

Unbiased Testing Under Weak Instrumental Variables

Abstract

This paper finds unbiased tests using three of Nagar's [1959] k -class estimators: two-stage least squares (2SLS), limited information maximum likelihood (LIML), and Fuller's [1977] modified LIML (FULL). Andrews et al. [2007] show that, using the conditional framework proposed by Moreira [2003], Wald tests based on these k -class estimators are biased and have poor power properties when instruments are weak. This paper introduces a new methodology that takes into account the asymmetry of the distribution of the t -statistic in the presence of weak instrumental variables. Using this framework, critical values that allow for unbiased testing using k -class estimators can be found. The power properties of the conditional t -test introduced in this paper are compared with that of other tests that are known to be robust to weak instrumental variables. The conditional t -test based on the three k -class estimators is unbiased and has good power. In particular, the conditional t -test based on the LIML estimator has power properties nearly identical to that of the conditional likelihood ratio test (CLR).

Benjamin Mills
Senior Honors Thesis
Advisor: Marcelo J. Moreira

1 Introduction

Economists are often interested in estimating and making inference about the parameter β in the linear model

$$y_1 = y_2\beta + X\gamma + u \quad (1.01)$$

with N observations, where $y_1, y_2 \in \mathbb{R}^N$ are endogenous variables, X is a $N \times l$ matrix of exogenous regressors, and $u \in \mathbb{R}^N$ is a vector of normally distributed i.i.d. random error terms with variance σ_u^2 . We let the subscript i denote the i th observation.

A commonly used estimator of β is the ordinary least squares (OLS) estimator $\hat{\beta}_{OLS}$:

$$\hat{\beta}_{OLS} = (y_2'y_2)^{-1} y_2'y_1 \quad (1.02)$$

For $\hat{\beta}_{OLS}$ to be consistent, it is necessary that y_2 is orthogonal to the error term, that is $E(y_{2i}u_i) = 0$ for any observation i . In the case of 1.0.1, $\hat{\beta}_{OLS}$ is not a consistent estimator for β since y_2 is assumed to be endogenous.

One way to overcome the problem of an endogenous regressor is the use of instrumental variables. A matrix of instrumental variables for 1.0.1 is a $N \times k$ matrix Z that is orthogonal to u and correlated with y_2 . Given valid instruments Z , a commonly used estimator of β is the 2SLS estimator

$$\hat{\beta}_{2SLS} = \frac{y_2'^{\perp} P_Z y_1^{\perp}}{y_2'^{\perp} P_Z y_2^{\perp}} \quad (1.03)$$

where $P_A = A(A'A)^{-1}A'$ is the projection matrix onto the column space of A and $B^{\perp} = (I_N - P_X)B$, is the projection of B onto the space orthogonal to the column space of X . It can be shown that if the instruments are strongly correlated with the endogenous regressor the distribution of $\hat{\beta}_{2SLS}$ nonstandard, even in large samples [Staiger and Stock, 1997]. This affects the distribution of any statistic used to do to inference, such as that of the t -statistic based on the two-stage least squares (2SLS) estimator,

$$t_{2SLS} = \frac{(\hat{\beta}_{2SLS} - \beta_0)}{\hat{\sigma}_u} \cdot (y_2'^{\perp} P_Z y_2^{\perp} - \kappa\omega_{22})^{1/2} \quad (1.04)$$

where $\hat{\sigma}_u$ is a consistent estimator of σ_u . Under strong instruments, the distribution of the t_{2SLS} is close to standard normal in large samples. When the instrumental variables and the endogenous variable are weakly correlated, properties of the normal distribution can not be used to conduct inference. In particular,

commonly used statistical tests, such as the Wald test, exhibit size distortion under weak instruments.

Particular attention has been paid to tests with correct size under weak instruments. Moreira [2003] proposes a conditional framework whereby the standard critical values used for hypothesis testing are replaced by critical values that are a function of the data. By conditioning on the weakness of the instruments, critical values that correct for the size distortion are derived from the conditional distribution of the test statistic. Andrews et al. [2006a] examined the conditional likelihood ratio test (CLR) proposed by Moreira [2003] and found it had correct size as well as good power compared to the Lagrange multiplier and the Anderson-Rubin tests. Andrews et al. [2007] numerically investigated the properties of the conditional Wald test based on four different estimators and found that the conditional Wald has correct size. However, the conditional Wald is biased weak instruments. That is, the test often rejects the null hypothesis with a higher probability under the null than under some alternatives. The goal of this paper is to introduce a methodology that corrects for the bias of the conditional Wald test in the presence of weak instruments.

Section 1.1 introduces the structural IV model and introduces the k -class estimators of Nagar [1959]. Section 2 gives a precise but brief overview of hypothesis testing and unbiasedness. Section 2.1 describes the conditional framework and how it applies to the IV model. Section 2.2 demonstrates numerically the asymmetry of the distribution of t_{2SLS} , both unconditional and conditional, under weak instruments. Section 3 introduces the critical values that will be central to the unbiased test. Section 3.1 provides the theoretical justification for the unbiasedness of the test and develops an algorithm to find the desired critical values. Section 3.2 the unbiased conditional t -test. Section 3.3 finds confidence intervals based on t_{2SLS} and investigates the behavior of the confidence intervals under a variety of parameters. Section 4 provides numerical power results for the conditional t -test and compares it to various other tests. Section 5 concludes the paper. Section 6 provides a comprehensive appendix of supplementary material. Section 6.1 provides the derivation of the t -statistic in a form required by the conditional framework. Sections 6.2 and 6.3 provide power curves for the conditional t -test under every parameter combination considered. Section 6.4 gives proofs to the theoretical results stated in the paper. Finally, section 6.5 provides a URL for all code and other files necessary to replicate all numerical results and graphs in the paper.

1.1 The Structural IV Model

Following Andrews et al. [2006a], we consider the structural equation equation for a single endogenous

regressor

$$y_1 = y_2\beta + X\gamma + u \quad (1.1.1)$$

$$y_2 = Z\pi + X\xi + v_2 \quad (1.1.2)$$

where $y_1, y_2 \in \mathbb{R}^N$ are endogenous variables, X is a $N \times l$ matrix of exogenous regressors, and Z as a $N \times k$ matrix of instrumental variables. We assume that the exogenous regressors X and the instruments Z are orthogonal, since if this were not the case, we could always redefine the instruments to be a projection of Z onto the space orthogonal to X . The corresponding reduced form model, written in matrix notation, is

$$Y = Z\pi a' + X\eta + V, \quad (1.1.3)$$

where

$$a = \begin{bmatrix} \beta_0 \\ 1 \end{bmatrix} \text{ and } \eta = [\gamma, \xi]$$

are parameters and

$$Y = [y_1, y_2] \text{ and } V = [v_1, v_2]$$

are random matrices. V is assumed to have a multivariate normal distribution $N(0, I_N \otimes \Omega)$, where

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21} \\ \omega_{12} & \omega_{22} \end{bmatrix}. \quad (1.1.4)$$

Since Ω can be consistently estimated (even when instruments are weak) by

$$\hat{\Omega} = \frac{Y'Z(Z'Z)^{-1}Z'Y}{n - k - l}, \quad (1.1.5)$$

[Andrews et al., 2006b] we assume that Ω is known. We define

$$\rho = \frac{\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}} \quad (1.1.6)$$

to be the correlation between the reduced form errors; that is, ρ measures the level of endogeneity of y_2 .

This paper investigates the two-sided hypothesis test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ under weak

instruments using k -class estimators of β . The k -class of estimators of β are defined by

$$\hat{\beta}_\kappa = \frac{(y_2' P_z y_1 - \kappa \omega_{22})}{(y_2' P_z y_2 - \kappa \omega_{12})} \quad (1.1.7)$$

where κ is a parameter that dictates the variety of k -class estimator. We focus on three k -class estimators: two-stage least squares (2SLS), limited information maximum likelihood (LIML), and a modified LIML proposed by Fuller [1977] (FULL). The corresponding κ for each estimator is given by

$$\begin{aligned} \kappa_{2SLS} &= 0 \\ \kappa_{LIML} &= \text{The smallest root of } f(\kappa) = \det(Y' P_Z Y - \kappa \Omega) = 0 \quad (1.1.8) \\ \kappa_{FULL} &= (N - k) (1 + \kappa_{LIML}/(N - k)) . \end{aligned}$$

2 Hypothesis Testing and Unbiasedness

In order to make inference about the parameters of a model $(\beta, \theta) \in \mathbb{R}^{m+1}$, where $\beta \in \mathbb{R}$ is a parameter of interest and $\theta \in \mathbb{R}^m$ is a vector of nuisance parameters, economists generally rely on hypothesis testing. Typically one will test a null hypothesis

$$H_0 : \beta = \beta_0$$

against an alternative hypothesis

$$H_1 : \beta \neq \beta_0.$$

This is equivalent to testing

$$H_0 : (\beta, \theta) \in \mathcal{B}_0 = \{\beta_0\} \times \mathbb{R}^m$$

against

$$H_1 : (\beta, \theta) \in \mathcal{B}_1 = (\mathbb{R} \setminus \{\beta_0\}) \times \mathbb{R}^m$$

We call \mathcal{B}_0 the *null set* and \mathcal{B}_1 the *alternative set*.

A test $\phi(X)$ is a function of the data X such that it takes on the value 1 to indicate rejection of the null hypothesis and the value 0 to indicate failure to reject the null hypothesis. Under the Neyman-Pearson framework, we fix the probability of rejecting the null hypothesis at a level α and seek a test that maximizes the probability of rejecting the null hypothesis when an alternative is true. The probability that a test will reject the null when β is true is called the *power* of the test, $E_\beta \phi(X)$. A test ϕ for which the power function

(the power of the test as a function of β) $E_\beta\phi(X)$ satisfies

$$E_\beta\phi(X) \leq \alpha \quad \text{if } (\beta, \theta_1, \dots, \theta_m) \in \mathcal{B}_0$$

$$E_\beta\phi(X) \geq \alpha \quad \text{if } (\beta, \theta_1, \dots, \theta_m) \in \mathcal{B}_1$$

is said to be *unbiased* [Lehmann and Romano, 2005]. If a test is biased then there exist alternatives where the test is more likely to accept the null than when the null is true. This is clearly an undesirable property, hence it is important that a test be unbiased when making inference about a parameter β .

2.1 The Conditioning Argument

The goal of the conditional framework is to control for the effect of π , which dictates the strength of the instruments. By Andrews et al. [2006a] Lemma 1e, $Z'Y$ is a sufficient statistic for $(\beta, \pi)'$, which eliminates the nuisance parameter η from the problem. Following Moreira [2003], we establish the one-to-one transformation of $Z'Y$:

$$S = (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0' \Omega b_0)^{-1/2} \quad (2.1.2)$$

$$T = (Z'Z)^{-1/2} Z'Y \Omega a_0 \cdot (a_0' \Omega a_0)^{-1/2} \quad (2.1.3)$$

where

$$a_0 = \begin{bmatrix} \beta_0 \\ 1 \end{bmatrix} \quad \text{and} \quad b_0 = \begin{bmatrix} 1 \\ -\beta_0 \end{bmatrix},$$

and define

$$Q = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix} = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} \quad (2.1.4)$$

where Q has a non-central Wishart distribution. The distribution of Q depends only on π through the nonnegative scalar

$$\lambda = \pi' Z' Z \pi \quad (2.1.5)$$

[Andrews et al., 2006a] that measures the strength of the instruments. The parameter λ has a direct connection with the first stage F-test statistic used to test for weak instruments. Define

$$\hat{\lambda} = \hat{\pi}' Z' Z \hat{\pi} \quad (2.1.6)$$

where $\hat{\pi}$ is the OLS estimate of π obtained by regressing y_2 on Z . The first stage F-test statistic is defined by

$$F = \frac{\hat{\lambda}}{k \cdot \hat{\omega}_{22}} \quad (2.1.7)$$

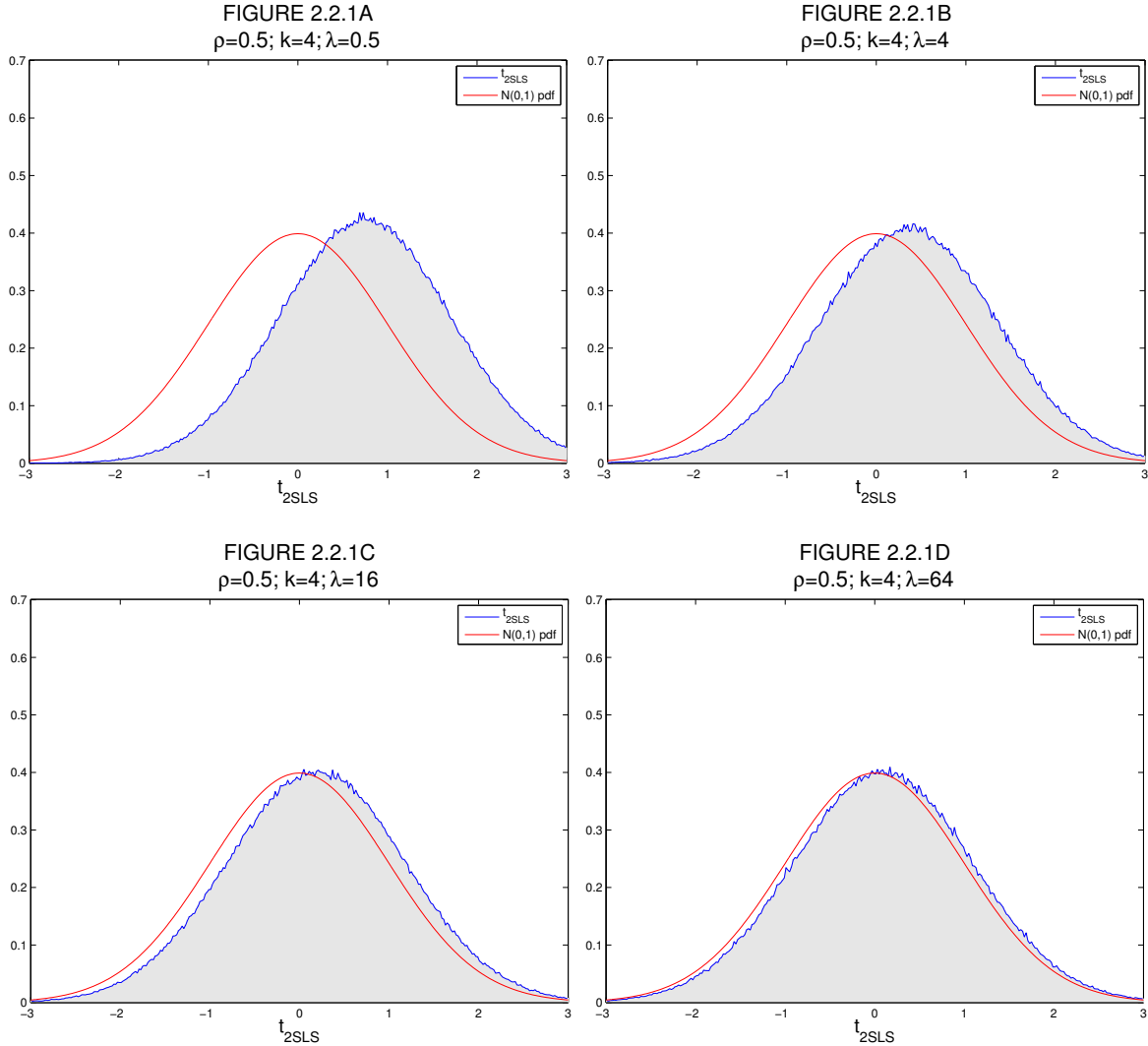
where $\hat{\omega}_{22}$ is a consistent estimator for the variance of the error term v_2 . Staiger and Stock [1997] proposed the rule-of-thumb that a value of the first stage F-test statistic that is less than 10 indicates that instruments are weak.

Because π represents the effect instruments have on the exogenous regressor, π determines the weakness of the instruments. The null rejection probability of conventional tests depend on π . We can eliminate π from the problem by establishing that the statistic Q_T is sufficient for λ . Hence by conditioning on $Q_T = q_T$, λ is eliminated, which in turn eliminates π from the problem. By conditioning on q_T , which is a function of the data, we can establish distributional properties of the parameter of interest β given the level of weakness, and thus make inference.

