sim N(0, I_k).$$

We note that for the linear programming, the only term which depends on $T = t$ is $h\left(s^{(j)}, t\right)$. The multipliers for this linear programming problem are the critical value functions $c_0(t)$ and $c(t)$. To speed up the numerical algorithm, we use the same sample $S^{(j)}$, $j = 1, ..., J$, for every level $T = t$.

Finally, we use the WAP test found in (2.4) to find a useful *power envelope*. The next proposition finds the optimal test for any given alternative which satisfies the SU condition.

Proposition 4 *The test which maximizes power for a given alternative* (β, μ) *given the constraints in (2.4) is*

$$\frac{\left(s' C_{\beta_0} \mu\right)^2}{\mu C_{\beta_0}^2 \mu} > q(1). \tag{2.5}$$

This test is denoted the Point Optimal Strongly Unbiased (POSU) test.

Comment: The POSU test does not depend on β but it does depend on the direction of the vector $C_{\beta_0}\mu$.

The power plot of the POSU test as β and μ change yields the power envelope. This proposition is analogous to Theorem 2-(c) of Moreira (2009) for the homoskedastic case within the class of SU tests.

3 Nearly Integrated Regressor

We want to test the null hypothesis $H_0 : \beta = \beta_0$. Consider the group of translation transformations on the data

$$\kappa \circ (y_{1,i}, y_{2,i}) = (y_{1,i} + \kappa, y_{2,i}),$$

where $\kappa \in \mathbb{R}$. The corresponding transformation on the parameter space is

$$\kappa \circ (\beta, \pi, \varphi) = (\beta, \pi, \varphi + \kappa).$$

This group action preserves the parameter of interest β. Because the group translation preserves the hypothesis testing problem, it is reasonable to focus on tests which are invariant to translation transformations on y_1.

Any invariant test can be written as a function of the maximal invariant statistic. Let $P = (P_1, P_2)$ be an orthogonal $N \times N$ matrix where the first column is given by $P_1 = 1_N/\sqrt{N}$. Algebraic manipulations show that $P_2 P_2' = M_{1_N}$, where $M_{1_N} = I_N - 1_N \left(1_N' 1_N\right)^{-1} 1_N'$ is the projection matrix to the space orthogonal to 1_N. Let $y_{2,-1}$ be the N-dimensional vector whose i-th entry is $y_{2,i-1}$, and define the $N-1$-dimensional vectors $\widetilde{y}_j = P_2' y_j$ for $j = 1, 2$. The maximal invariant statistic is given by $\widetilde{y}_1$ and y_2. Its density is given by

$$f_{\beta,\pi}\left(\widetilde{y}_1, y_2\right) = \left(2\pi\omega_{22}\right)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\omega_{22}} \sum_{i=1}^{N} \left(y_{2,i} - y_{2,i-1}\pi\right)^2\right\} \tag{3.6}$$

$$\times \left(2\pi\omega_{11.2}\right)^{-\frac{N-1}{2}} \exp\left\{-\frac{1}{2\omega_{11.2}} \sum_{i=1}^{N} \left(\widetilde{y}_{1,i} - \widetilde{y}_{2,i}\frac{\omega_{12}}{\omega_{22}} - \widetilde{y}_{2,i-1}\left[\beta - \pi\frac{\omega_{12}}{\omega_{22}}\right]\right)^2\right\},$$

where $\omega_{11.2} = \omega_{11} - \omega_{12}^2/\omega_{22}$ is the variance of $\epsilon_{1,i}$ not explained by $\epsilon_{2,i}$.

For testing $H_1 : \beta > \beta_0$, we find the WAP test which is similar at $\beta = \beta_0$ and has correct size:

$$\max_{\phi \in \mathbb{K}} \int \phi h, \text{ where } \int \phi f_{\beta_0,\pi} = \alpha \text{ and } \int \phi f_{\beta,\pi} \leq \alpha, \forall \beta \leq \beta_0, \pi. \tag{3.7}$$

We choose the prior $\Lambda_1(\beta, \mu)$ to be the product of $N(\beta_0, 1)$ conditional on $[\beta_0, \infty)$ and $Unif[\underline{\pi}, \overline{\pi}]$. The MM-1S statistic is the weighted average density

$$\int_{\underline{\pi}}^{\overline{\pi}} \int_{\beta_0}^{\infty} f_{\beta,\pi}(\widetilde{y}_1, y_2)(2pi\sigma^2)^{-1/2} \frac{2}{\overline{\pi} - \underline{\pi}} \exp\left[-\frac{(\beta - \beta_0)^2}{2\sigma^2}\right] \, d\beta \, d\pi.$$

As for the constraints in the maximization problem, there are two boundary conditions. The first one states that the test is similar. The second one asserts the test has correct size.

For testing $H_1 : \beta \neq \beta_0$, we seek the WAP-LU (locally unbiased) test:

$$\max_{\phi \in \mathbb{K}} \int \phi h, \text{ where } \int \phi f_{\beta_0,\pi} = \alpha \text{ and } \int \phi \left. \frac{\partial \ln f_{\beta,\pi}}{\partial \beta} \right|_{\beta=\beta_0} f_{\beta_0,\pi} = 0, \forall \pi. \tag{3.8}$$

We choose the prior $\Lambda_1(\beta, \mu)$ to be the product of $N(\beta_0, 1)$ and $Unif[\underline{\pi}, \overline{\pi}]$. The MM-2S statistic is the weighted average density becomes

$$h(\widetilde{y}_1, y_2) = \int_{\underline{\pi}}^{\overline{\pi}} \int_{-\infty}^{\infty} f_{\beta,\pi}(\widetilde{y}_1, y_2)(2pi\sigma^2)^{-1/2} \frac{1}{\overline{\pi} - \underline{\pi}} \exp\left[-\frac{(\beta - \beta_0)^2}{2\sigma^2}\right] \, d\beta \, d\pi.$$

There are two boundary conditions. The first one again states that the test is similar. The second constraint arises because the derivative of the power function of locally unbiased tests is zero at $\beta = \beta_0$:

$$\begin{aligned}
\left. \frac{\partial E_{\beta,\pi} \phi(\widetilde{y}_1, y_2)}{\partial \beta} \right|_{\beta=\beta_0} &= \int \phi(\widetilde{y}_1, y_2) \left. \frac{\partial f_{\beta,\pi}(\widetilde{y}_1, y_2)}{\partial \beta} \right|_{\beta=\beta_0} \\
&= \int \phi(\widetilde{y}_1, y_2) \left. \frac{\partial \ln f_{\beta,\pi}(\widetilde{y}_1, y_2)}{\partial \beta} \right|_{\beta=\beta_0} f_{\beta,\pi}(\widetilde{y}_1, y_2) \\
&= \int \phi(\widetilde{y}_1, y_2) \psi(\widetilde{y}_1, y_2) f_{\beta,\pi}(\widetilde{y}_1, y_2),
\end{aligned}$$

where the statistic $\psi(\widetilde{y}_1, y_2)$ is given by

$$\psi(\widetilde{y}_1, y_2) = \sum_{i=1}^{N} y_{2,i-1} \left(\widetilde{y}_{1,i} - \widetilde{y}_{2,i-1}\beta_0 - [\widetilde{y}_{2,i} - \widetilde{y}_{2,i-1}\pi] \frac{\omega_{12}}{\omega_{22}} \right).$$

We implement the one-sided WAP similar and the two-sided WAP locally unbiased tests by discretizing the number of boundary conditions in (3.7) and (3.8). To save space, we discuss only the implementation of the WAP locally unbiased test based on the MM-2S statistic. We then solve

$$\max_{\phi\in\mathbb{K}} \int \phi h, \text{ where } \int \phi f_{\beta_0,\pi_l} = \alpha \text{ and } \int \phi.\psi f_{\beta_0,\pi_l} = 0,$$

where $\underline{\pi} = \pi_1 < \pi_2 < ... < \pi_n = \overline{\pi}$.

To avoid numerical issues, we use the density[1]

$$\overline{f}_{\beta_0}(\widetilde{y}_1, y_2) = \frac{1}{n}\sum_{l=1}^{n} f_{\beta_0,\pi_l}(\widetilde{y}_1, y_2).$$

This density arises from generating the data

$$\begin{aligned} y_{1,i} &= y_{2,i-1}\beta + \epsilon_{1,i} \\ y_{2,i} &= y_{2,i-1}\pi_l + \epsilon_{2,i}, \end{aligned}$$

where we select π_l randomly among $\pi_1, \pi_2, ..., \pi_n$.

The two-sided maximization problem simplifies to

$$\max_{\phi\in\mathbb{K}} \int \phi \frac{h}{\overline{f}_{\beta_0}} \overline{f}_{\beta_0}, \text{ where } \int \phi \frac{f_{\beta_0,\pi_l}}{\overline{f}_{\beta_0}} \overline{f}_{\beta_0} = \alpha \text{ and } \int \phi.\psi \frac{f_{\beta_0,\pi_l}}{\overline{f}_{\beta_0}} \overline{f}_{\beta_0} = 0,$$

for $l = 1, ..., n$. Using SLLN, we solve the approximated two-sided testing problem

$$\max_{x^{(j)}\in\{0,1\}} \frac{1}{J}\sum_{j=1}^{J} x^{(j)} \frac{h\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}{\overline{f}_{\beta_0}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}, \tag{3.9}$$

$$\text{where } \frac{1}{J}\sum_{j=1}^{J} x^{(j)} \frac{f_{\beta_0,\pi_l}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}{\overline{f}_{\beta_0}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)} = \alpha$$

$$\text{and } \frac{1}{J}\sum_{j=1}^{J} x^{(j)} \psi\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right) \frac{f_{\beta_0,\pi_l}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}{\overline{f}_{\beta_0}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)} = 0,$$

[1]For the numerical results in this paper, we use the same densities in the boundary conditions for importance sampling (although there is no need to).

for $l = 1, ..., n$.

We can write (3.9) as

$$\max_{0 \leq x_j \leq 1} r'x$$
$$\text{s.t. } Ax \leq p,$$

for appropriate matrices A and vectors p and r. We then use standard linear programming algorithms to find x and $c = (c_1, ..., c_{2n})$ to the dual problem

$$\min_{c \in \mathbb{R}_+^{2n}} p'c$$
$$\text{s.t. } A'c \geq r.$$

The two-sided test rejects the null when

$$h(\widetilde{y}_1, y_2) > \sum_{l=1}^{n} \left[c_l + c_{n+l} \cdot \psi(\widetilde{y}_1, y_2)\right] f_{\beta_0, \pi_l}(\widetilde{y}_1, y_2).$$

4 Approximation in Hilbert Space

In this section, we show that power maximization is equivalent to norm minimization in Banach spaces. By modifying the original maximization problem, we characterize optimal tests in Hilbert spaces.

We would like to transform the maximization problem into a minimum norm problem. Let $\mathcal{L}_p(Y, h)$ be the Banach space of measurable functions ϕ such that $\int |\phi|^p h < \infty$ with norm $\|\phi\|_p^h = \left(\int |\phi|^p h\right)^{1/p}$. We then have

$$\sup_{\phi \in \mathcal{L}_1(Y,h)} \int \phi h \text{ where } 0 \leq \phi \leq 1 \text{ and } \int \phi g_v \in [\gamma_v^1, \gamma_v^2], \forall v \in \mathbf{V}. \tag{4.10}$$

Remark 1 *Consider the Banach space $\mathcal{L}_1(Y, h)$, where h is a density, and let $\Gamma_1(g, \gamma) = \{\phi \in \mathcal{L}_1(Y, h); 0 \leq \phi \leq 1$ and $\int \phi g_v \in [\gamma_v^1, \gamma_v^2], \forall v \in \mathbf{V}\}$. Then the maximization problem given by (4.10) is equivalent to the minimum norm problem*

$$1 - \inf_{\phi \in \Gamma_1(g,\gamma)} \|\phi - 1\|_1^h. \tag{4.11}$$

Norm minimization for general Banach spaces lacks geometric interpretation. Consider instead norm minimization in $\mathcal{L}_2(Y,h)$:

$$1 - \inf_{\phi \in \Gamma_2(g,\gamma)} \left(\|\phi - 1\|_2^h \right)^2, \tag{4.12}$$

where $\Gamma_2(g,\gamma) = \{\phi \in \mathcal{L}_2(Y,h); 0 \leq \phi \leq 1 \text{ and } \int \phi g_v \in [\gamma_v^1, \gamma_v^2], \forall v \in \mathbf{V}\}$. The following proposition provides a necessary and sufficient condition for maximization of (4.12).

Proposition 5 *If $g_v/h \in \mathcal{L}_2(Y,h)$ for $v \in \mathbf{V}$, then:*
(a) *There exists a unique $\overline{\phi}_2 \in \Gamma_2(g,\gamma)$ such that*

$$1 - \left\|\overline{\phi}_2 - 1\right\|_2^h \geq 1 - \|\phi_2 - 1\|_2^h$$

for all $\phi_2 \in \Gamma_2(g,\gamma)$.
(b) *A necessary and sufficient condition for $\overline{\phi}_2$ to solve (4.12) is that*

$$\int \left(1 - \overline{\phi}_2\right) \left(\phi_2 - \overline{\phi}_2\right) h \leq 0$$

for all $\phi_2 \in \Gamma_2(g,\gamma)$.

If the test ϕ is nonrandomized, the objective function equals the power function

$$1 - \left(\|\phi - 1\|_2^h \right)^2 = \int \phi h.$$

If the test ϕ is randomized, it distorts the objective function. Hence, the optimal test $\overline{\phi}$ for (4.11) can be different from $\overline{\phi}_2$ for (4.12). In the case of a similar test, it may even be possible that $\overline{\phi}$ is nonrandomized whereas $\overline{\phi}_2$ is randomized. When g_v is the density f_v and $\gamma_v^1 = \gamma_v^2 = \alpha \in (0,1)$, Proposition 5(b) guarantees that $\overline{\phi}$ is also optimal for problem (4.12) if and only if

$$\int \left(1 - \overline{\phi}\right) \phi_2 h \leq \int \left(1 - \overline{\phi}\right) \overline{\phi} h$$

for every $\phi_2 \in \Gamma_2(g,\gamma)$. If the optimal test $\overline{\phi}$ is nonrandomized, then

$$\int \left(1 - \overline{\phi}\right) \phi_2 h \leq 0$$

for all $\phi_2 \in \Gamma_2(g,\gamma)$. This clearly cannot be true for $\phi_2 = \alpha$ unless $\int \overline{\phi} h = 1$. This issue happens even for exponential family models, where an optimal nonrandomized test exists. Hence, minimizing (4.12) instead of minimizing (4.11) has unappealing consequences. An alternative is to consider a sequence of problems in Hilbert space whose objective function approaches $\int \phi h$.

The next lemma provides an approximation of $\overline{\phi}$ for the strong topology in $\mathcal{L}_2(Y,h)$.

Proposition 6 *Suppose that $g_v/h \in \mathcal{L}_2(Y,h)$ for $v \in \mathbf{V}$ and for any $\epsilon > 0$, define*

$$\sup_{\phi \in \Gamma_2(g,\gamma)} \int \phi h - \epsilon \int \phi^2 h. \tag{4.13}$$

(a) *There exists a unique $\overline{\phi}_\epsilon$ that solves (4.13). A necessary and sufficient condition for $\overline{\phi}_\epsilon$ is given by*

$$\int \left(\frac{1}{2\epsilon} - \overline{\phi}_\epsilon \right) \left(\phi - \overline{\phi}_\epsilon \right) h \leq 0,$$

for all $\phi \in \Gamma_2(g,\gamma)$.
(b) *As $\epsilon \downarrow 0$, the value function in (4.13) converges to the value function in*

$$\sup_{\phi \in \Gamma_2(g,\gamma)} \int \phi h.$$

(c) *The test $\overline{\phi}_\epsilon$ is continuous in ϵ and the power function $\int \overline{\phi}_\epsilon g_v \to \int \overline{\phi} g_v$ uniformly for every $v \in \mathbb{V}$.*

Comment: The optimal $\overline{\phi}$ is the limit of the net $(\overline{\phi}_\epsilon)$ in $\mathcal{L}_2(Y,h)$, hence is unique.

Suppose that g_v is the density f_v and $\gamma_v^1 = \gamma_v^2 = \alpha \in (0,1)$. Proposition 6 shows how to find a similar test $\overline{\phi}_\epsilon$ such that

$$\int \overline{\phi}_\epsilon h \geq \sup \int \phi h - \epsilon$$

among all α-similar tests on $\mathbf{V}$. Hence, $\overline{\phi}_\epsilon$ is an ϵ-optimal test in the sense of Linnik (2000). Using the necessary and sufficient condition from part (a)

to implement $\overline{\phi}_\epsilon$ is not straightforward. A simpler approach is to start with a collection of a finite number of similar tests $\phi^1, ..., \phi^n$, where $\phi^1 = \alpha$. For curved exponential family models, we can find these tests using the D-method of Wijsman (1958). Define the closed convex set

$$\widetilde{\Gamma}_2 = \left\{ \phi \in \mathcal{L}_2(Y, h); \phi = \sum_{l=1}^{n} \kappa_l \cdot \phi^l \text{ where } \sum_{l=1}^{n} \kappa_l = 1 \text{ and } \kappa_l \geq 0 \right\}.$$

Analogous to Proposition 6, there exists a unique test $\widetilde{\phi}_\epsilon$ that maximizes power in $\widetilde{\Gamma}_2$. The necessary and sufficient condition for $\widetilde{\phi}_\epsilon$ is given by

$$\int \left(\frac{1}{2\epsilon} - \widetilde{\phi}_\epsilon \right) \left(\widetilde{\phi} - \widetilde{\phi}_\epsilon \right) h \leq 0,$$

for all $\widetilde{\phi} \in \widetilde{\Gamma}_2$. The implementation of $\widetilde{\phi}_\epsilon$ can be done in the manner as Example 1 of Luenberger (1969, p. 71).

5 Proofs

Derivation of the Score Test. For the statistic $R = vec\,(Z'Z)^{-1/2} Z'Y$, the log-likelihood is proportional to

$$L(\beta, \mu) = -\frac{1}{2} (r - (a \otimes I_k)\mu)' \Sigma^{-1} (r - (a \otimes I_k)\mu).$$

Taking derivative with respect to μ:

$$\frac{\partial L(\beta, \mu)}{\partial \mu} = (a' \otimes I_k)\Sigma^{-1} (r - (a \otimes I_k)\mu) = 0$$

which implies that

$$\hat{\mu} = \left[(a' \otimes I_k)\Sigma^{-1}(a \otimes I_k) \right]^{-1} (a' \otimes I_k)\Sigma^{-1} r.$$

The concentrated log-likelihood function is

$$\begin{aligned} L_c(\beta) &= L(\beta, \hat{\mu}) \\ &= -\frac{1}{2} r' \Sigma^{-1/2} M_{\Sigma^{-1/2}(a \otimes I_k)} \Sigma^{-1/2} r \\ &= -\frac{1}{2} r' \Sigma^{-1} r + \frac{1}{2} r' \Sigma^{-1} (a \otimes I_k) \left[(a' \otimes I_k)\Sigma^{-1}(a \otimes I_k) \right]^{-1} (a' \otimes I_k)\Sigma^{-1} r. \end{aligned}$$

The score is given by

$$\begin{aligned}\frac{\partial L_c(\beta)}{\partial \beta} &= r'\Sigma^{-1}(a\otimes I_k)\left[(a'\otimes I_k)\Sigma^{-1}(a\otimes I_k)\right]^{-1}(e_1'\otimes I_k)\Sigma^{-1}r \\ &\quad -\frac{1}{2}r'\Sigma^{-1}(a\otimes I_k)\left[(a'\otimes I_k)\Sigma^{-1}(a\otimes I_k)\right]^{-1} \\ &\quad \times\left\{(e_1'\otimes I_k)\Sigma^{-1}(a\otimes I_k)+(a'\otimes I_k)\Sigma^{-1}(e_1\otimes I_k)\right\} \\ &\quad \times\left[(a'\otimes I_k)\Sigma^{-1}(a\otimes I_k)\right]^{-1}(a'\otimes I_k)\Sigma^{-1}r.\end{aligned}$$

At $\beta=\beta_0$:

$$\begin{aligned}\frac{\partial L_c(\beta_0)}{\partial \beta} &= r'\Sigma^{-1}(a_0\otimes I_k)\left[(a_0'\otimes I_k)\Sigma^{-1}(a_0\otimes I_k)\right]^{-1} \\ &\quad \times(e_1'\otimes I_k)\Sigma^{-1/2}M_{\Sigma^{-1/2}(a_0\otimes I_k)}\Sigma^{-1/2}\left[(e_1,e_2)\otimes I_k\right]r \\ &= r'\Sigma^{-1}(a_0\otimes I_k)\left[(a_0'\otimes I_k)\Sigma^{-1}(a_0\otimes I_k)\right]^{-1} \\ &\quad \times(e_1'\otimes I_k)\Sigma^{-1/2}M_{\Sigma^{-1/2}(a_0\otimes I_k)}\Sigma^{-1/2}\left((e_1,a_0-\beta_0e_1)\otimes I_k\right)r.\end{aligned}$$

Note that

$$\begin{aligned}C_{\beta_0}^{-1} &= (e_1'\otimes I_k)\Sigma^{-1}(e_1\otimes I_k)-(e_1'\otimes I_k)\Sigma^{-1}(a_0\otimes I_k) \\ &\quad \times\left[(a_0'\otimes I_k)\Sigma^{-1}(a_0\otimes I_k)\right]^{-1}(a_0'\otimes I_k)\Sigma^{-1}(e_1\otimes I_k).\end{aligned}$$

Indeed,

$$[X_1,X_2]=\left[\Sigma^{-1/2}(e_1\otimes I_k),\Sigma^{-1/2}(a_0\otimes I_k)\right]=\Sigma^{-1/2}\left[e_1,a_0\right]\otimes I_k.$$

$$\left([X_1,X_2]'[X_1,X_2]\right)^{-1}=\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2\end{pmatrix}^{-1}=\begin{pmatrix} X^{11} & X^{12} \\ X^{21} & X^{22}\end{pmatrix}.$$

Hence,

$$\begin{aligned}\left([X_1,X_2]'[X_1,X_2]\right)^{-1} &= \left[\left([e_1,a_0]'\otimes I_k\right)\Sigma^{-1}\left([e_1,a_0]\otimes I_k\right)\right]^{-1} \\ &= \left([e_1,a_0]^{-1}\otimes I_k\right)\Sigma\left([e_1,a_0]^{-1\prime}\otimes I_k\right) \\ &= \left(\begin{pmatrix} 1 & -\beta_0 \\ 0 & 1\end{pmatrix}\otimes I_k\right)\Sigma\left(\begin{pmatrix} 1 & 0 \\ -\beta_0 & 1\end{pmatrix}\otimes I_k\right).\end{aligned}$$

