FUNDAÇÃO GETÚLIO VARGAS
ESCOLA de PÓS-GRADUAÇÃO em ECONOMIA

Lucas Pimentel Vilela

Wald tests for IV Regression with Weak Instruments

Rio de Janeiro
2013

Lucas Pimentel Vilela

Wald tests for IV Regression with Weak Instruments

Dissertação para obtenção do grau de mestre apresentada à Escola de Pós-Graduação em Economia

Área de concentração: Econometria

Orientador: Marcelo J. Moreira

Rio de Janeiro
2013

Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Vilela, Lucas Pimentel
Wald tests for IV regression with weak instruments / Lucas Pimentel Vilela. – 2013.
44 f.

Dissertação (mestrado) - Fundação Getulio Vargas, Escola de Pós-Graduação em Economia.
Orientador: Marcelo J. Moreira.
Inclui bibliografia.

1. Variáveis instrumentais (Estatística). 2. Análise de regressão. 3. Testes de hipóteses estatísticas. I. Moreira, Marcelo J. II. Fundação Getulio Vargas. Escola de Pós-Graduação em Economia. III. Título.

CDD – 519.535

FUNDAÇÃO
GETULIO VARGAS

LUCAS PIMENTEL VILELA

WALD TESTS FOR IV REGRESSION WITH WEAK INSTRUMENTS.

Dissertação apresentada ao Curso de Mestrado em Economia da Escola de Pós-Graduação em Economia para obtenção do grau de Mestre em Economia.

Data da defesa: 17/09/2013

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

Marcelo Jovita Moreira
Orientador (a)

Caio Ibsen Rodrigues de Almeida

Marcelo Cunha Medeiros

Acknowledgements

I would like to express my sincere appreciation and gratitute to my advisor Marcelo J. Moreira for his guidance and support in my dissertation. I am also indebted to Donald Andrews, James Stock and Benjamim Mills for their help and suggestions.

Resumo

Esta dissertação trata do problema de inferência na presença de identificação fraca em modelos de regressão com variáveis instrumentais. Mais especificamente em testes de hipóteses com relação ao parâmetro da variável endógena quando os instrumentos são fracos. O principal foco é nos testes condicionais unilaterais baseados nas estatísticas de razão de máxima verossimilhança, score e Wald. Resultados téoricos e numéricos mostram que o teste t condicional unilateral baseado no estimador de mínimos quadrados em dois estágios tem uma boa performance mesmo na presença de instrumentos fracamente correlacionados com a variável endógena. A abordagem condicional corrige uniformemente o tamanho do teste t e quando a estatística F populacional é tão pequena quanto dois, o poder do teste é proximo ao *power envelope* tanto de testes similares quanto de não similares. Tal resultado é surpreendente visto a má performace dos testes t's condicionais bilaterais relatada em (6, Andrews, Moreira and Stock (2007)). Dado esse resultado aparentemente contra intuitivo, apresentamos novos testes t's condicionals bilaterais que são aproximadamente não viesados e performam, em alguns casos, tão bem quanto o teste condicional baseado na estatística de razão de verossimilhança de (19, Moreira (2003)).

Palavras-chave: Regressão com variáveis instrumentais, testes invariantes, testes ótimos, testes similares, testes não viesados, instrumentos fracos.

Abstract

This dissertation deals with the problem of making inference when there is weak identification in models of instrumental variables regression. More specifically we are interested in one-sided hypothesis testing for the coefficient of the endogenous variable when the instruments are weak. The focus is on the conditional tests based on likelihood ratio, score and Wald statistics. Theoretical and numerical work shows that the conditional t-test based on the two-stage least square (2SLS) estimator performs well even when instruments are weakly correlated with the endogenous variable.The conditional approach correct uniformly its size and when the population F-statistic is as small as two, its power is near the power envelopes for similar and non-similar tests. This finding is surprising considering the bad performance of the two-sided conditional t-tests found in (6, Andrews, Moreira and Stock (2007)). Given this counter intuitive result, we propose novel two-sided t-tests which are approximately unbiased and can perform as well as the conditional likelihood ratio (CLR) test of (19, Moreira (2003)).

Keywords: Instrumental variable regression, invariant tests, optimal tests, similar tests, unbiased tests, weak instruments.

List of Figures

Contents

[1] This is a work with Marcelo J. Moreira and Benjamin Mills.

Chapter 1

Introduction[1]

Instrumental variables (IVs) are commonly used to make inferences about the coefficient β of an endogenous regressor in a structural equation. When instruments are strongly correlated with the regressor, the tests based on the score (also known as lagrange multiplier (LM)), likelihood ratio (LR) and Wald (t-statistics) are asymptotically equivalent. This trinity of tests provides reliable inference as long as the instruments are strong. However, when identification is weak, the three approaches are no longer comparable. (16, Kleibergen (2002)) and (18, Moreira (2002)) show that the LM statistic has a standard chi-square distribution regardless of the strength of the instruments. (19, Moreira (2003)) proposes a conditional likelihood ratio (CLR) test which is shown by (4, Andrews, Moreira and Stock (2006)) (hereinafter, AMS06a) to be nearly optimal. However, most results in the literature on the performance of tests based on the commonly used t-statistics are negative: (13, Dufour (1997)) shows that standard tests based on t-statistics can have size arbitrarily close to one; (6, Andrews, Moreira and Stock (2007)) (hereinafter, AMS07) find that conditional t-tests are severely biased; and (2, Andrews and Guggenberger (2010)) prove that subsampling tests based on the two-stage least squares (2SLS) t-statistic do not have correct asymptotic size. See (22, Stock, Wright, and Yogo (2002)), (14, Dufour (2003)), and (8, Andrews and Stock (2007)) for surveys on weak IVs.

In this dissertation we present one-sided conditional t-tests for testing the null hypothesis $H_0 : \beta = \beta_0$ (or the augmented null $H_0 : \beta \leq \beta_0$) against the alternative $H_1 : \beta > \beta_0$ (the adjustment for $H_1 : \beta < \beta_0$ is straightforward). We consider t-statistics centered around the 2SLS, the limited information maximum likelihood (LIML), the bias-adjusted 2SLS (B2SLS) and the estimator proposed by (15, Fuller (1977)) (Fuller's estimator). We also introduce conditional tests based on an one-sided score (LM1) statistic, a likelihood ratio (LR1) statistic for $H_0 : \beta = \beta_0$, and a likelihood ratio statistic (MLR1) for $H_0 : \beta \leq \beta_0$. We develop a theory of optimal tests for one-sided alternatives that parallels the two-sided results of AMS06a. We adopt the same invariance condition as in AMS06a and (11, Chamberlain (2007)) under which inference is unchanged if the IVs are transformed by an orthogonal matrix, e.g., by changing the order in which the IVs appear. We develop the Gaussian power envelope for point-optimal invariant similar (POIS) tests. When the null hypothesis is $H_0 : \beta = \beta_0$, the conditional LR1 (CLR1) test is nearly optimal in the sense that its power function is numerically close to the power envelope. For the more relevant null $H_0 : \beta \leq \beta_0$, the CLR1 test does not control size uniformly. The conditional t-tests have correct size and the one based on the 2SLS estimator numerically outperforms the conditional MLR1 (CMLR1) test. The LM1 test is a POIS test and does not have good power overall.

The good performance of the one-sided conditional 2SLS t-test is somewhat surprising considering the bad performance of two-sided conditional t-tests found in AMS07. We show that the bad performance is due to the asymmetric distribution of t-statistics under the null $H_0 : \beta = \beta_0$ when instruments are weak. We consider two methods to improve power for two-sided tests

[1]This is a work with Marcelo J. Moreira and Benjamin Mills.

based on t-statistics. First, we propose novel tests which are by construction approximately unbiased. Second, we modify the t-statistics so that their null distribution is nearly symmetric. Both methods yield some t-tests whose power is close to the CLR test. Hence, this dissertation restores the triad of tests based on score, likelihood ratio, and t-statistics with reasonably good performance even when instruments are weak for two-sided hypothesis testing. By inverting the conditional t-tests, we can obtain informative confidence regions around different estimators –including the commonly used 2SLS estimator.

The foregoing results are developed under the assumption of normal reduced-form errors with known covariance matrix. The finite-sample theory is extended to non-normal errors with unknown variance at the cost of introducing asymptotic approximations. Under weak instrumental variable (WIV) asymptotics, the exact distributional results extend in large samples to feasible versions of the proposed tests. The finite-sample Gaussian power envelopes are also the asymptotic Gaussian power envelopes with unknown covariance matrix. Under strong-IV asymptotics, we derive consistency even when errors are nonnormal and asymptotic efficiency (AE) when errors are normal.[2]

The dissertation is organized as follows: Chapter 2 introduces the model with one endogenous regressor variable, multiple exogenous regressor variables, and multiple IVs. This chapter determines sufficient statistics for this model with normal errors and reduced-form covariance matrix. Chapter 3 introduces one-sided invariant similar tests. Chapter 4 finds the power envelope for similar and nonsimilar one-sided tests. Chapter 5 adjusts the tests to allow for an estimated error covariance matrix and analyzes their asymptotic properties under weak IVs. Chapter 6 obtains consistency and asymptotic efficiency for one-sided tests. Chapter 7 compares numerically the power of the tests considered in earlier chapters under WIV asympotics. Chapter 8 introduces novel unbiased two-sided tests. An appendix contains proofs of the results. The supplement presents: power comparisons for different one-sided and two-sided tests; similar and non-similar power envelopes which are numerically very close (this fact further strengthens our optimality results); and confidence intervals for returns to schooling using the data of (9, Angrist and Krueger (1991)).

[2] In principle, we could follow (10, Cattaneo, Crump and Jansson (2012)) to obtain efficient one-sided tests when errors are nonnormal, but we do not pursue this line of research here.

Chapter 2

Model and Sufficient Statistics

In this dissertation we study linear instrumental variable regression models with the objective of making inference about the coefficient of the endogenous variable when the instruments are possibly weak. More specifically, we want to make hypothesis tests of β in the following linear model:

$$\begin{aligned} y_1 &= y_2\beta + X\gamma_1 + u, \\ y_2 &= \widetilde{Z}\pi + X\xi_1 + v_2, \end{aligned} \tag{2.1}$$

where $y_1, y_2 \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, and $\widetilde{Z} \in \mathbb{R}^{n\times k}$ are observed variables; $u, v_2 \in \mathbb{R}^n$ are unobserved errors (possibily correlated); and $\beta \in \mathbb{R}$, $\gamma_1, \xi_1 \in \mathbb{R}^p$, and $\pi \in \mathbb{R}^k$ are unknown parameters. The matrices X and $\widetilde{Z}$ are taken to be fixed (i.e., non-stochastic) and $\overline{Z} = [X : \widetilde{Z}]$ has full column rank $p + k$.

In the first and main part of this dissertation we are interested with one-sided hypothesis testing of the coefficient β :[1]

$$H_0 : \beta = \beta_0 (\text{ or } H_0 : \beta \leq \beta_0) \text{ against } H_1 : \beta > \beta_0 \tag{2.2}$$

and in the last chapter we revise and deal with the two-sided hypothesis testing problem:

$$H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0 \tag{2.3}$$

It is convenient to transform the IV matrix $\widetilde{Z}$ into a matrix Z which is orthogonal to X: $Z'X = 0$. Since we are interested only in β , we can decompose the IV matrix $\widetilde{Z} = Z + P_X\widetilde{Z} = M_X\widetilde{Z} + P_X\widetilde{Z}$, where $M_A = I - P_A$ and $P_A = A(A'A)^{-1}A'$ for any full column matrix A, and work with the model:

$$y_1 = y_2\beta + X\gamma_1 + u, \tag{2.4}$$

$$y_2 = Z\pi + X\xi + v_2, \tag{2.5}$$

where $\xi = \xi_1 + (X'X)^{-1}X'\widetilde{Z}\pi$.

Furthermore, the model can be rewritten as a matricial reduced-form:

$$Y = Z\pi a' + X\eta + V, \tag{2.6}$$

where $Y = [y_1 : y_2]$, $V = [v_1 : v_2] = [u + v_2\beta : v_2]$, $a = (\beta, 1)'$, $\eta = [\gamma : \xi]$, and $\gamma = \gamma_1 + \xi\beta$.

The reduced-form errors V are assumed to be independently and identically distributed (i.i.d) across rows. To obtain exact distribution of the tests, we assume that each row has a mean zero bivariate normal distribution with *known* 2×2 nonsingular covariance matrix $\Omega = [\omega_{ij}]_{i,j=1,2}$. As

[1] The opposite inequality in the null and alternative hypothesis is a straightforward adaptation and we omit the results.

shown below, the normality and the knowledge of Ω assumptions can be relaxed when asymptotic approximations are considered.

The probability model for (2.6) is a member of the curved exponential family, and low dimensional sufficient statistics are available. Lemma 1 of AMS06a shows that $X'Y$ and $Z'Y$ are independent and sufficient for $\left(\gamma', \xi'\right)'$ and $(\beta, \pi')'$, respectively. Since the assessment of the performance of the tests is by its power we can focus only on tests based sufficient statistics, in particular on $Z'Y$. As shown by (19, Moreira (2003)), we can apply a one-to-one transformation to $Z'Y$ that yields the $k \times 2$ sufficient statistic $[S:T]$, where [2]

$$\begin{aligned} S &= (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0' \Omega b_0)^{-1/2} \text{ and} \\ T &= (Z'Z)^{-1/2} Z'Y \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2}, \end{aligned} \tag{2.7}$$

where $b_0 = (1, -\beta_0)'$ and $a_0 = (\beta_0, 1)'$.

The distribution of the sufficient statistic $[S:T]$ is multivariate normal,

$$vec[S:T] \sim N\left(h_\beta \otimes \mu_\pi, I_{2k}\right), \tag{2.8}$$

with first moment depending on the following quantities:

$$h_\beta = (c_\beta, d_\beta)' \in \mathbb{R}^2 \text{ and } \mu_\pi = (Z'Z)^{1/2}\pi \in \mathbb{R}^k, \tag{2.9}$$

where $c_\beta = (\beta - \beta_0) \cdot (b_0' \Omega b_0)^{-1/2}$ and $d_\beta = a' \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2}$.

[2] Henceforth, we use as the matrix square root as the (unique) symmetric square root.

Chapter 3

Invariant Similar Tests

Seems natural to suppose that our decision to reject or not the null hypothesis is invariant to changes in the coordinate system of the instrumental variables, i.e., the order in which each instrument appears, otherwise there is too much a priori information about the instruments and their relative relevance. The only exception to our knowledge that exclude specific instruments and consequently depend on the order in which the instruments appears is given by (12, Donald and Newey (2001)). In consequence we restrict our analysis to tests that are invariant to orthogonal transformations, i.e., let ϕ be a $[0,1]$-valued statistic depending on the sufficient statistics $[S : T]$ and F be a $k \times k$ orthogonal matrix, so the tests considered are such $\phi(FS, FT]) = \phi(S,T)$.[1]

By Theorem 6.2.1 of (17, Lehmann and Romano (2005)) and Theorem 1 of AMS06a, a test is invariant if and only if it can be written as a function of

$$Q = [S:T]'[S:T] = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix} = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix}. \tag{3.1}$$

The statistic Q has a Wishart distribution with rank one that depends on

$$\begin{aligned} \xi_\beta(q) &= h'_\beta q h_\beta = c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T, \text{ where} \\ q &= \begin{bmatrix} q_S & q_{ST} \\ q_{ST} & q_T \end{bmatrix} \in \mathbb{R}^{2\times 2}. \end{aligned} \tag{3.2}$$

Note that $\xi_\beta(q) \geq 0$ because q is positive semi-definite almost surely (a.s.). Define $Q_1 = (Q_S, Q_{ST})$. The density of Q evaluated at (q_1, q_T) is given by

$$\begin{aligned} f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) = K_1 \exp(-\lambda(c_\beta^2 + d_\beta^2)/2) \det(q)^{(k-3)/2} \\ \times \exp(-(q_S + q_T)/2)(\lambda\xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}(\sqrt{\lambda\xi_\beta(q)}), \end{aligned} \tag{3.3}$$

where K_1 is a constant, $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind of order ν, and

$$\lambda = \pi' Z' Z \pi \geq 0. \tag{3.4}$$

Examples of invariant test statistics are the (1, Anderson and Rubin (1949)) (AR), score and likelihood ratio statistics:

$$\begin{aligned} AR &= Q_S/k, \\ LM &= Q_{ST}^2/Q_T, \\ LR &= \frac{1}{2}\left(Q_S - Q_T + \sqrt{(Q_S - Q_T)^2 + 4Q_{ST}^2}\right). \end{aligned} \tag{3.5}$$

[1] (20, Moreira (2009)) shows that the group of transformation on $[S:T]$ is isomorphic to a group of transformations on the original data Y.

When the concentration parameter $\lambda/(\omega_{22}\cdot k)$ is small, most test statistics are not approximately distributed normal or chi-square. For example, under the weak instrument asymptotics of (22, Staiger and Stock (1997)) where $\pi = C/\sqrt{n}$, the LR statistic is not asymptotically pivotal. Its asymptotic distribution is nonstandard and depends on the nuisance and concentration parameter $\lambda/(\omega_{22}\cdot k)$ under the null. Consequently, the null rejection probability of the standard likelihood ratio test depends on the concentration parameter.

(19, Moreira (2003)) proposes similar tests which reject the null hypothesis when the test statistic ψ exceeds a critical value that depends on Q_T:

$$\psi(Q_S, Q_{ST}, Q_T) > \kappa_{\psi,\alpha}(Q_T), \tag{3.6}$$

where $\kappa_{\psi,\alpha}(q_T)$ is the $1-\alpha$ quantile of the distribution of ψ conditional on $Q_T = q_T$ when $\beta = \beta_0$:

$$P_{\beta_0}(\psi(Q_S, Q_{ST}, Q_T)) > \kappa_{\psi,\alpha}(q_T)) = \alpha, \tag{3.7}$$

In practice, the critical value function $\kappa_{\psi,\alpha}(Q_T)$ of the conditional test given in (3.6) is unknown and must be approximated. Given a statistic $\psi(Q_S, Q_{ST}, Q_T)$ write it as a function of Q_S, $\mathcal{S}_2 = Q_{ST}/(||S||\cdot||T||)$ and Q_T, by Lemma 3, (f) of AMS06a, $(Q_S, \mathcal{S}_2)$ is independent of Q_T and has a nuisance-parameter free distribution when $\beta = \beta_0$. The null distribution of $(Q_S, \mathcal{S}_2)$ can be approximated by simulating n_{MC} i.i.d random vectors $S_i \sim N(0, I_k)$ for $i = 1, ..., n_{MC}$ where n_{MC} is large. The approximation to $\kappa_{\psi,\alpha}(Q_T)$ is the $1-\alpha$ sample quantile of $\{\psi(S_i'S_i, S_i'e_1^k \cdot Q_T^{1/2}, Q_T) : i = 1, ..., n_{MC}\}$ and $e_1^k = (1, 0, ..., 0)' \in \mathbb{R}^k$.

We now introduce several new one-sided invariant similar tests for testing $H_0 : \beta = \beta_0$ (or $H_0 : \beta \leq \beta_0$) against $H_1 : \beta > \beta_0$. Each similar test will reject the null hypothesis when the one-sided statistic ψ is larger than the critical value function $\kappa_{\psi,\alpha}$.

The one-sided t-statistics are based on the $\underline{k}$-class estimators of β and its derivation are present in the appendix:[2]

$$\begin{aligned} t(\underline{k}) &= \frac{\beta(\underline{k}) - \beta_0}{\sigma_u(\underline{k})[f_2^2 Q_S + g_2^2 Q_T + 2 f_2 g_2 Q_{ST} - n(\underline{k}-1) w_{22}]^{-1/2}}, \text{ where} \\ \beta(\underline{k}) &= \frac{f_1 f_2 Q_S + g_1 g_2 Q_T + (g_1 f_2 + f_1 g_2) Q_{ST} - n(\underline{k}-1) w_{21}}{f_2^2 Q_S + g_2^2 Q_T + 2 f_2 g_2 Q_{ST} - n(\underline{k}-1) w_{22}}, \\ \sigma_u^2(\underline{k}) &= b(\underline{k})' \Omega b(\underline{k}) \text{ and } b(\underline{k}) = (1, -\beta(\underline{k}))', \end{aligned} \tag{3.8}$$

where $f_l = b_0'\Omega e_l/\sqrt{b_0'\Omega b_0}$, $g_l = a_0' e_l/\sqrt{a_0'\Omega^{-1}a_0}$, $e_1 = (1,0)'$ and $e_2 = (0,1)'$ for $l = 1, 2$.

Here we will use four $\underline{k}$-class: the 2SLS estimator, the limited information maximum likelihood for known Ω (LIMLK) estimator, the bias-adjusted 2SLS (B2SLS) estimator and the Fuller's estimator:

$$\begin{array}{ll} \text{2SLS:} & \underline{k} = 1, \\ \text{LIMLK:} & \underline{k} = \underline{k}_{\,LIMLK} = 1 + (Q_S - LR)/n \\ \text{B2SLS:} & \underline{k} = n/(n-k+2), \\ \text{Fuller:} & \underline{k} = \underline{k}_{\,LIMLK} - 1/(n-k-p). \end{array} \tag{3.9}$$

The finite-sample properties of the estimators $\beta(\underline{k})$ depend on $\underline{k}$. Consequently, the behavior of the $t(\underline{k})$ statistics can be sensitive to the choice of $\underline{k}$.

We construct two statistics from the likelihood of the model given in (2.6) with Ω known. The first one-sided statistic based on the standard LR statistic (i.e., -2 times the logarithm of the likelihood ratio) is for testing $H_0 : \beta = \beta_0$:

$$\begin{aligned} LR1 &= 2\left[\sup_{\beta\geq\beta_0} l_c(Y;\beta,\Omega) - l_c(Y;\beta_0,\Omega)\right] = R(\beta_0) - \inf_{\beta\geq\beta_0} R(\beta), \text{ where} \\ R(\beta) &= \frac{b'Y'P_Z Y b}{b'\Omega b} \text{ with } b = (1, -\beta)', \end{aligned} \tag{3.10}$$

[2] To avoid confusion with the number k of exogenous variables, we use $\underline{k}$ to define Theil's class of estimators rather than the more traditional k.

and $l_c(Y;\beta,\Omega)$ is the log-likelihood function for known Ω with all parameters concentrated out except β. In the Appendix, we show that $R(\beta)$ and $LR1$ depend on the observations only through Q and

$$LR1 = LR \times 1(\beta(\underline{k}_{LIMLK}) \geq \beta_0) + \max\{0, R(\beta_0) - R(\infty)\} \times 1(\beta(\underline{k}_{LIMLK}) < \beta_0), \quad (3.11)$$

where $R(\infty) = \lim_{\beta\to\infty} R(\beta)$. We will see later in the numerical results that the power function $P_{\beta,\lambda}(LR1 > \kappa_{LR1,\alpha}(Q_T))$ is not monotonic for $\beta < \beta_0$. As a result, the CLR1 test will not have correct size when the null hypothesis is $H_0 : \beta \leq \beta_0$.Given that, we present another one-sided statistic based on the likelihood function: the modified LR statistic for testing $H_0 : \beta \leq \beta_0$:

$$MLR1 = 2\left[\sup_{\beta} l_c(Y;\beta,\Omega) - \sup_{\beta\leq\beta_0} l_c(Y;\beta,\Omega)\right] = \inf_{\beta\leq\beta_0} R(\beta) - R(\beta(\underline{k}_{LIMLK})). \quad (3.12)$$

In the Appendix, we show that

$$MLR1 = [LR - \max\{0, R(\beta_0) - R(\infty)\}] \times 1(\beta(\underline{k}_{LIMLK}) \geq \beta_0). \quad (3.13)$$

(For $H_1 : \beta < \beta_0$, the inequalities in (3.11) and (3.13) are reversed.)

Chapter 4

Power Envelopes

In this chapter, we address the question of optimal invariant similar and nonsimilar tests when the IV's may be weak. To evaluate the performance of the novel one-sided conditional tests, we derive the power envelopes for similar and nonsimilar tests. The use of sufficiency and invariance reduces the dimension of the parameters from $1+k+2p$ for $\theta = (\beta, \pi', \xi', \gamma')'$ to just 2 for $(\beta, \lambda)'$. The dimension reduction allows the power envelope to meaningfully assess the performance of our one-sided tests. The envelope we derive here consists of upper bound for power and lower bound for size for either $H_0 : \beta = \beta_0$ or $H_0 : \beta \leq \beta_0$.

4.1 Similar Power Envelope

The following theorem is the main result of this section:

Theorem 1 *Define the statistic*

$$LR_{\beta^*\lambda^*}(Q_1, Q_T) = \frac{f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)}{f_{Q_T}(q_T; \beta^*, \lambda^*) f_{Q_1|Q_T}(q_1|q_T; \beta_0)} = \frac{\varphi_1(q_1, q_T; \beta^*, \lambda^*)}{\varphi_2(q_T; \beta^*, \lambda^*)}, \tag{4.1}$$

where

$$\begin{aligned} \varphi_1(q_1, q_T; \beta, \lambda) &= \exp(-\lambda c_\beta^2/2)(\lambda \xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda \xi_\beta(q)}\right) \text{ and} \\ \varphi_2(q_T; \beta, \lambda) &= \left(\lambda d_\beta^2 q_T\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda d_\beta^2 q_T}\right). \end{aligned} \tag{4.2}$$

Let $\kappa_{\beta^*\lambda^*,\alpha}(Q_T)$ be a shorthand for $\kappa_{LR_{\beta^*\lambda^*},\alpha}(Q_T)$. Then the following hold:
(a) *For (β^*, λ^*) with $\beta^* > \beta_0$, the test that rejects $H_0 : \beta = \beta_0$ when $LR_{\beta^*\lambda^*}(Q_1, Q_T) > \kappa_{\beta^*\lambda^*,\alpha}(Q_T)$ maximizes power over all level α invariant similar tests.*
(b) *For (β^*, λ^*) with $\beta^* < \beta_0$, the test that rejects $H_0 : \beta = \beta_0$ when $LR_{\beta^*\lambda^*}(Q_1, Q_T) < \kappa_{\beta^*\lambda^*,1-\alpha}(Q_T)$ minimizes the null rejection probability over all level α invariant similar tests.*

Comments: 1. We denote the test that rejects the null when $LR_{\beta^*\lambda^*}(Q_1, Q_T) > \kappa_{\beta^*\lambda^*,\alpha}(Q_T)$ as a point-optimal invariant similar (POIS) test. We determine the power upper bound by considering the POIS tests for arbitrary values (β^*, λ^*) when $\beta^* > \beta_0$. The power upper bound is for similar tests for $H_0 : \beta = \beta_0$. We do not impose the additional constraint that tests must have correct size, and so the upper bound could be conservative for $H_0 : \beta \leq \beta_0$. We shall see later that even for small values of λ, some tests for $H_0 : \beta \leq \beta_0$ do reach the upper bound.

2. The test which rejects the null when $LR_{\beta^*\lambda^*}(Q_1, Q_T) < \kappa_{\beta^*\lambda^*,1-\alpha}(Q_T)$ is called POIS0 test. We determine the null lower bound by finding the power of POIS0 tests for arbitrary values (β^*, λ^*) when $\beta^* < \beta_0$.

3. The power envelope is the union of the power upper bound and null lower bound. Both bounds are relevant because we would like to compare the probability of making the type I and type II errors for different tests.

4. The denominator $\varphi_2(q_T; \beta^*, \lambda^*)$ does not depend on q_1 and can be absorbed into the conditional critical value. Thus, the test based on $LR_{\beta^*\lambda^*}(Q_1, Q_T)$ is equivalent to a test based on the numerator of $\varphi_1(q_1, q_T; \beta^*, \lambda^*)$. For reasons of numerical stability, however, we recommend constructing critical values using $\ln(LR_{\beta^*\lambda^*}(Q_1, Q_T))$.

We now show that such tests do not depend on λ^*, so that the POIS and POIS0 tests are of a relatively simple form. Using a series expansion of $I_{(k-2)/2}(x)$, we can write

$$\varphi_1(q_1, q_T; \beta, \lambda) = 2^{-(k-2)/2} \exp(-\lambda c_\beta^2/2) \sum_{j=0}^{\infty} \frac{(\lambda \xi_\beta(q_1, q_T)/4)^j}{j!\Gamma((k-2)/2 + j + 1)} \tag{4.3}$$

The term $\varphi_2(q_T; \beta, \lambda)$ can be written analogously.

The function $\varphi_1(q_1, q_T; \beta, \lambda)$ is increasing in $\xi_\beta(q_1, q_T) \geq 0$. As a result, for a fixed value of β, say $\beta^* > \beta_0$, the optimal test for fixed alternative β^* rejects $H_0 : \beta = \beta_0$ when

$$\xi_{\beta^*}(Q_1, Q_T) > \kappa_{\beta^*,\alpha}(Q_T), \tag{4.4}$$

where $\kappa_{\beta^*,\alpha}(Q_T)$ is a shorthand for $\kappa_{\xi_{\beta^*},\alpha}(Q_T)$ as defined in (3.7). This POIS test is *one-sided* because it directs power at a single point β^* that is greater than the null value β_0. An analogous argument shows that the POIS0 test that minimizes rejection probabilities for fixed $\beta^* < \beta_0$ rejects H_0 when

$$\xi_{\beta^*}(Q_1, Q_T) < \kappa_{\beta^*,1-\alpha}(Q_T). \tag{4.5}$$

Corollary 2 *For $\beta^* > \beta_0$, the POIS test based on $\xi_{\beta^*}(Q_1, Q_T)$ is the uniformly most powerful test among invariant similar tests against the alternative distributions indexed by $\{(\beta^*, \lambda) : \lambda > 0\}$. For $\beta^* < \beta_0$, the POIS0 test based on $\xi_{\beta^*}(Q_1, Q_T)$ uniformly minimizes the null rejection probability among invariant similar tests against the alternative distributions indexed by $\{(\beta^*, \lambda) : \lambda > 0\}$.*

Comments: 1. Although the form of the POIS and POIS0 tests does not depend on λ^*, their power depends on the true value of λ. Hence, the power envelope depends on both parameters β and λ.

2. A test based on $\xi_{\beta^*}(Q_1, Q_T)$ is equivalent to a test that rejects when

$$\begin{aligned} POIS1_\delta &= \frac{Q_S + \delta \mathcal{S}_2 \sqrt{Q_S} - k}{\sqrt{2k + \delta^2}} > \kappa_{\delta,\alpha}(Q_T), \text{ where} \\ \delta &= (2d_{\beta^*}/c_{\beta^*})\sqrt{Q_T,}), \end{aligned} \tag{4.6}$$

and $\kappa_{\delta,\alpha}(Q_T)$ is a shorthand for $\kappa_{POIS1_\delta,\alpha}(Q_T)$ defined in (3.7). This formulation of the test is convenient because Q_S, $\mathcal{S}_2$, and Q_T are independent under $\beta = \beta_0$, which simplifies the calculation of critical values.

3. Provided $\omega_{12} - \omega_{22}\beta_0 \neq 0$, the quantity d_{β^*} is a linear function of β^* and equals zero if and only if $\beta^* = \beta_{AR}$, where

$$\beta_{AR} = \frac{\omega_{11} - \omega_{12}\beta_0}{\omega_{12} - \omega_{22}\beta_0}. \tag{4.7}$$

In this case, $\delta = 0$ and $POIS1_\delta$ reduces to $Q_S/\sqrt{2k}$, which is the AR statistic re-scaled. Hence, the AR test, usually conceived as a two-sided test, is one-sided POIS against the alternative $\beta = \beta_{AR}$.

4. The POIS test for β^* local to β_0 with $\beta^* > \beta_0$ (i.e., the LMPI test) is the one-sided LM test that rejects H_0 if

$$LM1 = Q_{ST}/Q_T^{1/2} > z_\alpha, \tag{4.8}$$

where z_α is the $1-\alpha$ quantile of the standard normal distribution. Analogously, if β^* is local to β_0 with $\beta^* < \beta_0$, then the LMPI test rejects H_0 if $-Q_{ST}/Q_T^{1/2} > z_\alpha$.

5. The sign of δ in (4.6) can change as β^* changes even for β^* values on the same side of the null hypothesis because d_{β^*} is a linear function of β^*. As a result, the form of the $POIS1_\delta$ statistic (and the power envelope) changes dramatically as β^* varies. The constant δ determines the weight put on the statistic $\mathcal{S}_2$. The optimal value of δ for small values of $\beta > \beta_0$ has the wrong sign for large values of β and vice versa. This fact has adverse consequences for the overall one-sided power properties of POIS tests.

6. The optimal one-sided test for β^* arbitrarily large rejects H_0 if

$$Q_S + 2(\det(\Omega))^{-1/2}(\beta_0\omega_{22} - \omega_{12})Q_{ST} > \kappa_{\infty,\alpha}(Q_T) \tag{4.9}$$

for $\kappa_{\infty,\alpha}(\cdot)$ as defined in (3.7). Remarkably, the same test is the optimal one-sided test for β^* negative and arbitrarily large in absolute value for any λ^*. Consequently, the optimal two-sided test for $|\beta^* - \beta_0|$ arbitrarily large is the test in (4.9).

Corollary 2 shows that the POIS test for an alternative (β^*, λ^*) depends only on β^*. Because the true parameter β is unknown, we could construct an empirical version of the standardized optimal statistic:

$$\widetilde{\xi}_{\widehat{\beta}} = x'_{\widehat{\beta}} Q x_{\widehat{\beta}}, \tag{4.10}$$

where $x_{\widehat{\beta}} = (c_{\widehat{\beta}}/||h_{\widehat{\beta}}||, d_{\widehat{\beta}}/||h_{\widehat{\beta}}||)'$ and $\widehat{\beta}$ is the maximum likelihood estimator of β under $H_1 : \beta > \beta_0$. The next theorem shows that the empirical POIS test is equivalent to those based on $CLR1$ test.

Theorem 3 *The statistics $\widetilde{\xi}_{\widehat{\beta}}$ and $LR1$ are equivalent up to strictly increasing transformations (possibly depending on Q_T). In particular,*

$$P_{\beta,\lambda}(\widetilde{\xi}_{\widehat{\beta}} > \kappa_{\widetilde{\xi}_{\widehat{\beta}},\alpha}(Q_T)) = P_{\beta,\lambda}\left(LR1 > \kappa_{LR1,\alpha}(Q_T)\right). \tag{4.11}$$

Comment: This theorem and Comment 5 of Corollary 2 indicate that the CLR1 test does not have correct size when the null hypothesis is $H_0 : \beta \leq \beta_0$ instead of $H_0 : \beta = \beta_0$. See chapter 7 below for numerical simulations on size and power of the CLR1 test.

4.2 Non-Similar Power Envelope

Non-similar tests have null rejection probability below the significance level for some values of the nuisance parameter λ. Due to the continuity of the power function, for such values of λ, the power of a non-similar test is less than the power of a similar test for alternatives close enough to the null hypothesis. However, for other values of λ, or for more distant alternatives, non-similar tests can have greater power than similar tests. For this reason, we also consider optimal invariant non-similar tests of the hypothesis $H_0 : \beta = \beta_0$ against point alternatives.

Our construction of point-optimal invariant (POI) non-similar tests follows Section 3.8 of (17, Lehmann and Romano (2005)). Consider the composite null hypothesis

$$H_0 : (\beta, \lambda) \in \{(\beta_0, \lambda) : 0 \leq \lambda < \infty\}, \tag{4.12}$$

and the point alternative

$$H_1 : (\beta, \lambda) = (\beta^*, \lambda^*). \tag{4.13}$$

Let Λ be a probability measure over $\{\lambda : 0 \leq \lambda < \infty\}$ and h_Λ be the weighted pdf,

$$h_\Lambda(q) = \int f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) d\Lambda(\lambda), \tag{4.14}$$

where $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ is given in (3.3). The effect of weighting by Λ under the null is to turn the composite null into a point null, so that the most powerful test can be obtained using the Neyman-Pearson Lemma. Specifically, let ϕ_Λ be the most powerful test of h_Λ against $f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)$, so that ϕ_Λ rejects the null when

$$NP_\Lambda(q) = \frac{f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)}{h_\Lambda(q)} > d_{\Lambda,\alpha}, \tag{4.15}$$

where $d_{\Lambda,\alpha}$ is the critical value of the test, chosen so that $NP_\Lambda(q)$ rejects the null with probability α under the distribution h_Λ.

If the test ϕ_Λ has size α for the null hypothesis H_0 in (4.12), i.e.,

$$\sup_{0 \le \lambda < \infty} P_{\beta_0,\lambda}(NP_\Lambda(Q) > d_{\Lambda,\alpha}) = \alpha, \tag{4.16}$$

then the test ϕ_Λ is most powerful for testing H_0 against H_1, and the distribution Λ is least favorable; cf. Thm. 3.8.1 and Cor. 3.8.1 of (17, Lehmann and Romano (2005)).

Given a distribution Λ, condition (4.16) is easily checked numerically. What proves more computationally difficult is finding the distribution that satisfies (4.16). In the numerical work we consider distributions Λ that put point mass on some point λ_0. In this case, we have

$$NP_\Lambda = \frac{f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)}{f_{Q_1,Q_T}(q_1, q_T; \beta_0, \lambda_0)} \tag{4.17}$$

Let $\mathcal{R}(\beta_0, \lambda_0, \beta^*, \lambda^* | \beta, \lambda)$ be the rejection probability of the test based on the statistic in (4.17) when the true values are β and λ. The numerical problem is to find the value of λ_0 such that the test has size α. Denote this value of λ_0 by λ_0^{LF}; then λ_0^{LF} solves

$$\begin{aligned} \mathcal{R}(\beta_0, \lambda_0^{LF}, \beta^*, \lambda^* | \beta_0, \lambda_0^{LF}) &= \alpha \text{ and} \\ \sup_{0 \le \lambda < \infty} \mathcal{R}(\beta_0, \lambda_0^{LF}, \beta^*, \lambda^* | \beta_0, \lambda) &\le \alpha. \end{aligned} \tag{4.18}$$

If there is a $\lambda_0^{LF}(\beta_0, \beta^*, \lambda^*)$ that satisfies (4.18), then the test based on $NP_{\lambda_0^{LF}}$ is the POI non-similar test.

The power upper bound for invariant non-similar tests is $\mathcal{R}(\beta_0, \lambda_0^{LF}(\beta_0, \beta^*, \lambda^*), \beta^*, \lambda^* | \beta^*, \lambda^*)$ (an analogous argument yields a null lower bound). We find numerically that the power envelopes for similar and non-similar tests are essentially the same, up to numerical accuracy. The reason for this is twofold. On one hand, the conditional critical values for the POIS tests depend on q_T only weakly in the range of q_T that is most likely to occur under the alternative. Thus, the POIS tests are very nearly unconditional. On the other hand, the POI non-similar tests have null rejection rates that are very nearly equal to α for all values of λ; thus, the POI non-similar tests are very nearly similar. Because POI similar tests are nearly unconditional and the POI non-similar tests are nearly similar, the two types of tests have nearly the same rejection regions. An analogous result is described by (7, Andrews, Moreira and Stock (2008)) for two-sided testing.

Chapter 5

Weak IV Asymptotics

Here, we consider the same model and hypotheses as in chapter 2, but with non-normal reduced-form errors with unknown covariance matrix. We show that the finite-sample distribution of the tests and statistics holds asymptotically under the same high-level assumptions as in (21, Staiger and Stock (1997)). To model weak IV asymptotics and fixed alternatives (WIV-FA), we let π be local to zero and the alternative β be fixed, not local to the null value β_0:

Assumption WIV-FA. (a) $\pi = C/n^{1/2}$ for some non-stochastic k-vector C.

(b) β is a fixed constant for all $n \geq 1$.

(c) k is a fixed positive integer that does not depend on n.

We now specify the asymptotic behavior of the instruments, exogenous regressors, and reduced-form errors.

Assumption 1. $n^{-1}\overline{Z}'\overline{Z} \rightarrow_p D$ for some positive definite $(k+p) \times (k+p)$ matrix D.

Assumption 2. $n^{-1}V'V \rightarrow_p \Omega$ for some positive definite 2×2 matrix Ω.

Assumption 3. $n^{-1/2}vec(\overline{Z}'V) \rightarrow_d N(0, \Phi)$ for some positive definite $2(k+p) \times 2(k+p)$ matrix Φ, where $vec(\cdot)$ denotes the column by column vec operator.

Assumption 4. $\Phi = \Omega \otimes D$.

The quantities C, D, and Ω are assumed to be unknown. (3, Andrews, Moreira and Stock (2004)) (hereinafter, AMS04) show that Assumptions 1-3 hold under general conditions. Assumption 4 holds under Assumptions 1-3 and homoskedasticity of the errors V_i, i.e., $E(V_iV_i'|\overline{Z}_i) = EV_iV_i' = \Omega$ $a.s.$

We now introduce tests that are suitable for (possibly) non-normal, homoskedastic, uncorrelated errors with unknown covariance matrix. See AMS04 for tests and results for cases in which the errors are not homoskedastic or are correlated. For clarity of the asymptotics results, we write S, T, Q, etc. of chapter 2, as S_n, T_n, Q_n, etc.

5.1 Tests for Unknown Ω and Possibly Non-normal Errors

For feasible tests, when the reduced-form error covariance matrix Ω is unknown, we need to estimate consistently Ω and it is achieved by the estimator:[1]

$$\widehat{\Omega}_n = (n-k-p)^{-1}\widehat{V}'\widehat{V}, \text{ where } \widehat{V} = M_{Z,X}Y = Y - P_ZY - P_XY, \tag{5.1}$$

see Lemma 1 of (5, Andrews, Moreira and Stock (2006,b)) (hereinafter, AMS06b).

[1] This definition of $\widehat{\Omega}_n$ is suitable if Z or X contains a vector of ones, as is usually the case. If not, then $\widehat{\Omega}_n$ is defined with the sample mean of $\widehat{V}$ subtracted from it.

Given that, we can replace Ω by $\widehat{\Omega}_n$ and obtain modified versions of the statistics S_n, T_n, $Q_{S,n}$, $Q_{ST,n}$ and $Q_{T,n}$:

$$\begin{aligned} \widehat{S}_n &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2}, \\ \widehat{T}_n &= (Z'Z)^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0 \cdot (a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2}, \\ \widehat{Q}_{S,n} &= \widehat{S}_n'\widehat{S}_n,\ \widehat{Q}_{ST,n} = \widehat{S}_n'\widehat{T}_n \text{ and } \widehat{Q}_{T,n} = \widehat{T}_n'\widehat{T}_n \end{aligned} \tag{5.2}$$

The feasible one-sided statistics for Ω unknown are derived in the appendix and we present here:

$$\begin{aligned} \widehat{t}\left(\underline{\widehat{k}}\right)_n &= \frac{\widehat{\beta}(\underline{\widehat{k}}) - \beta_0}{\widehat{\sigma}_u(\underline{\widehat{k}})[\widehat{f}_2^2\widehat{Q}_{S,n} + \widehat{g}_2^2\widehat{Q}_{T,n} + 2\widehat{f}_2\widehat{g}_2\widehat{Q}_{ST,n} - n(\underline{\widehat{k}}-1)\widehat{w}_{22}]^{-1/2}}, \text{ where} \\ \widehat{\beta}(\underline{\widehat{k}}) &= \frac{\widehat{f}_1\widehat{f}_2\widehat{Q}_{S,n} + \widehat{g}_1\widehat{g}_2\widehat{Q}_{T,n} + (\widehat{g}_1\widehat{f}_2 + \widehat{f}_1\widehat{g}_2)\widehat{Q}_{ST,n} - n(\underline{\widehat{k}}-1)\widehat{w}_{21}}{\widehat{f}_2^2\widehat{Q}_{S,n} + \widehat{g}_2^2\widehat{Q}_{T,n} + 2\widehat{f}_2\widehat{g}_2\widehat{Q}_{ST,n} - n(\underline{\widehat{k}}-1)\widehat{w}_{22}}, \text{ and} \\ \widehat{\sigma}_u^2\left(\underline{\widehat{k}}\right) &= \left(1, -\widehat{\beta}(\underline{\widehat{k}})\right)\widehat{\Omega}_n\left(1, -\widehat{\beta}(\underline{\widehat{k}})\right)'. \end{aligned}$$

where $\widehat{f}_l = b_0'\widehat{\Omega}_n e_l/\sqrt{b_0'\widehat{\Omega}_n b_0}$ and $\widehat{g}_l = a_0'e_l/\sqrt{a_0'\widehat{\Omega}_n^{-1}a_0}$ for $l = 1, 2$, and the $\underline{k}$'s for the four k-class estimators are given by:

$$\begin{array}{ll} \text{2SLS:} & \underline{\widehat{k}} = 1 \\ \text{LIML:} & \underline{\widehat{k}} = \underline{\widehat{k}}_{\,LIML} = 1 + (\widehat{Q}_{S,n} - \widehat{LR}_n)/(n-k-p) \\ \text{B2SLS:} & \underline{\widehat{k}} = n/\left(n-k+2\right) \\ \text{Fuller:} & \underline{\widehat{k}} = \underline{\widehat{k}}_{\,LIML} - 1/\left(n-k-p\right). \end{array} \tag{5.3}$$

For all remaining test statistics, we just need to replace Q_S, Q_{ST} and Q_T by their analogues in which Ω is estimated by $\widehat{\Omega}_n$. For example, the LR test statistic for unknown Ω is defined as in (3.5), but with Q_S, Q_{ST} and Q_T replaced by $\widehat{Q}_{S,n}$,$\widehat{Q}_{ST,n}$ and $\widehat{Q}_{T,n}$. We denote the resulting test statistic by $\widehat{LR}_n$. The analogue of $LR1$, $MLR1$ and $LM1$ are denoted by $\widehat{LR1}_n$, $\widehat{MLR1}_n$ and $\widehat{LM1}_n$, respectively.

The critical value function for each test statistic ψ is simply $\kappa_{\psi,\alpha}(\widehat{Q}_{T,n})$, as defined in (3.7).

5.2 Weak IV Asymptotic Results

In this section we derive the limit distribution of each one-sided test present in the previous section. In particular, we show that all tests converges in distribution to their respective finite sample distribution under normality.

Under Assumptions WIV-FA and 1-4, Lemma 4 of AMS06a shows that

$$\begin{aligned} &(S_n, T_n) \to_d (S_\infty, T_\infty), \\ &(\widehat{S}_n, \widehat{T}_n, \widehat{\Omega}_n) - (S_n, T_n, \Omega) \to_p 0, \\ &(\widehat{S}_n, \widehat{T}_n, \widehat{\Omega}_n) \to_d (S_\infty, T_\infty, \Omega), \end{aligned} \tag{5.4}$$

where S_∞ and T_∞ are independent k-vectors which are defined as follows:

$$\begin{aligned} vec(N_Z) &\sim N(vec(D_Z Ca'), \Omega \otimes D_Z), \\ S_\infty &= D_Z^{-1/2}N_Z b_0 \cdot (b_0'\Omega b_0)^{-1/2} \sim N(c_\beta D_Z^{1/2}C, I_k), \\ T_\infty &= D_Z^{-1/2}N_Z\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \sim N(d_\beta D_Z^{1/2}C, I_k), \text{ where} \\ D_Z &= D_{11} - D_{12}D_{22}^{-1}D_{21}, \\ D &= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \ D_{11} \in \mathbb{R}^{k\times k},\ D_{12} \in \mathbb{R}^{k\times p}, \text{ and } D_{22} \in \mathbb{R}^{p\times p}. \end{aligned} \tag{5.5}$$

The matrix D_Z is the probability limit of $n^{-1}Z'Z$. Under $H_0 : \beta = \beta_0$, S_∞ has mean zero, but T_∞ does not. Let

$$\begin{aligned} Q_\infty &= [S_\infty : T_\infty]'[S_\infty : T_\infty], \\ Q_{S,\infty} &= S'_\infty S_\infty, Q_{ST,\infty} = S'_\infty T_\infty,\ Q_{T,\infty} = T'_\infty T_\infty, \\ \mathcal{S}_{2,\infty} &= S'_\infty T_\infty/(||S_\infty|| \cdot ||T_\infty||),\ \text{and} \\ \lambda_\infty &= C'D_Z C. \end{aligned} \tag{5.6}$$

By (5.5), we find that the finite-sample distribution of $(Q_{S,\infty}, Q_{ST,\infty}, Q_{T,\infty})$ is the same as that of $(Q_{S,n}, Q_{ST,n}, Q_{T,n})$ with λ_n replaced by λ_∞. Then the asymptotic distribution of the feasible tests and statistics are the same as finite-sample case:

Theorem 4 *Under Assumptions* WIV-FA *and* 1-4 *with* $\widehat{\underline{k}}$ *defined in (5.3)*
(a) (i)$(\widehat{t}(\widehat{\underline{k}})_n, \kappa_{t(\underline{k}_\infty),\alpha}(\widehat{Q}_{T,n})) \to_d (t(\underline{k}_\infty), \kappa_{t(\underline{k}_\infty),\alpha}(Q_{T,\infty}))$,
(ii) $(\widehat{LR1}_n, \kappa_{LR1,\alpha}(\widehat{Q}_{T,n})) \to_d (LR1(Q_{1,\infty}, Q_{T,\infty}), \kappa_{LR1,\alpha}(Q_{T,\infty}))$,
(iii)$(\widehat{MLR1}_n, \kappa_{MLR1,\alpha}(\widehat{Q}_{T,n})) \to_d (MLR1(Q_{1,\infty}, Q_{T,\infty}), \kappa_{MLR1,\alpha}(Q_{T,\infty}))$
(iv)$(\widehat{LM1}_n, \kappa_{LM1,\alpha}(\widehat{Q}_{T,n})) \to_d (LM1(Q_{1,\infty}, Q_{T,\infty}), \kappa_{LM1,\alpha}(Q_{T,\infty}))$
(b) (i) $P(\widehat{t}(\widehat{\underline{k}})_n > \kappa_{t(\underline{k}_\infty),\alpha}(\widehat{Q}_{T,n})) \to P(t(\underline{k}_\infty) > \kappa_{t(\underline{k}_\infty),\alpha}(Q_{T,\infty}))$,
(ii) $P(\widehat{LR1}_n > \kappa_{LR1,\alpha}(\widehat{Q}_{T,n})) \to P(LR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{LR1,\alpha}(Q_{T,\infty}))$,
(iii) $P(\widehat{MLR1}_n > \kappa_{MLR1,\alpha}(\widehat{Q}_{T,n})) \to P(MLR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{MLR1,\alpha}(Q_{T,\infty}))$,
(iv) $P(\widehat{LM1}_n > \kappa_{LM1,\alpha}(\widehat{Q}_{T,n})) \to P(LM1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{LM1,\alpha}(Q_{T,\infty}))$,
(c) $P(t(\underline{k}_\infty) > \kappa_{t(\underline{k}_\infty),\alpha}(Q_{T,\infty})) = P(LR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{LR1,\alpha}(Q_{T,\infty})) = P(MLR1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{MLR1,\alpha}(Q_{T,\infty})) = P(LM1(Q_{1,\infty}, Q_{T,\infty}) > \kappa_{LM1,\alpha}(Q_{T,\infty})) = \alpha$ *under* $\beta = \beta_0$.

