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

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

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,

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

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

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

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

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

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

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

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:

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

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 $q\left(t\right)$ 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

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

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 $\phi\left(S,T\right)$ 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 $F_\phi(z_1,z_2)\neq0$, 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

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

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 $c_0\left(t\right)$ and $c_1\left(t\right)$ using linear programming based on simulations for the statistic S. Consider the approximated problem

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

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

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

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 $c_0\left(t\right)$ and $c\left(t\right)$. 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

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

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

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]

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

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

$$\begin{aligned} &\max_{x^{(j)} \in \{0,1\}} \frac{1}{J} \sum_{j=1}^{J} x^{(j)} \frac{h\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}{\overline{f}_{\beta_0}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}, \\ &\text{where } \frac{1}{J} \sum_{j=1}^{J} x^{(j)} \frac{f_{\beta_0,\pi_l}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}{\overline{f}_{\beta_0}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)} = \alpha \\ &\text{and } \frac{1}{J} \sum_{j=1}^{J} x^{(j)} \psi\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right) \frac{f_{\beta_0,\pi_l}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)}{\overline{f}_{\beta_0}\left(\widetilde{y}_1^{(j)}, y_2^{(j)}\right)} = 0, \end{aligned} \tag{3.9}$$

[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

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

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

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

The two-sided test rejects the null when

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

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

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

Remark 1 *Consider the Banach space $\mathcal{L}_1(Y, h)$, where h is a density, and let $\Gamma_1(g, \gamma) = \{\phi \in \mathcal{L}_1(Y, h); 0 \le \phi \le 1$ 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 $\underline{f}_v$ 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 $\beta = \beta_0$:

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

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. $\square$

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),

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

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

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

By taking sequences β_N approaching β_0, we show that

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

It also must be the case that

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

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

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

Algebraic manipulations show that (5.17) simplifies to

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

The test ϕ must be uncorrelated with the statistic T:

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

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

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

as we wanted to prove. □

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. $\square$

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. $\square$

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. $\square$

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

$$\Omega = \left[\begin{array}{cc} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{array} \right] P \left[\begin{array}{cc} 1+\rho & 0 \\ 0 & 1-\rho \end{array} \right] P' \left[\begin{array}{cc} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{array} \right],$$

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:

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

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) $k = 2, \rho = -0.5$

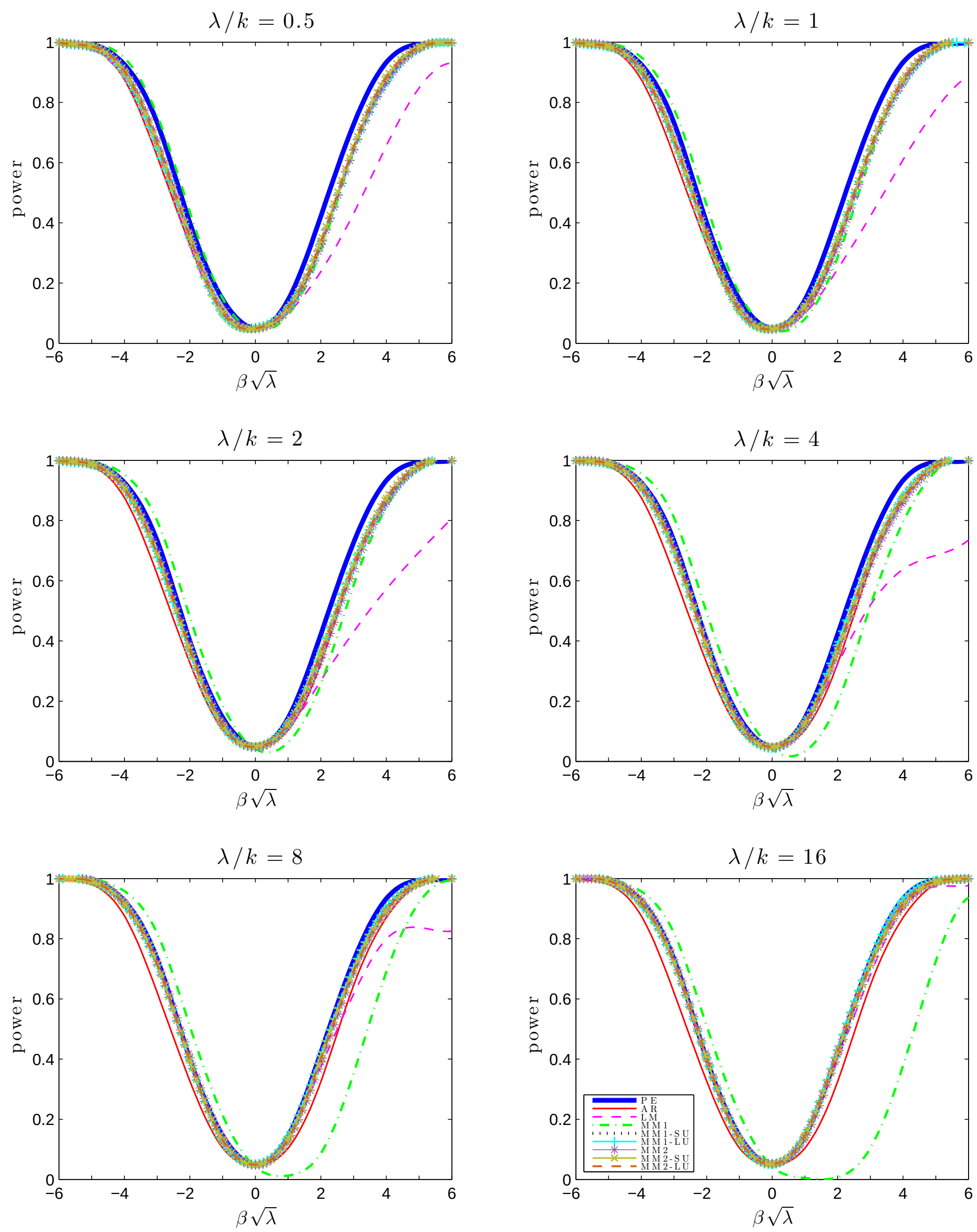


Figure 2: Power Comparison (Kronecker Covariance) $k = 2, \rho = 0.2$

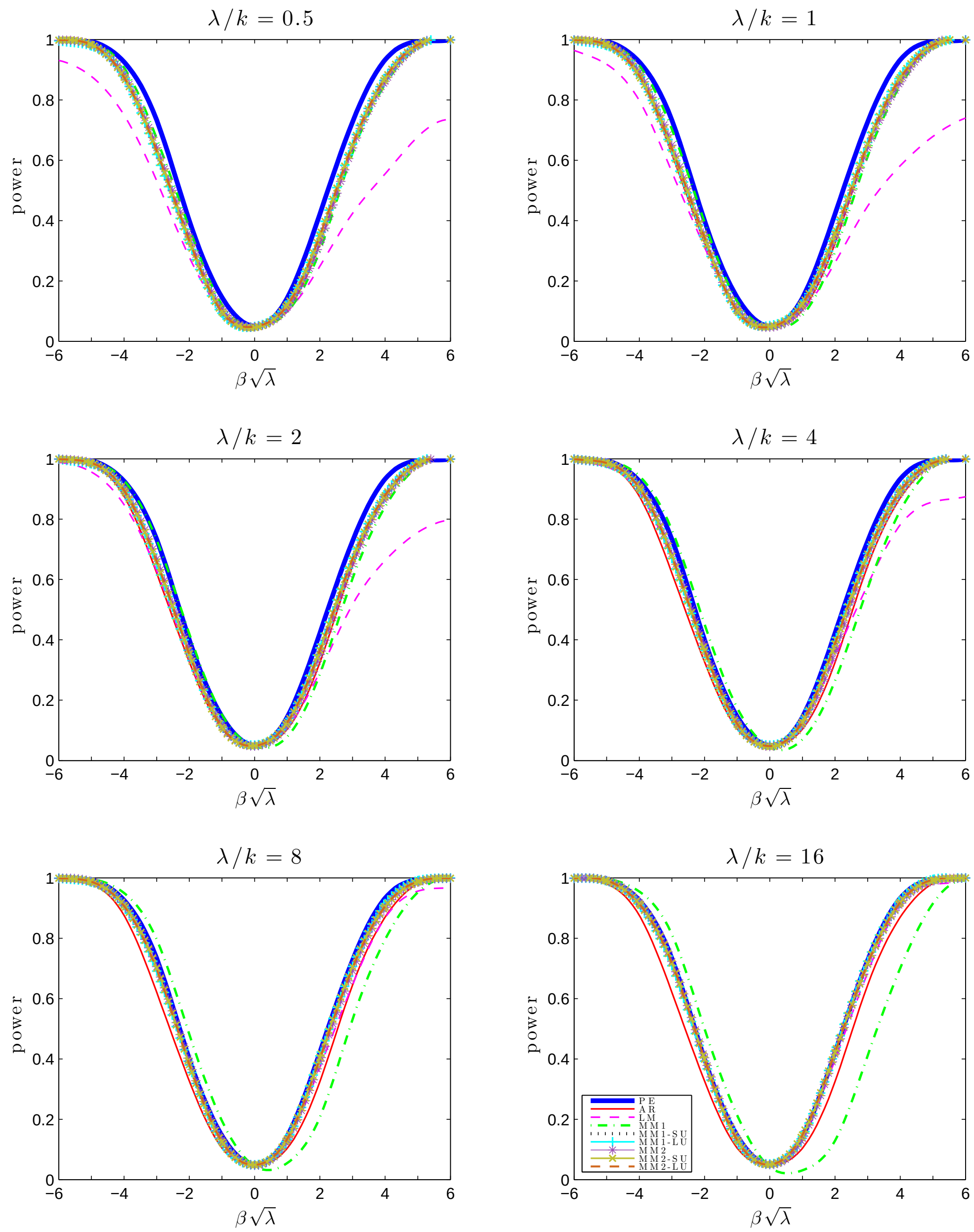


Figure 3: Power Comparison (Kronecker Covariance) $k = 2, \rho = 0.5$

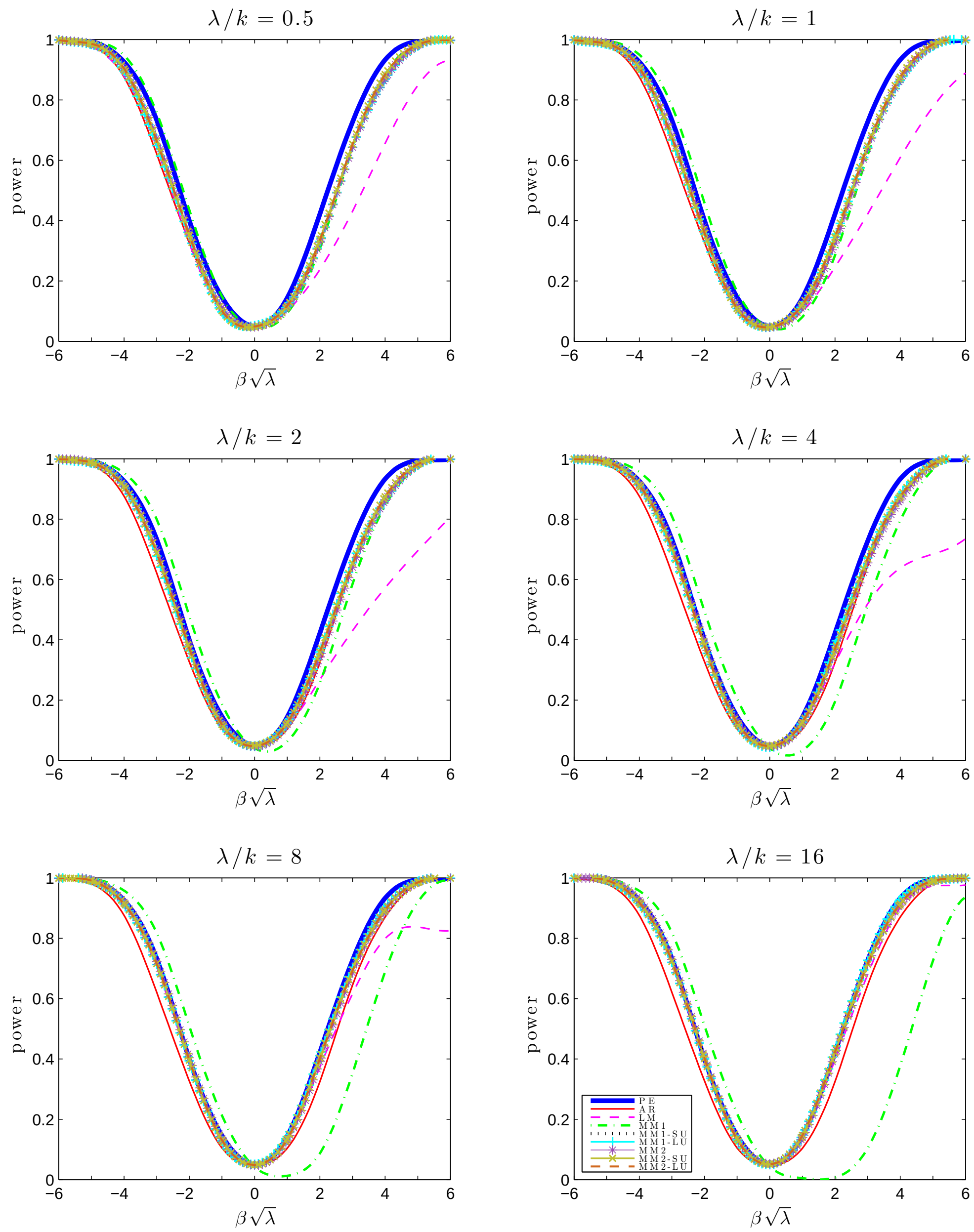


Figure 4: Power Comparison (Kronecker Covariance) $k = 2, \rho = 0.9$

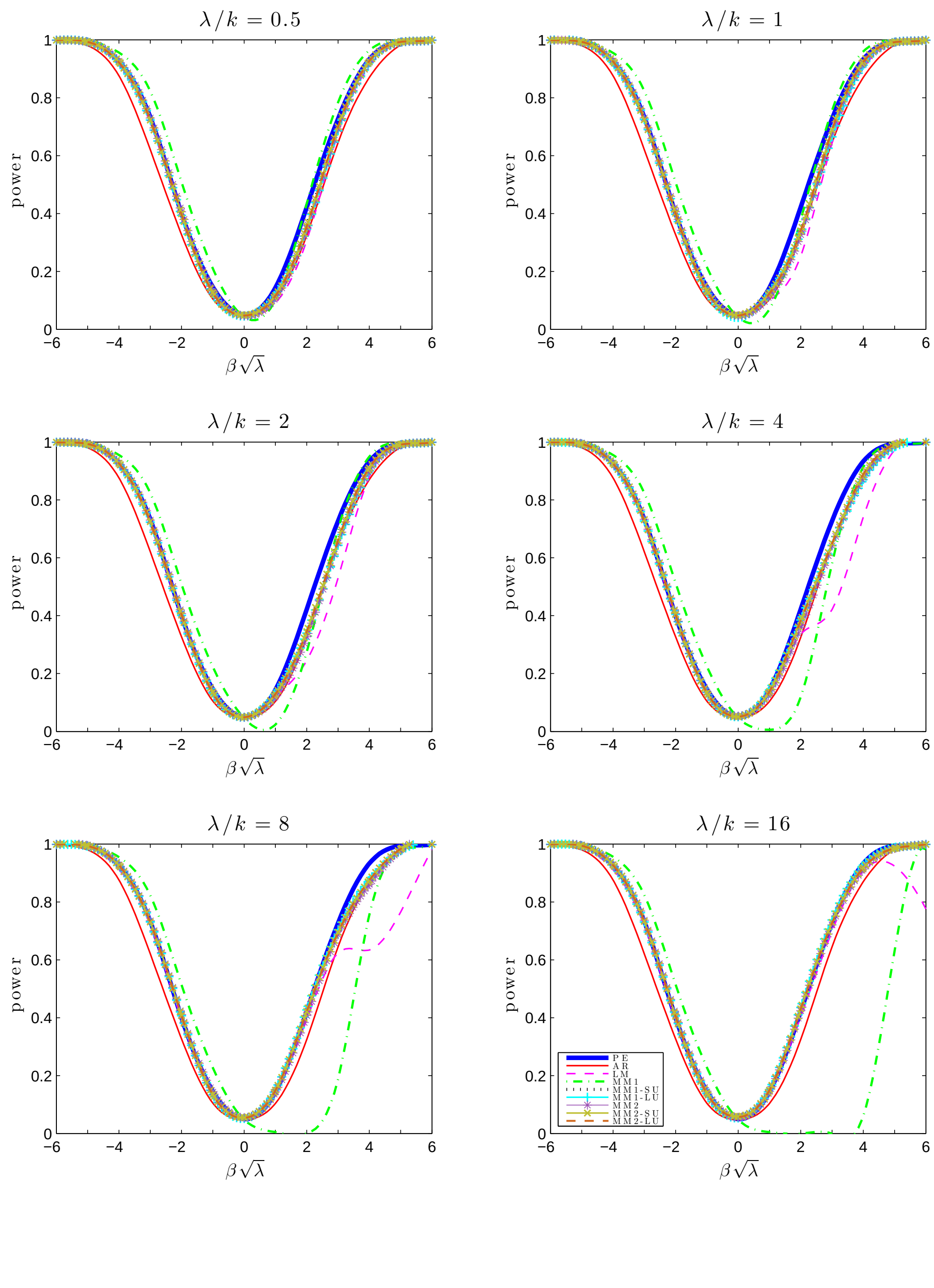


Figure 5: Power Comparison (Kronecker Covariance) $k = 5, \rho = -0.5$