2.2 Conditional t-Statistics

Staiger and Stock [1997] demonstrated numerically the distortion that occurs under weak instruments to the asymptotic probability distribution functions of the t -statistic based on the 2SLS estimator, t_{2SLS} . The distribution of t_{2SLS} is asymmetric when instruments are weak. For fixed parameters $\omega_{11} = \omega_{22} = 1$, $\rho = \omega_{12} = \omega_{21} = 0.5$ and $k = 4$, we set four values of λ : 0.5, 4, 16, and 64, where $\lambda = 0.5$ represents weak instruments and $\lambda = 64$ represents strong instruments. FIGURE 2.2.1 displays the sample probability distribution of t_{2SLS} as instruments get progressively stronger. The sample probability distributions were generated from 1,000,000 simulated values of t_{2SLS} each.

FIGURE 2.2.1: UNCONDITIONAL PDFS OF THE t_{2SLS} UNDER WEAK INSTRUMENTS



To illustrate this distortion in the conditional framework, this section contains numerically approximations of the probability distribution function of the t_{2SLS} given a value of q_T . By writing t_{2SLS} in terms of the sufficient statistics (Q_{ST}, Q_S, Q_T) , under the null, given $Q_T = q_T$, values of Q_{ST} and Q_S are randomly generated to produce simulated t_{2SLS} statistics conditional on q_T . The parameters are $\sigma = 1$, $\omega_{11} = \omega_{22} = 1$, $\rho = \omega_{12} = \omega_{21} = 0.5$ and $k = 4$. We set four values of q_T , defined by $\ln(q_T/k) = 0, 2, 4$, and 8 , where $\ln(q_T/k) = 0$ represents weak instruments and $\ln(q_T/k) = 8$ represents strong instruments. FIGURE 2.2.2 displays the sample probability distribution of t_{2SLS} as instruments get progressively stronger. The sample probability distributions were generated from 1,000,000 simulated values of t_{2SLS} each.

FIGURE 2.2.2: CONDITIONAL PDFS OF THE t_{2SLS} UNDER WEAK INSTRUMENTS

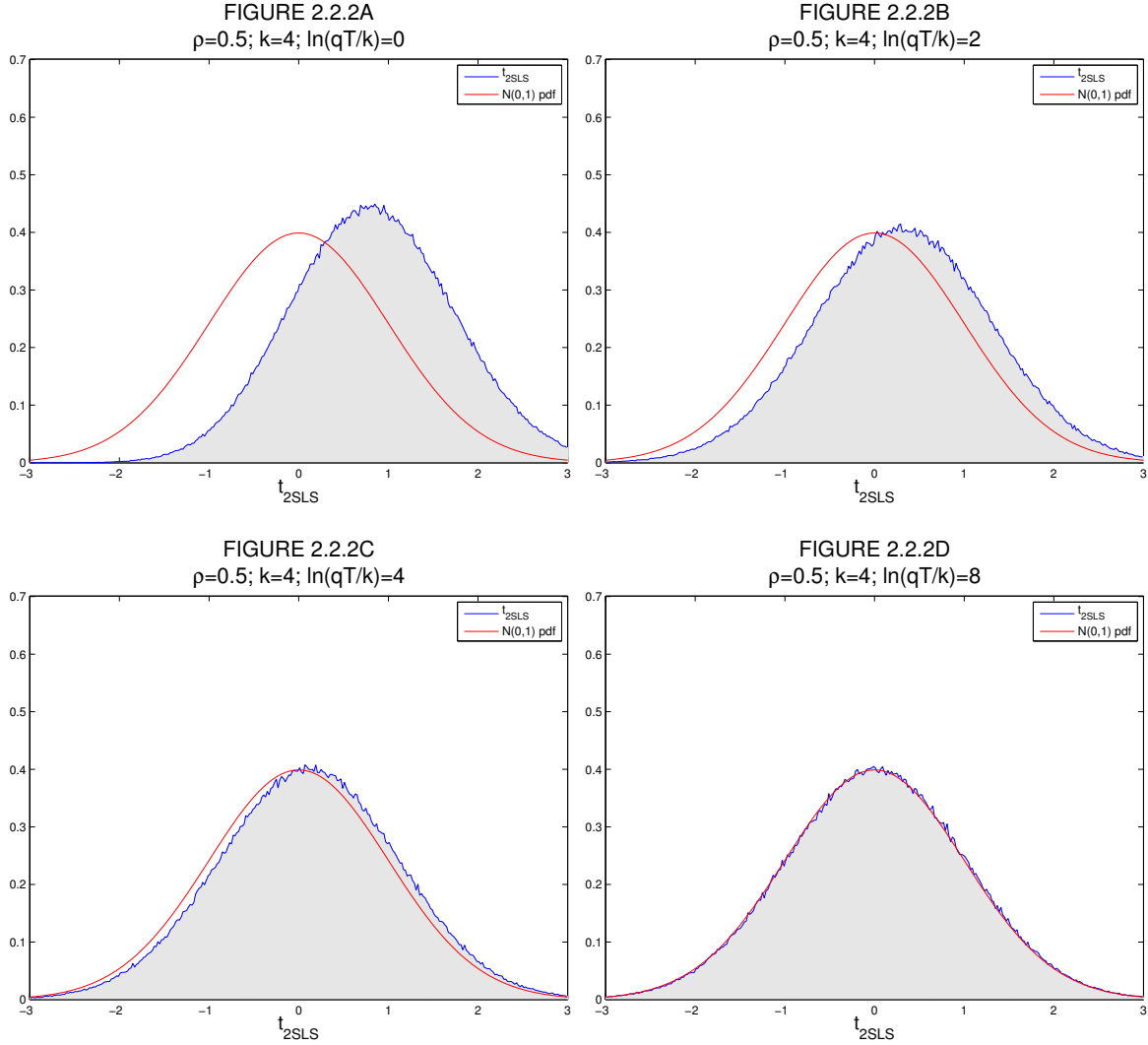


FIGURE 2.2.1A shows that under weak instruments the pdf of the t_{2SLS} statistic, even when conditioned on q_T , is asymmetric. As instruments get stronger, the pdf of the statistic tends towards that of the standard normal. Eventually when instruments are strong, the pdf of the statistic is almost perfectly that of a standard normal.

3 Critical Values

The primary goal of this paper is to find an unbiased two-sided test based on the k -class of estimators that is robust to weak instruments. To accomplish this we require necessary and sufficient conditions for unbiasedness. Following Andrews et al. [2006a], we consider the class of tests that are invariant to orthonormal transformations of the instrument variables. We consider the invariant test $\phi(Q_S, LM, Q_T)$, where

$LM = Q_{ST}/\sqrt{Q_T}$ is the test statistic of the one-sided Lagrange multiplier test of [Breusch and Pagan, 1980].

Andrews et al. [2006b] show that the two conditions

$$E_{\beta_0} \left(\phi(Q_S, LM, Q_T) \middle| Q_T = q_T \right) = \alpha \quad (3.0.1)$$

$$E_{\beta_0} \left(\phi(Q_S, LM, Q_T) \cdot LM \middle| Q_T = q_T \right) = 0, \quad (3.0.2)$$

for almost all q_T , are necessary for the invariant test to be unbiased with size α . A test that satisfies condition 3.0.1 is said to be *similar*. Condition 3.0.2 indicates that the derivative of the power function with respect to β under the null hypothesis is zero; a necessary condition for the power function of the of test to achieve its minimum at β_0 . Given q_T , these conditions can be written as

$$E_{\beta_0} (\phi(Q_S, LM, q_T)) = \alpha \quad (3.0.3)$$

$$E_{\beta_0} (\phi(Q_S, LM, q_T) \cdot LM) = 0 \quad . \quad (3.0.4)$$

For ease of notation, we can write these conditions in terms of acceptance

$$E_{\beta_0} (\varphi(Q_S, LM, q_T)) = 1 - \alpha \quad (3.0.6)$$

$$E_{\beta_0} (\varphi(Q_S, LM, q_T) \cdot LM) = 0 \quad (3.0.7)$$

where $\varphi(Q_S, LM, q_T) = 1 - \phi(Q_S, LM, q_T)$.

ASSUMPTION 3.0.1: *Given q_T and a statistic based on a k -class estimator $\psi_\kappa(Q_S, LM, q_T)$, there exist unique values $C_1(q_T), C_2(q_T) \in \mathbb{R}$ such that the test*

$$\varphi(Q_S, LM, q_T) = \begin{cases} 1 & \text{when } C_1(q_T) < \psi_\kappa(Q_S, LM, q_T) < C_2(q_T) \\ 0 & \text{when } C_1(q_T) > \psi_\kappa(Q_S, LM, q_T) \text{ or } C_2(q_T) < \psi_\kappa(Q_S, LM, q_T) \end{cases}, \quad (3.0.8)$$

satisfies 3.0.6 and 3.0.7.

This Assumption 3.0.1 establishes the test for which 3.0.6 and 3.0.7 are sufficient conditions for unbiasedness. Because the joint distribution of the sufficient statistics (Q_S, LM, Q_T) are in the exponential family, by Lehmann and Romano [2005] Section 4.2, $E_\beta(\varphi(Q_S, LM, q_T))$ has a maximum at β_0 and is strictly decreasing as β tends away from β_0 in either direction. Then the test $\phi(Q_S, LM, q_T) = 1 - \varphi(Q_S, LM, q_T)$ is necessarily unbiased since it reaches a minimum at β_0 and is strictly increasing as β tends away from β_0 in

either direction. In sum, by Assumption 3.0.1, an unbiased $(1 - \alpha)$ % confidence interval for $\psi_\kappa(Q_S, LM, q_T)$ under the null is defined by the critical values $C_1(q_T)$ and $C_2(q_T)$ such that

$$E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < \psi_\kappa(Q_S, LM, q_T) < C_2(q_T) \} \right) = 1 - \alpha \quad (3.0.9)$$

$$E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < \psi_\kappa(Q_S, LM, q_T) < C_2(q_T) \} LM \right) = 0 \quad (3.0.10)$$

The goal is then to find a suitable statistic ψ_κ and the critical values $C_1(q_T)$ and $C_2(q_T)$ in order to implement the test in practice. A possible candidate statistic ψ_κ is the conditional Wald statistic, such as the square of the t_{2SL_S} constructed in Section 2.2. Because we are testing a two-sided hypothesis, it is natural to expect that we will reject for large values of the Wald statistic. In such a case, finding $C_1(q_T)$ and $C_2(q_T)$ that satisfy 3.0.9 is straightforward, namely $C_1(q_T) = 0$ and $C_2(q_T)$ the $1 - \alpha$ quantile of the distribution of the conditional Wald statistic given q_T . However these values of $C_1(q_T)$ and $C_2(q_T)$ do not satisfy 3.0.10 in general. The problem is more severe when we consider the asymmetry of the distribution of the conditional t -statistic when instruments are weak.

3.1 An Algorithm for Finding Critical Values

Using the framework of the previous section, we develop an algorithm to approximate the critical values $C_1(q_T)$ and $C_2(q_T)$ that give an unbiased $(1 - \alpha)$ % confidence interval for the t -statistic based on the k -class estimators, t_κ . We define a statistic testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$: a t_κ -statistic under the null conditioned on q_T , written in terms of Q_S and Q_{ST} ¹. Since $LM = Q_{ST}/\sqrt{Q_T}$ we can write the conditional t_κ -statistic as a function of Q_S and LM . Under the null, the distributions of LM and Q_S , respectively, are known,

$$LM \sim N(0, 1)$$

$$Q_S = LM^2 + Q_{k-1}, \quad \text{where } Q_{k-1} \sim \chi_{k-1}^2,$$

given q_T [Moreira, 2003]. Thus an unbiased $(1 - \alpha)$ % confidence interval under the null based on $t_\kappa(Q_S, LM, q_T)$ would be defined by $C_1(q_T)$ and $C_2(q_T)$, such that

$$E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} \right) = 1 - \alpha \quad (3.1.2)$$

$$E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} LM \right) = 0 \quad (3.1.3)$$

¹Derivation of the conditional t -statistic can be found in Appendix 6.1.

Then by Assumption 3.0.1, $C_1(q_T)$ and $C_2(q_T)$ are defined by the unique solution to the minimization problem

$$\begin{aligned} & \min_{(C_1(q_T), C_2(q_T))} \left| E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} LM \right) \right| \\ & \text{such that } E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} \right) = 1 - \alpha \end{aligned} \quad (3.1.4)$$

Because $C_1(q_T)$ and $C_2(q_T)$ can not be calculated directly, we rely on finding consistent estimators. The following results provide a theoretical basis for estimating $C_1(q_T)$ and $C_2(q_T)$.

LEMMA 3.1.1: *The left hand sides of 3.1.2 and 3.1.3 exist.*

LEMMA 3.1.2:

a) *The function $g(C_1(q_T), C_2(q_T)) = E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} \right)$ is a continuous function of $(C_1(q_T), C_2(q_T))$.*

b) *The function $g^*(C_1(q_T), C_2(q_T)) = E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} LM \right)$ is a continuous function of $(C_1(q_T), C_2(q_T))$.*

THEOREM 3.1.3:

a) *Given an i.i.d. sequence of random variables $\{Q_S^j, LM^j\}$,*

$$\begin{aligned} & \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} LM^j \\ & \xrightarrow{p} E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} LM \right) \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} \\ & \xrightarrow{p} E_{\beta_0} \left(\mathbb{I} \{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \} \right) . \end{aligned}$$

b)

$$\begin{aligned} & \text{plim}_{J \rightarrow \infty} \operatorname{argmin}_{(C_1(q_T), C_2(q_T))} \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} LM^j \right| \\ & \text{such that } \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} = 1 - \alpha \end{aligned}$$

$$\begin{aligned}
&= \underset{(C_1(q_T), C_2(q_T))}{\operatorname{argmin}} \operatorname{plim}_{J \rightarrow \infty} \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} LM^j \right| \\
&\quad \text{such that} \quad \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} = 1 - \alpha .
\end{aligned}$$

DEFINITION 3.1.4: Given q_T and a random i.i.d. sample of J observations of LM and Q_S , let $C_1^J(q_T)$ and $C_2^J(q_T)$ be defined by

$$\begin{aligned}
\left(C_1^J(q_T), C_2^J(q_T) \right) &= \underset{(C_1(q_T), C_2(q_T))}{\operatorname{argmin}} \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} LM^j \right| \\
\text{such that} \quad &\frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} = 1 - \alpha .
\end{aligned} \tag{3.1.5}$$

COROLLARY 3.1.5:

$$\left(C_1^J(q_T), C_2^J(q_T) \right) \xrightarrow{p} \left(C_1(q_T), C_2(q_T) \right) \quad \text{as } J \rightarrow \infty$$

where $(C_1(q_T), C_2(q_T))$ is the unique solution the minimization problem 3.1.4.

Corollary 3.1.5 tells us that $C_1^J(q_T)$ and $C_2^J(q_T)$ are consistent approximations of the respective critical values $C_1(q_T)$ and $C_2(q_T)$.

By generating a sample of J values of Q_S and LM , we define the vectors $\mathbf{Q}_S = (Q_S^1, \dots, Q_S^J)$ and $\mathbf{LM} = (LM^1, \dots, LM^J)$ of respective values. Let

$$\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}; q_T) = (t_\kappa(Q_S^1, LM^1; q_T), \dots, t_\kappa(Q_S^J, LM^J; q_T)) \tag{3.1.6}$$

be a vector of J t_κ -statistics and let $\mathcal{Q}_z(\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T))$ be the z th quantile of $\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T)$. Then to control for the constraint in 3.1.5 we let

$$C_1^J(q_T) = \mathcal{Q}_x(\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T)) . \tag{3.1.7}$$

Then for the constraint in 3.1.5 to hold, it must be the case that

$$C_2^J(q_T) = \mathcal{Q}_{(1-\alpha)+x}(\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T)) \tag{3.1.8}$$

since by the definition of \mathcal{Q}_z ,

$$\frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ \mathcal{Q}_x(\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T)) < t_\kappa(Q_S^j, LM^j, q_T) < \mathcal{Q}_{(1-\alpha)+x}(\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T)) \right\} = 1 - \alpha \quad (3.1.9)$$

for any $x \in [0, \alpha]$. Hence, approximating the desired $C_1(q_T)$ and $C_2(q_T)$ is equivalent to finding the x that solves the constrained minimization problem

$$\min_{x \in [0, \alpha]} \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ \mathcal{Q}_x(\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T)) < t_\kappa(Q_S^j, LM^j, q_T) < \mathcal{Q}_{(1-\alpha)+x}(\mathbf{T}_\kappa(\mathbf{Q}_S, \mathbf{LM}, q_T)) \right\} LM^j \right|,$$

a function of one bounded variable on a compact set².