Comments. 1. The $\underline{k}'_\infty$s are the limiting distribution of the statistics $\underline{k}$ and for all the four $\underline{k}$-class considered $\underline{k}_\infty = 1$. And the t-statistics $t(\underline{k}_\infty)$ are the limiting distribution of $t(\underline{k})_n$, which are functions of Q_∞ given in (5.6).

2. Part (c) asserts that conditional tests derived from the $\widehat{LR1}_n$ and $\widehat{MLR1}_n$ statistics and on the t-statistics based on the 2SLS, LIML, B2SLS, and Fuller estimators are asymptotically similar at level α under WIV-FA.

Chapter 6

Strong IV Asymptotics

Two important large sample properties of tests are *pointwise consistency*: for a fixed alternative, the power of the test goes to one as the sample size increases; and *asymptotic efficiency* (AE): the test uniformly maximize the asymptotic power among the asymptotically unbiased tests (see (17, Lehmann and Romano (2005))). Given that, we analyze the asymptotic properties of the conditional tests based on $\widehat{LM1}_n$, $\widehat{LR1}_n$, $\widehat{MLR1}_n$, and $\widehat{t}(\widehat{k})_n$ statistics for both local alternatives (SIV-LA) and fixed alternatives (SIV-FA). Under SIV-LA, we establish AE for the one-sided tests and under SIV-FA, we address consistency.

Before analyzing the asymptotic properties of the one-sided conditional tests, we need to establish the limit behaviour of the critical value functions under SIV-LA and SIV-FA. Under both asymptotics, SIV-LA and SIV-FA, $Q_{T,n}$ diverges in probability to ∞ as the sample size increases. Given that we provide the following preliminary results:

Lemma 5 *Let z_α be the $1-\alpha$ quantile of the standard normal distribution. As $q_T \to_p \infty$,*
(a) *for any $\underline{k}$ in (3.9), $\kappa_{t(\underline{k}),\alpha}(q_T) \to z_\alpha$,*
(b) *for $\alpha \in (0, 1/2)$, $\kappa_{LR1,\alpha}(q_T) \to z_\alpha$.*
(c) *for $\alpha \in (0, 1/2)$, $\kappa_{MLR1,\alpha}(q_T) \to z_\alpha$.*

Comment. As the sample size increases the statistic $Q_{T,n}$ diverges to infinity in probability which implies, by the previous result, that the critical value functions of the tests considered converge in probability to the $1-\alpha$ quantile of the standard normal distribution.

6.1 Local Alternatives

For local alternatives, β is local to the null value β_0 as the sample size increases:

Assumption SIV-LA. (a) $\beta = \beta_0 + B/n^{1/2}$ for some constant $B > 0$.
(b) π is a fixed non-zero k-vector for all $n \geq 1$.
(c) k is a fixed positive integer that does not depend on n.

We use Lemma 6 of AMS06a to establish the strong IV-local alternative limiting distribution of tests. Under Assumptions SIV-LA and 1-4:

$$\begin{aligned}
&(S_n, T_n/n^{1/2}) \to_d (S_{B\infty}, \alpha_T), \\
&(\widehat{S}_n, \widehat{T}_n/n^{1/2}, \widehat{\Omega}_n) - (S_n, T_n/n^{1/2}, \Omega) \to_p 0, \\
&(\widehat{S}_n, \widehat{T}_n/n^{1/2}, \widehat{\Omega}_n) \to_d (S_{B\infty}, \alpha_T, \Omega),
\end{aligned} \tag{6.1}$$

where $S_{B\infty}$ and α_T are k-vectors defined as follows:

$$\begin{aligned}
S_{B\infty} &\sim N(\alpha_S, I_k), \text{ where} \\
\alpha_S &= D_Z^{1/2}\pi B(b_0'\Omega b_0)^{-1/2} \text{ and} \\
\alpha_T &= D_Z^{1/2}\pi(a_0'\Omega^{-1}a_0)^{1/2}.
\end{aligned} \tag{6.2}$$

These definitions allow us to determine the behavior of the $\widehat{LM1}_n$, $\widehat{LR1}_n$, $\widehat{MLR1}_n$, and $\widehat{t}(\widehat{k})_n$ statistics under SIV-LA asymptotics.

Theorem 6 *Under Assumptions* SIV-LA *and* 1-4 *with* $\widehat{\underline{k}}$ *defined in (5.3):*
(a) $\widehat{t}(\widehat{\underline{k}}) = t(\underline{k}) + o_p(1) \to_d (\alpha_T' S_{B\infty})/||\alpha_T||$.
(b) $\widehat{LM1}_n = LM1_n + o_p(1) \to_d (\alpha_T' S_{B\infty})/||\alpha_T||$,
(c) $\widehat{LR1}_n^{1/2} = LR1_n^{1/2} + o_p(1) \to_d \max\{(\alpha_T' S_{B\infty})/||\alpha_T||, 0\}$,
(d) $\widehat{MLR1}_n^{1/2} = MLR1_n^{1/2} + o_p(1) \to_d \max\{(\alpha_T' S_{B\infty})/||\alpha_T||, 0\}$,

Comments. 1. Since the $\underline{k}$-class estimators considered satisfies $\widehat{\underline{k}} = \underline{k} + O_p(n^{-1}) = 1 + O_p\left(n^{-1}\right)$ we can replacement $\widehat{\underline{k}}$ by $\underline{k}$ such that the t-statistics do not suffer any asymptotic effect.

Together with Lemma 5, Theorem 6 yields the following optimality result for a sequence of experiments under SIV-LA and i.i.d normal errors with unknown covariance matrix Ω. Under SIV-LA, the curvature of the model (2.6) vanishes asymptotically and standard local asymptotically normal (LAN) likelihood ratio theory is applicable. For one-sided alternatives, the usual one-sided LM test is AE under the SIV-LA asymptotics and i.i.d normal errors. The others one-sided tests that we propose are also AE.

Theorem 7 *Suppose Assumptions SIV-LA and* 1 *hold and the reduced-form errors* $\{V_i : i \geq 1\}$ *are i.i.d normal, independent of* $\{\overline{Z}_i : i \geq 1\}$ *with mean zero and p.d variance matrix* Ω *which may be known or unknown. Then the score test based on* $\widehat{LM1}_n$ *and the conditional tests based on* $\widehat{t}\,(\underline{k})_n$ *are one-sided AE. If* $\alpha \in (0, 1/2)$*, the conditional tests based on* $\widehat{LR1}_n^{1/2}$ *and* $\widehat{MLR1}_n^{1/2}$ *are also AE.*

6.2 Fixed Alternatives

We now analyze properties of the tests under strong IV fixed alternative (SIV-FA) asymptotics. This asymptotic framework is novel in the weak-instrument literature and determines the *pointwise consistency* of tests.

Assumption SIV-FA. (a) $\beta \neq \beta_0$ is a fixed scalar for all $n \geq 1$.
(b) π is a fixed non-zero k-vector for all $n \geq 1$.
(c) k is a fixed positive integer that does not depend on n.

The strong IV-fixed alternative (SIV-FA) asymptotic behavior of tests depends on the random vector $\varsigma_k \sim N(0, I_k)$ and

$$\lambda_{FA} = \pi' D_Z \pi, \tag{6.3}$$

where D_Z is defined in (5.5).

Lemma 8 *Under Assumptions* SIV-FA *and* 1-3,
(i) $(S_n/n^{1/2}, T_n/n^{1/2}) \to_p (c_\beta D_Z^{1/2}\pi,\ d_\beta D_Z^{1/2}\pi)$,
(ii) $(\widehat{S}_n/n^{1/2}, \widehat{T}_n/n^{1/2}) - (S_n/n^{1/2}, T_n/n^{1/2}) \to_p 0$, *and*
(iii) *if* $\beta = \beta_{AR}$ *and Assumption* 4 *holds, then* $T_n \to_d \varsigma_k$ *and* $\widehat{T}_n - T_n \to_p 0$.

Lemma 8 allows us to determine the limiting behavior of the one-sided conditional tests based on $\widehat{t}\left(\widehat{\underline{k}}\right)_n$, $\widehat{LR1}_n$, $\widehat{MLR1}_n$ and $\widehat{LM1}_n$ statistics under SIV-FA asymptotics.

Theorem 9 *Under Assumptions SIV-FA and* 1-3 *with* $\widehat{\underline{k}}$ *defined in (5.3)*
(a) $\widehat{t}(\widehat{\underline{k}})/n^{1/2} = t(\underline{k})/n^{1/2} + o_p(1) \to_p c_\beta \lambda_{FA}^{1/2} \times (b_0'\Omega b_0/b'\Omega b)^{1/2}$.
(b) *if* $\beta \neq \beta_{AR}$*, then* $\widehat{LM1}_n/n^{1/2} = LM1_n/n^{1/2} + o_p(1) \to_p c_\beta \lambda_{FA}^{1/2}$,

(c) *if* $\beta = \beta_{AR}$ *and Assumption* 4 *holds, then* $\widehat{LM1}_n/n^{1/2} = LM1_n/n^{1/2} + o_p(1) \to_d c_\beta \pi' D_Z^{1/2} \varsigma_k/||\varsigma_k||$,
(d) *if* $\beta > \beta_0$, $\widehat{LR1}_n^{1/2}/n^{1/2} = LR1_n^{1/2}/n^{1/2} + o_p(1) \to_p c_\beta \lambda_{FA}^{1/2}$,
(e) *if* $\beta < \beta_0$, $\widehat{LR1}_n^{1/2}/n^{1/2} = LR1_n^{1/2}/n^{1/2} + o_p(1) \to_p \sqrt{max\left(c_\beta^2 - \omega_{22}^{-1}, 0\right)}\lambda_{FA}^{1/2}$,
(f) *if* $\beta > \beta_0$, $\widehat{MLR1}_n^{1/2}/n^{1/2} = MLR1_n^{1/2}/n^{1/2} + o_p(1) \to_p \sqrt{min\left(c_\beta^2, \omega_{22}^{-1}\right)}\lambda_{FA}^{1/2}$,
(g) *if* $\beta < \beta_0$, $\widehat{MLR1}_n^{1/2}/n^{1/2} = MLR1_n^{1/2}/n^{1/2} + o_p(1) \to_p 0$,

Comments. 1. When $\beta \neq \beta_{AR}$, the critical values of the conditional tests are either constants or converge in probability to constants as $n \to \infty$ (see the comments following Lemma 5). When $\beta = \beta_{AR}$, the critical value functions of these tests (for each β_0) are bounded. Therefore, this theorem addresses the consistency of each test.

2. The one-sided LM test rejects the null when $\widehat{LM1}_n/n^{1/2} > z_\alpha/n^{1/2}$. Because $z_\alpha/n^{1/2}$ converges to zero and the probability of $c_\beta \pi' D_Z^{1/2} \varsigma_k/||\varsigma_k||$ being smaller than zero equals 50%, the LM1 test is *not* consistent at $\beta = \beta_{AR} > \beta_0$.

3. Part (d) shows that the $CLR1$ test is consistent against any alternative $\beta > \beta_0$ for the null hypothesis $H_0 : \beta = \beta_0$. Part (e) shows that the $CLR1$ test asymptotically rejects the null with probability one for some value of $\beta < \beta_0$. Hence, the $CLR1$ test has asymptotic size equal to one once we augment the null hypothesis to $H_0 : \beta \leq \beta_0$.

4. Parts (a), (f) and (g) establish consistency of the conditional tests based on the $MLR1$ statistic and t-statistics whether $H_0 : \beta = \beta_0$ or $H_0 : \beta \leq \beta_0$.

Chapter 7

Numerical Results

This chapter reports numerical results for power envelopes and comparative powers of tests developed earlier for known Ω and normal errors. By transforming variables and parameters in the model (2.6), we can without loss of generality set $\beta_0 = 0$ and the reduced-form errors to be normal with unit variances and correlation ρ.[11] Without loss of generality, no X matrix is included. The parameters characterizing the distribution of the tests are λ $(= \pi' Z' Z \pi)$, the number of IVs k, the correlation between the reduced form errors ρ, and the structural coefficient β.

The numerical simulations apply asymptotically to feasible tests which replace Ω with $\widehat{\Omega}_n$ for stochastic regressors and nonnormal errors. Following Section 6.4 of AMS06a, the power envelopes obtained here are asymptotically valid when the errors are i.i.d normal with *unknown* covariance matrix. Numerical results have been computed for $\lambda/k = 0.5, 1, 2, 4, 8, 16$, which span the range from weak to strong instruments, $\rho = 0.2$, 0.5, 0.9, and $k = 2$, 5, 10, 20. To conserve space, we report only results for $\lambda/k = 1, 2$, $\rho = 0.5$ and $k = 5$. Additional numerical simulations are available in the supplement.

The simulations are presented as plots of power envelopes and power functions against various alternative values of β and λ. Power is plotted as a function of the re-scaled alternative $\beta\lambda^{1/2}$. This can be thought of as a local power plot, where the local neighborhood is $1/\lambda^{1/2}$ instead of the usual $1/n^{1/2}$, since λ measures the effective sample size.

Figures 7.1 and 7.2 plots the power curves for the conditional t-tests as well as both CLR1 and CMLR1 tests. Conditional critical values for all test statistics are computed based on 100,000 Monte Carlo simulations for each observed value $Q_T = q_T$. In the absence of a UMPI test, we consider tests whose power functions may be near the one-sided power envelope for invariant similar tests based on Corollary 2. In the supplement, we provide numerical evidence that the power upper bounds for similar and non-similar tests are alike.

The CLR1 test has rejection probabilities close to the power upper bound for alternatives $\beta > \beta_0$. However, this test has null rejection probabilities close to one for small enough values of $\beta < \beta_0$. This bad behavior is in accordance with Theorem 3 which shows that the CLR1 test is an empirical version of POIS tests. Hence, this test is not very useful for applied researchers. Perhaps surprisingly, all four conditional t-tests perform at least as well as the CMLR1 test which has correct size for $H_0 : \beta \leq \beta_0$. When instruments are very weak ($\lambda/k = 1$), the conditional t-test based on the 2SLS seems to dominate the remaining tests. As λ increases, the power of the conditional t-tests approaches the conservative power envelope. This result is in accordance with chapter 6, which shows that the conditional t-tests are asymptotically efficient under strong-instrument asymptotics (SI-AE). Numerically, the conditional t-tests perform near the power envelope even when λ/k is as small as 2. The supplement provides further evidence for

[11] There is no loss of generality in taking $\beta_0 = 0$ because the structural equation $y_1 = y_2\beta + X\gamma_1 + u$ and hypothesis $H_0 : \beta = \beta_0$ can be transformed into $\widetilde{y}_1 = y_2\widetilde{\beta} + X\gamma_1 + u$ and $H_0 : \widetilde{\beta} = 0$, where $\widetilde{y}_1 = y_1 - y_2\beta_0$ and $\widetilde{\beta} = \beta - \beta_0$.

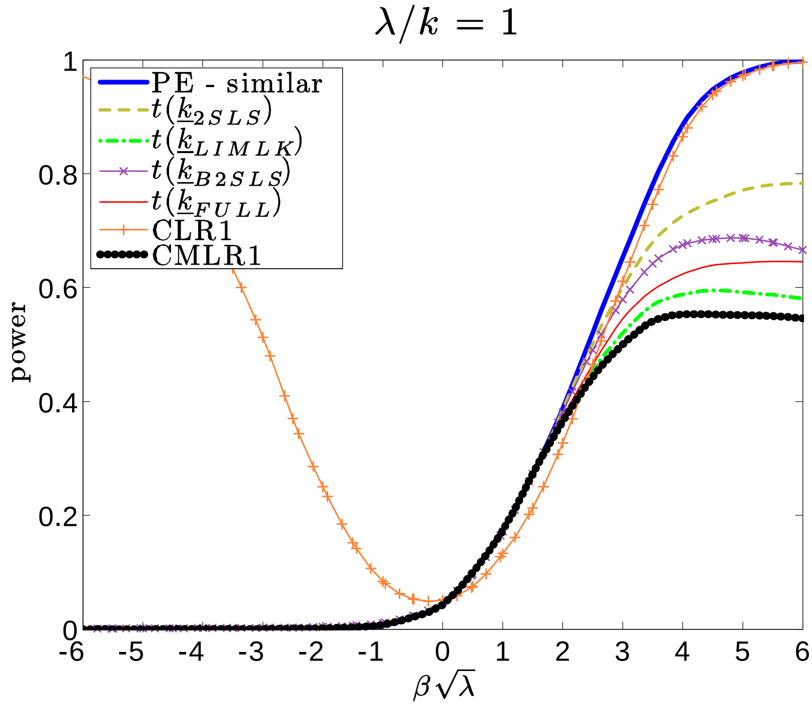


Figure 7.1: Power curves for the one-sided conditional tests when $\lambda\backslash k = 1$

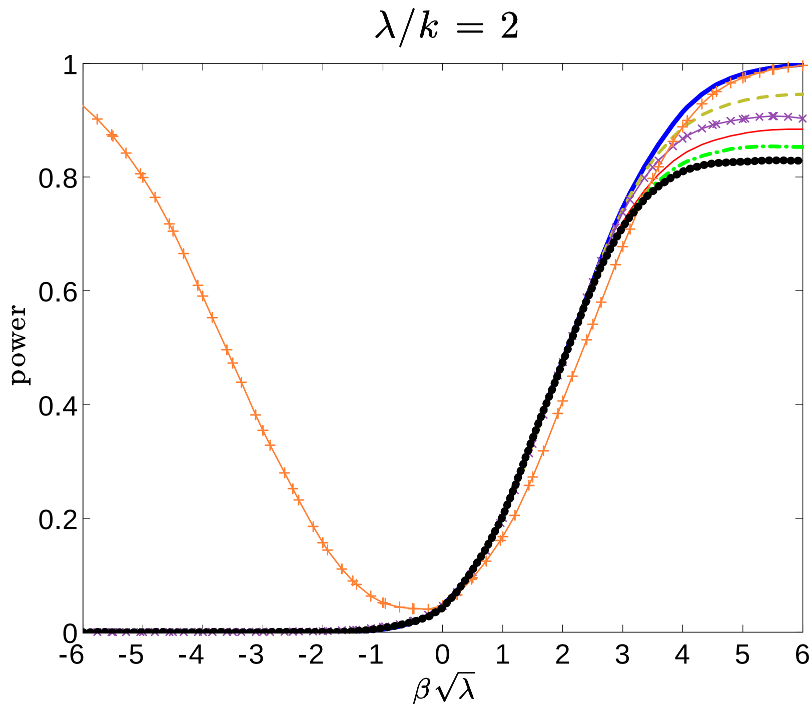


Figure 7.2: Power curves for the one-sided conditional tests when $\lambda\backslash k = 2$

the use of the one-sided conditional t-tests (in particular, the one based on the 2SLS estimator) in empirical practice.

Chapter 8

Two-Sided Tests

In the present chapter we will revise the AMS07's finding for the conditional t-tests in the two-sided hypothesis:

$$H_0 : \beta = \beta_0 \;\text{ against } H_1 : \beta \neq \beta_0. \tag{8.1}$$

In chapter 7 and in the supplements we presented numerical results showing a good performance for the one-sided conditional t-tests, this is striking given the bad performance of the two-sided conditional t-tests documented by AMS07. The goal of this chapter is to solve this apparent counterintuitive result between one-sided and two-sided conditional t-tests.

It turns out that AMS07's finding strongly relies on the asymmetry of the null conditional distribution of the t-statistics considered. Figures 8.1 and 8.2 plots the standard normal distribution with the null conditional distributions for the $t\left(\underline{k}\right)$ statistics based on the 2SLS, LIMLK, Fuller and B2SLS to constrast this asymmetry. When the instruments are strong, ilustrated in Figure 8.1, the null conditional distribution of the t-statistics are symmetric around zero and we can use a symmetric critical value function in the two-sided test without loss of power.

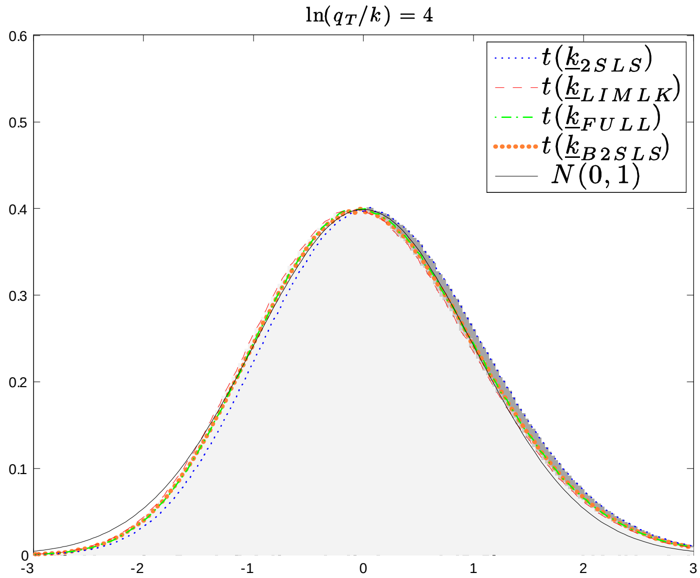


Figure 8.1: Null conditional distribution of $t(\underline{k})$ when $\ln(q_T/k) = 4$.

However, Figure 8.2 shows that when the instruments are weak the null conditional distribution of the t-statistics turn to be asymmetric and the use of symmetric critical value function will generate biased tests.

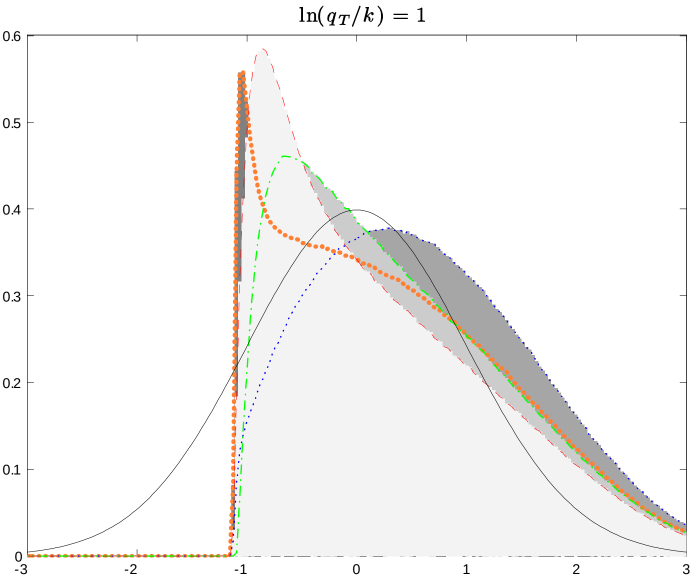


Figure 8.2: Null conditional distribution of $t(\underline{k})$ when $\ln(q_T/k) = 1$.

To overcome this asymmetry we propose two methods: the first method augment the conditional argument of (19, Moreira (2003)) in a manner to obtain two critical value functions; the second method use other t-statistics, which are approximately symmetric, in the construction of the t-tests.

Hereinafter, it is convenient to work with the statistics

$$Q_{(k-1)} = S' M_T S,\ LM1 = Q_{ST}/Q_T^{1/2},\ \text{and } Q_T, \tag{8.2}$$

which are a one-to-one transformation of Q.

For testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, Theorem 1 of AMS06b proves that an unbiased test ϕ must satisfy

$$E_{\beta_0}\left(\phi\left(Q_{(k-1)}, LM1, q_T\right)\right) = \alpha \text{ and} \tag{8.3}$$

$$E_{\beta_0}\left(\phi\left(Q_{(k-1)}, LM1, q_T\right) \cdot LM1\right) = 0 \tag{8.4}$$

for almost all values of q_T. By Corollary 1 of AMS06b, the CLR test satisfies both boundary conditions. Other conditional tests – such as tests based on $t\left(\underline{k}\right)^2$– do not necessarily satisfy (8.4). This places considerable limits on the applicability of conditional method of generating unbiased tests.

For this consider the two-sided unbiased tests based on one-sided statistics $t\left(\underline{k}\right)$ in (3.8) which reject the null when

$$t\left(\underline{k}\right) < \kappa_{t(\underline{k}),1-x_\alpha}\left(q_T\right) \text{ or } t\left(\underline{k}\right) > \kappa_{t(\underline{k}),\alpha-x_\alpha}\left(q_T\right), \tag{8.5}$$

where $\kappa_{t(\underline{k}),x_\alpha}\left(q_T\right)$ is the $1-$ x_α quantile of the conditional distribution and $x_\alpha \in [0, \alpha]$ is chosen to approximately satisfy (8.3) and (8.4). Inverting the approximately unbiased t-tests in (8.5) allows us to construct confidence regions around a chosen estimator (we do not obtain equal-tailed two-sided intervals, otherwise the test would be biased). In particular, we can construct confidence regions based on the 2SLS estimator, which is commonly used in applied research.

Implicit the conditional t-test used in AMS07 consider a symmetric null distribution of $t\left(\underline{k}\right)$ conditional on q_T. Given that and $x_\alpha = \alpha/2$ we have

$$\kappa_{t(\underline{k}),1-x_\alpha}\left(q_T\right) = -\kappa_{t(\underline{k}),\alpha-x_\alpha}\left(q_T\right), \tag{8.6}$$

Consequently, the test that reject the null when (8.5) is equivalent as the test that rejects the null when $t(\underline{k})^2 > \left(\kappa_{t(\underline{k}),\alpha/2}(q_T)\right)^2$. That is, we would have obtained the conditional test based on $t(\underline{k})^2$ where the critical value function is $\kappa_{t(\underline{k})^2,\alpha}(q_T) = \left(\kappa_{t(\underline{k}),\alpha/2}(q_T)\right)^2$.

The second method considered use t-tests based on modifications of the t-statistic that are approximately symmetric. Figures 8.3 and 8.4 plots the null conditional distribution for the modified versions of t-statistics:

$$t_0(\underline{k}) = \frac{\beta(\underline{k}) - \beta_0}{\sigma_0 \cdot \left[y_2' P_Z y_2 - n(\underline{k} - 1)\omega_{22}\right]^{-1/2}}, \tag{8.7}$$

which we use $\sigma_0^2 = (1, -\beta_0)'\Omega(1, -\beta_0)$ as the estimator of the variance of structural error.

The conditional distributions for the $t_0(\underline{k})$ statistics based on the 2SLS and Fuller estimators are also asymmetric around zero when q_T is small. However, the $t_0(\underline{k})$ statistic for the LIMLK estimator is nearly symmetric around zero for any value of q_T. Hence, the conditional test based on $t_0(\underline{k}_{LIMLK})^2$ is nearly unbiased and should not suffer the bad power properties found by AMS07 for the $t(\underline{k})^2$ statistics.

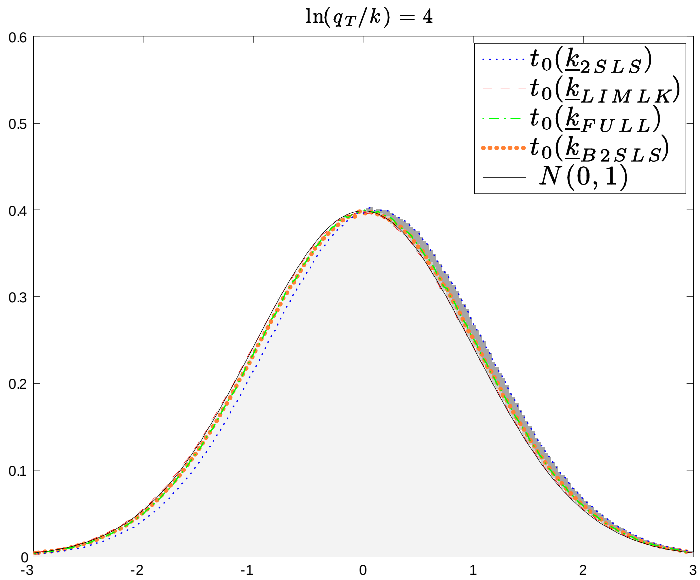


Figure 8.3: Null conditional distribution of $t_0(\underline{k})$ when $\ln(q_T/k) = 4$.

In the supplement, we provide numerical results showing that the conditional t-test based on the $t_0(\underline{k}_{LIMLK})^2$ statistic and some of the unbiased t-tests can perform as well as the CLR test. Hence, the conclusion of AMS07 is only valid for a smaller class of t-tests.

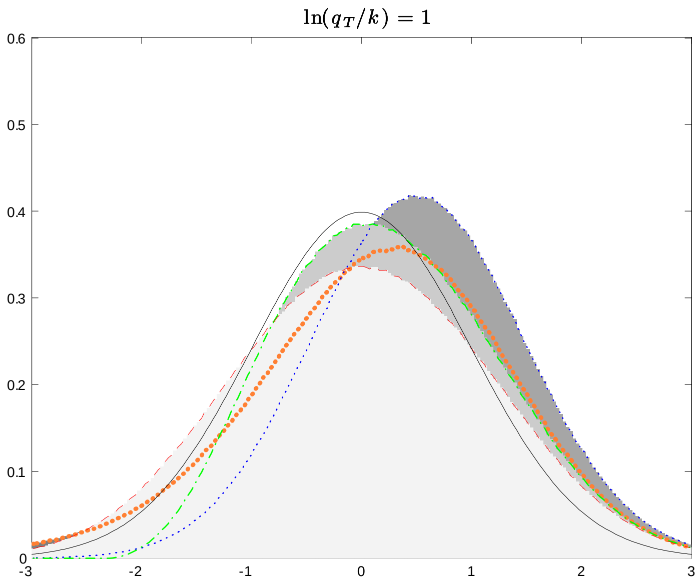


Figure 8.4: Null conditional distribution of $t_0(\underline{k})$ when $\ln(q_T/k) = 1$.

Part I

Appendix

Derivation of the t-Statistics

The k-class estimator for the model (2.6) is given by:

$$\beta(\underline{k}) = \frac{[y_2^{\perp\prime}(I-\underline{k}M_Z)y_1^{\perp}]}{[y_2^{\perp\prime}(I-\underline{k}M_Z)y_2^{\perp}]}, \tag{8.8}$$

where $Y^{\perp} = [y_1^{\perp}, y_2^{\perp}] = M_X Y$ and the four $\underline{k}$-class used are:

$$\begin{array}{ll} \text{2SLS:} & \underline{k} = 1, \\ \text{LIMLK:} & \underline{k} = \underline{k}_{\ LIMLK} = \text{Smallest root } \kappa \text{ of } |(Y'P_Z Y/n + \Omega) - \kappa\Omega| = 0 \\ \text{B2SLS:} & \underline{k} = n/\left(n-k+2\right), \\ \text{Fuller:} & \underline{k} = \underline{k}_{\ LIMLK} - 1/\left(n-k-p\right). \end{array} \tag{8.9}$$

Consequently the t-statistics follows:

$$\begin{aligned} t(\underline{k}) &= \frac{\beta(\underline{k}) - \beta_0}{\sigma_u(\underline{k})[y_2^{\perp\prime}(I-\underline{k}M_Z)y_2^{\perp}]^{-1/2}}, \text{ where} \\ \sigma_u^2(\underline{k}) &= b(\underline{k})'\Omega b(\underline{k}) \text{ and } b(\underline{k}) = (1, -\beta(\underline{k}))'. \end{aligned} \tag{8.10}$$

Since we known Ω we can simplify some expressions:

$$\begin{aligned} [y_2^{\perp\prime}(I-\underline{k}M_Z)y_l^{\perp}] &= [y_2^{\perp\prime}P_Z y_l^{\perp} - (\underline{k}-1)y_2^{\perp\prime}M_Z y_l^{\perp}] \\ &= [y_2 P_Z y_l - n(\underline{k}-1)(n-k-p)/n \cdot e_2'Y'M_X M_Z Y e_l/(n-k-p)] \\ &= [y_2 P_Z y_l - n(\underline{k}-1)(n-k-p)/n \cdot \widehat{w}_{2l}] \\ &= [y_2 P_Z y_l - n(\underline{k}-1)w_{2l}] + o_p(1), \end{aligned} \tag{8.11}$$

for $l = 1, 2$ and where the last equality follows from the fact that $\hat{\Omega}$ is a consistent estimator for Ω, $(n-k-p)/n = 1 + o(1)$ and $n(\underline{k}-1) = O_p(1)$ (this last result is derived from (8.9), (8.15) and (3.3)).

So the t-statistic can be simplified too:

$$\begin{aligned} t(\underline{k}) &= \frac{\beta(\underline{k}) - \beta_0}{\sigma_u(\underline{k})[y_2' P_Z y_2 - n(\underline{k}-1)\omega_{22}]^{-1/2}}, \text{ where} \\ \beta(\underline{k}) &= \frac{[y_2' P_Z y_1 - n(\underline{k}-1)\omega_{21}]}{[y_2' P_Z y_2 - n(\underline{k}-1)\omega_{22}]}, \\ \sigma_u^2(\underline{k}) &= b(\underline{k})'\Omega b(\underline{k}) \text{ and } b(\underline{k}) = (1, -\beta(\underline{k}))'. \end{aligned} \tag{8.12}$$

Now we will derive the t-statistic as a function of Q. First we will prove that the $\underline{k}$'s for the k-class estimators are function of Q, in particular we just need to prove that the $\underline{k}_{LIMLK}$ is (the others follows trivially). By definition $\underline{k}_{LIMLK}$ is the smallest root κ of $|(Y'P_Z Y/n+\Omega)-\kappa\Omega| = 0$, given that and considering the orthogonal matrix:

$$J = \left[\frac{\Omega^{1/2}b_0}{\sqrt{b_0'\Omega b_0}} \vdots \frac{\Omega^{-1/2}a_0}{\sqrt{a_0'\Omega^{-1}a_0}} \right], \tag{8.13}$$

we can find the roots:

$$\begin{aligned} & |(Y'P_Z Y/n + \Omega) - \kappa\Omega| = 0 \\ \Leftrightarrow\ & |J'\Omega^{-1/2}Y'P_Z Y\Omega^{-1/2}J - n(\kappa-1)I_2| = 0 \text{ because } |J| = 1 \\ & (n(\kappa-1))^2 - (n(\kappa-1))(Q_S + Q_T) + (Q_S Q_T - Q_{ST}^2) = 0. \end{aligned} \tag{8.14}$$

Consequently the smallest root gives us the $\underline{k}_{LIMLK}$:

$$\begin{aligned}
\underline{k}_{LIMLK} &= 1+\frac{1}{2n}(Q_S+Q_T-\sqrt{(Q_T-Q_S)^2+4Q_{ST}^2}) \\
&= 1+\frac{1}{n}(Q_S-LR).
\end{aligned} \tag{8.15}$$

Second, define:

$$\begin{aligned}
(Z'Z)^{-\frac{1}{2}}Z'y_l &= [S:T]J'\Omega^{\frac{1}{2}}e_l \\
&= f_lS+g_lT\ ,
\end{aligned} \tag{8.16}$$

where $f_l=b_0'\Omega e_l/\sqrt{b_0'\Omega b_0}$ and $g_l=a_0'e_l/\sqrt{a_0'\Omega^{-1}a_0}$ for $l=1,2$. So we can write the t-statistics as:

$$\begin{aligned}
t(\underline{k}) &= \frac{\beta(\underline{k})-\beta_0}{\sigma_u(\underline{k})[f_2^2Q_S+g_2^2Q_T+2f_2g_2Q_{ST}-n(\underline{k}-1)w_{22}]^{-1/2}}, \text{ where} \\
\beta(\underline{k}) &= \frac{f_1f_2Q_S+g_1g_2Q_T+(g_1f_2+f_1g_2)Q_{ST}-n(\underline{k}-1)w_{21}}{f_2^2Q_S+g_2^2Q_T+2f_2g_2Q_{ST}-n(\underline{k}-1)w_{22}}, \\
\sigma_u^2(\underline{k}) &= b(\underline{k})'\Omega b(\underline{k}) \text{ and } b(\underline{k})=(1,-\beta(\underline{k}))'.
\end{aligned} \tag{8.17}$$

When we do not known Ω the tests statistics need to consider the estimation of Ω. The t-statistics and the $\underline{k}$-class estimator of β follows:

$$\begin{aligned}
\widehat{t}(\widehat{\underline{k}}) &= \frac{\widehat{\beta}(\widehat{\underline{k}})-\beta_0}{\widehat{\sigma}_u(\widehat{\underline{k}})[y_2^{\perp\prime}(I-\widehat{\underline{k}}M_Z)y_2^{\perp}]^{-1/2}}, \\
\widehat{\beta}(\widehat{\underline{k}}) &= \frac{y_2^{\perp\prime}(I-\widehat{\underline{k}}M_Z)y_1^{\perp}}{y_2^{\perp\prime}(I-\widehat{\underline{k}}M_Z)y_2^{\perp}}, \text{ where } \widehat{\sigma}_u^2(\widehat{\underline{k}})=\frac{[y_1^{\perp}-y_2^{\perp}\widehat{\beta}(\widehat{\underline{k}})]'[y_1^{\perp}-y_2^{\perp}\widehat{\beta}(\widehat{\underline{k}})]}{n-1},
\end{aligned} \tag{8.18}$$

where $\underline{k}$'s are given by:

$$\begin{array}{ll}
\text{2SLS:} & \widehat{\underline{k}}=1 \\
\text{LIML:} & \widehat{\underline{k}}=\widehat{\underline{k}}_{\ LIML}=\text{smallest root } \kappa \text{ of } \left|Y^{\perp\prime}Y^{\perp}-\kappa Y^{\perp\prime}M_ZY^{\perp}\right|=0 \\
\text{B2SLS:} & \widehat{\underline{k}}=n/\left(n-k+2\right) \\
\text{Fuller:} & \widehat{\underline{k}}=\widehat{\underline{k}}_{\ LIML}-1/\left(n-k-p\right).
\end{array} \tag{8.19}$$

Using the same arguments as before, the $\underline{k}$ for LIML estimator can be written as function of $\widehat{Q}_n$:

$$\widehat{\underline{k}}_{\ LIML}=1+\frac{1}{n-k-p}(\widehat{Q}_{S,n}-\widehat{LR}_n), \tag{8.20}$$

as $\widehat{\beta}(\widehat{\underline{k}})$:

$$\begin{aligned}
\widehat{\beta}(\widehat{\underline{k}}) &= \frac{y_2P_Zy_1-(\widehat{\underline{k}}-1)(n-k-p)\cdot\widehat{w}_{21}}{y_2P_Zy_2-(\widehat{\underline{k}}-1)(n-k-p)\cdot\widehat{w}_{22}} \\
&= \frac{\widehat{f}_1\widehat{f}_2\widehat{Q}_{S,n}+\widehat{g}_1\widehat{g}_2\widehat{Q}_{T,n}+(\widehat{g}_1\widehat{f}_2+\widehat{f}_1\widehat{g}_2)\widehat{Q}_{ST,n}-(\widehat{\underline{k}}-1)(n-k-p)\cdot\widehat{w}_{21}}{\widehat{f}_2^2\widehat{Q}_{S,n}+\widehat{g}_2^2\widehat{Q}_{T,n}+2\widehat{f}_2\widehat{g}_2\widehat{Q}_{ST,n}-(\widehat{\underline{k}}-1)(n-k-p)\widehat{w}_{22}},
\end{aligned} \tag{8.21}$$

where $\widehat{f}_l=b_0'\widehat{\Omega}_ne_l/\sqrt{b_0'\widehat{\Omega}b_0}$ and $\widehat{g}_l=a_0'e_l/\sqrt{a_0'\widehat{\Omega}_n^{-1}a_0}$ for $l=1,2$, and $\widehat{\sigma}_u^2(\widehat{\underline{k}})$:

$$\begin{aligned}
\widehat{\sigma}_u^2(\widehat{\underline{k}}) &= \frac{[y_1^{\perp}-y_2^{\perp}\widehat{\beta}(\widehat{\underline{k}})]'[y_1^{\perp}-y_2^{\perp}\widehat{\beta}(\widehat{\underline{k}})]}{n-1} \\
&= \widehat{b}(\widehat{\underline{k}})'(\frac{Y'P_ZY}{n-1}+\frac{Y'M_XM_ZY}{n-1})\widehat{b}(\widehat{\underline{k}}) \\
&= \widehat{b}(\widehat{\underline{k}})'(\frac{\widehat{\Omega}^{1/2}\widehat{J}\widehat{Q}J'\widehat{\Omega}^{1/2}}{n-1}+\widehat{\Omega}\cdot\frac{(n-k-p)}{n-1})\widehat{b}(\widehat{\underline{k}}),
\end{aligned} \tag{8.22}$$

where $\widehat{b}(\widehat{\underline{k}}) = (1, -\widehat{\beta}(\widehat{\underline{k}}))$. Finally, t-statistic is a function of $\widehat{Q}_n$:

$$\widehat{t}(\widehat{\underline{k}}) = \frac{\widehat{\beta}(\widehat{\underline{k}}) - \beta_0}{\widehat{\sigma}_u(\widehat{\underline{k}})[\widehat{f}_2^2 \widehat{Q}_{S,n} + \widehat{g}_2^2 \widehat{Q}_{T,n} + 2\widehat{f}_2 \widehat{g}_2 \widehat{Q}_{ST,n} - (\widehat{\underline{k}} - 1)(n-k-p)\widehat{w}_{22}]^{-1/2}}. \tag{8.23}$$

Derivation of the One-sided Likelihood Ratio Statistics

Ignoring an additive constant, the log-likelihood function for known Ω with all parameters concentrated out except β is

$$l_c(Y;\beta,\Omega) = -\frac{n}{2}\ln\det(\Omega) - \frac{1}{2}\left(tr(\Omega^{-1}Y'M_XY) + R(\beta)\right). \tag{8.24}$$

Hence, we have

$$LR1 = 2\left[\sup_{\beta\geq\beta_0} l_c(Y;\beta,\Omega) - l_c(Y;\beta_0,\Omega)\right] = R(\beta_0) - \inf_{\beta\geq\beta_0} R(\beta). \tag{8.25}$$

We now determine $\inf_{\beta\geq\beta_0} R(\beta)$. By definition, $\beta(\underline{k}_{LIMLK})$ maximizes $l_c(Y;\beta,\Omega)$ over $\beta\in\mathbb{R}$. Equivalently, $\beta(\underline{k}_{LIMLK})$ minimizes $R(\beta)$ over $\beta\in\mathbb{R}$. If $\beta(\underline{k}_{LIMLK}) \geq \beta_0$, then $\inf_{\beta\geq\beta_0} R(\beta) = \inf_{\beta\in\mathbb{R}} R(\beta) = R(\beta(\underline{k}_{LIMLK}))$ and $LR1 = R(\beta_0) - \inf_{\beta\in\mathbb{R}} R(\beta) = LR$. If $\beta(\underline{k}_{LIMLK}) < \beta_0$, then $\inf_{\beta\geq\beta_0} R(\beta)$ equals either $R(\beta_0)$ or $R(\infty)$ because $R(\beta)$ is the ratio of two quadratic forms in β with positive definite weight matrices. Hence, the second equality in (3.11) holds. A similar reasoning yields (3.13).

To see that $LR1$ and $MLR1$ are function of Q we just need to prove that $R(\beta)$ is a function of Q, for this note that

$$R(\beta) = \frac{b\Omega^{1/2}JQJ'\Omega^{1/2}b}{b'\Omega b}. \tag{8.26}$$

Furthermore, we have that:

$$\begin{aligned} R(\beta_0) &= Q_S \text{ and} \\ R(\infty) &= \frac{y_2'P_Zy_2}{w_{22}} = \frac{f_2^2Q_S + g_2^2Q_T + 2f_2g_2Q_{ST}}{w_{22}}. \end{aligned} \tag{8.27}$$

Proof of Theorem 1: The power function is given by

$$K(\phi;\beta,\lambda) = \int_{\mathbb{R}^+\times\mathbb{R}\times\mathbb{R}^+} \phi(q_1,q_T) f_{Q_1,Q_T}(q_1,q_T;\beta,\lambda)dq_1dq_T. \tag{8.28}$$

We want to find a test that maximizes power at (β^*,λ^*) among all level α invariant similar tests. By Theorem 2 of AMS06a, invariant similar tests must be similar conditional on $Q_T = q_T$ for almost all q_T. In addition, the unconditional power equals the expected conditional power given Q_T. Hence, it is sufficient to determine the test that maximizes conditional power given $Q_T = q_T$ among invariant tests that are similar conditional on $Q_T = q_T$, for each q_T. By the Neyman-Pearson Lemma, the test of significance level α that maximizes conditional power given $Q_T = q_T$ is of the likelihood ratio (LR) form and rejects H_0 when the LR is sufficiently large (part a) or small (part b). In particular, the conditional LR test statistic is

$$LR_{\beta^*\lambda^*}(q_S,q_{ST},q_T) = \frac{f_{Q_1|Q_T}(q_1|q_T;\beta^*,\lambda^*)}{f_{Q_1|Q_T}(q_1|q_T;\beta_0)} = \frac{f_{Q_1,Q_T}(q;\beta^*,\lambda^*)}{f_{Q_T}(q_T;\beta^*,\lambda^*)f_{Q_1|Q_T}(q_1|q_T;\beta_0)}. \tag{8.29}$$

From the density $f_{Q_1,Q_T}(q_1,q_T;\beta,\lambda)$ given in (3.3), we can determine $f_{Q_T}(q_T;\beta^*,\lambda^*)$ and $f_{Q_1|Q_T}(q_1|q_T;\beta_0)$ to provide the explicit expression for $LR_{\beta^*\lambda^*}(Q_1,Q_T)$ that appears in (4.1); see Lemma 3 of AMS06a. □

Proof of Theorem 3. First, we rewrite $LR1$ in a form that is closer to that of $\widetilde{\xi}_{\widehat{\beta}}$. Ignoring an additive constant, the log-likelihood function (after concentrating out η) can be written as

$$l(Y;\beta,\pi,\Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}tr\left(\Omega^{-1}V'M_X V\right), \text{ where } V = Y - Z\pi a' - X\eta. \tag{8.30}$$

Maximizing (8.30) with respect to π, one finds that $\pi(\beta) = (Z'Z)^{-1} Z'Y\Omega^{-1}a/a'\Omega^{-1}a$, where $a \equiv (\beta\ ,1)'$. The concentrated log-likelihood function, $l_c(Y;\beta,\Omega)$, defined as $l(Y;\beta,\pi(\beta),\Omega)$, is given by

$$l_c(Y;\beta,\Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}\left[tr(\Omega^{-1}Y'M_{\overline{Z}}Y) - \frac{a'\Omega^{-1}Y'P_Z Y\Omega^{-1}a}{a'\Omega^{-1}a}\right]. \tag{8.31}$$

We can simplifies the last term in (8.31):

$$\begin{aligned} \tau(Q;\beta,\Omega) &= \frac{a'\Omega^{-1}Y'P_Z Y\Omega^{-1}a}{a'\Omega^{-1}a} \qquad (8.32)\\ &= \frac{a'\Omega^{-1/2}J(J'\Omega^{-1/2}Y'P_Z Y\Omega^{-1/2}J)J'\Omega^{-1/2}a}{a'\Omega^{-1/2}JJ'\Omega^{-1/2}a} \\ &= \frac{\widetilde{a}'Q\widetilde{a}}{\widetilde{a}'\widetilde{a}}, \text{ where } \widetilde{a} = J\Omega^{-1/2}a. \end{aligned}$$

When evaluated at $\widehat{\beta}$, the maximum likelihood estimator of β under $H_1 : \beta > \beta_0$, (8.31) becomes

$$l_c(Y;\widehat{\beta},\Omega) = -\frac{n}{2}\ln|\Omega| - \frac{1}{2}\left[tr(\Omega^{-1}Y'Y) - \tau(Q;\widehat{\beta},\Omega)\right]. \tag{8.33}$$

Because $LR1$ is defined as $2[l_c(Y;\widehat{\beta},\Omega) - l_c(Y;\beta_0,\Omega)]$, it follows that

$$LR1 = 2[l_c(Y;\widehat{\beta},\Omega) - l_c(Y;\beta_0,\Omega)] = \tau(Q;\widehat{\beta},\Omega) - Q_T. \tag{8.34}$$

Since the term Q_T can be ignored for conditional testing, the CLR1 test is equivalent to rejecting H_0 when $\tau(Q;\widehat{\beta},\Omega) > \kappa_{\tau,\alpha}(Q_T)$. Because the vector $\widehat{a} = (\widehat{\beta},1)'$ maximizes (8.31),

$$\tau(Q;\widehat{\beta},\Omega) = x'_{\widehat{\beta}}Qx_{\widehat{\beta}}, \text{ where } x_{\widehat{\beta}} = (c_{\widehat{\beta}}/||h_{\widehat{\beta}}||, d_{\widehat{\beta}}/||h_{\widehat{\beta}}||)'. \tag{8.35}$$

This proves the equivalence between the CLR1 test and the empirical POIS test based on $\widetilde{\xi}_{\widehat{\beta}} = x'_{\widehat{\beta}}Qx_{\widehat{\beta}}$. □

Proof of Theorem 4.

For parts (a)(i) and (b)(i) note that $\widehat{t}(\widehat{\underline{k}})_n$ in (8.23) is a (almost everywhere) continuous function of $\widehat{Q}_n$ and $\kappa_{t(\underline{k}_\infty),\alpha}(.)$ is a continuous function of $\widehat{Q}_{T,n}$. The result follows from the continuous mapping theorem.

By the same arguments we have the results of parts (a)(ii)-(iv) and (b)(ii)-(iv).

The tests are asymptotically similar at level α by definition of the critical value functions.□

Proof of Lemma 5. For part (a), we need to analyze t-statistics based on the 2SLS, B2SLS, LIMLK and Fuller estimators. The null distribution of the t-statistics conditional on $Q_T = q_T$ depends on the null distribution of Q_S and $\mathcal{S}_2$.

After some calculation, the t-statistic can be written as:

$$\begin{aligned} t(\underline{k}) &= \frac{[f_2 Q_S + g_2 Q_{ST}]}{[f_2^2 Q_S + g_2^2 Q_T + 2f_2 g_2 Q_{ST} - n(\underline{k}-1)w_{22}]^{1/2}} \cdot \frac{\sigma_0}{\sigma_u(\underline{k})} \\ &\quad - \frac{n(\underline{k}-1)(w_{12} - w_{22}\beta_0)}{\sigma_u(\underline{k})[f_2^2 Q_S + g_2^2 Q_T + 2f_2 g_2 Q_{ST} - n(\underline{k}-1)w_{22}]^{1/2}}. \qquad (8.36) \end{aligned}$$

For the 2SLS estimator the second term is zero and the null conditional distribution on $Q_T = q_T$ of the t-statistic is:

$$
\begin{aligned}
t(1) &= \frac{[\frac{f_2}{g_2}\frac{Q_S}{q_T^{1/2}} + \mathcal{S}_2 Q_S^{1/2}]}{[1+\frac{f_2^2}{g_2^2}\frac{Q_S}{q_T} + 2\frac{f_2}{g_2}\mathcal{S}_2\frac{Q_S^{1/2}}{q_T^{1/2}}]^{1/2}} \cdot \frac{\sigma_0}{\sigma_u(1)} \\
&\to {}_d \mathcal{S}_2 Q_S^{1/2} = LM1 \text{ as } q_T \to_p \infty \qquad (8.37)
\end{aligned}
$$

because $\beta(1) \to_p \beta_0$ and $\sigma_u(1) \to_p (b_0'\Omega b_0)^{1/2} = \sigma_0$ as $q_T \to_p \infty$

The same limiting result holds for the t-statistic based on the B2SLS estimator.