Therefore, the top-left submatrix X^{11} of the matrix $\left([X_1, X_2]' [X_1, X_2]\right)^{-1}$ equals C_{β_0}:

$$(e_1' \otimes I_k)\Sigma^{-1/2} M_{\Sigma^{-1/2}(a \otimes I_k)} \Sigma^{-1/2} e_1 = \left[(b_0' \otimes I_k)\Sigma(b_0 \otimes I_k)\right]^{-1}.$$

We obtain

$$\begin{aligned} \frac{\partial \mathcal{L}_c}{\partial \beta} &= r'\Sigma^{-1}(a_0 \otimes I_k)\left[(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)\right]^{-1} \\ &\quad \times \left[(b_0' \otimes I_k)\Sigma(b_0 \otimes I_k)\right]^{-1}(b_0' \otimes I_k) r \\ &= s' C_{\beta_0}^{-1/2} D_{\beta_0}^{-1/2} t. \end{aligned}$$

We can standardize it by a consistent estimator of the asymptotic variance. In particular, we can choose

$$LM = \frac{S' C_{\beta_0}^{-1/2} D_{\beta_0}^{-1/2} T}{\sqrt{T' D_{\beta_0}^{-1/2} C_{\beta_0}^{-1} D_{\beta_0}^{-1/2} T}},$$

as we wanted to prove. □

Proof of Proposition 1. For part (a),

$$\begin{aligned} (s', t')\, \Psi_\beta^{-1}\, (s', t')' &= tr\left([S, T]\left(I_2 + \sigma^2 l_\beta l_\beta{}'\right)^{-1}[S, T]'\right) \\ &= tr\left(\left(I_2 + \sigma^2 l_\beta l_\beta{}'\right)^{-1} Q\right). \end{aligned}$$

For part (b),

$$\begin{aligned} (s', t')\, \Psi_{\beta,\zeta}^{-1}\, (s', t')' &= tr\left(\left(I_2 + \zeta \frac{l_\beta l_\beta{}'}{\|l_\beta\|^2}\right)^{-1} Q\right) \\ &= tr\left(\left(I_2 - \frac{\zeta}{1+\zeta}\frac{l_\beta l_\beta{}'}{\|l_\beta\|^2}\right) Q\right) \\ &= Q_S + Q_T - \frac{\zeta}{1+\zeta}\frac{l_\beta Q l_\beta}{\|l_\beta\|^2}. \end{aligned} \tag{5.14}$$

Using the change of variables,

$$\begin{aligned} (s', t')\, \Psi_{\beta,\zeta}^{-1}\, (s', t')' &= Q_S + Q_T - \frac{\zeta}{1+\zeta}\left[\cos(\theta), \sin(\theta)\right] Q \left[\cos(\theta), \sin(\theta)\right]' \\ = \left(1 - \frac{\zeta}{1+\zeta}\cos^2(\theta)\right) Q_S &+ \left(1 - \frac{\zeta}{1+\zeta}\sin^2(\theta)\right) Q_T - \frac{\zeta}{1+\zeta}\sin(2\theta)\, Q_{ST}. \end{aligned}$$

The determinant of the matrix $\Psi_{\beta,\zeta}$ simplifies to

$$|\Psi_{\beta,\zeta}| = \left| I_2 + \zeta \frac{l_\beta l_\beta{}'}{\|l_\beta\|^2} \right|^{k/2} = (1+\zeta)^{k/2}.$$

Hence integrating out θ with respect to a uniform distribution on $[-pi, pi]$:

$$\begin{aligned} h_2(s,t) &= (2pi)^{-(k+1)} \int_{-pi}^{pi} (1+\zeta)^{-k/2} \exp\left(\frac{\zeta}{2(1+\zeta)} \sin(2\theta) Q_{ST} \right) \\ &\times \exp\left(-\frac{1}{2}\left(1 - \frac{\zeta}{1+\zeta}\cos^2(\theta)\right) Q_S - \frac{1}{2}\left(1 - \frac{\zeta}{1+\zeta}\sin^2(\theta)\right) Q_T \right) d\theta. \end{aligned}$$

The function $h_2(s,t)$ depends on the data only through Q_S, $|Q_{ST}|$, and Q_T, because

$$\cosh(\kappa) = \frac{\exp(\kappa) + \exp(-\kappa)}{2}$$

depends only on $|\kappa|$. $\square$

Proof of Proposition 2. In order to preserve the model (the null and the alternative hypothesis), it is necessary and sufficient to find $a_2 = (\beta_2, 1)'$ and μ such that

$$(a_0' \otimes I_k)\Sigma^{-1}(a \otimes I_k)\mu = (a_0' \otimes I_k)\Sigma^{-1}(a_2 \otimes I_k)\mu, \tag{5.15}$$

where $\mu(\beta - \beta_0) = -\mu(\beta_2 - \beta_0)$. Condition (5.15) is equivalent to

$$(a_0' \otimes I_k)\Sigma^{-1}\left[(a_2 \otimes I_k)(\beta - \beta_0) + (a \otimes I_k)(\beta_2 - \beta_0)\right]\mu = 0$$

or, alternatively, to

$$(a_0' \otimes I_k)\Sigma^{-1} \begin{pmatrix} \beta_2(\beta - \beta_0) + \beta(\beta_2 - \beta_0) \\ \beta - \beta_0 + \beta_2 - \beta_0 \end{pmatrix} \otimes \mu = 0. \tag{5.16}$$

We use the identity

$$\begin{aligned} \beta_2(\beta - \beta_0) + \beta(\beta_2 - \beta_0) &= (\beta_2 - \beta_0)(\beta - \beta_0) + \beta_0(\beta - \beta_0) \\ &+ (\beta - \beta_0)(\beta_2 - \beta_0) + \beta_0(\beta_2 - \beta_0) \end{aligned}$$

to write expression (5.16) as

$$\begin{aligned} -(\beta - \beta_0)(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)\mu &= (\beta_2 - \beta_0)(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)\mu \\ &+ 2(\beta_2 - \beta_0)(\beta - \beta_0)(a_0' \otimes I_k)\Sigma^{-1}(e_1 \otimes I_k)\mu. \end{aligned}$$

Therefore,

$$\begin{aligned}(\beta_2-\beta_0)\mu &= -(\beta-\beta_0)F_\beta^{-1}D_{\beta_0}^2\mu, \text{ where}\\ F_\beta &= D_{\beta_0}^2+2(\beta-\beta_0)(a_0'\otimes I_k)\Sigma^{-1}(e_1\otimes I_k).\end{aligned}$$

Because $\beta \neq \beta_0$ and μ is generic, we must have $F_\beta^{-1}D_{\beta_0}^2 = I_k$, which is impossible. □

Proof of Proposition 3. If a test is unbiased, then

$$E_{\beta_0,\mu}\phi\left(S,T\right) \leq \alpha \leq E_{\beta,\mu}\phi\left(S,T\right).$$

By taking sequences β_N approaching β_0, we show that

$$E_{\beta_0,\mu}\phi\left(S,T\right) = \alpha, \forall\mu.$$

It also must be the case that

$$\frac{\partial E_{\beta_0,\mu}\phi\left(S,T\right)}{\partial\beta} = 0, \forall\mu, \tag{5.17}$$

otherwise the power is smaller than zero for some value β close enough to β_0. The derivative of the power function is

$$\frac{\partial E_{\beta,\mu}\phi\left(S,T\right)}{\partial\beta} = \int\phi\left(s,t\right)\frac{\partial\ln f_{\beta,\mu}\left(s,t\right)}{\partial\beta}f_{\beta,\mu}\left(s,t\right).$$

Algebraic manipulations show that (5.17) simplifies to

$$E_{\beta_0,\mu}\phi\left(S,T\right)\cdot\left(S'C_{\beta_0}\mu+\left(T-D_{\beta_0}\mu\right)D_{\beta_0}^{1/2}\left(a_0'\otimes I_k\right)\Sigma^{-1}\left(e_1\otimes I_k\right)\mu\right)=0.$$

The test ϕ must be uncorrelated with the statistic T:

$$\begin{aligned}E_{\beta_0,\mu}\phi\left(S,T\right)\cdot\left(T-D_{\beta_0}\mu\right) &= E_{\beta_0,\mu}^T\left[E_{\beta_0}^S\phi\left(S,T\right)\right]\cdot\left(T-D_{\beta_0}\mu\right)\\ &= E_{\beta_0,\mu}^T\alpha\cdot\left(T-D_{\beta_0}\mu\right)\\ &= 0,\end{aligned}$$

where the second equality uses the fact that the test is similar and T is sufficient and complete under the null. Consequently, expression (5.17) holds if and only if

$$E_{\beta_0,\mu}\phi\left(S,T\right)S'C_{\beta_0}\mu = 0, \forall\mu,$$

as we wanted to prove. $\square$

Proof of Lemma 1. To simplify the notation, we will omit the explicit dependence of F_ϕ and G_ϕ on the test ϕ in the notation. Without loss of generality we can assume that $G_\phi(s,t,z_1,z_2) = \chi(s,t)'z_1 \cdot exp\,(t'z_2)$, where $\chi(s,t) = \phi(s,t) \cdot exp\,(-\,(s's+t't)\,/2)\, s'C_{\beta_0}$ (notice that we have eliminated the term $exp\,(-z_2'z_2/2)$).

The first differential of F_ϕ at (z_1,z_2) evaluated at the vector (u_1,u_2) is

$$DF_\phi(z_1,z_2)(u_1,u_2) = \int (\chi'u_1 + \chi'z_1 \cdot t'u_2)\, exp\,(t'z_2)\, d(s,t).$$

The second differential of F_ϕ at (z_1,z_2) evaluated at the vector (u_1,u_2) is

$$D^2F_\phi(z_1,z_2)(u_1,u_2) = \int (2\chi'u_1 + \chi'z_1 \cdot t'u_2)\, t'u_2 \cdot exp\,(t'z_2)\, d(s,t).$$

By finite induction, the n-th differential of F_ϕ at (z_1,z_2) evaluated at the vector (u_1,u_2) is

$$D^nF_\phi(z_1,z_2)(u_1,u_2) = \int (n\chi'u_1 + \chi'z_1 \cdot t'u_2)\,(t'u_2)^{n-1} \cdot exp\,(t'z_2)\, d(s,t).$$

Since $F_\phi(z_1,z_2)$ is an analytic function of (z_1,z_2), $F_\phi(z_1,z_1) = 0$ for all z_1 is equivalent to $0 = D^nF_\phi(0,0)(u_1,u_1)$, for all u_1 and $n = 0,1,2....$ Using the above expression of $D^nF_\phi(z_1,z_2)(u_1,u_2)$, this last condition is equivalent to

$$D^nF_\phi(0,0)(u_1,u_1) = \int \chi'u_1\,(t'u_1)^{n-1}\, d(s,t) = 0, \text{ for all } u_1 \text{ and } n = 1,2,... \tag{5.18}$$

To prove the lemma it is enough to show that there exists $\phi \in K$ for which condition (5.18) holds and $D^{n_0}F_\phi(0,0)(u_1^0,u_2^0) \neq 0$, for some $u_1^0, u_2^0 \in \mathbb{R}^k$ and $n_0 \in \mathbb{N}$. Defining the measure $d\nu(s,t) = exp\,(-\,(s's+t't)\,/2)\, d(s,t)$ and using the definition of χ, this is equivalent to

$$\int \phi(s,t)s'C_{\beta_0}u_1^0\left(t'u_2^0\right)^{n_0-1} d\nu(s,t) \neq 0. \tag{5.19}$$

Consider the following subspaces of $\mathcal{L}_1(\mathbb{R}^{2k})$:

$$\begin{aligned}\mathcal{M} &= \overline{\mathcal{L}}\left(\{s'u_1(t'u_1)^{n-1}; u_1 \in \mathbb{R}^k \text{ and } n \in \mathbb{N}\}\right)\\ \mathcal{N} &= \overline{\mathcal{L}}\left(\{s'u_1(t'u_2)^{n-1}; u_1,u_2 \in \mathbb{R}^k \text{ and } n \in \mathbb{N}\}\right),\end{aligned}$$

where the symbol $\overline{\mathcal{L}}(X)$ means the closure of the subspace generated by X. We claim that $\mathcal{M} \subsetneqq \mathcal{N}$. Indeed, since $k > 2$, the function $f_0(s,t) = s_1 t_2 \in \mathcal{N}$ is a non-null function orthogonal to all functions in $\mathcal{M}$: for each $i = 1, ..., k$ and $n \in \mathbb{N}$ we have that

$$\int f_0(s,t) s_i t_i^{n-1} d\nu = \begin{cases} \int t_2 s_1^2 t_1^{n-1} d\nu = 0, & \text{if } i = 1 \\ \int s_2 t_2^n d\nu = 0, & \text{if } i = 2 \\ \int s_1 t_2 s_i t_i^{n-1} d\nu = 0, & \text{if } i > 2. \end{cases}$$

That is, the function $f_0(s,t) = s_1 t_2$ is orthogonal to the generator set of $\mathcal{M}$ and then to all functions in $\mathcal{M}$. In particular, $f_0 \in \mathcal{N} \backslash \mathcal{M}$. By the Hahn-Banach Theorem (see Dunford and Schwartz (1988, p. 62)), there exists $\phi_0 \in \mathcal{L}_\infty(\mathbb{R}^{2k})$ such that

$$0 = \int \phi_0 f d\nu < \int \phi_0 f_0 d\nu,$$

for all $f \in \mathcal{M}$. Taking $\epsilon > 0$ sufficiently small, $\phi = \alpha + \epsilon \phi_0 \in \mathbb{K}$ and, since $\int f d\nu = 0$, for all $f \in \mathcal{N}$,

$$0 = \int \phi f d\nu < \int \phi f_0 d\nu,$$

for all $f \in \mathcal{M}$. Indeed, $0 = \int \phi f d\nu$, for all $f \in \mathcal{M}$, is equivalent to condition (5.18). Moreover, since $f_0 \in \mathcal{N}$ and $0 < \int \phi f_0 d\nu$, there must exist at least one element of generator set of $\mathcal{N}$, say $s' u_1^0 (t' u_2^0)^{n_0 - 1}$ for some $u_1^0, u_2^0 \in \mathbb{R}^k$ and $n_0 \in \mathbb{N}$, such that condition (5.19) holds. □

Proof of Proposition 4. Let us find the test which maximizes power for an alternative (β, μ') given the constraints in (2.4). This test rejects the null when

$$\frac{f_{\beta,\mu}(s,t)}{f_{\beta_0}(s,t)} > \widetilde{c_0}(t) + \widetilde{c_1}(t)' s,$$

where the multipliers $\widetilde{c_0}(t)$ and $\widetilde{c_1}(t)$ satisfy the boundary restrictions in (2.4). Algebraic manipulations show that the rejection region is given by

$$\exp\left(s' (\beta - \beta_0) C_{\beta_0} \mu\right) > c_0(t) + s' c_1(t),$$

where $\widetilde{c_0}(t)$ and $\widetilde{c_1}(t)$ are chosen to satisfy (2.4). For $\widetilde{c_0}(t) = c_0$ and $\widetilde{c_1}(t) = c_1 \times (\beta - \beta_0) C_{\beta_0} \mu$,

$$\exp\left[s' (\beta - \beta_0) C_{\beta_0} \mu\right] > c_0 + c_1 \times s' (\beta - \beta_0) C_{\beta_0} \mu.$$

We now choose the constants c_0 and c_1 so that the rejection region is given by

$$\frac{\left(s'C_{\beta_0}\mu\right)^2}{\mu C_{\beta_0}^2\mu} > q(1),$$

where $q(1)$ is the α quantile of a chi-square-one distribution. □

Proof of Remark 1. Expression (4.10) is equivalent to

$$1 - \inf_{\phi\in\Gamma_1(g,\gamma)} 1 - \int \phi h.$$

Because h is a density and $0 \leq \phi \leq 1$, this is the same as

$$1 - \inf_{\phi\in\mathcal{L}_1(Y,h)} \int |\phi - 1|\, h \text{ where } 0 \leq \phi \leq 1 \text{ and } \int \phi g_v \in [\gamma_v^1, \gamma_v^2],\ \forall v \in \mathbf{V},$$

as we wanted to show. Furthermore, define $H(\phi^*) = \sup_{\phi\in\Gamma_1(g,\gamma)} \langle \phi, \phi^* \rangle$ as the support functional of $\Gamma_1(g,\gamma)$. By Theorem 1 of Luenberger (1969, p. 136), this expression is equal to

$$1 - \max_{\|\phi^*\|_\infty^h \leq 1} \left[\langle \phi, \phi^* \rangle - H(\phi^*)\right],$$

and the maximum is attained by some $\overline{\phi}^* \in \mathcal{L}_\infty(Y,h)$. Furthermore,

$$\int \left(\overline{\phi} - 1\right)\left(-\overline{\phi}^*\right) h = \left\|1 - \overline{\phi}\right\|_1^h \cdot \left\|\overline{\phi}^*\right\|_\infty^h$$

for $\overline{\phi} \in \mathcal{L}_1(Y,h)$ which solves the optimization problem. □

Proof of Proposition 5. It is straightforward to show that $\Gamma_2(g,\gamma)$ is convex. Now, let (ϕ_n) be any sequence in $\Gamma_2(g,\gamma)$ that converges to ϕ in the $\mathcal{L}_2(Y,h)$ topology: $\int (\phi_n - \phi)^2 h \to 0$. It is trivial to show that $0 \leq \phi \leq 1$. We need to show that $\int \phi g_v \in [\gamma_v^1, \gamma_v^2]$, $\forall v \in \mathbf{V}$. We note that

$$\int \phi g_v = \int \phi \frac{g_v}{h} h \leq \left(\int \phi^2 h\right)^{1/2} \left(\int \left(\frac{g_v}{h}\right)^2 h\right)^{1/2} < \infty$$

and $\int (\phi_n)^2 h \leq 1$. By the Banach-Alaoglu Theorem, we select a subsequence $\left(\phi_{n_k}\right)$ that converges in the weak* topology: $\int \phi_{n_k} g_v \to \int \phi g_v$ for every

$v \in \mathbf{V}$. Trivially, $\int \phi g_v \in [\gamma_v^1, \gamma_v^2]$, $\forall v \in \mathbf{V}$. Hence, $\Gamma_2(g, \gamma)$ is also closed. The result now follows from Theorem 1 of Luenberger (1969, p. 69). $\square$

Proof of Proposition 6. For part (a), consider the problem

$$\inf_{\phi \in \Gamma_2(g,\gamma)} \int \left(\epsilon\phi - \frac{1}{2}\right)^2 h. \tag{5.20}$$

By the Banach-Alaoglu Theorem, the set of all measurable functions ϕ where $0 \leq \phi \leq 1$ and $\int \phi g_v \in [\gamma_v^1, \gamma_v^2]$, $\forall v \in \mathbf{V}$, is closed in $\mathcal{L}_2(Y, h)$. This is a minimum-norm problem in a Hilbert space. By Theorem 1 of Luenberger (1969, p. 69), there exists a unique $\overline{\phi}_\delta$ that attains the infimum of (5.20). Now,

$$\int \left(\epsilon\phi - \frac{1}{2}\right)^2 h = \epsilon^2 \int \phi^2 h - \epsilon \int \phi h + \frac{1}{4} \int h.$$

Finding its minimum is the same as finding the minimum of

$$\epsilon \int \phi^2 h - \int \phi h.$$

A necessary and sufficient condition for $\overline{\phi}_\epsilon$ to be optimal is that

$$\int \left(\frac{1}{2\epsilon} - \overline{\phi}_\epsilon\right) \left(\phi - \overline{\phi}_\epsilon\right) h \leq 0,$$

for all $\phi \in \Gamma_2(g, \gamma)$.

For part (b), $\int |\phi_n - \phi| \, h \to 0$ holds when $\int (\phi_n - \phi)^2 h \to 0$. Therefore, $\int \phi_n h \to \int \phi h$ and $\int \phi_n^2 h \to \int \phi^2 h$. The objective function given in (4.13) is then continuous in (ϕ, δ). The result now follows from the Maximum Theorem of Berge (1997, p. 116).

For part (c), continuity of $\overline{\phi}_\epsilon$ follows again from the Maximum Theorem. Since $\overline{\phi}_\epsilon$ is bounded and $\overline{\phi}_\epsilon \to \overline{\phi}$ in $\mathcal{L}_2(Y, h)$, we have $\overline{\phi}_\epsilon \to \overline{\phi}$ in $\mathcal{L}_\infty(Y, h)$. This implies that $\int \overline{\phi}_\epsilon g_v \to \int \overline{\phi} g_v$ for every $v \in \mathbf{V}$; see Theorem 4.3 of Rudin (1991). $\square$

Appendix A: HAC-IV

We can write

$$\Omega = \begin{bmatrix} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{bmatrix} P \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix} P' \begin{bmatrix} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{bmatrix},$$

where P is an orthogonal matrix and $\rho = \omega_{12}/\omega_{11}^{1/2}\omega_{22}^{1/2}$. For the numerical simulations, we specify $\omega_{11} = \omega_{22} = 1$.

We use the decomposition of Ω to perform numerical simulations for a class of covariance matrices:

$$\Sigma = P \begin{bmatrix} 1+\rho & 0 \\ 0 & 0 \end{bmatrix} P' \otimes diag\left(\varsigma_1\right) + P \begin{bmatrix} 0 & 0 \\ 0 & 1-\rho \end{bmatrix} P' \otimes diag\left(\varsigma_2\right),$$

where ς_1 and ς_2 are k-dimensional vectors.

We consider two possible choices for ς_1 and ς_2. For the first design, we set $\varsigma_1 = \varsigma_2 = (1/\varepsilon - 1, 1, ..., 1)'$. The covariance matrix then simplifies to a Kronecker product: $\Sigma = \Omega \otimes diag\left(\varsigma_1\right)$. For the non-Kronecker design, we set $\varsigma_1 = (1/\varepsilon - 1, 1, ..., 1)'$ and $\varsigma_2 = (1, ..., 1, 1/\varepsilon - 1)'$. This setup captures the data asymmetry in extracting information about the parameter β from each instrument. For small ε, the angle between ς_1 and ς_2 is nearly zero. We report numerical simulations for $\varepsilon = (k+1)^{-1}$. As k increases, the vector ς_1 becomes orthogonal to ς_2 in the non-Kronecker design.