3.2 Constructing the Conditional t-Test

Under strong instruments, conducting a t -test with a k -class estimator testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ proceeds in the following manner: given data (Y, Z) , and assuming Ω is known, we construct a t -statistic

$$t_\kappa(Y, Z) = \frac{1}{\hat{\sigma}_u(Y, Z)} \left(\hat{\beta}_\kappa(Y, Z) - \beta_0 \right) \sqrt{y_2' P_Z y_2 - \kappa(Y, Z) \omega_{22}} \quad (3.2.1)$$

where $\hat{\sigma}_u(Y, Z)$ is a consistent estimate of σ_u . Given a size α and critical value $C_{\alpha/2}$, the test is defined by

$$\varphi_t(Y, Z) = 1 - \mathbb{I}(-C_{\alpha/2} < t_\kappa(Y, Z) < C_{\alpha/2}) \quad (3.2.2)$$

where 1 indicates rejection of the null and 0 indicates failure to reject.

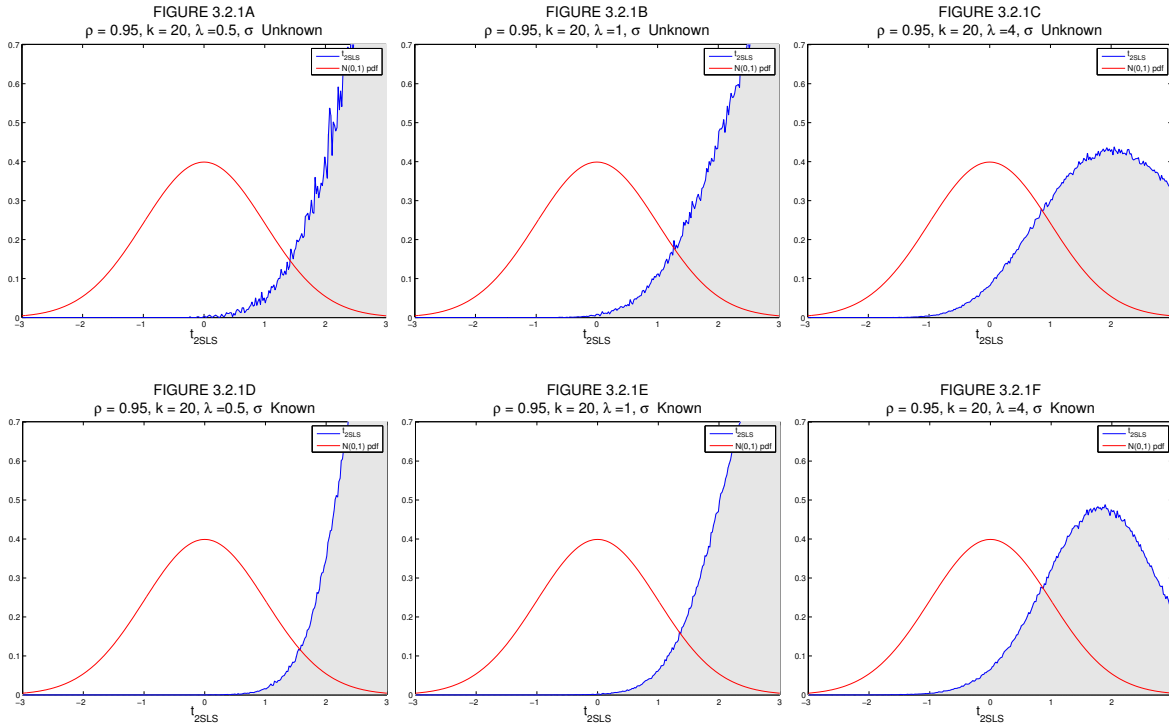
Under weak instruments, σ_u is not consistently estimable in general. As a result, the distribution of the t -statistic using a standard estimator for σ_u can differ significantly from the distribution when σ_u is known. Figures 3.2.1, 3.2.2, and 3.2.3 illustrates how estimating σ_u affects the distribution of the t -statistic based on the *2SLS*, *FULL*, and *LIML* estimators, respectively. Each figure represents a sample distribution generated from 1,000,000 simulated values of t_κ with $\rho = 0.95$ and $k = 20$, and λ/k ranging over 0.5, 1, and 4. In each figure, panels *A*, *B*, and *C* give the distribution when σ_u is unknown and estimated by

$$\hat{\sigma}_u = \sqrt{\begin{bmatrix} 1, -\hat{\beta}_\kappa \end{bmatrix} \Omega \begin{bmatrix} 1, -\hat{\beta}_\kappa \end{bmatrix}'}$$

Panels *D*, *E*, and *F* give the distribution when σ_u is known, which in this case is $\sigma_u = 1$.

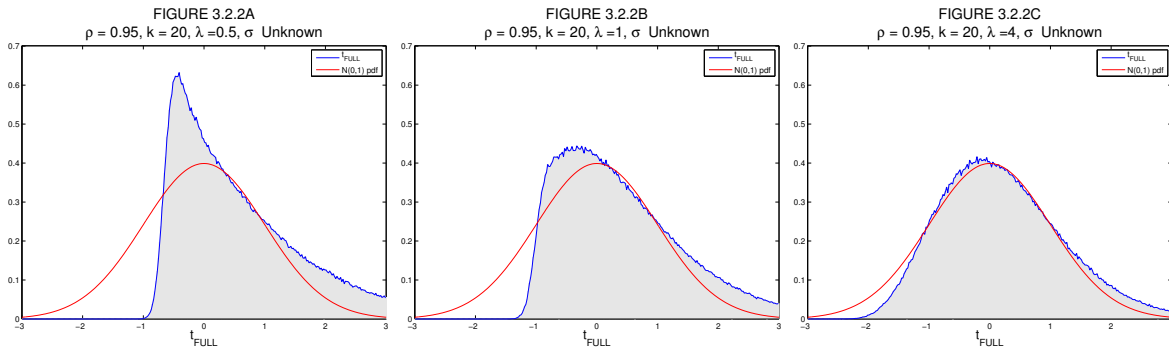
²Optimizing this function is straightforward with any suitable numerical package. Using the `fminbnd` function in Matlab with $\alpha = 0.05$ and a sample of 100,000 observations, finding a minimum takes less than 1 second.

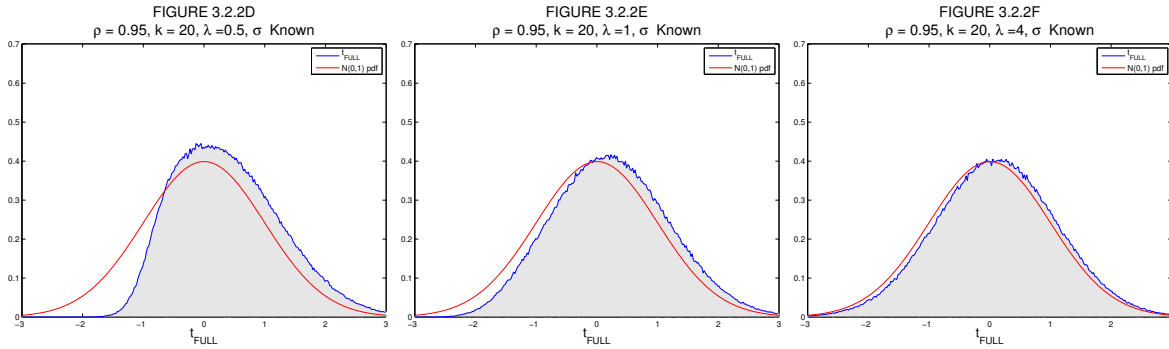
FIGURE 3.2.1: PDFS OF THE t_{2SLS} WITH σ_u KNOWN AND σ_u UNKNOWN



Comparing panels A to D, B to E, and C to F in figure 3.2.1, it appears that there is some difference between the pdfs. However, the general shapes are not dramatically different.

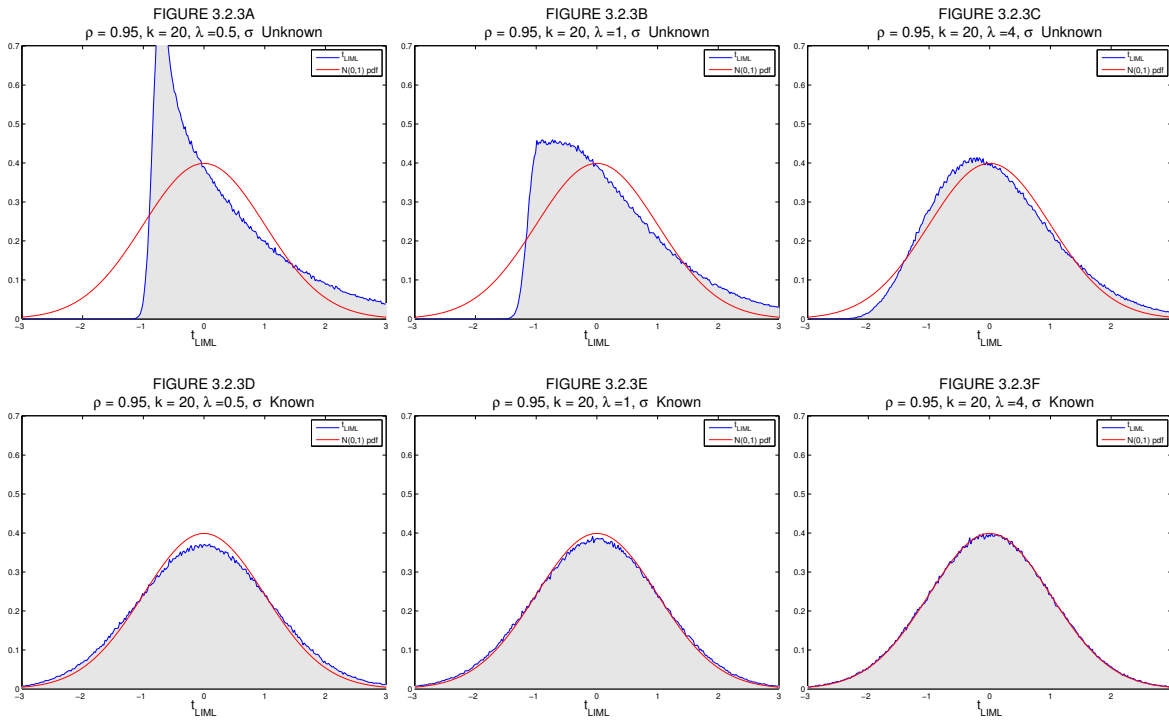
FIGURE 3.2.2: PDFS OF THE t_{FULL} WITH σ_u KNOWN AND σ_u UNKNOWN





Comparing panels *A* to *D*, *B* to *E*, and *C* to *F* in figure 3.2.2, there is a greater difference in the shape of the pdfs than in the case of t_{2SLS} . The difference is most dramatic when instruments are weakest, in panels *A* and *D*.

FIGURE 3.2.3: PDFS OF THE t_{LIML} WITH σ_u KNOWN AND σ_u UNKNOWN



Comparing panels *A* to *D*, *B* to *E*, and *C* to *F* in figure 3.2.3, the shapes of the pdfs are markedly different, especially under weaker instruments. Of particular note is that the pdf of t_{LIML} is the only distribution that remains symmetric when σ_u is known.

Ideally we would like to know the true value of σ_u and thus avoid the distortion demonstrated above, but in practice is this not possible. However, we can avoid estimating σ_u by making use of the form that critical

values take in the conditional framework. To do this we first assume that the true value of σ_u is known. Then given data (Y, Z) , and assuming Ω is known, the t_κ -statistic is given by

$$t_\kappa(Y, Z) = \frac{1}{\sigma_u} \left(\hat{\beta}_\kappa(Y, Z) - \beta_0 \right) \sqrt{y_2' P_Z y_2 - \kappa(Y, Z) \omega_{22}} . \quad (3.2.3)$$

Letting

$$\tilde{t}_\kappa(Y, Z) = \left(\hat{\beta}_\kappa(Y, Z) - \beta_0 \right) \sqrt{y_2' P_Z y_2 - \kappa(Y, Z) \omega_{22}} \quad (3.2.4)$$

and noting that $\tilde{t}_\kappa(Y, Z)$ can be written as a function of Q_S , LM , and Q_T , we get that

$$t_\kappa(Q_S, LM, Q_T) = \frac{1}{\sigma_u} \tilde{t}_\kappa(Q_S, LM, Q_T) . \quad (3.2.5)$$

Conditioning on q_T , we apply the algorithm from section 3.1 to find the critical values

$$\begin{aligned} C_1(q_T) &= t'_\kappa(Q'_S, LM', q_T) \\ C_2(q_T) &= t''_\kappa(Q''_S, LM'', q_T) \end{aligned} \quad (3.2.6)$$

where (Q'_S, LM') and (Q''_S, LM'') are the simulated values of Q_S and LM that correspond to $C_1(q_T)$ and $C_2(q_T)$, respectively. Thus the test is given by

$$\varphi_t(Y, Z) = 1 - \mathbb{I} \left(t'_\kappa(Q'_S, LM', q_T) < t_\kappa(Y, Z) < t''_\kappa(Q''_S, LM'', q_T) \right) . \quad (3.2.7)$$

By 3.2.3, 3.2.4, and 3.2.6,

$$\varphi_t(Y, Z) = 1 - \mathbb{I} \left(\frac{1}{\sigma_u} \tilde{t}'_\kappa(Q'_S, LM', q_T) < \frac{1}{\sigma_u} \tilde{t}_\kappa(Y, Z) < \frac{1}{\sigma_u} \tilde{t}''_\kappa(Q''_S, LM'', q_T) \right) . \quad (3.2.8)$$

Since σ_u is a positive scalar,

$$\varphi_t(Y, Z) = 1 - \mathbb{I} \left(\tilde{t}'_\kappa(Q'_S, LM', q_T) < \tilde{t}_\kappa(Y, Z) < \tilde{t}''_\kappa(Q''_S, LM'', q_T) \right) . \quad (3.2.9)$$

Hence σ_u is eliminated from the test.

3.3 Confidence Intervals

This section employs the algorithm established in Section 3.1 to produce 95% confidence intervals for the t_{2SLS} conditioned on different values of q_T and fixing other parameters $\sigma = 1$, $k = \{1, 2, 5, 10, 20\}$ and $\rho = \{0.2, 0.5, 0.95\}$. The critical values were calculated by generating a sample of $J = 1,000,000$. $\ln(q_T/k)$ can be seen as a measurement of the weakness of the instruments where -6 is very weak and 6 is very strong.