Now lets analyse the t-statistic based on the LIMLK estimator. First, by expression (A.12) of AMS06a, the null conditional distribution of the LR statistics converges to chi-square-one as $q_T \to_p \infty$:

$$
\begin{aligned}
LR &= \frac{1}{2}\left[Q_S - q_T + (q_T - Q_S)\left(1 + \frac{2q_T}{(q_T - Q_S)^2} Q_S \mathcal{S}_2^2\right)\right] + o_p(1) \\
&= Q_S \mathcal{S}_2^2 + o_p(1), \qquad (8.38)
\end{aligned}
$$

and then $n(\underline{k}_{LIMLK} - 1)/q_T^{1/2} \to_p 0$ as $q_T \to_p \infty$. Second, the null conditional distribution of the LIMLK estimator of β converges in probability to β_0 as $q_T \to_p \infty$:

$$
\begin{aligned}
\beta(\underline{k}_{LIMLK}) &= \frac{f_1 f_2 \frac{Q_S}{q_T} + g_1 g_2 + (g_1 f_2 + f_1 g_2)\frac{\mathcal{S}_2 Q_S^{1/2}}{q_T^{1/2}} - \frac{n(\underline{k}_{LIMLK}-1)w_{21}}{q_T}}{\frac{f_2^2 Q_S}{q_T} + g_2^2 + 2 f_2 g_2 \frac{\mathcal{S}_2 Q_S^{1/2}}{q_T^{1/2}} - \frac{n(\underline{k}_{LIMLK}-1)w_{22}}{q_T}} \\
&\to {}_p \frac{g_1}{g_2} = \beta_0, \qquad (8.39)
\end{aligned}
$$

consequently $\sigma_u(\underline{k}_{LIMLK}) \to_p \sigma_0$ when $q_T \to_p \infty$. Finally the null conditional distribution of $t(\underline{k}_{LIMLK})$ converges in distribution to standard normal distribution:

$$
\begin{aligned}
t(\underline{k}_{LIMLK}) &= \frac{[\frac{f_2 Q_S}{g_2 q_T^{1/2}} + \mathcal{S}_2 Q_S^{1/2}]}{[1+\frac{f_2^2}{g_2^2}\frac{Q_S}{q_T} + 2\frac{f_2}{g_2}\mathcal{S}_2\frac{Q_S^{1/2}}{q_T^{1/2}} - \frac{n(\underline{k}_{LIMLK}-1)}{q_T}w_{22}]^{1/2}} \cdot \frac{\sigma_0}{\sigma_u(\underline{k}_{LIMLK})} \\
&\quad - \frac{\frac{n(\underline{k}_{LIMLK}-1)}{g_2 q_T^{1/2}}(w_{12} - w_{22}\beta_0)}{\sigma_u(\underline{k}_{LIMLK})[1+\frac{f_2^2}{g_2^2}\frac{Q_S}{q_T} + 2\frac{f_2}{g_2}\mathcal{S}_2\frac{Q_S^{1/2}}{q_T^{1/2}} - \frac{n(\underline{k}_{LIMLK}-1)}{q_T}w_{22}]^{1/2}} \\
&\to {}_d \; LM1 \text{ as } q_T \to_p \infty. \qquad (8.40)
\end{aligned}
$$

We can obtain the same result for the t-statistic based on the Fuller estimator.

For part (b), note first that the null conditional distribution of $\max\{R(\beta_0) - R(\infty), 0\}$ goes to zero in probability as $q_T \to_p \infty$:

$$
\max\left\{Q_S(1 - \frac{f_2^2}{w_{22}}) - g_2 q_T^{1/2}(\frac{g_2 q_T^{1/2} + 2 f_2 \mathcal{S}_2 Q_S^{1/2}}{w_{22}}), 0\right\} \to_p 0, \qquad (8.41)
$$

and so we can apply the continuous mapping theorem to obtain:

$$
\begin{aligned}
LR1^{1/2} &= LR^{1/2} \times 1(t(\underline{k}_{\,LIMLK}) > 0) + o_p(1) \\
&= \mathcal{S}_2 Q_S^{1/2} \times 1(\mathcal{S}_2 Q_S^{1/2} > 0) + o_p(1) \\
&\to {}_d \; \max\{\mathcal{S}_2 Q_S^{1/2}, 0\} \text{ as } q_T \to \infty. \qquad (8.42)
\end{aligned}
$$

The critical value for $\max\{\mathcal{S}_2 Q_S^{1/2}, 0\}$ at level α (with $0 < \alpha < 1/2$) is z_α because

$$P\left(\max\{\mathcal{S}_2 Q_S^{1/2}, 0\} \geq z_\alpha\right) = P\left(\mathcal{S}_2 Q_S^{1/2} \geq z_\alpha\right) = \alpha. \tag{8.43}$$

Part (c) also follows from (8.41), (8.42) and (8.43).□

Proof of Theorem 6. To prove part (a) note that the $\underline{k}$-class estimator are consistent:

$$\begin{aligned} \beta(\underline{k}) &= \frac{f_1 f_2 \frac{Q_{S,n}}{n} + g_1 g_2 \frac{Q_{T,n}}{n} + (g_1 f_2 + f_1 g_2) \frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{21}}{f_2^2 \frac{Q_{S,n}}{n} + g_2^2 \frac{Q_{T,n}}{n} + 2 f_2 g_2 \frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{22}} \\ &\to_p \frac{g_1}{g_2} = \beta_0, \end{aligned}$$

because $(\underline{k} - 1) = O_p(n^{-1})$ for each $\underline{k}$-class considered. Consequently, $\sigma_u(\underline{k}) \to_p \sigma_0$ and

$$\begin{aligned} t(\underline{k})_n &= \frac{[f_2 \frac{Q_{S,n}}{n^{1/2}} + g_2 \frac{Q_{ST,n}}{n^{1/2}}]}{[f_2^2 \frac{Q_{S,n}}{n} + g_2^2 \frac{Q_{T,n}}{n} + 2 f_2 g_2 \frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{22}]^{1/2}} \cdot \frac{\sigma_0}{\sigma_u(\underline{k})} \\ &\quad - \frac{n^{1/2} (\underline{k} - 1)(w_{12} - w_{22} \beta_0)}{\sigma_u(\underline{k}) [f_2^2 \frac{Q_{S,n}}{n} + g_2^2 \frac{Q_{T,n}}{n} + 2 f_2 g_2 \frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{22}]^{1/2}} \\ &\to_d \alpha_T' S_{B\infty} / ||\alpha_T||. \end{aligned} \tag{8.44}$$

Its easy to see that the $\widehat{t}\left(\widehat{\underline{k}}\right)_n$ statistics are asymptotically equivalent to the corresponding $t(\underline{k})_n$ statistics.

Part (b) trivially follows from (6.1).

For part (c), recall that the $LR1_n$ statistic is

$$LR1_n = LR_n \times 1(\beta(\underline{k}_{LIMLK}) > \beta_0) + \max\{R(\beta_0) - R(\infty), 0\} \times 1(\beta(\underline{k}_{LIMLK}) < \beta_0). \tag{8.45}$$

Under local alternatives, $\max\{R(\beta_0) - R(\infty), 0\} \to_p 0$ because

$$\begin{aligned} R(\beta_0) &= Q_{S,n} = O_p(1) \text{ and} \\ R(\infty) &= (f_2^2 Q_{S,n} + g_2^2 Q_{T,n} + 2 f_2 g_2 Q_{ST,n}) / w_{22} \to_p \infty. \end{aligned} \tag{8.46}$$

Therefore,

$$\begin{aligned} LR1_n^{1/2} &= LR_n^{1/2} \times 1(t(\underline{k}_{\ LIMLK}) > 0) + o_p(1) \\ &\to_d (\alpha_T' S_{B\infty}) / ||\alpha_T|| \times 1[\alpha_T' S_{B\infty} / ||\alpha_T|| > 0], \end{aligned} \tag{8.47}$$

where the third equality follows from the continuous mapping theorem and the joint convergence of LR and $t(\underline{k}_{\ LIMLK})$ to $(\alpha_T' S_{B\infty})^2 / ||\alpha_T||^2$ and $\alpha_T' S_{B\infty} / ||\alpha_T||$, respectively; see AMS06a, Theorem 6(c), regarding the convergence in distribution of LR to $(\alpha_T' S_{B\infty})^2 / ||\alpha_T||^2$. By (6.1) we have the same asymptotic result for $\widehat{LR1}_n$.

Part (d) also follows from (8.47) because $\max\{R(\beta_0) - R(\infty), 0\}$ converges in probability to zero.□

Proof of Theorem 7. For part (a), following the proof of Theorem 7 of AMS06a, we know that the one-sided LM statistic for known Ω is $LM1_n = Q_{ST,n} / Q_{T,n}^{1/2}$, which is asymptotically efficient by standard results.

By Lemma 5, the critical values of conditional tests based on $LR1_n$, $MLR1_n$, and t-statistics converge to a standard normal $1 - \alpha$ quantile (provided $\alpha \in [0, 1/2)$ for the likelihood ratio statistics).

The $LR1_n^{1/2}$ and $MLR1_n^{1/2}$ statistics are not asymptotically equivalent to $LM1_n$. However, the asymptotic power of the one-sided tests based on $LR1_n^{1/2}$, $MLR1_n^{1/2}$, and $LM1_n$ are the same:

$$\begin{aligned} P\left(\max\{(\alpha_T' S_{B\infty})/||\alpha_T||, 0\} \geq z_\alpha\right) &= P(\max\{\varsigma_1 + \lambda^{1/2} B\left(b_0'\Omega b_0\right)^{-1/2}, 0\} \geq z_\alpha) \\ &= P(\varsigma_1 + \lambda^{1/2} B\left(b_0'\Omega b_0\right)^{-1/2} \geq z_\alpha) \\ &= P\left(\alpha_T' S_{B\infty}/||\alpha_T|| \geq z_\alpha\right), \end{aligned} \tag{8.48}$$

where $\varsigma_1 \sim N(0,1)$, $B > 0$, and z_α is a positive critical value.

By Theorem 6, the asymptotic behavior of the tests above are the same when $\widehat{\Omega}_n$ replaces Ω. Hence, these tests are asymptotically efficient when Ω is estimated. $\square$

Proof of Lemma 8. Part (i) of the Lemma is established as follows:

$$\begin{aligned} S_n/n^{1/2} &= (n^{-1}Z'Z)^{-1/2} n^{-1} Z'Y b_0 \cdot (b_0'\Omega b_0)^{-1/2} \\ &\rightarrow_p D_Z^{1/2} \pi a' b_0 \cdot (b_0'\Omega b_0)^{-1/2} = D_Z^{1/2} \pi c_\beta. \end{aligned} \tag{8.49}$$

Similarly,

$$\begin{aligned} T_n/n^{1/2} &= (n^{-1}Z'Z)^{-1/2} n^{-1} Z'Y \Omega^{-1} a_0 \cdot (a_0'\Omega^{-1} a_0)^{-1/2} \\ &\rightarrow_p D_Z^{1/2} \pi a' \Omega^{-1} a_0 \cdot (a_0'\Omega^{-1} a_0)^{-1/2} = D_Z^{1/2} \pi d_\beta. \end{aligned} \tag{8.50}$$

Part (ii) of the Lemma follows from Lemma 1 of AMS06b and part (i).

Next, we prove part (iii) of the Lemma. If $\beta = \beta_{AR}$, then $a'\Omega^{-1}a_0 = 0$ and using Assumption 4, we have

$$T_n = (n^{-1}Z'Z)^{-1/2} n^{-1/2} Z'V\Omega^{-1} a_0 \cdot (a_0'\Omega^{-1} a_0)^{-1/2} \tag{8.51}$$

$$\rightarrow_d \varsigma_k \sim N(0, I_k). \tag{8.52}$$

Part (iii) now follows from Lemma 1 of AMS06b. $\square$

Proof of Theorem 9: For part (a), we use Lemma 8 in (3.9) to note that $(\underline{k} - 1) = o_p(1)$ and consequently the $\underline{k}$-class estimators are consistent:

$$\begin{aligned} \beta(\underline{k}) &= \frac{f_1 f_2 \frac{Q_{S,n}}{n} + g_1 g_2 \frac{Q_{T,n}}{n} + (g_1 f_2 + f_1 g_2)\frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{21}}{f_2^2 \frac{Q_{S,n}}{n} + g_2^2 \frac{Q_{T,n}}{n} + 2 f_2 g_2 \frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{22}} \\ &\rightarrow_p \frac{f_1 c_\beta + g_1 d_\beta}{f_2 c_\beta + g_2 d_\beta} = \beta, \end{aligned} \tag{8.53}$$

if $\beta \neq \beta_{AR}$ or $\beta = \beta_{AR}$, therefore $\sigma_u(\underline{k}) \rightarrow_p (b'\Omega b)^{1/2} = \sigma_u$.

Now we can show the asymptotic distribution of $t(\underline{k})_n/n^{1/2}$. For either $\beta \neq \beta_{AR}$ or $\beta = \beta_{AR}$:

$$\begin{aligned} \frac{t(\underline{k})_n}{\sqrt{n}} &= \frac{[f_2 \frac{Q_{S,n}}{n} + g_2 \frac{Q_{ST,n}}{n}]}{[f_2^2 \frac{Q_{S,n}}{n} + g_2^2 \frac{Q_{T,n}}{n} + 2 f_2 g_2 \frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{22}]^{1/2}} \cdot \frac{\sigma_0}{\sigma_u(\underline{k})} \\ &\quad - \frac{(\underline{k} - 1)(w_{12} - w_{22}\beta_0)}{\sigma_u(\underline{k})[f \frac{Q_{S,n}}{n} + g_2^2 \frac{Q_{T,n}}{n} + 2 f_2 g_2 \frac{Q_{ST,n}}{n} - (\underline{k} - 1) w_{22}]^{1/2}} \\ &\rightarrow_p c_\beta \lambda_{FA}^{1/2} \frac{\sigma_0}{\sigma_u}. \end{aligned} \tag{8.54}$$

By the same arguments, $\widehat{t}(\widehat{\underline{k}})_n/n^{1/2}$ converges in probability to $c_\beta \lambda_{FA}^{1/2} \sigma_0/\sigma_u$.

Part (b) follows trivially.

For parts (c) and (d), we write the $LR1_n$ statistic as in (3.11). To study the behavior of $LR1_n$ under fixed alternatives, we use the following:

$$\begin{aligned} \frac{R(\infty)}{n} &= (f_2^2\frac{Q_{S,n}}{n} + g_2^2\frac{Q_{T,n}}{n} + 2f_2g_2\frac{Q_{ST,n}}{n})w_{22}^{-1} \to_p \lambda_{FA}w_{22}^{-1}, \\ \frac{R(\beta_0)}{n} &= \frac{Q_{S,n}}{n} \to_p c_\beta^2\lambda_{FA}, \text{ and} \\ \frac{LR_n}{n} &= \frac{1}{2}(\frac{Q_{S,n}}{n} - \frac{Q_{T,n}}{n} + \sqrt{(\frac{Q_{T,n}}{n} - \frac{Q_{S,n}}{n})^2 + 4\frac{Q_{ST,n}}{n}} \to_p c_\beta^2\lambda_{FA}. \end{aligned} \tag{8.55}$$

which holds if: $\beta = \beta_{AR}$ or $\beta \neq \beta_{AR}$.

If $\beta > \beta_0$, $1(\beta(\underline{k}_{LIMLK}) \geq \beta_0) \to_p 1$ by continuous mapping theorem and so

$$\frac{LR1_n}{n} \to_p c_\beta^2\lambda_{FA}. \tag{8.56}$$

If $\beta < \beta_0$, $1(\beta(\underline{k}_{LIMLK}) \geq \beta_0) \to_p 0$ and

$$\frac{LR1_n}{n} \to_p \max\{c_\beta^2 - w_{22}^{-1}, 0\}\lambda_{FA}. \tag{8.57}$$

The same is true for $\widehat{LR1_n}/n$.

For parts (e) and (f), we write the $MLR1_n$ statistic as in (3.13).

If $\beta > \beta_0$, $1(\beta(\underline{k}_{LIMLK}) \geq \beta_0) \to_p 1$ and

$$\frac{MLR1_n}{n} \to_p c_\beta^2\lambda_{FA} - \max\{c_\beta^2\lambda_{FA} - w_{22}^{-1}\lambda_{FA}, 0\} + c_\beta^2\lambda_{FA} = \min\{c_\beta^2, w_{22}^{-1}\}\lambda_{FA}. \tag{8.58}$$

If $\beta > \beta_0$, $1(\beta(\underline{k}_{LIMLK}) \geq \beta_0) \to_p 0$ and

$$\frac{MLR1_n}{n} \to_p 0.$$

For $\widehat{MLR1_n}/n$ we have the same result.□

Bibliography

[1] 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.

[2] Andrews, D. W. K., and P. Guggenberger (2010): "Applications of subsampling, hybrid, and size-correction methods," *Journal of Econometrics*, 158, 285–305.

[3] Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression," NBER Working Paper t0299.

[4] 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.

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

[6] Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2007): "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 139, 116–132.

[7] Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2008): "Optimal Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 146, 241–254.

[8] Andrews, D. W. K., and J. H. Stock (2007): "Inference with Weak Instruments,"in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, ed. by T. P. R. Blundell, W. K. Whitney, vol. 3, chap. 6. Cambridge University Press, Cambridge.

[9] Angrist, J., and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106, 979–1014.

[10] Cattaneo, M., R. Crump, and M. Jansson (2012): "Optimal Inference for Instrumental Variables Regression with Non-Gaussian Errors," *Journal of Econometrics*, 167, 1–15.

[11] Chamberlain, G. (2007): "Decision Theory Applied To an Instrumental Variables Model," *Econometrica*, 75, 609–652.

[12] Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1192.

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

[14] Dufour, J. M (2003): "Presidential Address: Identification, Weak Instruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36, 767–808.

[15] Fuller, W. A (1977): "Some properties of a modification of the limited information estimator," *Econometrica*, 45, 939–954.

[16] Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

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

[18] Moreira, M. J. (2002): "Tests with Correct Size in the Simultaneous Equations Model," Ph.D. thesis, UC Berkeley.

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

[20] Moreira, M. J. (2009): "A Maximum Likelihood Method for the Incidental Parameter Problem," *Annals of Statistics*, 37, 3660–3696.

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

[22] Stock, J. H., J. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.

1 Asymptotic Power

Figures 1 to 36 provide power curves for all one-sided designs considered. For the one-sided power curves, we plot graphs of the power envelope for similar tests (PE - similar), the power envelope for nonsimilar tests (PE - nonsimilar), one-sided conditional $t\left(\underline{k}\right)$ and $t_0\left(\underline{k}\right)$ tests based on the two-stage least squares (2SLS), limited information maximum likelihood (LIML), Fuller (FULL; Fuller, 1977), and bias adjusted two-stage least squares (B2SLS; Nagar, 1959, Rothenberg, 1984) estimators, CLR1, and CMLR1 tests. Figures 37 to 60 provide power curves for all two-sided designs considered. For the two-sided power curves, we plot graphs of the power envelope for locally unbiased invariant tests (PE-LU; Andrews, Moreira, and Stock, 2006a), unbiased $t\left(\underline{k}\right)$, unbiased $t_0\left(\underline{k}\right)$, Moreira's (2003) conditional $t\left(\underline{k}\right)^2$, and conditional $t_0\left(\underline{k}\right)^2$ tests.

The critical values for the power envelope for one-sided similar tests and two-sided locally unbiased invariant tests, conditional t-tests, and the unbiased t-tests are calculated based on 100,000 simulations of $Q_{(k-1)}$ and $LM1$. For the one-sided nonsimilar power envelope, least favorable distributions were approximated using a single-point distribution. The critical values for the PE-LU are calculated using the algorithm described in Section 8.3 of Andrews, Moreira, and Stock (2006b).

For two-sided unbiased tests, we employ an algorithm which approximates the critical value functions $\kappa_{t(\underline{k}),1-x_\alpha}\left(q_T\right)$ and $\kappa_{t(\underline{k}),\alpha-x_\alpha}\left(q_T\right)$. We first write the t-statistics as functions of $Q_{(k-1)}$, $LM1$, and Q_T, e.g., $t\left(\underline{k}\right) = \psi\left(Q_{k-1}, LM1, q_T, \beta_0, \Omega\right)$. We generate J replications drawn from $\chi^2_{(k-1)}$ and standard normal distributions: $\{Q^j_{(k-1)}, LM1^j\}^J_{j=1}$. This allows us to approximate the conditional null distribution of $t\left(\underline{k}\right)$ by

$$P^J\left(\psi\left(Q_{k-1}, LM1, q_T, \beta_0, \Omega\right) \leq x\right) = J^{-1}\sum_{j=1}^{J} I\left(\psi\left(Q^j_{k-1}, LM1^j, q_T, \beta_0, \Omega\right) \leq x\right). \quad (1)$$

To find the approximate critical values $\kappa^J_{t(\underline{k}),1-x_\alpha}\left(q_T\right) = c^J_1$ and $\kappa^J_{t(\underline{k}),\alpha-x_\alpha}\left(q_T\right) = c^J_2$, we follow three steps for each level of q_T. First, we compute the quantiles based on the empirical distribution $P^J\left(\psi\left(Q_{k-1}, LM1, q_T, \beta_0, \Omega\right) \leq Q^J_z\right) = z$. Then, we find x^J_α in the interval $[0, \alpha]$ which solves

$$\min_x \left| J^{-1}\sum_{j=1}^{J} I\left(Q^J_{1-x} < \psi\left(Q^j_{k-1}, LM1^j, q_T, \beta_0, \Omega\right) < Q^J_{\alpha-x}\right)^c LM1^j \right|. \quad (2)$$

Finally, we use $Q^J_{1-x^J_\alpha}$ and $Q^J_{\alpha-x^J_\alpha}$ respectively as approximations to $\kappa_{t(\underline{k}),1-x_\alpha}\left(q_T\right)$ and $\kappa_{t(\underline{k}),\alpha-x_\alpha}\left(q_T\right)$.

The power function for each design is computed at 13 evenly spaced points for $\beta/\sqrt{\lambda}$ ranging from -6 to 6. Results are based on 5,000 simulations. The parameter combinations considered are $k = 2, 5, 10, 20$, $\rho = 0.2, 0.5, 0.9$, and $\lambda/k = 0.5, 1, 2, 4, 8, 16$. We note that the cases of $\rho = -0.2, -0.5, -0.9$ are covered as well. For one-sided tests, the power curves with $\rho = \rho_0$ for $H_1 : \beta < 0$ are the mirror image

of the power curves with $\rho = -\rho_0$ for $H_1 : \beta > 0$. Similarly for two-sided tests, the power curves with $\rho = \rho_0$ are the mirror image of the power curves with $\rho = -\rho_0$.

2 Empirical Example

We revisit Angrist and Krueger's (1991) work on returns to schooling for the 1920-29, 1930-39, and 1940-49 cohorts using the U.S. Census. The data consist of the original 1970 and 1980 U.S. Census sample of males born in the United States used by Angrist and Krueger (1991)[1]. The 1920-29 cohort is from the 1970 Census sample, the 1930-39 and 1940-49 cohorts are both from the 1980 Census sample. For the 1920-29 cohort we excluded 3000 individuals whose place-of-birth was ambiguous, leaving us with a sample size of 244,099 observations. All individuals in the 1930-39 and 1940-49 cohorts had clearly identified places-of-birth and none were eliminated. The sample sizes of the 1930-39 and 1940-49 cohorts are 329,509 and 486,926 observations, respectively.

We include a constant, race, metropolitan area, marital status, age, age-squared, and dummies for year-of-birth, state-of-birth, and regions as covariates. The log-weekly earning variable is imputed from weeks worked and annual earnings. The metropolitan area variable equals one if the individual lives in a Standard Metropolitan Statistical Area (SMSA), as determined by the U.S. Office of Management and Budget and recorded by the Census. The 50 state-of-birth dummies are constructed from the place-of-birth Census code variable, ignoring the code value of 11 (District of Columbia). The race variable equals one if the individual is black and zero if not. The 8 regional dummy variables are constructed from the Census region code variable that represents the 9 Regional divisions used by the U.S. Census Bureau. The age variable is imputed from the year-of-birth and quarter-of-birth of individuals. A detailed description of how the data were constructed from the original Census data can be found in Appendix 1 of Angrist and Krueger (1991).

Tables 1 to 3 provide specifications for the empirical example. Specifications I and II use quarter-of-birth and quarter-of-birth $\times$ year-of-birth respectively as instruments, and include a constant, race, metropolitan area and marital status, nine year-of-birth and eight regional dummies as controls. Specification III adds age and age^2 as covariates and allows interaction between quarter-of-birth and year-of-birth. Specification IV replaces year-of-birth dummies used in specification III by state-of-birth dummies.

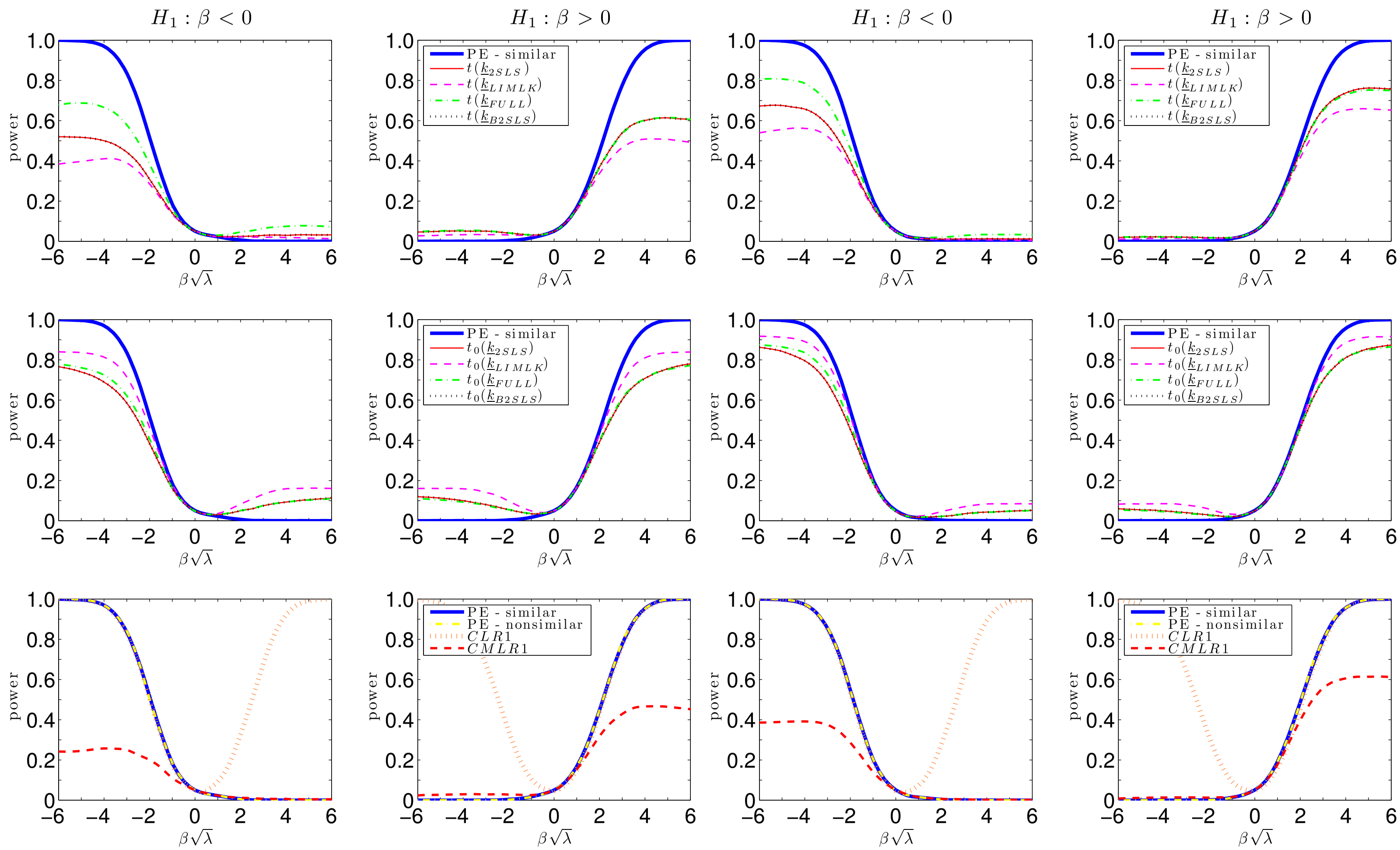
The first rows present the OLS, LIML, 2SLS, Fuller and B2SLS estimators with their respective standard errors. We next present 95% confidence confidence intervals for one-sided tests: CMLR1 and unbiased t-tests based on the LIML, and 2SLS, Fuller and B2SLS estimators. Finally we present 95% confidence intervals for two-sided tests: CLR and unbiased t-tests based on the LIML, 2SLS, Fuller and B2SLS estimators.

We focus on specification IV, shown by Cruz and Moreira (2005) to be informative

[1]The data are available at http://economics.mit.edu/faculty/angrist/data1/data.

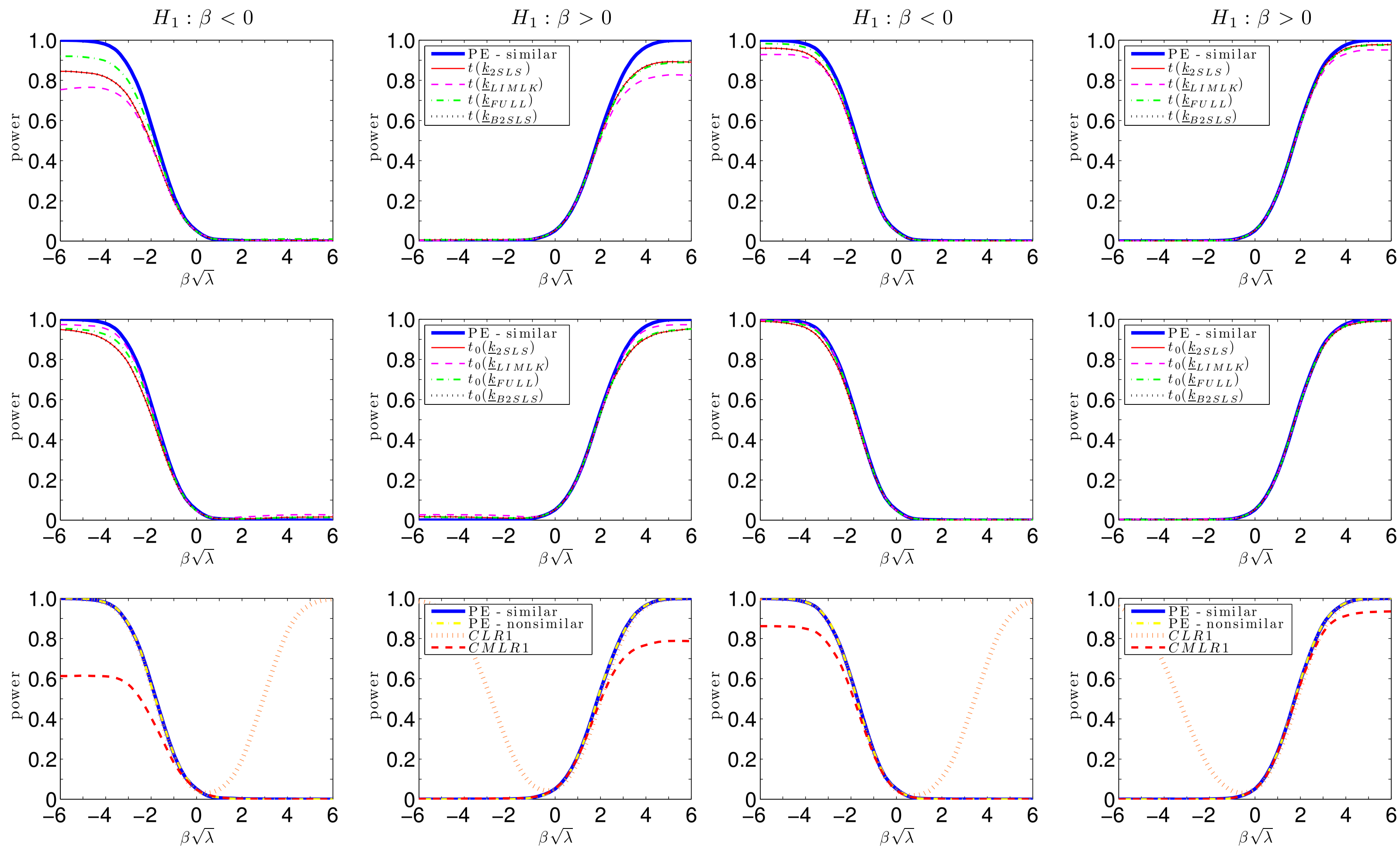
for returns to schooling despite the low first-stage F-statistic and 178 instruments. The LIML estimator is larger than the 2SLS estimator for the three cohorts. We first report confidence intervals for one-sided tests: CMLR1 and the unbiased t-tests based on the 2SLS and LIML estimators. One-sided tests may be appropriate in this application. Because returns to schooling β are non-negative, we can test against the alternative $H_1 : \beta > \beta_0$. The confidence regions for the t-tests are comparable despite the associated estimates being quite different. For example, take the 1930-39 cohort. The 2SLS estimates returns to schooling to be 8.11% while the LIML estimate is 9.82%. The lower bound for our confidence regions is about the same (6.1%-6.3%). In the one-sided case we do not make a statement about the upper bound on returns to schooling.

Next we report confidence intervals for two-sided tests: the CLR test and unbiased t-tests based on the 2SLS and LIML estimators. As in the one-sided case, confidence regions for the t-tests and CLR test are comparable. The lower bound for our confidence regions of the 1930-39 cohort is about the same across two-sided tests (5.4%-5.6%), while the upper bound for the unbiased t-tests using the 2SLS estimator is slightly larger than using the LIML estimator (15.5% instead of 14.5%). The unbiased t-tests and the CLR test produce comparable confidence intervals. In particular, the confidence regions centered around the 2SLS estimator can be informative even when the first-stage F-statistic is low. This empirical exercise supports our theoretical work on the use of the 2SLS and confidence intervals based on unbiased t-tests in applied work.



(a) $\lambda/k = 0.5$ (b) $\lambda/k = 1$

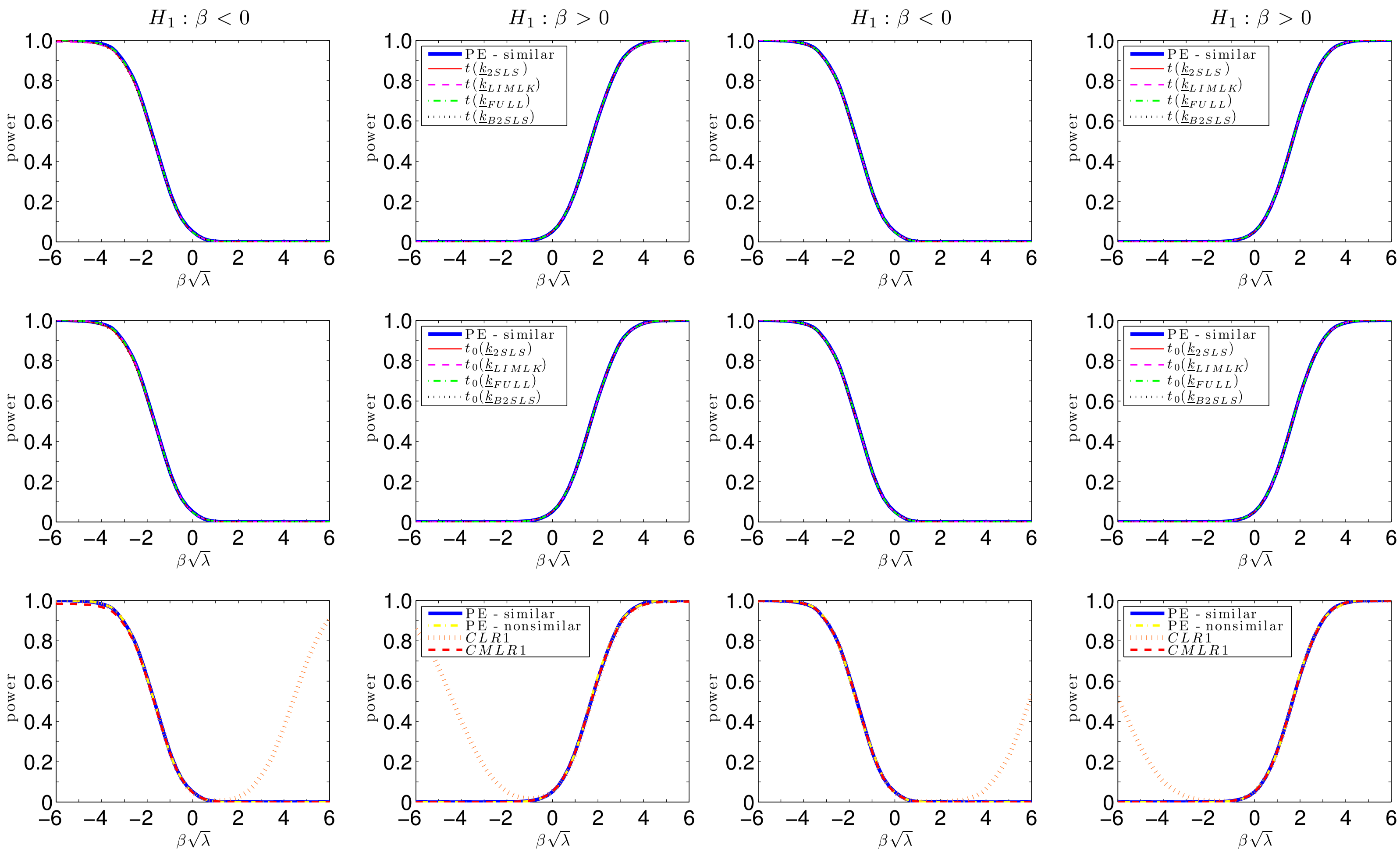
Figure 1: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 2$.



(a) $\lambda/k = 2$

(b) $\lambda/k = 4$

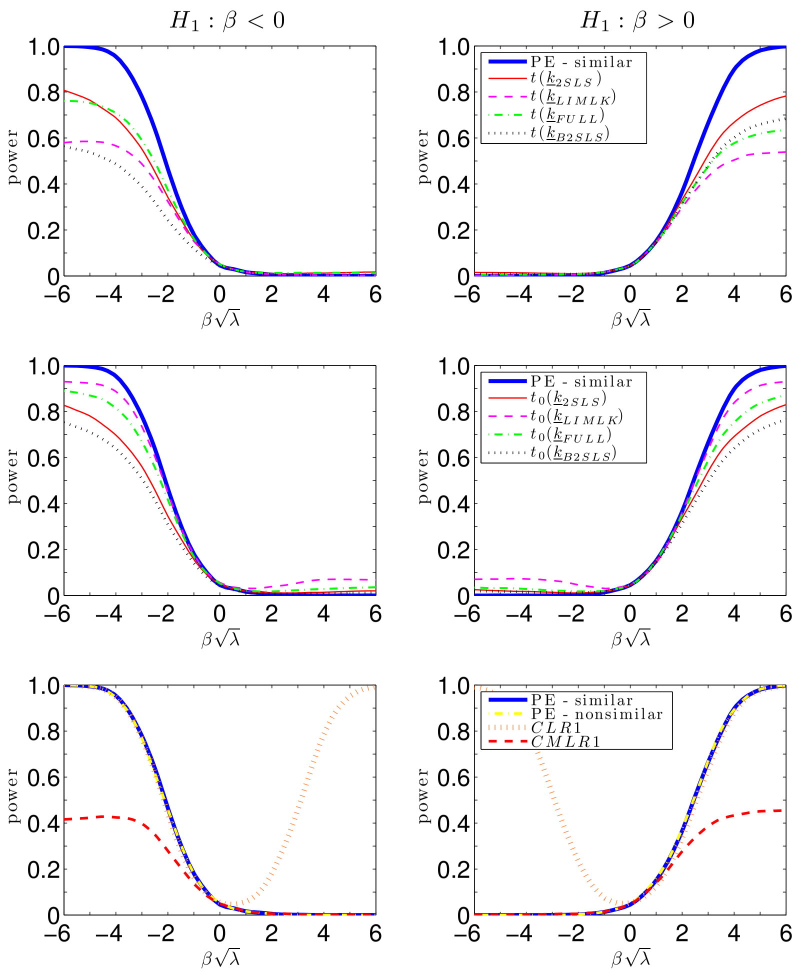
Figure 2: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 2$.



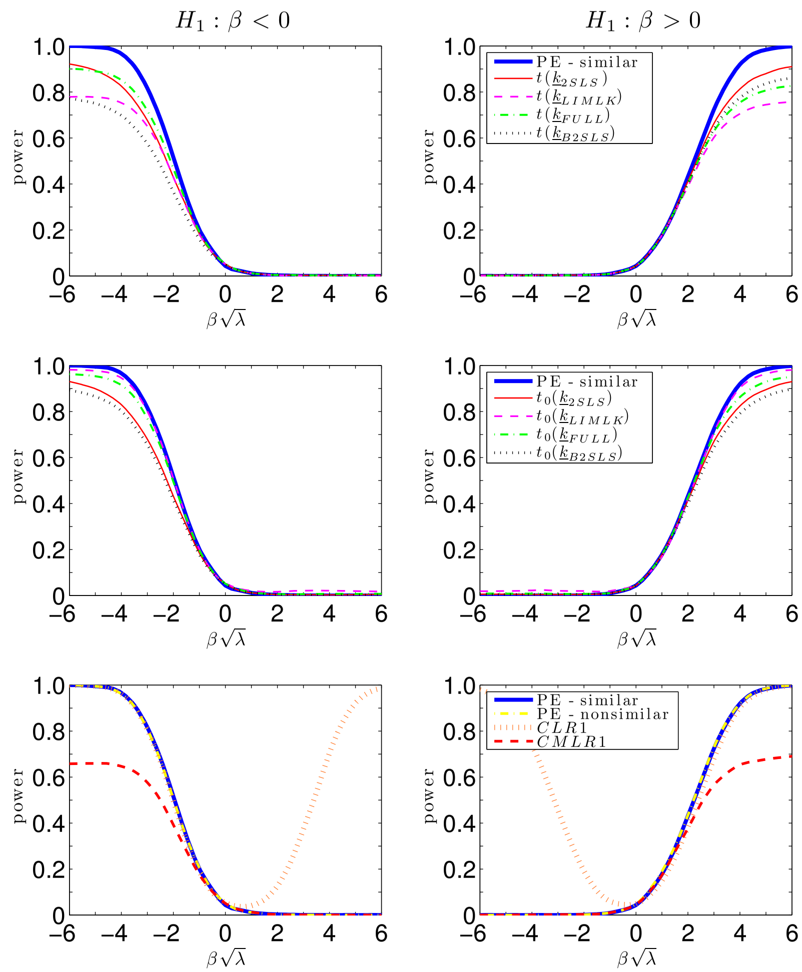
(a) $\lambda/k = 8$

(b) $\lambda/k = 16$

Figure 3: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 2$.

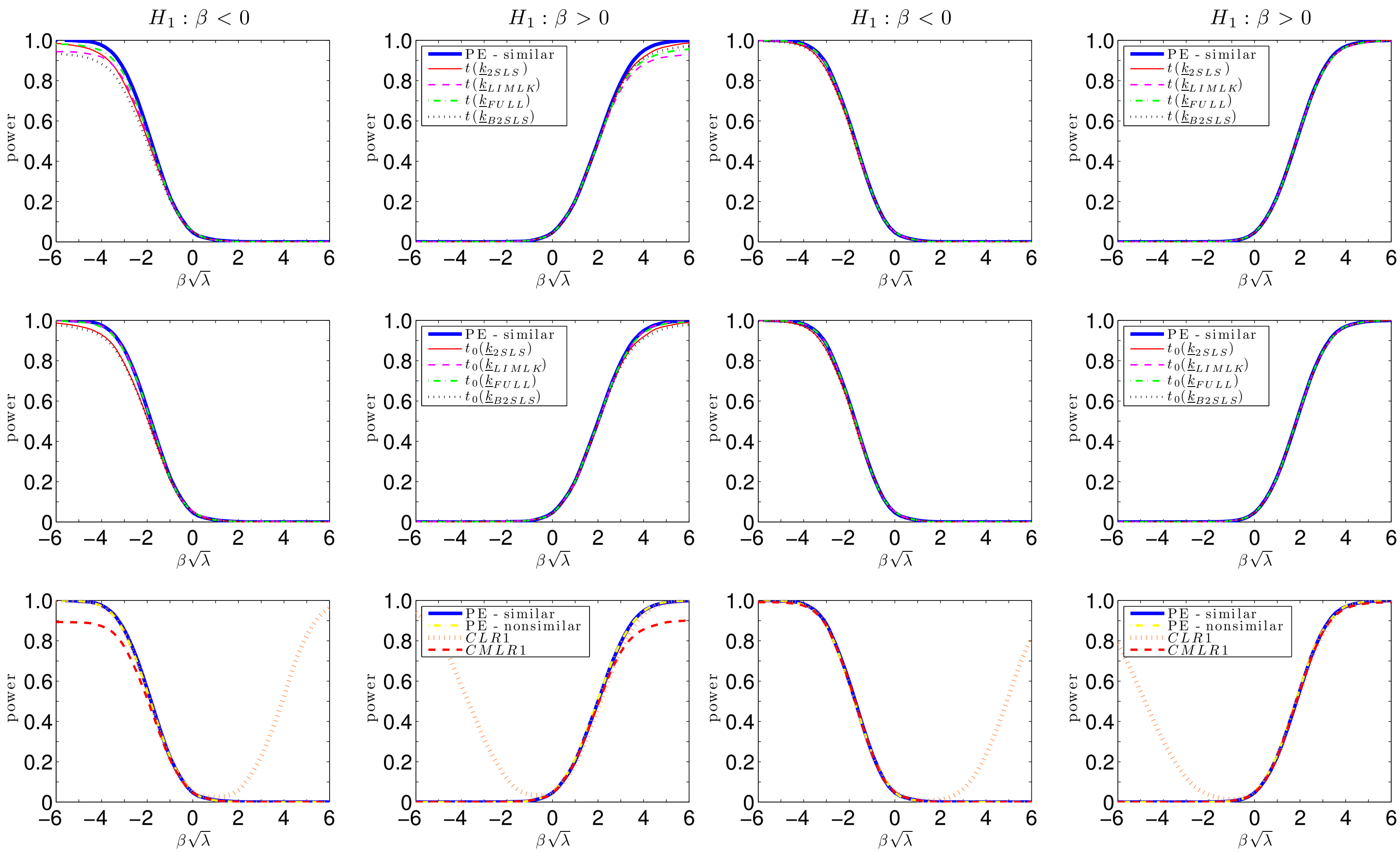


(a) $\lambda/k = 0.5$



(b) $\lambda/k = 1$

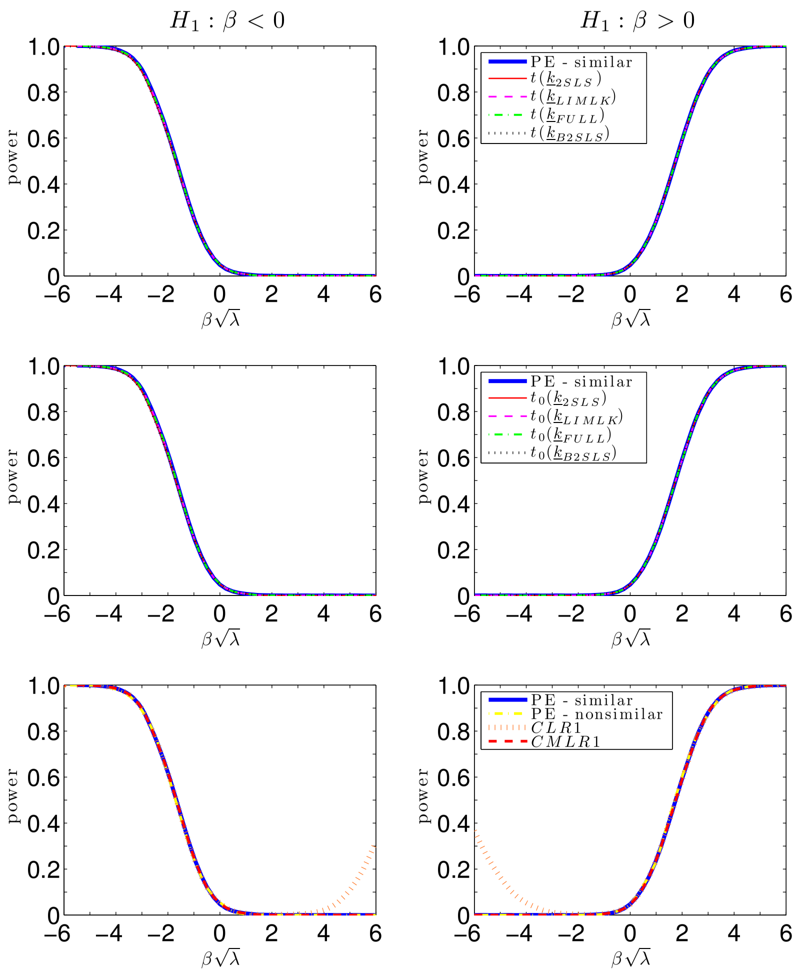
Figure 4: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 5$.