We set the parameter $\mu = \left(\lambda^{1/2}/\sqrt{k}\right) 1_k$ for $k = 2, 5, 10, 20$ and $\rho = -0.5, 0.2, 0.5, 0.9$. We choose $\lambda/k = 0.5, 1, 2, 4, 8, 16$, which span the range from weak to strong instruments. We focus on tests with significance level 5% for testing $\beta_0 = 0$.

We report power plots for the power envelope (thick solid dark blue line) and the following tests: AR (thin solid red line), LM (dashed pink line), MM1 (dash-dot green line), MM1-SU (dotted black line), MM1-LU (solid light blue line with bars), MM2 (thin purple line with asterisks), MM2-SU (thick light brown line with asterisks), MM2-LU (dark brown dashed line).

Summary of findings.

1. The AR test has power close to the power envelope when k is small. When the number of instruments is large ($k = 10, 20$), its power is considerably lower than the power envelope. These two facts about the AR test are true for the Kronecker and non-Kronecker designs.

2. The LM test has power considerably below the power envelope when λ/k is small for both Kronecker and non-Kronecker designs. Its power is also non-monotonic as β increases (in absolute value). This test has power close to the power envelope for alternatives near $\beta_0 = 0$ when instruments are strong ($\lambda/k = 8, 16$).

3. In both Kronecker and non-Kronecker designs, the MM1 similar test is biased. This test behaves more like a one-sided test for alternatives near the null with bias increasing as λ/k grows.

4. In the Kronecker design, the MM2 similar test has power considerably closer to the power envelope than the AR and LM tests. In the non-Kronecker design, the MM2 similar tests is biased. This test behaves more like a one-sided test with bias increasing as λ/k grows.

5. The MM1-LU and MM2-LU tests have power closer to the power envelope than the AR and LM tests for both Kronecker and non-Kronecker designs.

6. The MM1-SU and MM2-SU tests have power very close to the MM1-LU and MM2-LU tests in most designs. Hence, the potential power loss in using the SU condition seems negligible. This suggests that the MM1-SU and MM2-SU tests are nearly admissible. Because the SU tests are easier to implement than the LU tests, we recommend the use of MM1-SU and MM2-SU tests in empirical practice.