TABLE 3.3A: 95% CONFIDENCE INTERVALS FOR t_{2SLS} WITH $\rho = 0.2$

$\ln(q_T/k)$	$k = 1$	$k = 2$	$k = 5$	$k = 10$	$k = 20$
-6	(-0.24, 1.96)	(-0.30, 2.43)	(0.17, 3.29)	(1.58, 4.38)	(2.67, 5.58)
-5	(-0.40, 1.97)	(-0.31, 2.46)	(0.27, 3.27)	(1.36, 4.27)	(2.57, 5.51)
-4	(-0.66, 1.96)	(-0.57, 2.43)	(0.17, 3.24)	(1.01, 4.08)	(2.14, 5.22)
-3	(-1.09, 1.96)	(-0.94, 2.36)	(-0.24, 3.04)	(0.44, 3.74)	(1.34, 4.62)
-2	(-1.80, 1.96)	(-1.31, 2.28)	(-0.80, 2.76)	(-0.31, 3.23)	(0.35, 3.89)
-1	(-1.96, 1.96)	(-1.63, 2.16)	(-1.27, 2.48)	(-0.93, 2.81)	(-0.50, 3.24)
0	(-1.96, 1.96)	(-1.79, 2.09)	(-1.56, 2.30)	(-1.35, 2.50)	(-1.07, 2.78)
1	(-1.96, 1.96)	(-1.86, 2.04)	(-1.72, 2.18)	(-1.60, 2.30)	(-1.41, 2.48)
2	(-1.96, 1.96)	(-1.90, 2.01)	(-1.82, 2.09)	(-1.74, 2.17)	(-1.63, 2.28)
3	(-1.96, 1.96)	(-1.92, 1.99)	(-1.88, 2.04)	(-1.83, 2.09)	(-1.76, 2.16)
4	(-1.96, 1.96)	(-1.94, 1.98)	(-1.91, 2.01)	(-1.88, 2.04)	(-1.84, 2.07)
5	(-1.96, 1.96)	(-1.95, 1.97)	(-1.93, 1.98)	(-1.92, 2)	(-1.89, 2.03)
6	(-1.96, 1.96)	(-1.95, 1.96)	(-1.95, 1.97)	(-1.93, 1.99)	(-1.93, 2.00)

TABLE 3.3.2B: 95% CONFIDENCE INTERVALS FOR t_{2SL_S} WITH $\rho = 0.5$

$\ln(q_T/k)$	$k = 1$	$k = 2$	$k = 5$	$k = 10$	$k = 20$
-6	(-0.09, 1.96)	(0.16, 2.61)	(0.95, 3.69)	(1.73, 4.46)	(3.22, 6.20)
-5	(-0.14, 1.96)	(0.23, 2.91)	(0.50, 3.34)	(1.22, 4.27)	(2.92, 5.71)
-4	(-0.23, 1.97)	(-0.22, 2.44)	(0.81, 3.58)	(1.69, 4.50)	(2.44, 5.54)
-3	(-0.38, 1.98)	(-0.27, 2.47)	(0.42, 3.32)	(1.44, 4.33)	(2.62, 5.54)
-2	(-0.63, 1.97)	(-0.57, 2.43)	(0.20, 3.25)	(1.04, 4.09)	(2.17, 5.24)
-1	(-1.05, 1.96)	(-0.91, 2.37)	(-0.21, 3.06)	(0.48, 3.75)	(1.40, 4.66)
0	(-1.73, 1.96)	(-1.28, 2.3)	(-0.76, 2.77)	(-0.26, 3.27)	(0.43, 3.96)
1	(-1.96, 1.96)	(-1.61, 2.18)	(-1.25, 2.5)	(-0.89, 2.84)	(-0.44, 3.29)
2	(-1.96, 1.96)	(-1.78, 2.09)	(-1.55, 2.31)	(-1.32, 2.52)	(-1.03, 2.81)
3	(-1.96, 1.96)	(-1.85, 2.05)	(-1.71, 2.18)	(-1.58, 2.31)	(-1.4, 2.49)
4	(-1.96, 1.96)	(-1.90, 2.02)	(-1.81, 2.10)	(-1.73, 2.17)	(-1.62, 2.28)
5	(-1.96, 1.97)	(-1.92, 1.99)	(-1.88, 2.04)	(-1.83, 2.09)	(-1.76, 2.16)
6	(-1.97, 1.95)	(-1.94, 1.98)	(-1.91, 2.01)	(-1.87, 2.04)	(-1.84, 2.08)

TABLE 3.3.2C: 95% CONFIDENCE INTERVALS FOR t_{2SL_S} WITH $\rho = 0.95$

$\ln(q_T/k)$	$k = 1$	$k = 2$	$k = 5$	$k = 10$	$k = 20$
-6	(0.03, 2.19)	(0.20, 2.65)	(0.87, 3.52)	(1.84, 4.58)	(3.21, 6.06)
-5	(0.02, 2.17)	(0.21, 2.67)	(0.84, 3.49)	(1.80, 4.53)	(2.90, 5.69)
-4	(0.05, 2.68)	(0.26, 2.89)	(0.95, 3.68)	(1.91, 4.77)	(3.19, 6.03)
-3	(-0.07, 1.96)	(0.16, 2.59)	(0.98, 3.78)	(1.66, 4.40)	(2.79, 5.65)
-2	(-0.11, 1.99)	(0.22, 2.82)	(0.86, 3.54)	(1.54, 4.34)	(3.03, 5.80)
-1	(-0.20, 1.96)	(-0.16, 2.45)	(0.68, 3.41)	(1.52, 4.33)	(2.38, 5.55)
0	(-0.33, 1.96)	(-0.34, 2.43)	(0.49, 3.34)	(1.50, 4.35)	(2.66, 5.56)
1	(-0.54, 1.96)	(-0.58, 2.41)	(0.35, 3.32)	(1.17, 4.17)	(2.29, 5.32)
2	(-0.89, 1.96)	(-0.78, 2.40)	(-0.04, 3.15)	(0.70, 3.90)	(1.69, 4.88)
3	(-1.47, 1.96)	(-1.17, 2.32)	(-0.56, 2.88)	(-0.02, 3.42)	(0.73, 4.16)
4	(-1.96, 1.96)	(-1.53, 2.21)	(-1.11, 2.58)	(-0.71, 2.96)	(-0.19, 3.49)
5	(-1.96, 1.96)	(-1.74, 2.12)	(-1.47, 2.36)	(-1.21, 2.61)	(-0.87, 2.94)
6	(-1.96, 1.96)	(-1.85, 2.05)	(-1.67, 2.21)	(-1.51, 2.37)	(-1.30, 2.58)

As expected, when instruments become stronger the 95% confidence intervals tend to $(-1.96, 1.96)$, that of the standard normal distribution. We also note that with more instruments, such as $k = 20$, the distortion of the confidence intervals persists for “longer” in the sense that instruments have to be stronger to converge to $(-1.96, 1.96)$ than when there are less instruments. Similarly when ρ is larger, such as $\rho = 0.95$, we find a similar effect.

4 Power of the Conditional t-Test

This section compares the power properties of the conditional t_κ -test with other known tests that are robust to weak instrumental variables, testing $H_0 : \beta = \beta_0$ against local alternatives. Following Andrews et al. [2006a] the local neighborhood we look at is $1/\sqrt{\lambda}$. Results are computed for 36 different parameter combinations of ρ , k , and λ/k . In particular we look at $\rho = 0.2, 0.5, \text{ and } 0.95$; $k = 2, 5, \text{ and } 20$; and $\lambda/k = 0.5, 1, 4, \text{ and } 16$, noting that the smaller is λ/k , the weaker the instruments, and the larger is λ/k , the stronger the instruments. Following Andrews et al. [2007], without loss, we set $\beta_0 = 0$ and

$$\Omega = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

The power function for each parameter combination has been computed at 25 evenly spaced points from $\beta/\sqrt{\lambda} = -6$ to $\beta/\sqrt{\lambda} = 6$. Results are based on 1,000 Monte Carlo draws. The size of the test in each case is $\alpha = 0.05$. The critical values for the conditional t -test and conditional Wald test were calculated based on 100,000 simulations of Q_S and LM . Evaluation of the CLR was implemented using p-values using the algorithm described in Andrews et al. [2006a].

4.1 Power Comparison of the 2SLS Conditional t-Test with the 2SLS Conditional Wald

This section compares the power of the conditional t -test against that of the conditional Wald. Figure 4.1.1 gives power results for $\rho = 0.5$ and $k = 5$. Complete results for all 36 designs are found in Section 6.2.

FIGURE 4.1.1 POWER CURVES OF CONDITIONAL t AND WALD TESTS, $\rho = 0.5, k = 5$

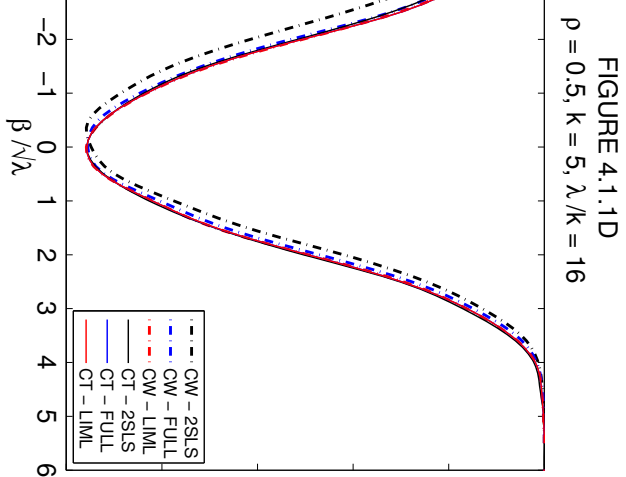
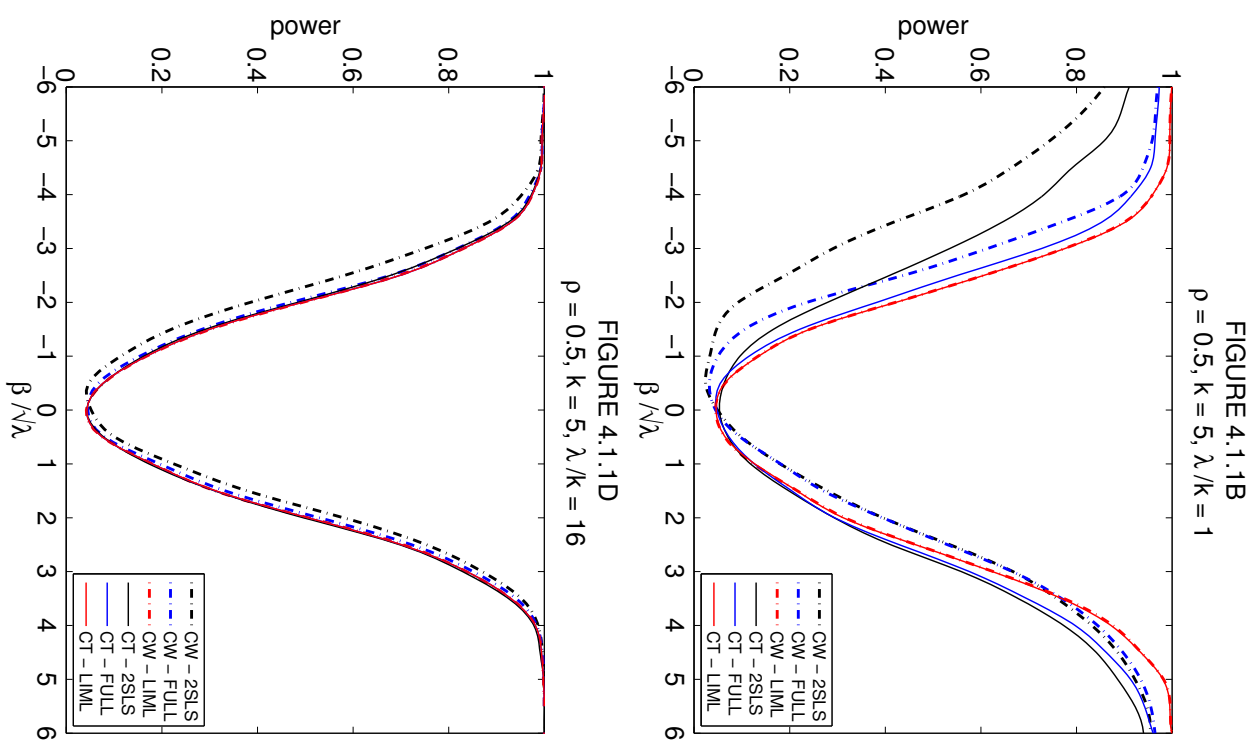
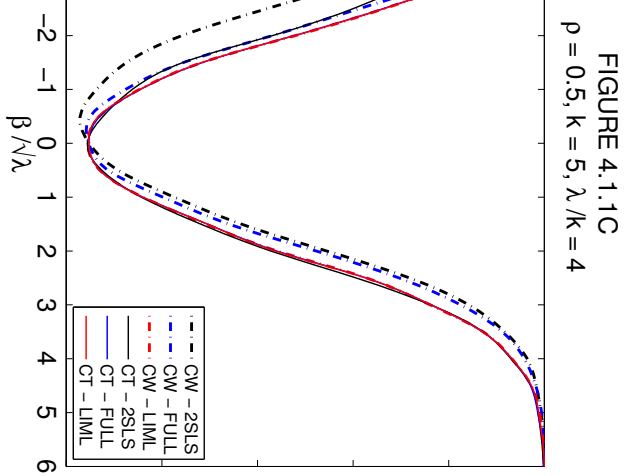
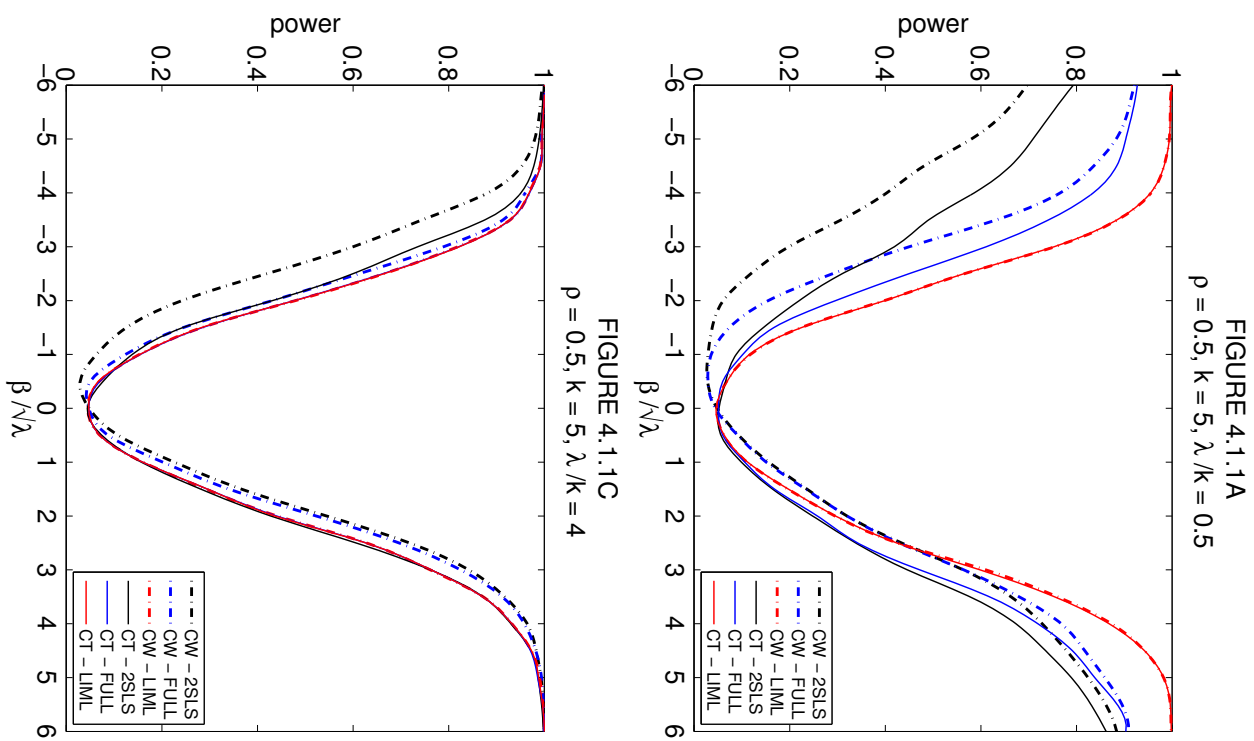