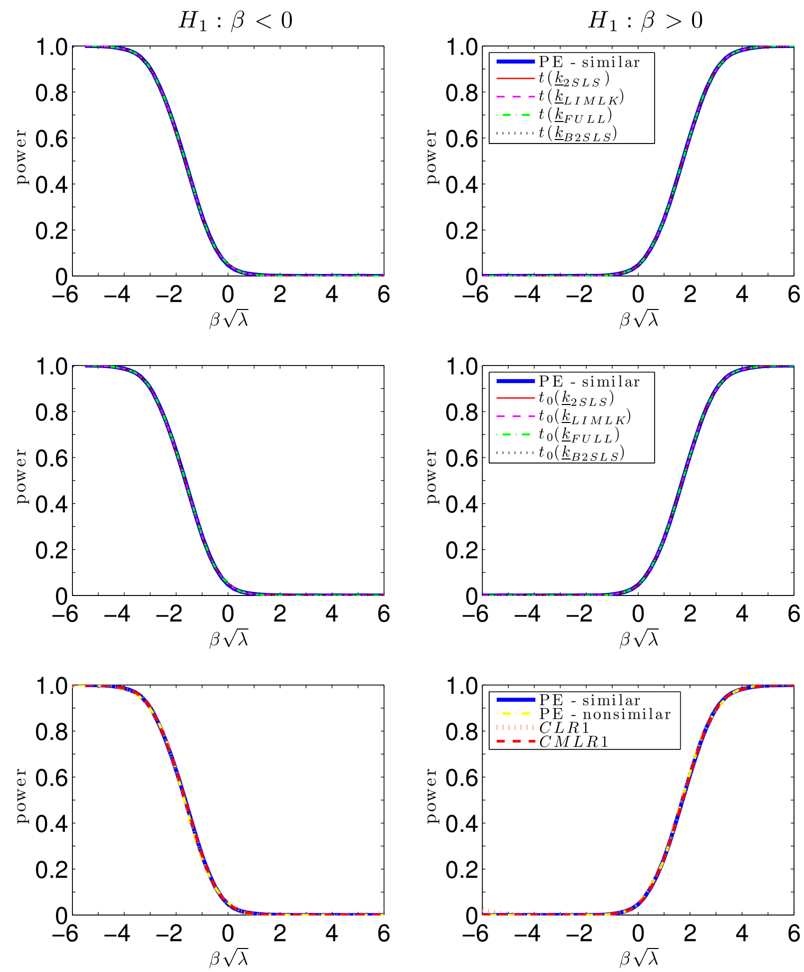
(a) $\lambda/k = 2$

(b) $\lambda/k = 4$

Figure 5: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 5$.

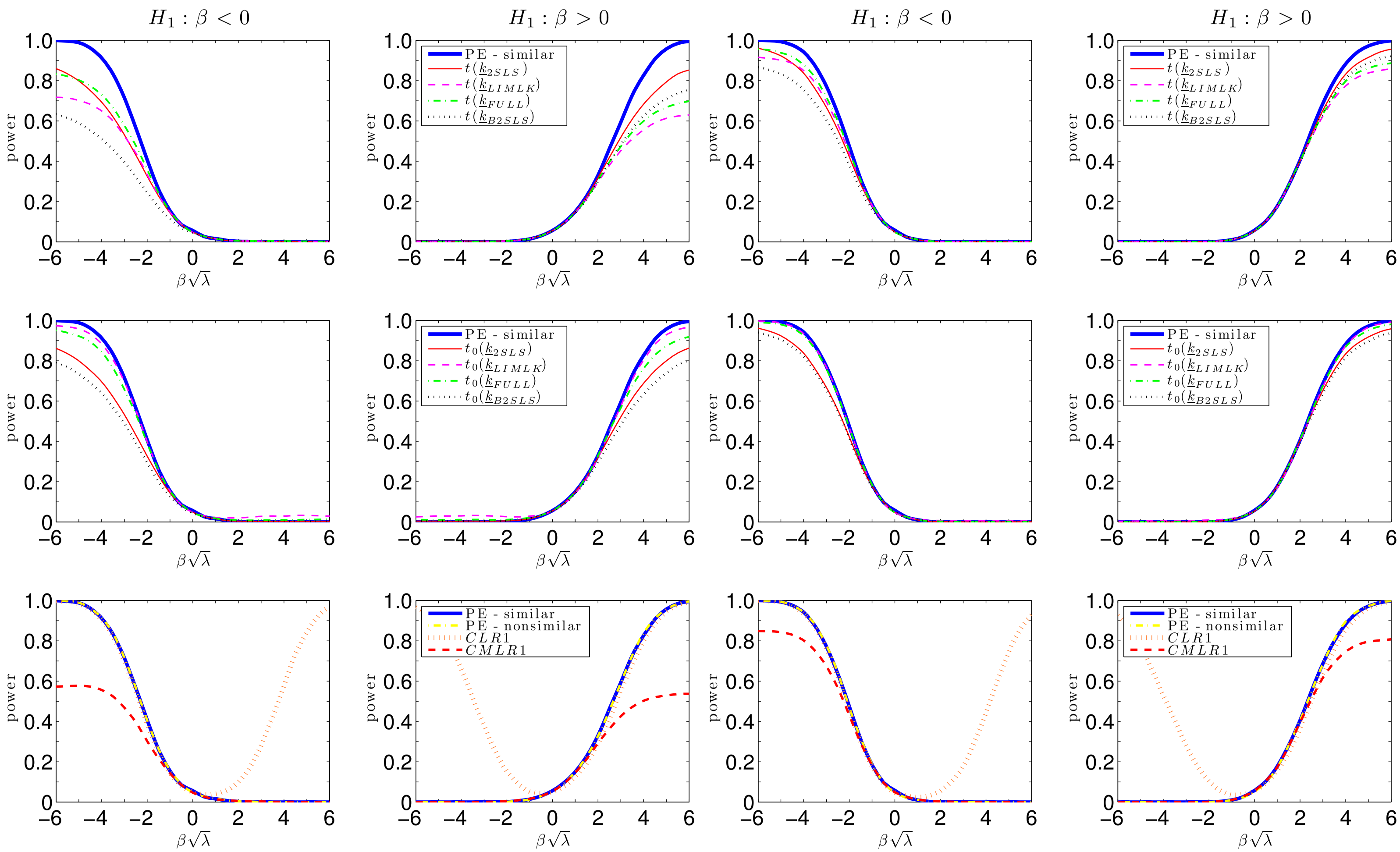


(a) $\lambda/k = 8$



(b) $\lambda/k = 16$

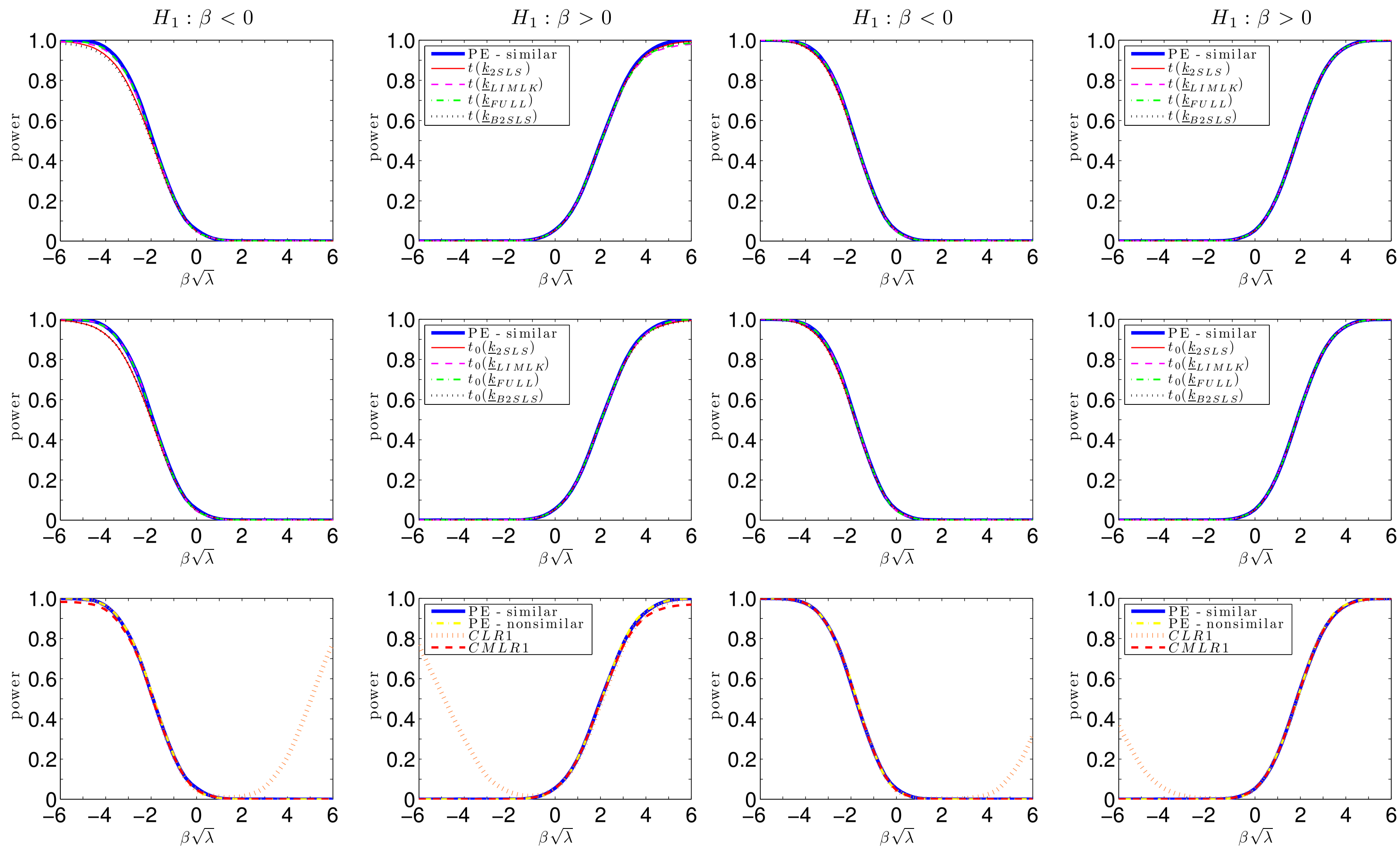
Figure 6: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 5$.



(a) $\lambda/\mathrm{k} = 0.5$

(b) $\lambda/\mathrm{k} = 1$

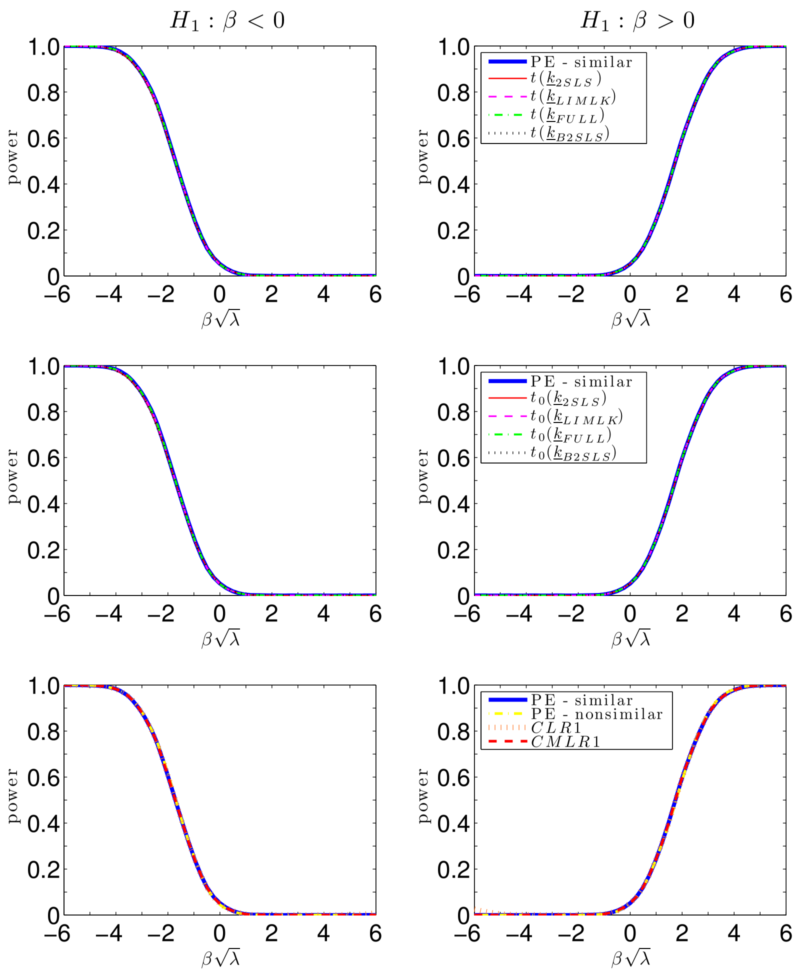
Figure 7: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 10$.



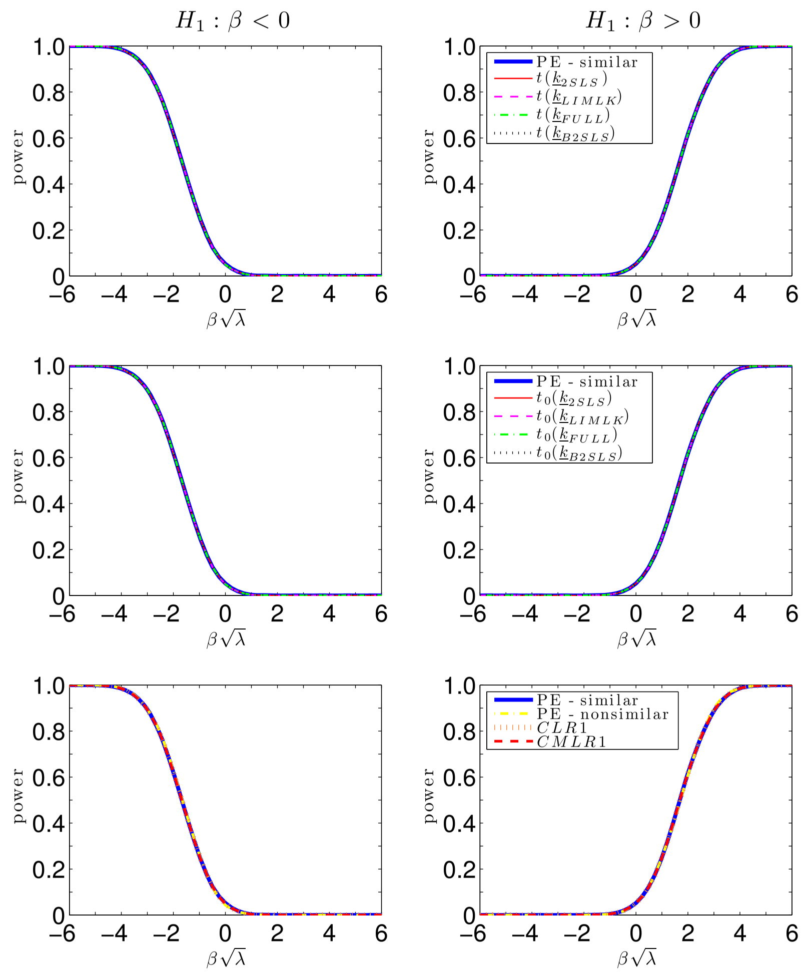
(a) $\lambda/\mathrm{k} = 2$

(b) $\lambda/\mathrm{k} = 4$

Figure 8: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 10$.

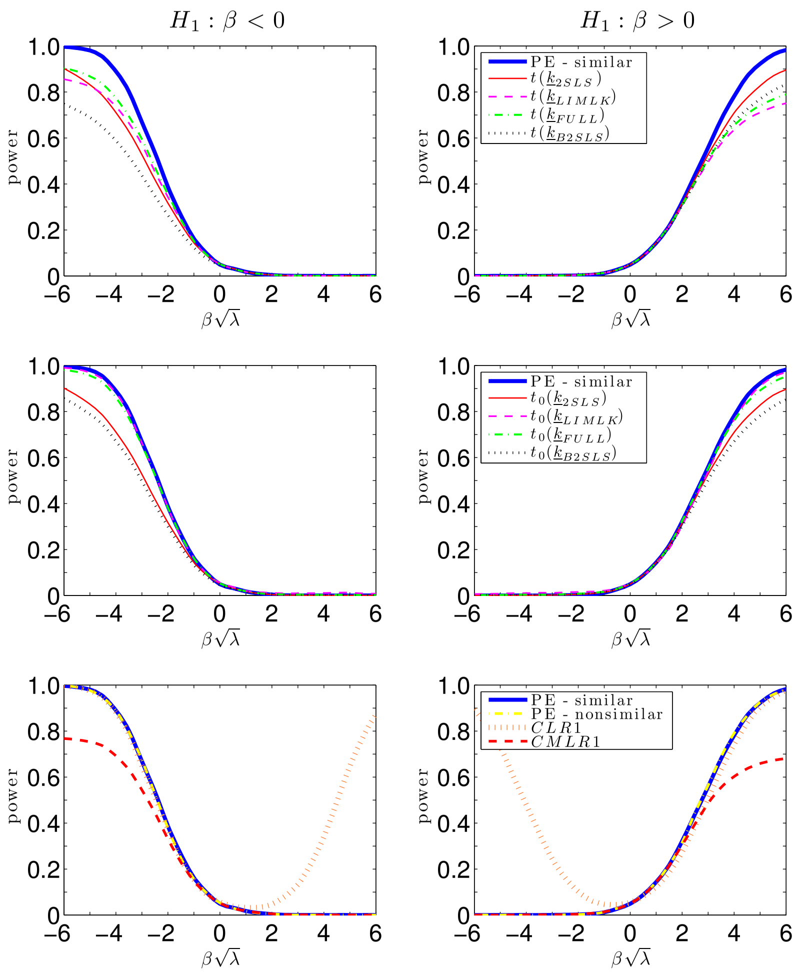


(a) $\lambda/k = 8$

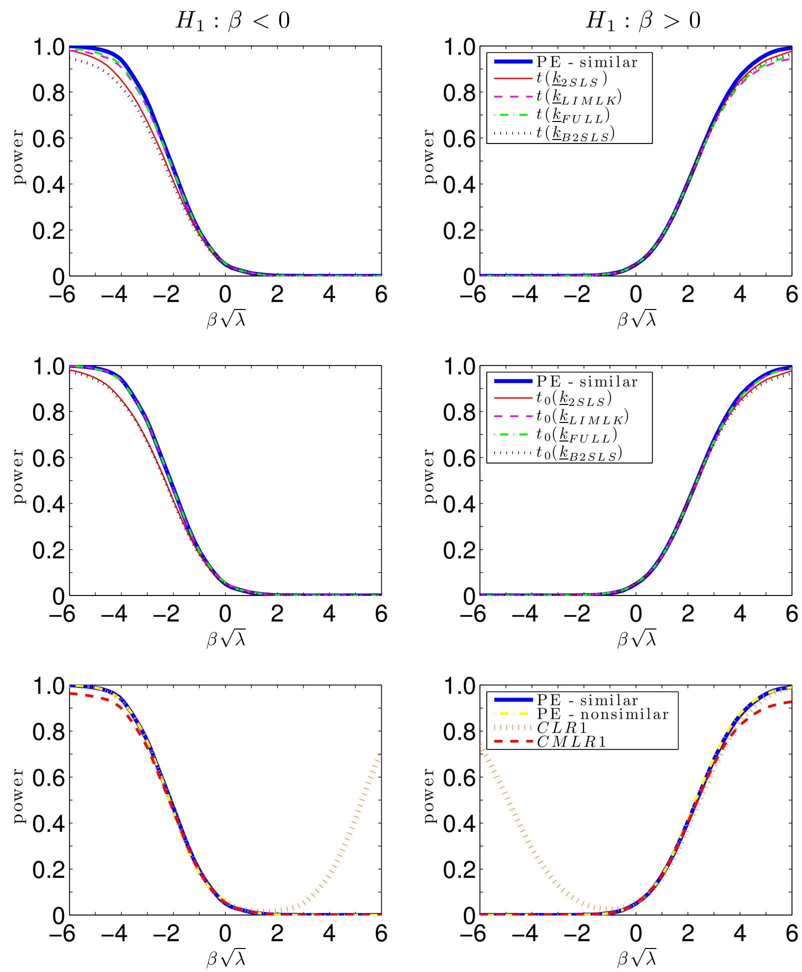


(b) $\lambda/k = 16$

Figure 9: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 10$.

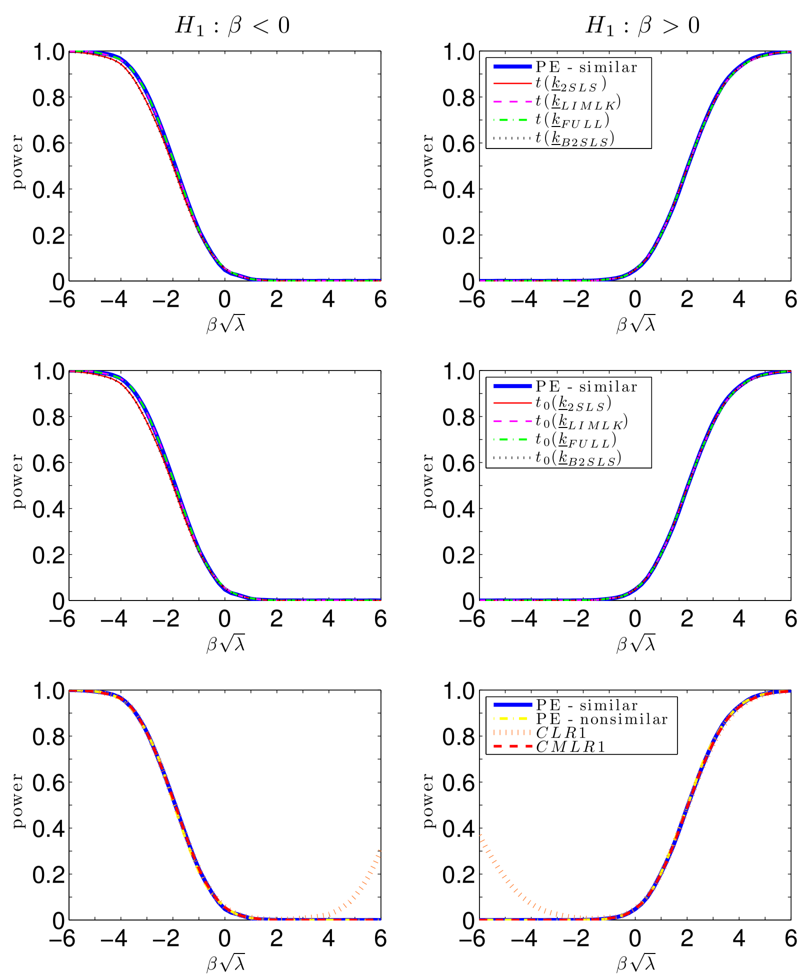


(a) $\lambda/\mathrm{k} = 0.5$

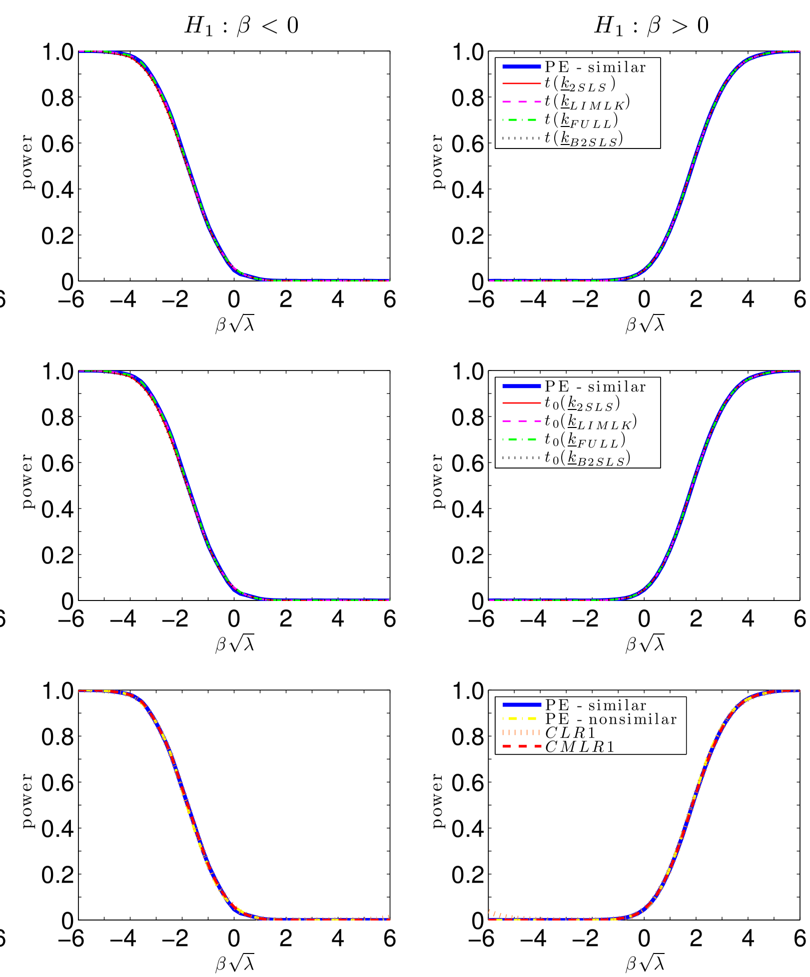


(b) $\lambda/\mathrm{k} = 1$

Figure 10: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 20$.

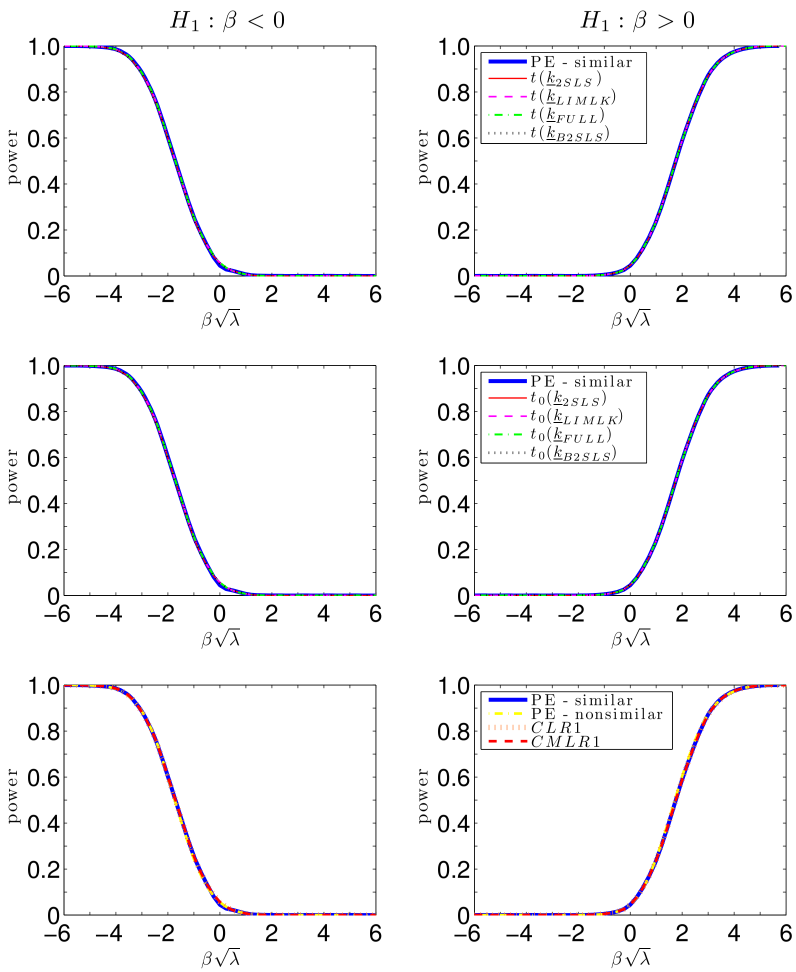


(a) $\lambda/\mathrm{k} = 2$

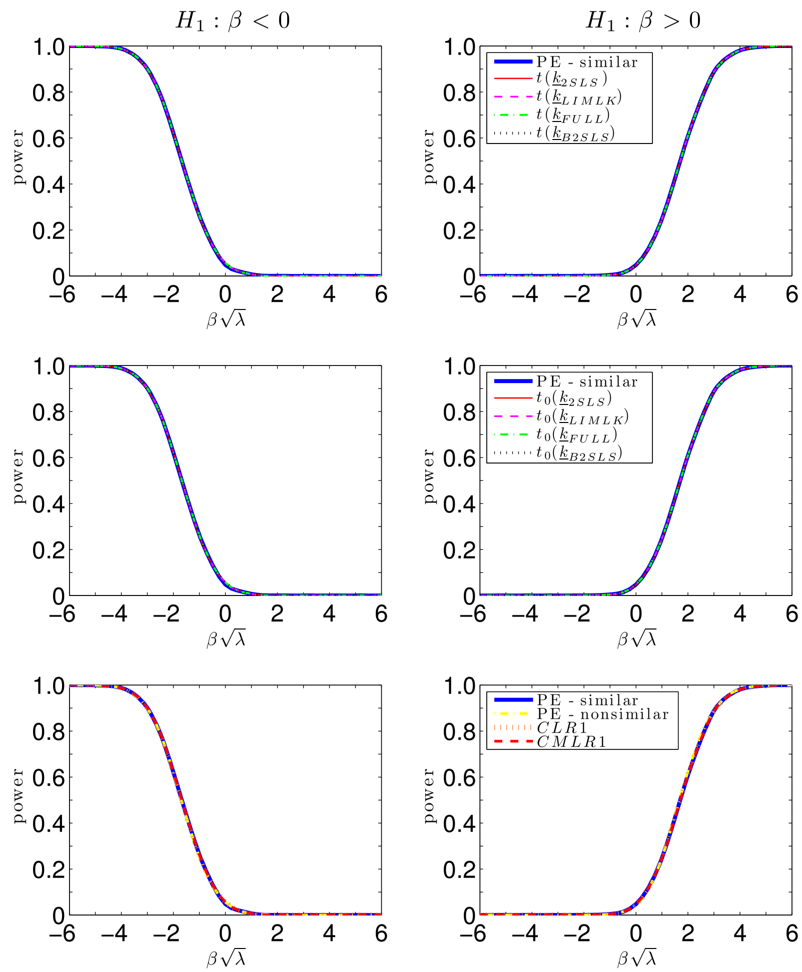


(b) $\lambda/\mathrm{k} = 4$

Figure 11: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 20$.

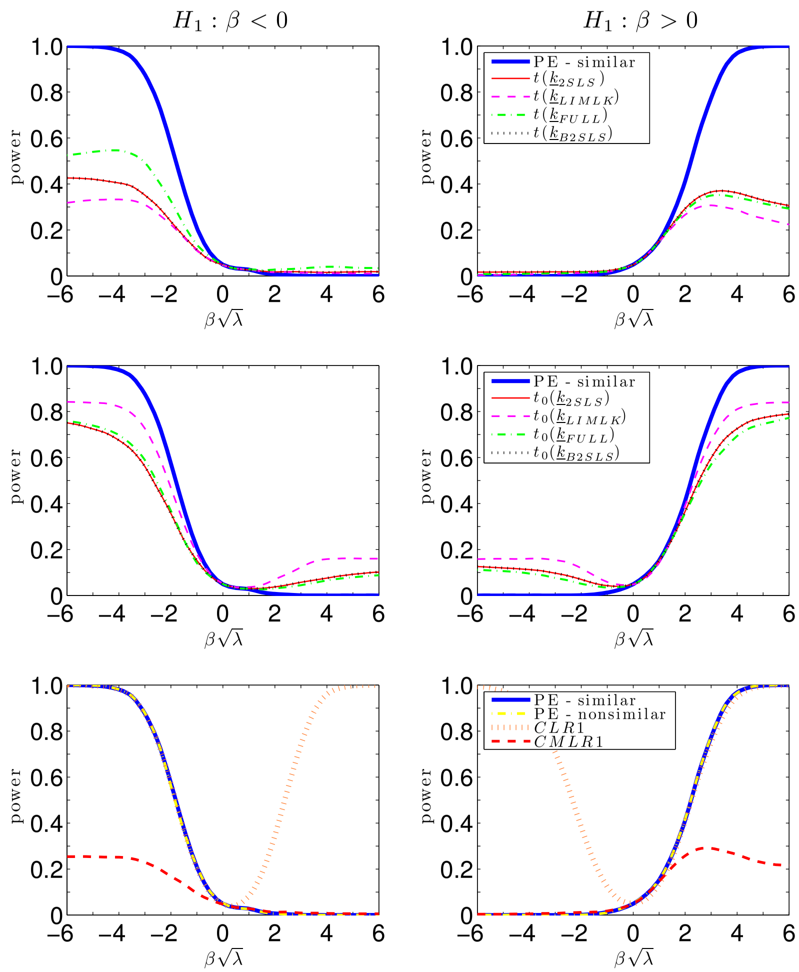


(a) $\lambda/k = 8$

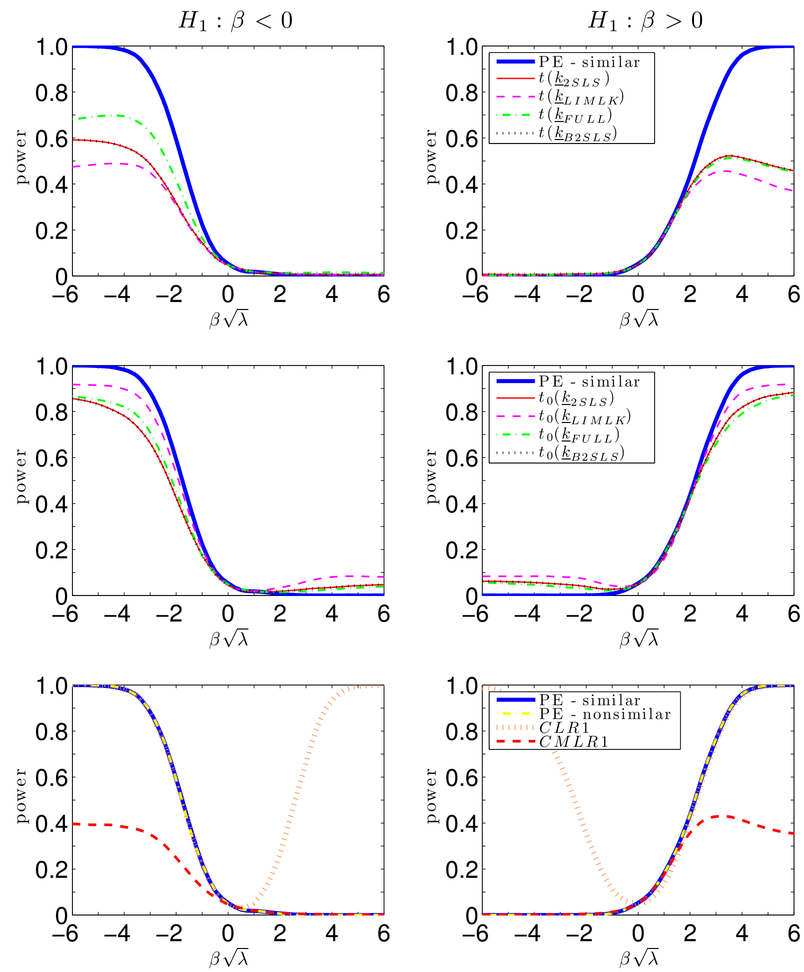


(b) $\lambda/k = 16$

Figure 12: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.2$, $k = 20$.

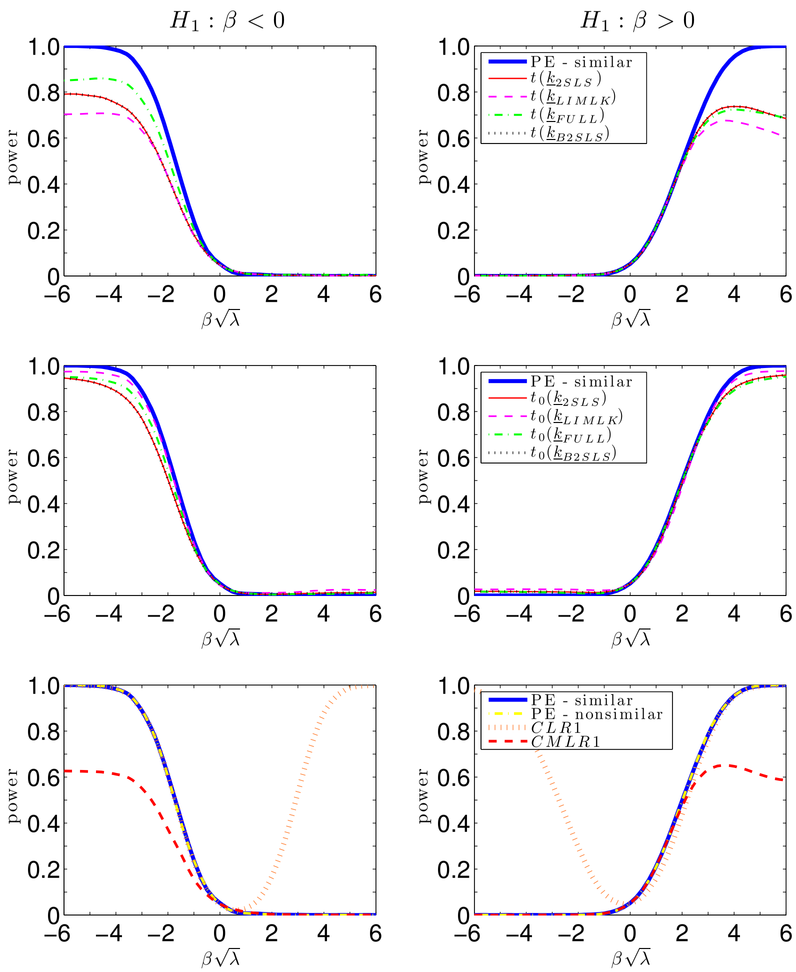


(a) $\lambda/k = 0.5$

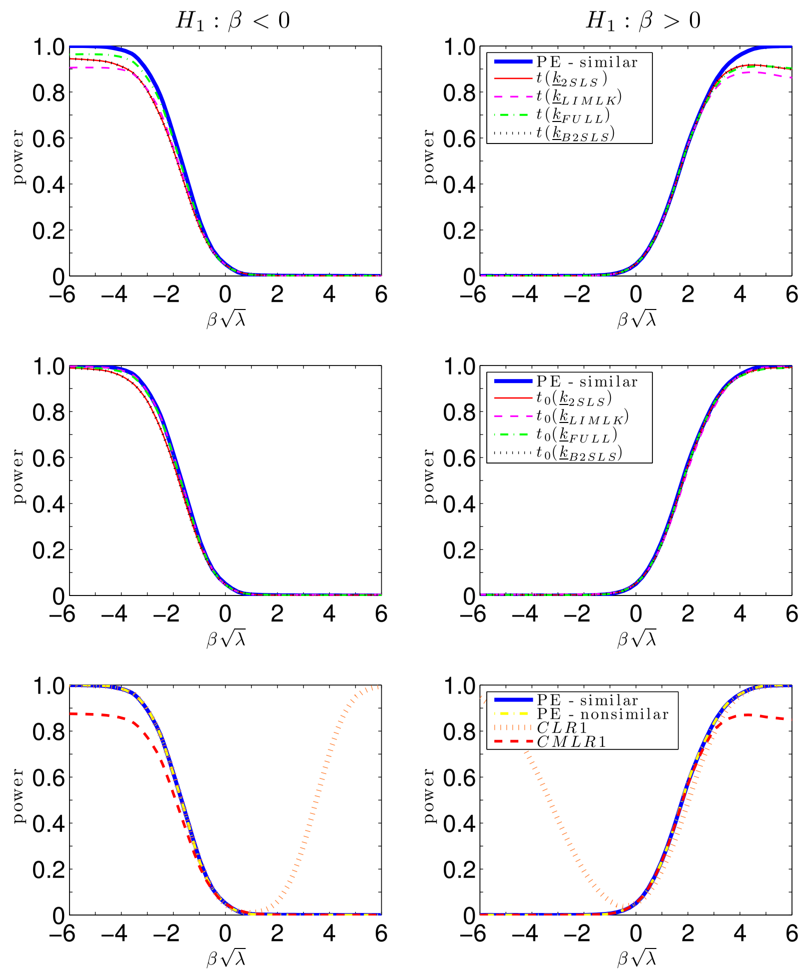


(b) $\lambda/k = 1$

Figure 13: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 2$.

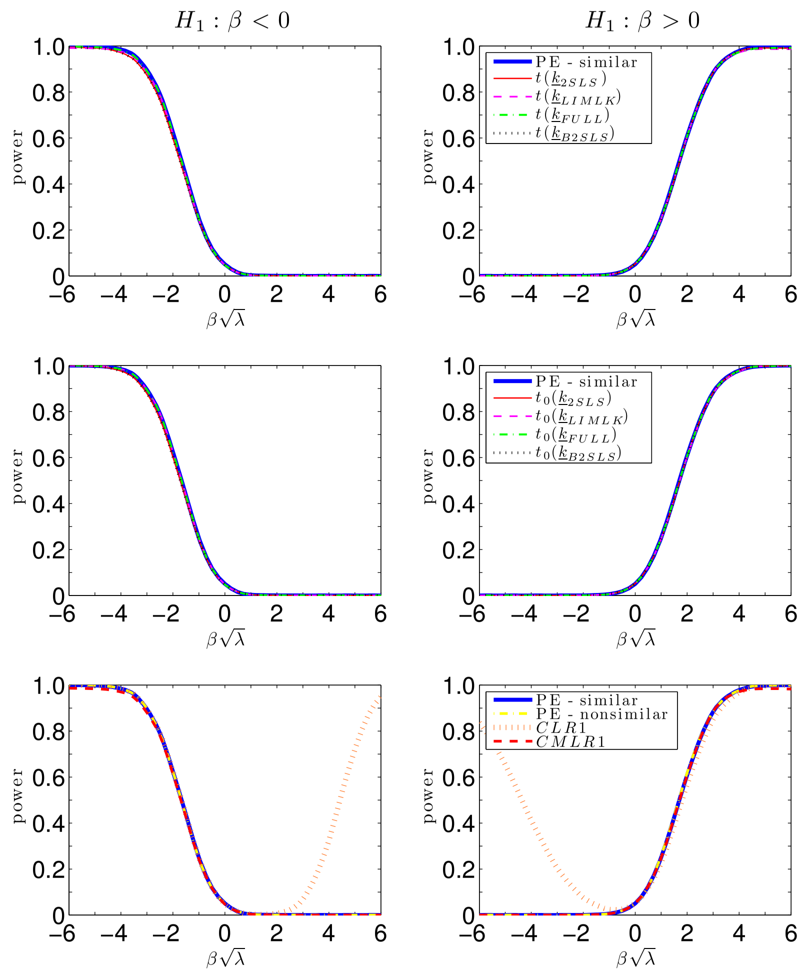


(a) $\lambda/\mathrm{k} = 2$

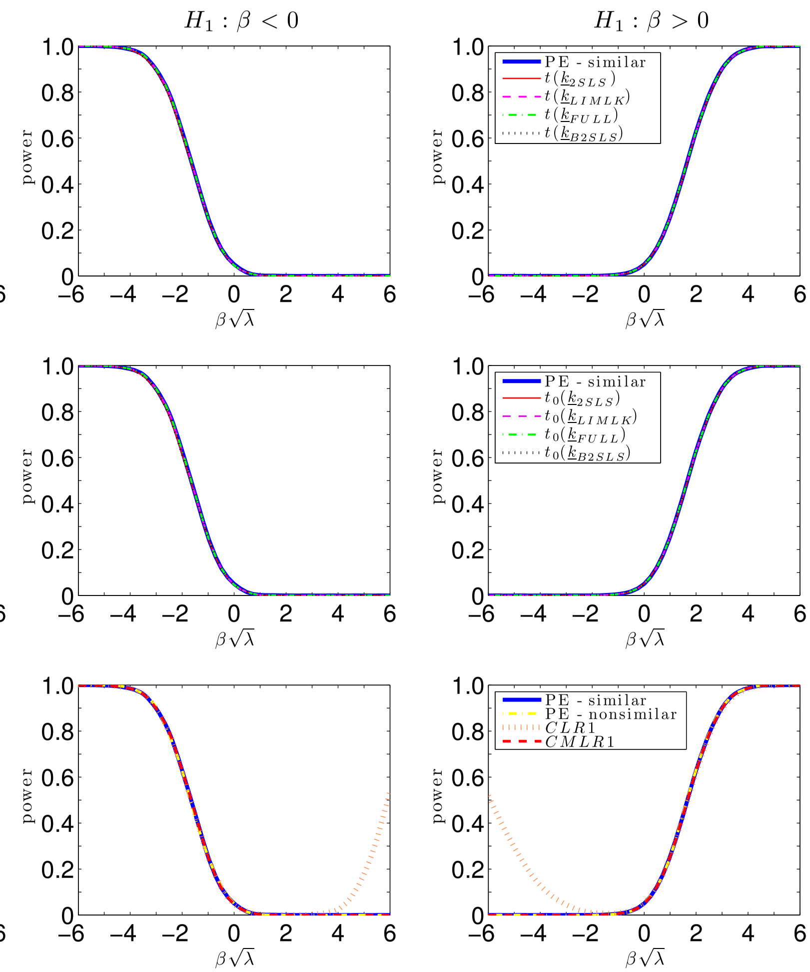


(b) $\lambda/\mathrm{k} = 4$

Figure 14: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 2$.

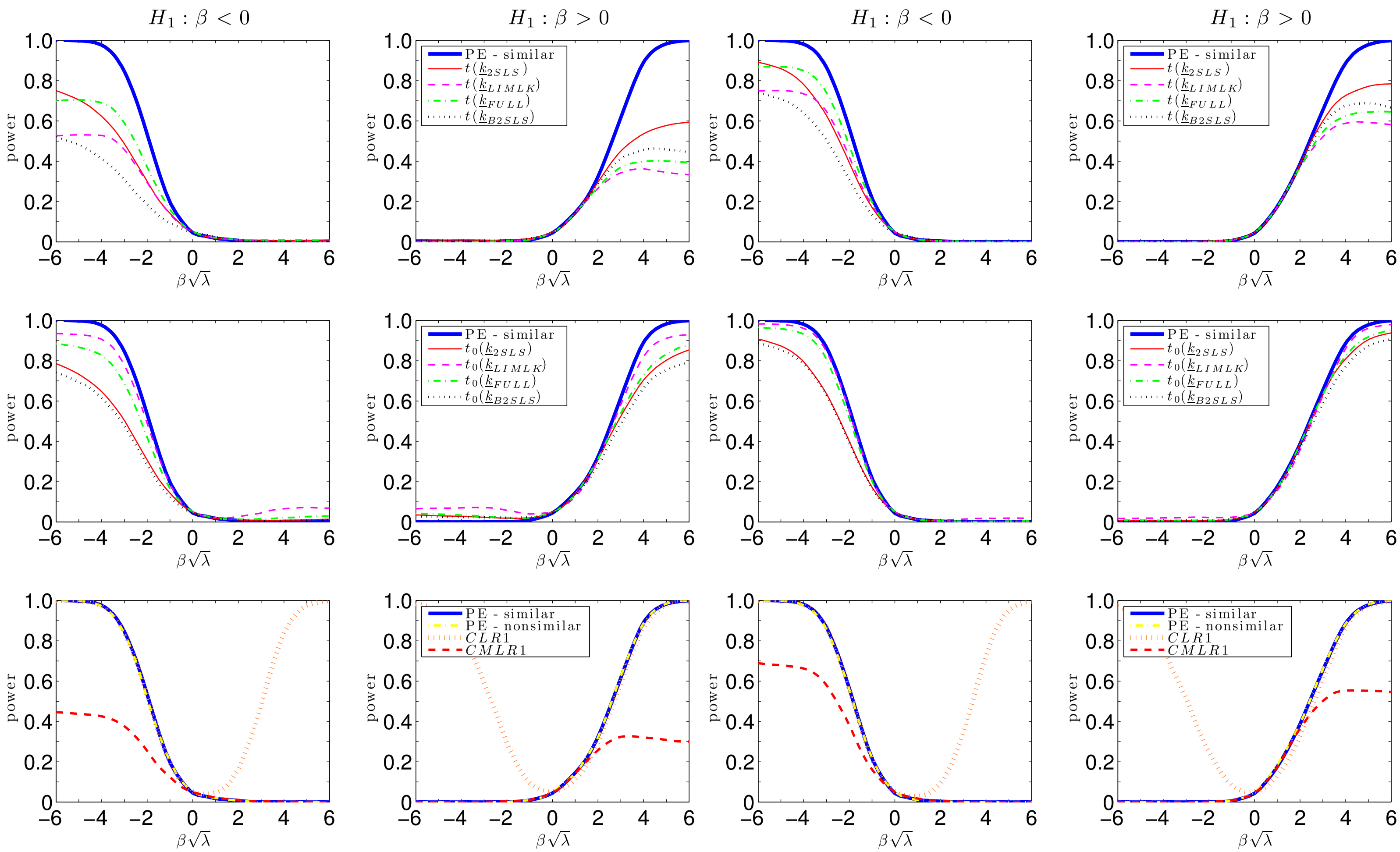


(a) $\lambda/k = 8$



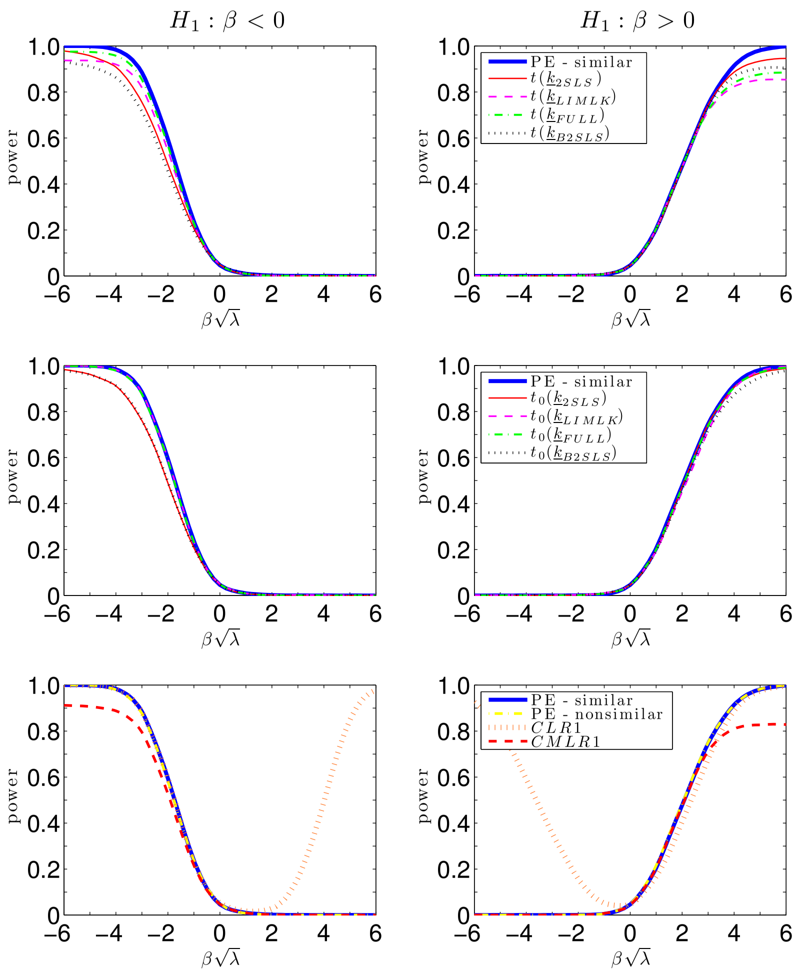
(b) $\lambda/k = 16$

Figure 15: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 2$.