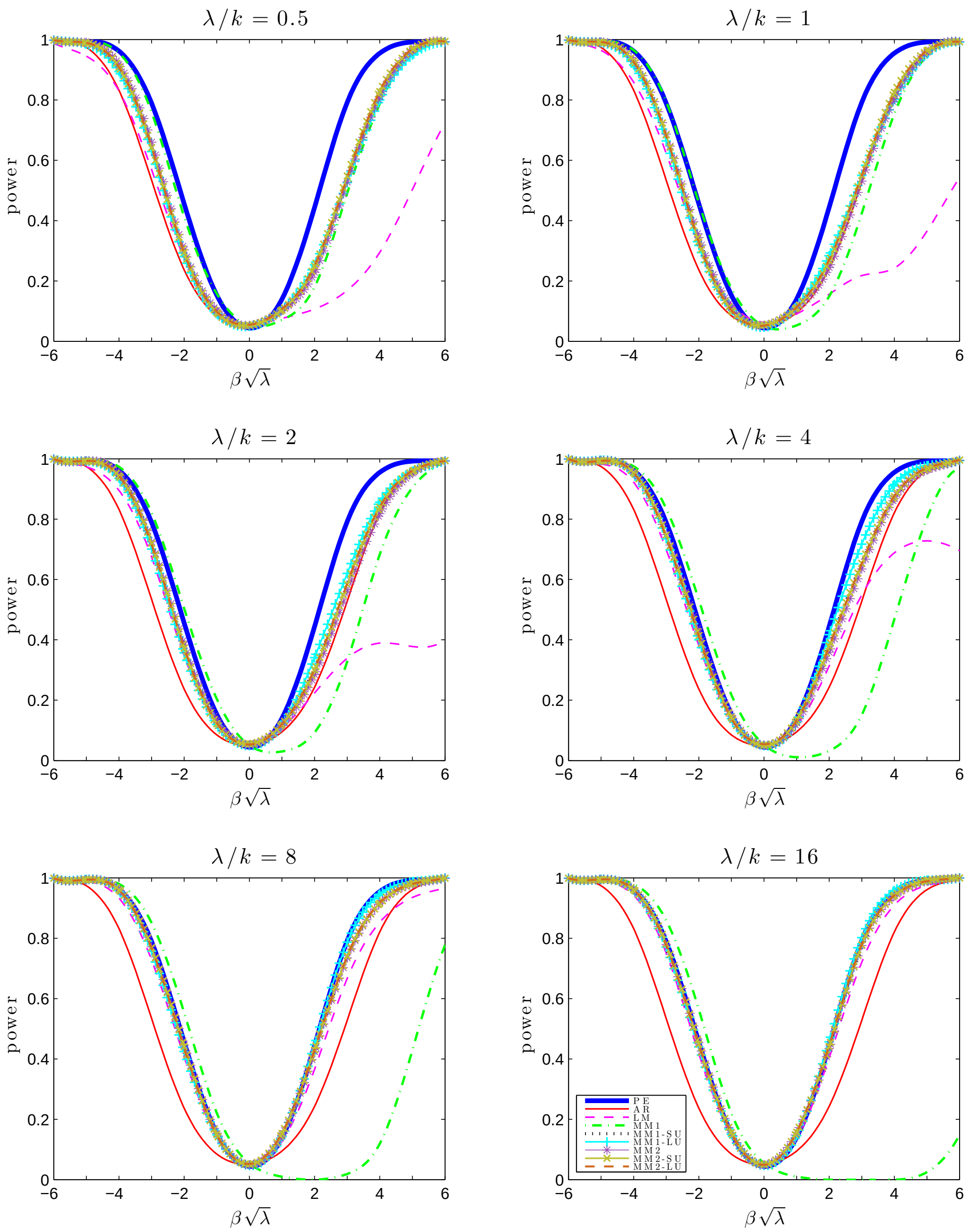


Figure 6: Power Comparison (Kronecker Covariance) $k = 5, \rho = 0.2$

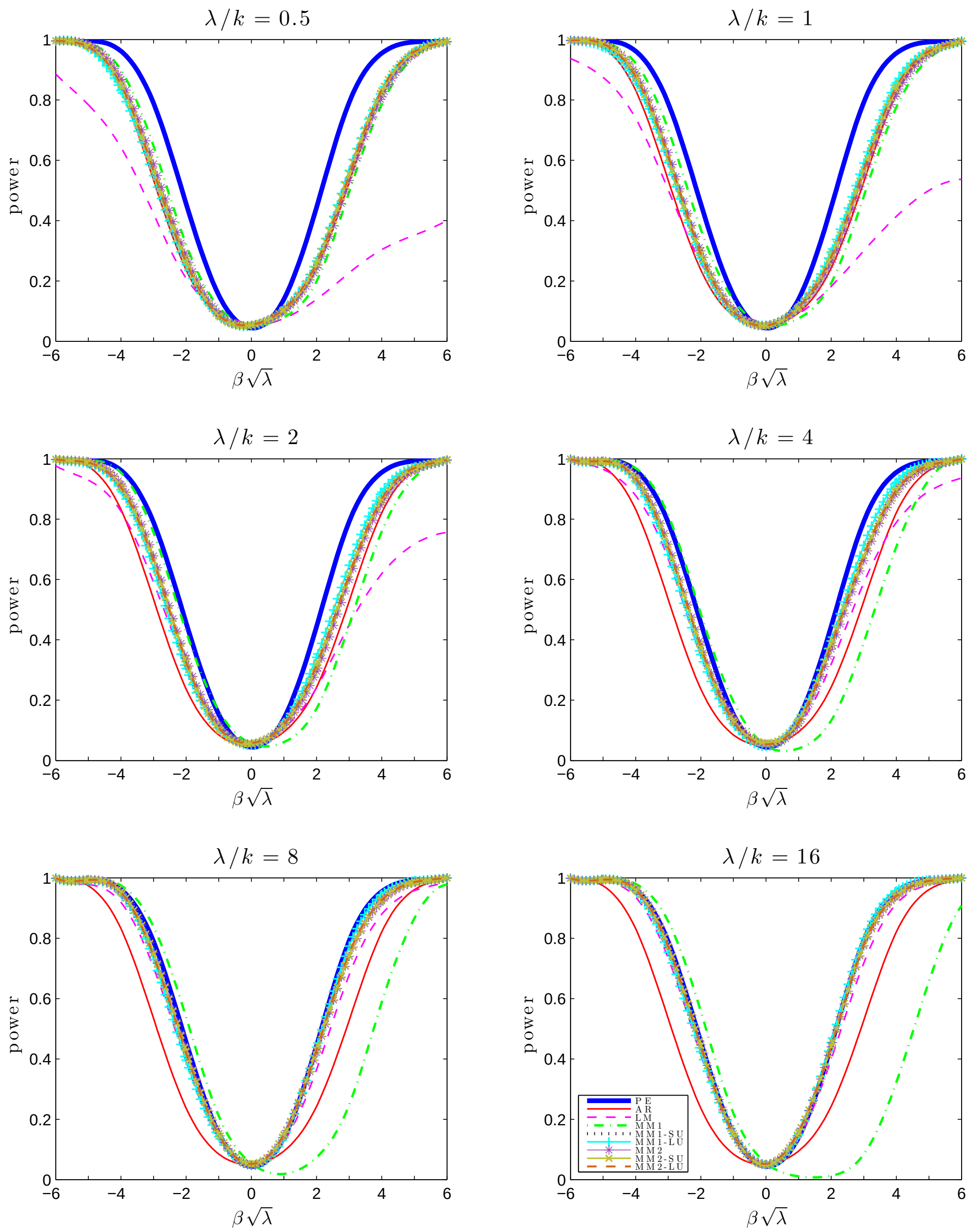


Figure 7: Power Comparison (Kronecker Covariance) $k = 5, \rho = 0.5$

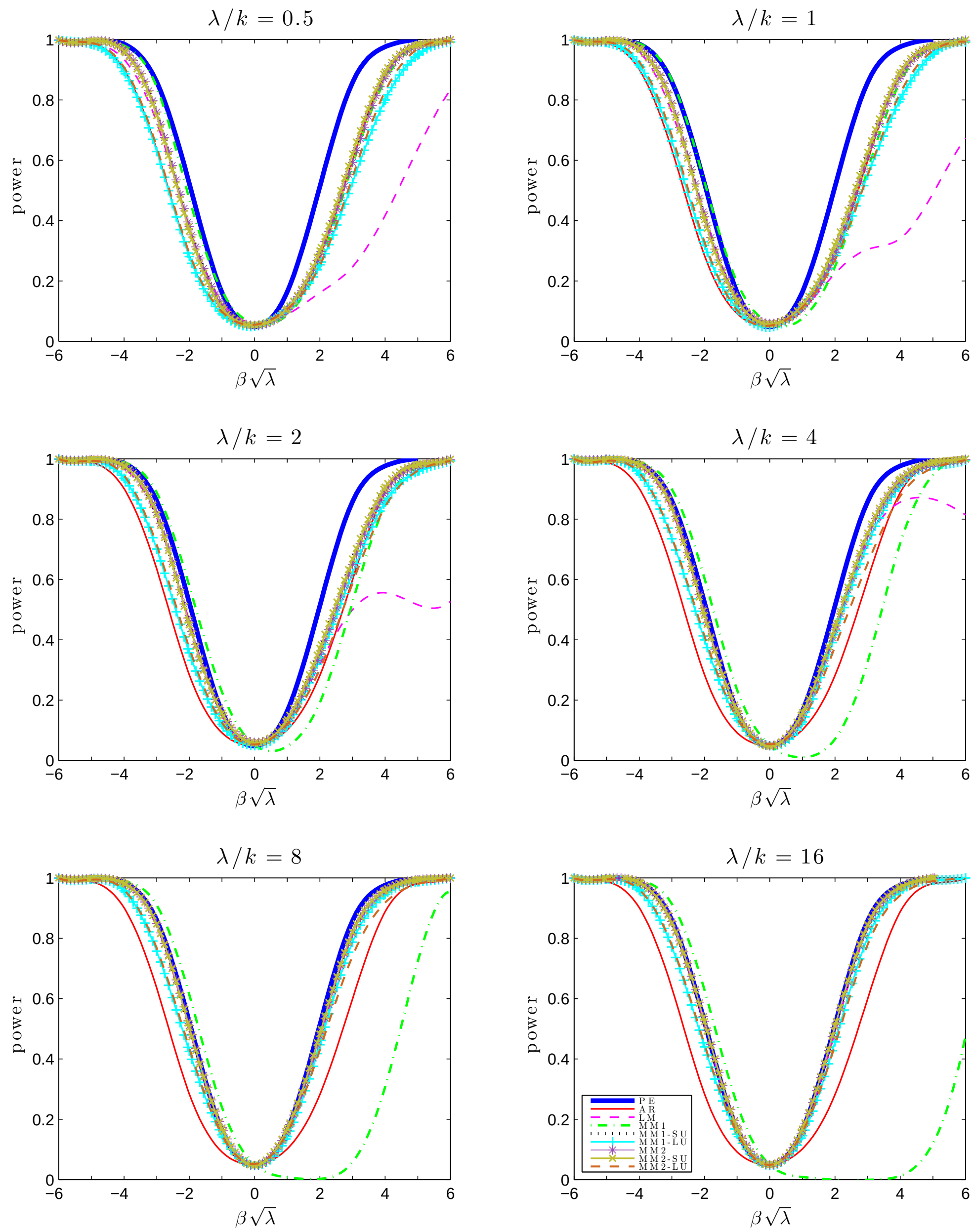


Figure 8: Power Comparison (Kronecker Covariance) $k = 5, \rho = 0.9$

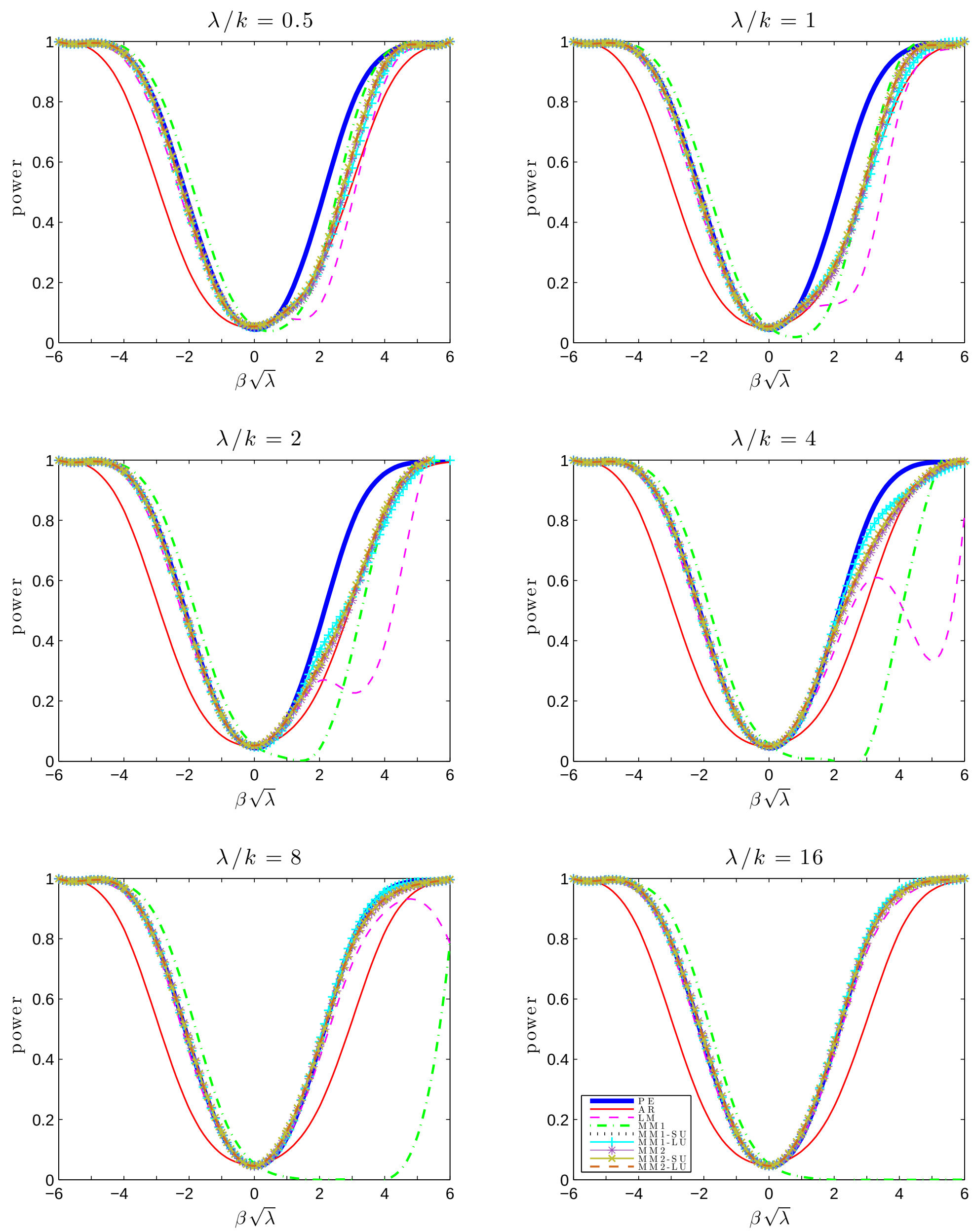


Figure 9: Power Comparison (Kronecker Covariance) $k = 10, \rho = -0.5$

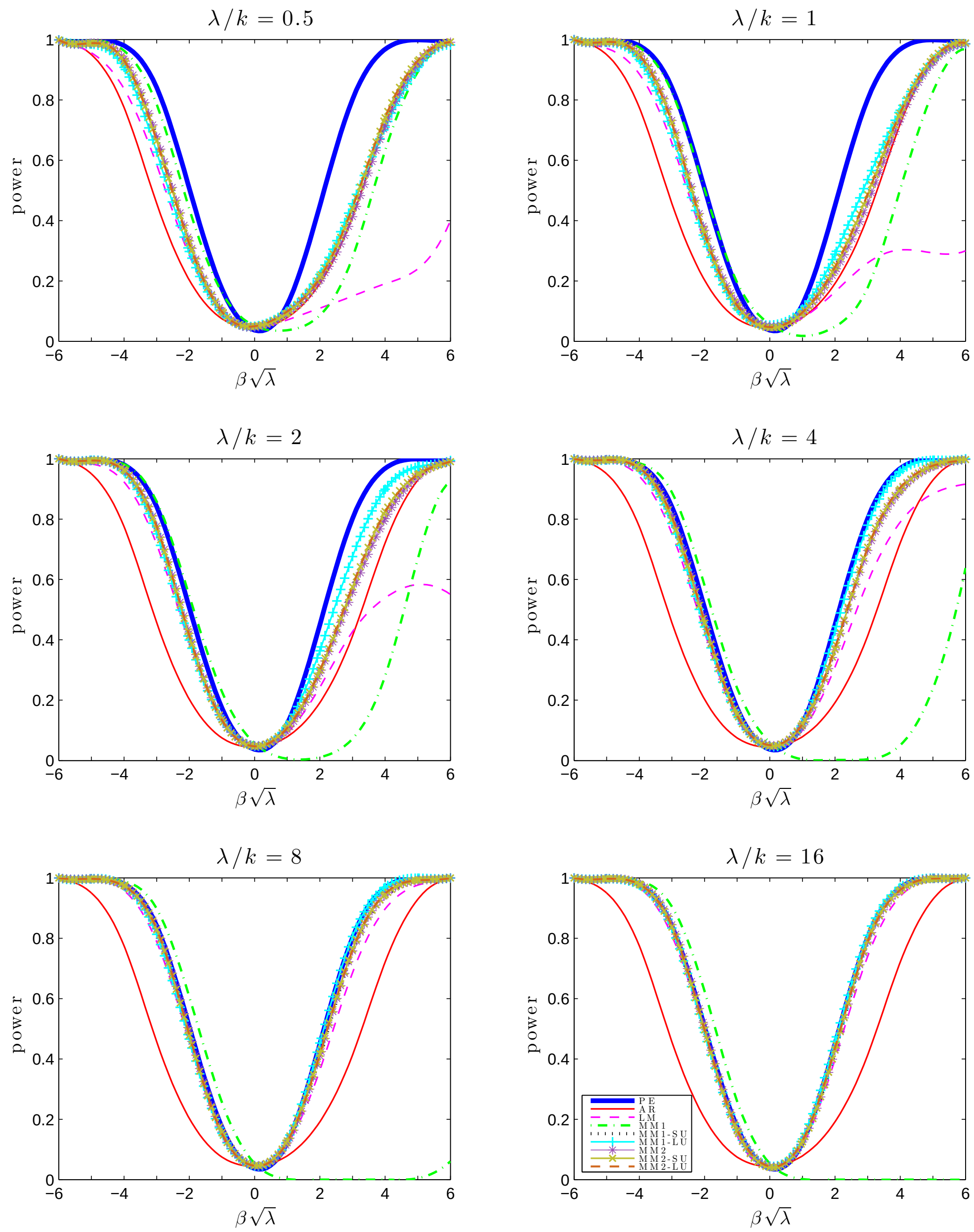


Figure 10: Power Comparison (Kronecker Covariance) $k = 10, \rho = 0.2$