FIGURE 4.1.2 POWER CURVES OF CONDITIONAL t AND WALD TESTS, $\rho = 0.95$, $k = 20$

FIGURE 4.1.2A
 $\rho = 0.95$, $k = 20$, $\lambda/k = 0.5$

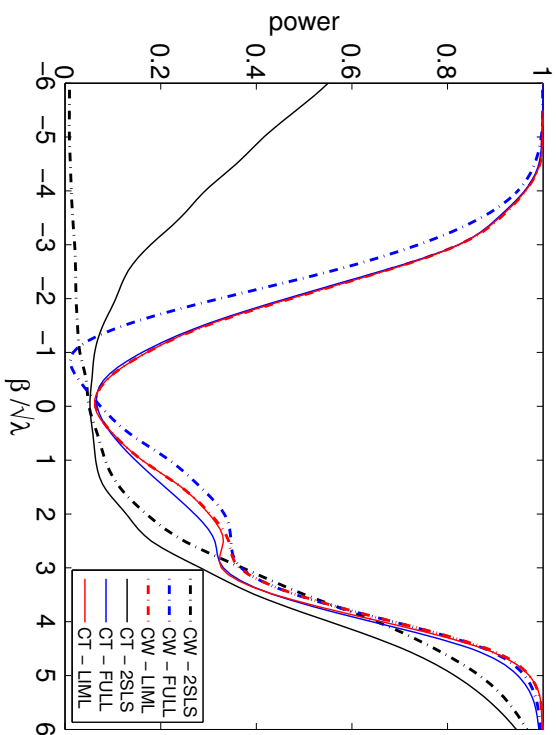


FIGURE 4.1.2C
 $\rho = 0.95$, $k = 20$, $\lambda/k = 4$

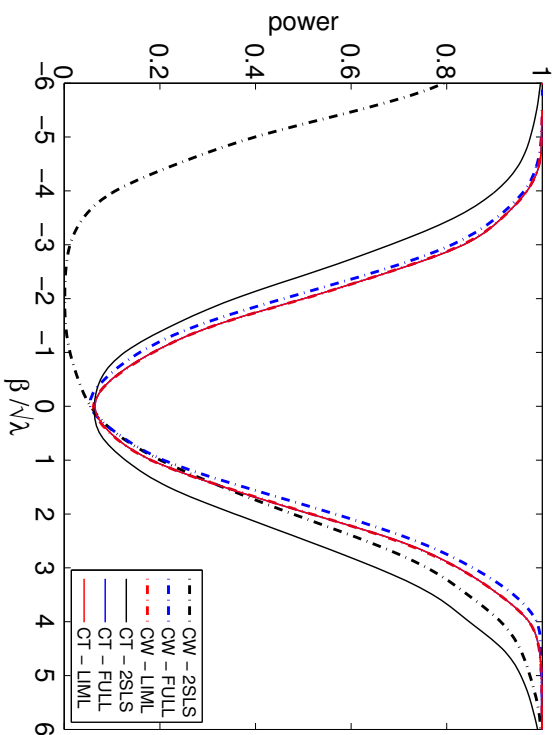


FIGURE 4.1.2B
 $\rho = 0.95$, $k = 20$, $\lambda/k = 1$

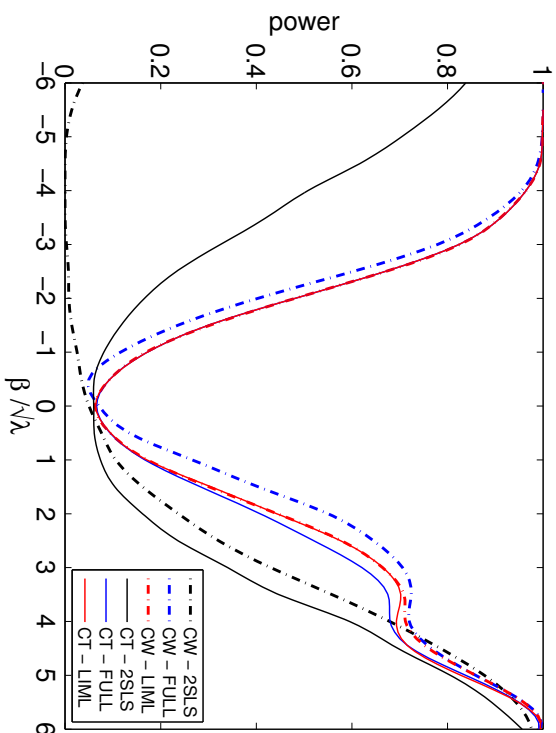


FIGURE 4.1.2D
 $\rho = 0.95$, $k = 20$, $\lambda/k = 16$

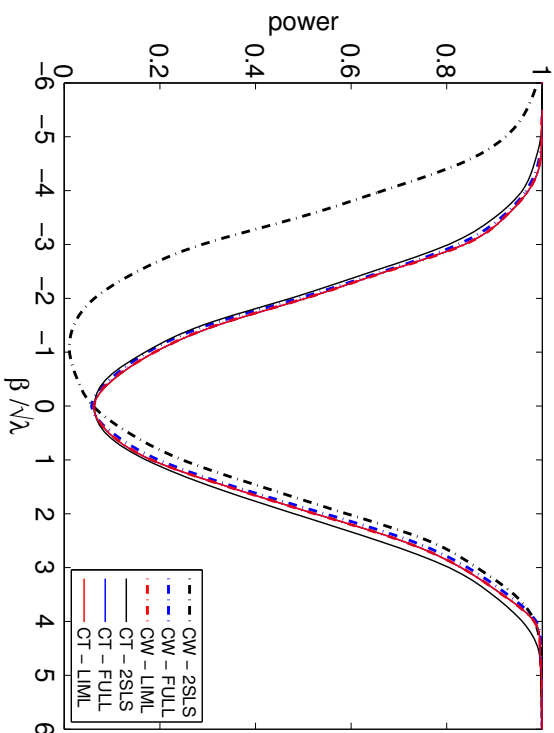


Figure 4.1.1 shows that with $\rho = 0.5$, a moderate degree of endogeneity, and $k = 5$. In all cases the conditional t -test constructed in this paper is unbiased for every k -class estimator considered. In the weak instrument cases, as expected, the conditional Wald based on the 2SLS and FULL estimators are biased for negative alternatives. However, of note is the fact that the conditional Wald based on the LIML estimator is unbiased, and in fact has the same power as the conditional t based on the LIML estimator. The discrepancy between this result and the that of Andrews et al. [2007] is likely due to the elimination of the parameter σ_u from the test. As shown in Section 3.2, eliminating σ_u from the t_{LIML} results in a symmetric distribution under weak instruments. This symmetry may explain why the critical values of the conditional Wald based on the LIML result in an unbiased test. On the other hand, eliminating σ_u from t_{2SLS} and t_{FULL} do not result in a symmetric distribution. Hence one would not expect the conditional Wald based on the 2SLS and FULL to be unbiased.

Figure 4.1.2 shows that with $\rho = 0.95$, a strong degree of endogeneity, and $k = 20$. Again the conditional Wald based on the 2SLS and FULL estimators are biased for negative alternatives, only more so than in 4.1.1. Even under strong instruments, the conditional Wald based on the 2SLS estimator is biased. By contrast, every variety of conditional t -test is unbiased. Although it is unbiased, the conditional t -test based on the 2SLS estimator appears to lose power compared with figure 4.1.1. By contrast, the conditional t based on the FULL estimator appears to show better power properties under weak instruments compared to 4.1.1. Panels 4.1.2A and 4.1.2B are noteworthy in that they show the strongest divergence between the conditional t and conditional Wald based on the LIML estimator. In both 4.1.2A and 4.1.2B, the conditional Wald has slightly better power around $\beta/\lambda = 3$ and $\beta/\lambda = 4$, respectively.

4.2 Power Comparison with the AR, LM, and CLR

This section compares the power properties of the conditional t test using the three k -class estimators with those of the conditional likelihood ratio test (CLR) of Moreira [2003], the AR test of Anderson and Rubin [1949], and the LM test of Breusch and Pagan [1980]; independently shown by Kleibergen [2002] and Moreira [2002] to have correct size. Complete results for all 36 designs are found in Section 6.3.

FIGURE 4.2.1 POWER CURVES OF CONDITIONAL t , AR, LM, AND CLR TESTS, $\rho = 0.5, k = 5$

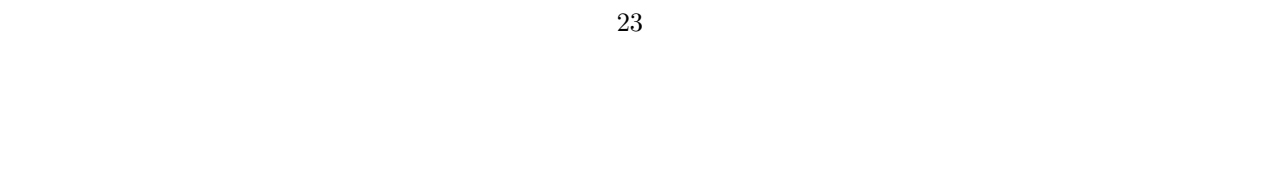
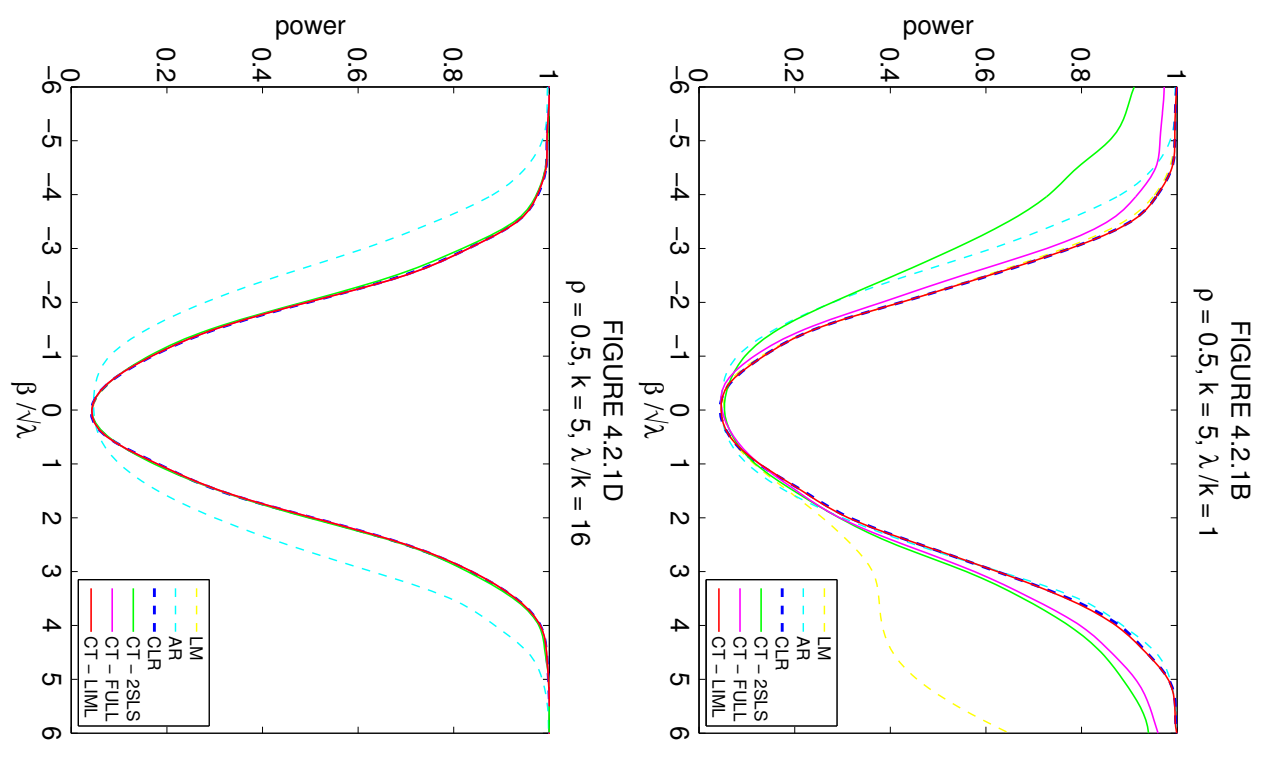
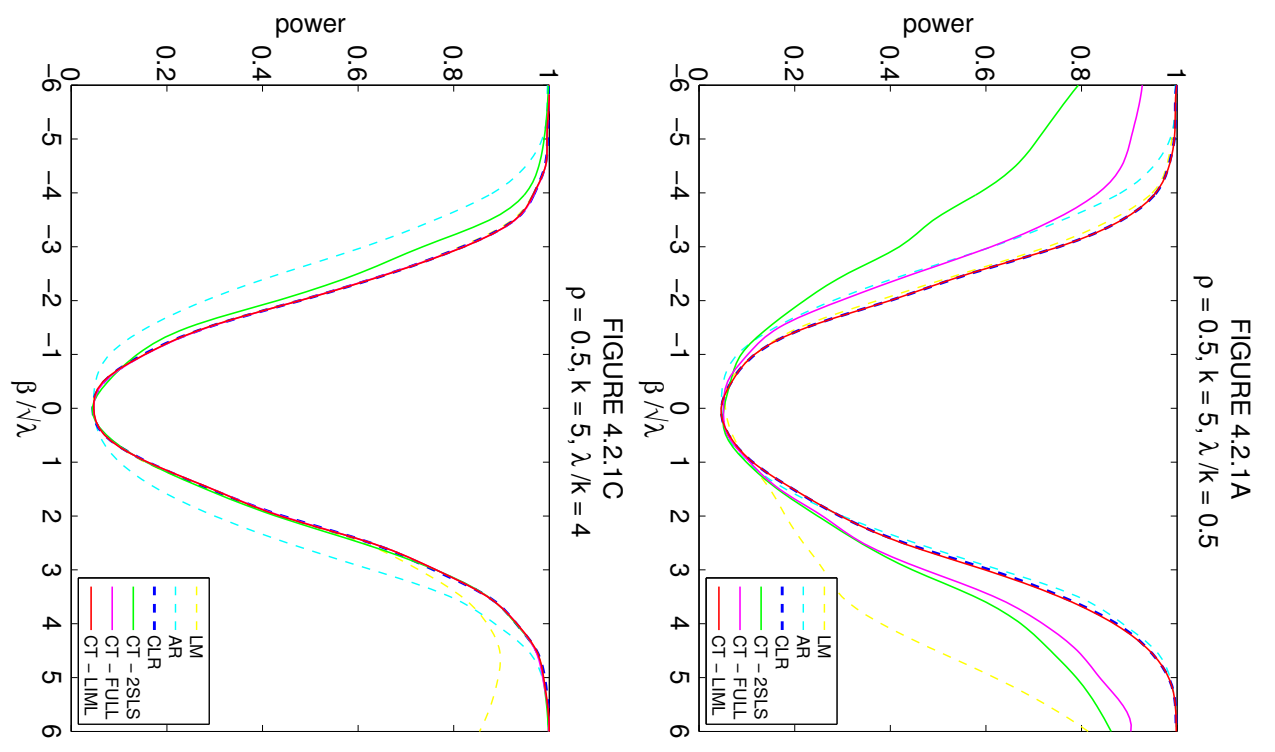