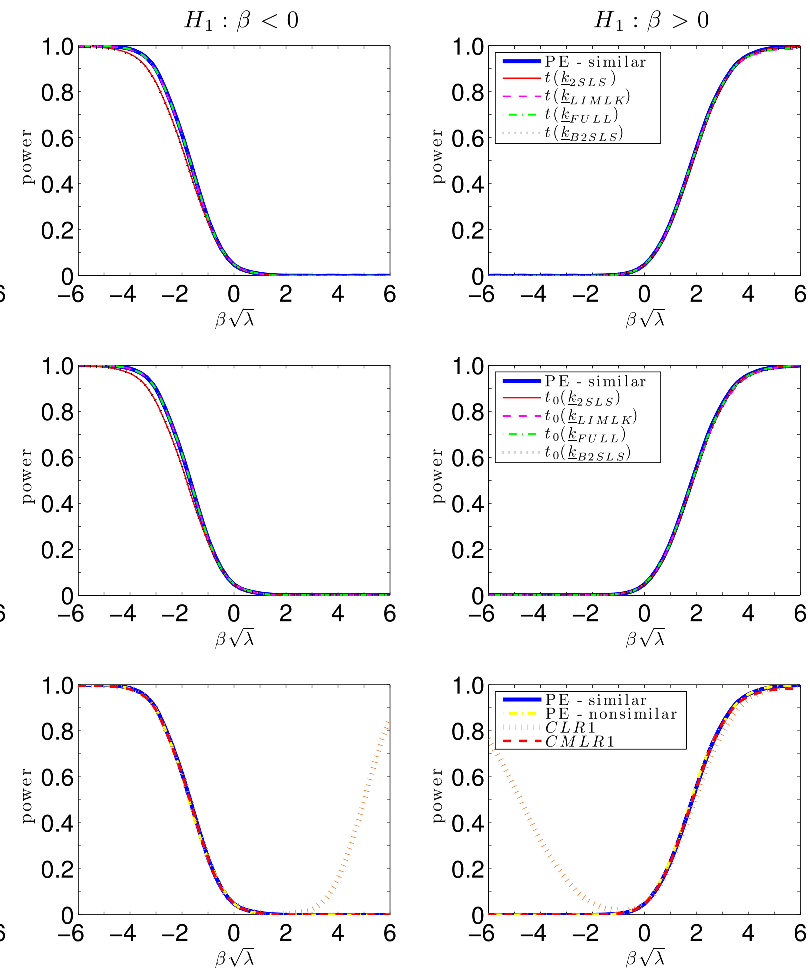


(a) $\lambda/k = 0.5$ (b) $\lambda/k = 1$

Figure 16: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 5$.

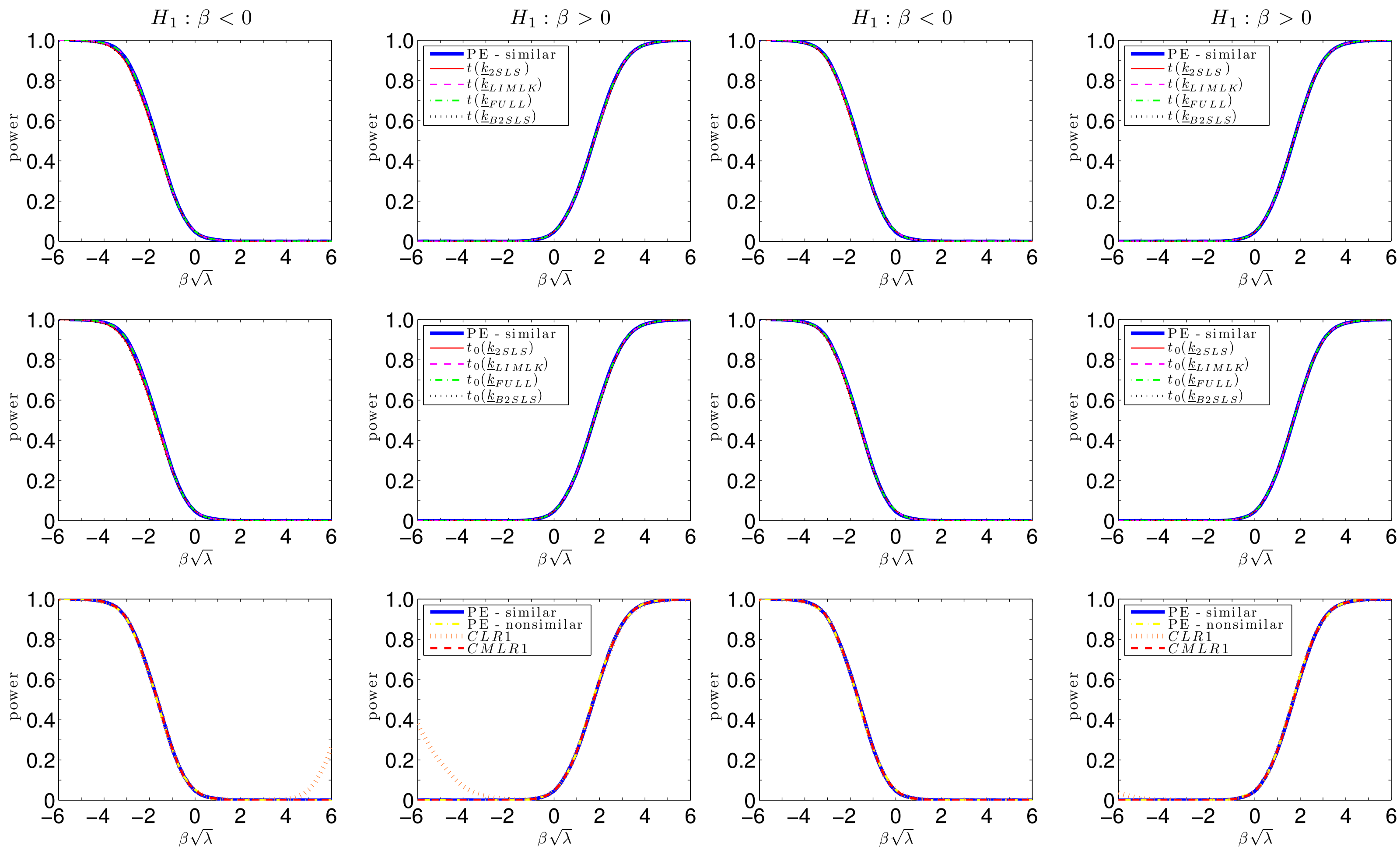


(a) $\lambda/\mathrm{k} = 2$



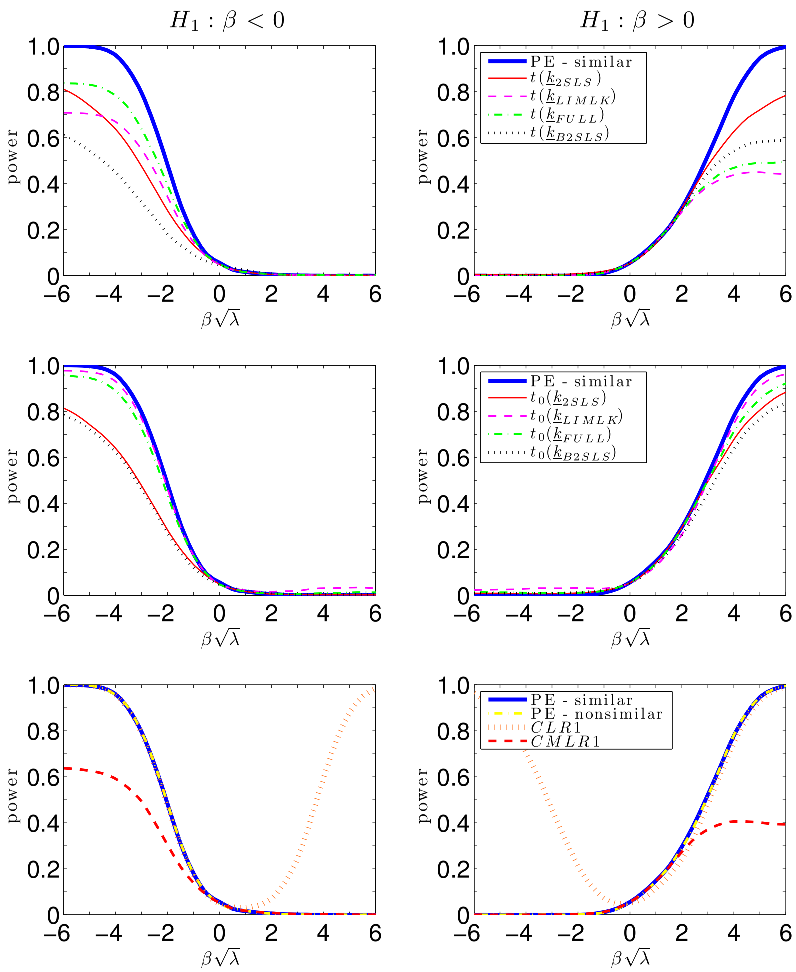
(b) $\lambda/\mathrm{k} = 4$

Figure 17: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 5$.

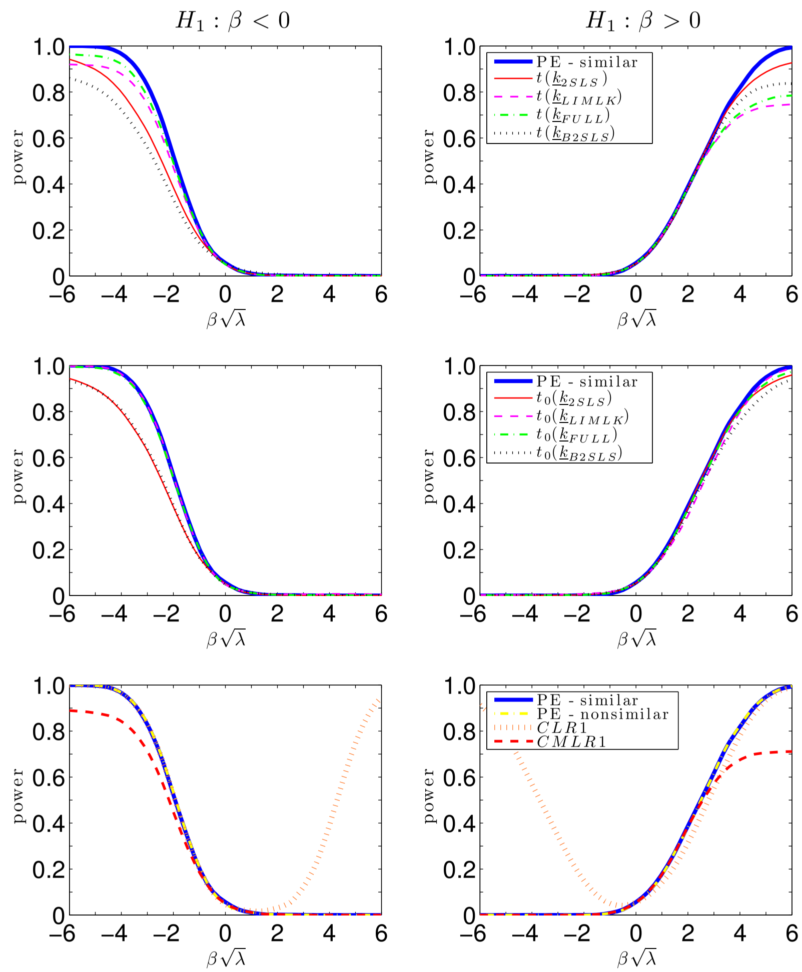


(a) $\lambda/k = 8$ (b) $\lambda/k = 16$

Figure 18: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 5$.

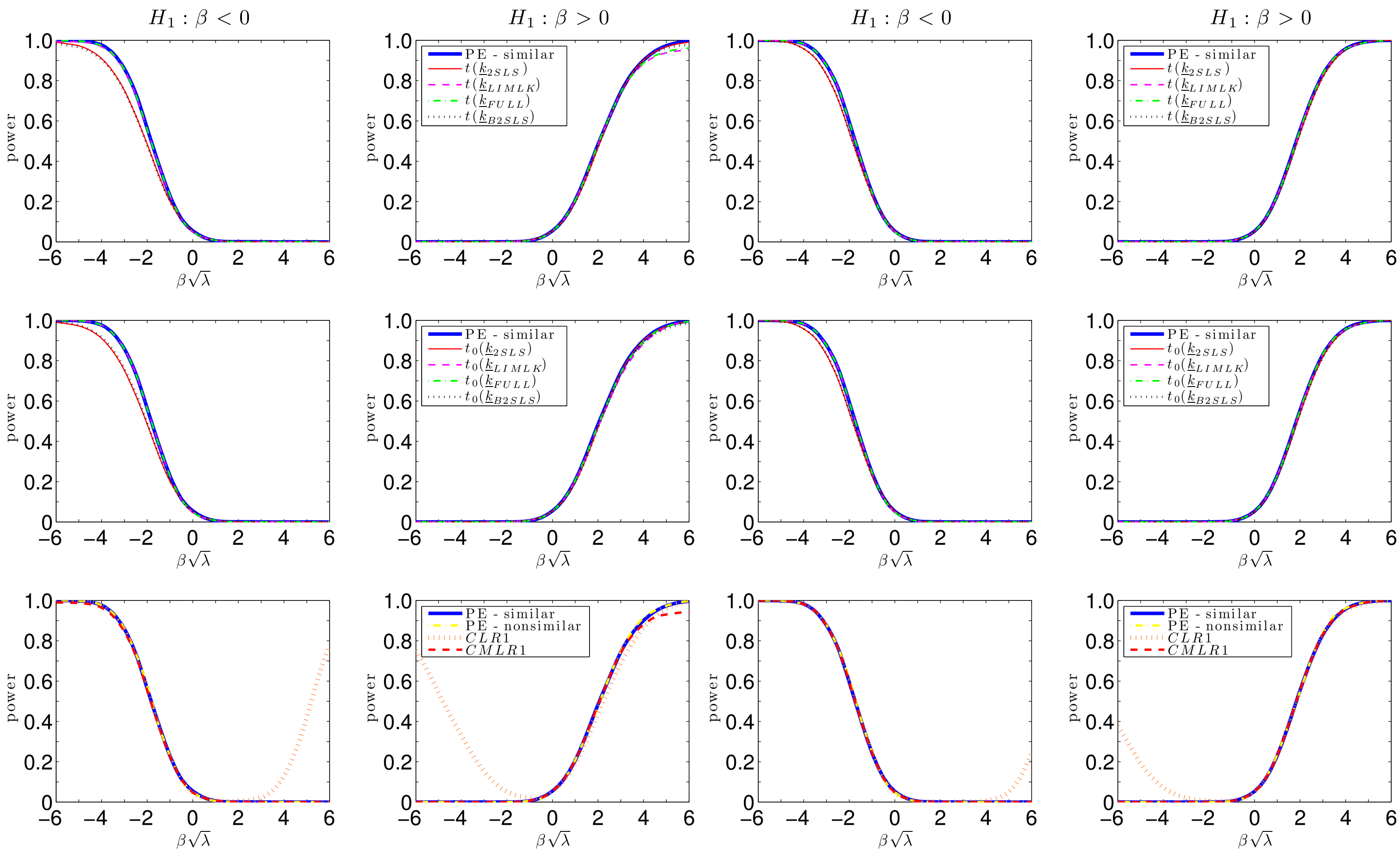


(a) $\lambda/\mathrm{k} = 0.5$



(b) $\lambda/\mathrm{k} = 1$

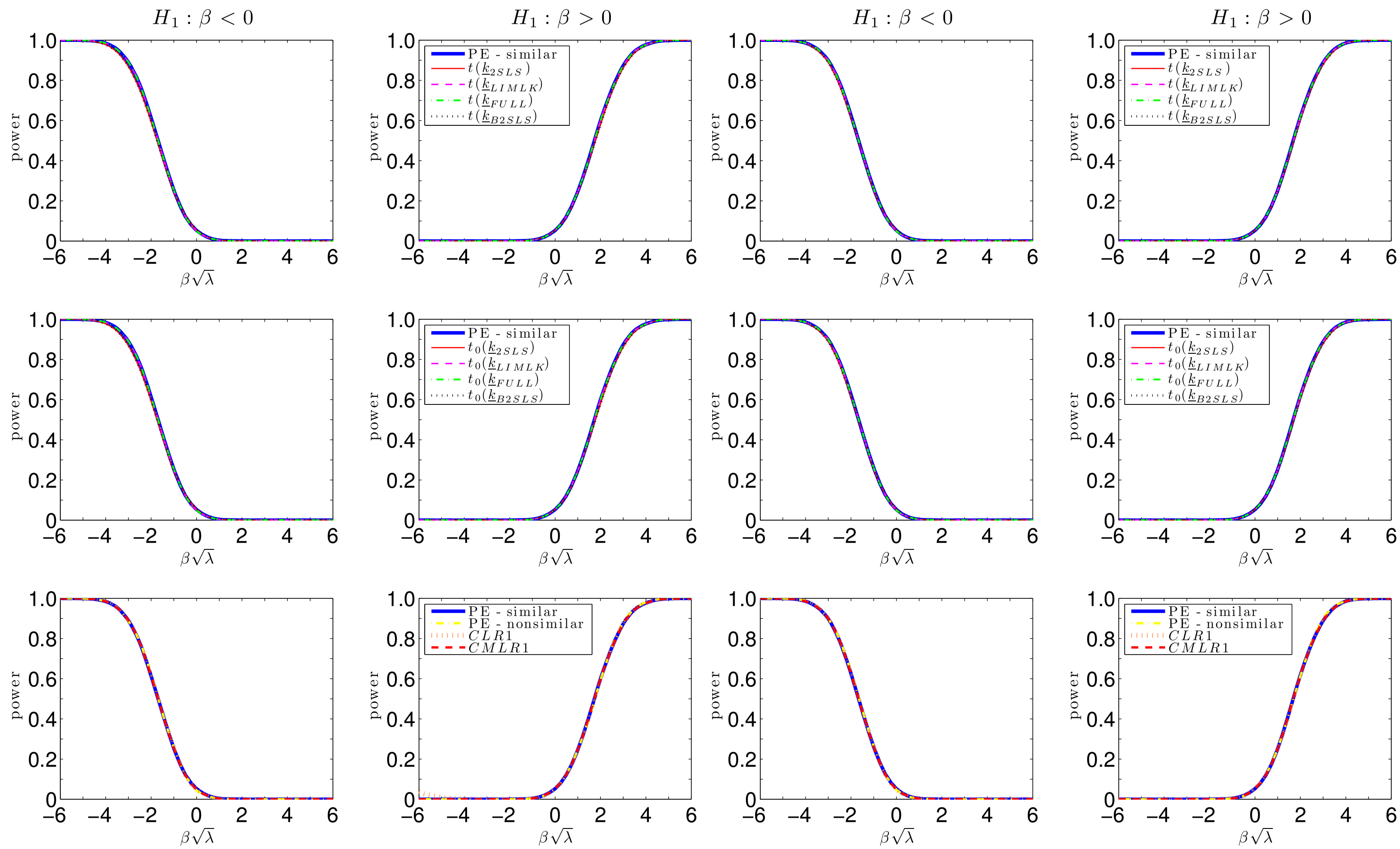
Figure 19: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 10$.



(a) $\lambda/\mathrm{k} = 2$

(b) $\lambda/\mathrm{k} = 4$

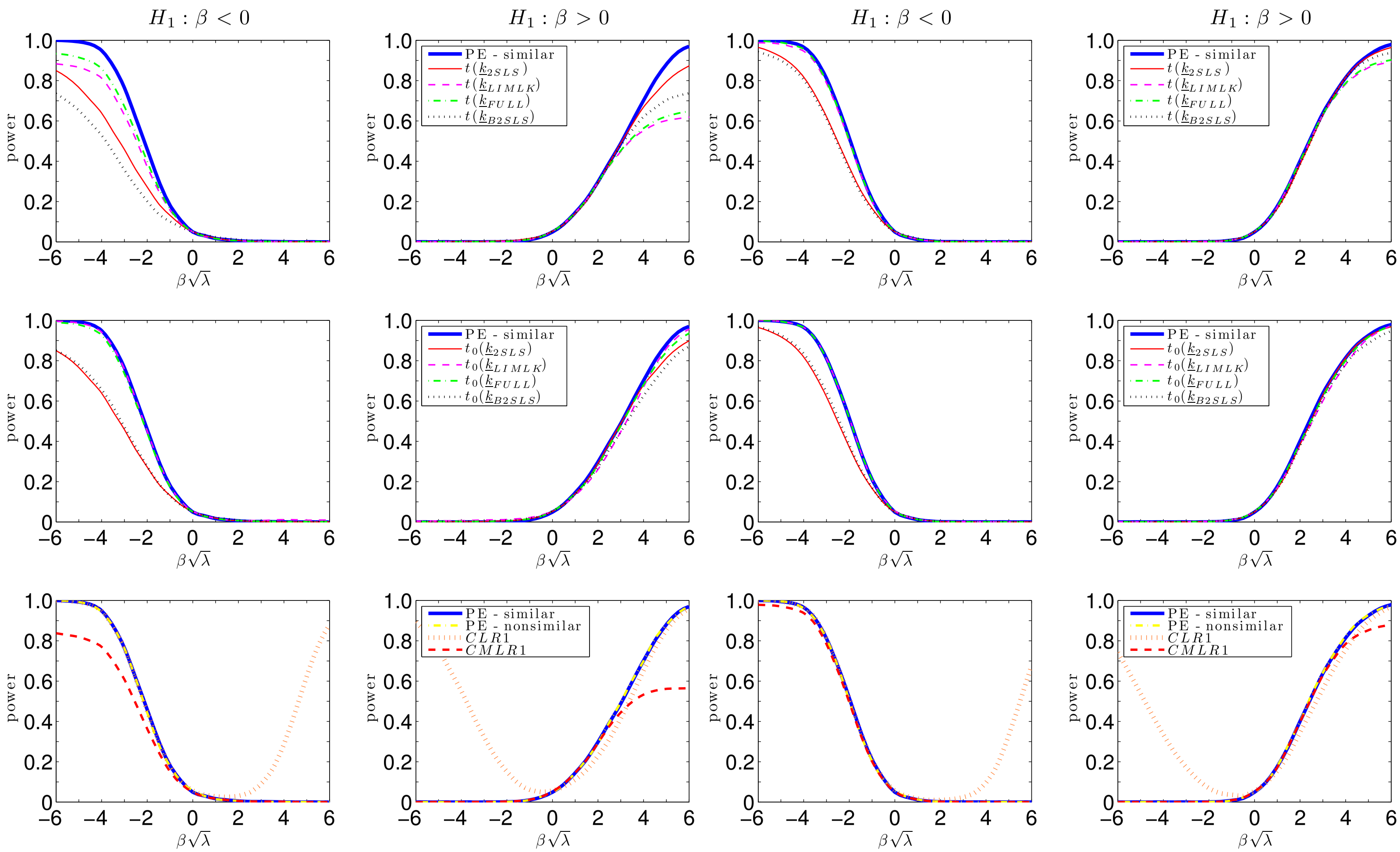
Figure 20: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 10$.



(a) $\lambda/\mathrm{k} = 8$

(b) $\lambda/\mathrm{k} = 16$

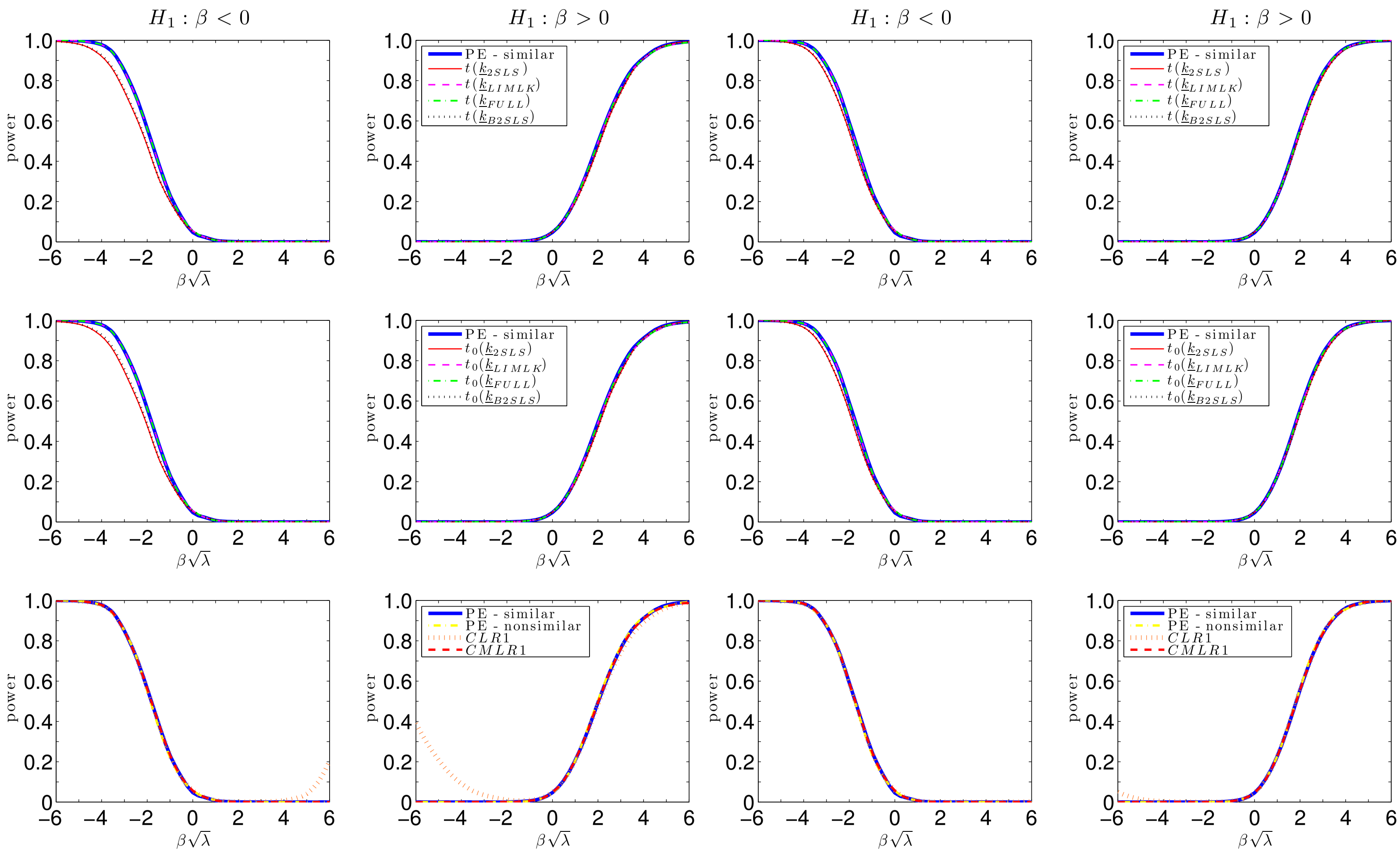
Figure 21: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 10$.



(a) $\lambda/k = 0.5$

(b) $\lambda/k = 1$

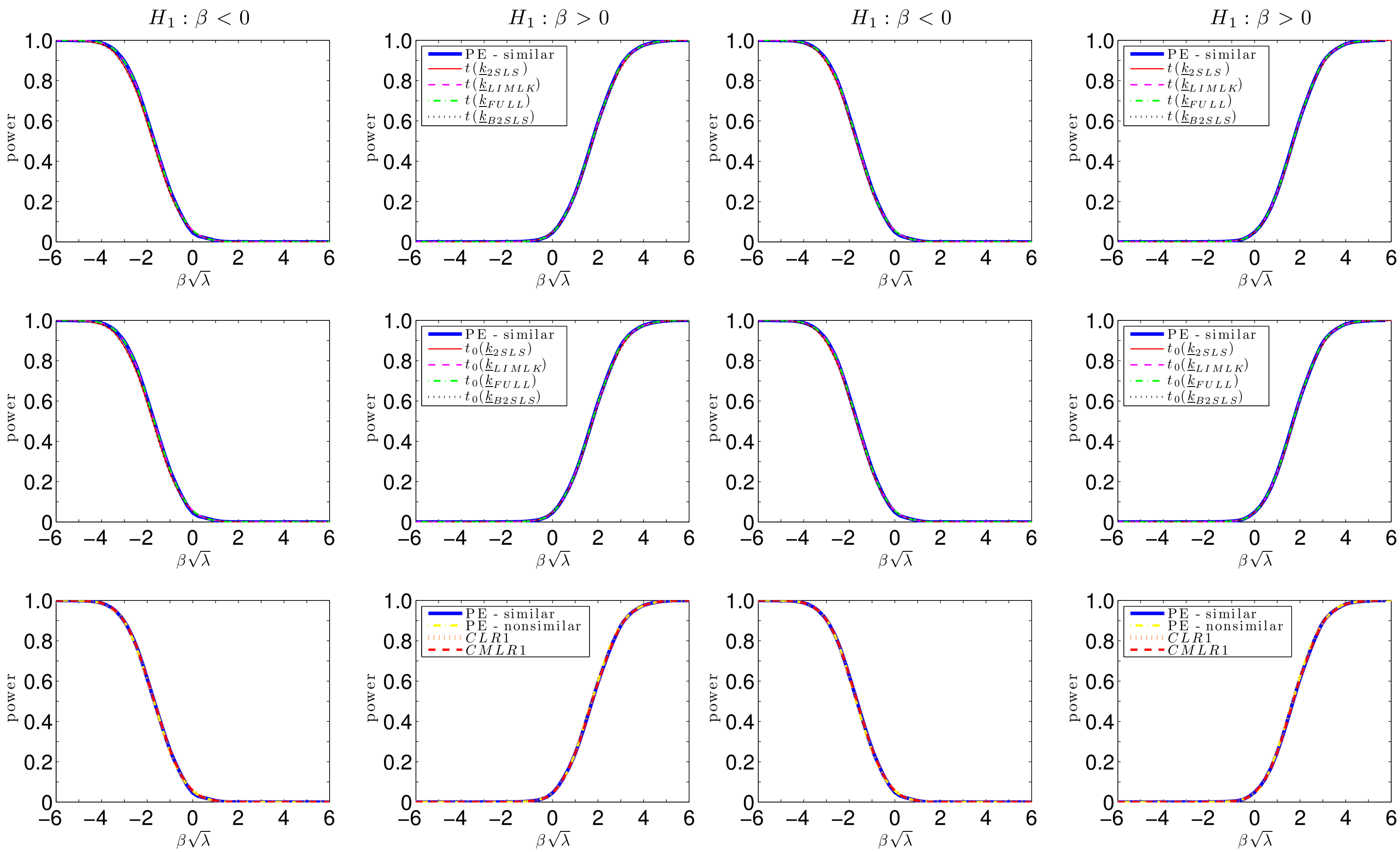
Figure 22: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 20$.



(a) $\lambda/\mathrm{k} = 2$

(b) $\lambda/\mathrm{k} = 4$

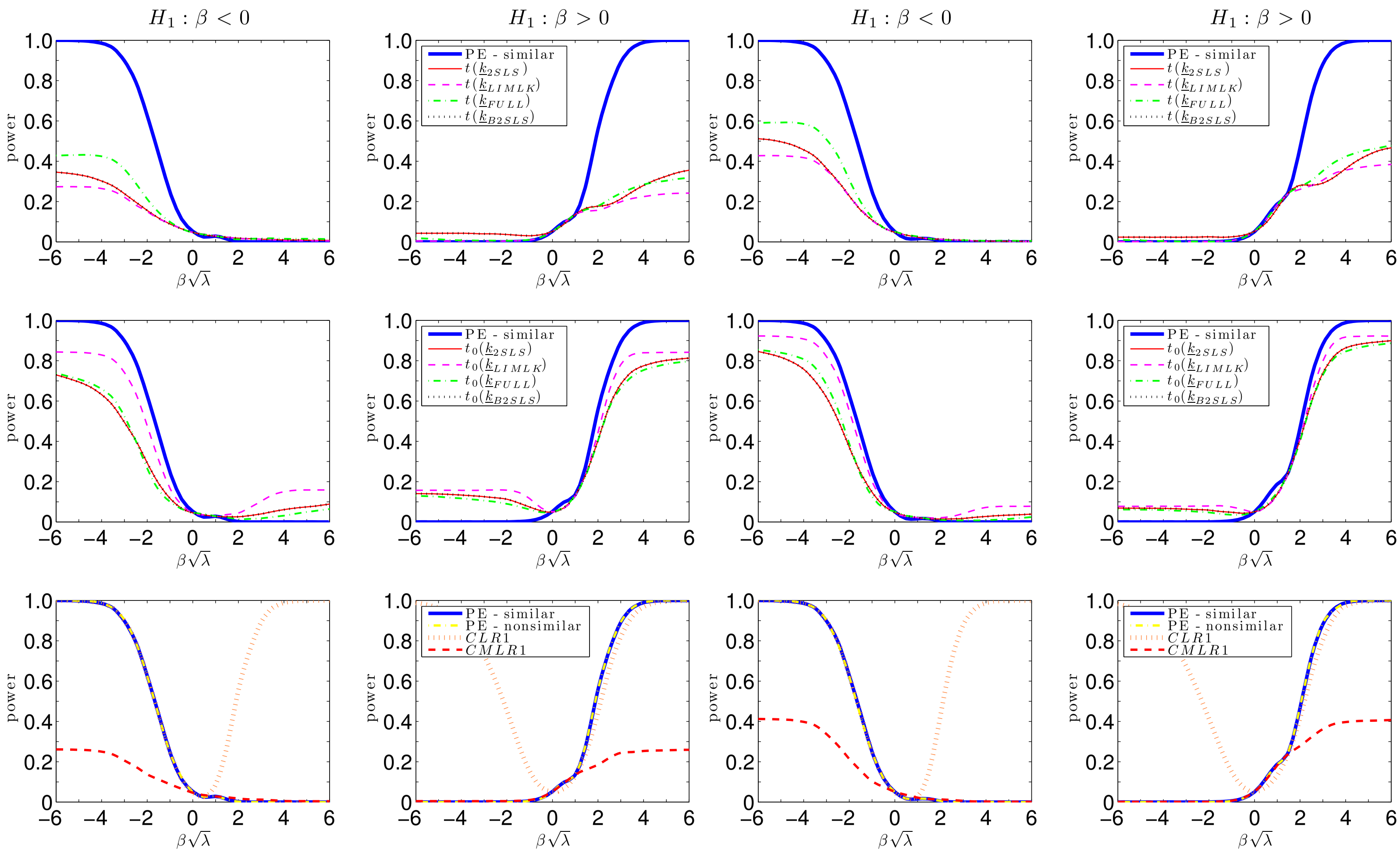
Figure 23: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 20$.



(a) $\lambda/k = 8$

(b) $\lambda/k = 16$

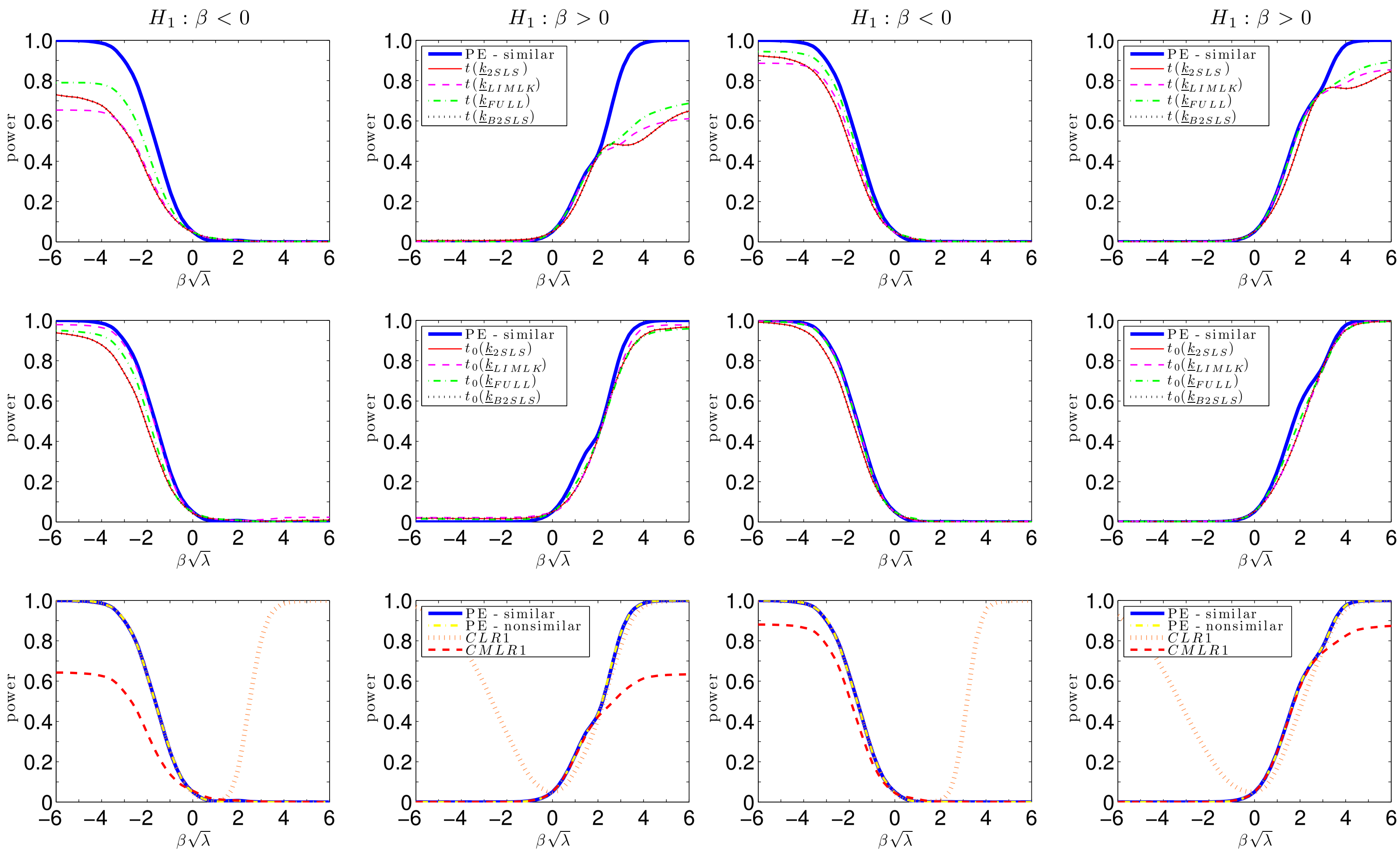
Figure 24: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.5$, $k = 20$.



(a) $\lambda/k = 0.5$

(b) $\lambda/k = 1$

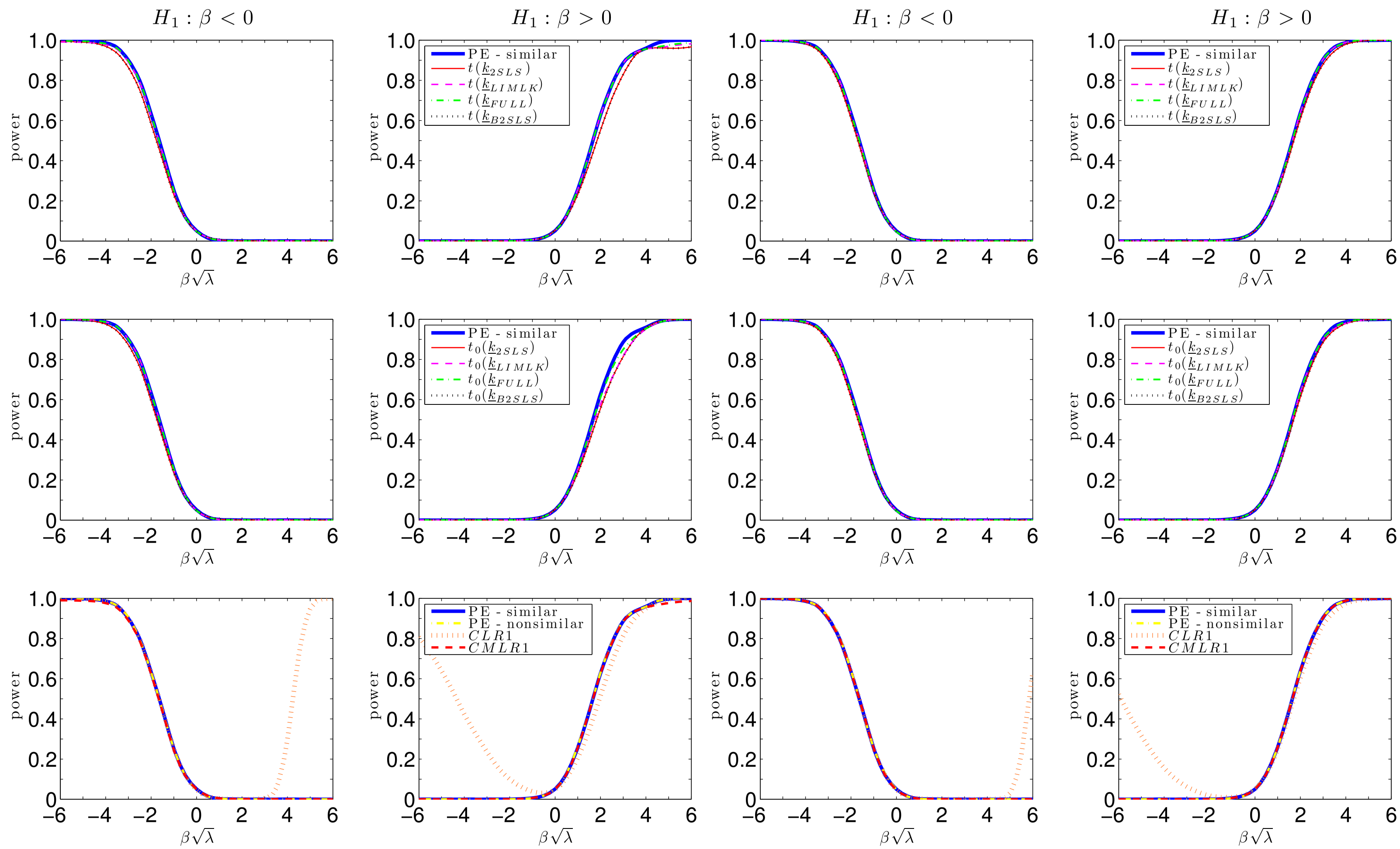
Figure 25: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 2$.



(a) $\lambda/\mathrm{k} = 2$

(b) $\lambda/\mathrm{k} = 4$

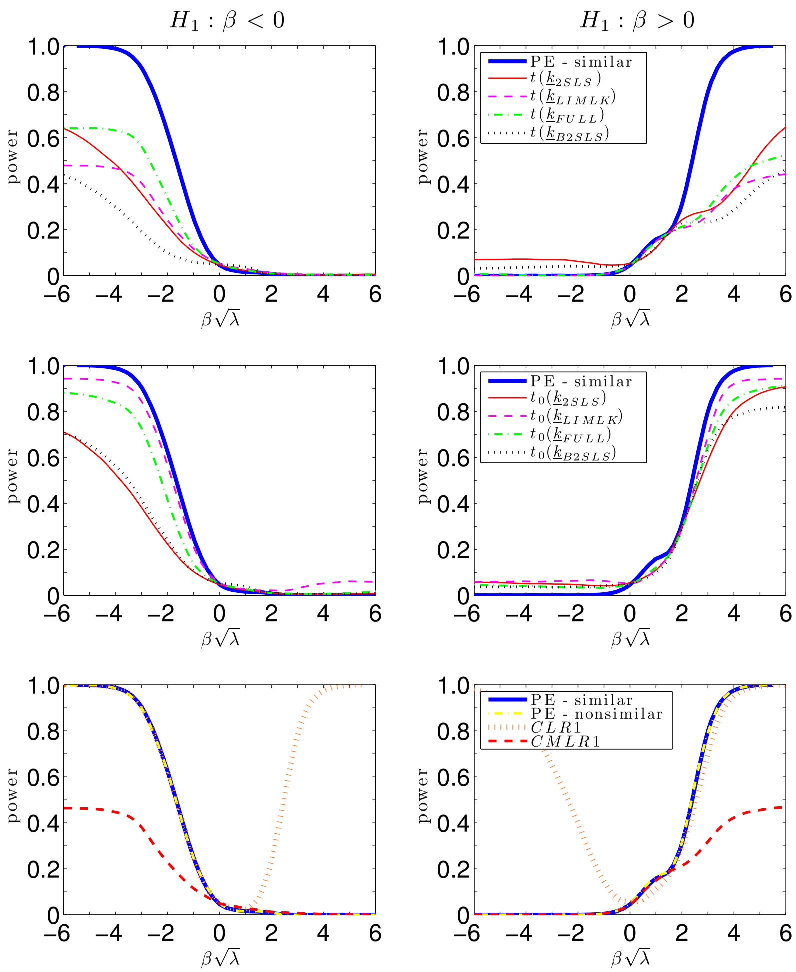
Figure 26: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 2$.



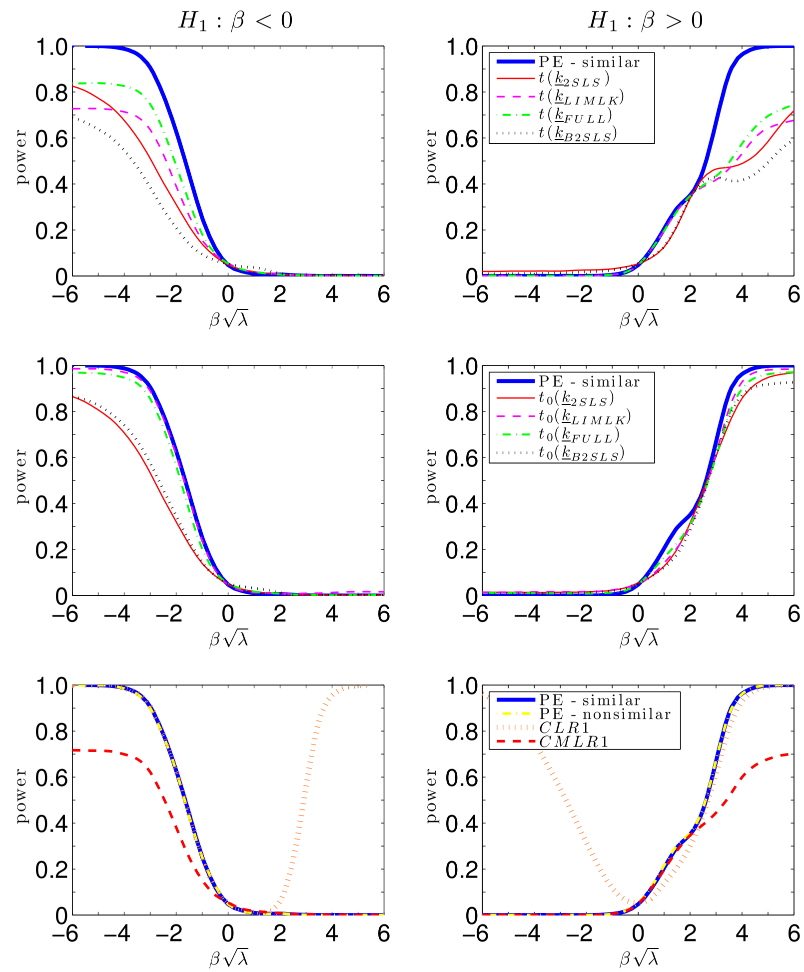
(a) $\lambda/k = 8$

(b) $\lambda/k = 16$

Figure 27: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 2$.

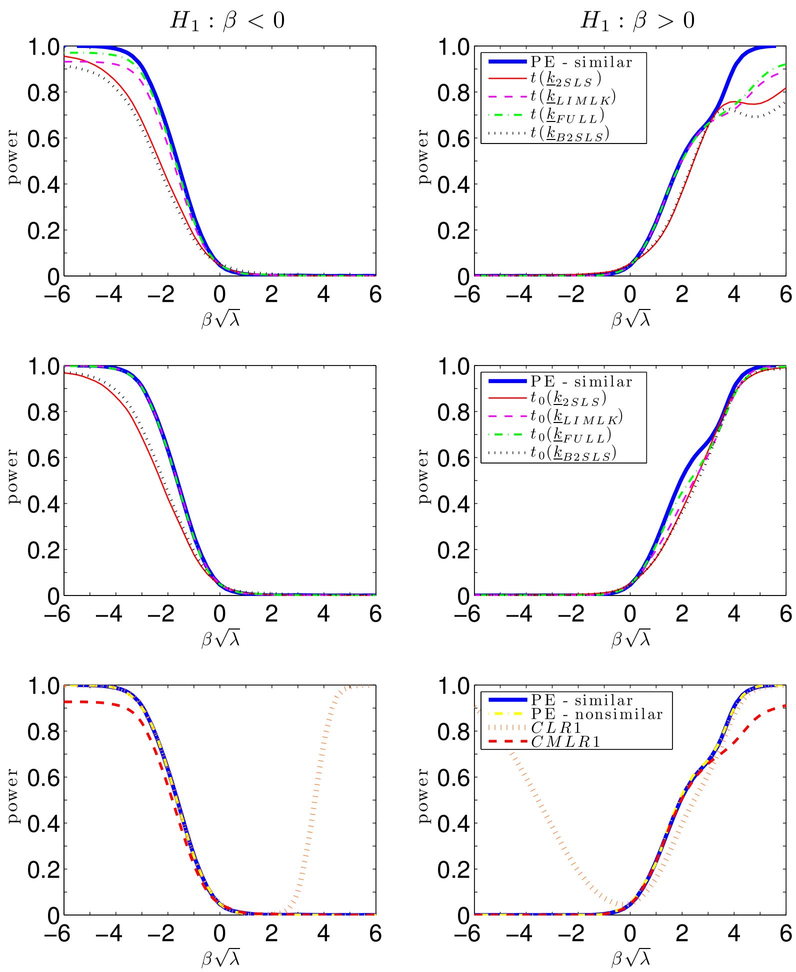


(a) $\lambda/k = 0.5$

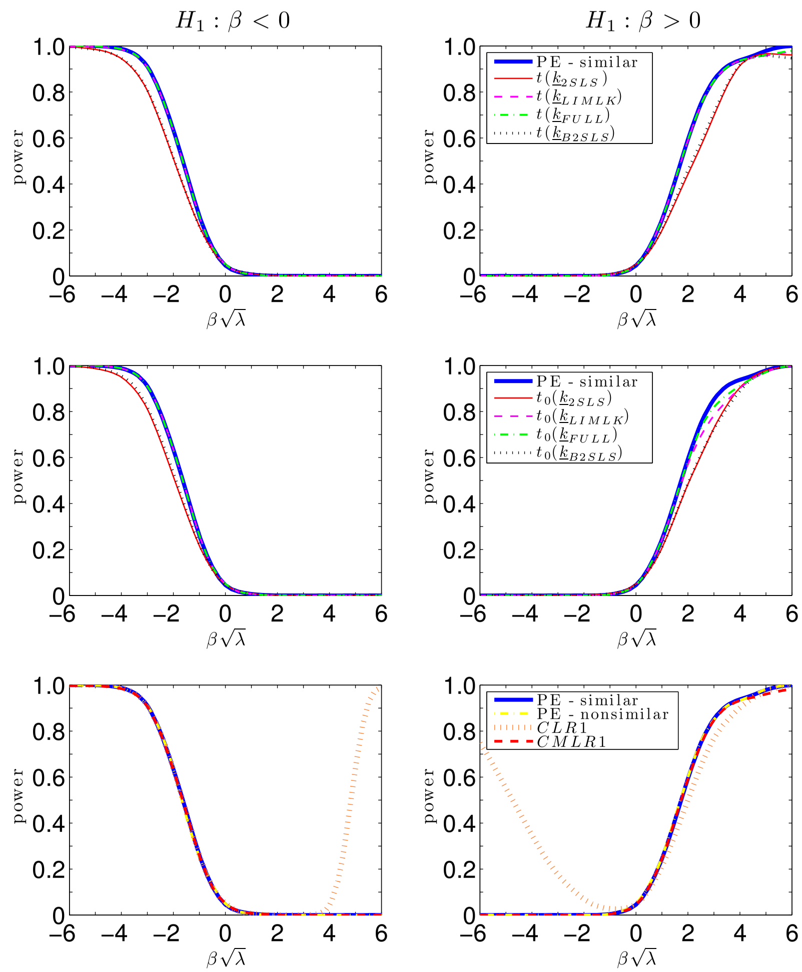


(b) $\lambda/k = 1$

Figure 28: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 5$.