Figure 1: Power Comparison (Kronecker Covariance) $k = 2, \rho = -0.5$

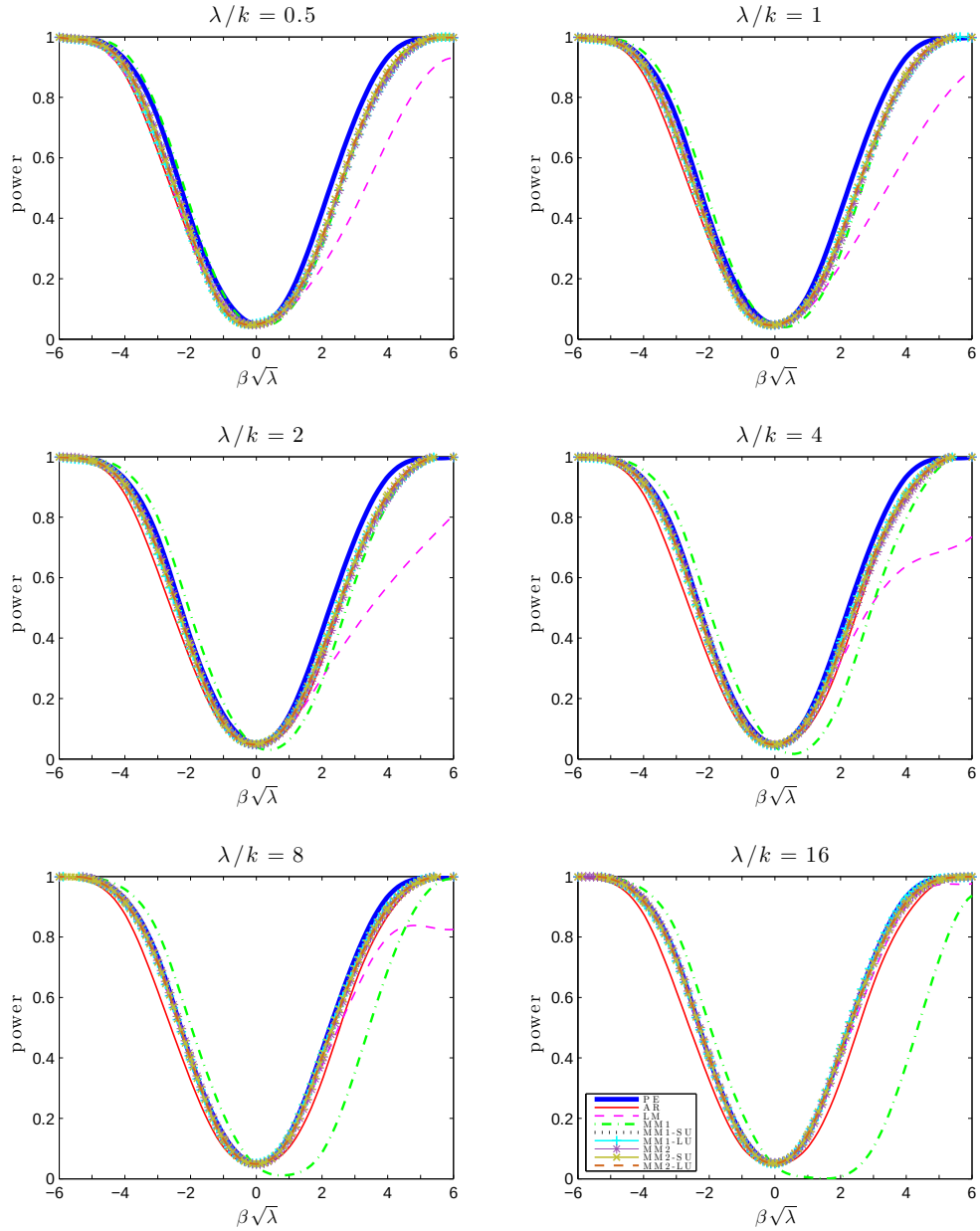


Figure 2: Power Comparison (Kronecker Covariance) $k = 2, \rho = 0.2$

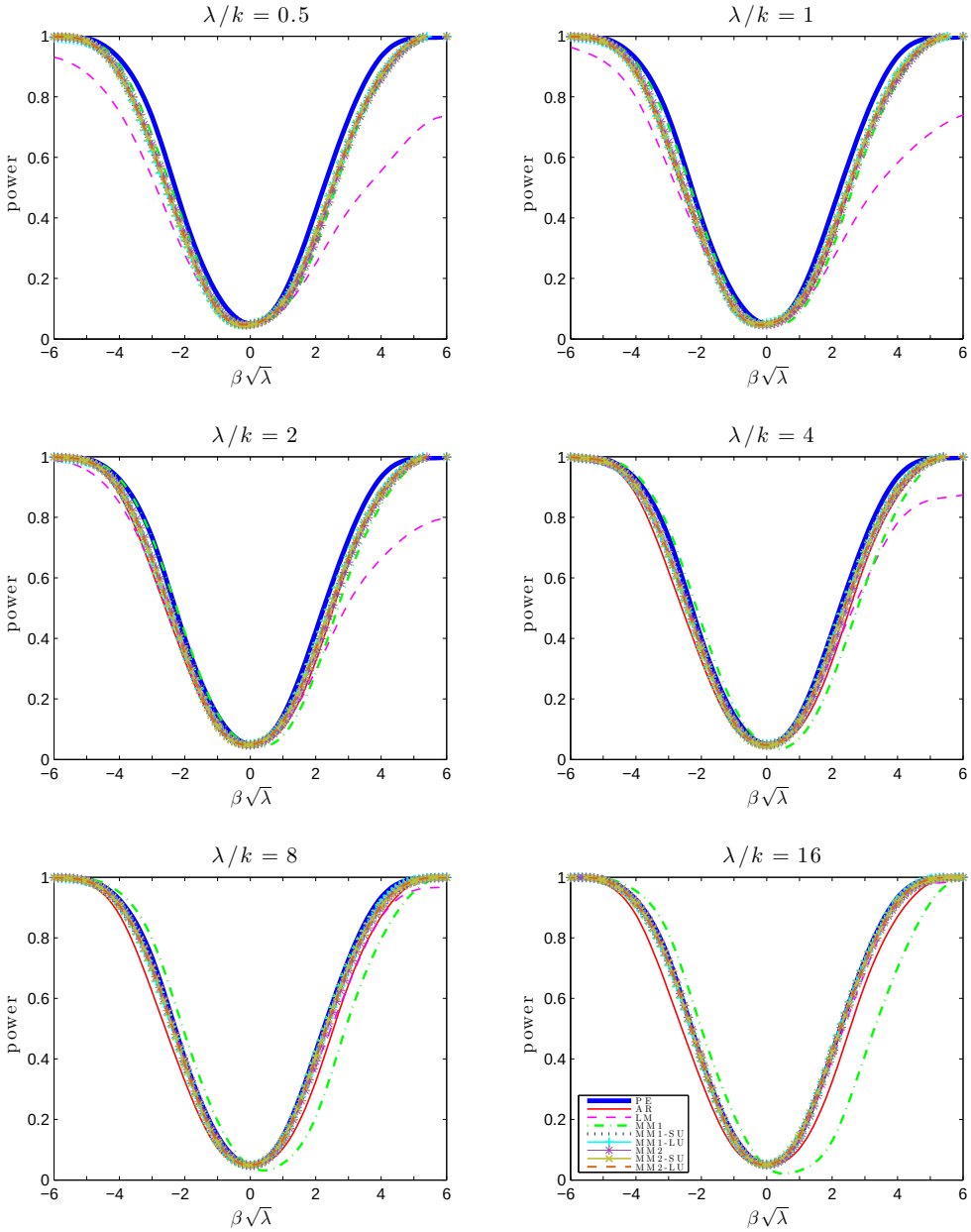


Figure 3: Power Comparison (Kronecker Covariance) $k = 2, \rho = 0.5$

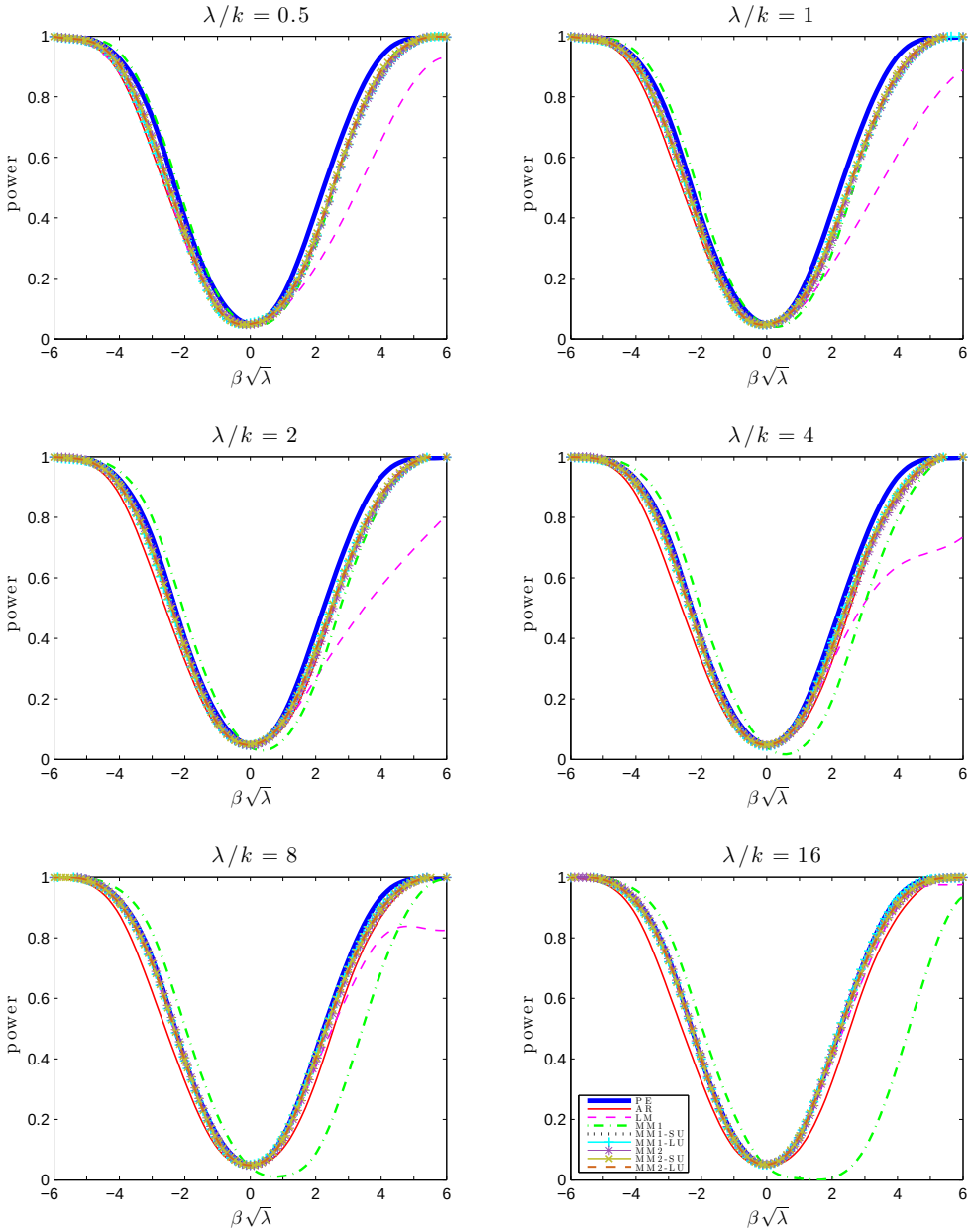


Figure 4: Power Comparison (Kronecker Covariance) $k = 2, \rho = 0.9$

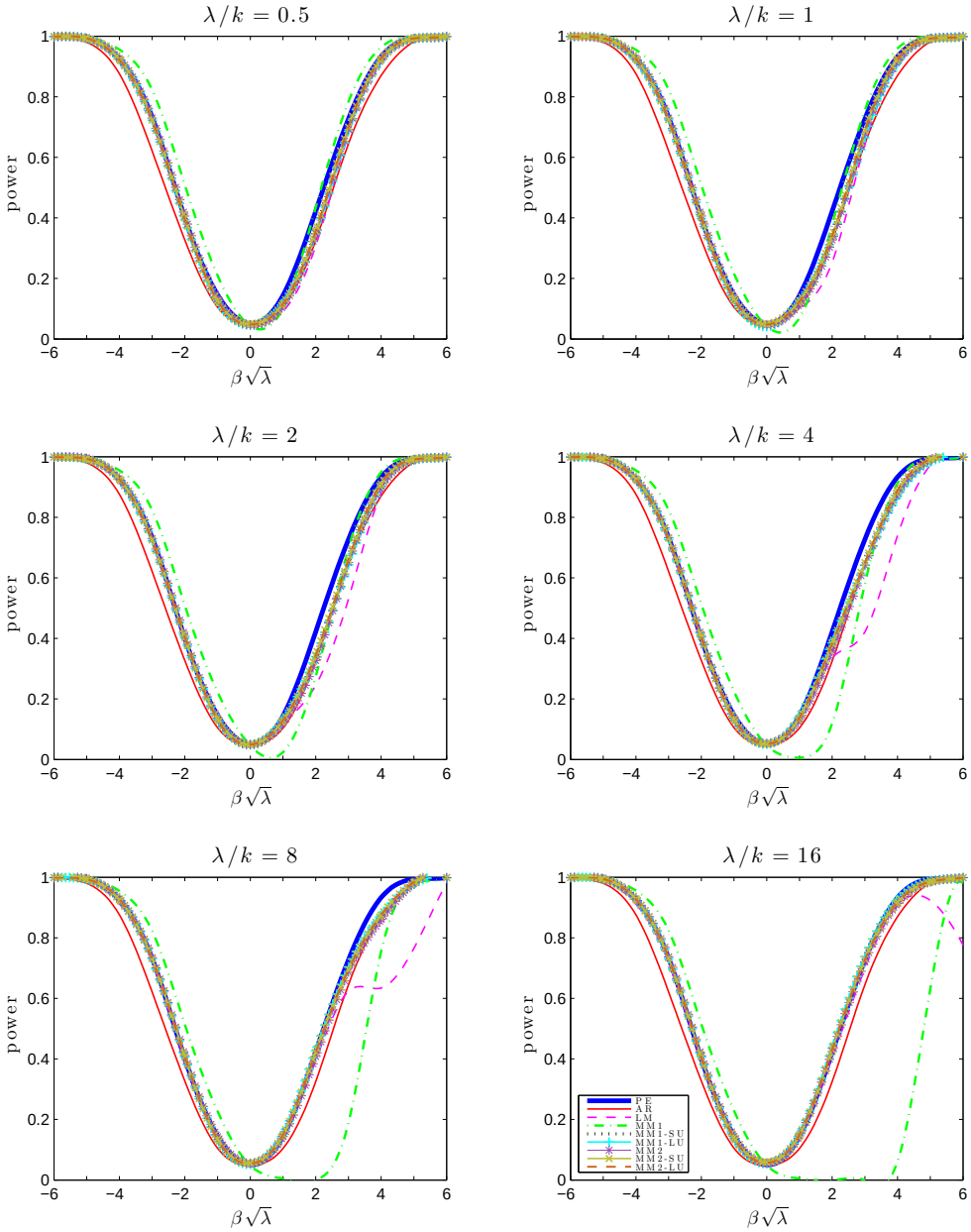


Figure 5: Power Comparison (Kronecker Covariance) $k = 5, \rho = -0.5$

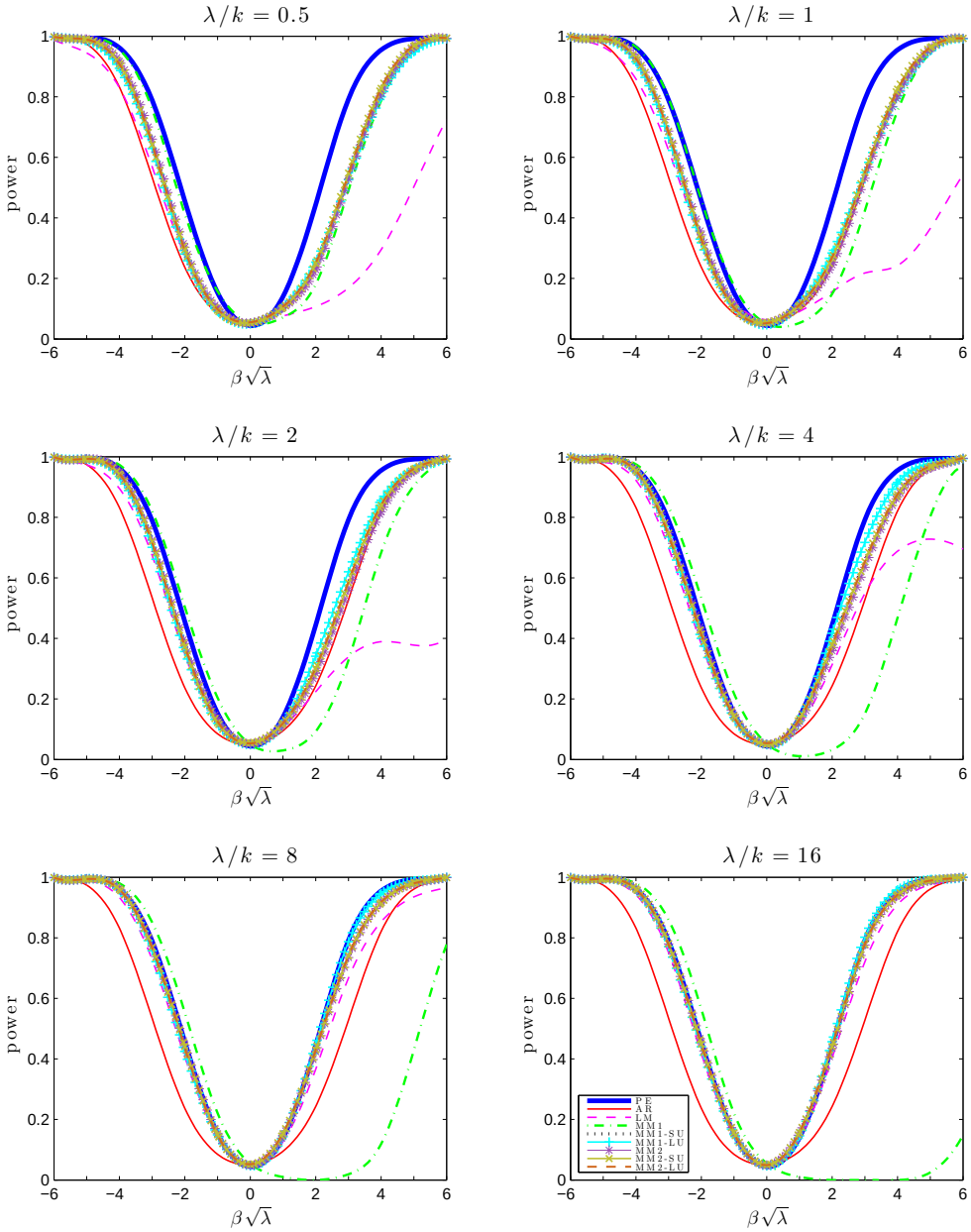


Figure 6: Power Comparison (Kronecker Covariance) $k = 5, \rho = 0.2$

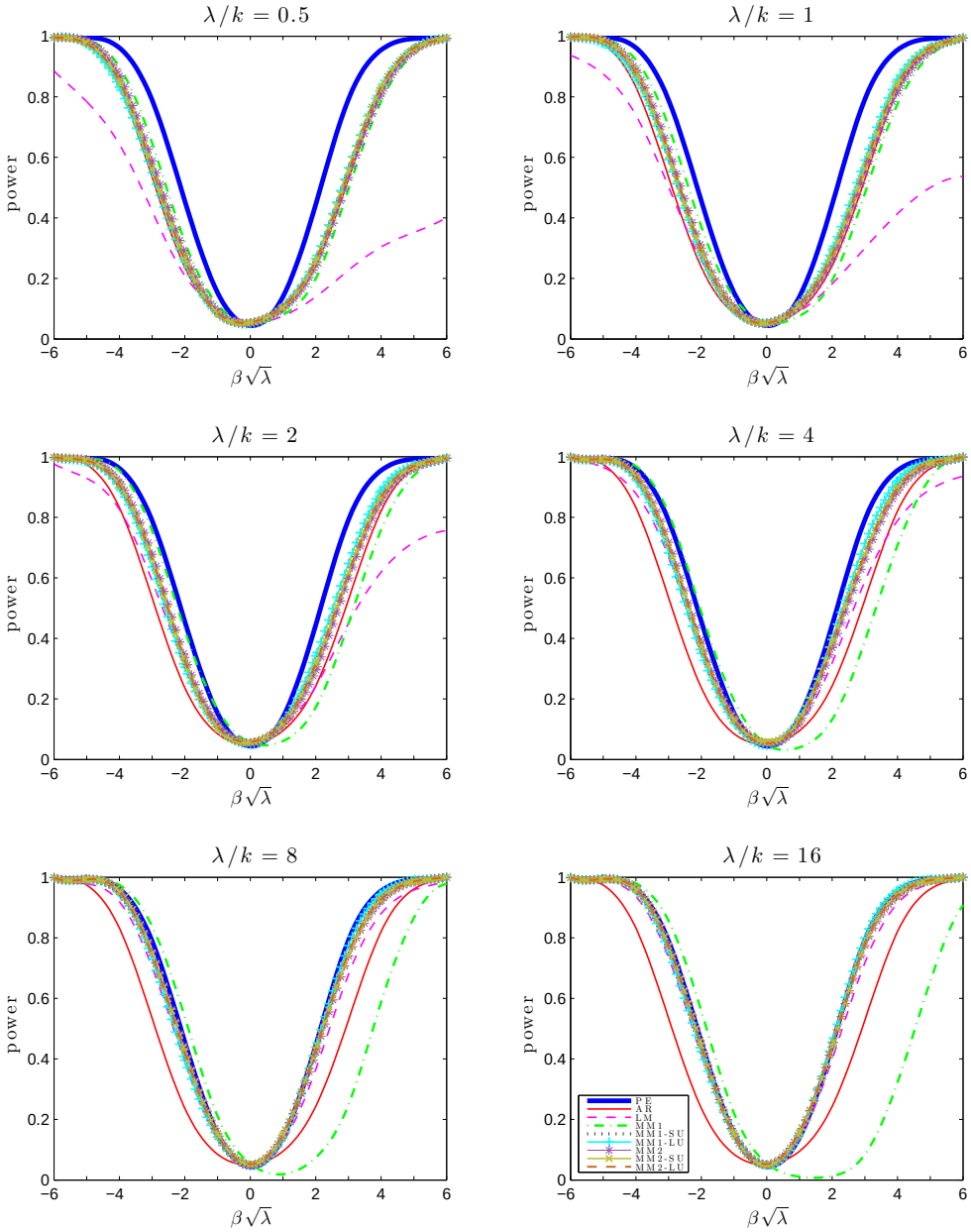


Figure 7: Power Comparison (Kronecker Covariance) $k = 5, \rho = 0.5$

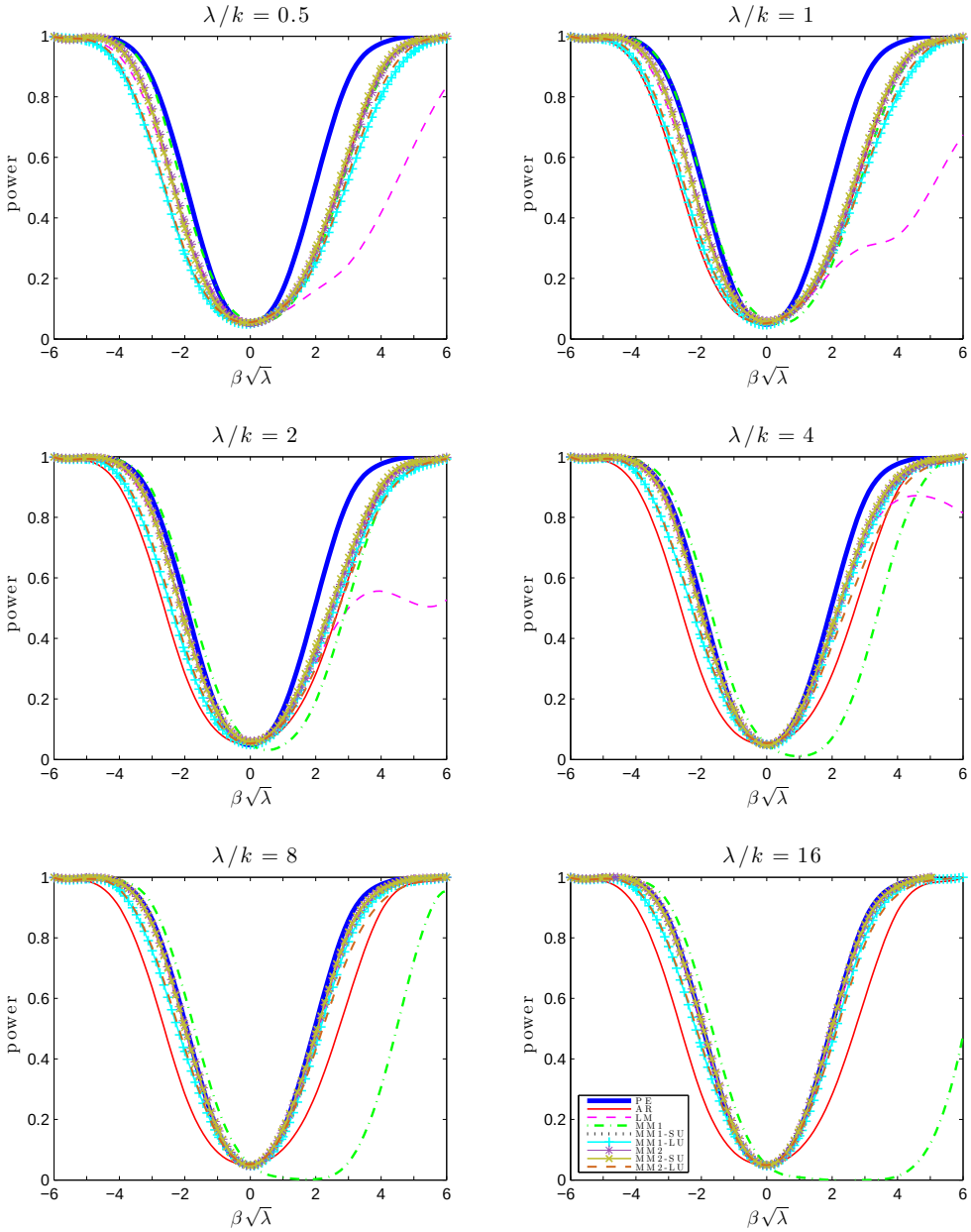


Figure 8: Power Comparison (Kronecker Covariance) $k = 5, \rho = 0.9$

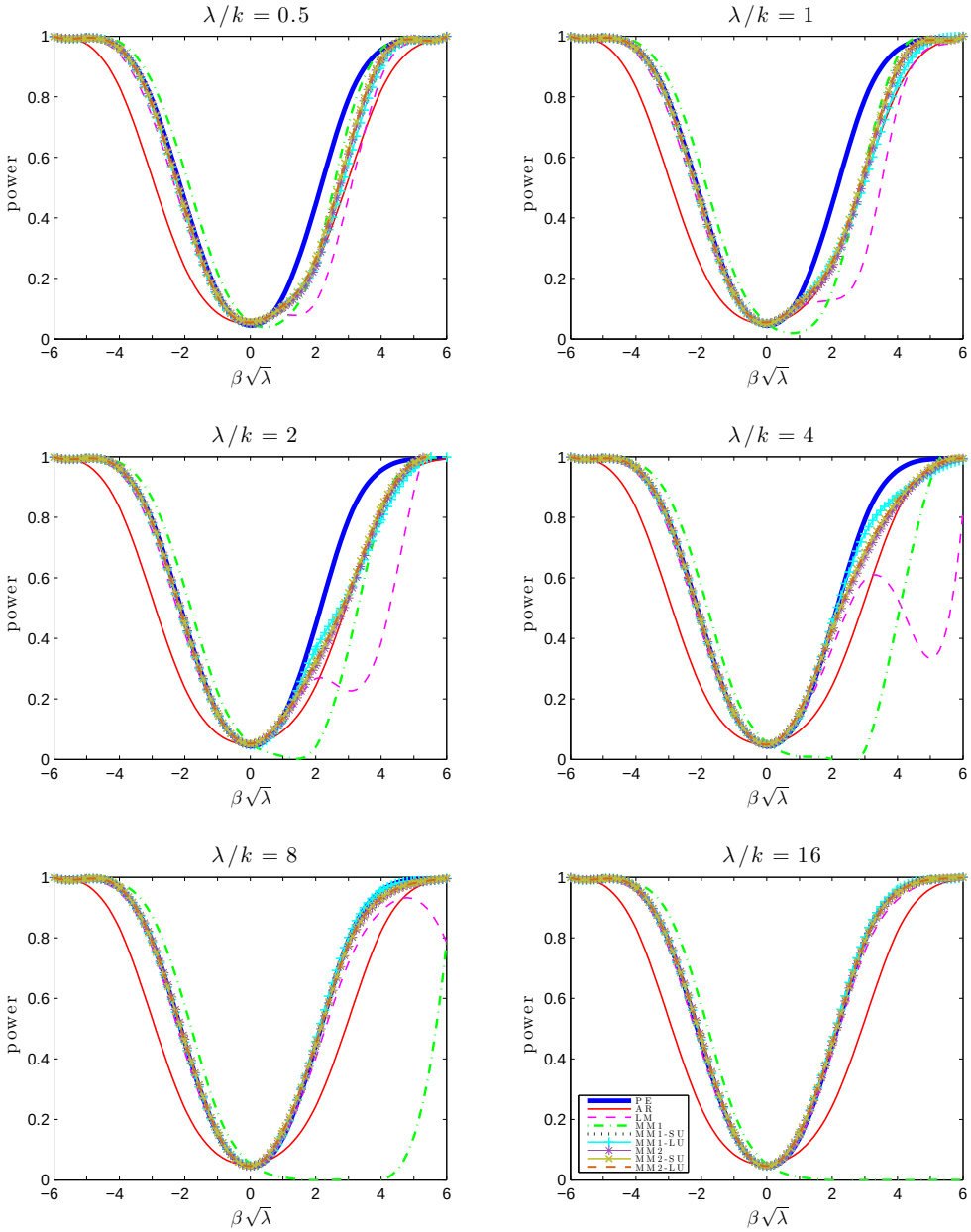


Figure 9: Power Comparison (Kronecker Covariance) $k = 10, \rho = -0.5$

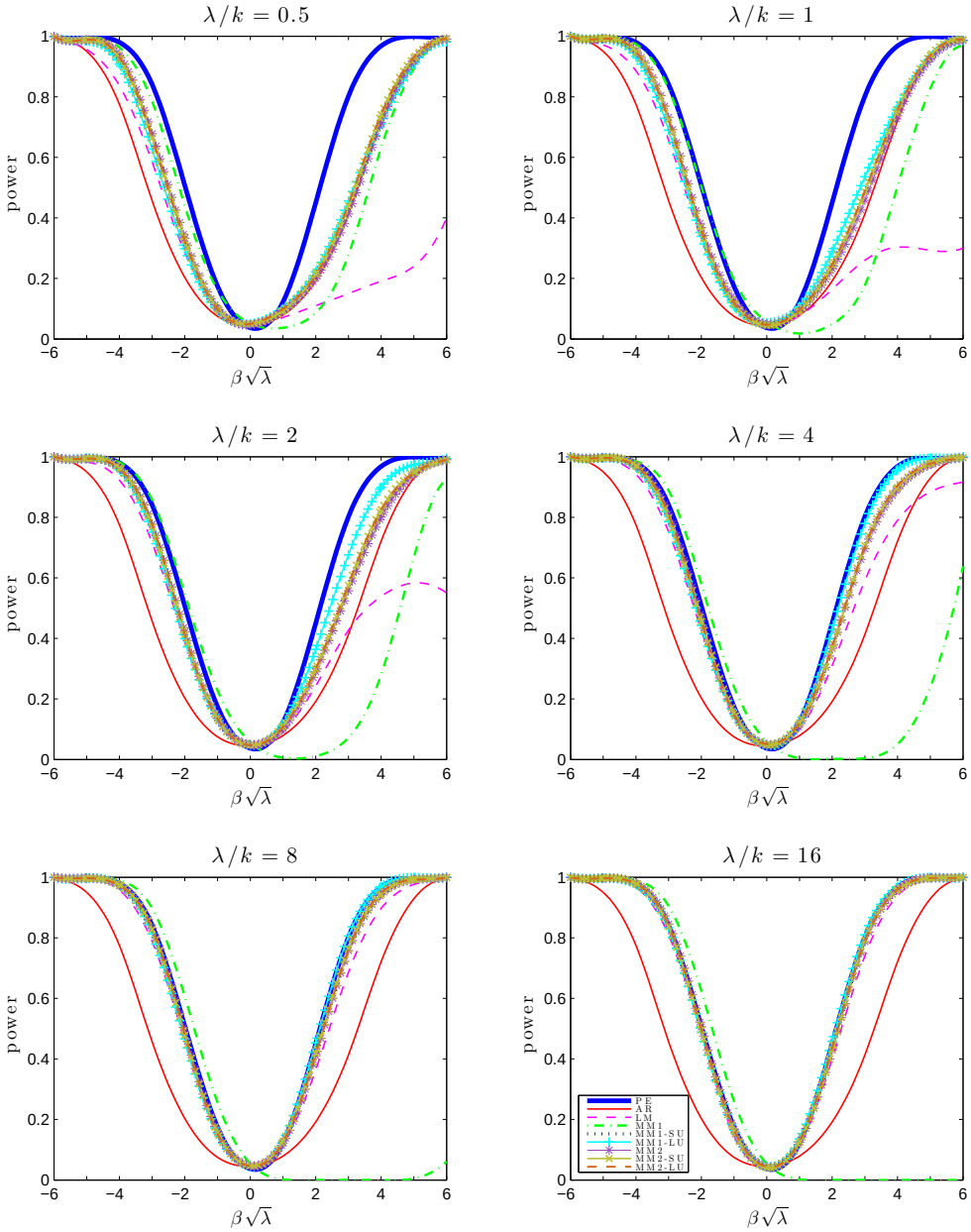


Figure 10: Power Comparison (Kronecker Covariance) $k = 10, \rho = 0.2$

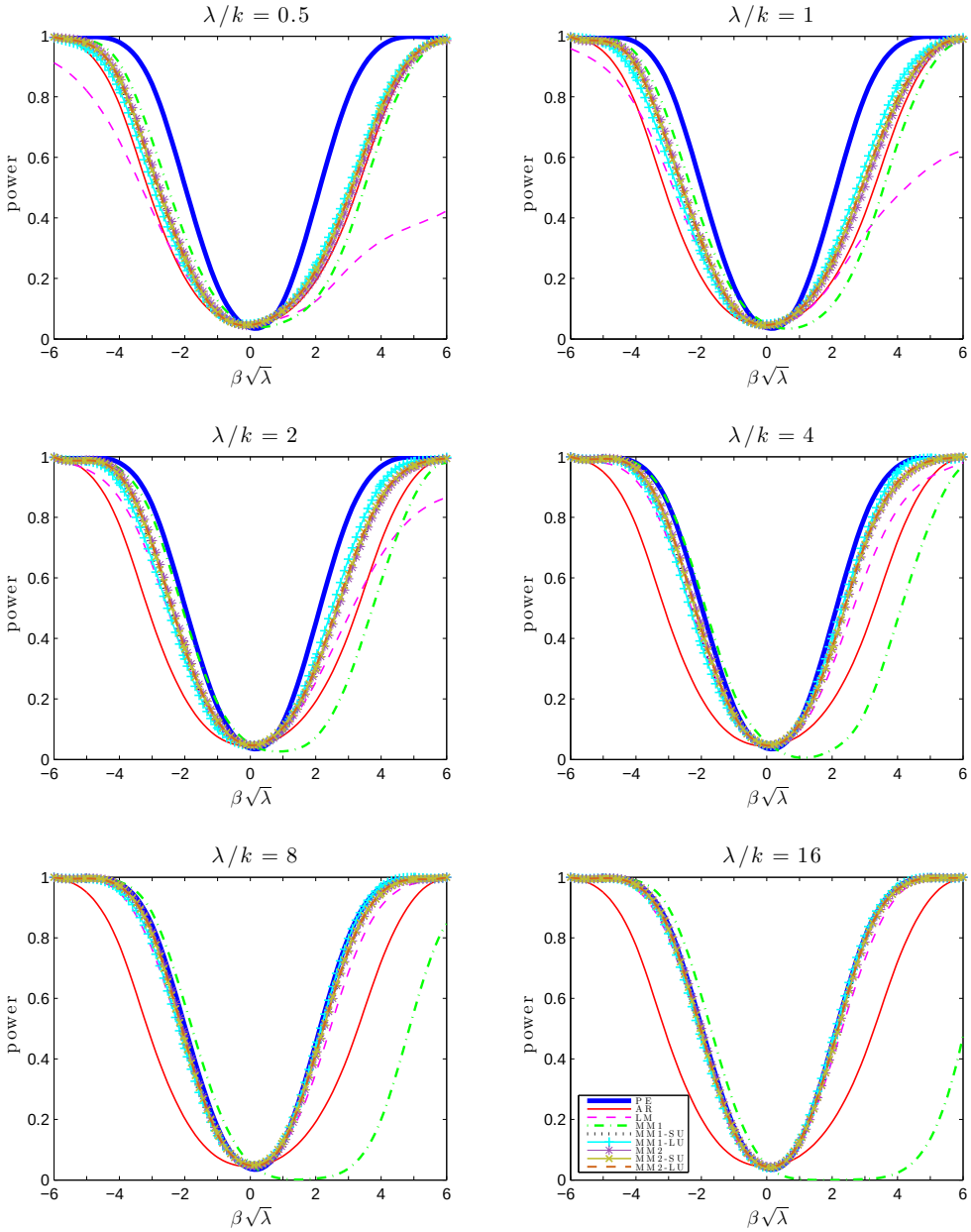


Figure 11: Power Comparison (Kronecker Covariance) $k = 10, \rho = 0.5$

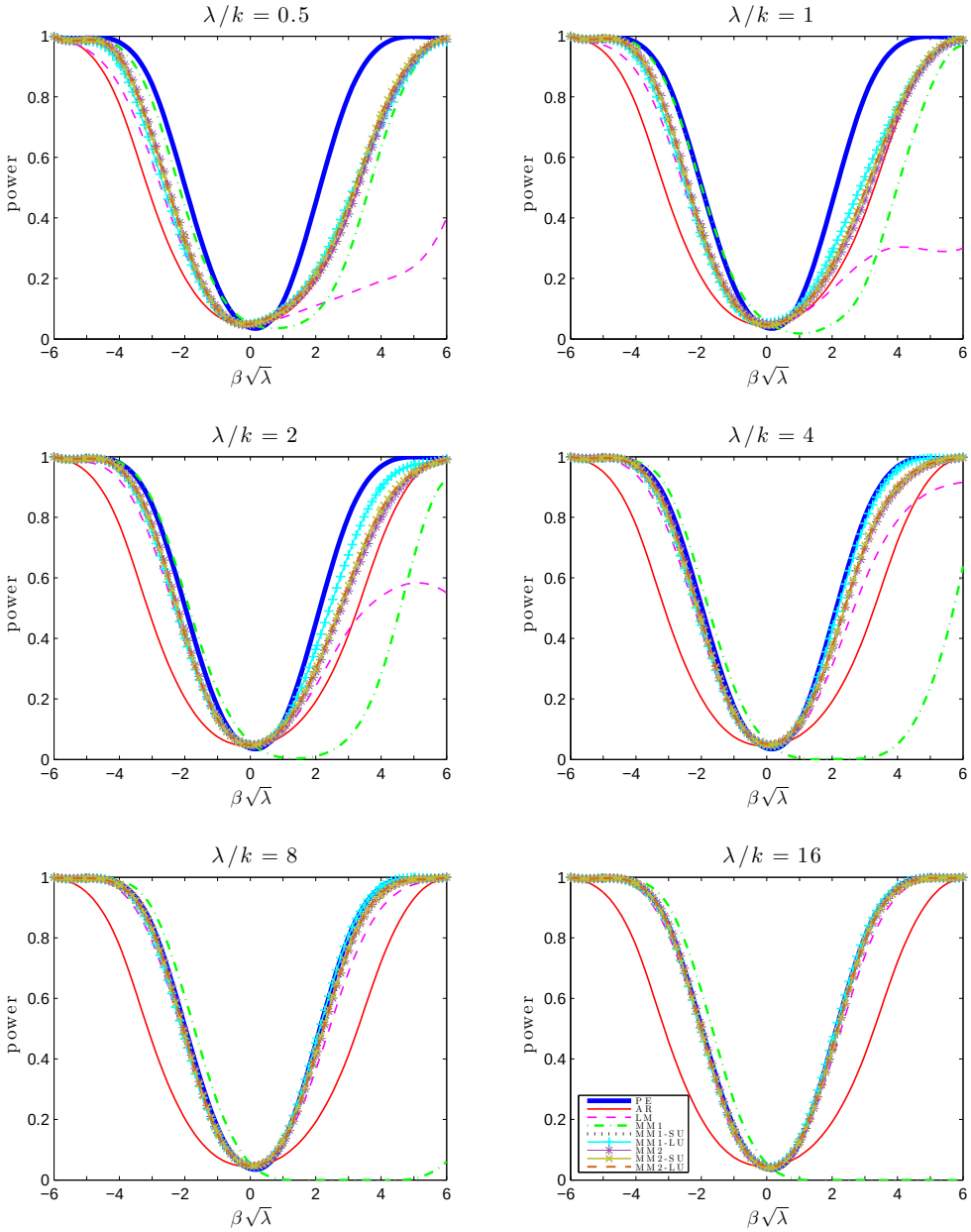


Figure 12: Power Comparison (Kronecker Covariance) $k = 10, \rho = 0.9$

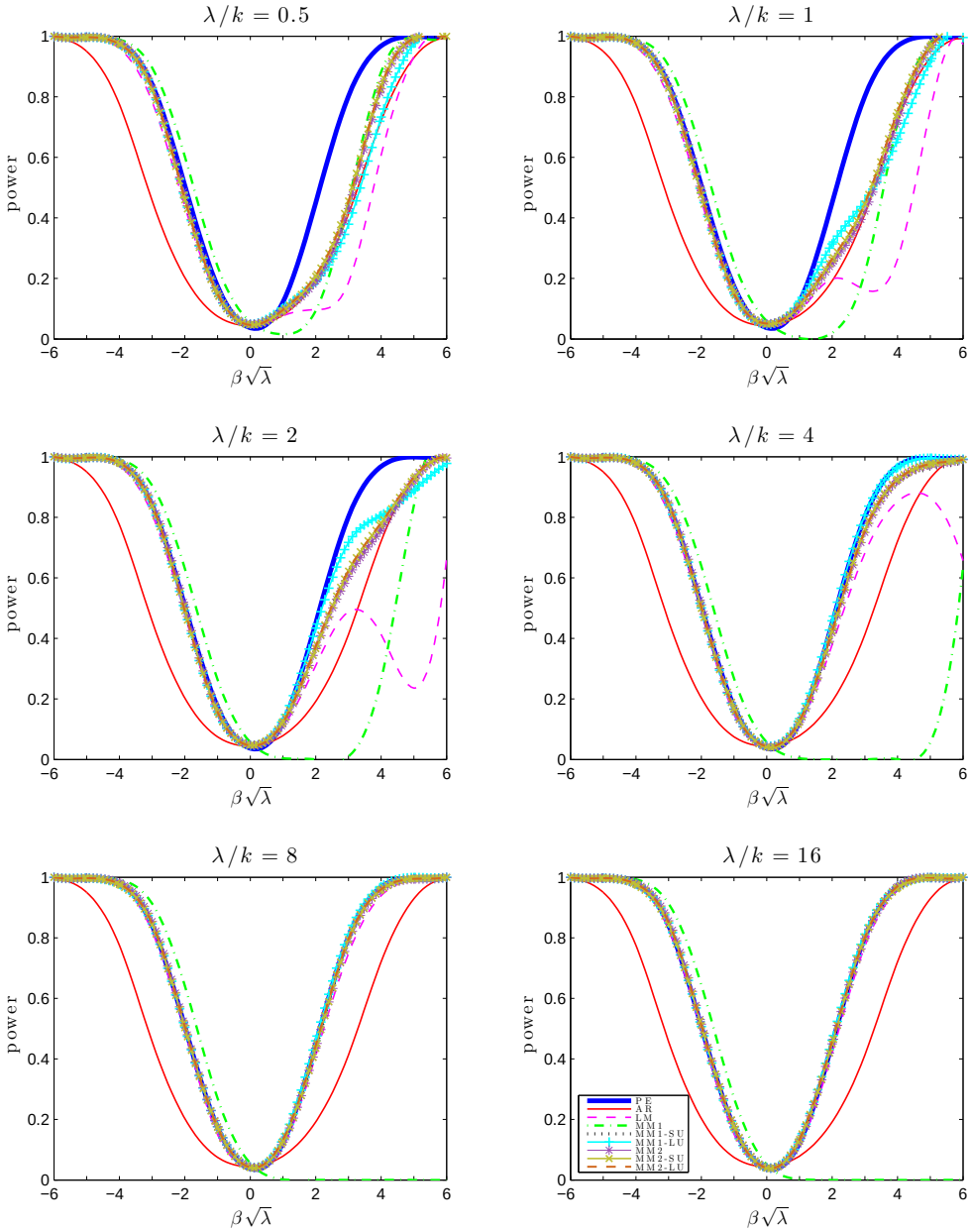


Figure 13: Power Comparison (Kronecker Covariance) $k = 20, \rho = -0.5$

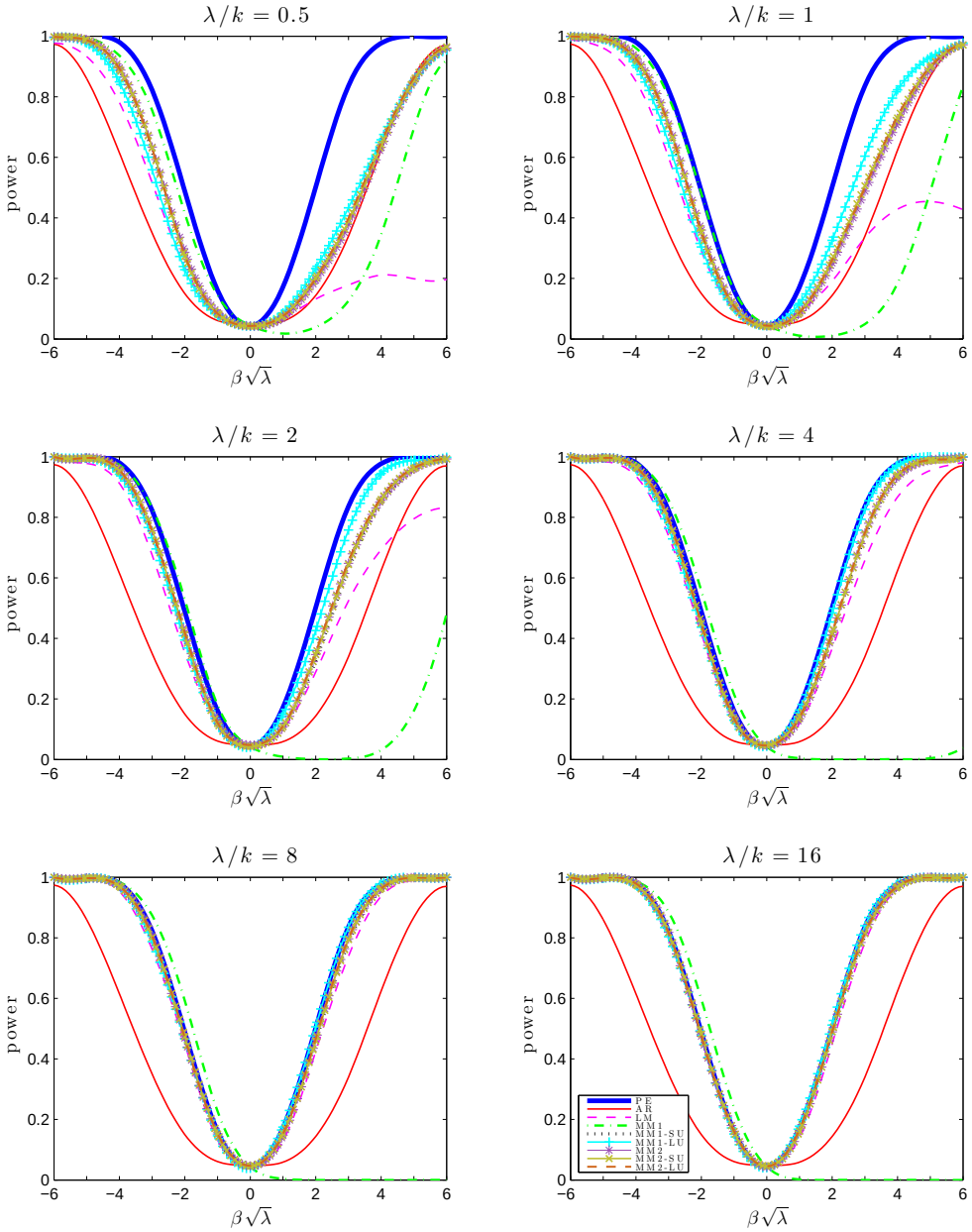


Figure 14: Power Comparison (Kronecker Covariance) $k = 20, \rho = 0.2$