FIGURE 4.2.2 POWER CURVES OF CONDITIONAL t , AR, LM, AND CLR TESTS, $\rho = 0.95, k = 20$

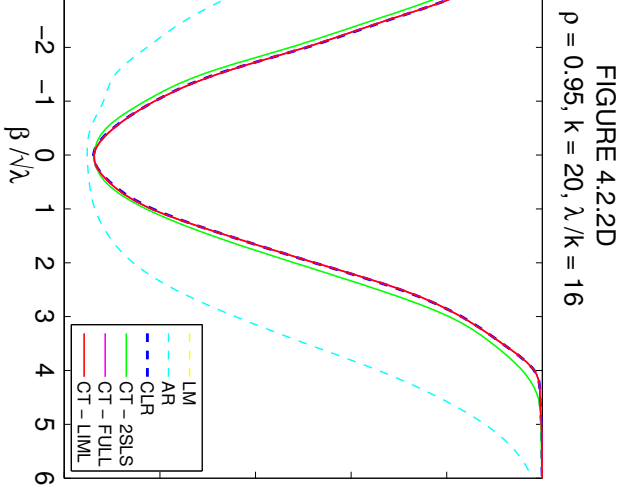
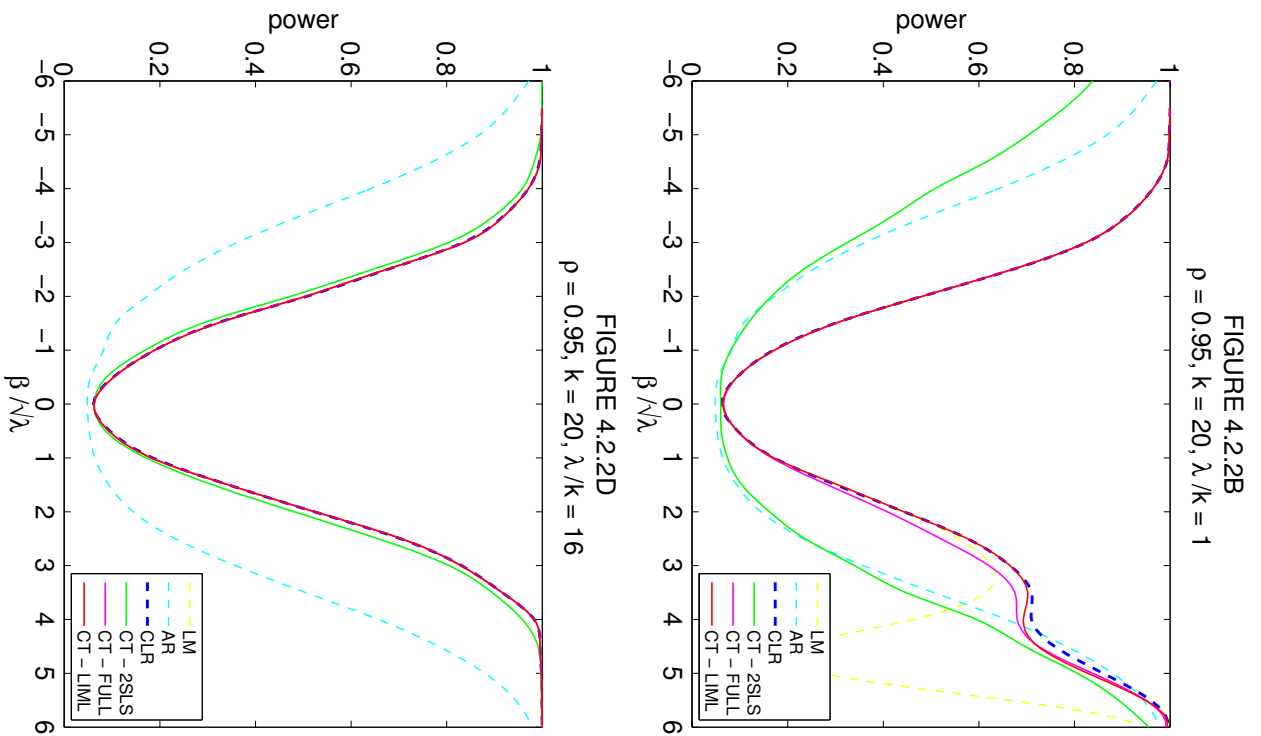
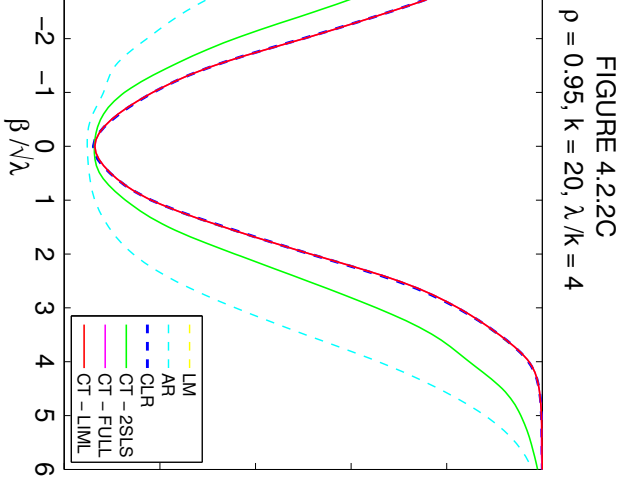
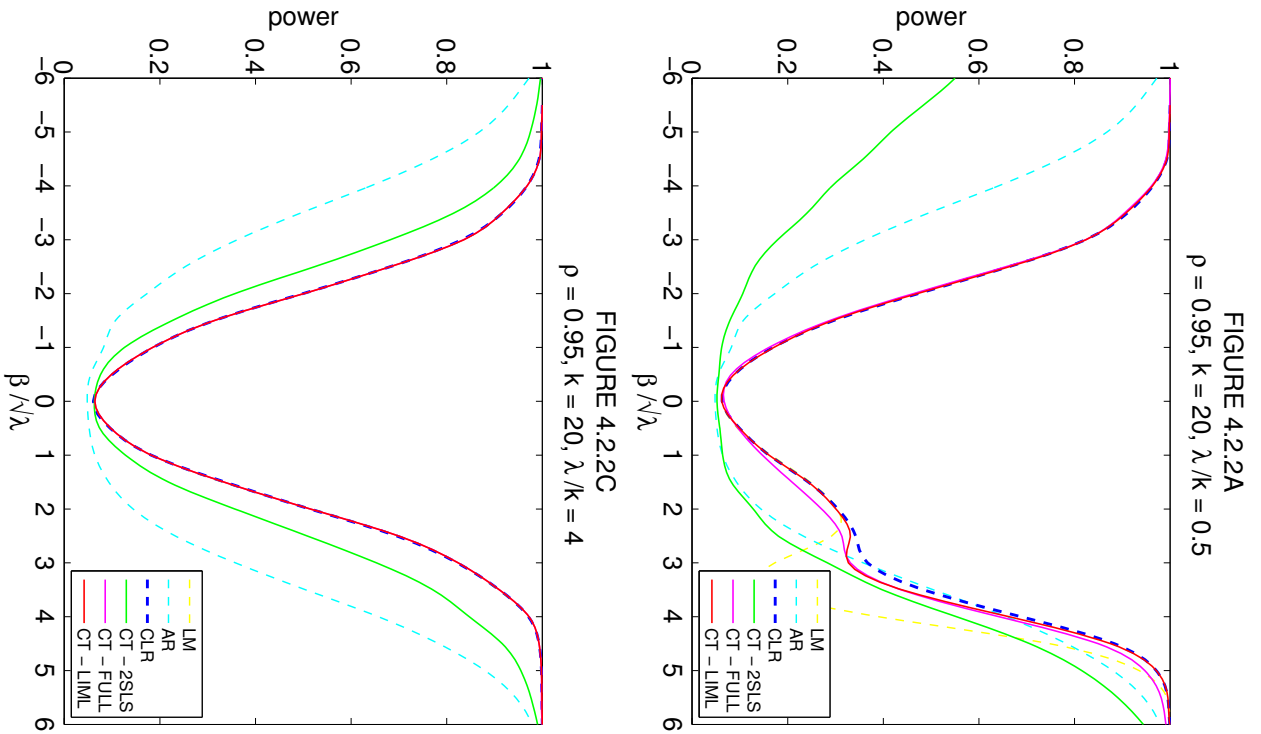


Figure 4.2.1 shows that the conditional t -test based on LIML performs well against the AR, LM, and CLR tests in every case. In particular, the conditional t -test based on LIML has power properties almost identical to that of the CLR test.

Figure 4.2.2 shows, analogous to 4.1.2, the strongest divergence between the conditional t test based the LIML and the CLR test. In particular, in both 4.2.2A and 4.2.2B, the CLR has slightly better power compared with the conditional t based the LIML around $\beta/\lambda = 3$ and $\beta/\lambda = 4$, respectively. This is precisely where the divergence between the conditional t test based the LIML and the conditional Wald test based the LIML occurred. It appears that the conditional Wald based the LIML, after eliminating σ_u from the test, has nearly identical power properties to the CLR in every case considered.

5 Conclusion

Previous authors, e.g. Andrews et al. [2007] and Moreira [2003], only required that the conditional Wald test have correct size. By not taking into account the asymmetry of the distribution of the conditional t -statistic, the critical values for the conditional Wald result in bias under weak instruments. By imposing unbiasedness and taking into account this asymmetry, unbiased conditional Wald tests based on k -class estimators can be constructed. In addition, power is improved by eliminating the unknown structural form error variance from the test. Numerical results appear to confirm the validity of the methodology as well as well as the algorithm used to construct the test.

6 Appendix

6.1 Derivation of Statistics

DERIVATION OF T-STATISTICS BASED ON K-CLASS ESTIMATORS AS FUNCTIONS OF Q_S , Q_{ST} , and Q_T :

This section derives k -class t -statistics, two-stage least squares, LIML, and the modified LIML of Fuller [1977] as functions of Q_S , Q_{ST} , and Q_T . For simplicity, we assume that Y has already been projected onto the space orthogonal to the column space of any exogenous regressors X .

$$[S, T] = (Z'Z)^{-1/2} Z'Y\Omega^{-1/2} \left[\frac{\Omega^{1/2}b_0}{\sqrt{b_0'\Omega b_0}}, \frac{\Omega^{-1/2}a_0}{\sqrt{a_0'\Omega^{-1}a_0}} \right]$$

$$\begin{aligned}
\Rightarrow [S, T] & \begin{bmatrix} \frac{b'_0 \Omega^{1/2}}{\sqrt{b'_0 \Omega b_0}} \\ \frac{a'_0 \Omega^{-1/2}}{\sqrt{a'_0 \Omega^{-1} a_0}} \end{bmatrix} \Omega^{1/2} = (Z'Z)^{-1/2} Z'Y \\
\Rightarrow [S, T] & \begin{bmatrix} \frac{b'_0 \Omega}{\sqrt{b'_0 \Omega b_0}} \\ \frac{a'_0}{\sqrt{a'_0 \Omega^{-1} a_0}} \end{bmatrix} = (Z'Z)^{-1/2} Z'Y \\
\Rightarrow [S, T] & \begin{bmatrix} \frac{b'_0 \Omega \mathbf{e}_1}{\sqrt{b'_0 \Omega b_0}} \\ \frac{a'_0 \mathbf{e}_1}{\sqrt{a'_0 \Omega^{-1} a_0}} \end{bmatrix} = (Z'Z)^{-1/2} Z'y_1 \\
\Rightarrow [S, T] & \begin{bmatrix} \frac{b'_0 \Omega \mathbf{e}_2}{\sqrt{b'_0 \Omega b_0}} \\ \frac{a'_0 \mathbf{e}_2}{\sqrt{a'_0 \Omega^{-1} a_0}} \end{bmatrix} = (Z'Z)^{-1/2} Z'y_2 \\
& \Rightarrow (Z'Z)^{-1/2} Z'y_2 = c_1 S + c_2 T
\end{aligned}$$

where

$$c_1 = \frac{b'_0 \Omega \mathbf{e}_2}{\sqrt{b'_0 \Omega b_0}}, c_2 = \frac{a'_0 \mathbf{e}_2}{\sqrt{a'_0 \Omega^{-1} a_0}}, d_1 = \frac{b'_0 \Omega \mathbf{e}_1}{\sqrt{b'_0 \Omega b_0}}, \text{ and } d_2 = \frac{a'_0 \mathbf{e}_1}{\sqrt{a'_0 \Omega^{-1} a_0}}$$

$$\begin{aligned}
\Rightarrow y'_2 P_Z y_2 &= y'_2 Z (Z'Z)^{-1} Z' y_2 \\
&= \left((Z'Z)^{-1/2} Z' y_2 \right)' \left((Z'Z)^{-1/2} Z' y_2 \right) \\
&= (c_1 S + c_2 T)' (c_1 S + c_2 T) \\
&= (c_1 S' + c_2 T') (c_1 S + c_2 T) \\
&= c_1^2 S' S + c_1 c_2 S' T + c_1 c_2 T' S + c_2^2 T' T \\
&= c_1^2 Q_S + 2c_1 c_2 Q_{ST} + c_2^2 Q_T
\end{aligned}$$

$$\begin{aligned}
\Rightarrow y_2' P_Z y_1 &= y_2' Z (Z' Z)^{-1} Z' y_1 \\
&= \left((Z' Z)^{-1/2} Z' y_2 \right)' \left((Z' Z)^{-1/2} Z' y_1 \right) \\
&= (c_1 S + c_2 T)' (d_1 S + d_2 T) \\
&= (c_1 S' + c_2 T') (d_1 S + d_2 T) \\
&= c_1 d_1 S' S + c_1 d_2 S' T + c_2 d_1 T' S + c_2 d_2 T' T \\
&= c_1 d_1 Q_S + (c_1 d_2 + c_2 d_1) Q_{ST} + c_2 d_2 Q_T
\end{aligned}$$

Using the expression for $\hat{\beta}_\kappa$ given in 2.0.5, we have

$$\begin{aligned}
t(\kappa) &= \frac{(\hat{\beta}(\kappa) - \beta_0)}{\hat{\sigma}_u} \cdot (y_2' P_Z y_2 - \kappa \omega_{22})^{1/2} \\
&= \frac{1}{\hat{\sigma}_u} \left(\frac{y_2' P_Z y_1 - \kappa \omega_{12}}{y_2' P_Z y_2 - \kappa \omega_{22}} - \beta_0 \right) \cdot (y_2' P_Z y_2 - \kappa \omega_{22})^{1/2} \\
&= \frac{1}{\hat{\sigma}_u} \left(\frac{y_2' P_Z y_1 - \kappa \omega_{12}}{(y_2' P_Z y_2 - \kappa \omega_{22})^{1/2}} - \beta_0 (y_2' P_Z y_2 - \kappa \omega_{22})^{1/2} \right) \\
&= \frac{1}{\hat{\sigma}_u} \left(\frac{c_1 d_1 Q_S + (c_1 d_2 + c_2 d_1) Q_{ST} + c_2 d_2 Q_T - \kappa \omega_{12}}{(c_1^2 Q_S + 2c_1 c_2 Q_{ST} + c_2^2 Q_T - \kappa \omega_{22})^{1/2}} - \beta_0 (c_1^2 Q_S + 2c_1 c_2 Q_{ST} + c_2^2 Q_T - \kappa \omega_{22})^{1/2} \right)
\end{aligned}$$

Next we find an expression for κ in terms of Q_S , Q_{ST} , and Q_T . Let

$$p_1 = \frac{b_0' \Omega}{\sqrt{b_0' \Omega b_0}} \text{ and } p_2 = \frac{a_0'}{\sqrt{a_0' \Omega^{-1} a_0}}$$

By Andrews et al. [2007] equation 12, $Y' P_Z Y$ can be written in terms of Q_S , Q_{ST} , and Q_T , in the following manner:

$$\begin{aligned}
Y' P_Z Y &= \Omega^{1/2} \begin{bmatrix} \Omega^{1/2} b_0 & \Omega^{-1/2} a_0 \\ \sqrt{b_0' \Omega b_0} & \sqrt{a_0' \Omega^{-1} a_0} \end{bmatrix}'^{-1} \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} \begin{bmatrix} \Omega^{1/2} b_0 & \Omega^{-1/2} a_0 \\ \sqrt{b_0' \Omega b_0} & \sqrt{a_0' \Omega^{-1} a_0} \end{bmatrix}^{-1} \Omega^{1/2} \\
&= \begin{bmatrix} D_{11}(Q_S, Q_{ST}, Q_T) & D_{12}(Q_S, Q_{ST}, Q_T) \\ D_{21}(Q_S, Q_{ST}, Q_T) & D_{22}(Q_S, Q_{ST}, Q_T) \end{bmatrix} \\
&= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow h(\kappa) &= \det(Y'P_ZY - \kappa\Omega) \\
&= \det\left(\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} - \kappa \begin{bmatrix} \omega_{11} & \omega_{21} \\ \omega_{12} & \omega_{22} \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} D_{11} - \kappa\omega_{11} & D_{12} - \kappa\omega_{21} \\ D_{21} - \kappa\omega_{12} & D_{22} - \kappa\omega_{22} \end{bmatrix}\right) \\
&= (D_{11} - \kappa\omega_{11})(D_{22} - \kappa\omega_{22}) - (D_{21} - \kappa\omega_{12})(D_{12} - \kappa\omega_{21}) \\
&= D_{11}D_{22} - \omega_{22}D_{11}\kappa - \omega_{11}D_{22}\kappa + \omega_{11}\omega_{22}\kappa^2 - D_{21}D_{12} + \omega_{21}D_{21}\kappa + \omega_{12}D_{12}\kappa - \omega_{21}\omega_{12}\kappa^2 \\
&= \underbrace{(\omega_{11}\omega_{22} - \omega_{21}\omega_{12})\kappa^2}_{P_2} + \underbrace{(\omega_{21}D_{21} + \omega_{12}D_{12} - \omega_{22}D_{11} - \omega_{11}D_{22})\kappa}_{P_1(Q_S, Q_{ST}, Q_T)} + \underbrace{(D_{11}D_{22} - D_{21}D_{12})}_{P_0(Q_S, Q_{ST}, Q_T)} \\
&= P_2\kappa^2 + P_1(Q_S, Q_{ST}, Q_T)\kappa + P_0(Q_S, Q_{ST}, Q_T)
\end{aligned}$$