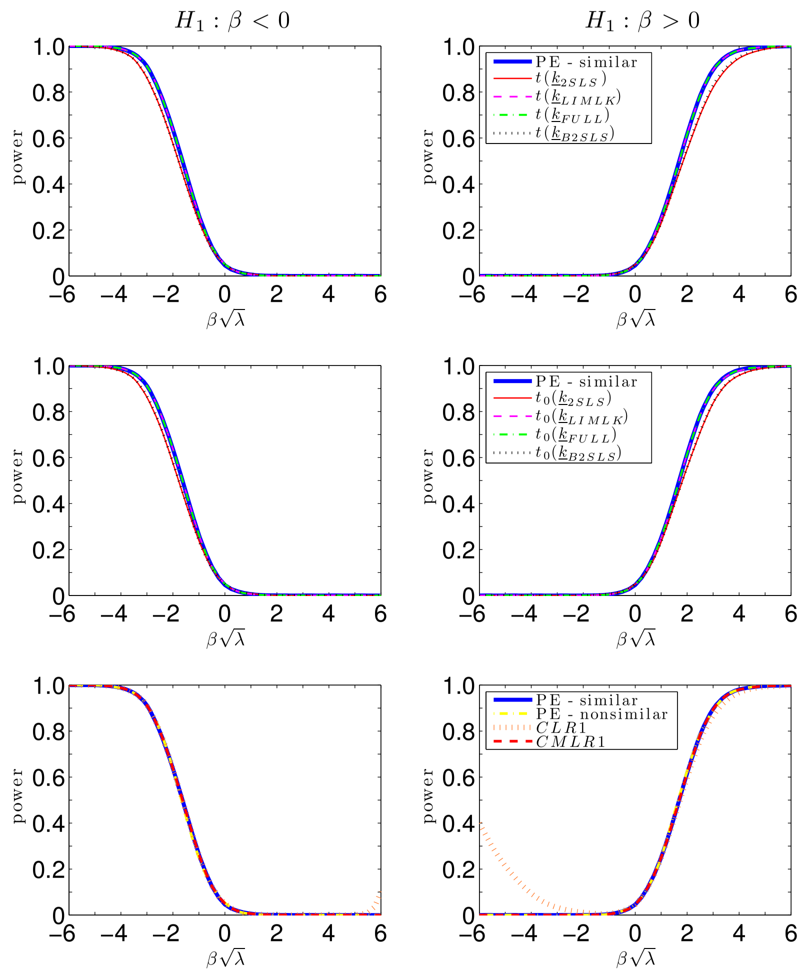


(a) $\lambda/k = 2$

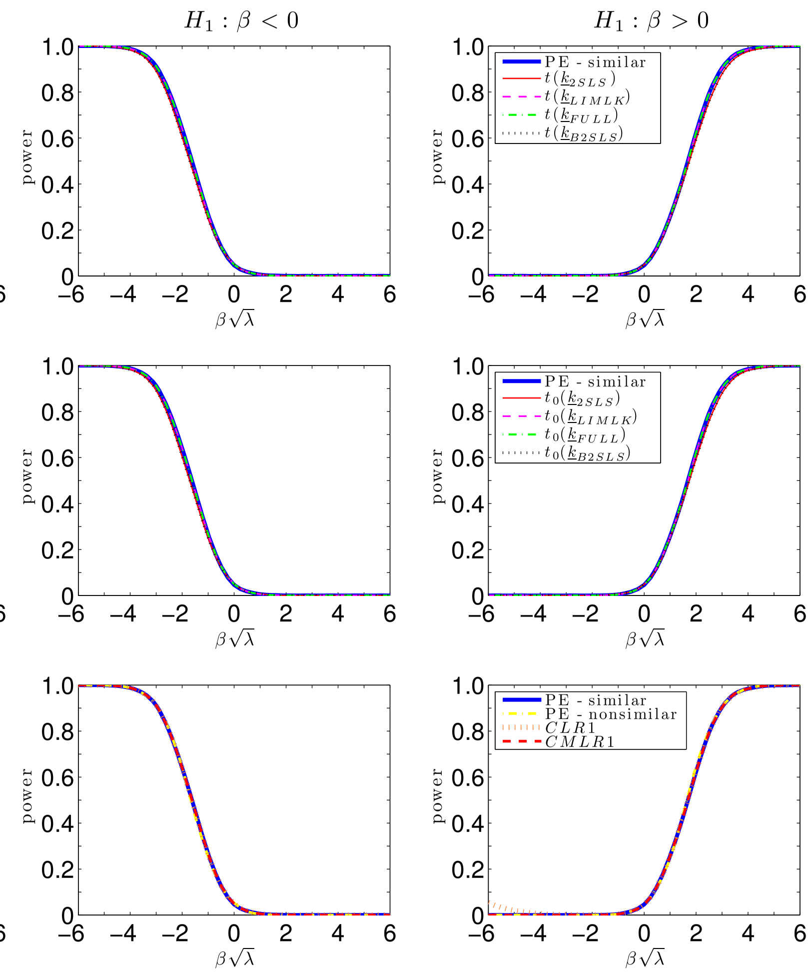


(b) $\lambda/k = 4$

Figure 29: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 5$.

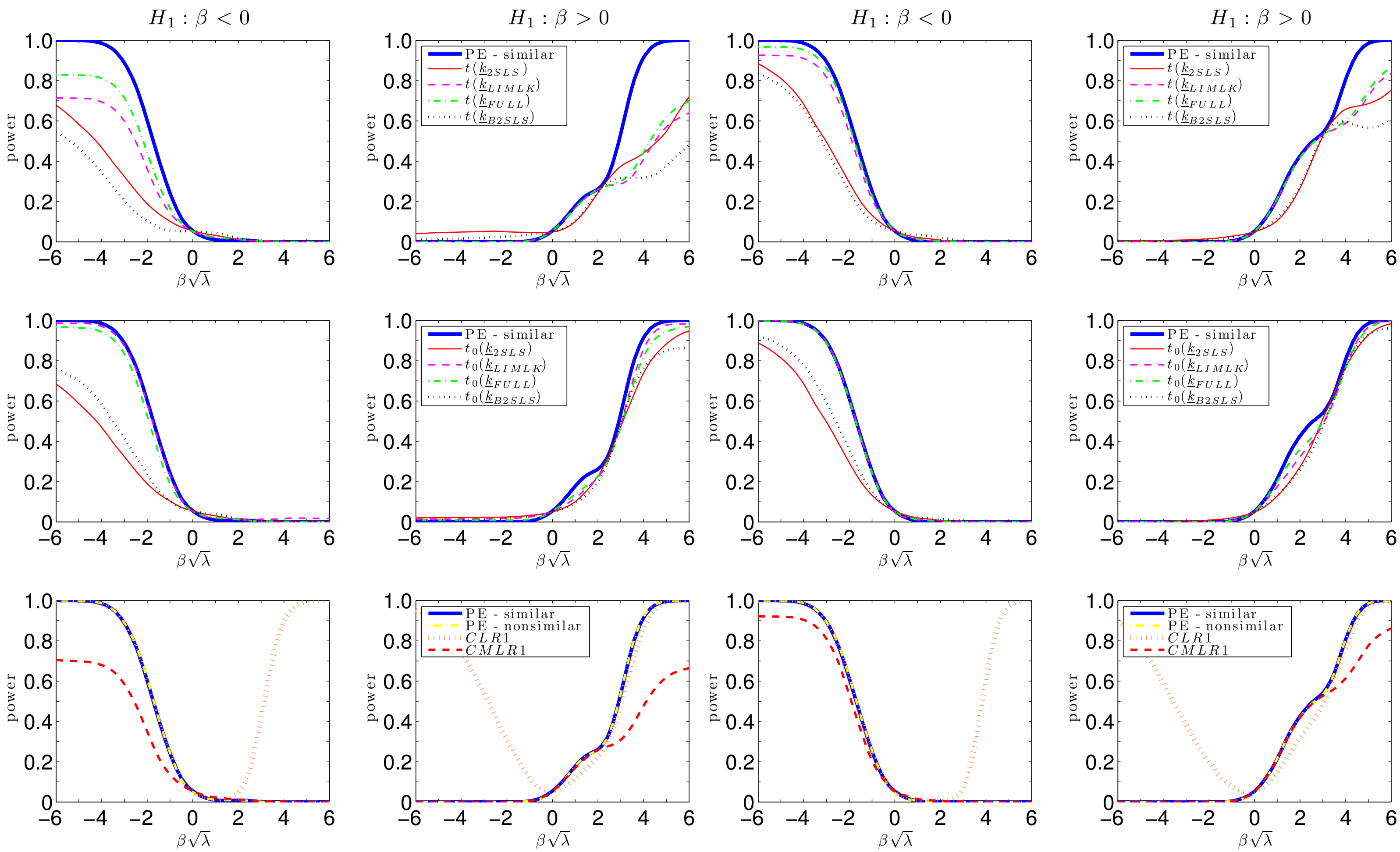


(a) $\lambda/k = 8$



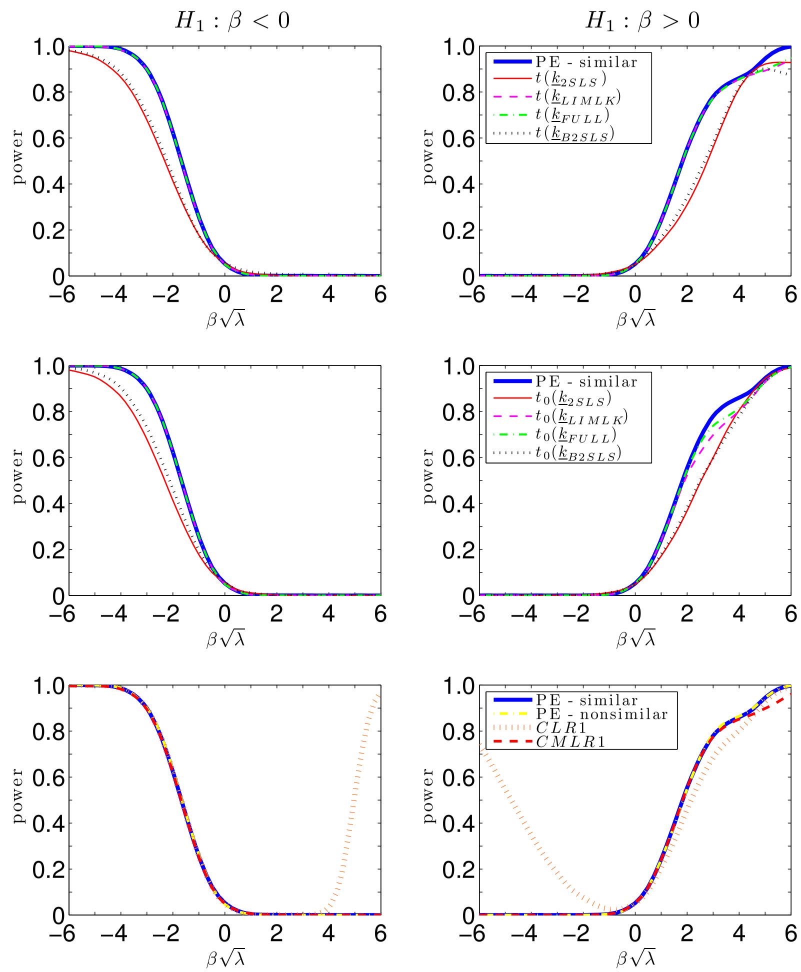
(b) $\lambda/k = 16$

Figure 30: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 5$.

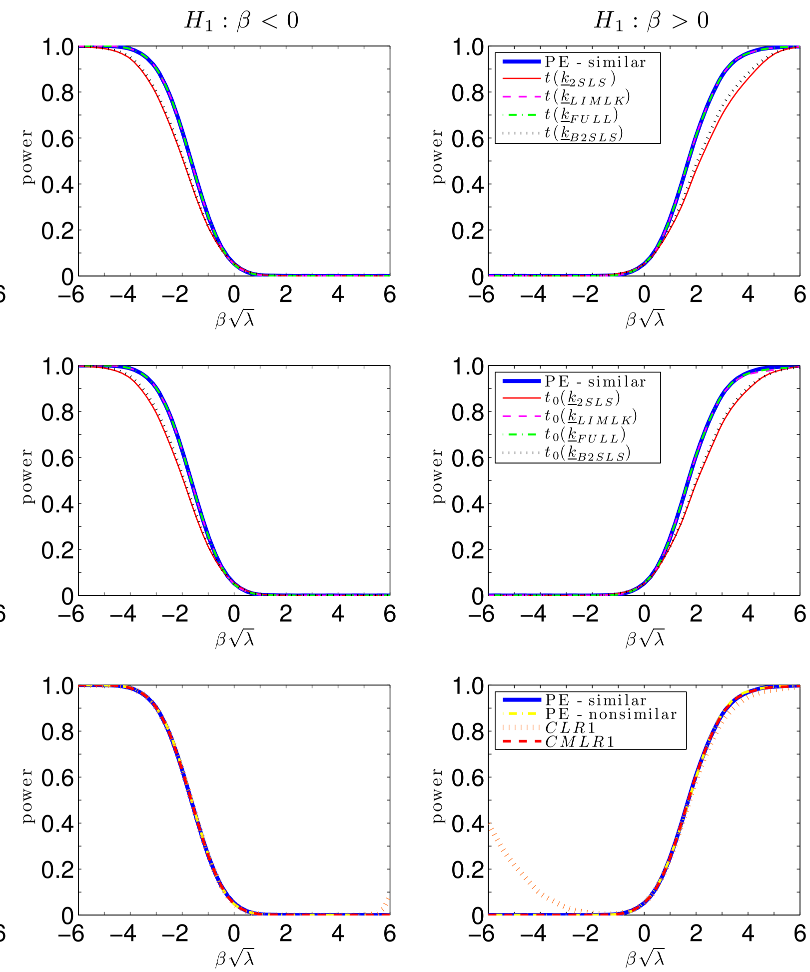


(a) $\lambda/k = 0.5$ (b) $\lambda/k = 1$

Figure 31: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 10$.

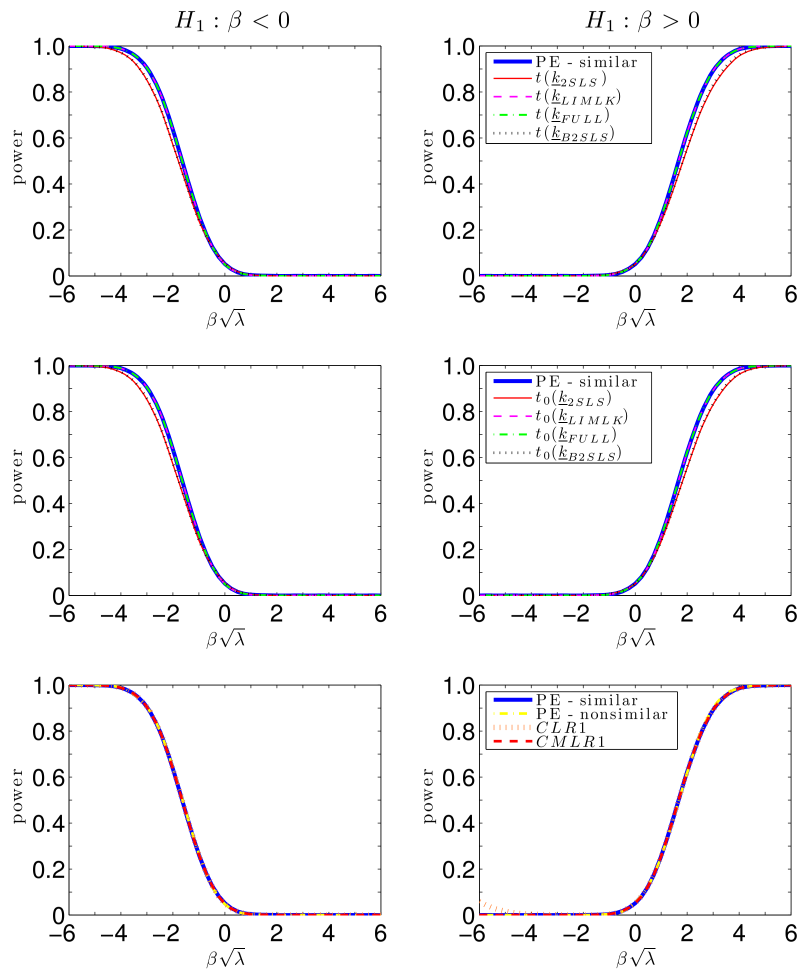


(a) $\lambda/k = 2$

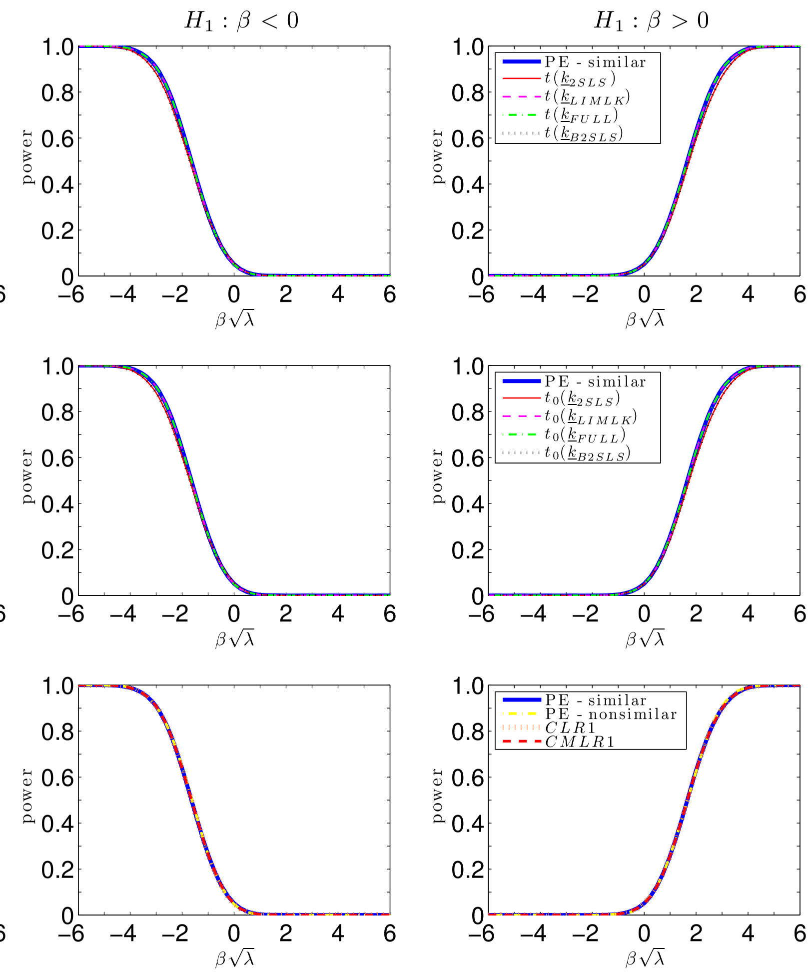


(b) $\lambda/k = 4$

Figure 32: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 10$.

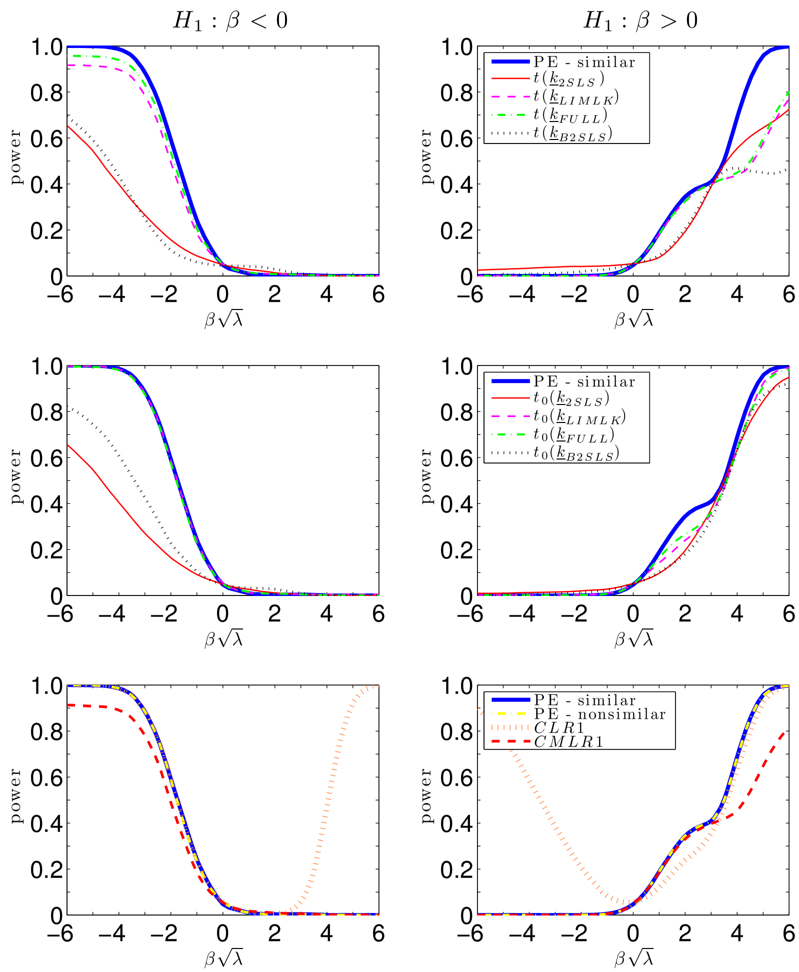


(a) $\lambda/k = 8$

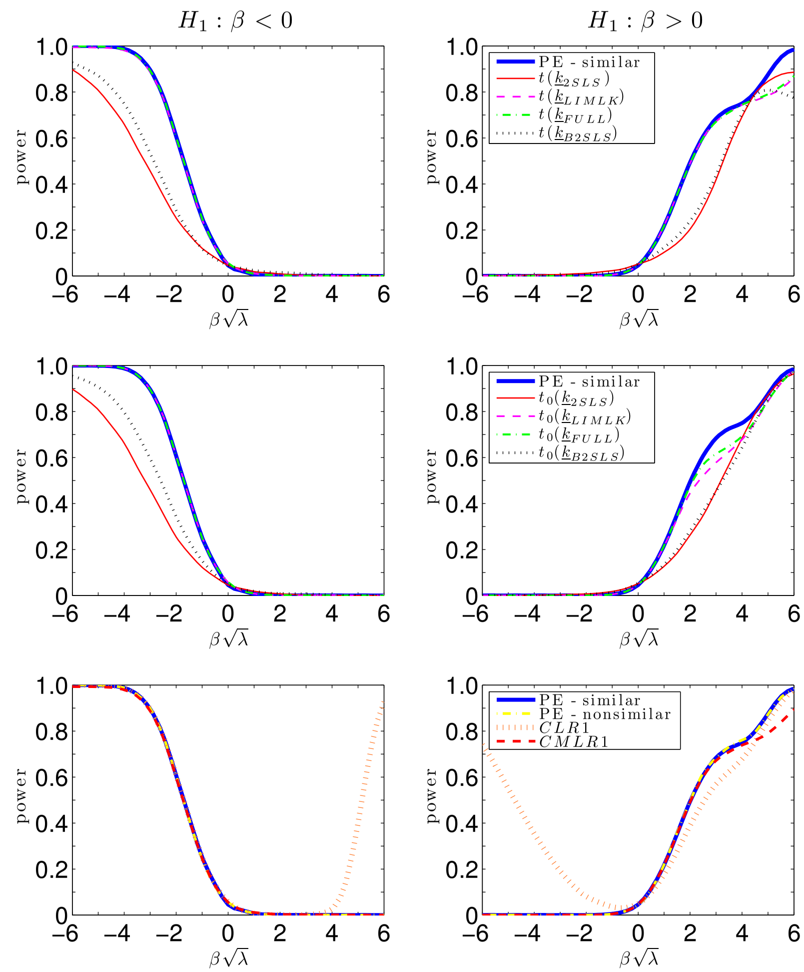


(b) $\lambda/k = 16$

Figure 33: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 10$.

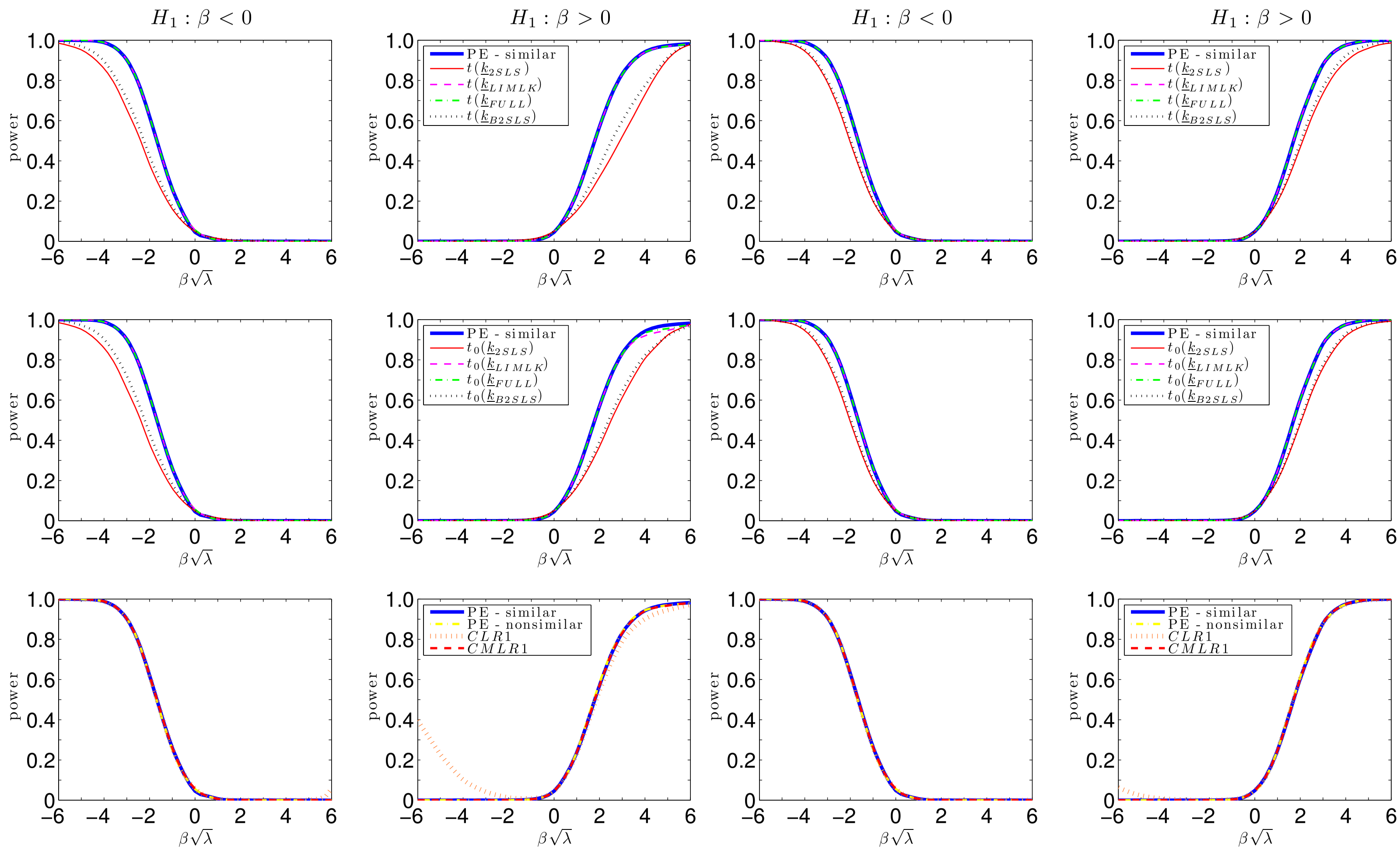


(a) $\lambda/k = 0.5$



(b) $\lambda/k = 1$

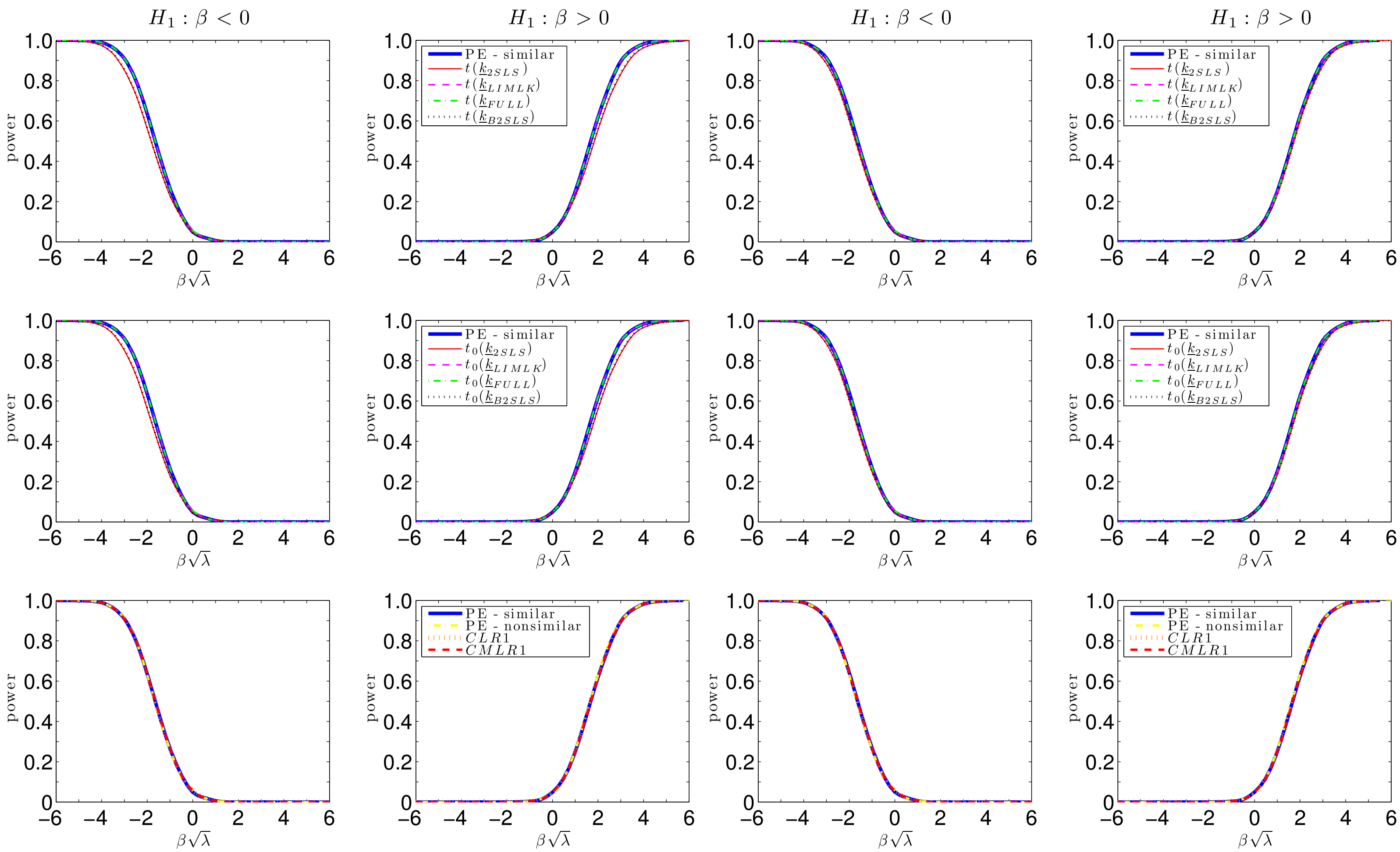
Figure 34: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 20$.



(a) $\lambda/\mathrm{k} = 2$

(b) $\lambda/\mathrm{k} = 4$

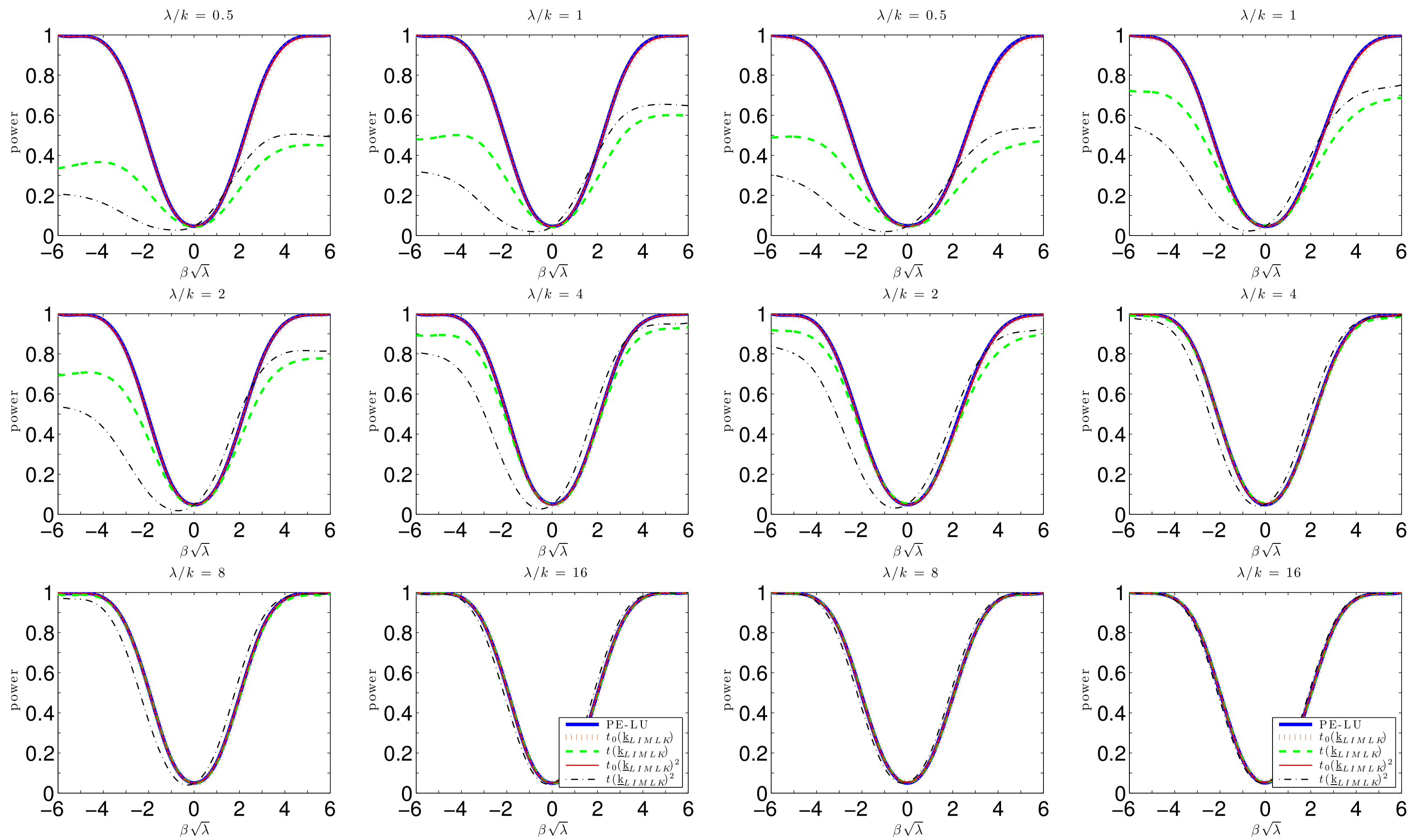
Figure 35: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 20$.



(a) $\lambda/k = 8$

(b) $\lambda/k = 16$

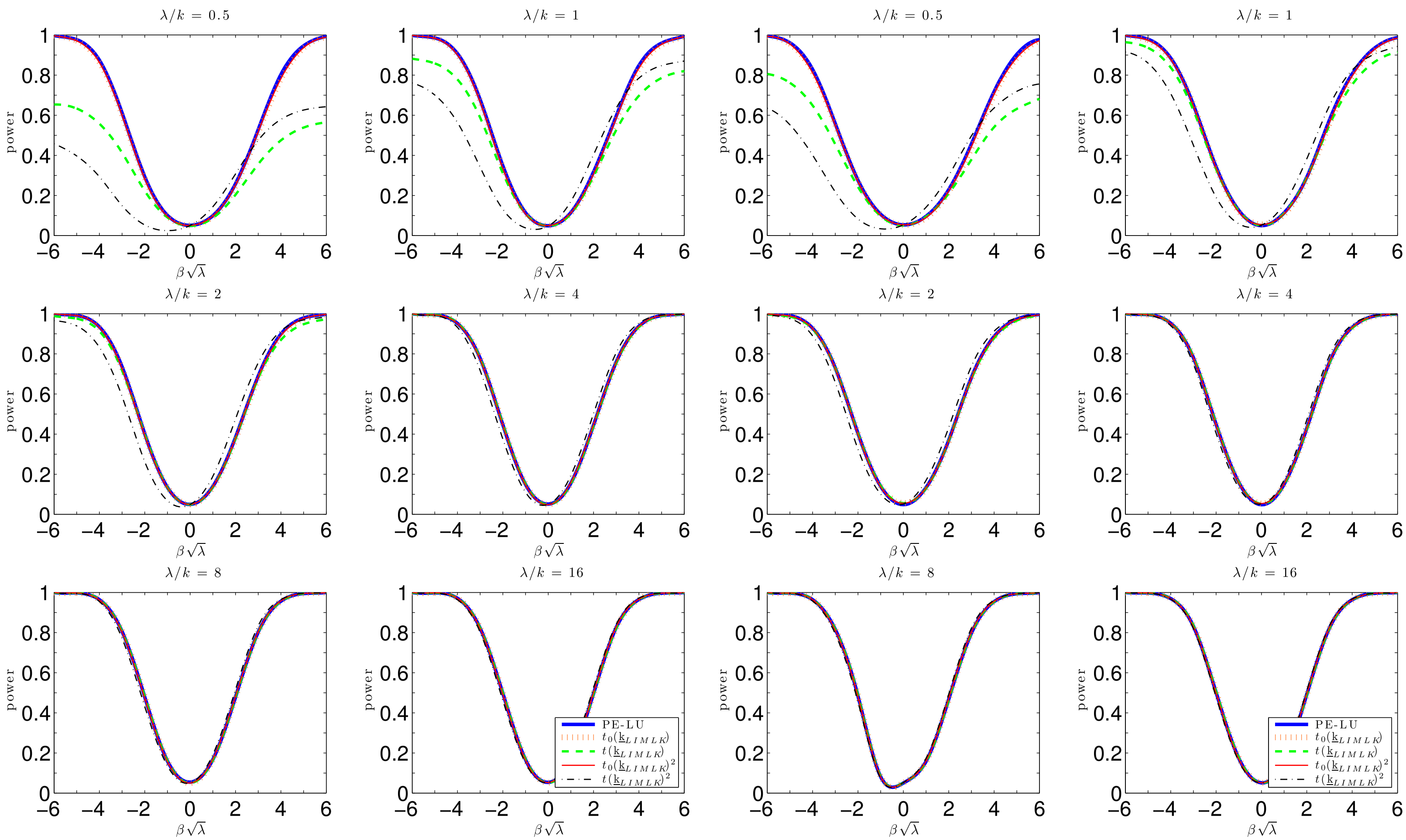
Figure 36: Power curves for one-sided unbiased LR and t-tests: $\rho = 0.9$, $k = 20$.



(a) $k = 2$

(b) $k = 5$

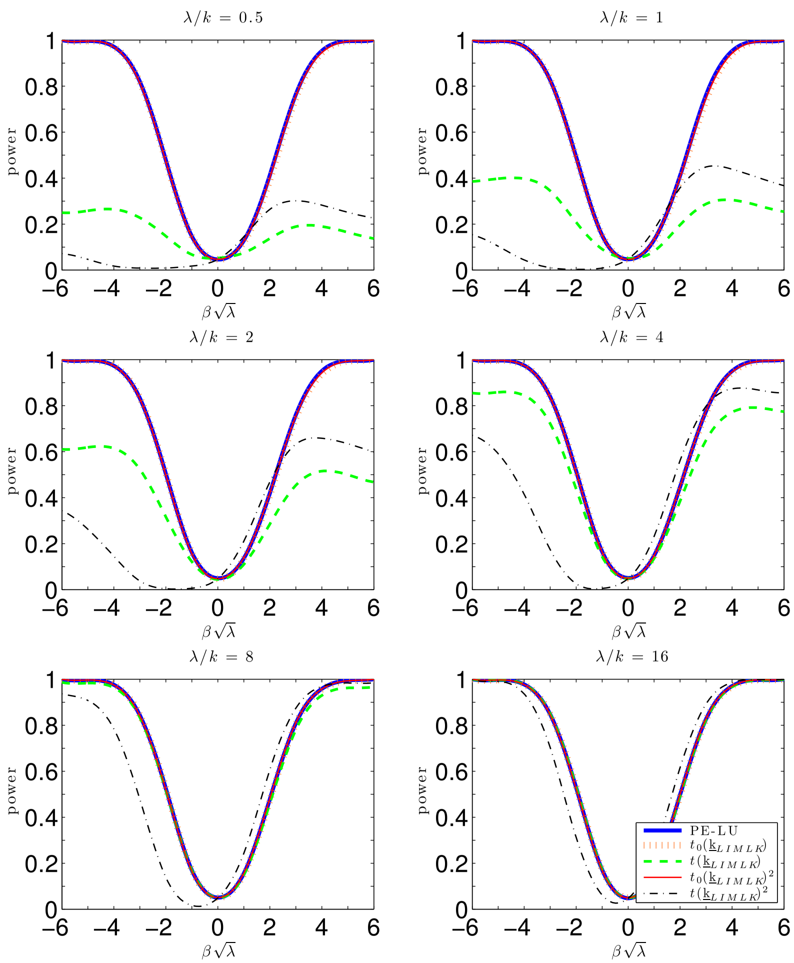
Figure 37: Power curves for two-sided conditional and unbiased t-tests based on the LIML estimator: $\rho = 0.2$.



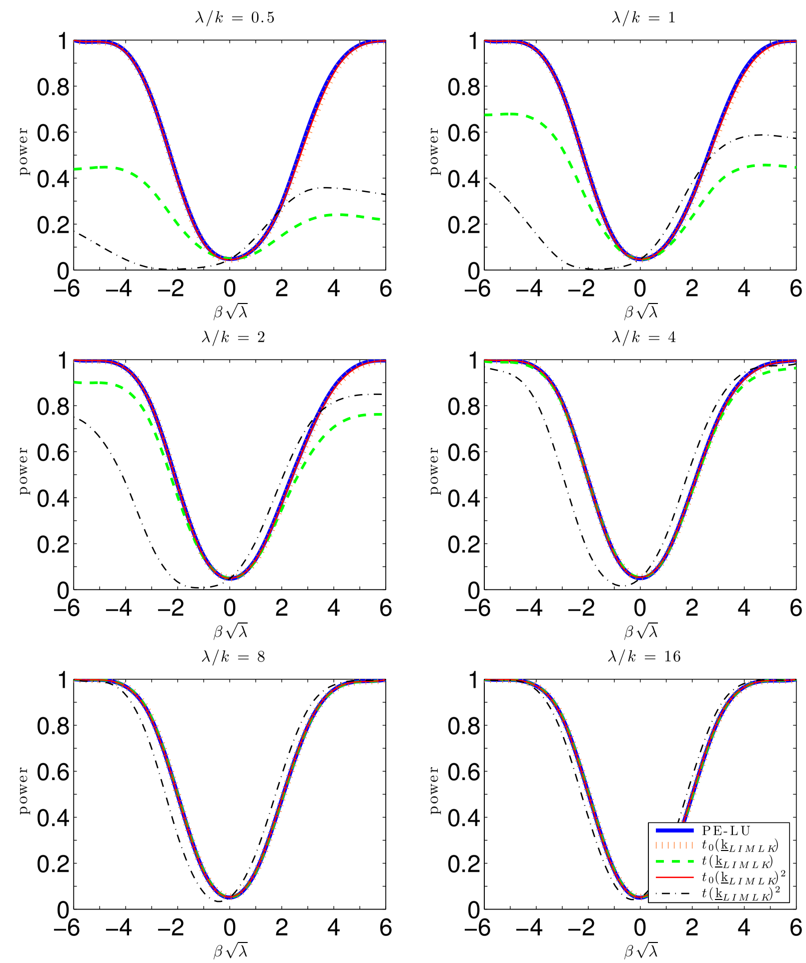
(a) $k = 10$

(b) $k = 20$

Figure 38: Power curves for two-sided conditional and unbiased t-tests based on the LIML estimator: $\rho = 0.2$.

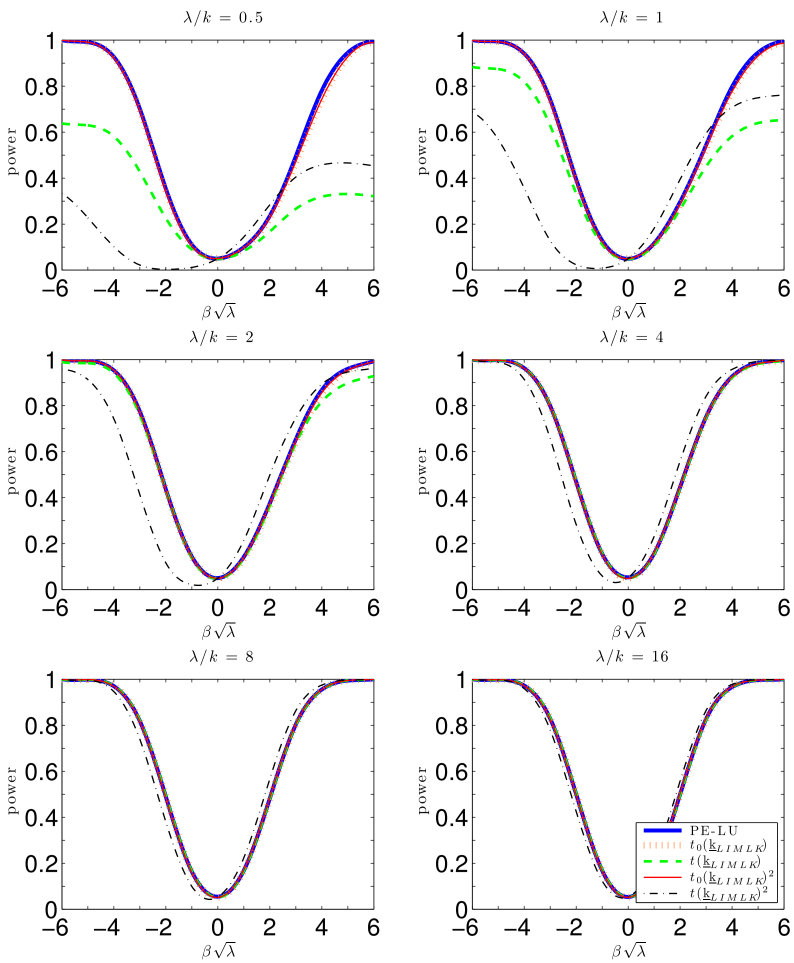


(a) $k = 2$

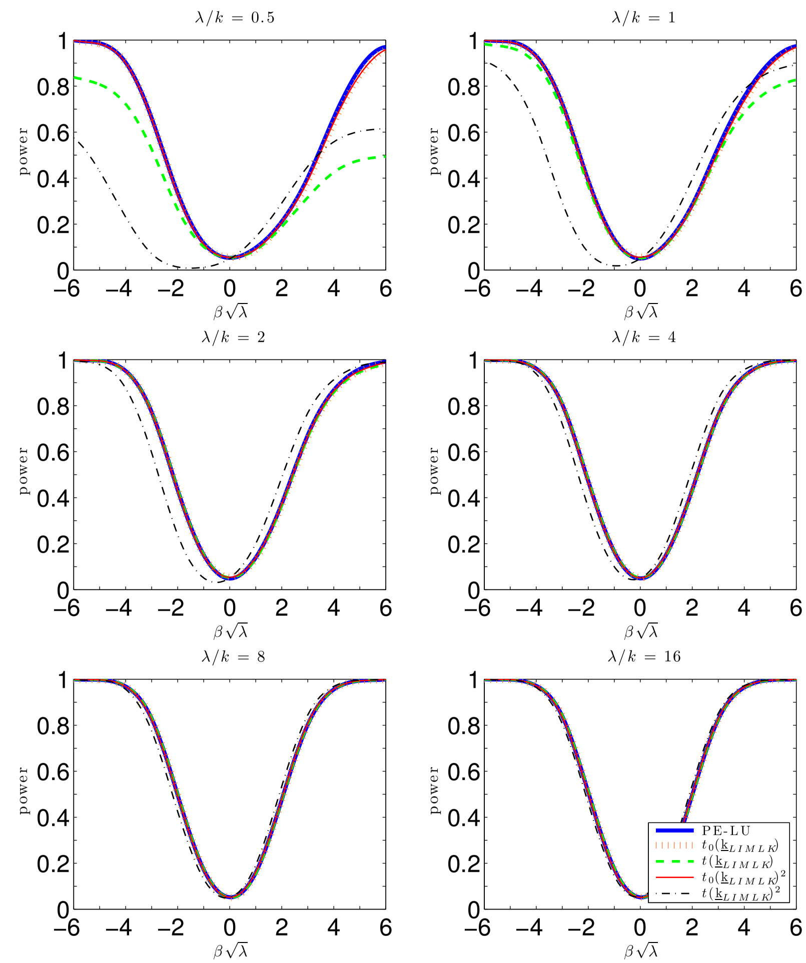


(b) $k = 5$

Figure 39: Power curves for two-sided conditional and unbiased t-tests based on the LIML estimator: $\rho = 0.5$.