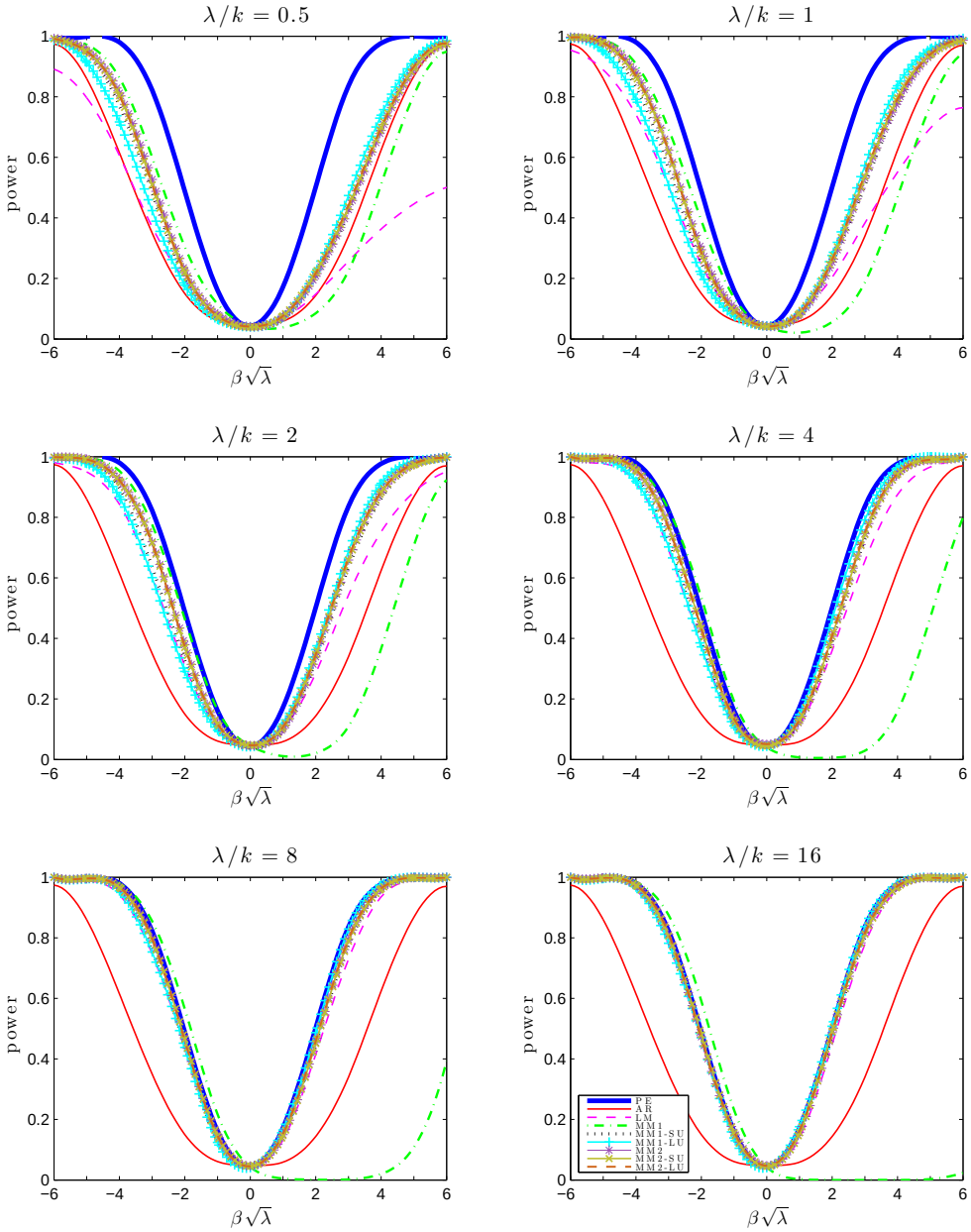


Figure 15: Power Comparison (Kronecker Covariance) $k = 20, \rho = 0.5$

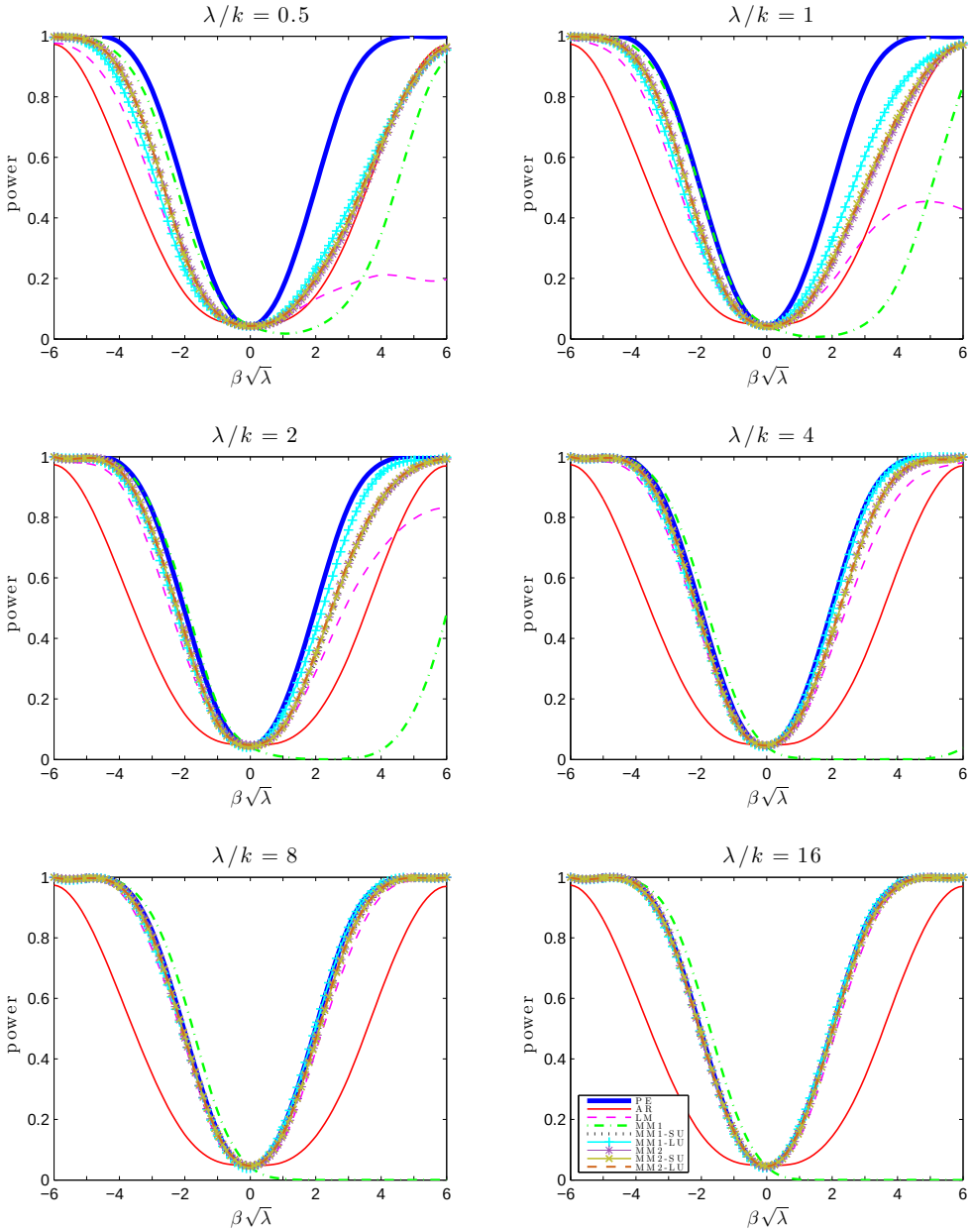


Figure 16: Power Comparison (Kronecker Covariance) $k = 20, \rho = 0.9$

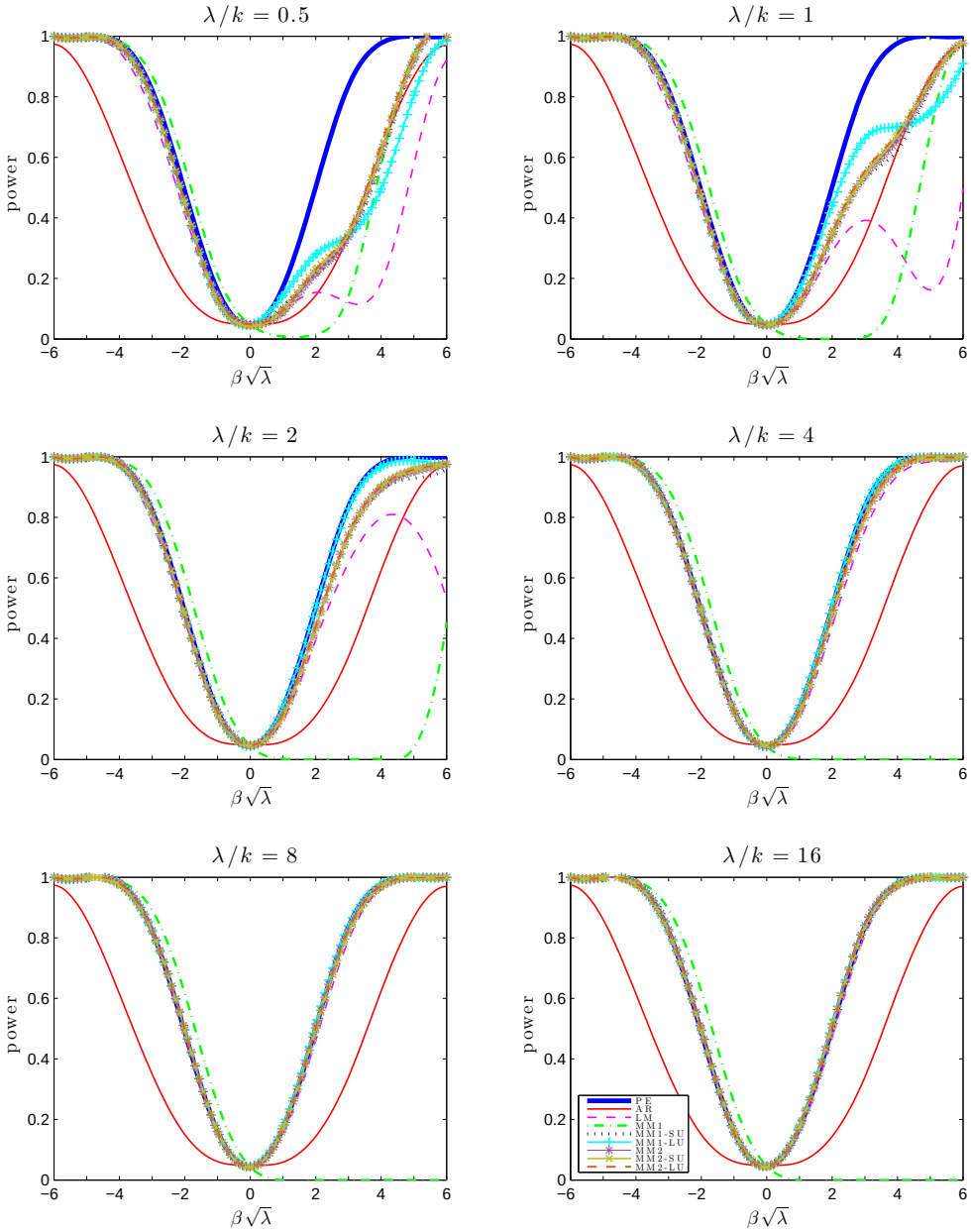


Figure 17: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = -0.5$

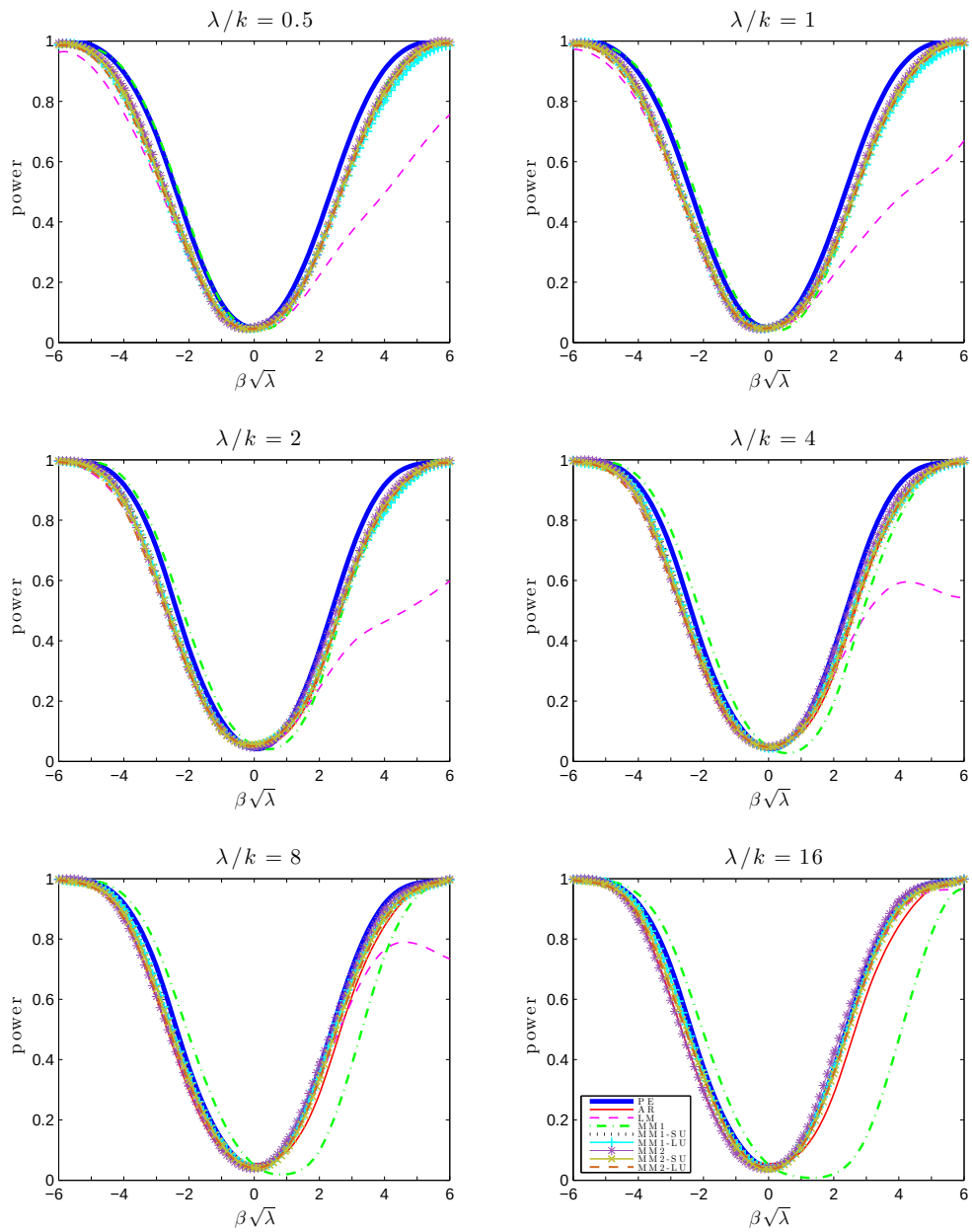


Figure 18: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = 0.2$

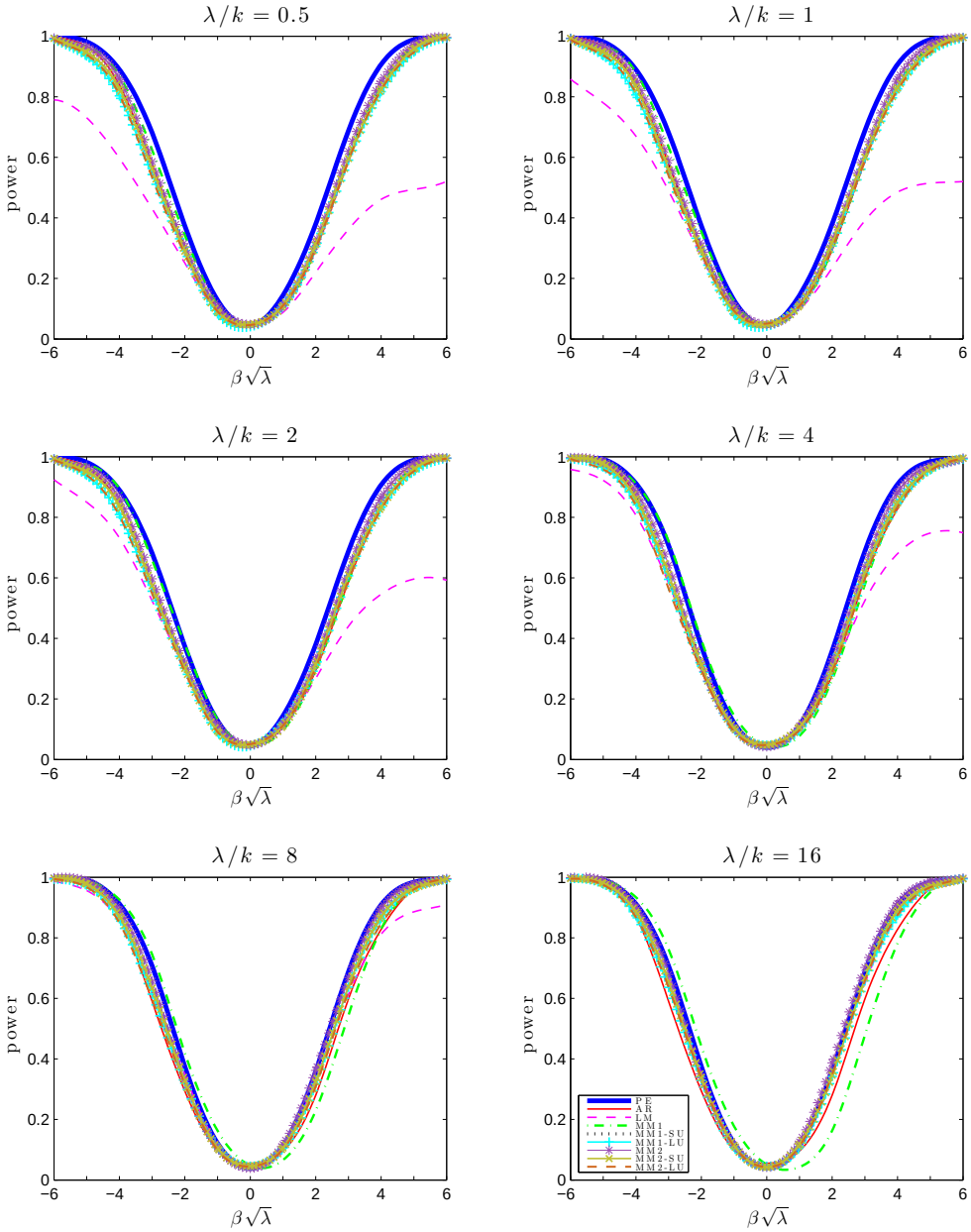


Figure 19: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = 0.5$

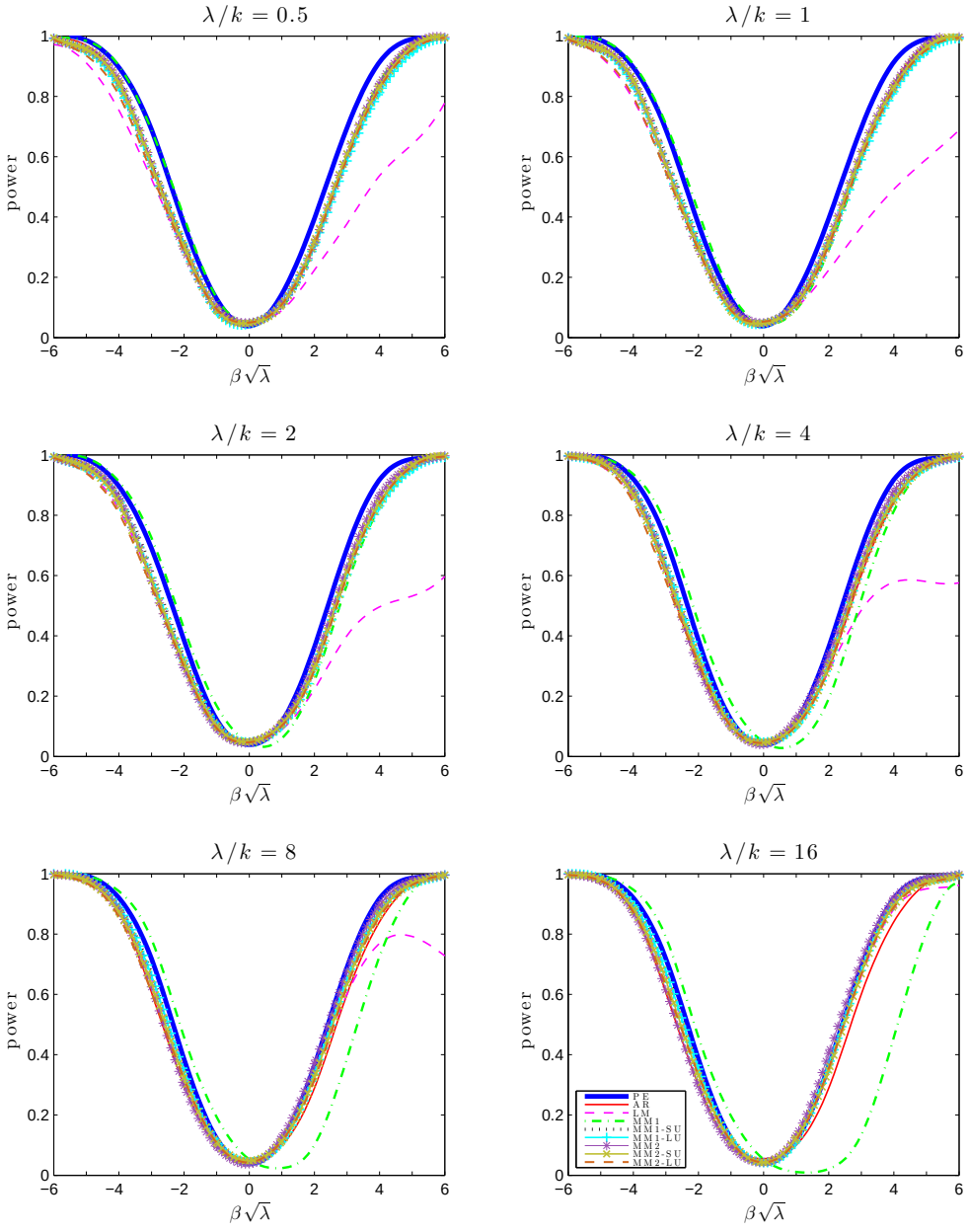


Figure 20: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = 0.9$

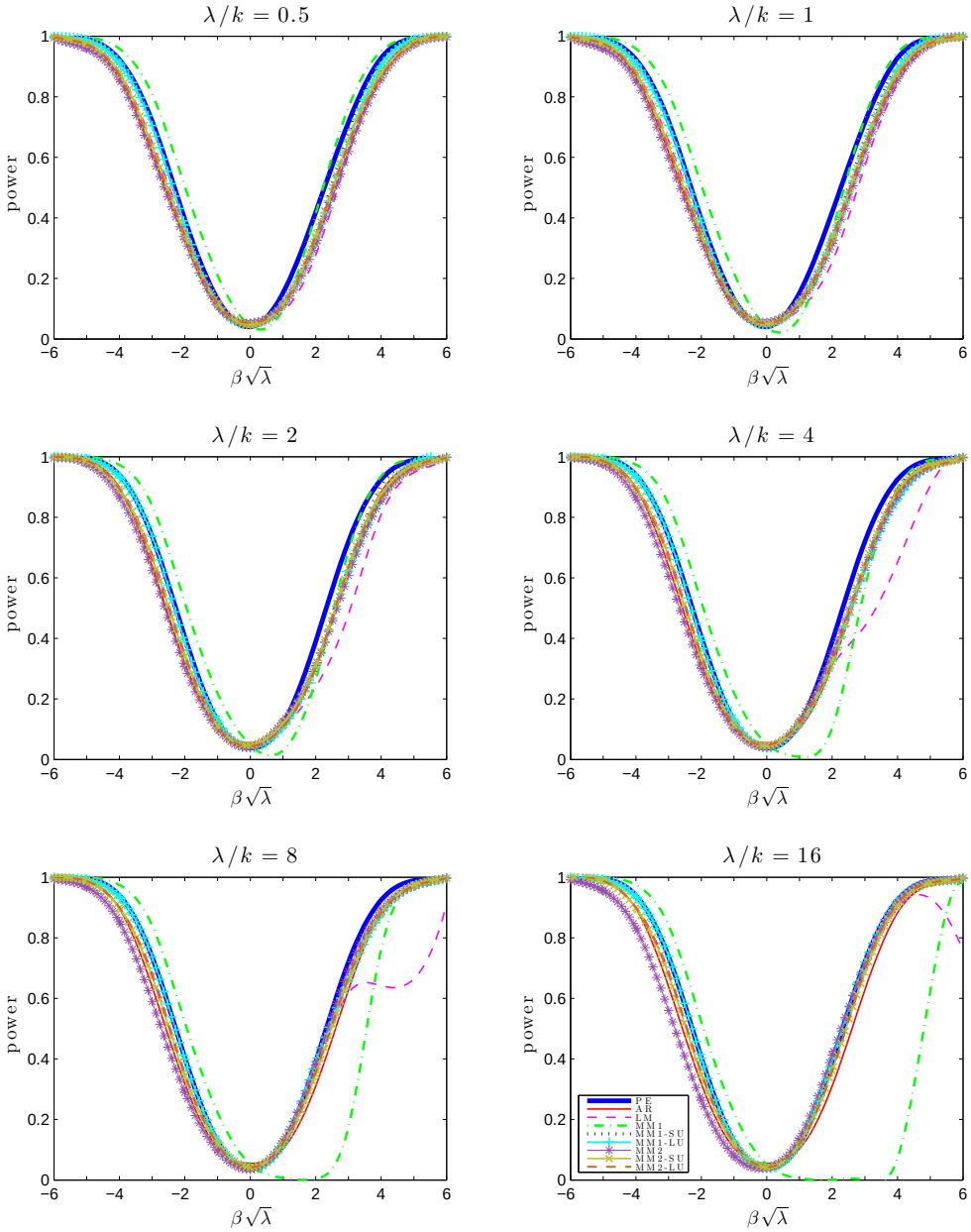


Figure 21: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = -0.5$

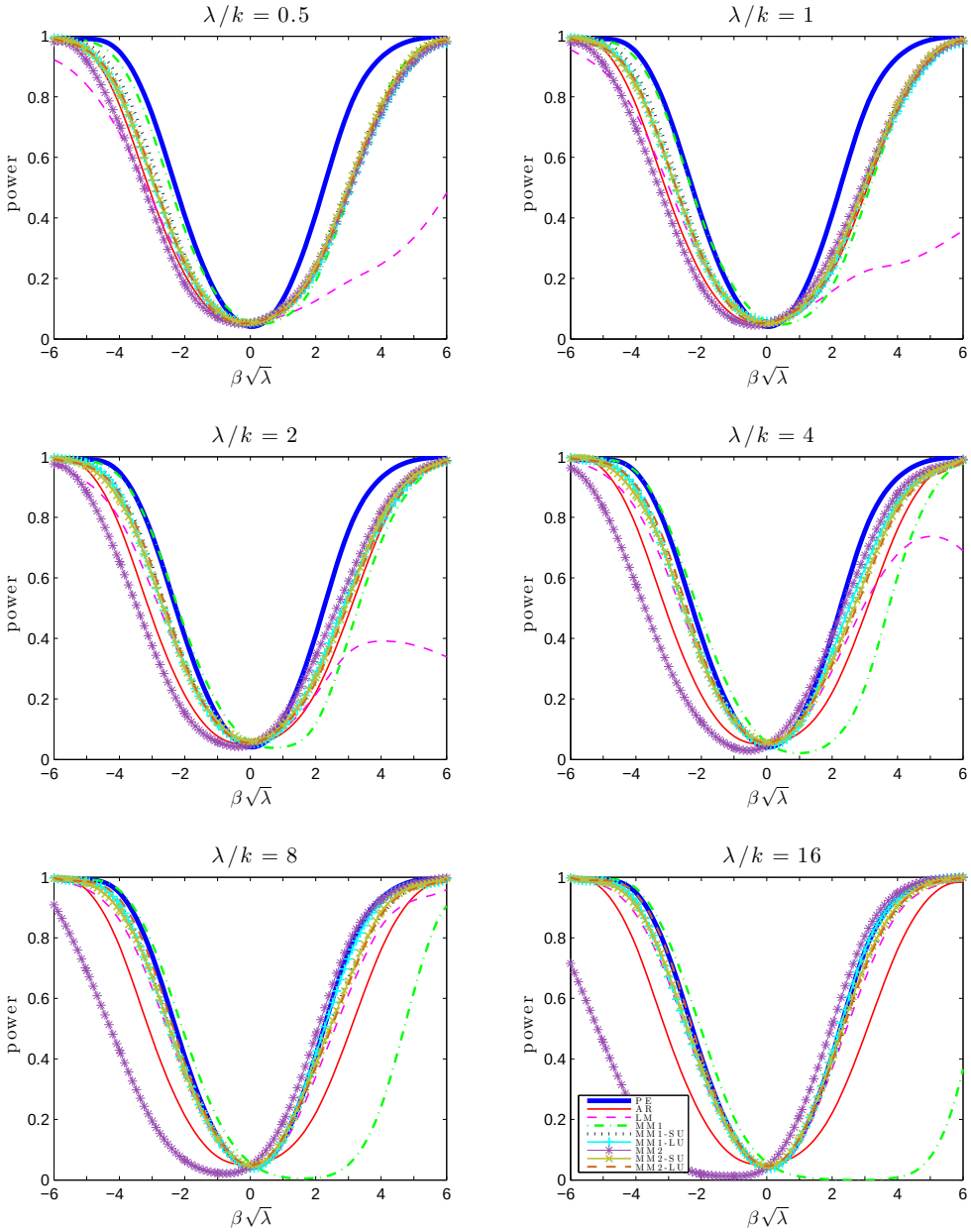


Figure 22: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = 0.2$

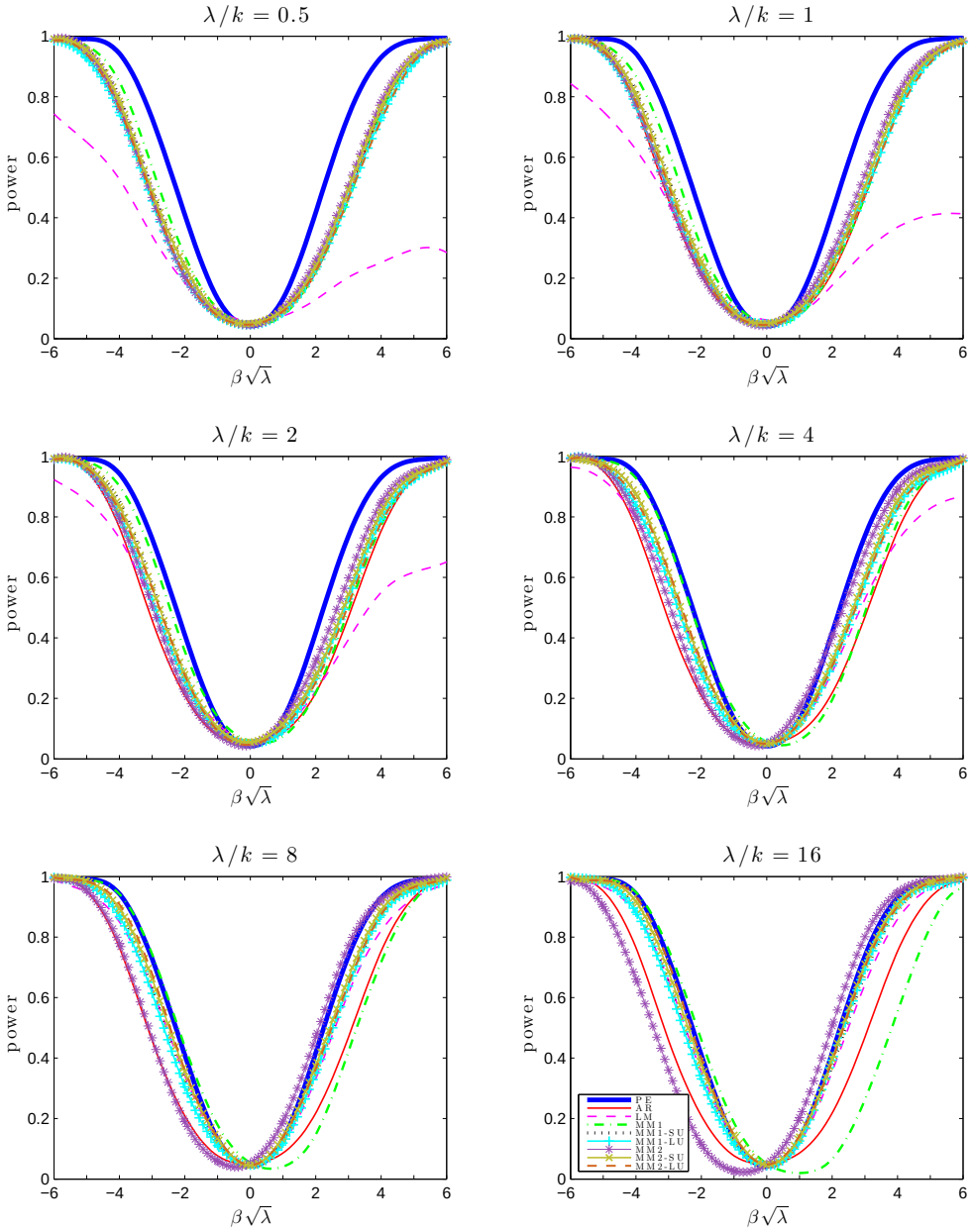


Figure 23: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = 0.5$

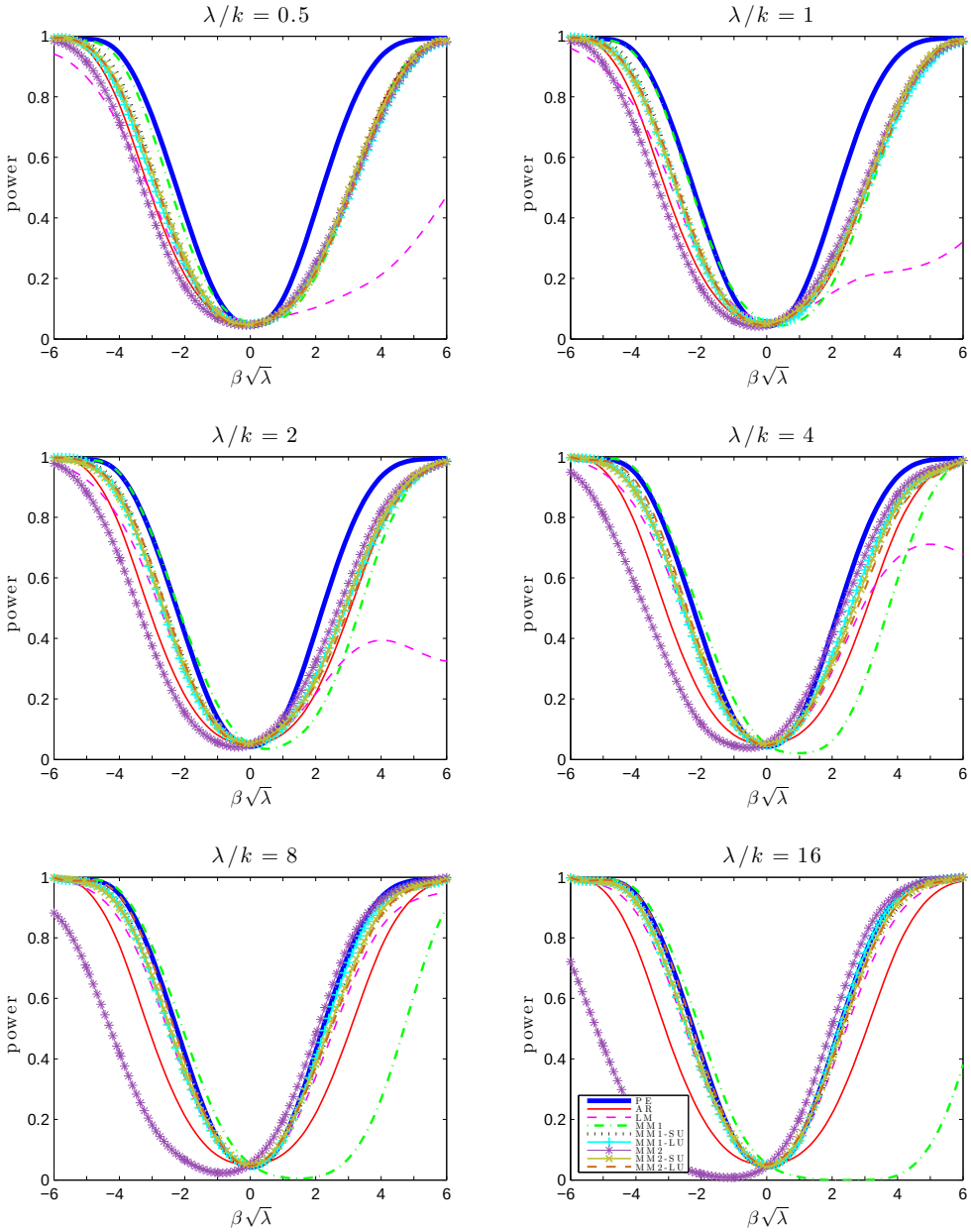


Figure 24: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = 0.9$

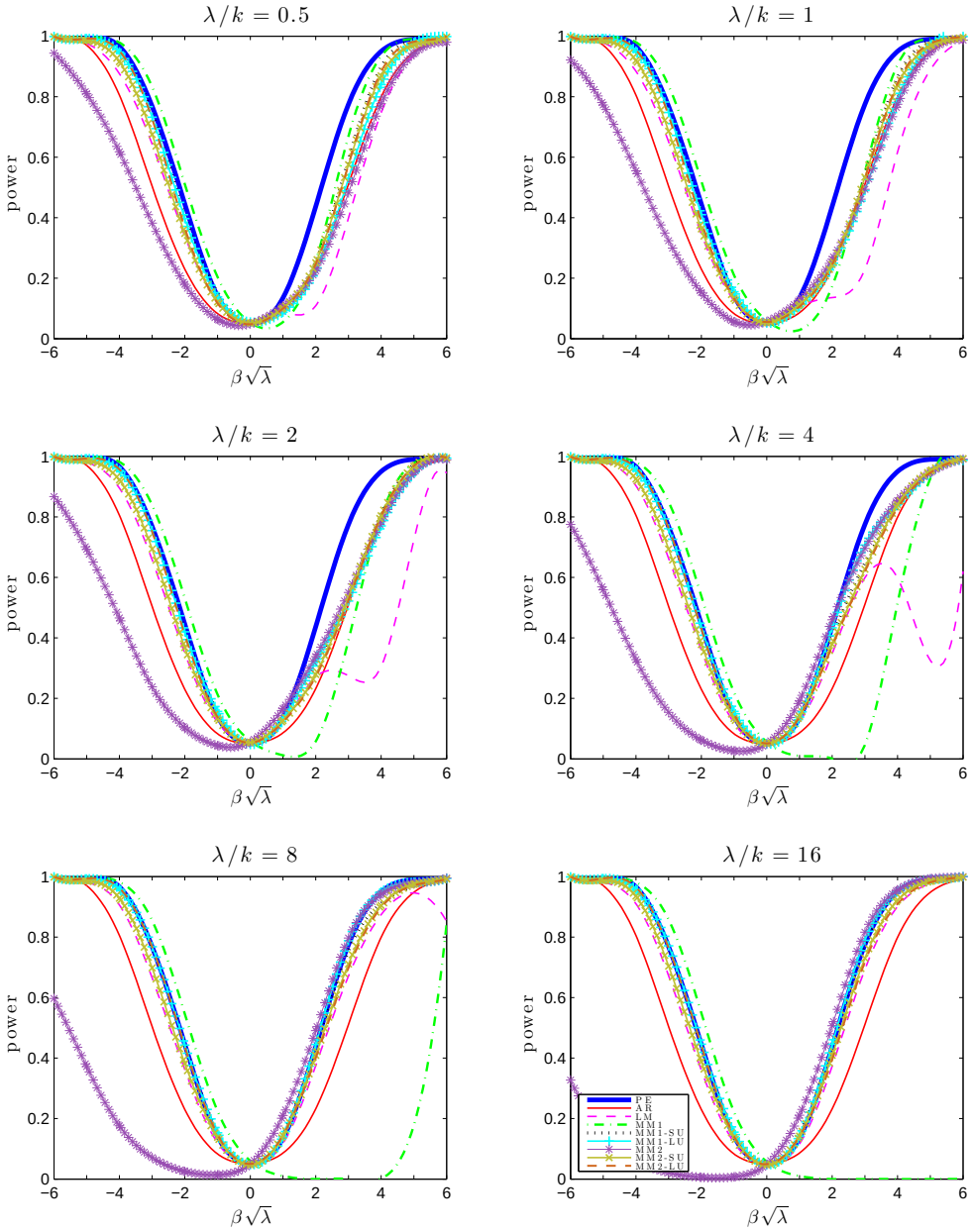


Figure 25: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = -0.5$

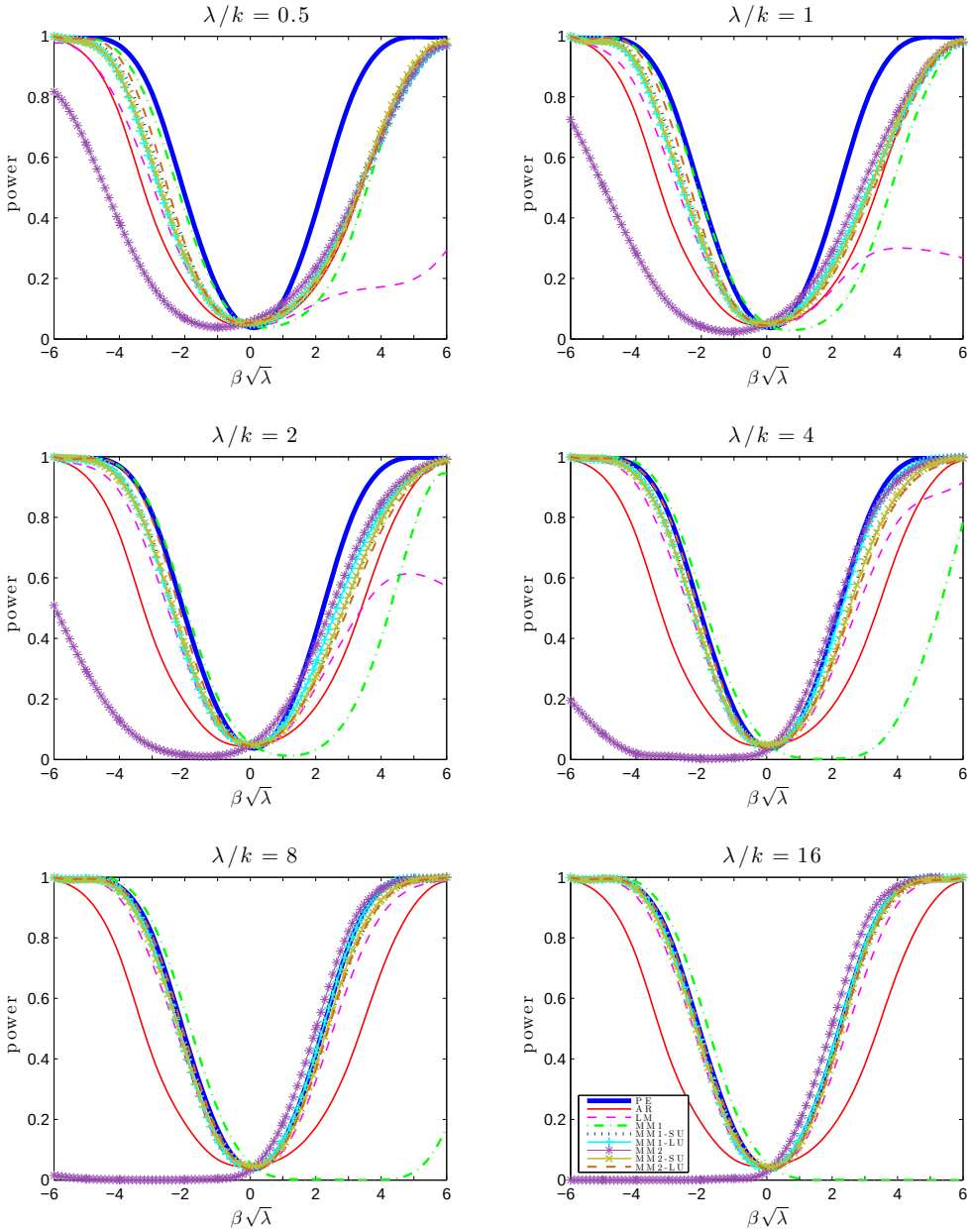


Figure 26: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = 0.2$

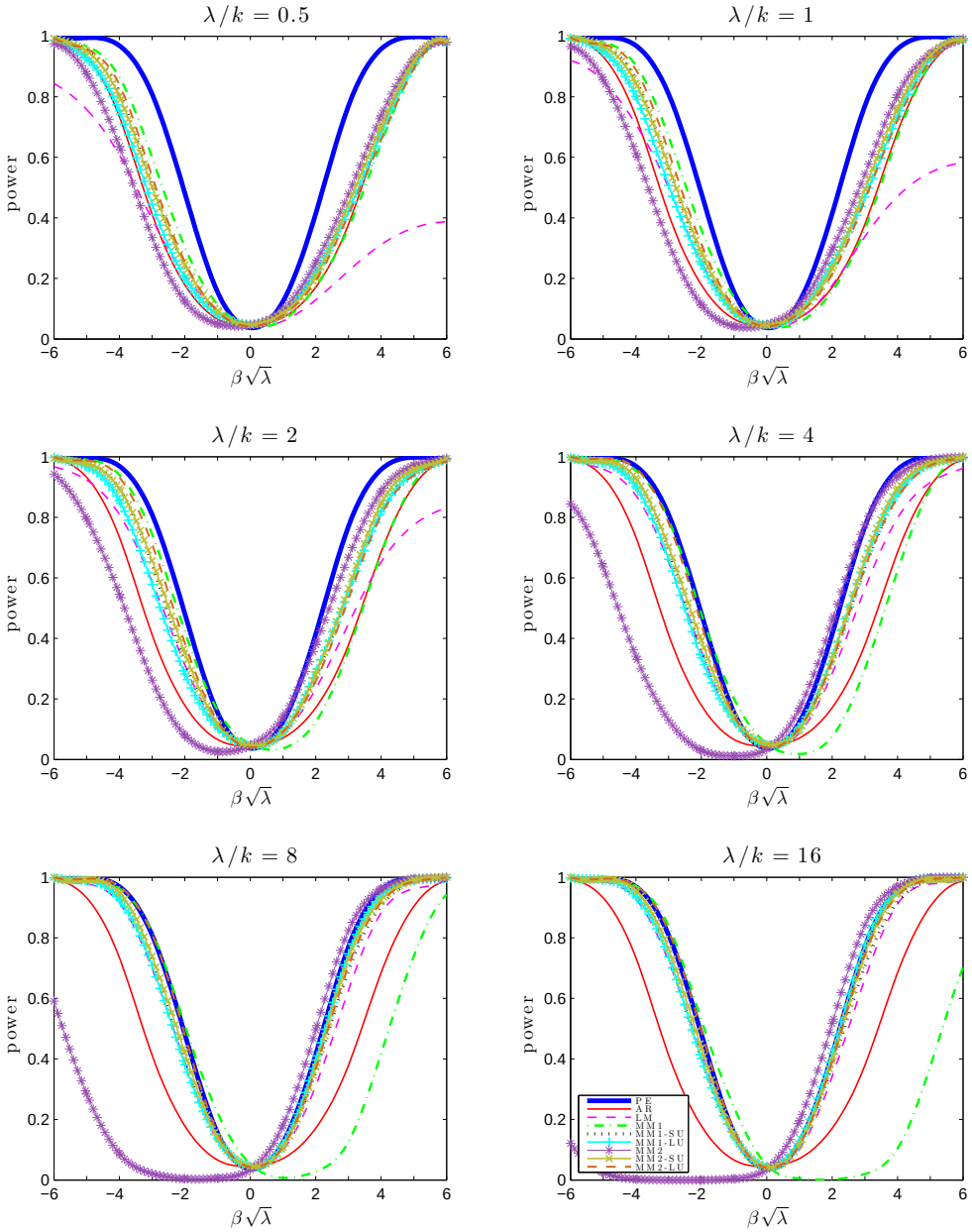


Figure 27: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = 0.5$

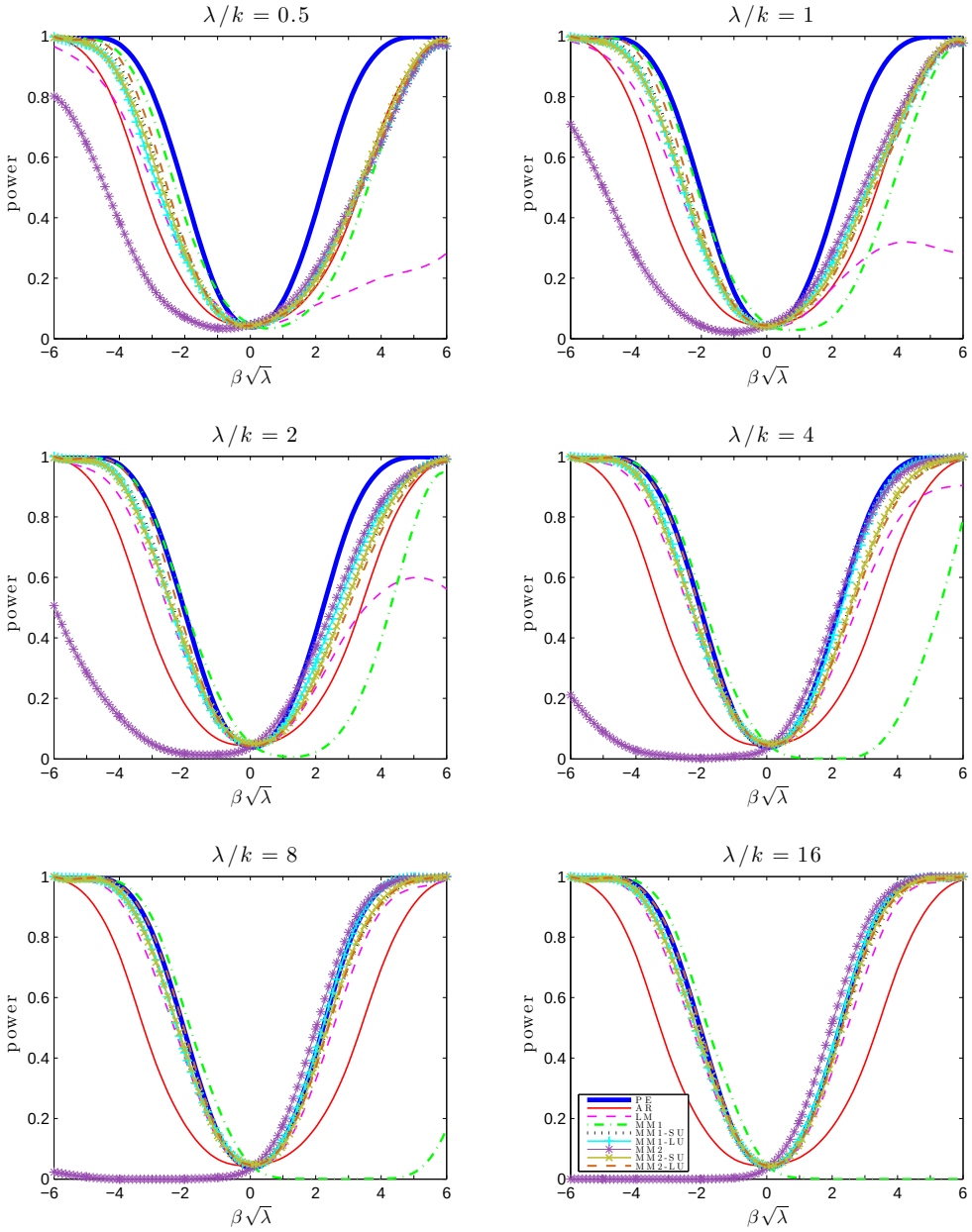


Figure 28: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = 0.9$