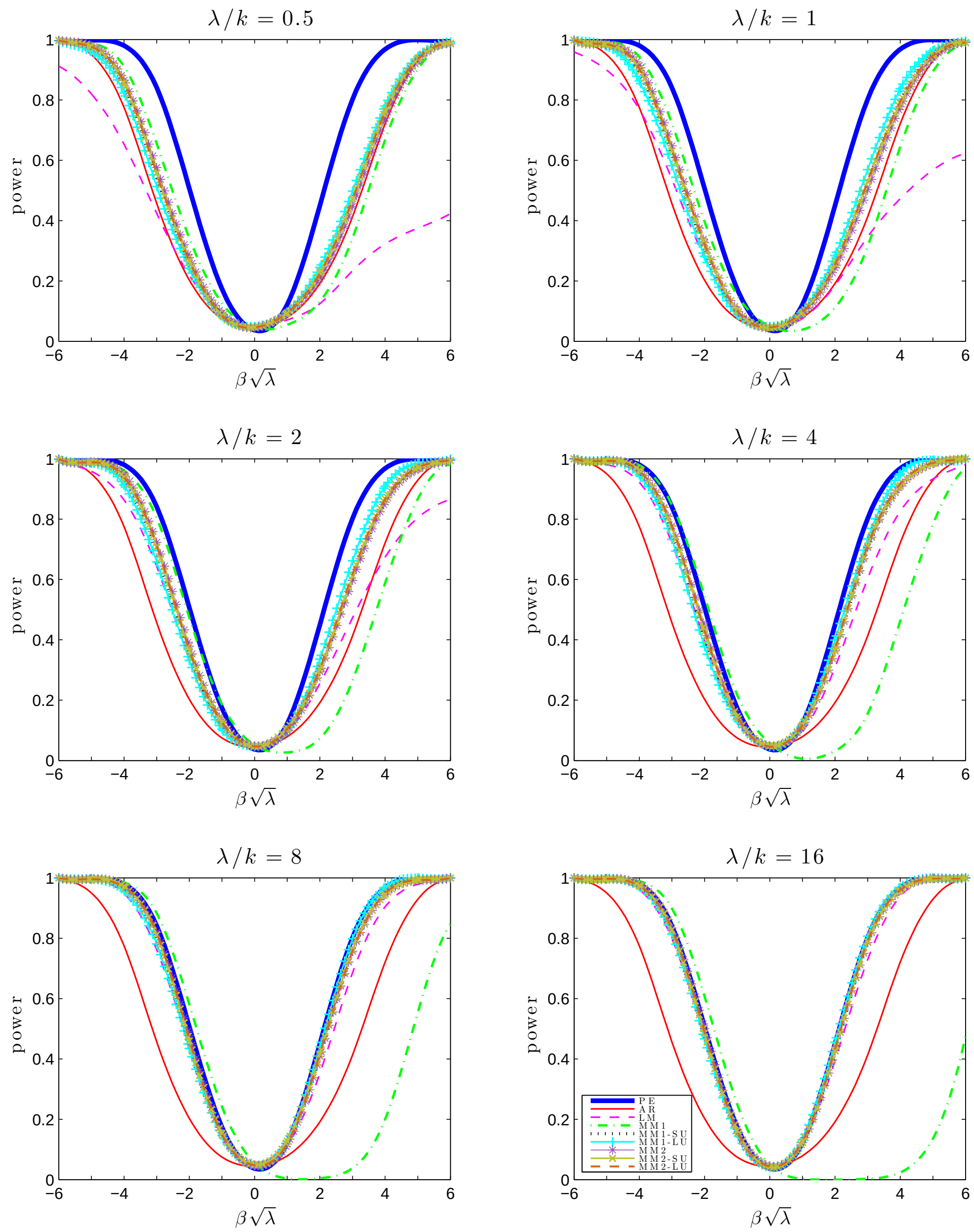


Figure 11: Power Comparison (Kronecker Covariance) $k = 10, \rho = 0.5$

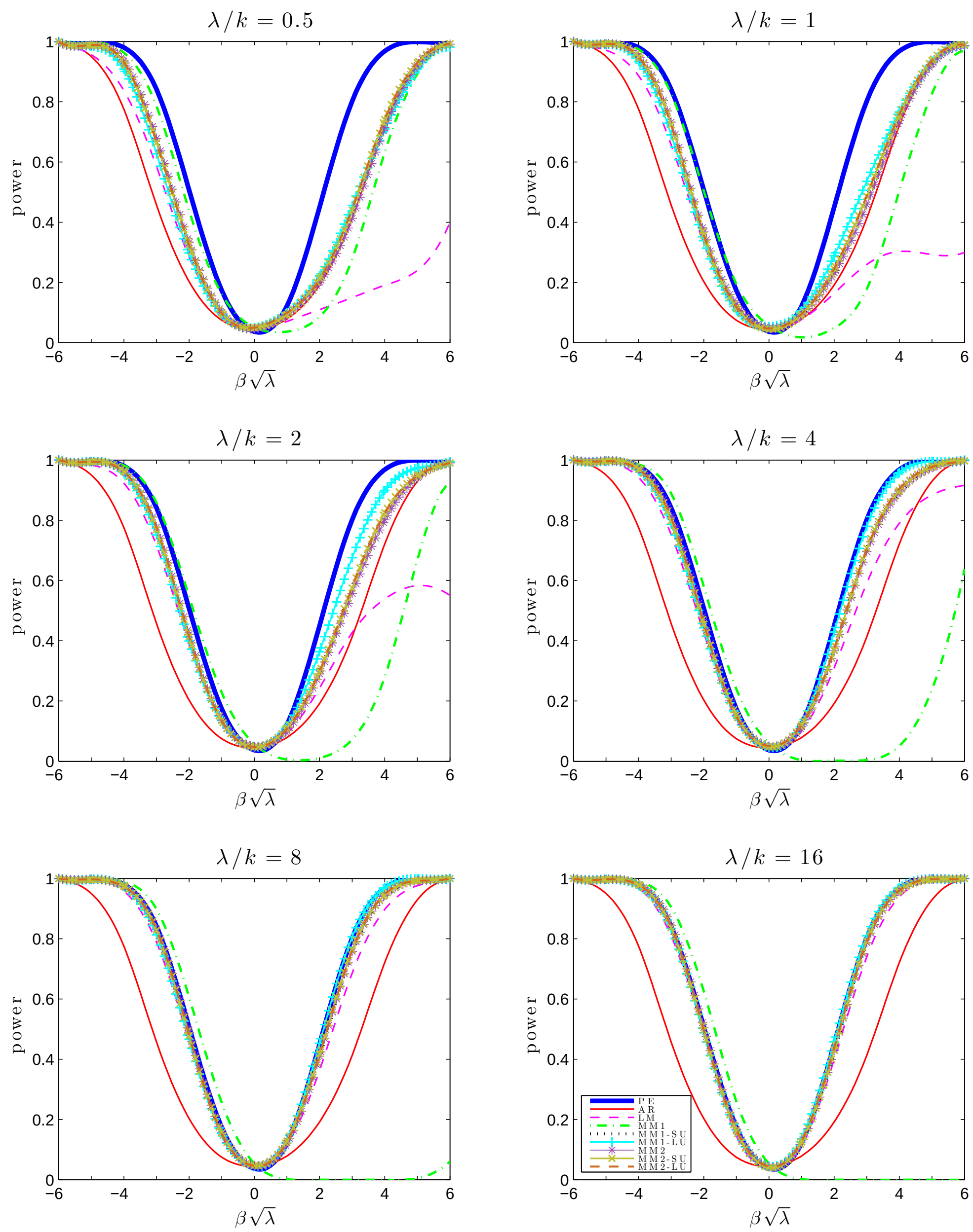


Figure 12: Power Comparison (Kronecker Covariance) $k = 10, \rho = 0.9$

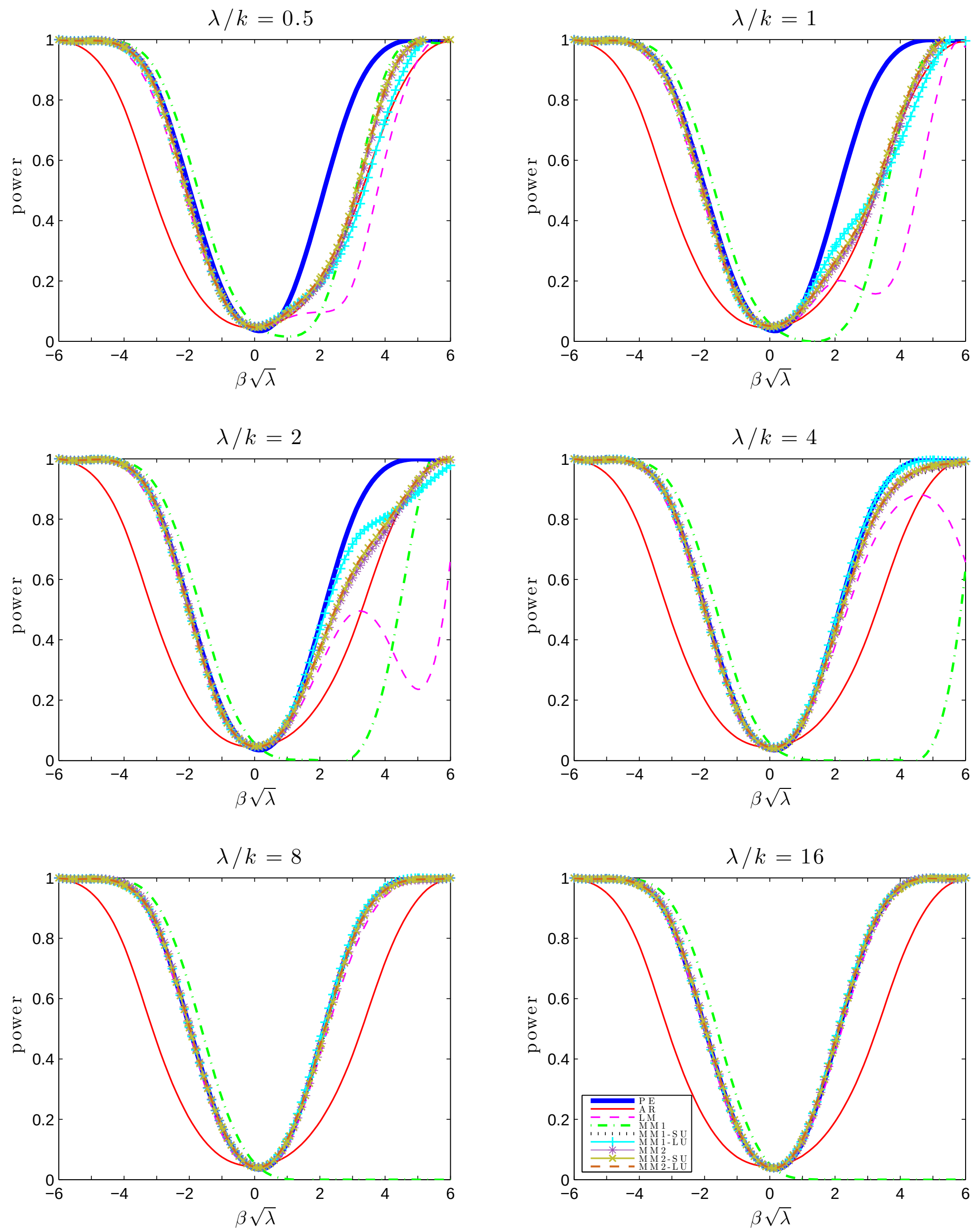


Figure 13: Power Comparison (Kronecker Covariance) $k = 20, \rho = -0.5$

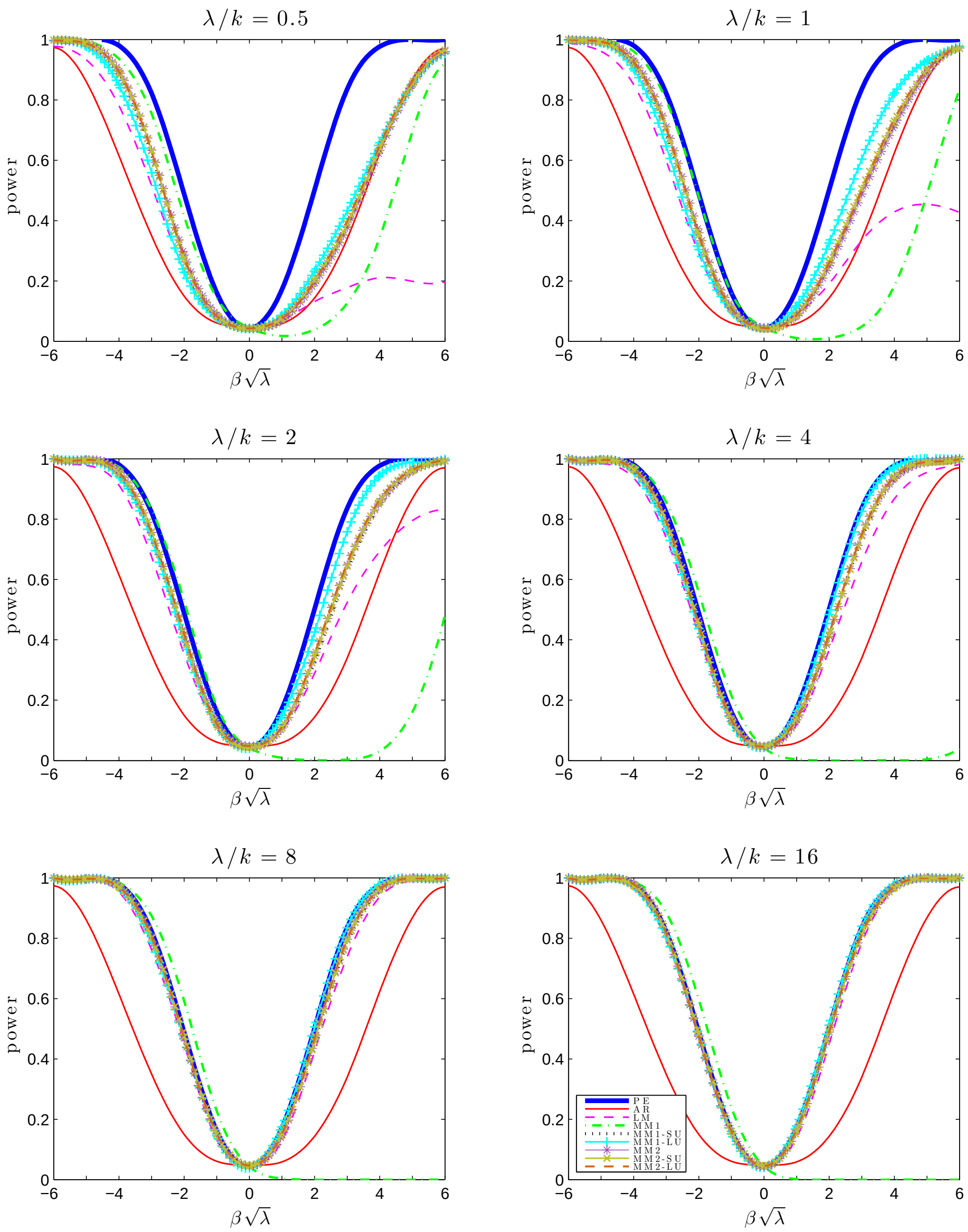


Figure 14: Power Comparison (Kronecker Covariance) $k = 20, \rho = 0.2$

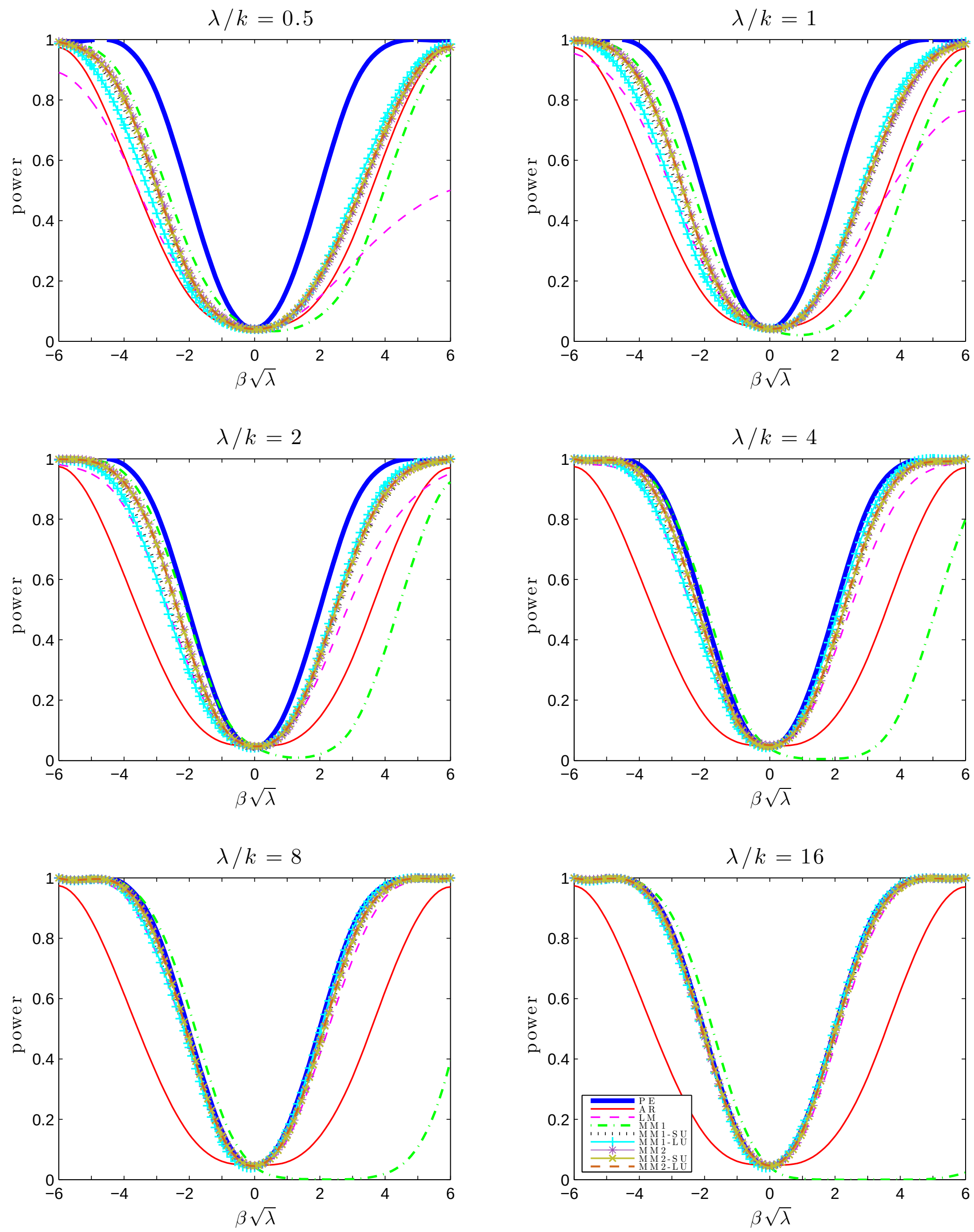


Figure 15: Power Comparison (Kronecker Covariance) $k = 20, \rho = 0.5$