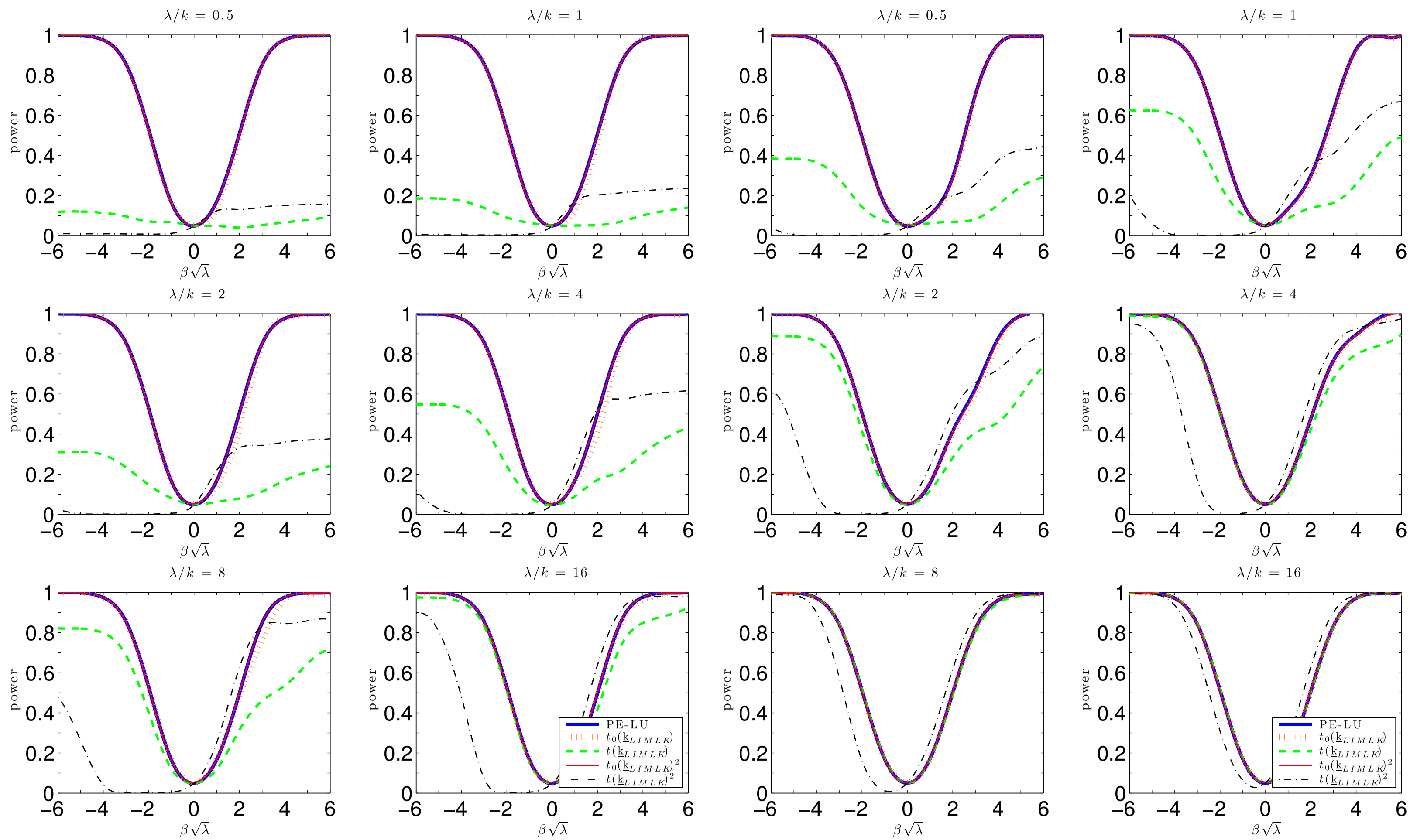


(a) $k = 10$



(b) $k = 20$

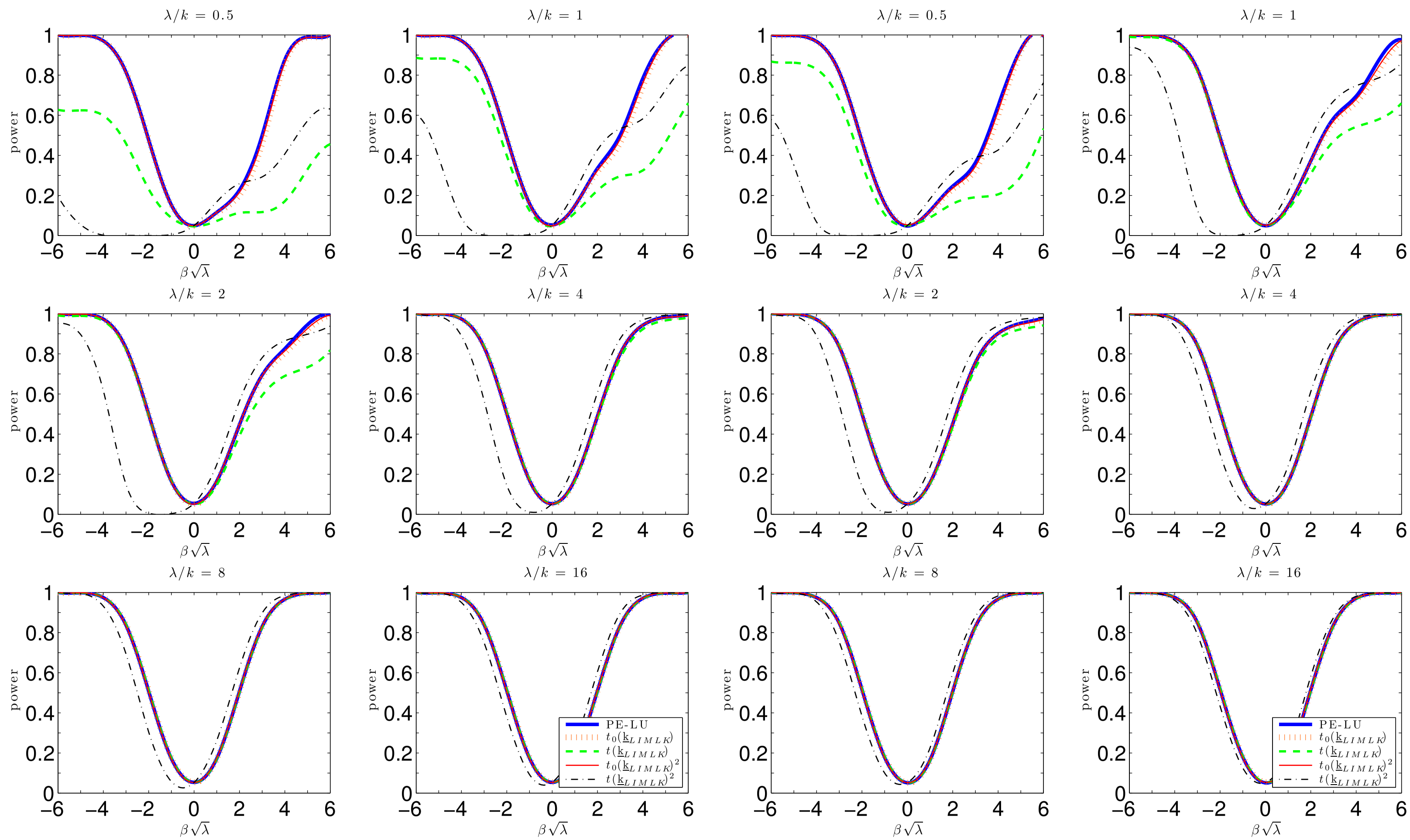
Figure 40: Power curves for two-sided conditional and unbiased t-tests based on the LIML estimator: $\rho = 0.5$.



(a) $k = 2$

(b) $k = 5$

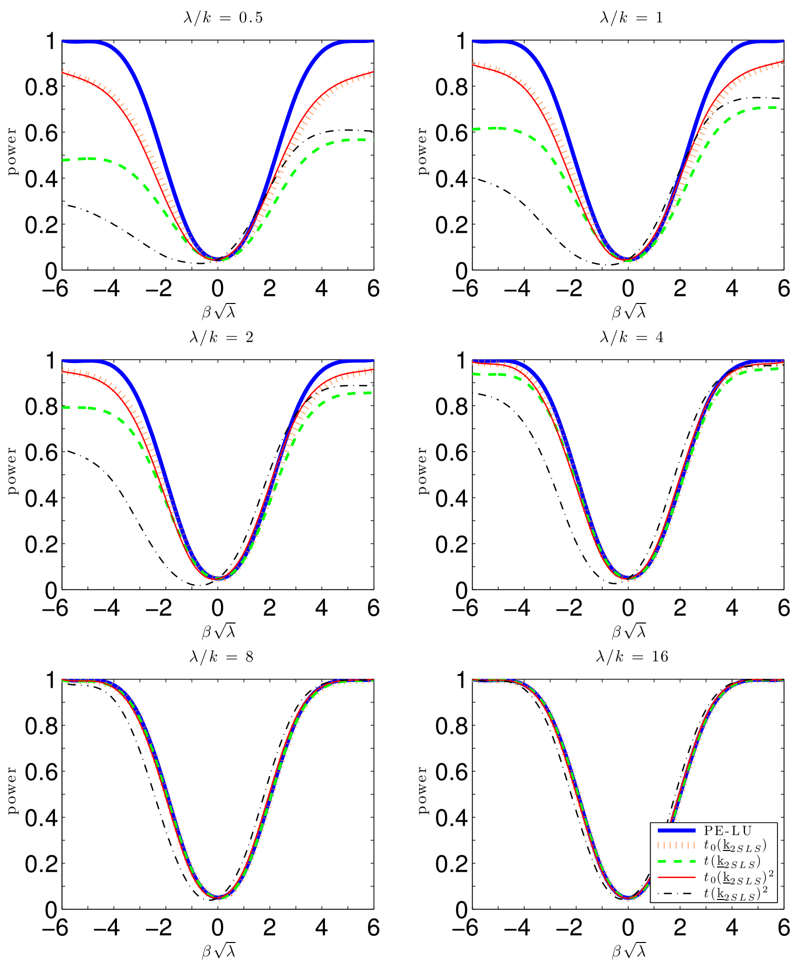
Figure 41: Power curves for two-sided conditional and unbiased t-tests based on the LIML estimator: $\rho = 0.9$.



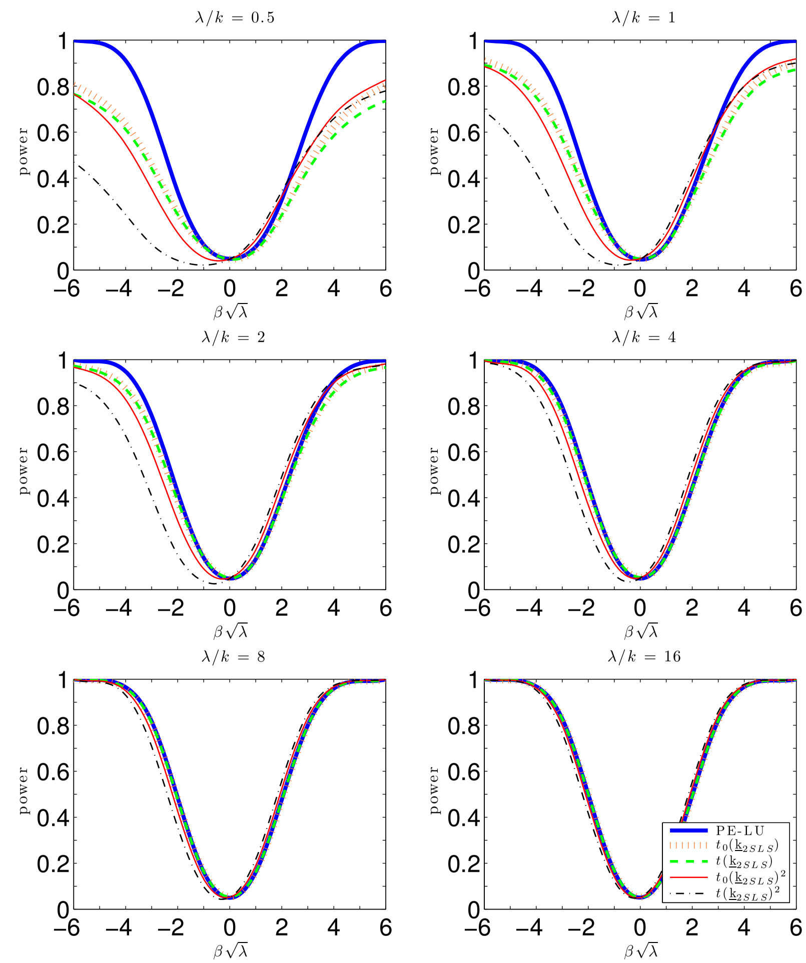
(a) $k = 10$

(b) $k = 20$

Figure 42: Power curves for two-sided conditional and unbiased t-tests based on the LIML estimator: $\rho = 0.9$.

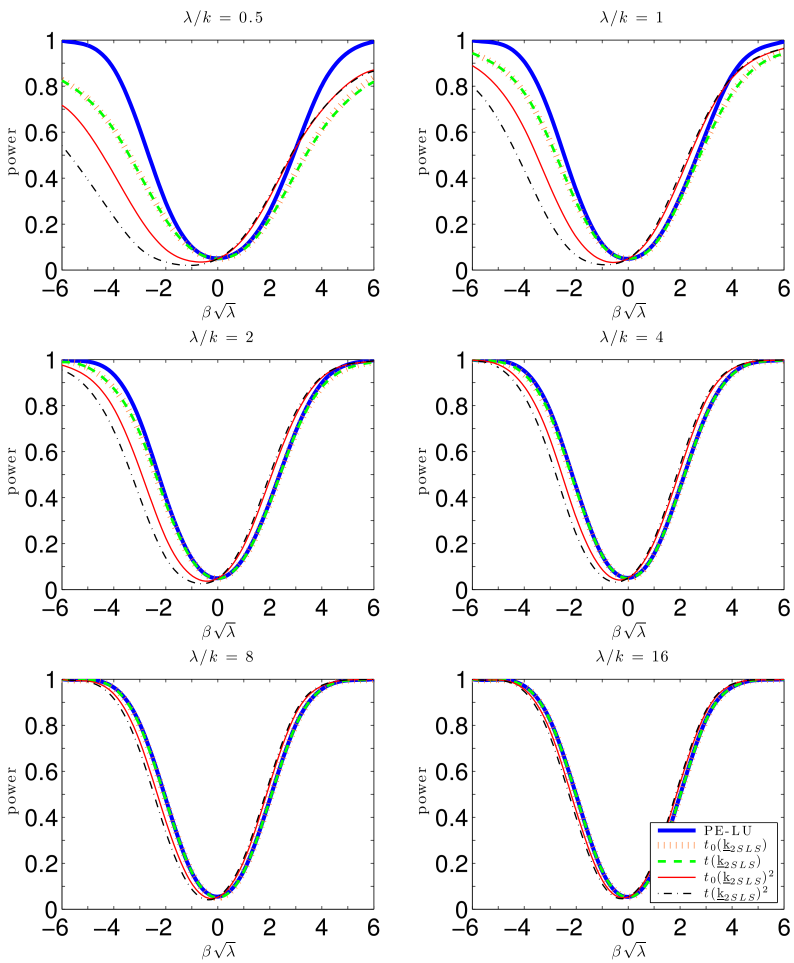


(a) $k = 2$

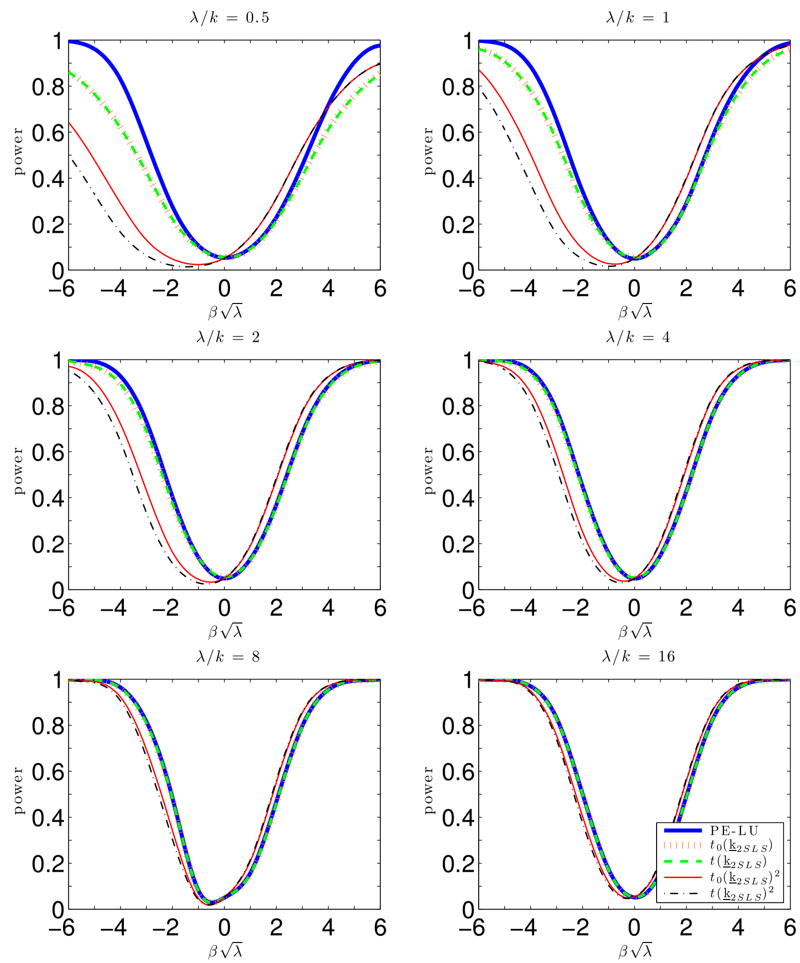


(b) $k = 5$

Figure 43: Power curves for two-sided conditional and unbiased t-tests based on the 2SLS estimator: $\rho = 0.2$.

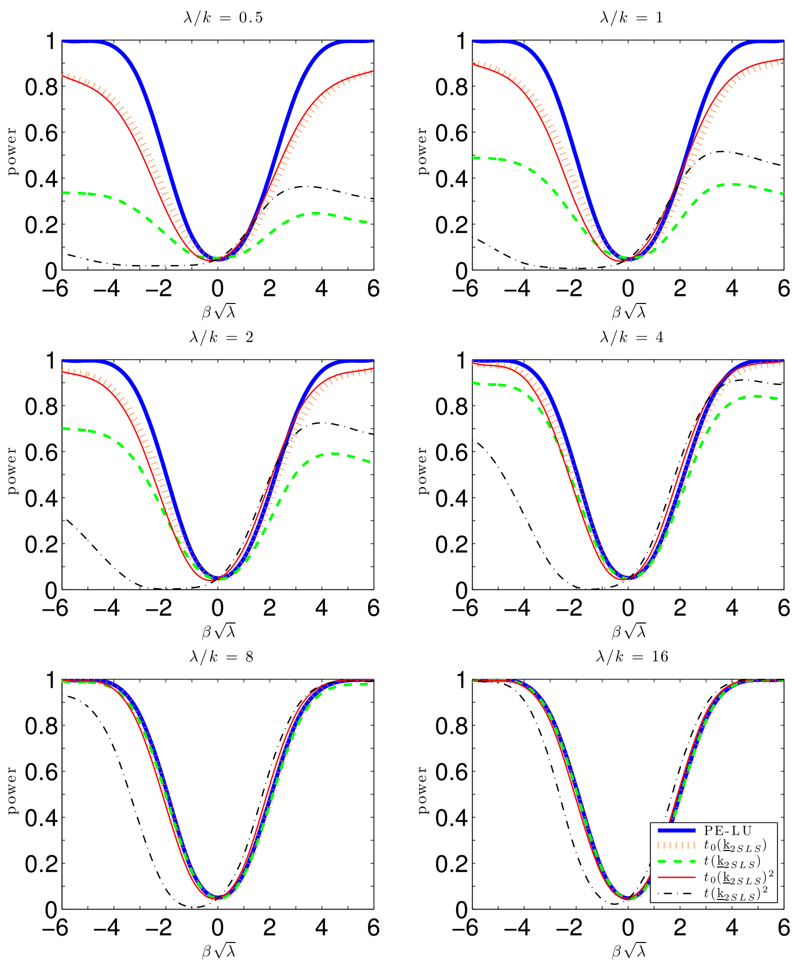


(a) $k = 10$

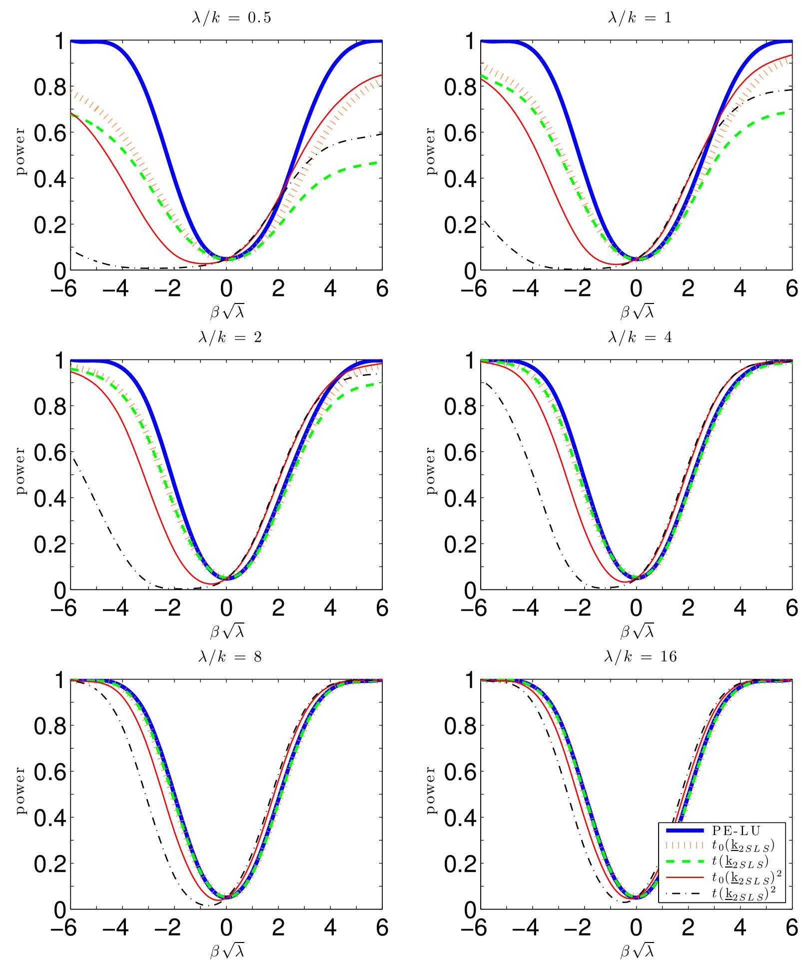


(b) $k = 20$

Figure 44: Power curves for two-sided conditional and unbiased t-tests based on the 2SLS estimator: $\rho = 0.2$.

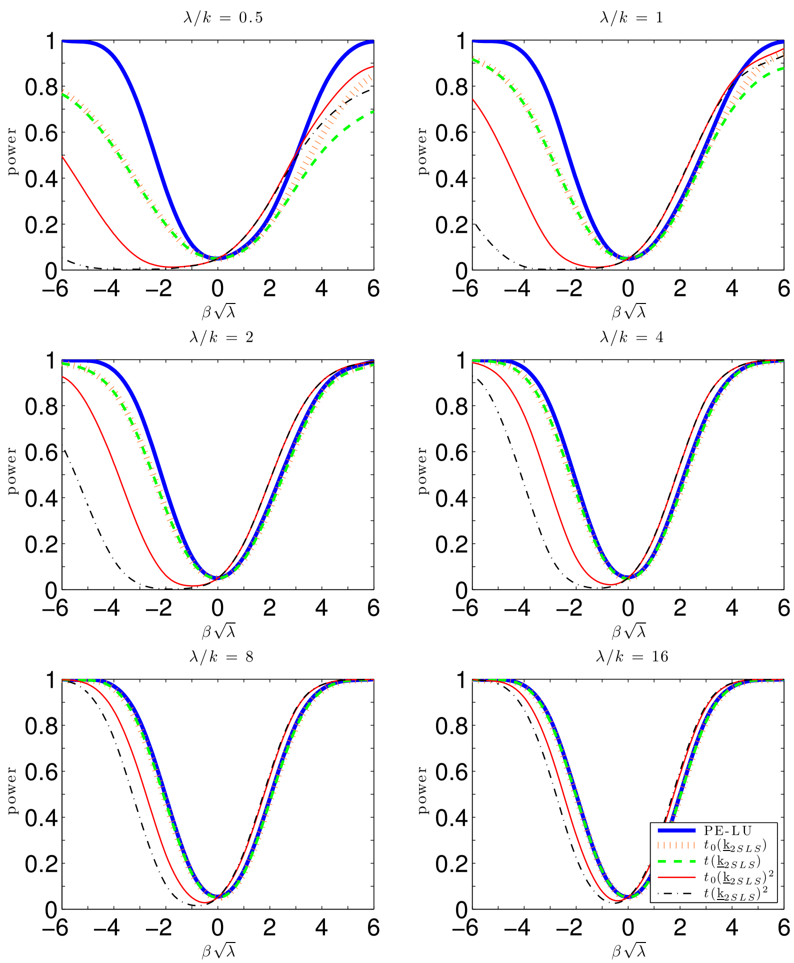


(a) $k = 2$

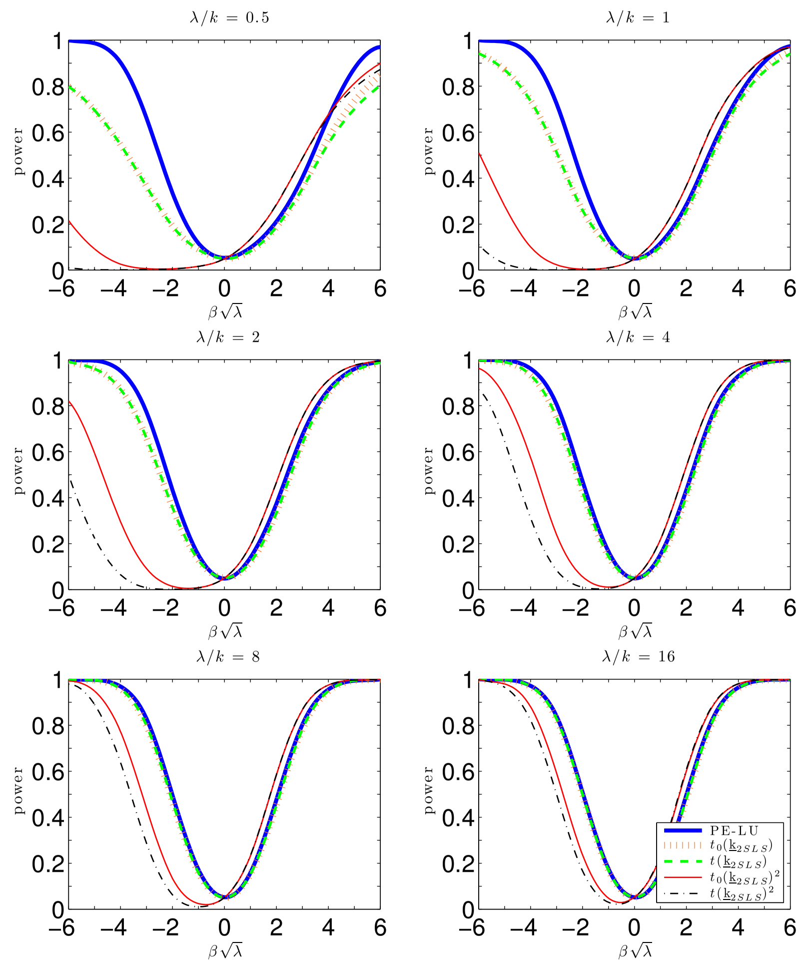


(b) $k = 5$

Figure 45: Power curves for two-sided conditional and unbiased t-tests based on the 2SLS estimator: $\rho = 0.5$.

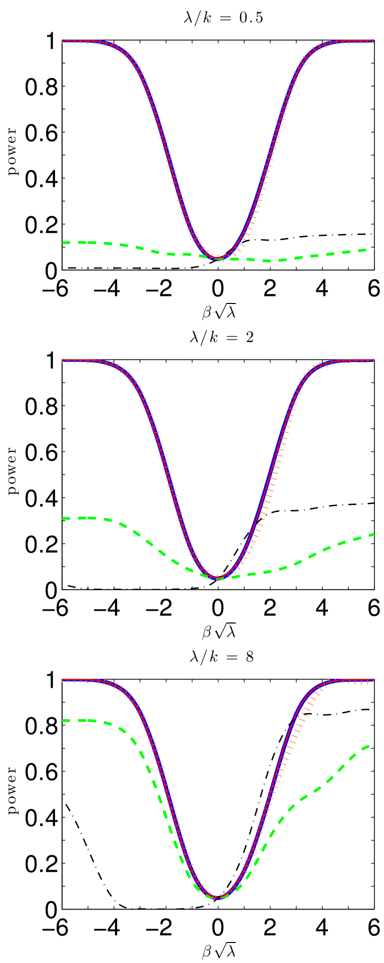


(a) $k = 10$

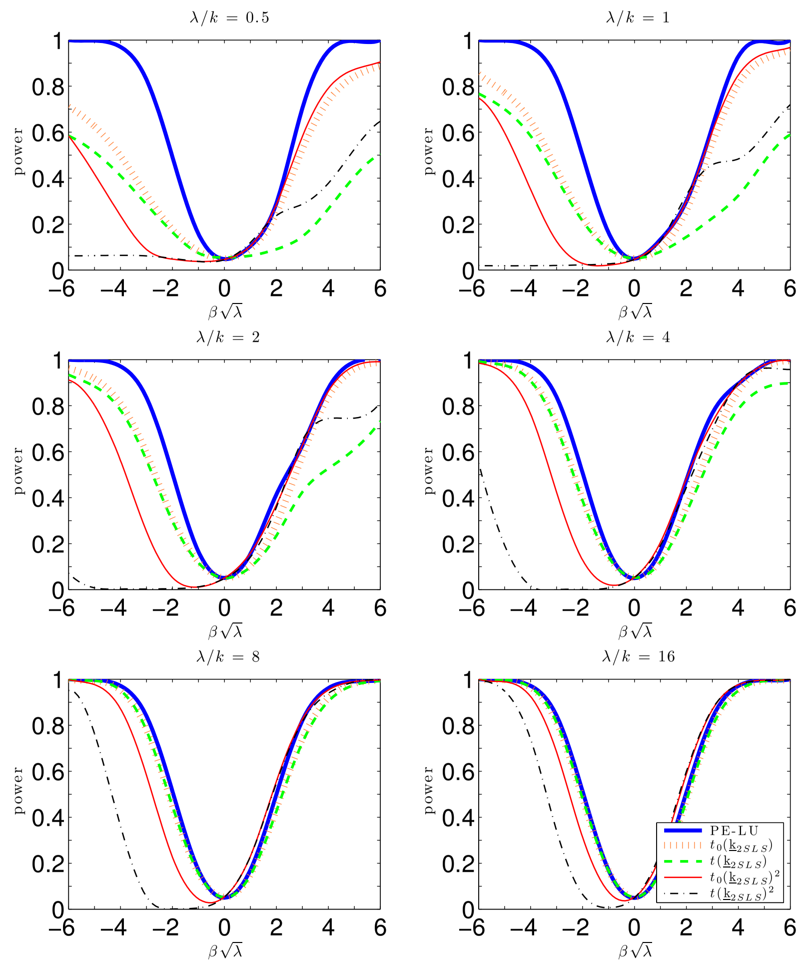


(b) $k = 20$

Figure 46: Power curves for two-sided conditional and unbiased t-tests based on the 2SLS estimator: $\rho = 0.5$.

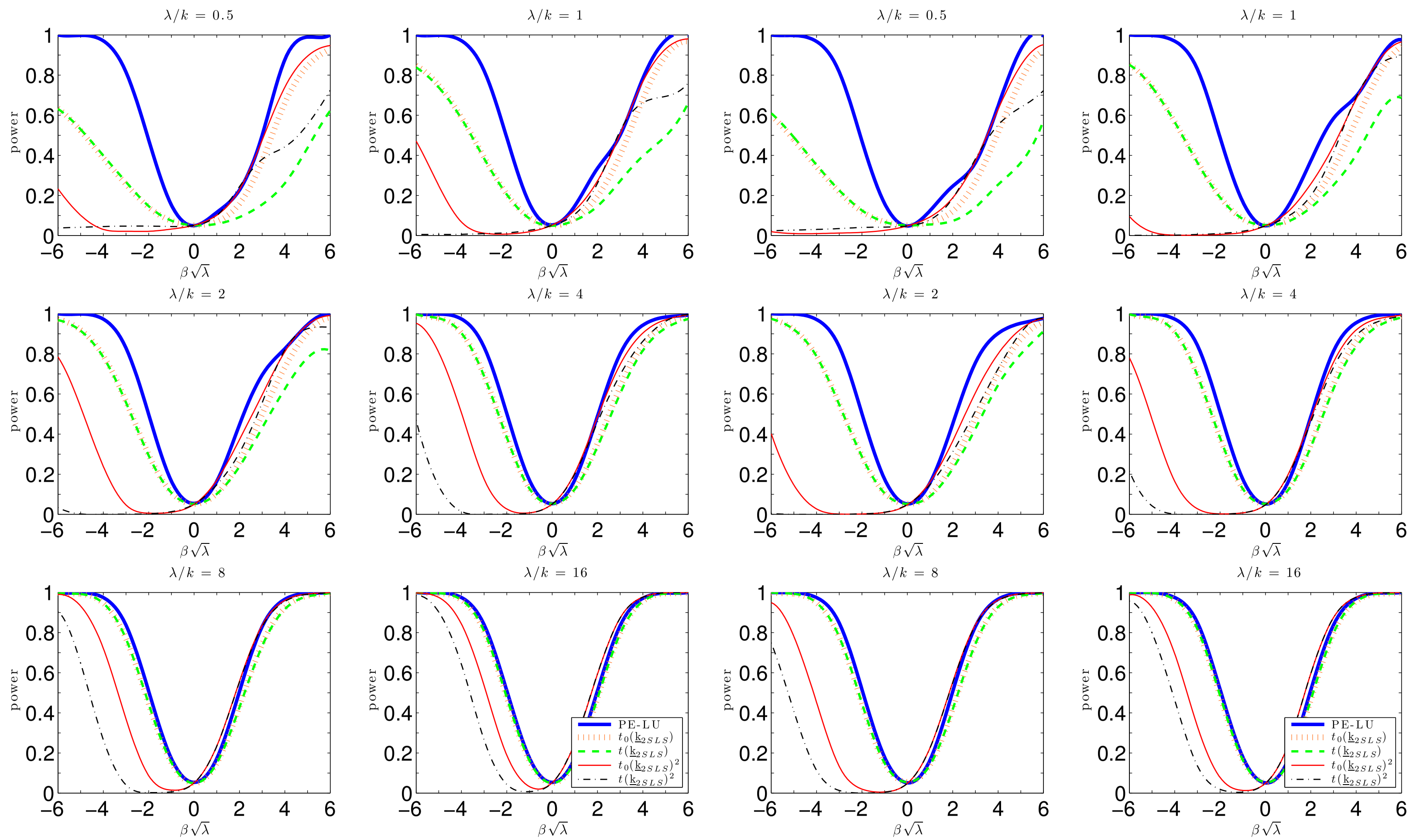


(a) $k = 2$



(b) $k = 5$

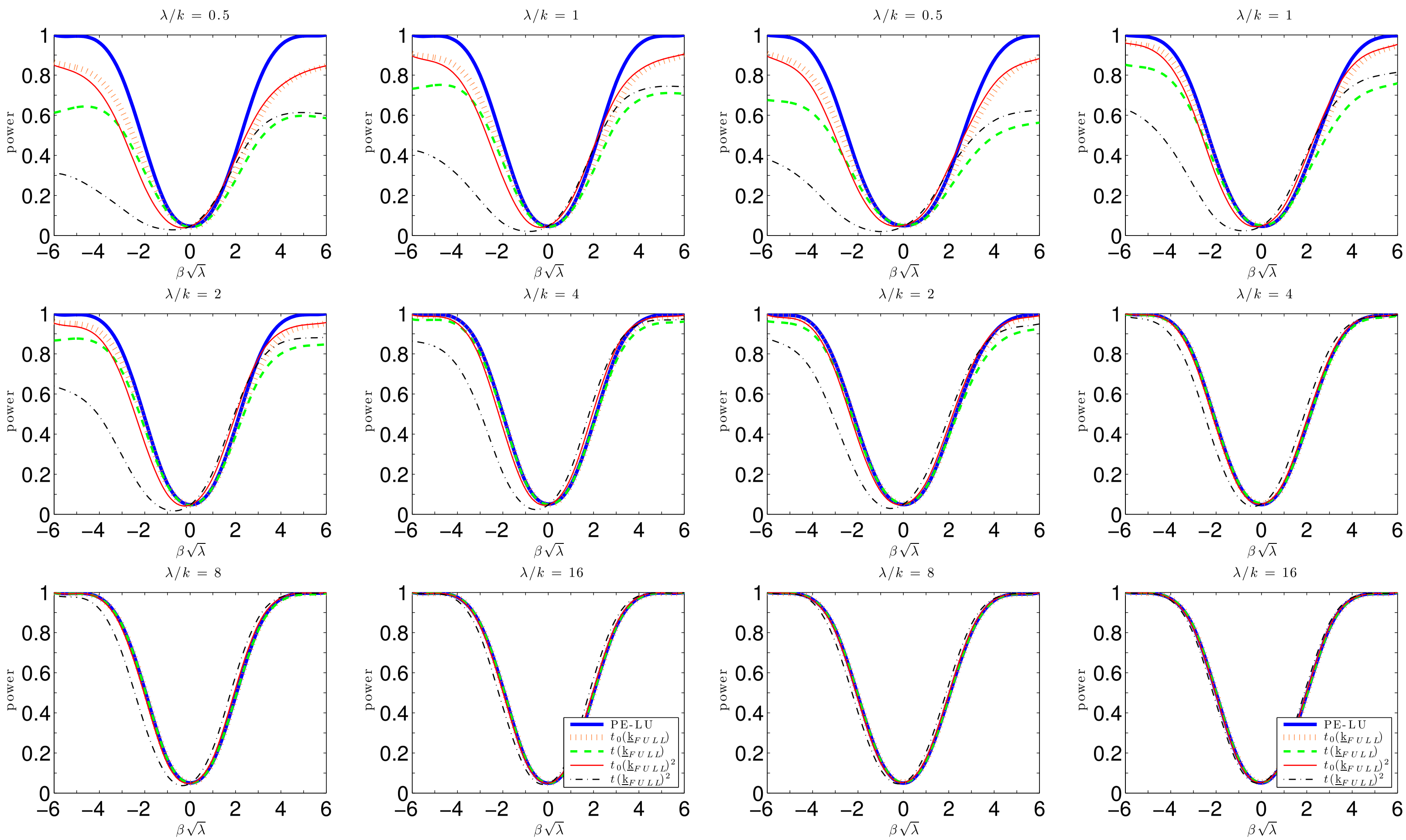
Figure 47: Power curves for two-sided conditional and unbiased t-tests based on the 2SLS estimator: $\rho = 0.9$.



(a) $k = 10$

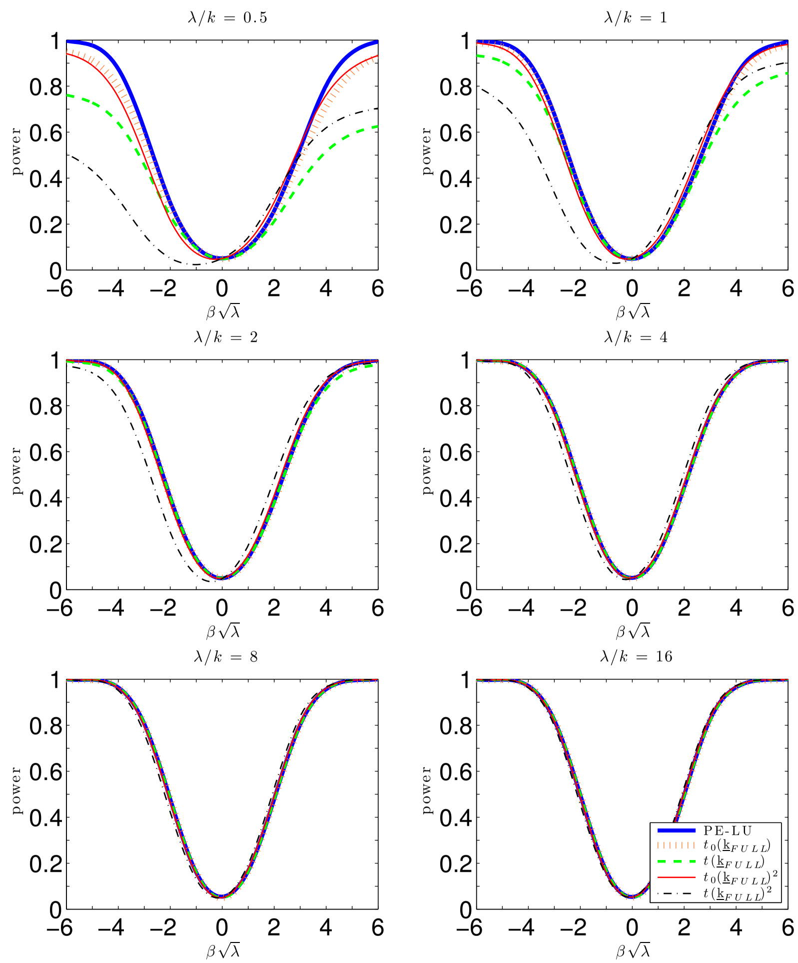
(b) $k = 20$

Figure 48: Power curves for two-sided conditional and unbiased t-tests based on the 2SLS estimator: $\rho = 0.9$.

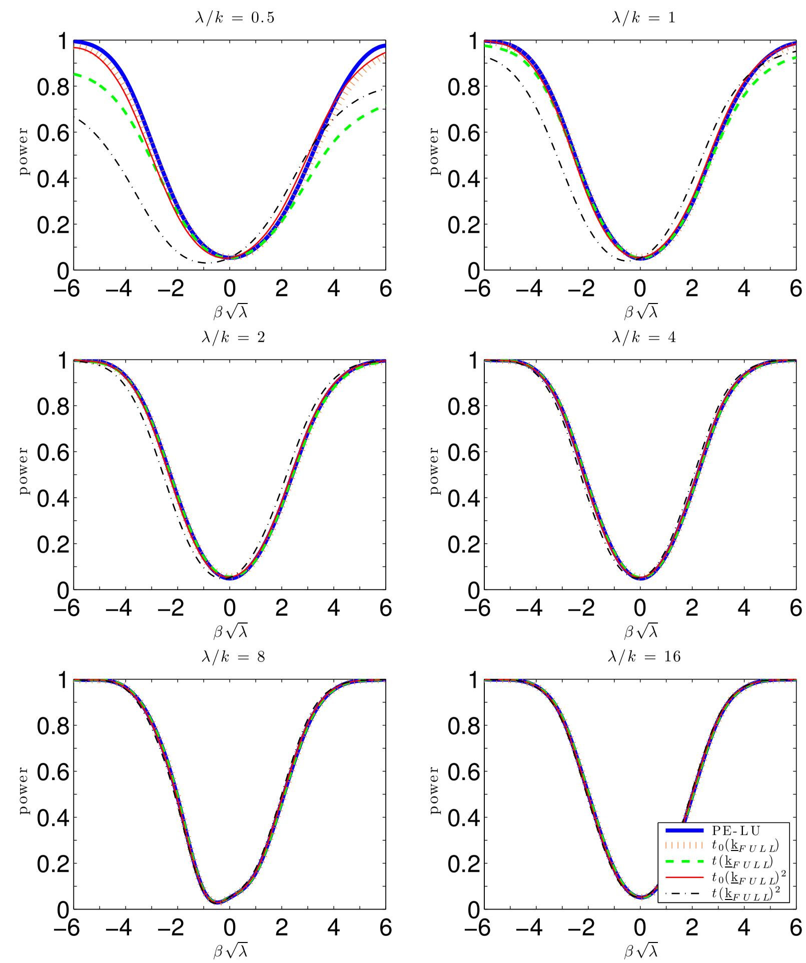


(a) $k = 2$ (b) $k = 5$

Figure 49: Power curves for two-sided conditional and unbiased t-tests based on the Fuller estimator: $\rho = 0.2$.

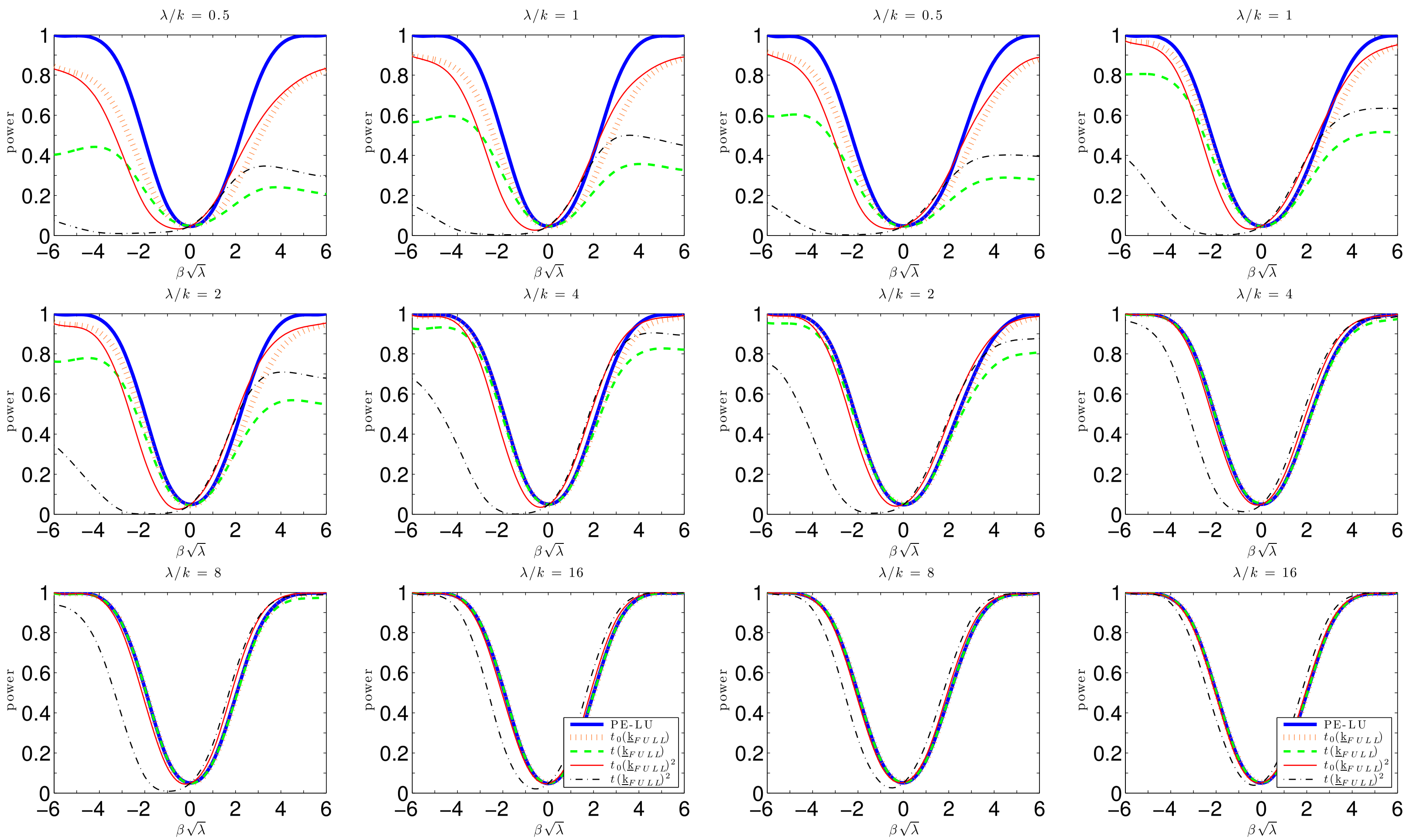


(a) $k = 10$



(b) $k = 20$

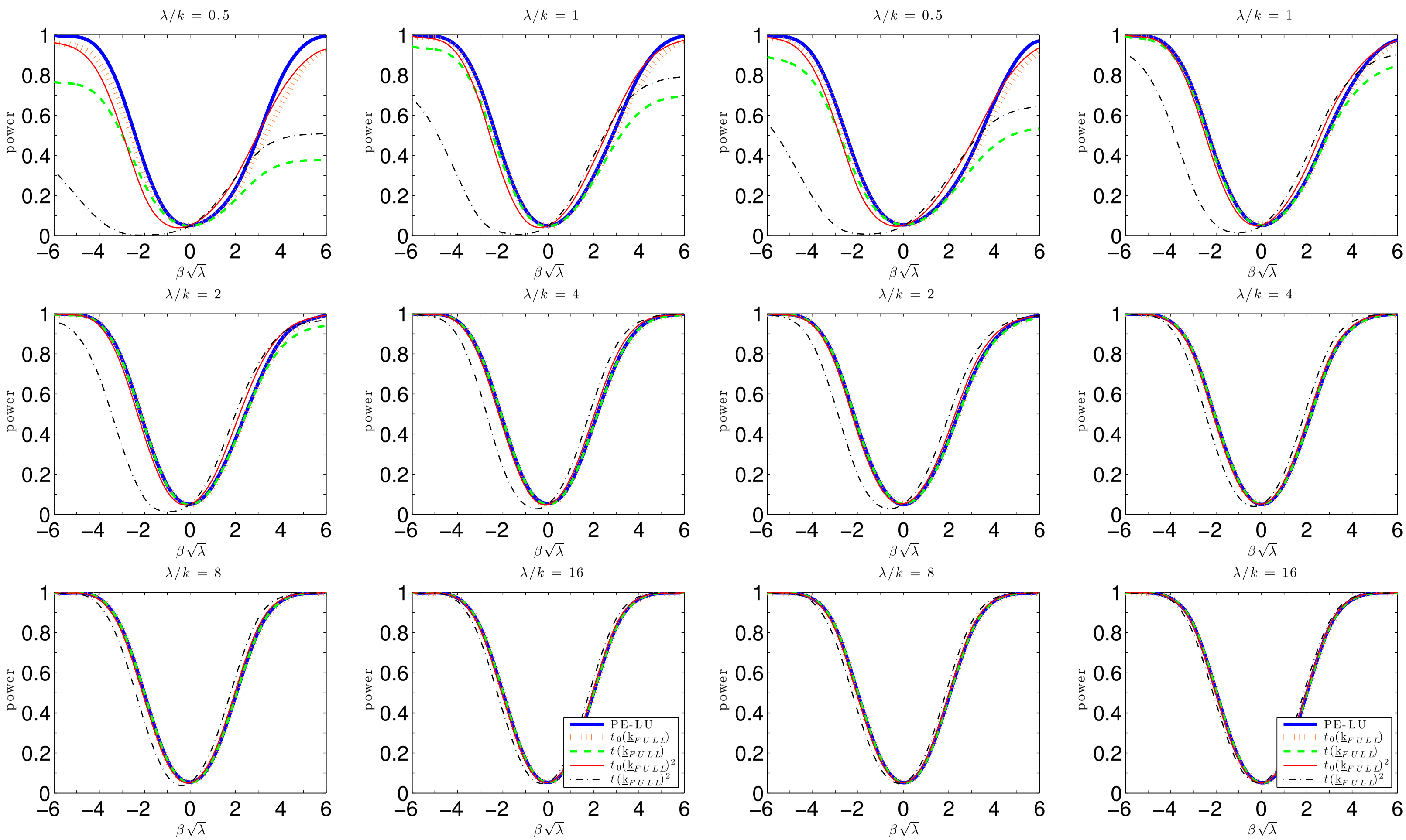
Figure 50: Power curves for two-sided conditional and unbiased t-tests based on the Fuller estimator: $\rho = 0.2$.