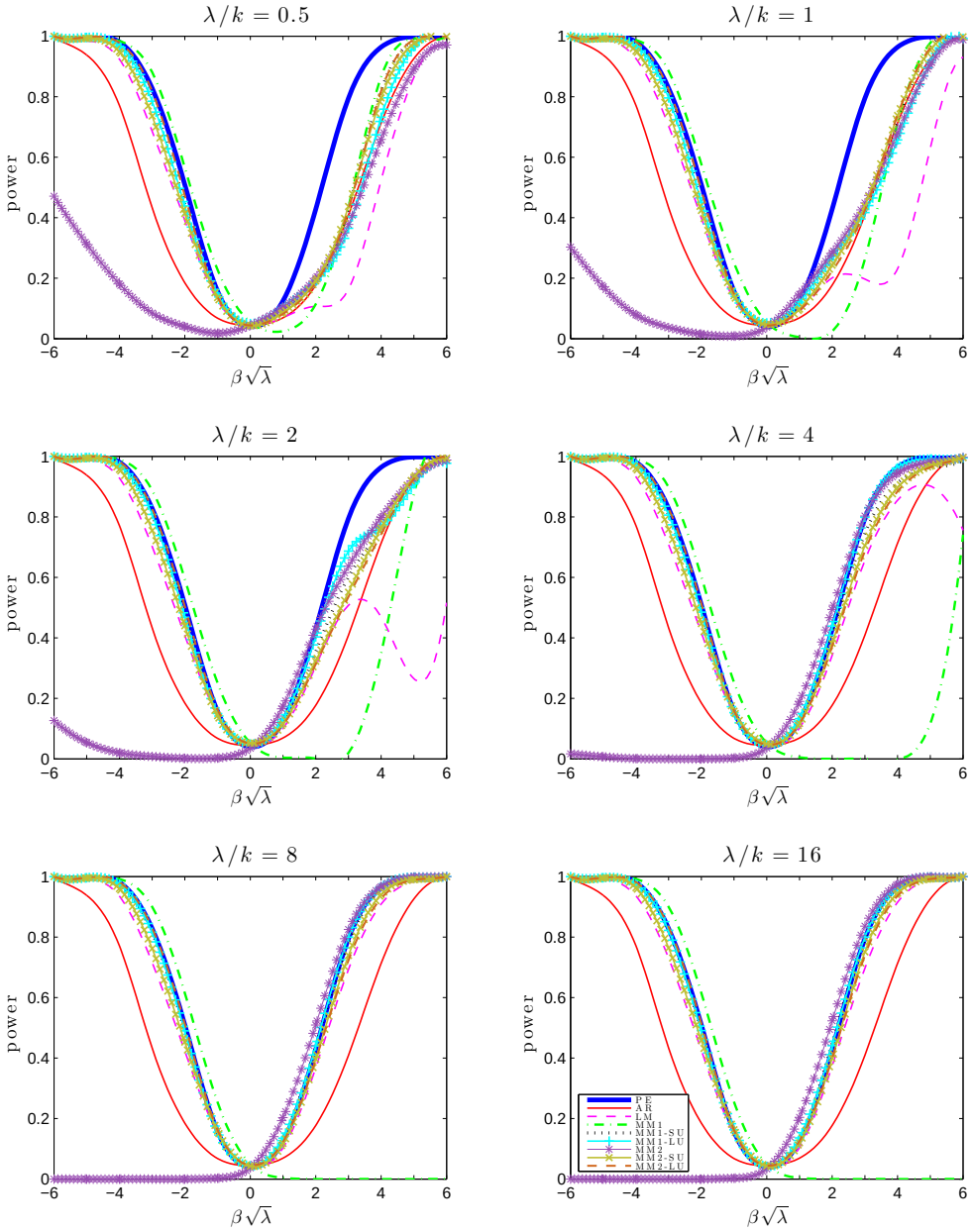


Figure 29: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = -0.5$

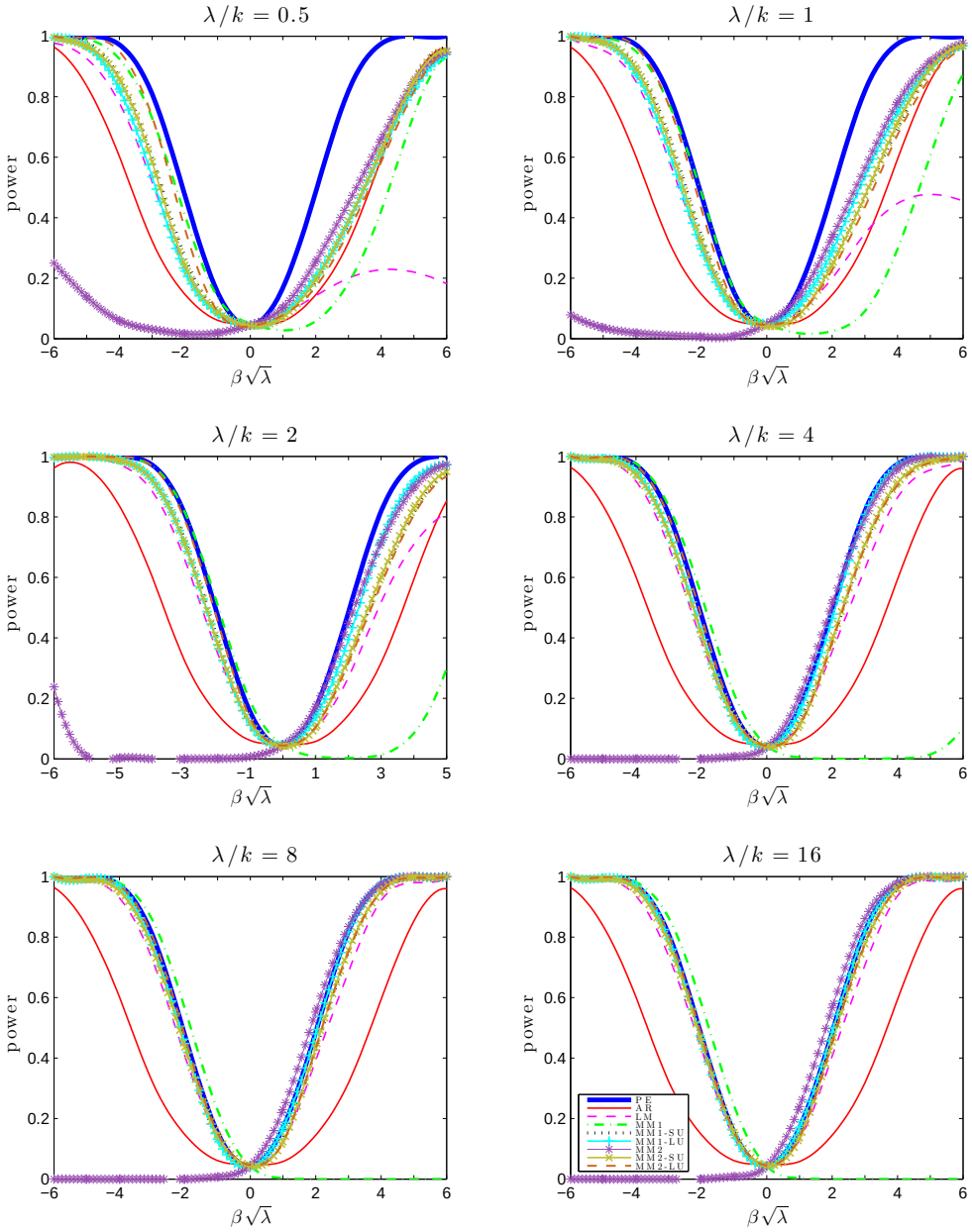


Figure 30: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = 0.2$

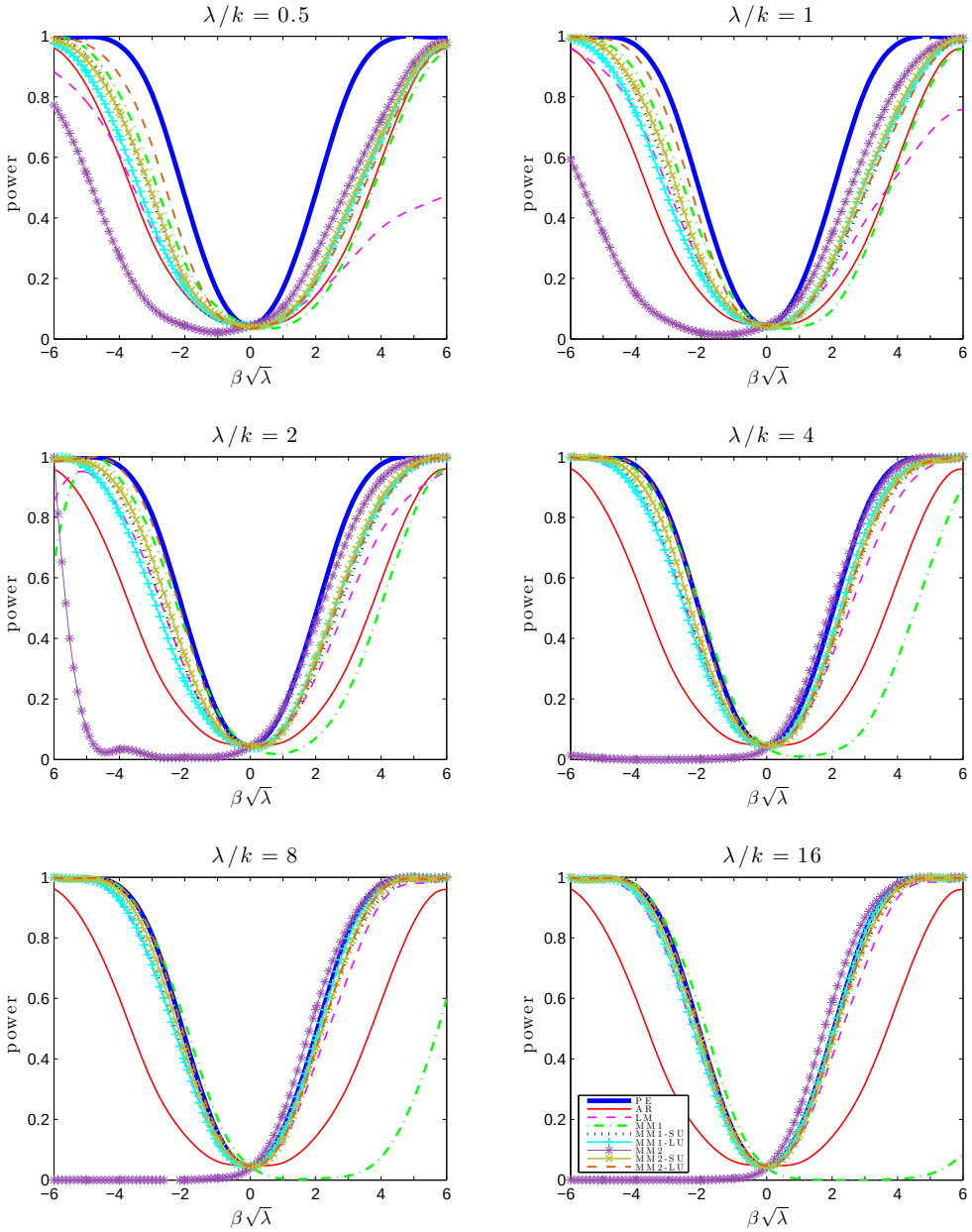


Figure 31: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = 0.5$

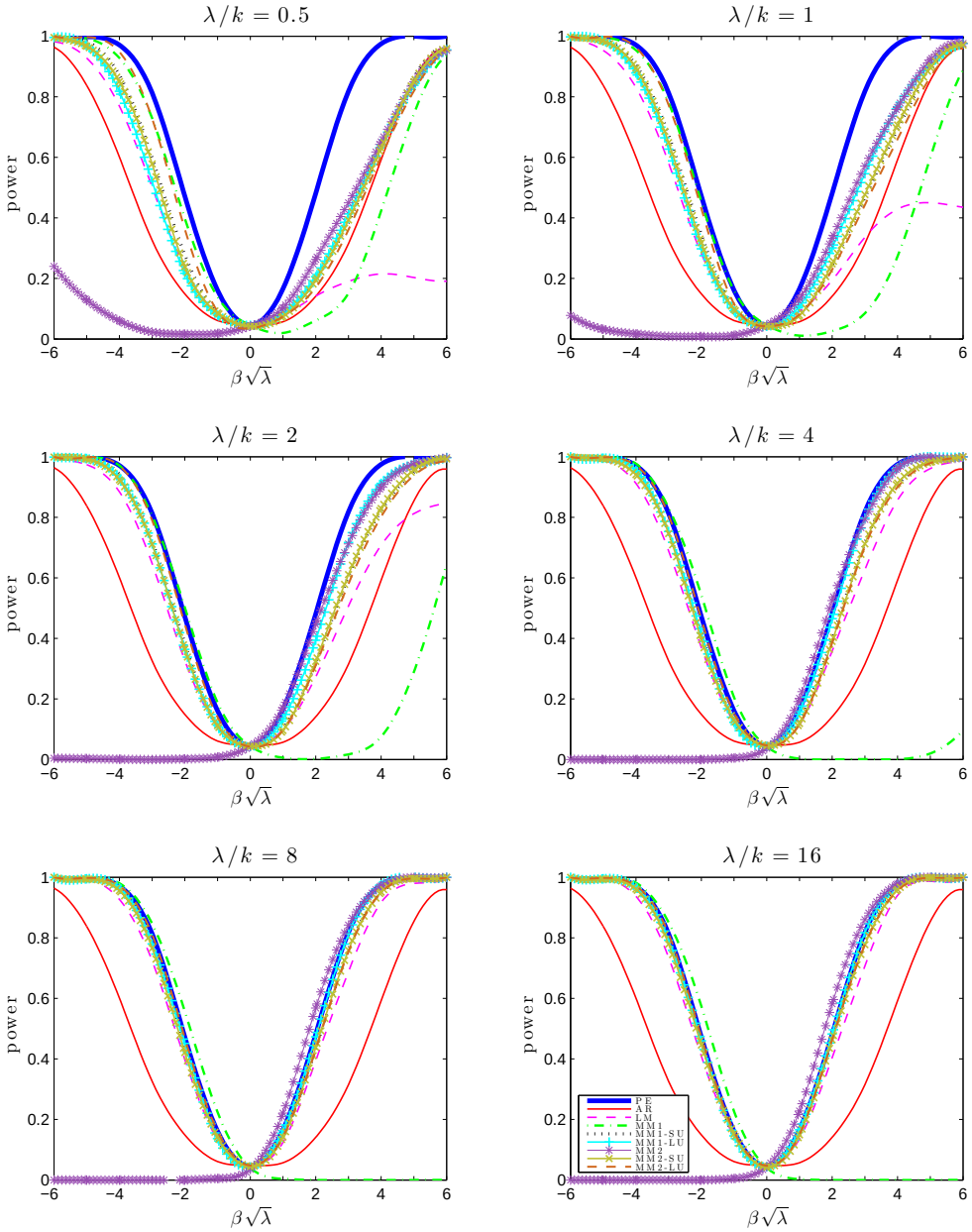
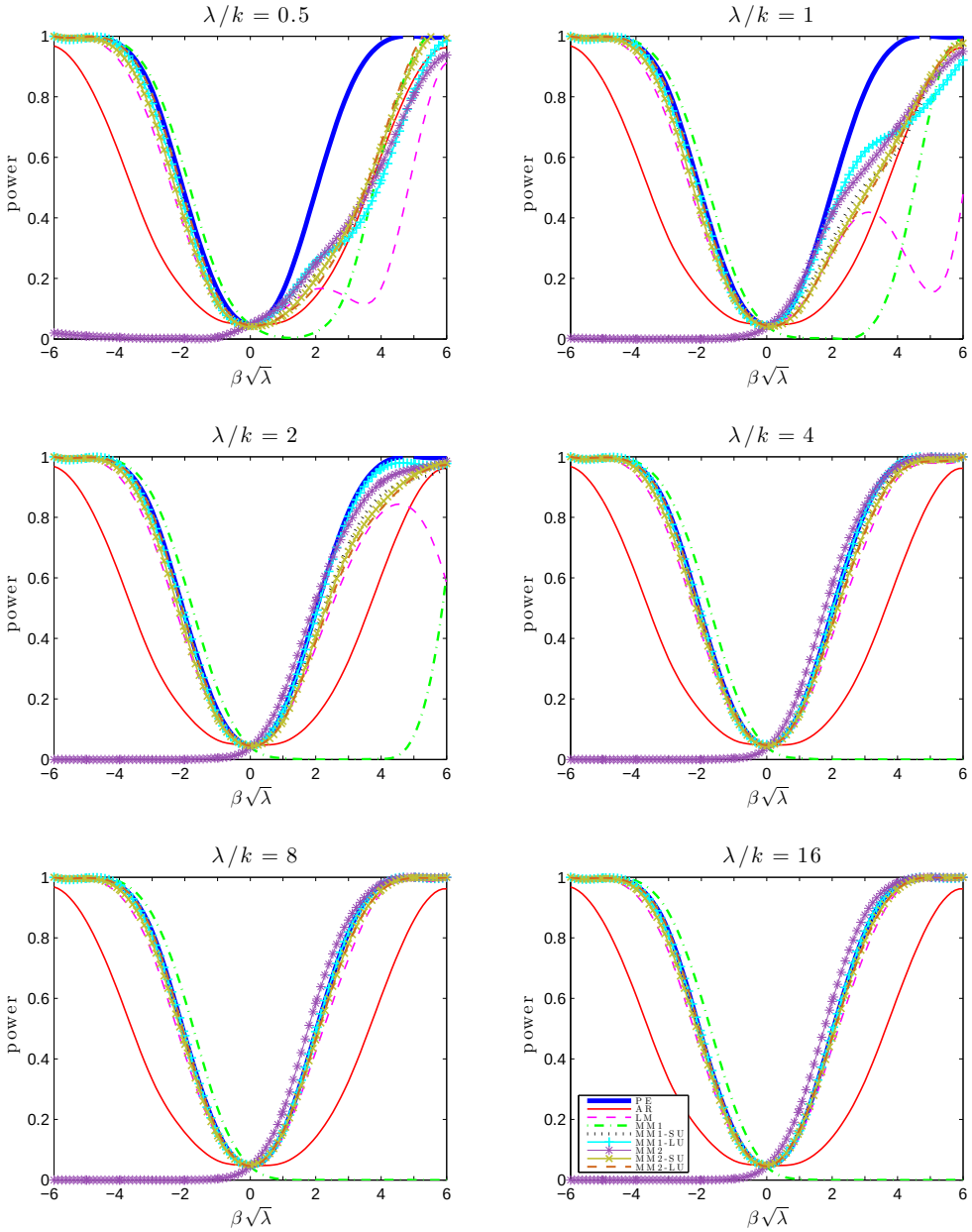


Figure 32: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = 0.9$



6 Appendix B: Nearly Integrated Regressor

To evaluate rejection probabilities, we perform 1,000 Monte Carlo simulations following the design of Jansson and Moreira (2006). The disturbances ε_t^y and ε_t^x are serially iid, with variance one and correlation $\rho = \omega_{12}/\omega_{11}^{1/2}\omega_{22}^{1/2}$. We use 1,000 replications to find the Lagrange multipliers using linear programming (LP).

The numerical simulations are done for $\rho = -0.5, 0.5$, $\gamma_N = 1 + c/N$ for $c = 0, -5, -10, -15, -25, -40$, and $\beta = b \cdot \sigma_{yy.x} g_T(\gamma_N)$ for $b = -6, -5, ..., 6$. The scaling function $g(\gamma_N) = \left(\sum_{i=1}^{N-1} \sum_{l=0}^{i-1} {\gamma_N}^{2l}\right)^{-1/2}$ allows us to look at the relevant power plots as γ_N changes. The value $b = 0$ corresponds to the null hypothesis $H_0 : \beta = 0$.

We report power plots for the power envelope (thick solid dark blue line) and the following tests: L_2 (thin solid red line), WAP size-corrected (light brown dashed line), WAP similar (black dotted line), and WAP-LU (thick purple line with rectangles).