Then by 2.0.6,

$$\kappa_{LIML} = \text{The smallest root of } h(\kappa) = P_2\kappa^2 + P_1(Q_S, Q_{ST}, Q_T)\kappa + P_0(Q_S, Q_{ST}, Q_T) = 0$$

and thus we have

$$\begin{aligned}
t_{SLS} &= \frac{1}{\hat{\sigma}_u} \left(\frac{c_1d_1Q_S + (c_1d_2 + c_2d_1)Q_{ST} + c_2d_2Q_T}{(c_1^2Q_S + 2c_1c_2Q_{ST} + c_2^2Q_T)^{1/2}} - \beta_0 (c_1^2Q_S + 2c_1c_2Q_{ST} + c_2^2Q_T)^{1/2} \right) \\
t_{LIML} &= \frac{1}{\hat{\sigma}_u} \left(\frac{c_1d_1Q_S + (c_1d_2 + c_2d_1)Q_{ST} + c_2d_2Q_T - \kappa_{LIML}\omega_{12}}{(c_1^2Q_S + 2c_1c_2Q_{ST} + c_2^2Q_T - \kappa_{LIML}\omega_{22})^{1/2}} - \beta_0 (c_1^2Q_S + 2c_1c_2Q_{ST} + c_2^2Q_T - \kappa_{LIML}\omega_{22})^{1/2} \right) \\
t_{FULL} &= \frac{1}{\hat{\sigma}_u} \left(\frac{c_1d_1Q_S + (c_1d_2 + c_2d_1)Q_{ST} + c_2d_2Q_T - \kappa_{FULL}\omega_{12}}{(c_1^2Q_S + 2c_1c_2Q_{ST} + c_2^2Q_T - \kappa_{FULL}\omega_{22})^{1/2}} - \beta_0 (c_1^2Q_S + 2c_1c_2Q_{ST} + c_2^2Q_T - \kappa_{FULL}\omega_{22})^{1/2} \right). \quad \square
\end{aligned}$$