(a) $k = 2$

(b) $k = 5$

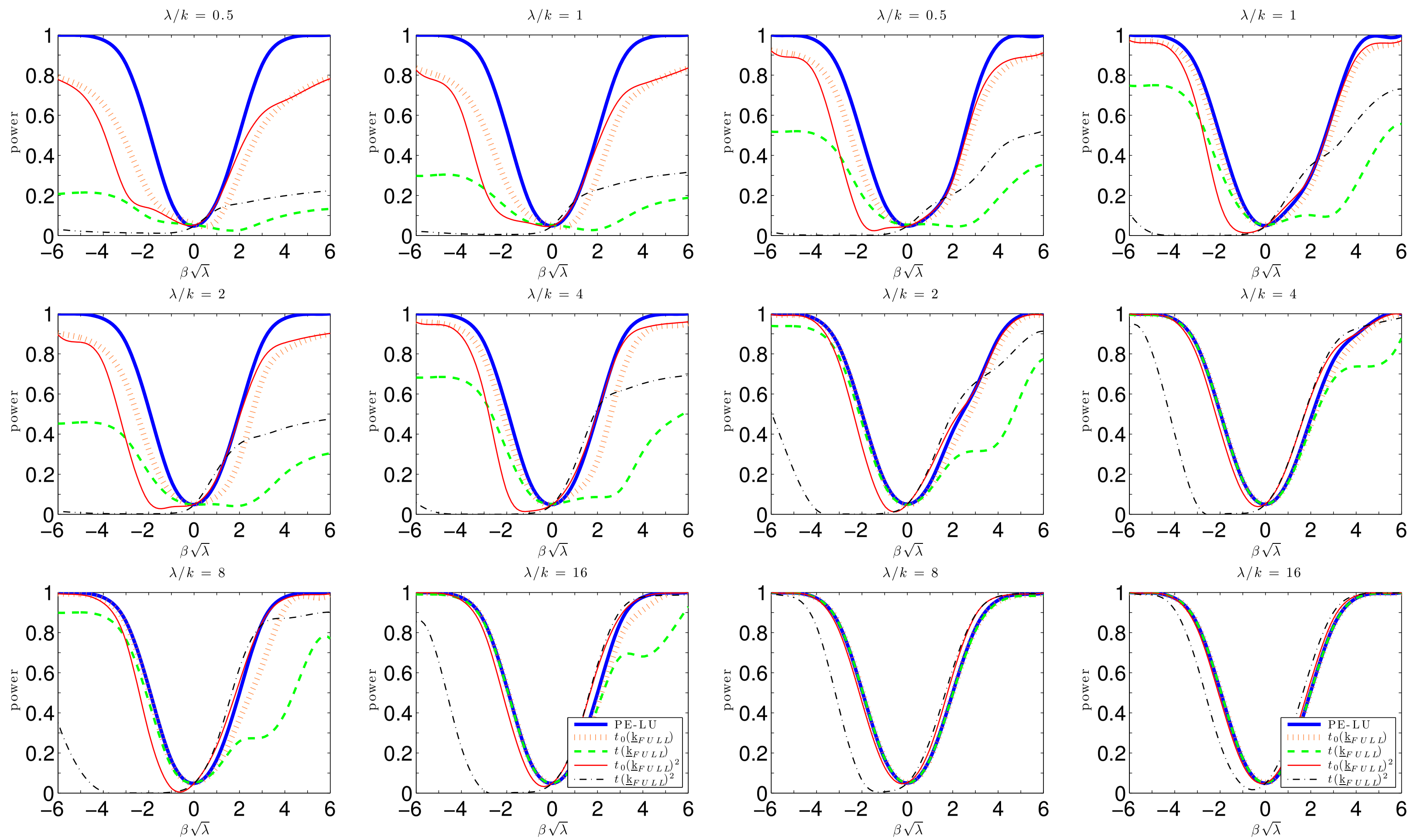
Figure 51: Power curves for two-sided conditional and unbiased t-tests based on the Fuller estimator: $\rho = 0.5$.



(a) $k = 10$

(b) $k = 20$

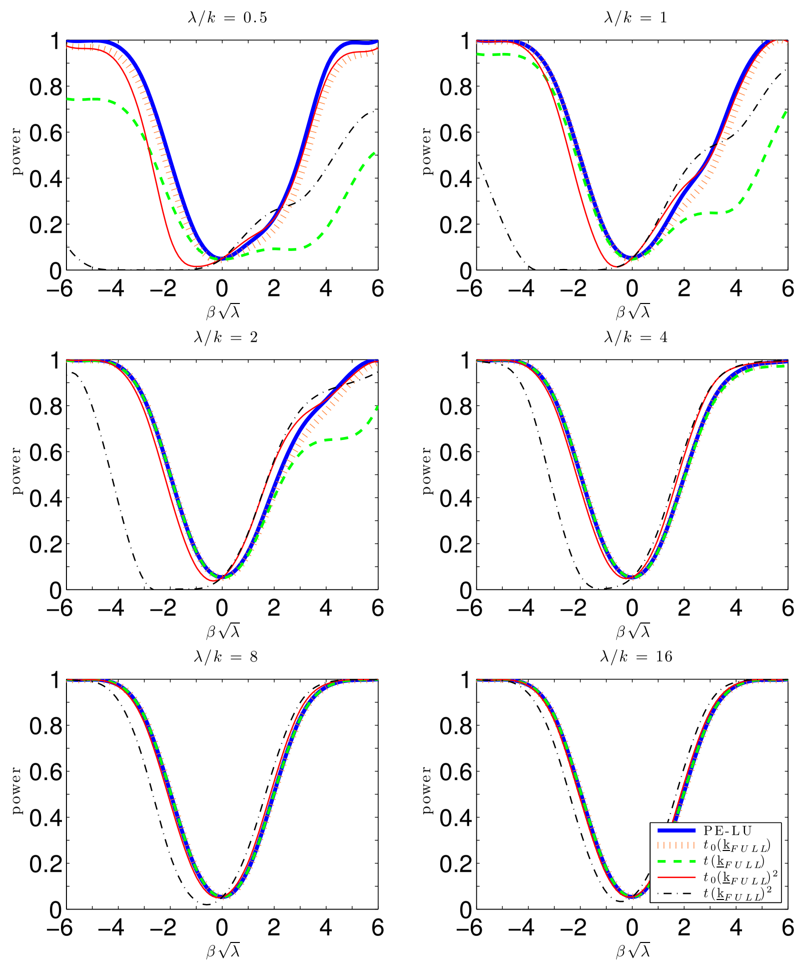
Figure 52: Power curves for two-sided conditional and unbiased t-tests based on the Fuller estimator: $\rho = 0.5$.



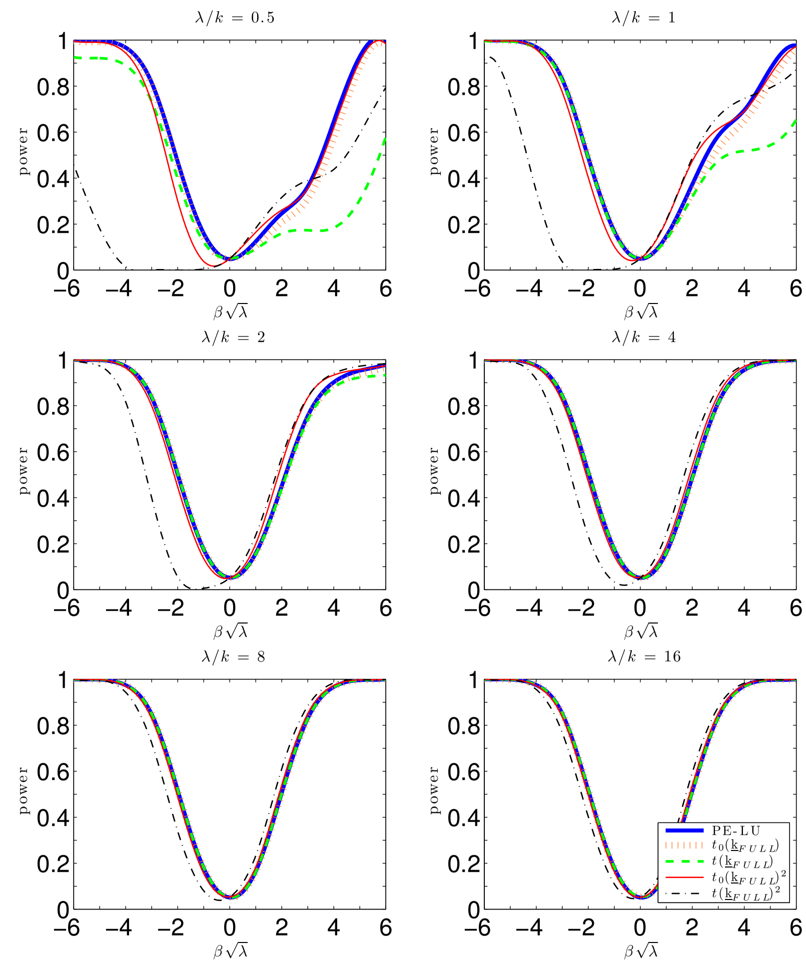
(a) $k = 2$

(b) $k = 5$

Figure 53: Power curves for two-sided conditional and unbiased t-tests based on the Fuller estimator: $\rho = 0.9$.

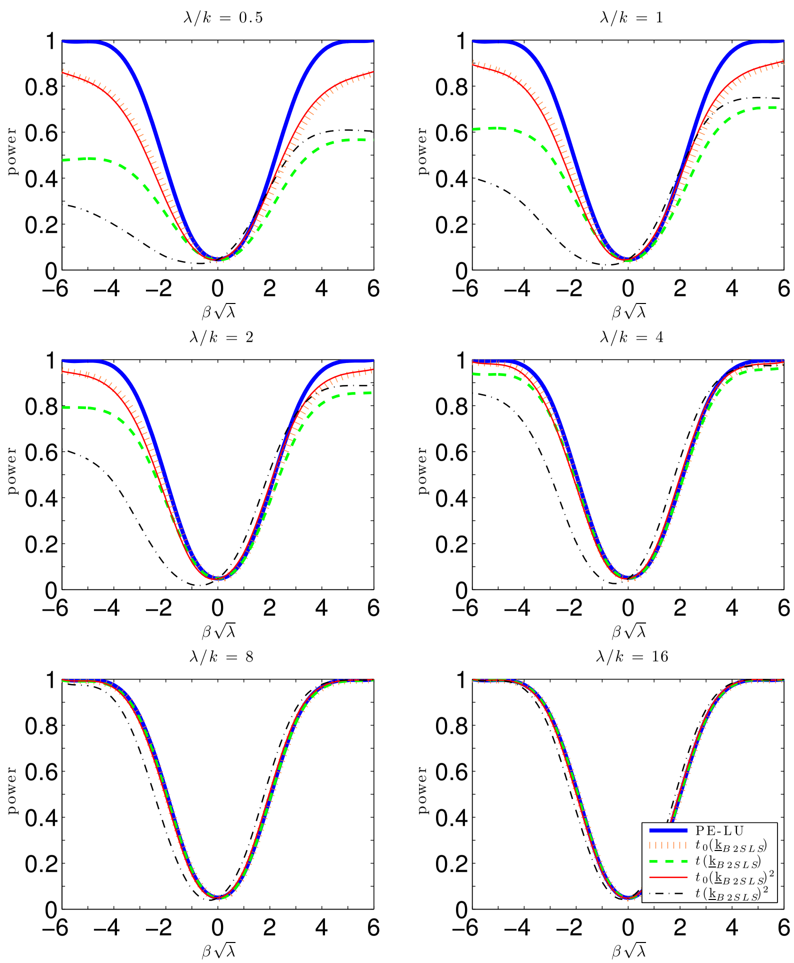


(a) $k = 10$

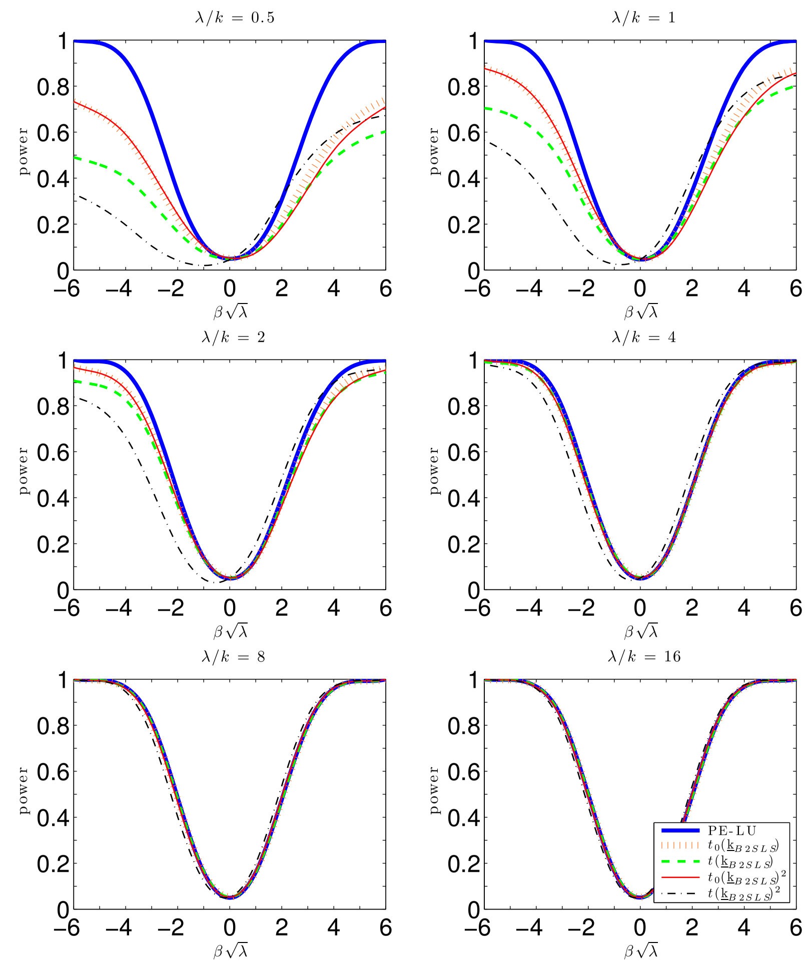


(b) $k = 20$

Figure 54: Power curves for two-sided conditional and unbiased t-tests based on the Fuller estimator: $\rho = 0.9$.

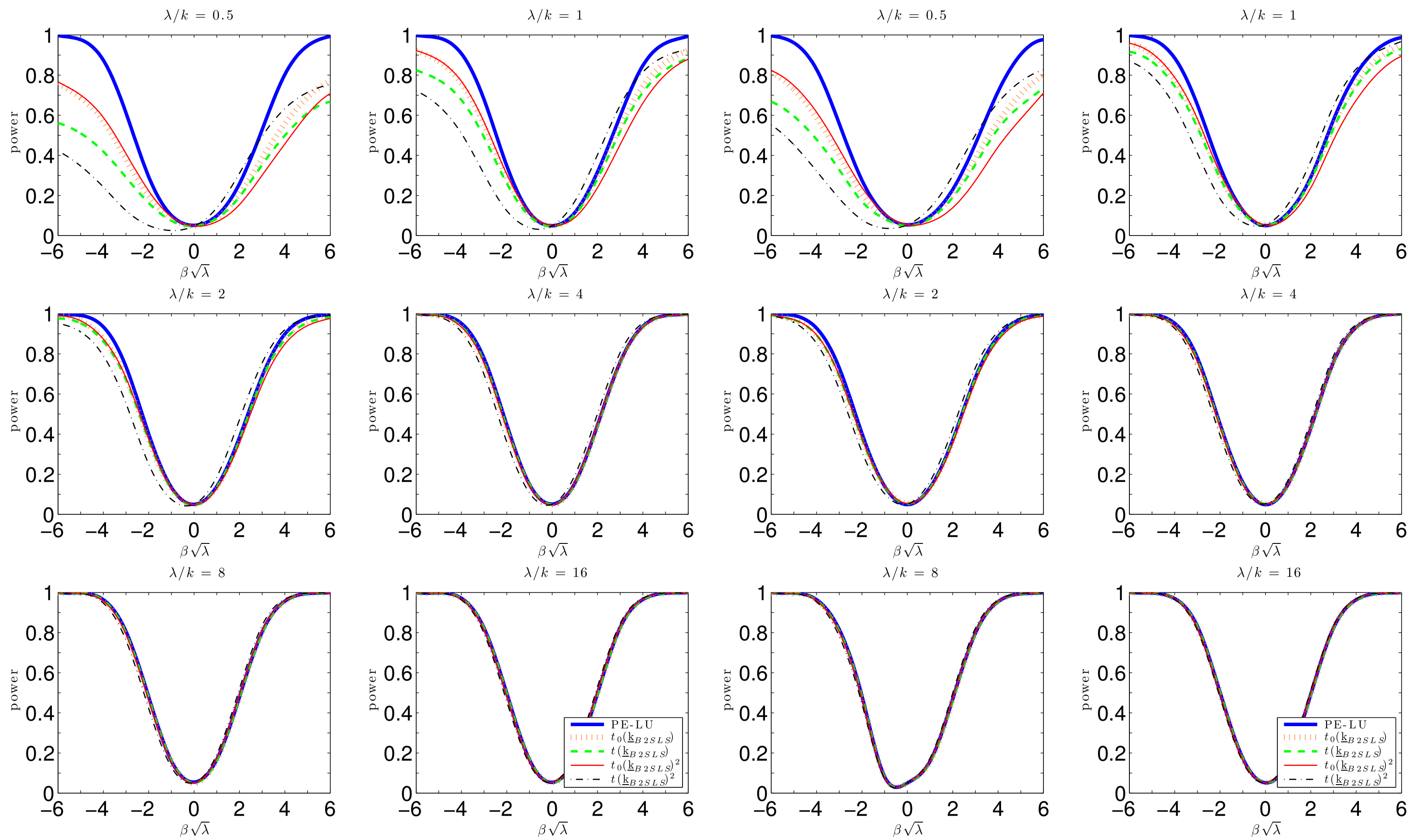


(a) $k = 2$



(b) $k = 5$

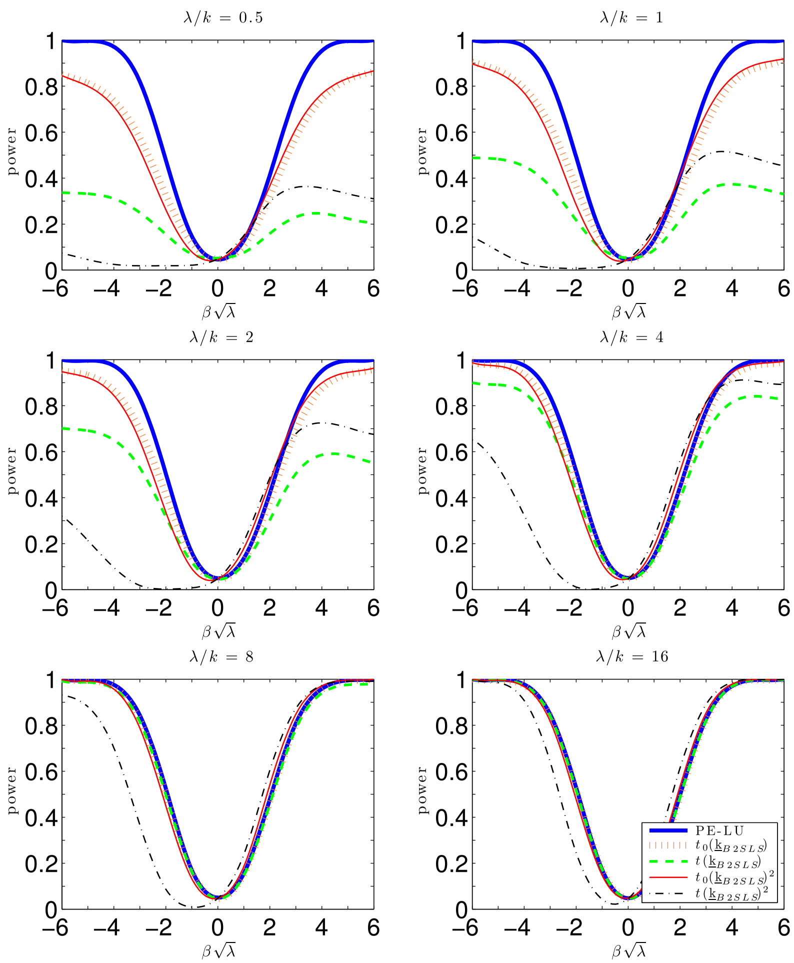
Figure 55: Power curves for two-sided conditional and unbiased t-tests based on the B2SLS estimator: $\rho = 0.2$.



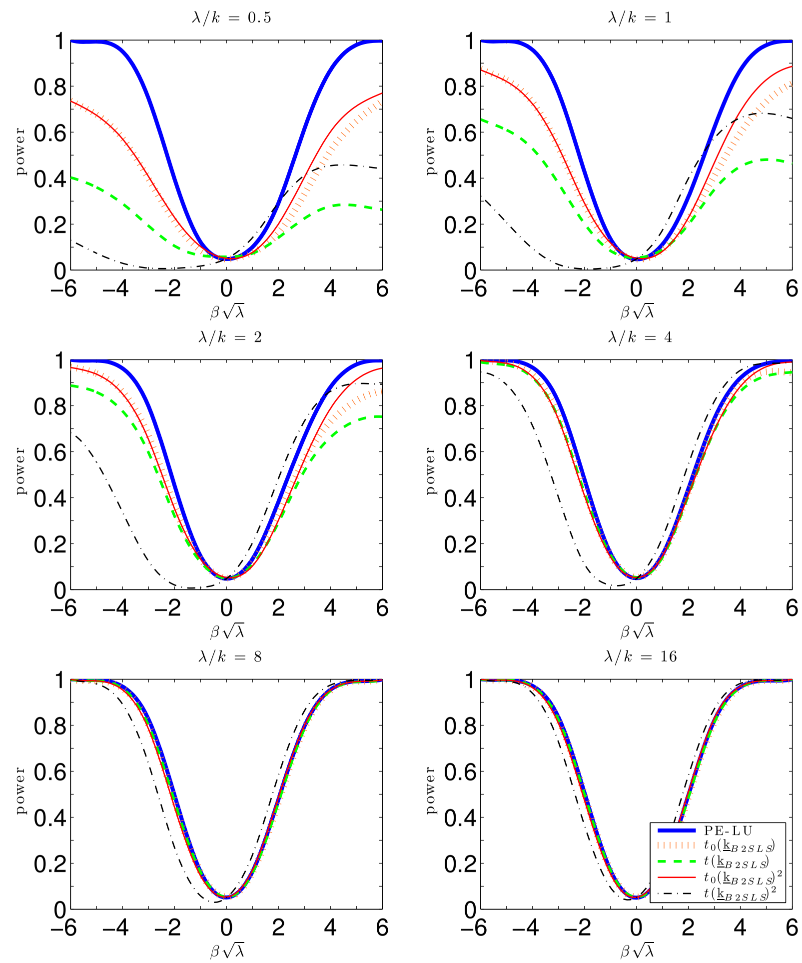
(a) $k = 10$

(b) $k = 20$

Figure 56: Power curves for two-sided conditional and unbiased t-tests based on the B2SLS estimator: $\rho = 0.2$.

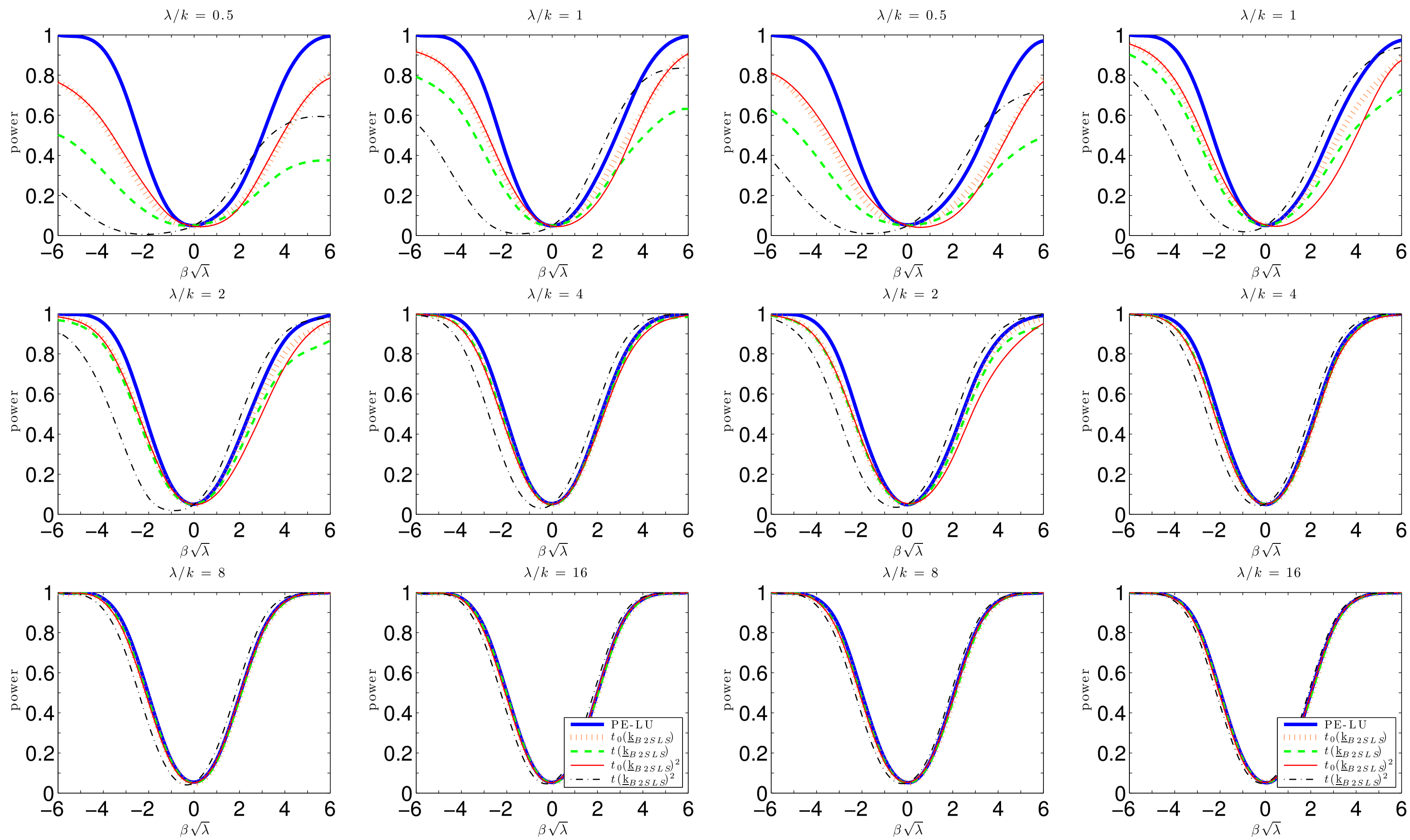


(a) $k = 2$



(b) $k = 5$

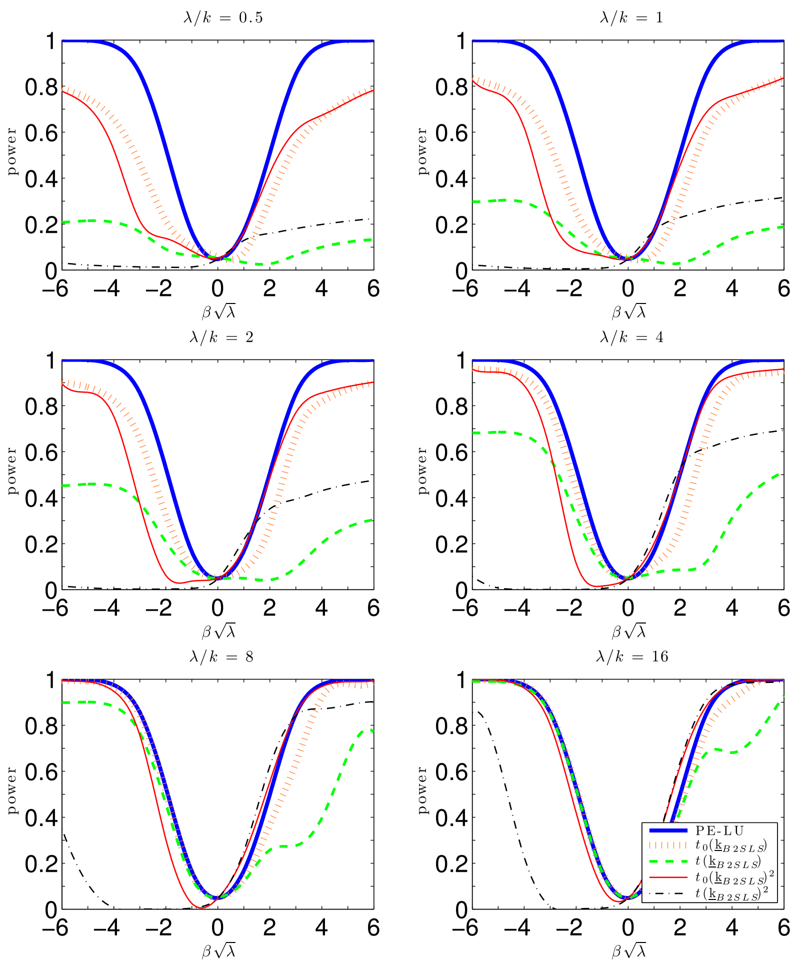
Figure 57: Power curves for two-sided conditional and unbiased t-tests based on the B2SLS estimator: $\rho = 0.5$.



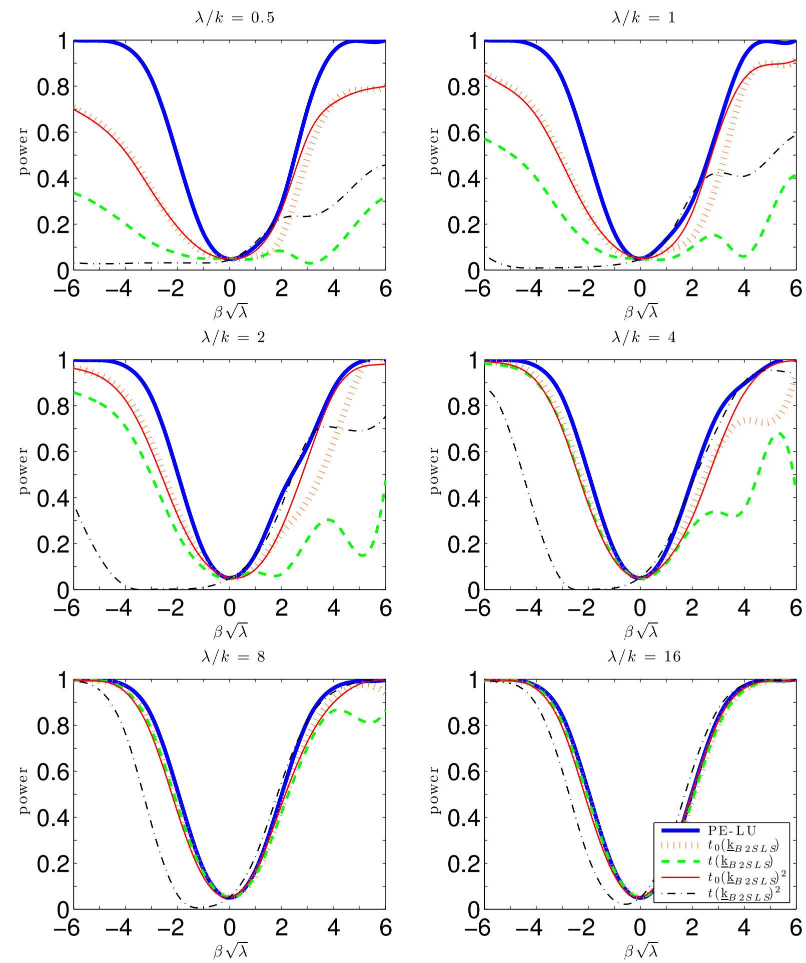
(a) $k = 10$

(b) $k = 20$

Figure 58: Power curves for two-sided conditional and unbiased t-tests based on the B2SLS estimator: $\rho = 0.5$.

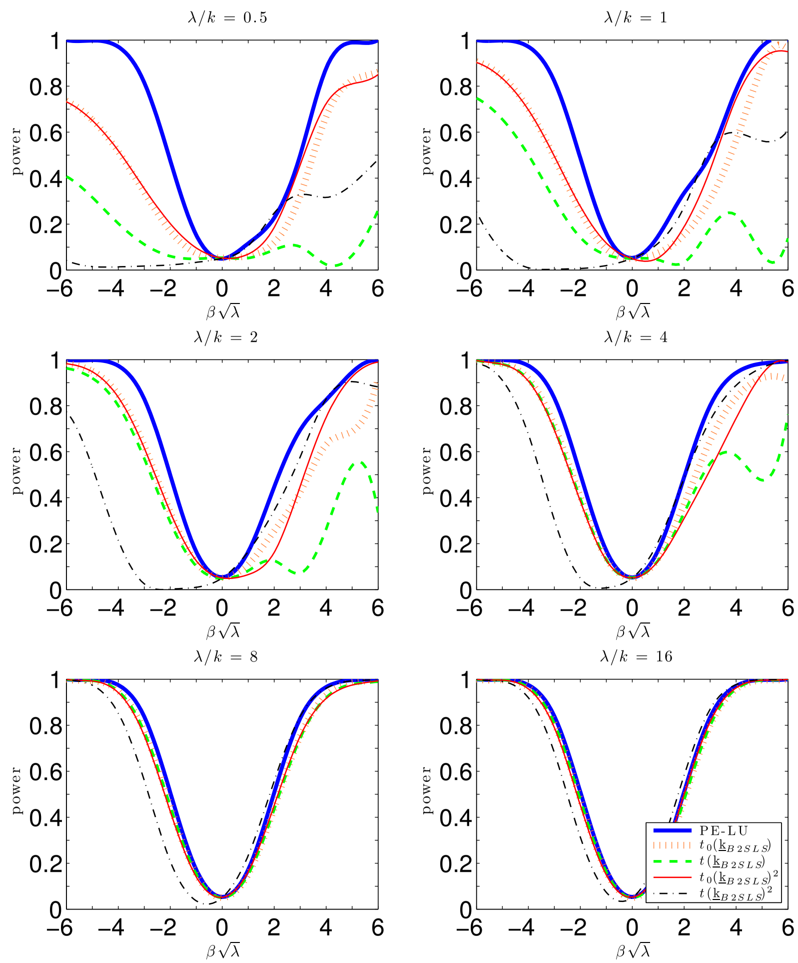


(a) $k = 2$

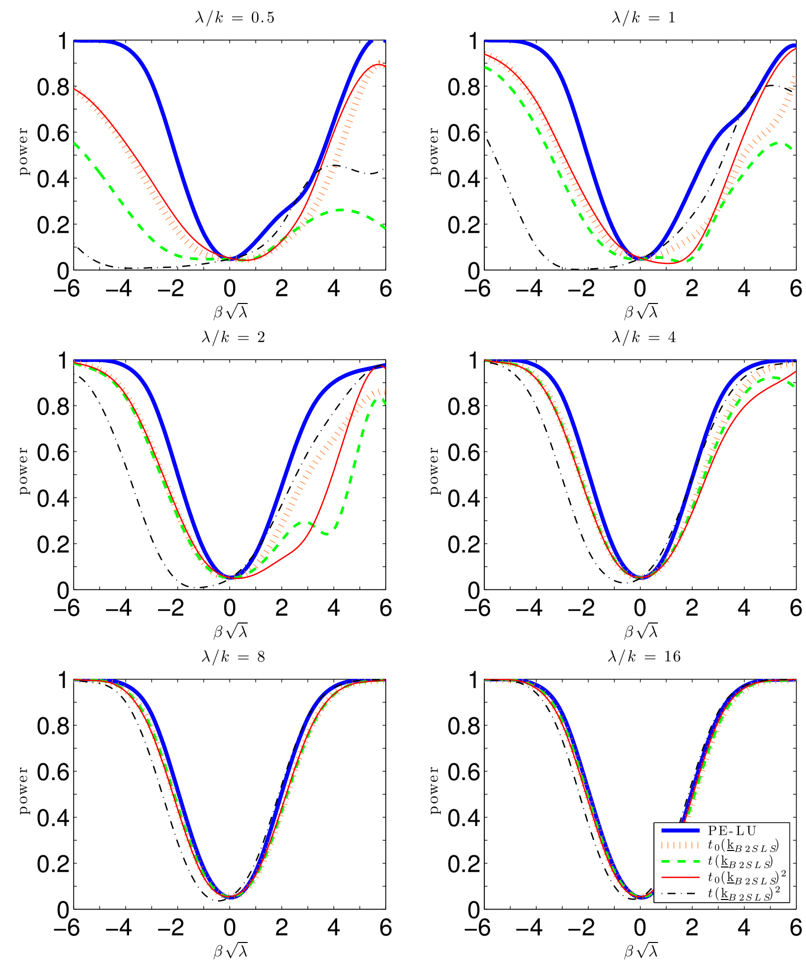


(b) $k = 5$

Figure 59: Power curves for two-sided conditional and unbiased t-tests based on the B2SLS estimator: $\rho = 0.9$.



(a) $k = 10$



(b) $k = 20$

Figure 60: Power curves for two-sided conditional and unbiased t-tests based on the B2SLS estimator: $\rho = 0.9$.

Table 1: Effects of Years of Education on Log Weekly Earnings (1920-29 cohort)

	I	II	III	IV
$\hat{\beta}_{OLS}$	0.0701	0.0701	0.0701	0.0692
(s.e.)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
$\hat{\beta}_{LIML}$	0.0595	0.0691	0.2615	0.1390
(s.e.)	(0.0166)	(0.0151)	(0.1116)	(0.0196)
$\hat{\beta}_{2SLS}$	0.0594	0.0688	0.1053	0.0936
(s.e.)	(0.0166)	(0.0175)	(0.0331)	(0.0109)
$\hat{\beta}_{FULL}$	0.0595	0.0688	0.2318	0.1383
(s.e.)	(0.0165)	(0.0174)	(0.1026)	(0.0195)
$\hat{\beta}_{B2SLS}$	0.0594	0.0689	0.3640	0.1365
(s.e.)	(0.0166)	(0.0170)	(0.1821)	(0.0191)
	One-sided 95% Confidence Intervals			
$CMLR1$	$[0.031, \infty]$	$[0.035, \infty]$	$[-\infty, \infty]$	$[0.089, \infty]$
$t_0\,(\underline{k}_{LIML})$	$[0.031, \infty]$	$[0.035, \infty]$	$[-\infty, \infty]$	$[0.089, \infty]$
$t\,(\underline{k}_{LIML})$	$[0.031, \infty]$	$[0.035, \infty]$	$[-\infty, \infty]$	$[0.089, \infty]$
$t_0\,(\underline{k}_{2SLS})$	$[0.030, \infty]$	$[0.036, \infty]$	$[-\infty, \infty]$	$[0.087, \infty]$
$t\,(\underline{k}_{2SLS})$	$[0.030, \infty]$	$[0.036, \infty]$	$[-\infty, \infty]$	$[0.087, \infty]$
$t_0\,(\underline{k}_{FULL})$	$[0.031, \infty]$	$[0.035, \infty]$	$[-\infty, \infty]$	$[0.089, \infty]$
$t\,(\underline{k}_{FULL})$	$[0.031, \infty]$	$[0.035, \infty]$	$[-\infty, \infty]$	$[0.089, \infty]$
$t_0\,(\underline{k}_{B2SLS})$	$[0.030, \infty]$	$[0.036, \infty]$	$[-\infty, \infty]$	$[0.087, \infty]$
$t\,(\underline{k}_{B2SLS})$	$[0.030, \infty]$	$[0.040, \infty]$	$[-\infty, \infty]$	$[0.106, \infty]$
	Two-sided 95% Confidence Intervals			
CLR	$[0.025, 0.093]$	$[0.028, 0.109]$	$[-\infty, \infty]$	$[0.080, 0.214]$
$t_0\,(\underline{k}_{LIML})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$[-\infty, \infty]$	$[0.080, 0.213]$
$t\,(\underline{k}_{LIML})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$[-\infty, \infty]$	$[0.080, 0.213]$
$t_0\,(\underline{k}_{2SLS})$	$[0.025, 0.093]$	$[0.029, 0.107]$	$[-\infty, \infty]$	$[0.077, 0.236]$
$t\,(\underline{k}_{2SLS})$	$[0.025, 0.093]$	$[0.029, 0.107]$	$[-\infty, \infty]$	$[0.077, 0.236]$
$t_0\,(\underline{k}_{FULL})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$[-\infty, \infty]$	$[0.080, 0.213]$
$t\,(\underline{k}_{FULL})$	$[0.025, 0.093]$	$[0.028, 0.109]$	$[-\infty, \infty]$	$[0.080, 0.213]$
$t_0\,(\underline{k}_{B2SLS})$	$[0.025, 0.093]$	$[0.029, 0.108]$	$[-\infty, \infty]$	$[0.101, 0.228]$
$t\,(\underline{k}_{B2SLS})$	$[0.025, 0.093]$	$[0.034, 0.103]$	$[-\infty, \infty]$	$[0.077, 1830]$
F (first stage)	37.971	4.538	1.055	1.553
Number of instruments	3	30	28	178

Table 2: Effects of Years of Education on Log Weekly Earnings (1930-39 cohort)

	I	II	III	IV
$\hat{\beta}_{OLS}$	0.0632	0.0632	0.0632	0.0628
(s.e.)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
$\hat{\beta}_{LIML}$	0.0999	0.0838	0.0574	0.0982
(s.e.)	(0.0210)	(0.0179)	(0.0385)	(0.0153)
$\hat{\beta}_{2SLS}$	0.0990	0.0806	0.0600	0.0811
(s.e.)	(0.0207)	(0.0164)	(0.0290)	(0.0109)
$\hat{\beta}_{FULL}$	0.0995	0.0836	0.0577	0.0980
(s.e.)	(0.0209)	(0.0178)	(0.0378)	(0.0153)
$\hat{\beta}_{B2SLS}$	0.0994	0.0848	0.0555	0.1016
(s.e.)	(0.0208)	(0.0183)	(0.0445)	(0.0161)
	One-sided 95% Confidence Intervals			
$CMLR1$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.184, \infty]$	$[0.061, \infty]$
$t_0\,(\underline{k}_{LIML})$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.184, \infty]$	$[0.061, \infty]$
$t\,(\underline{k}_{LIML})$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.075, \infty]$	$[0.062, \infty]$
$t_0\,(\underline{k}_{2SLS})$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.378, \infty]$	$[0.063, \infty]$
$t\,(\underline{k}_{2SLS})$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.270, \infty]$	$[0.063, \infty]$
$t_0\,(\underline{k}_{FULL})$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.105, \infty]$	$[0.061, \infty]$
$t\,(\underline{k}_{FULL})$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.076, \infty]$	$[0.062, \infty]$
$t_0\,(\underline{k}_{B2SLS})$	$[0.065, \infty]$	$[0.051, \infty]$	$[-0.946, \infty]$	$[0.063, \infty]$
$t\,(\underline{k}_{B2SLS})$	$[0.065, \infty]$	$[0.054, \infty]$	$[-0.384, \infty]$	$[0.075, \infty]$
	Two-sided 95% Confidence Intervals			
CLR	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-0.187, 0.285]$	$[0.055, 0.145]$
$t_0\,(\underline{k}_{LIML})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-0.182, 0.335]$	$[0.054, 0.145]$
$t\,(\underline{k}_{LIML})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-\infty, \infty]$	$[0.054, 0.145]$
$t_0\,(\underline{k}_{2SLS})$	$[0.059, 0.145]$	$[0.044, 0.128]$	$[-\infty, \infty]$	$[0.056, 0.155]$
$t\,(\underline{k}_{2SLS})$	$[0.059, 0.145]$	$[0.044, 0.128]$	$[-\infty, \infty]$	$[0.056, 0.155]$
$t_0\,(\underline{k}_{FULL})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-\infty, \infty]$	$[0.054, 0.145]$
$t\,(\underline{k}_{FULL})$	$[0.059, 0.145]$	$[0.044, 0.125]$	$[-\infty, \infty]$	$[0.054, 0.145]$
$t_0\,(\underline{k}_{B2SLS})$	$[0.059, 0.145]$	$[0.044, 0.128]$	$[-\infty, \infty]$	$[0.055, 0.136]$
$t\,(\underline{k}_{B2SLS})$	$[0.059, 0.145]$	$[0.048, 0.123]$	$[-\infty, \infty]$	$[0.070, 0.155]$
F (first stage)	30.528	4.748	1.613	1.870
Number of instruments	3	30	28	178

Table 3: Effects of Years of Education on Log Weekly Earnings (1940-49 cohort)

	I	II	III	IV
$\hat{\beta}_{OLS}$	0.0632	0.0632	0.0632	0.0516
(s.e.)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
$\hat{\beta}_{LIML}$	-0.0902	0.0286	0.1243	0.0878
(s.e.)	(0.0301)	(0.0197)	(0.0420)	(0.0178)
$\hat{\beta}_{2SLS}$	-0.0734	0.0393	0.0779	0.0666
(s.e.)	(0.0273)	(0.0145)	(0.0239)	(0.0113)
$\hat{\beta}_{FULL}$	-0.0882	0.0289	0.1218	0.0875
(s.e.)	(0.0299)	(0.0196)	(0.0412)	(0.0177)
$\hat{\beta}_{B2SLS}$	-0.0750	0.0373	0.0912	0.08725
(s.e.)	(0.0275)	(0.0156)	(0.0297)	(0.0164)
	One-sided 95% Confidence Intervals			
$CMLR1$	$[-0.148, \infty]$	$[-0.008, \infty]$	$[0.043, \infty]$	$[0.046, \infty]$
$t_0\,(\underline{k}_{LIML})$	$[-0.148, \infty]$	$[-0.008, \infty]$	$[0.043, \infty]$	$[0.046, \infty]$
$t\,(\underline{k}_{LIML})$	$[-0.148, \infty]$	$[-0.007, \infty]$	$[0.041, \infty]$	$[0.046, \infty]$
$t_0\,(\underline{k}_{2SLS})$	$[-0.130, \infty]$	$[0.009, \infty]$	$[0.031, \infty]$	$[0.044, \infty]$
$t\,(\underline{k}_{2SLS})$	$[-0.130, \infty]$	$[0.009, \infty]$	$[0.031, \infty]$	$[0.044, \infty]$
$t_0\,(\underline{k}_{FULL})$	$[-0.148, \infty]$	$[-0.008, \infty]$	$[0.043, \infty]$	$[0.046, \infty]$
$t\,(\underline{k}_{FULL})$	$[-0.148, \infty]$	$[-0.007, \infty]$	$[0.040, \infty]$	$[0.046, \infty]$
$t_0\,(\underline{k}_{B2SLS})$	$[-0.130, \infty]$	$[0.008, \infty]$	$[0.031, \infty]$	$[0.044, \infty]$
$t\,(\underline{k}_{B2SLS})$	$[-0.130, \infty]$	$[0.011, \infty]$	$[0.043, \infty]$	$[0.055, \infty]$
	Two-sided 95% Confidence Intervals			
CLR	$[-0.161, -0.036]$	$[-0.015, 0.071]$	$[0.026, 0.262]$	$[0.038, 0.142]$
$t_0\,(\underline{k}_{LIML})$	$[-0.161, -0.036]$	$[-0.015, 0.070]$	$[0.028, 0.264]$	$[0.037, 0.141]$
$t\,(\underline{k}_{LIML})$	$[-0.160, -0.036]$	$[-0.015, 0.070]$	$[0.024, 0.256]$	$[0.037, 0.141]$
$t_0\,(\underline{k}_{2SLS})$	$[-0.141, -0.026]$	$[0.003, 0.070]$	$[0.019, 0.179]$	$[0.036, 0.133]$
$t\,(\underline{k}_{2SLS})$	$[-0.141, -0.026]$	$[0.003, 0.070]$	$[0.019, 0.178]$	$[0.036, 0.133]$
$t_0\,(\underline{k}_{FULL})$	$[-0.161, -0.036]$	$[-0.015, 0.070]$	$[0.027, 0.263]$	$[0.037, 0.141]$
$t\,(\underline{k}_{FULL})$	$[-0.160, -0.036]$	$[-0.015, 0.070]$	$[0.023, 0.255]$	$[0.037, 0.141]$
$t_0\,(\underline{k}_{B2SLS})$	$[-0.142, -0.025]$	$[0.003, 0.071]$	$[0.019, 0.180]$	$[0.036, 0.134]$
$t\,(\underline{k}_{B2SLS})$	$[-0.141, -0.026]$	$[0.005, 0.068]$	$[0.034, 0.160]$	$[0.050, 0.117]$
F (first stage)	26.316	6.849	2.736	1.929
Number of instruments	3	30	28	178

1. References

D. W. K. Andrews, M. J. Moreira, and J. H. Stock. Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression. *Econometrica*, 74(3):715–752, 2006a.

D. W. K. Andrews, M. J. Moreira, and J. H. Stock. Supplement to 'Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression'. *Econometrica*, 74(3):pp. 715–752, 2006b.

J. Angrist and A. B. Krueger. Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, 106:979–1014, 1991.

L. M. Cruz and M. J. Moreira. On the Validity of Econometric Techniques With Weak Instruments: Inference on Returns to Education Using Compulsory School Attendance Laws. *Journal of Human Resources*, 40:393–410, 2005.

W. A. Fuller. Some Properties of a Modification of the Limited Information Estimator. *Econometrica*, 45:939–953, 1977.

M. J. Moreira. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, 71 (4):1027–1048, 2003.

A. Nagar. The Bias and Moment Matrix of the General k-Class Estimators of Parameters in Simultaneous Equations. *Econometrica*, 27:575–595, 1959.

T. J. Rothenberg. Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches and M. D. Intriligator, editors, *Handbook of Econometrics*, volume 2, chapter 15, pages 881–935. Elsevier Science, Amsterdam, 1984.