Summary of findings.

1. As expected, the power curves for $\rho = -0.5$ are mirror images to the power plots for $\rho = 0.5$.

2. The L_2 test has correct size but not great power. When $c = 0$, this test behaves like a two-sided test. As c decreases, this test starts resembling a one-sided test. In particular, this test has power close to zero for some alternatives far from the null.

3. The WAP size-corrected test is biased when the regressor is integrated ($c = 0$) or nearly integrated ($c = -5, -10$). As c decreases and the regressor becomes stationary, the bias goes away.

4. The WAP similar test presents similar to the WAP size-corrected test (with slightly smaller bias).

5. The WAP-LU test decreases the bias of the two other WAP tests considerably (even though we evaluate the boundary conditions at only 15 points).

6. When c is small, the power of the WAP-LU test based on the MM-2S statistic is very close to the power envelope for b negative and $\rho = 0.5$ (or b positive and $\rho = -0.5$). However, the power curve of the WAP-LU test is smaller than the power envelope for b positive and $\rho = 0.5$ (or b negative and

$\rho = -0.5$). This suggests there may be some power gains using a weighted average density different from the MM-2S statistic.

7. The WAP-LU has overall better power than the other WAP tests, and numerically dominates the L_2 test. We recommend the use of the WAP-LU test in empirical practice.

Figure 33: Power Comparison: $\rho = -0.5$

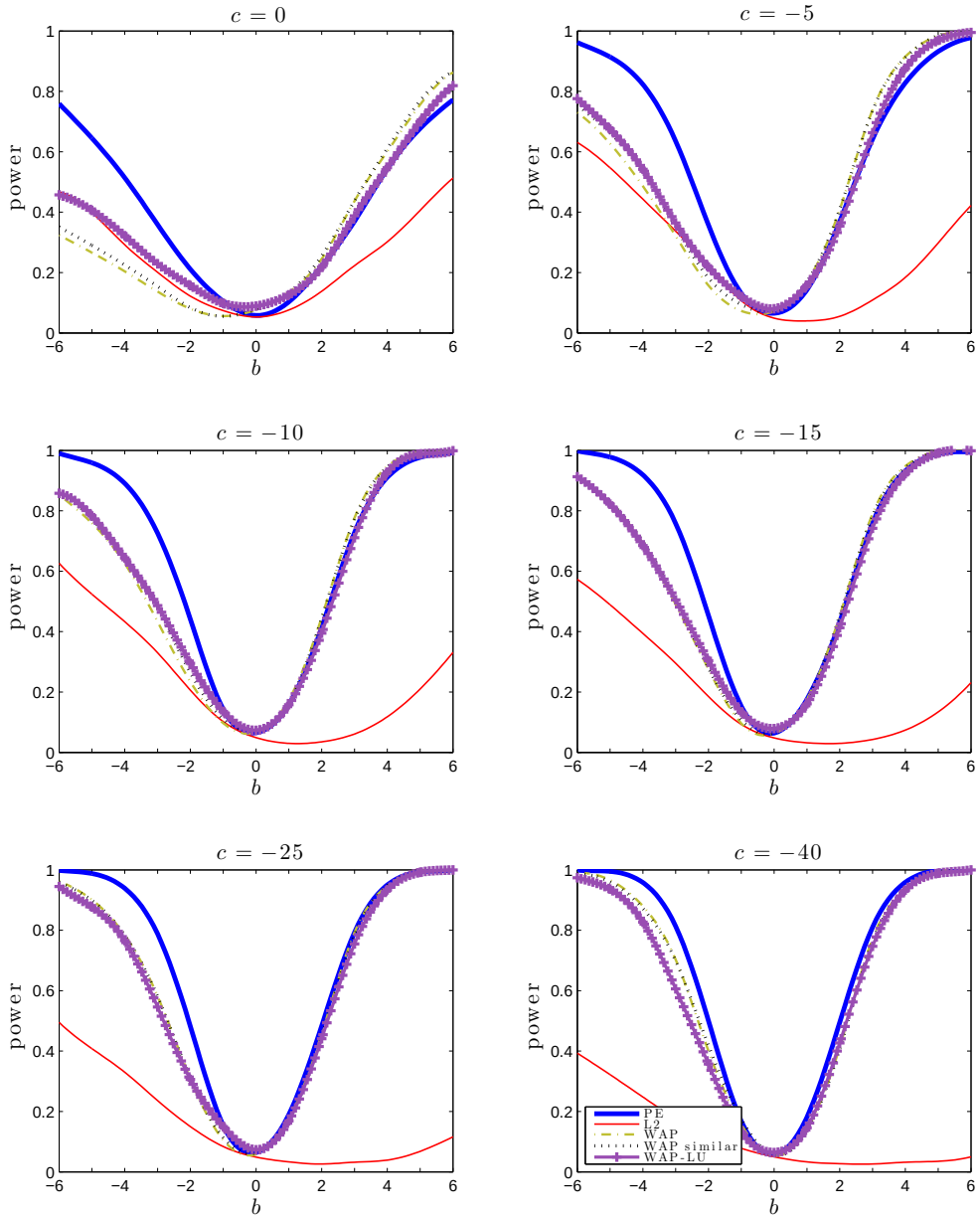