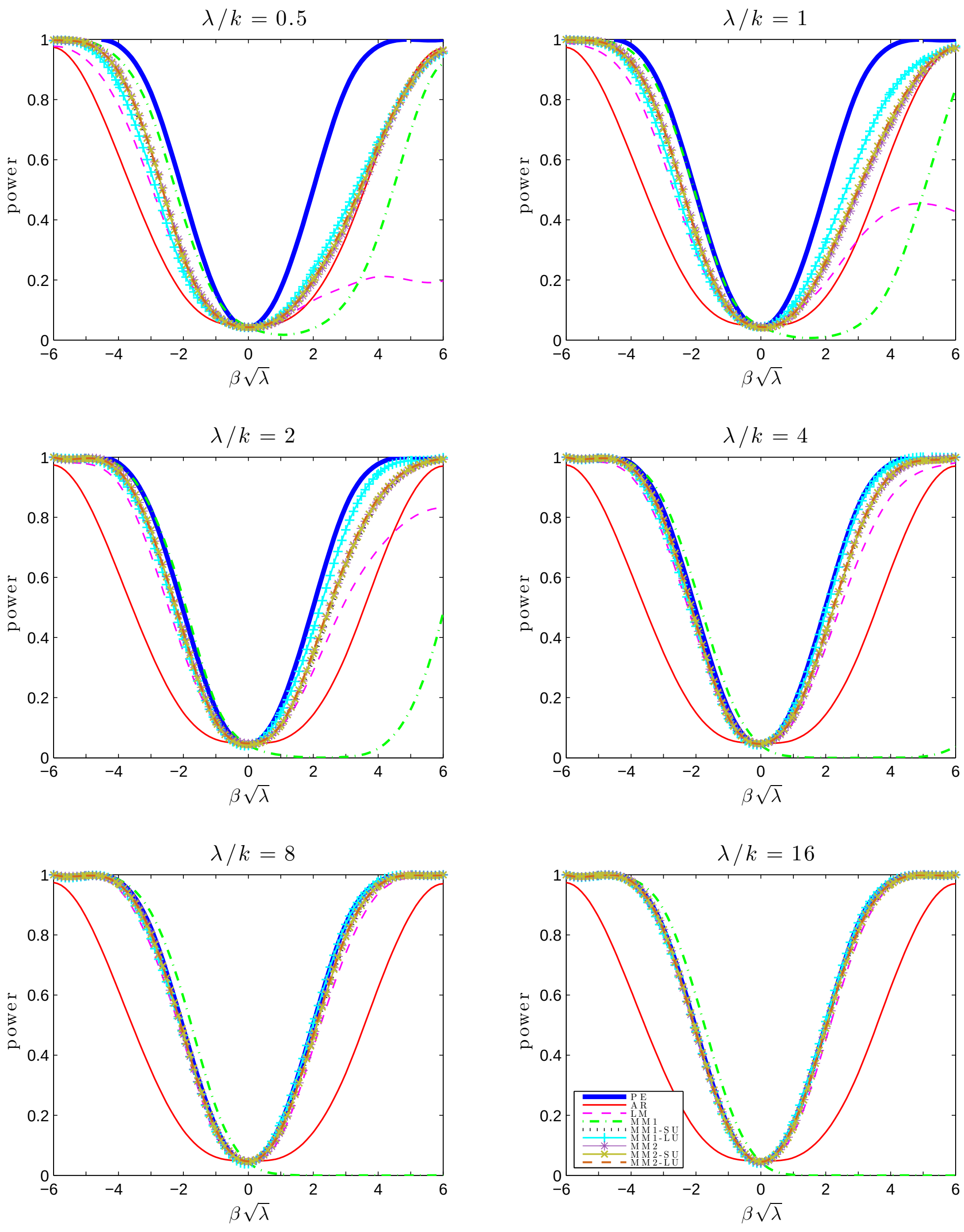


Figure 16: Power Comparison (Kronecker Covariance) $k = 20, \rho = 0.9$

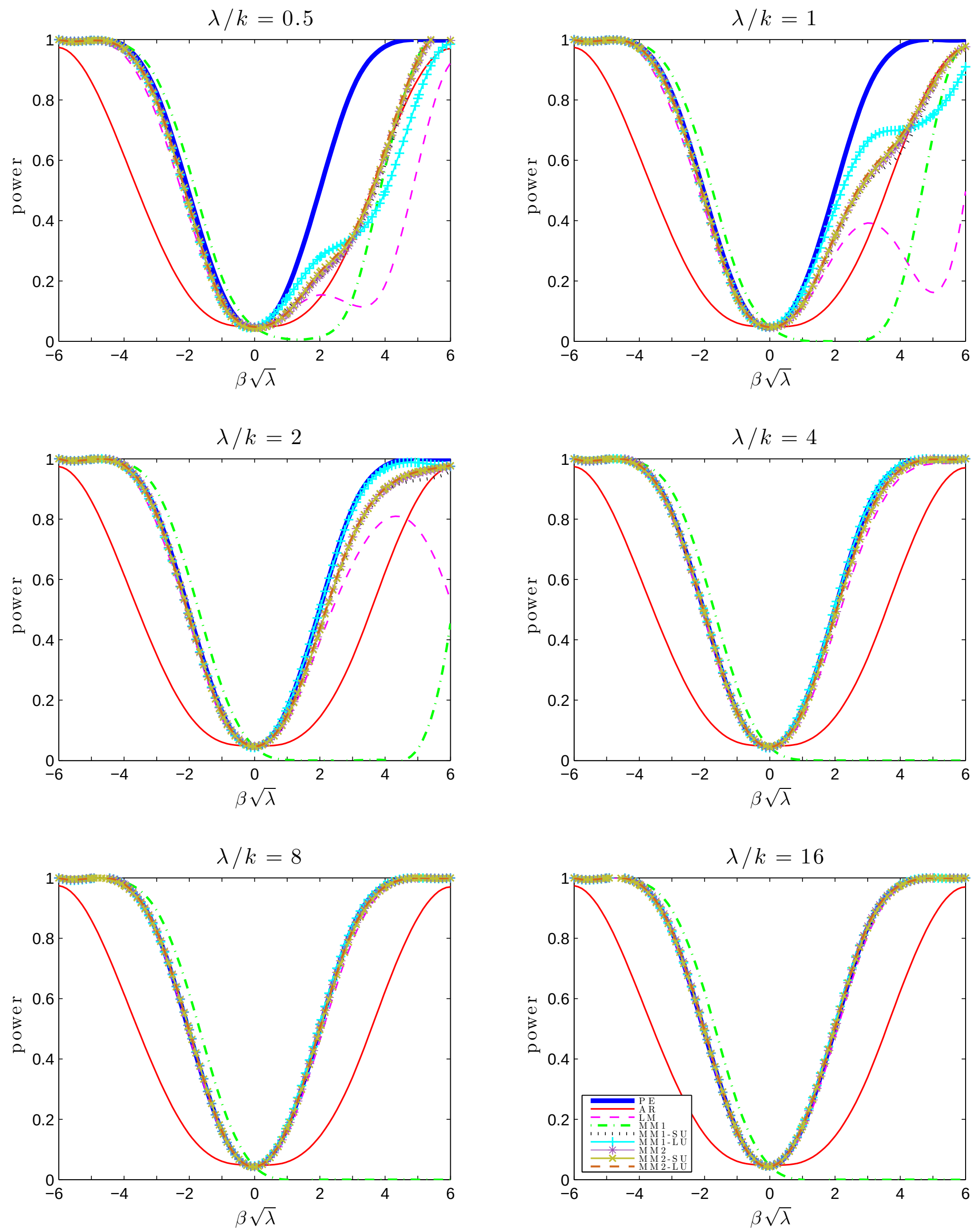


Figure 17: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = -0.5$

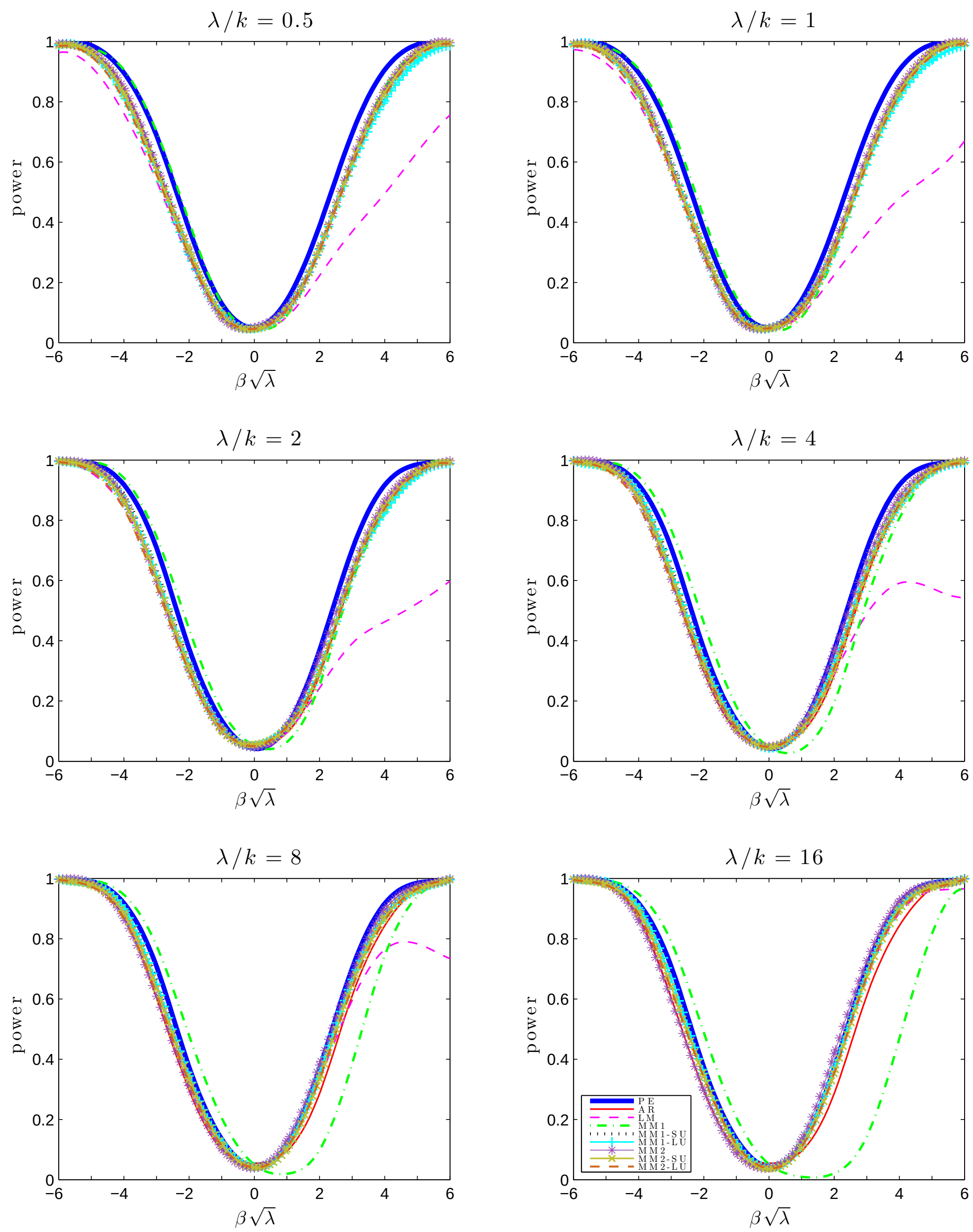


Figure 18: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = 0.2$

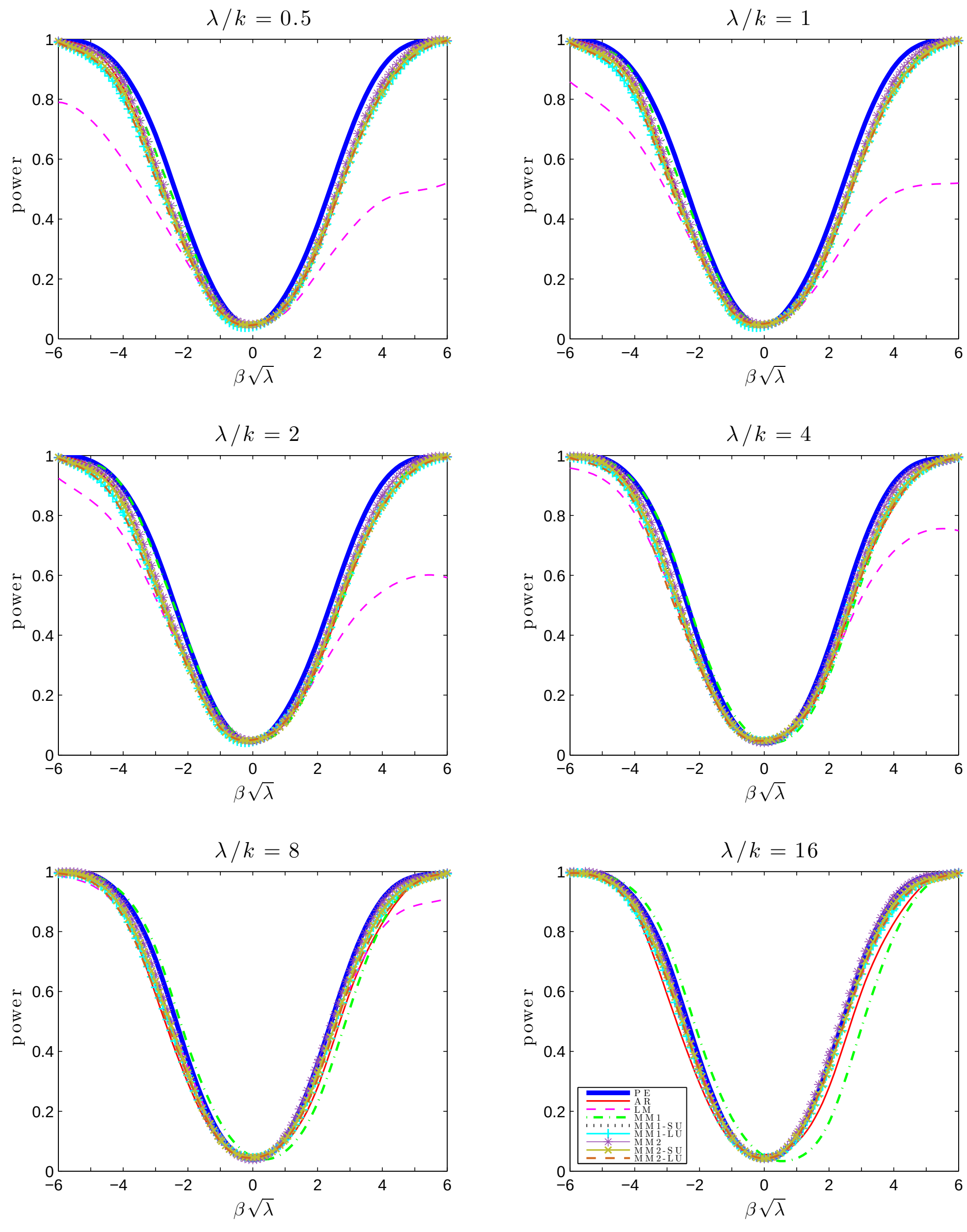


Figure 19: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = 0.5$

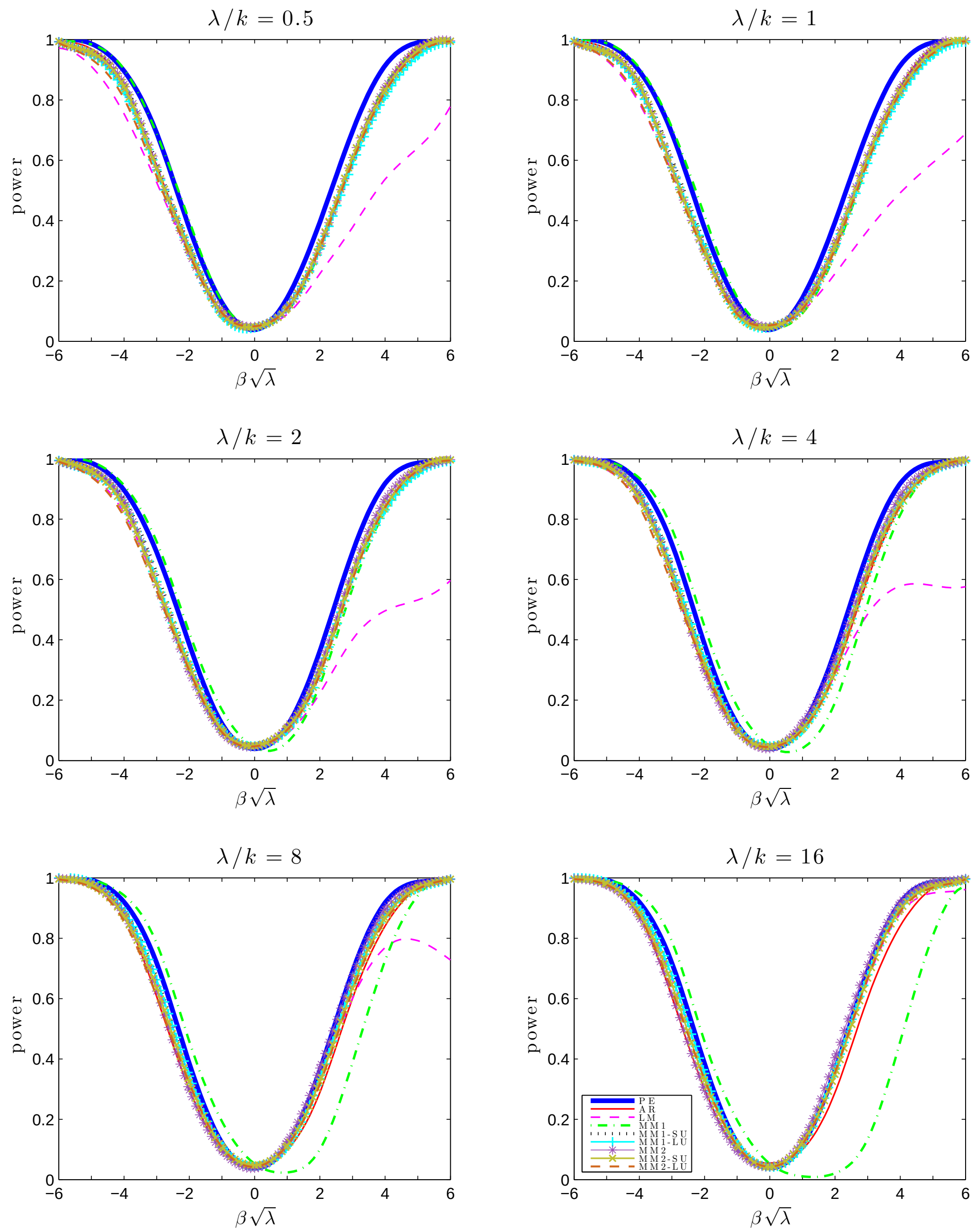


Figure 20: Power Comparison (Non-Kronecker Covariance) $k = 2, \rho = 0.9$