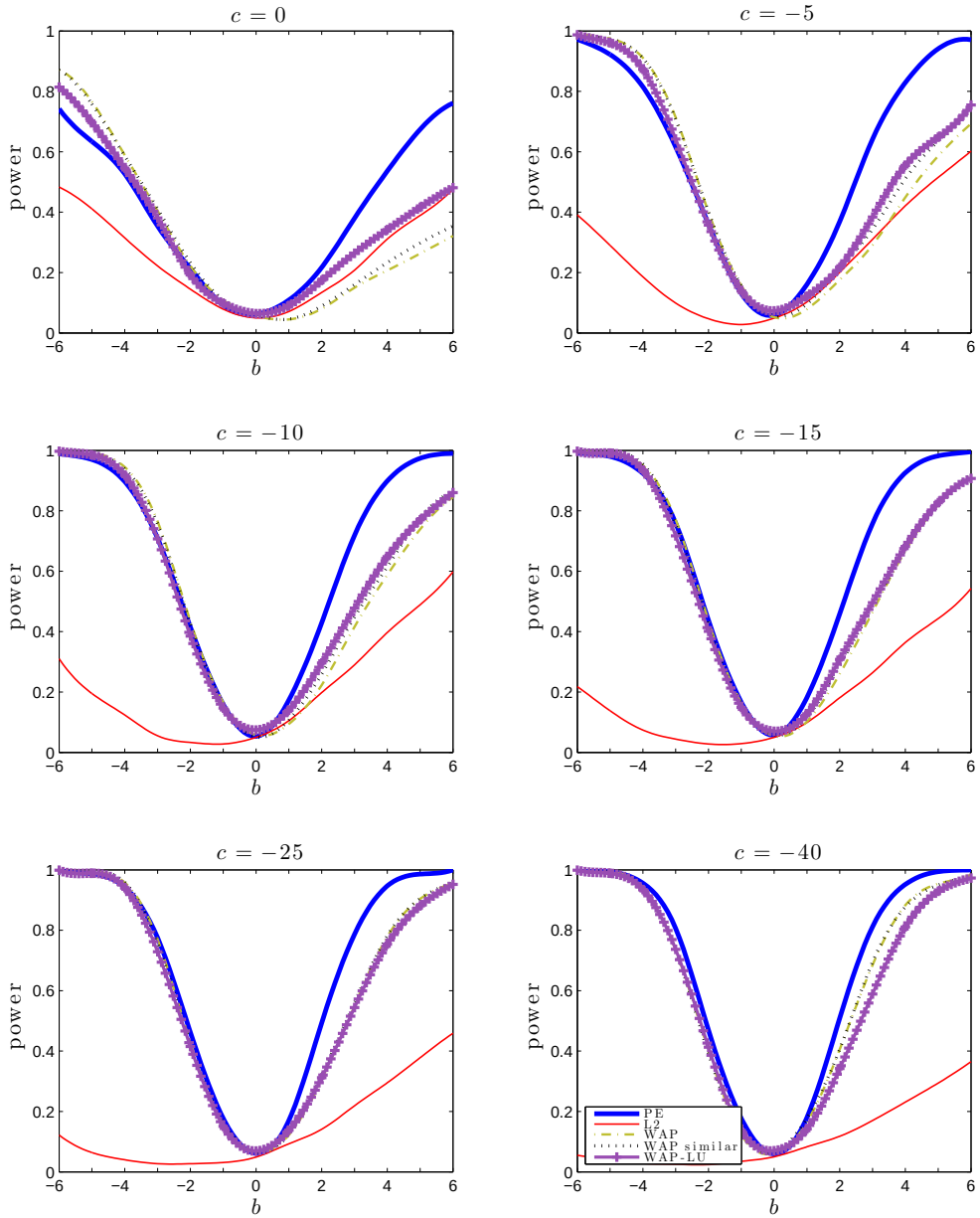


Figure 34: Power Comparison: $\rho = 0.5$



References

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752, Supplement.

Berge, C. (1997): *Topological Spaces*. Dover Books on Mathematics.

Campbell, J. Y., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60.

Dunford, N., and J. T. Schwartz (1988): *Linear Operators Part I: General Theory*. Wiley Classics Library. John Wiley and Sons.

Jansson, M., and M. J. Moreira (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74, 681–715.

Linnik, J. V. (2000): *Statistical Problems with Nuisance Parameters*. Translations of Mathematical Monographs, vol. 20.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*. John Wiley and Sons, Inc.

Moreira, H., and M. J. Moreira (2011): "Inference with Persistent Regressors," *Unpublished Manuscript*, Columbia University.

——— (2013): "Contributions to the Theory of Optimal Tests," *Unpublished Manuscript*, FGV/EPGE.

Moreira, M. J. (2009): "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," *Journal of Econometrics*, 152, 131–140.

Rudin, W. (1991): *Functional Analysis*. Second edn., McGraw-Hill.

Van Loan, C., and N. Ptsianis (1993): "Approximation with Kronecker Products," in *Linear Algebra for Large Scale and Real-Time Applications*, ed. by M. Moonen, and G. Golub, pp. 293–314. Springer.

Wijsman, R. A. (1958): "Incomplete Sufficient Statistics and Similar Tests," *Annals of Mathematical Statistics*, 29, 1028–1045.

Wright, J. H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business and Economic Statistics*, 18, 211–222.