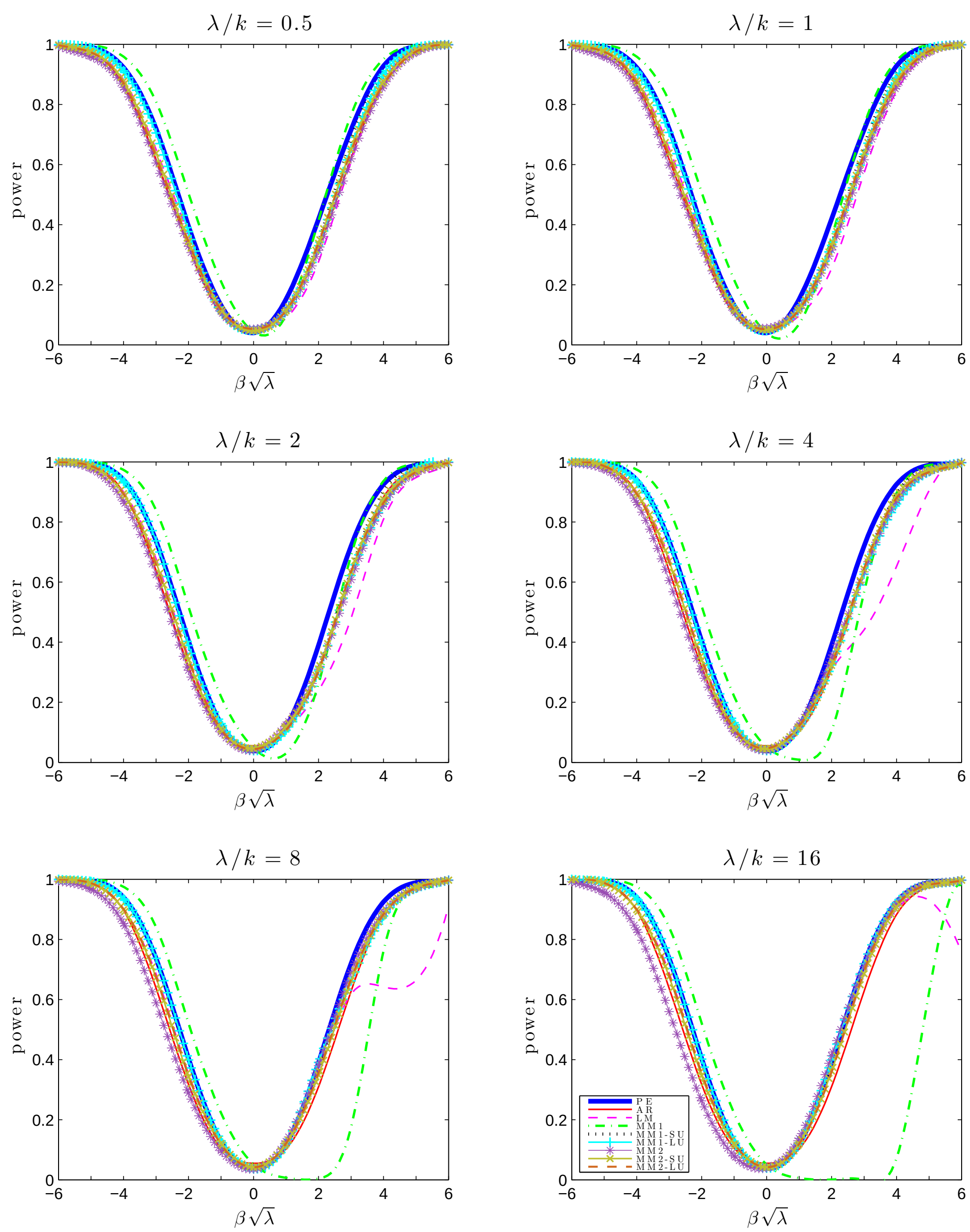


Figure 21: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = -0.5$

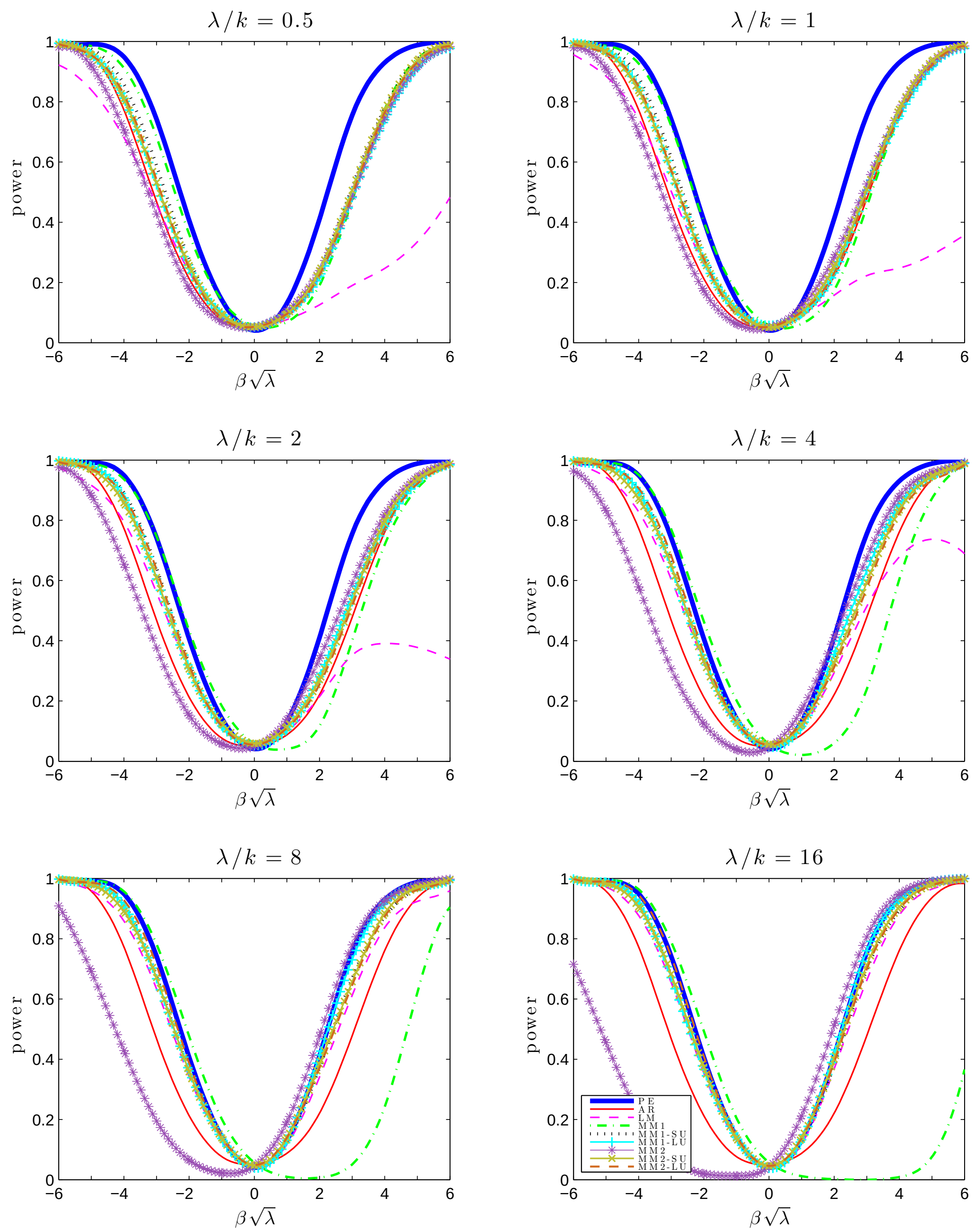


Figure 22: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = 0.2$

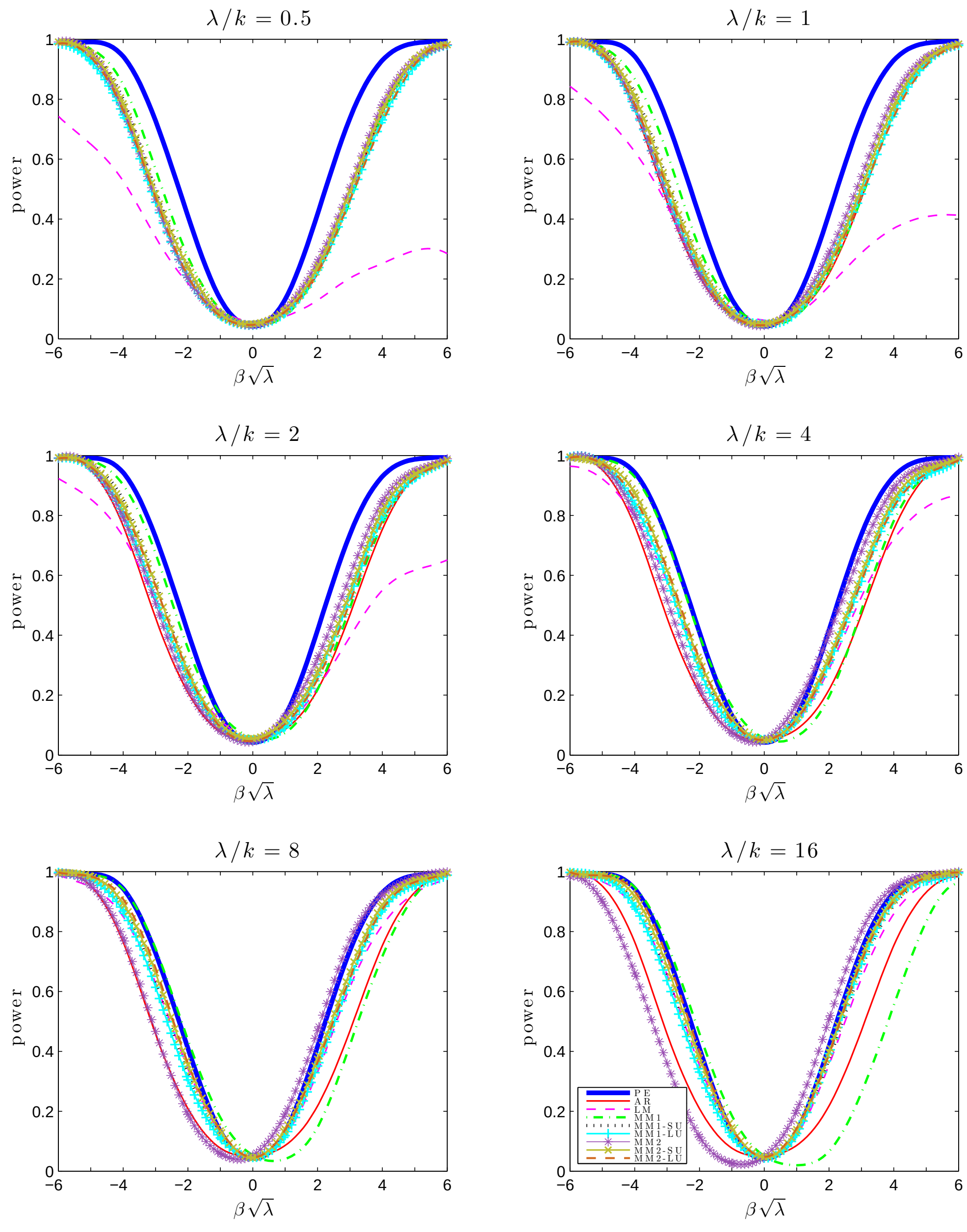


Figure 23: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = 0.5$

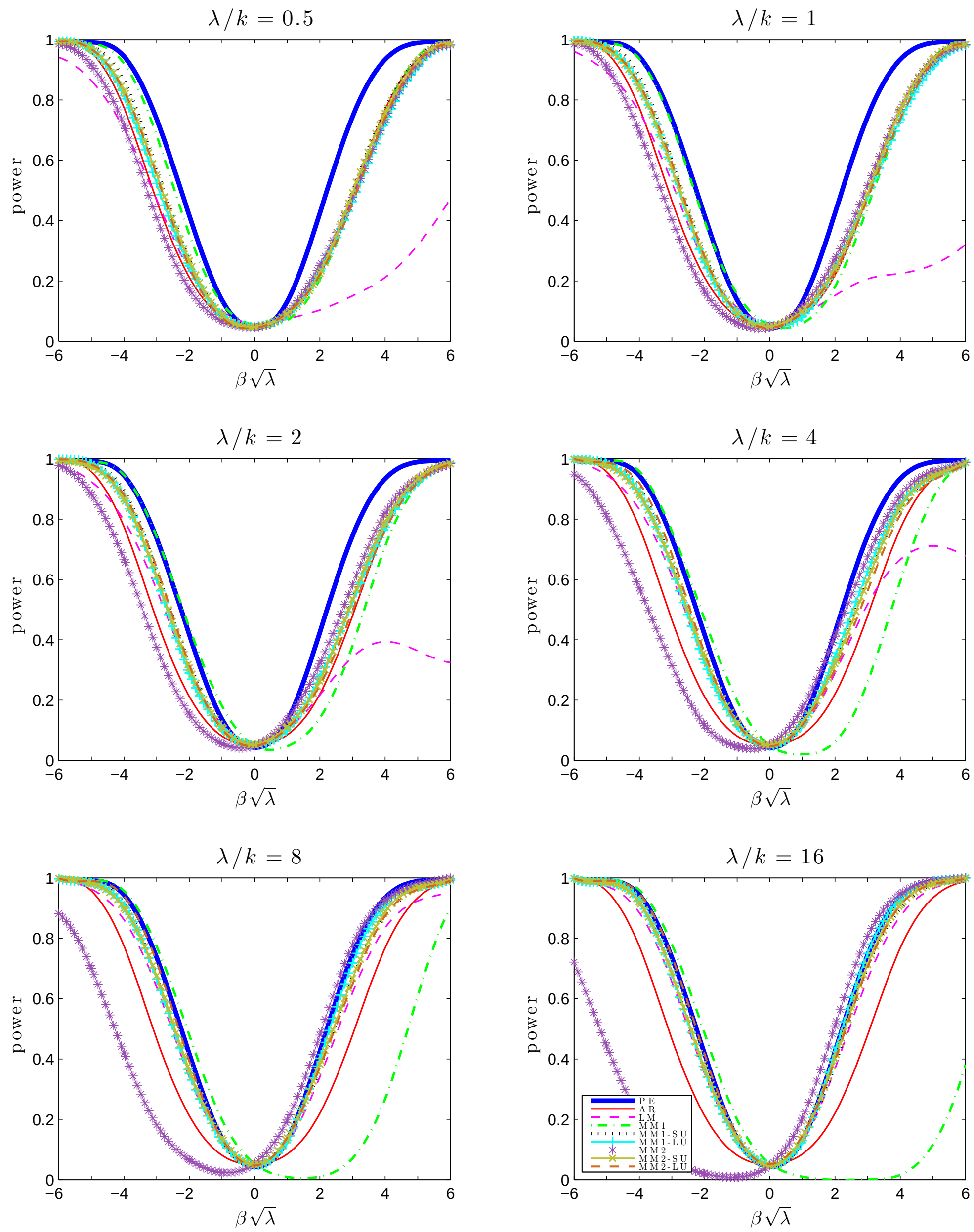


Figure 24: Power Comparison (Non-Kronecker Covariance) $k = 5, \rho = 0.9$

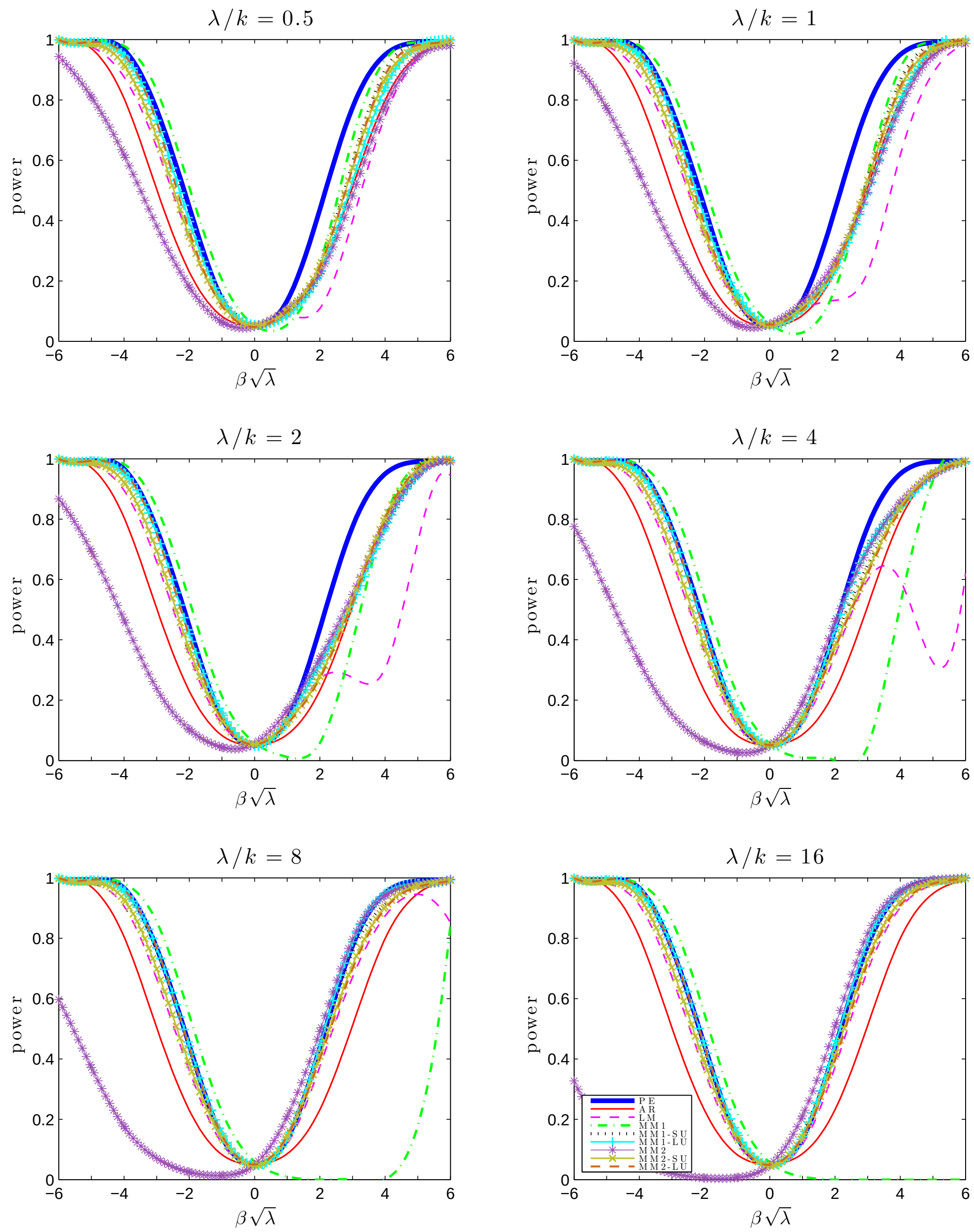


Figure 25: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = -0.5$

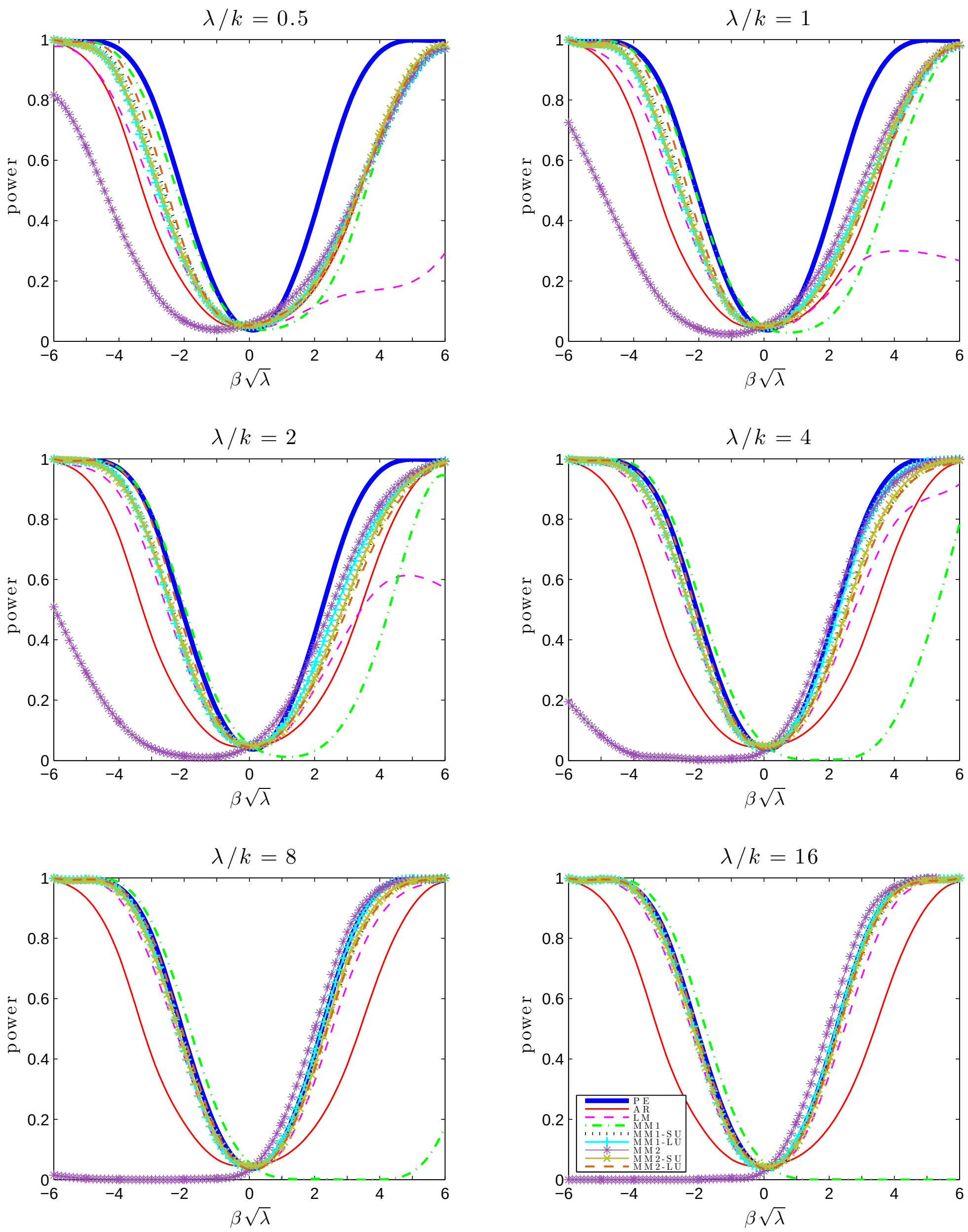


Figure 26: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = 0.2$

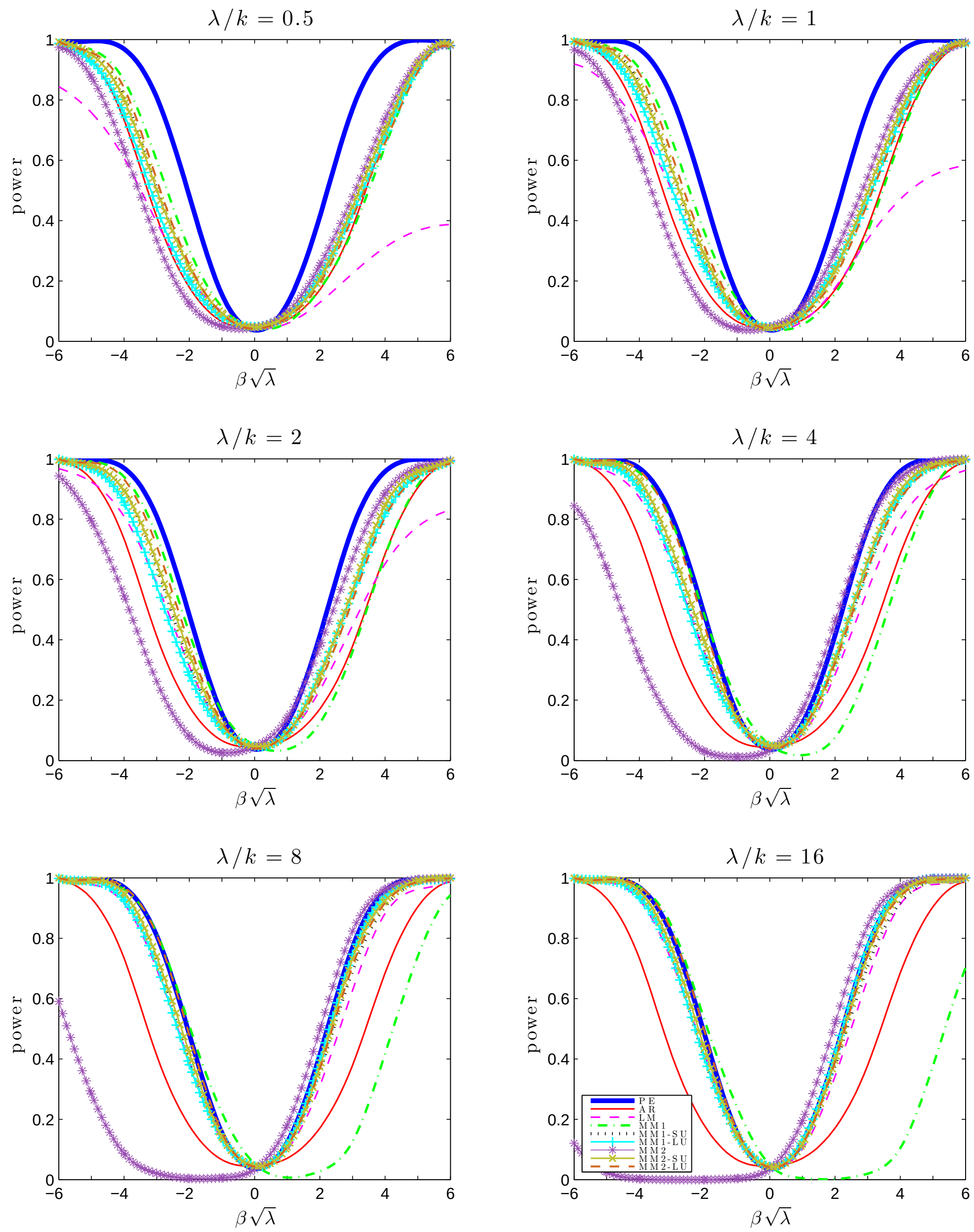


Figure 27: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = 0.5$

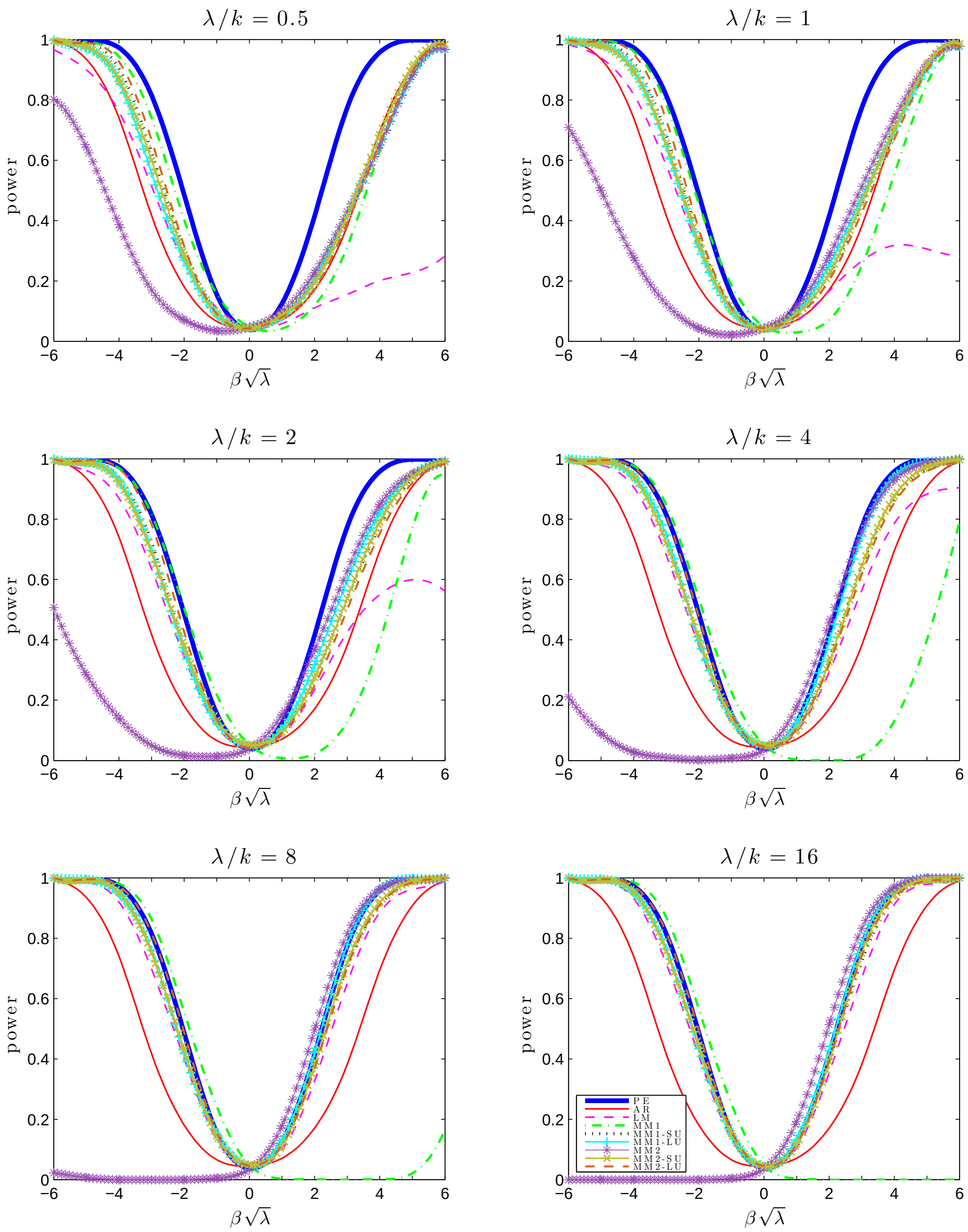


Figure 28: Power Comparison (Non-Kronecker Covariance) $k = 10, \rho = 0.9$

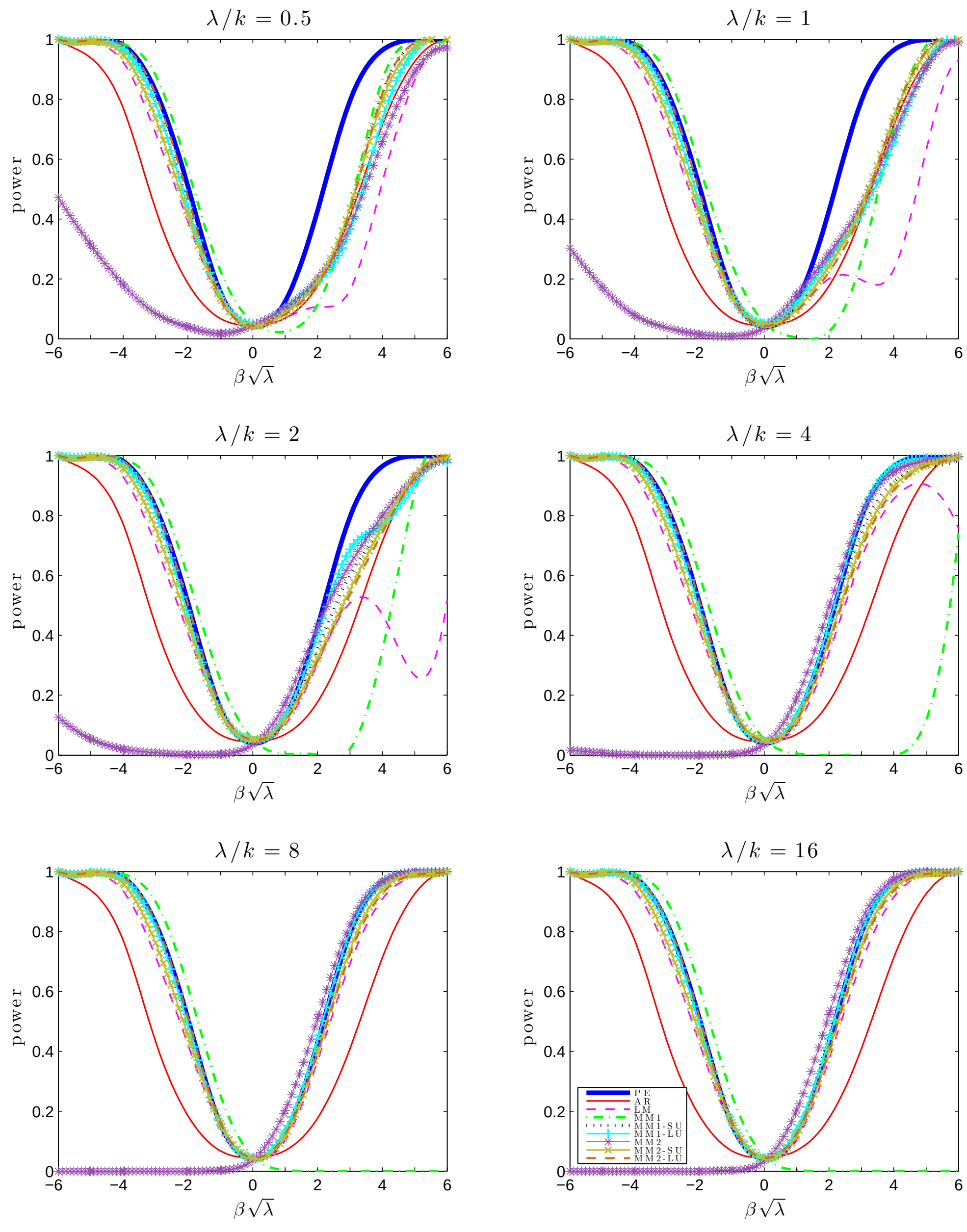


Figure 29: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = -0.5$

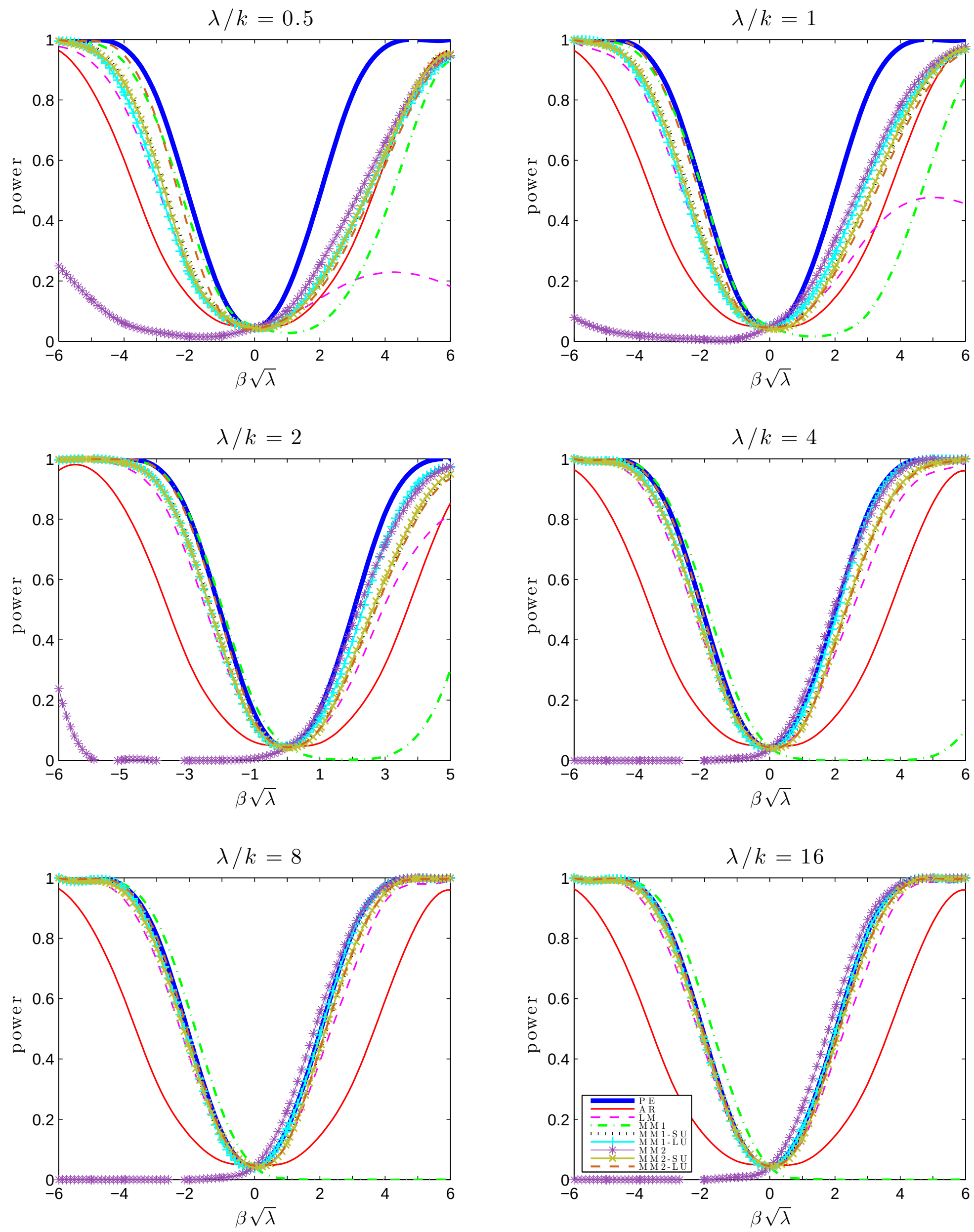


Figure 30: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = 0.2$

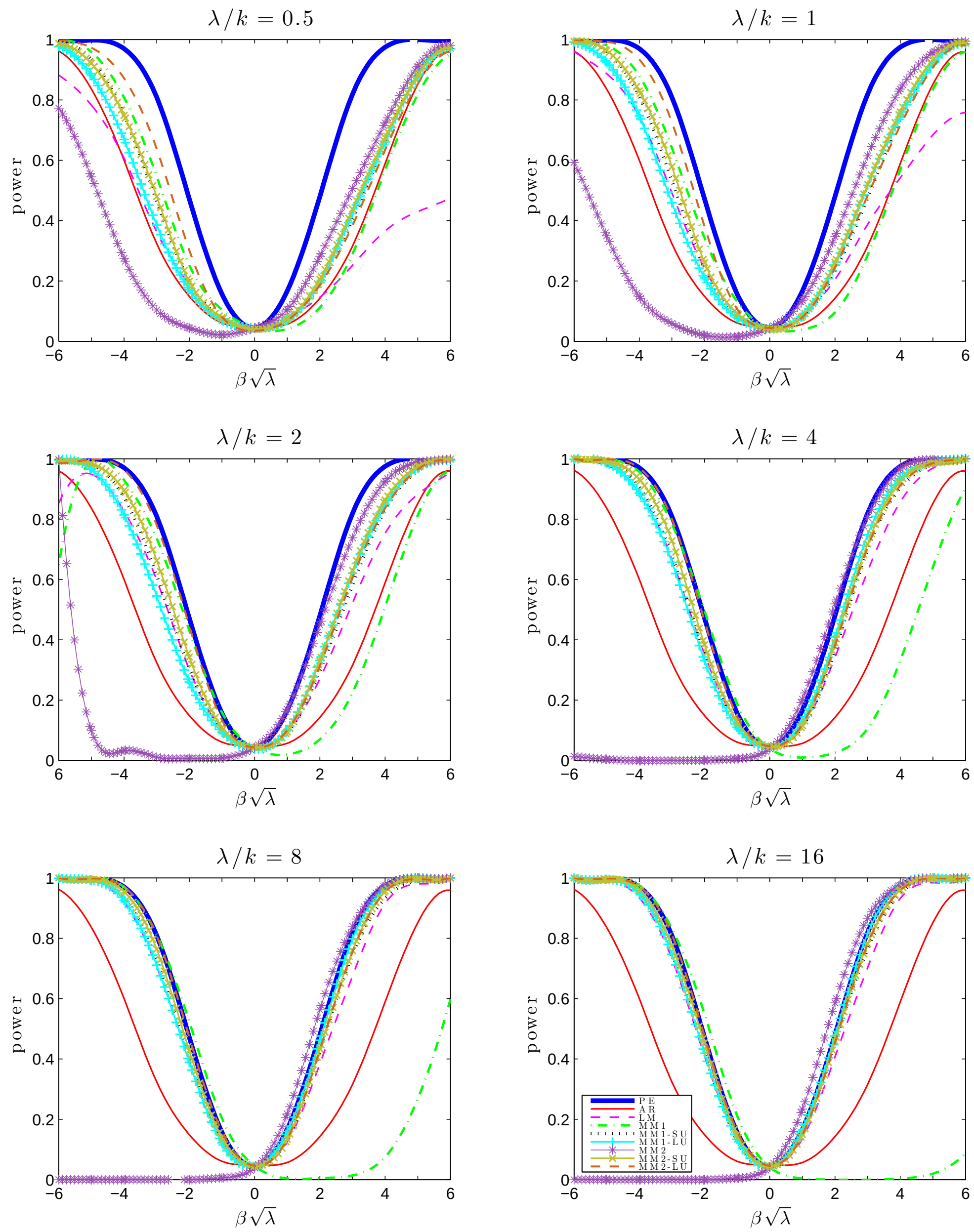


Figure 31: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = 0.5$

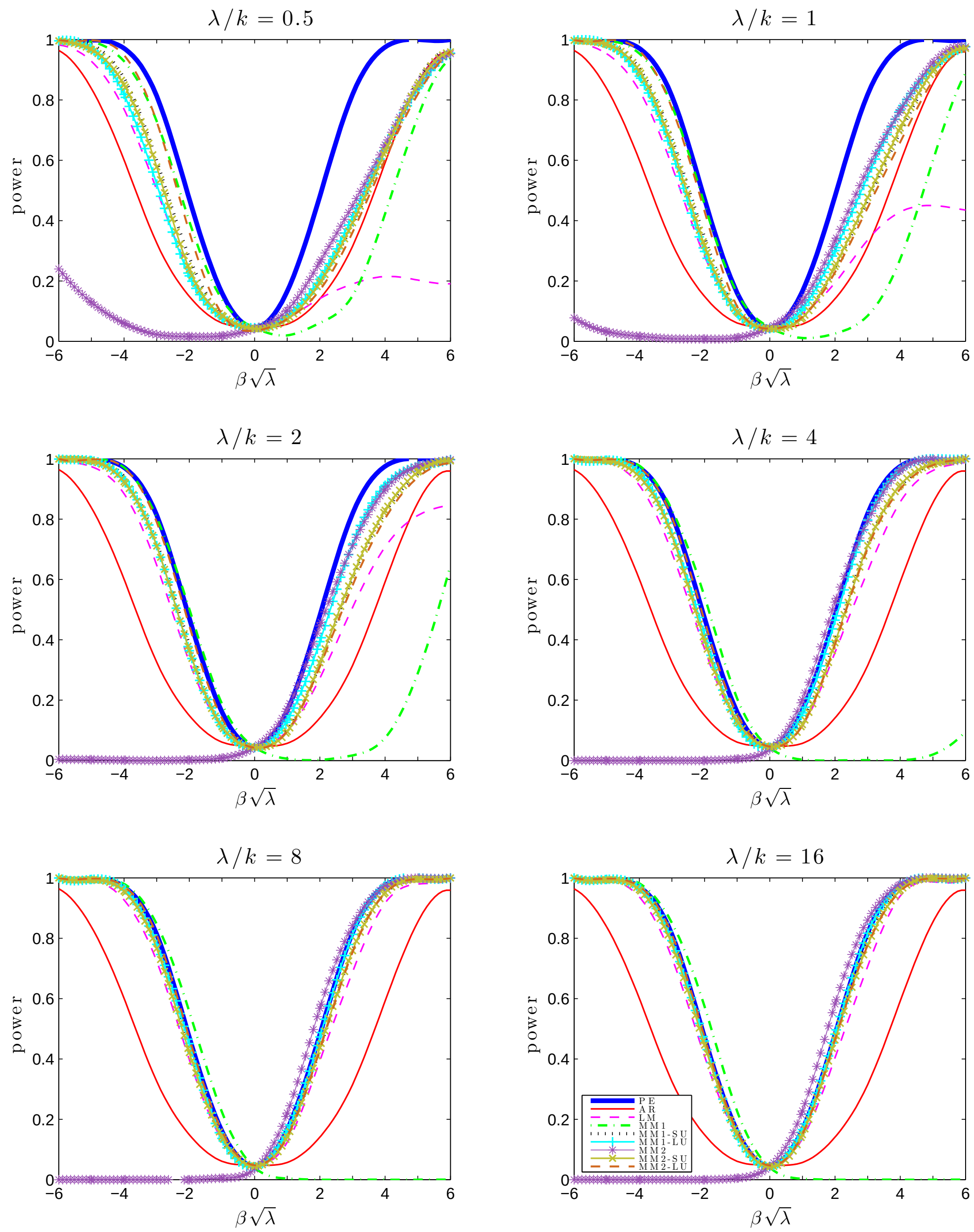
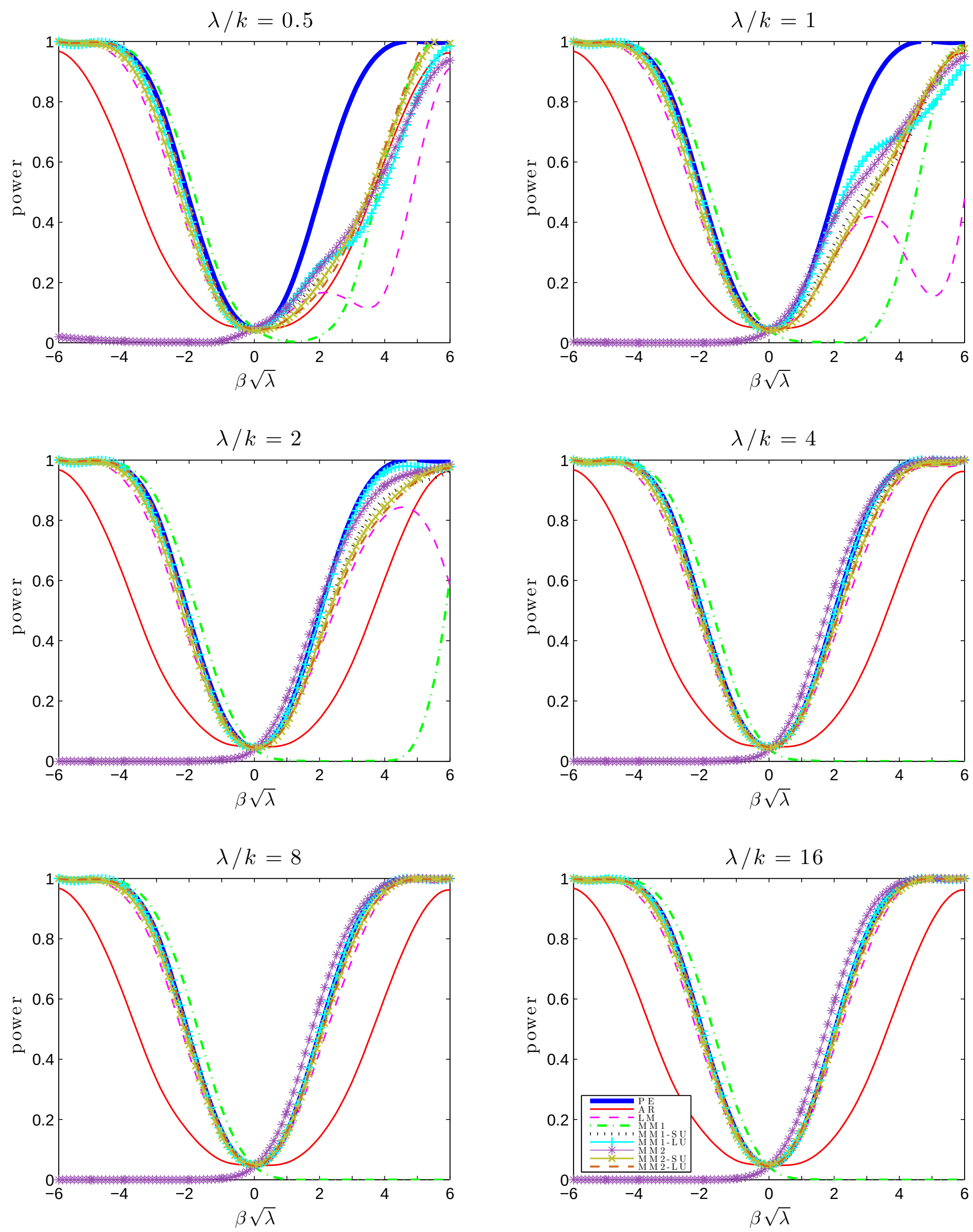


Figure 32: Power Comparison (Non-Kronecker Covariance) $k = 20, \rho = 0.9$



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: $\rho = -0.5$

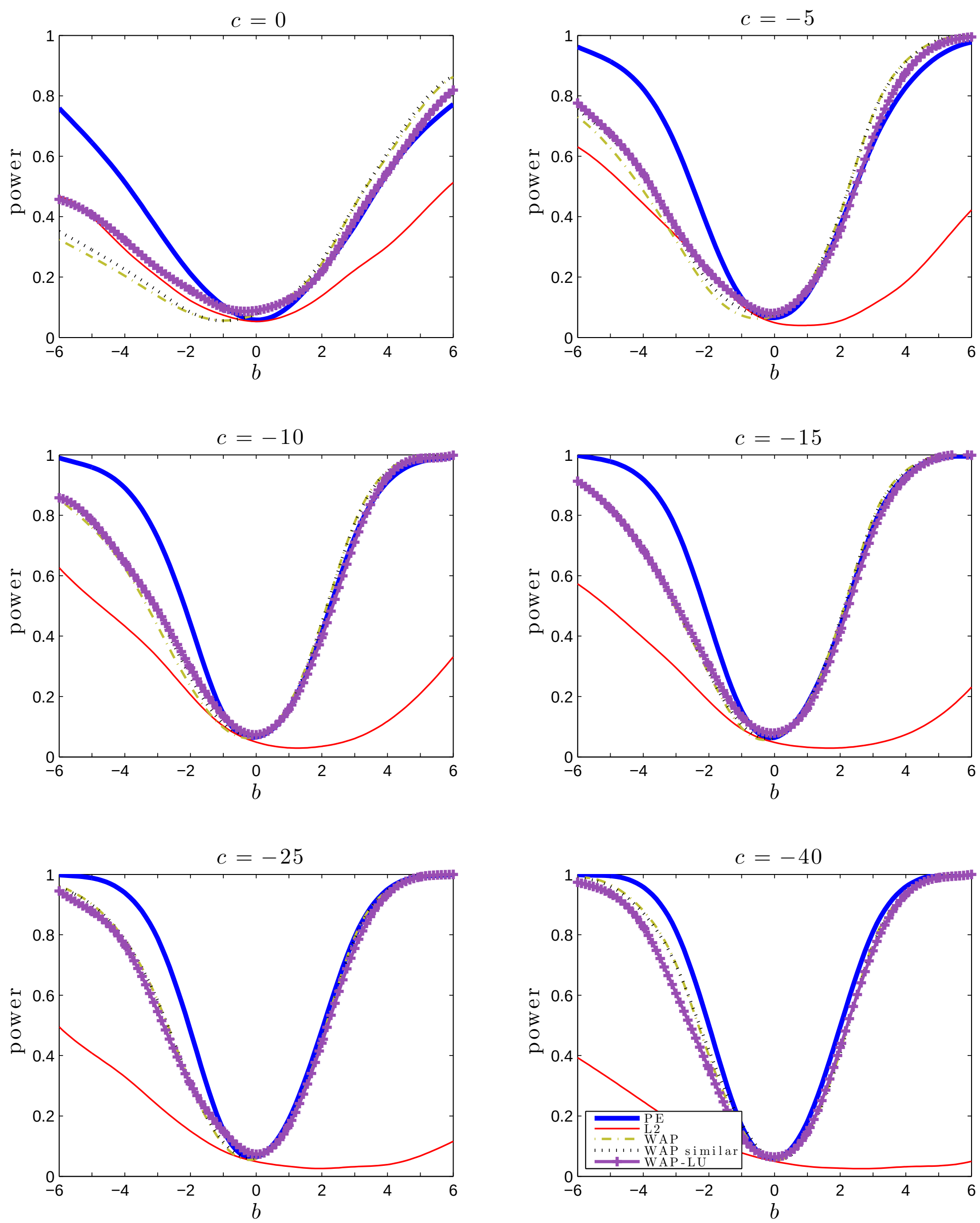
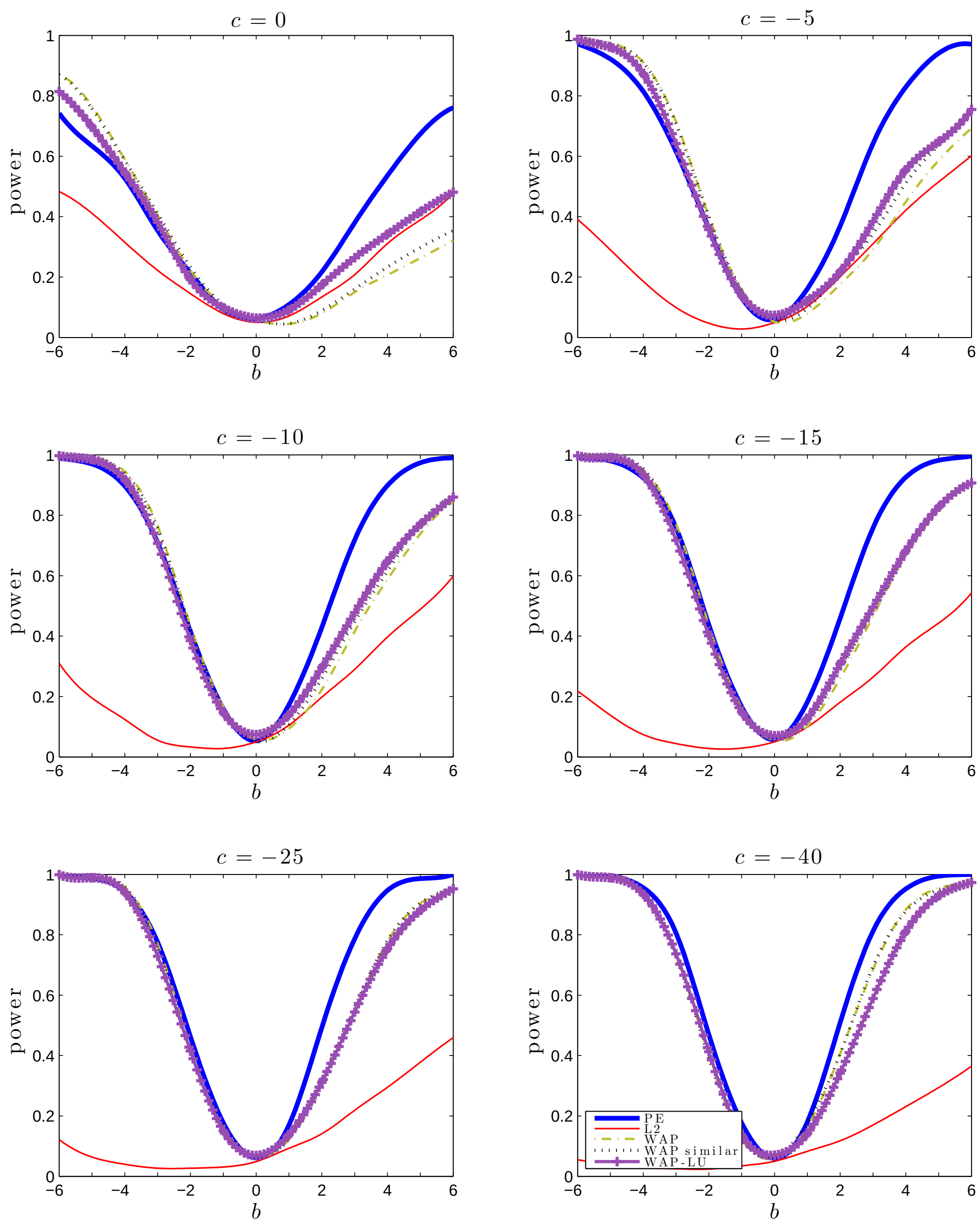


Figure 34: Power Comparison: $\rho = 0.5$



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.