6.2 Power Curves for the Conditional t-Test Compared with the Conditional Wald

FIGURE 6.2.1A
 $\rho = 0.2, k = 2, \lambda / k = 0.5$

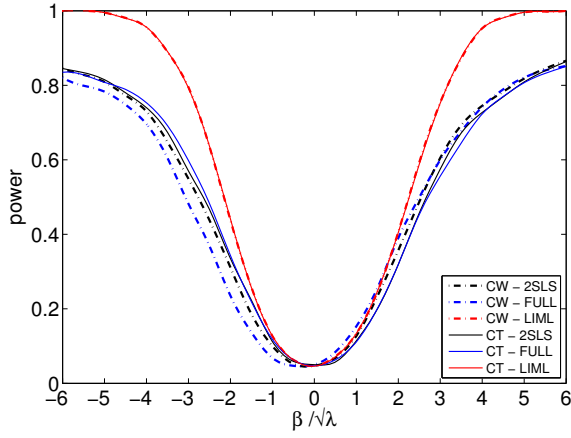


FIGURE 6.2.1B
 $\rho = 0.2, k = 2, \lambda / k = 1$

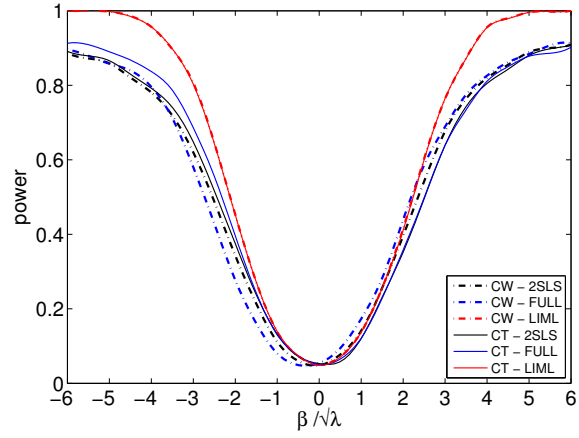


FIGURE 6.2.1C
 $\rho = 0.2, k = 2, \lambda / k = 4$

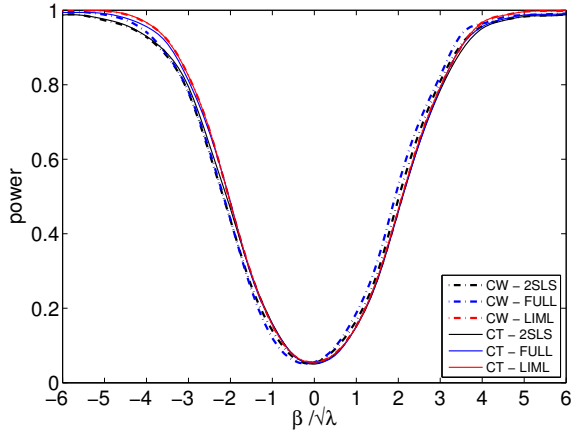


FIGURE 6.2.1D
 $\rho = 0.2, k = 2, \lambda / k = 16$

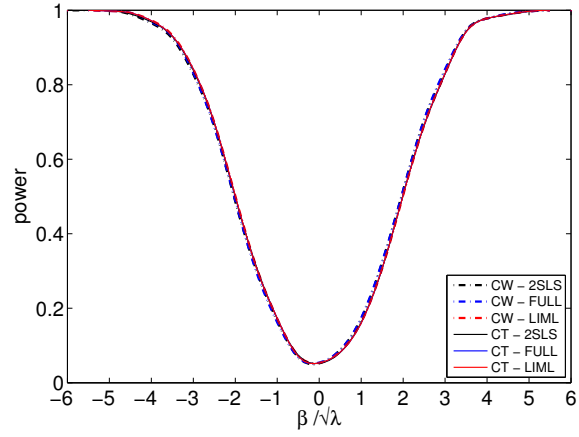


FIGURE 6.2.2A
 $\rho = 0.2, k = 5, \lambda / k = 0.5$

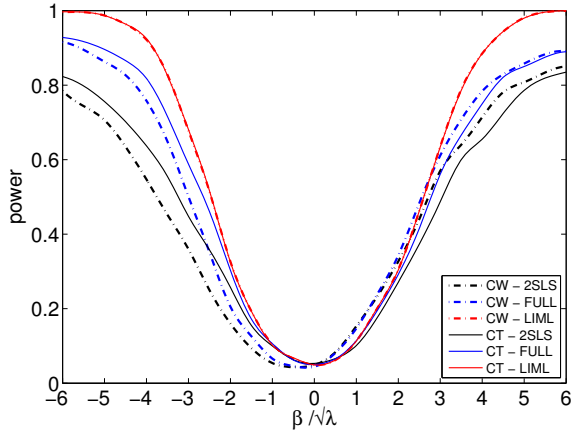


FIGURE 6.2.2B
 $\rho = 0.2, k = 5, \lambda / k = 1$

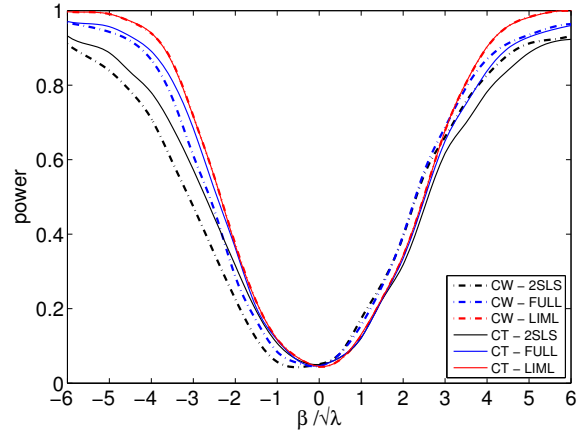


FIGURE 6.2.2C
 $\rho = 0.2, k = 5, \lambda / k = 4$

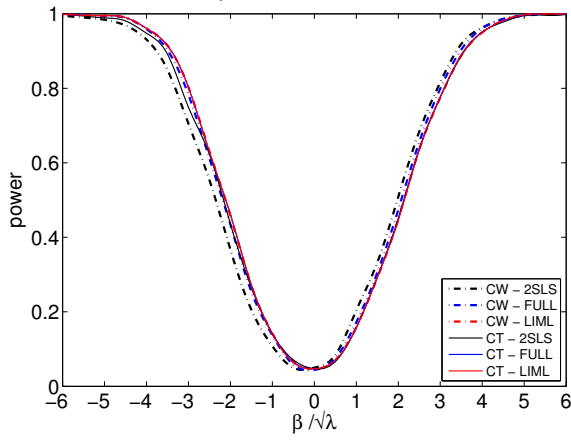


FIGURE 6.2.2D
 $\rho = 0.2, k = 5, \lambda / k = 16$

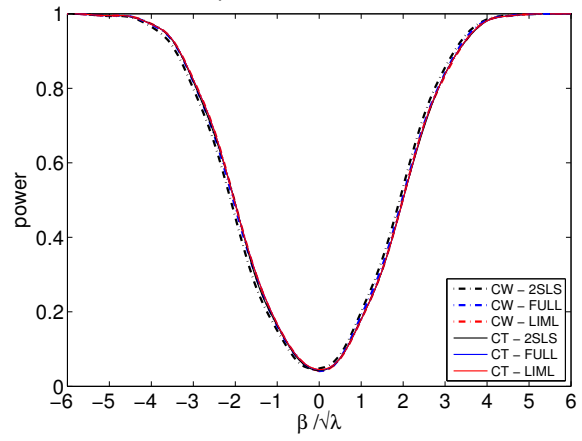


FIGURE 6.2.3A
 $\rho = 0.2, k = 20, \lambda / k = 0.5$

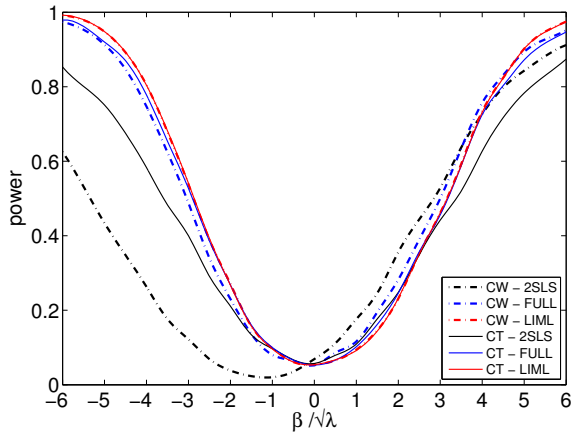


FIGURE 6.2.3B
 $\rho = 0.2, k = 20, \lambda / k = 1$

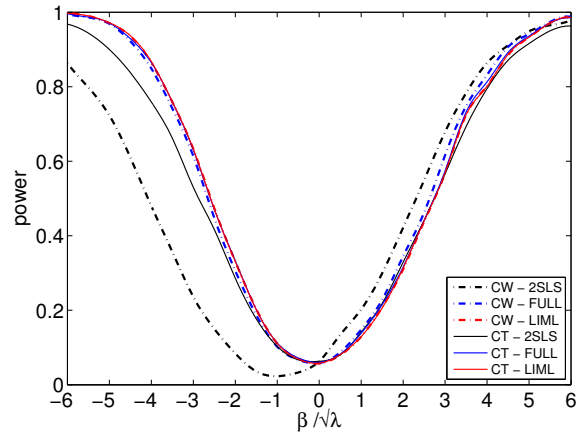


FIGURE 6.2.3C
 $\rho = 0.2, k = 20, \lambda / k = 4$

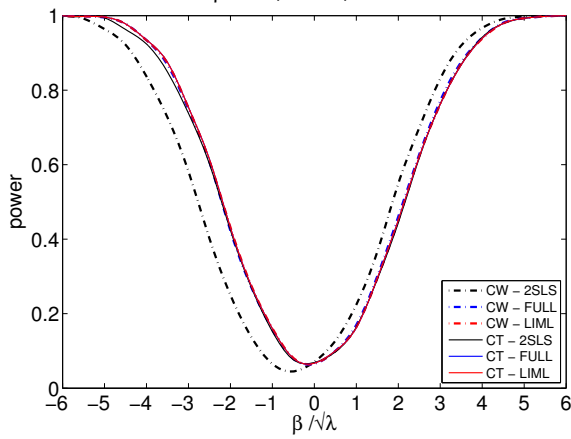


FIGURE 6.2.3D
 $\rho = 0.2, k = 20, \lambda / k = 16$

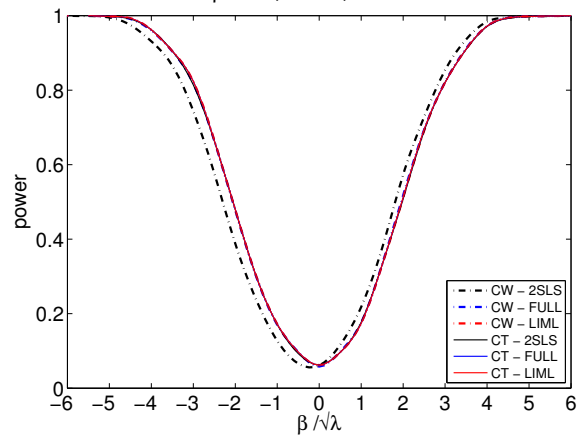


FIGURE 6.2.4A
 $\rho = 0.5, k = 2, \lambda / k = 0.5$

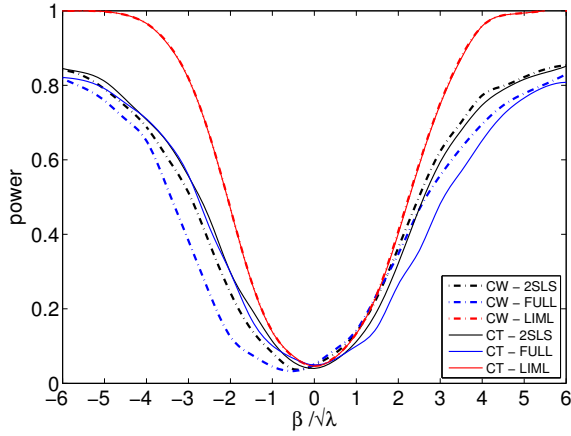


FIGURE 6.2.4B
 $\rho = 0.5, k = 2, \lambda / k = 1$

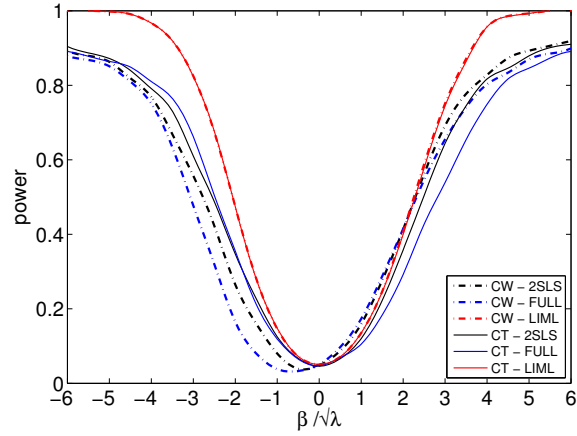


FIGURE 6.2.4C
 $\rho = 0.5, k = 2, \lambda / k = 4$

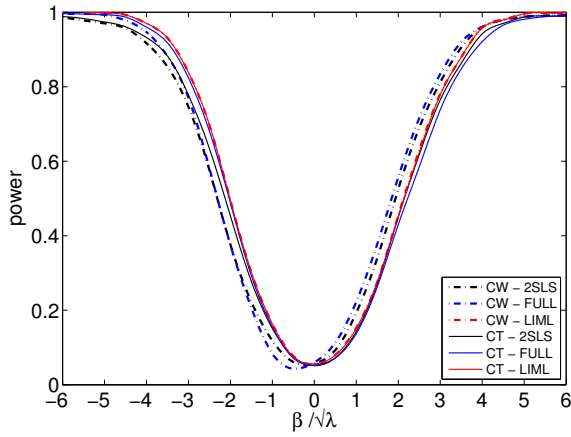


FIGURE 6.2.4D
 $\rho = 0.5, k = 2, \lambda / k = 16$

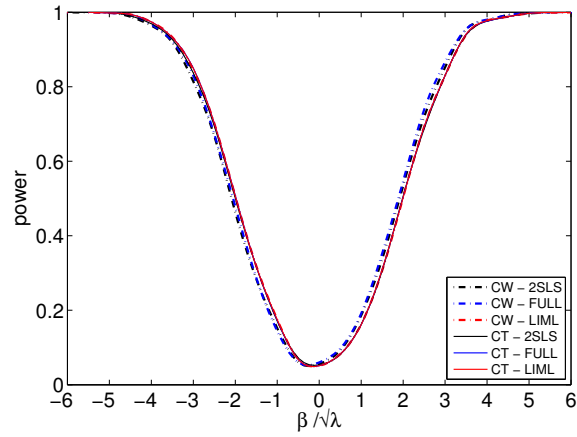


FIGURE 6.2.5A
 $\rho = 0.5, k = 5, \lambda / k = 0.5$

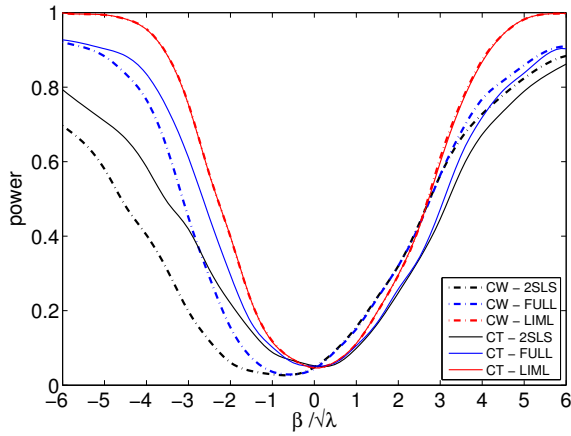


FIGURE 6.2.5B
 $\rho = 0.5, k = 5, \lambda / k = 1$

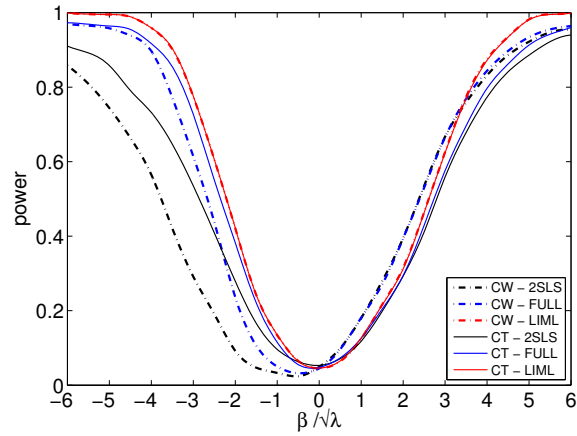


FIGURE 6.2.5C
 $\rho = 0.5, k = 5, \lambda / k = 4$

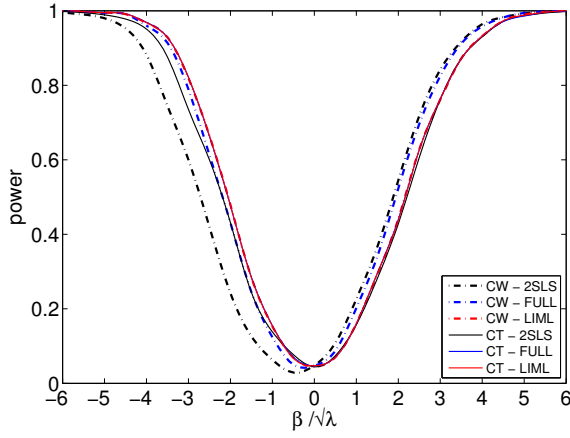


FIGURE 6.2.5D
 $\rho = 0.5, k = 5, \lambda / k = 16$

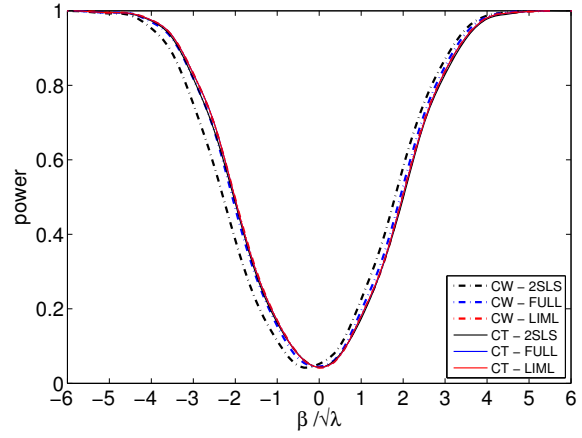


FIGURE 6.2.6A
 $\rho = 0.5, k = 20, \lambda / k = 0.5$

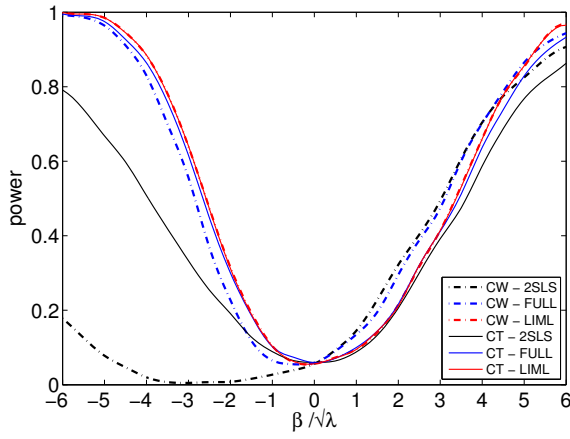


FIGURE 6.2.6B
 $\rho = 0.5, k = 20, \lambda / k = 1$

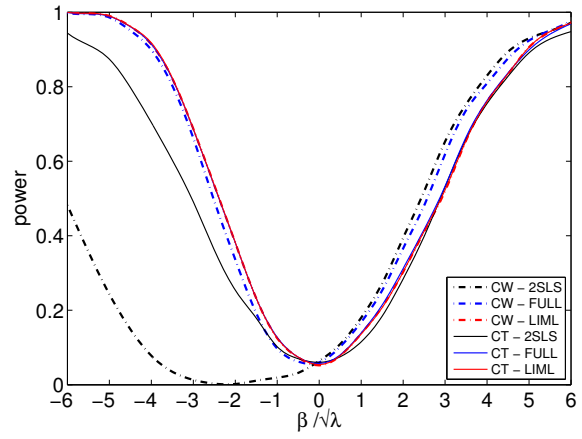


FIGURE 6.2.6C
 $\rho = 0.5, k = 20, \lambda / k = 4$

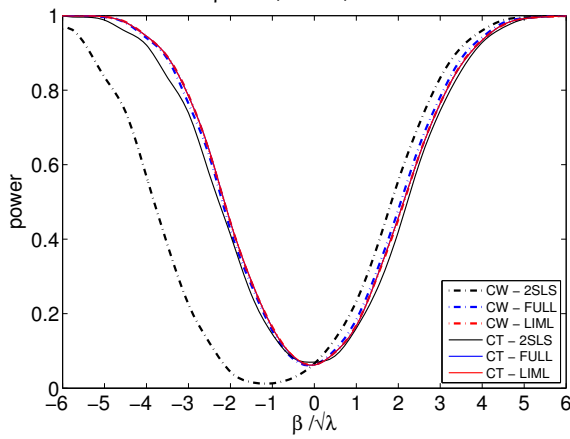


FIGURE 6.2.6D
 $\rho = 0.5, k = 20, \lambda / k = 16$

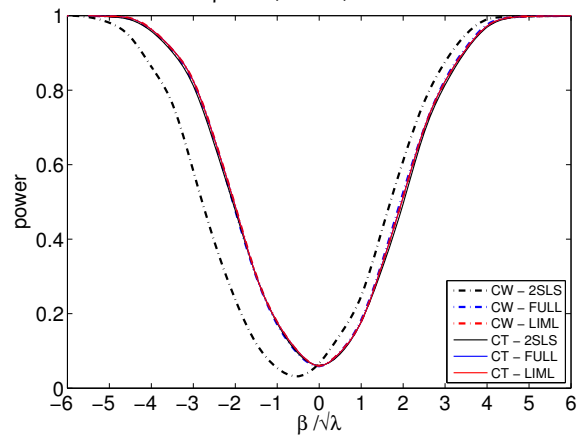


FIGURE 6.2.7A
 $\rho = 0.95, k = 2, \lambda / k = 0.5$

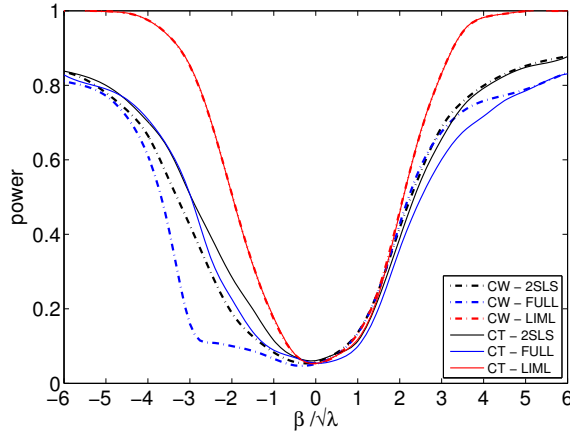


FIGURE 6.2.7B
 $\rho = 0.95, k = 2, \lambda / k = 1$

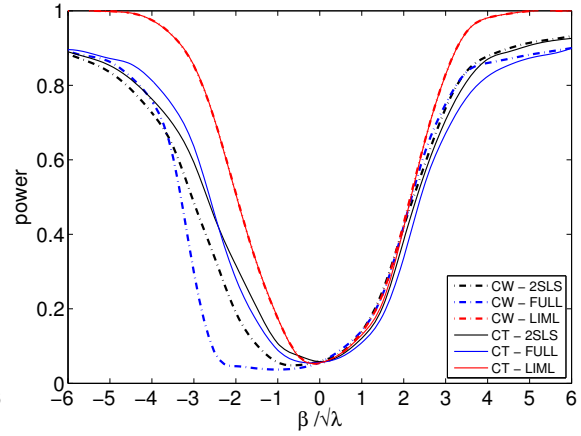


FIGURE 6.2.7C
 $\rho = 0.95, k = 2, \lambda / k = 4$

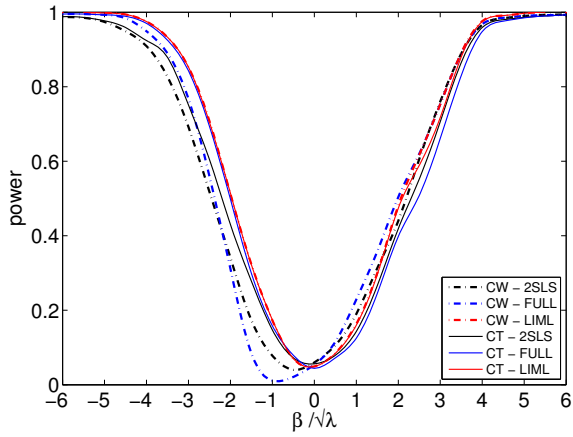


FIGURE 6.2.7D
 $\rho = 0.95, k = 2, \lambda / k = 16$

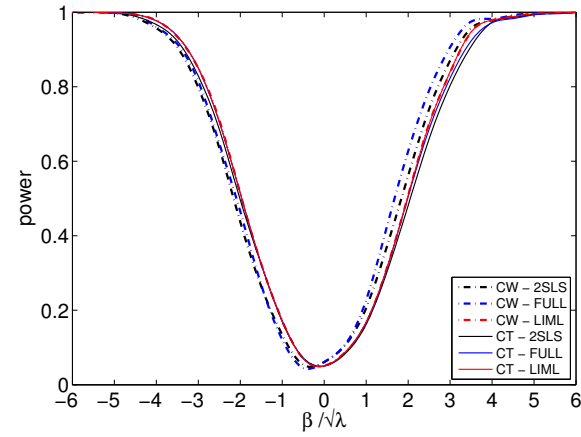


FIGURE 6.2.8A
 $\rho = 0.95, k = 5, \lambda / k = 0.5$

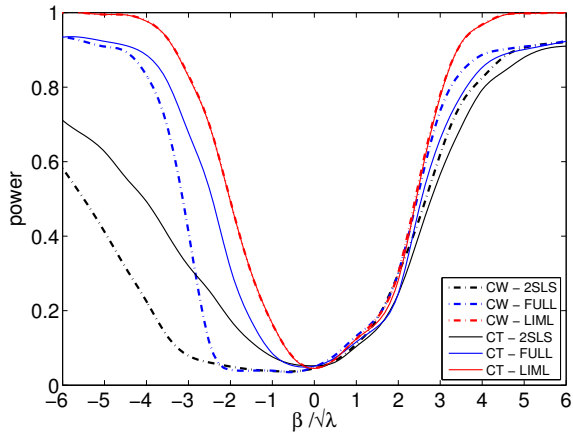


FIGURE 6.2.8B
 $\rho = 0.95, k = 5, \lambda / k = 1$

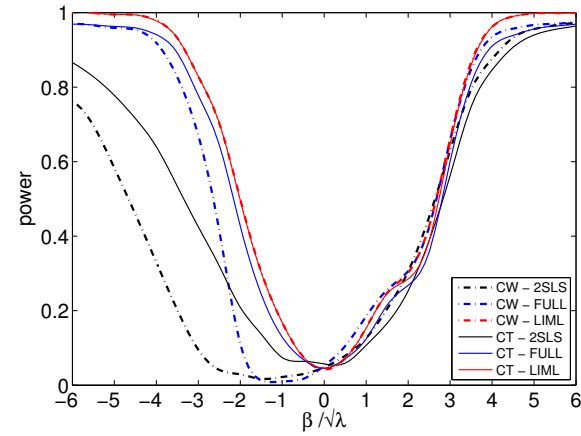


FIGURE 6.2.8C
 $\rho = 0.95, k = 5, \lambda / k = 4$

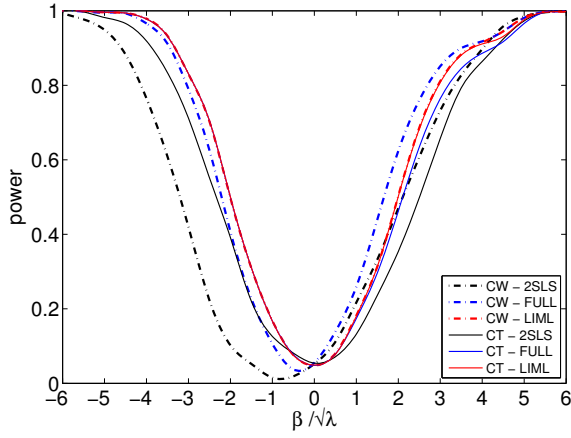


FIGURE 6.2.8D
 $\rho = 0.95, k = 5, \lambda / k = 16$

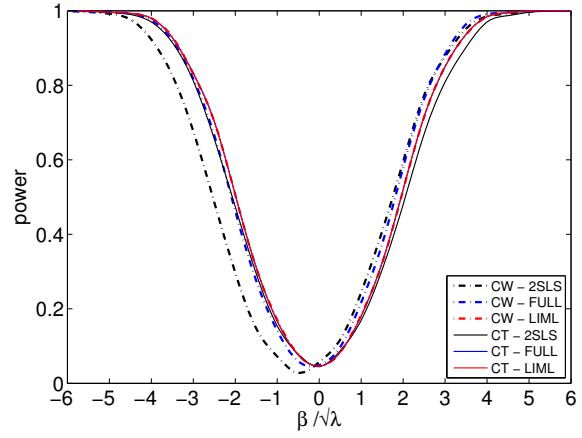


FIGURE 6.2.9A
 $\rho = 0.95, k = 20, \lambda / k = 0.5$

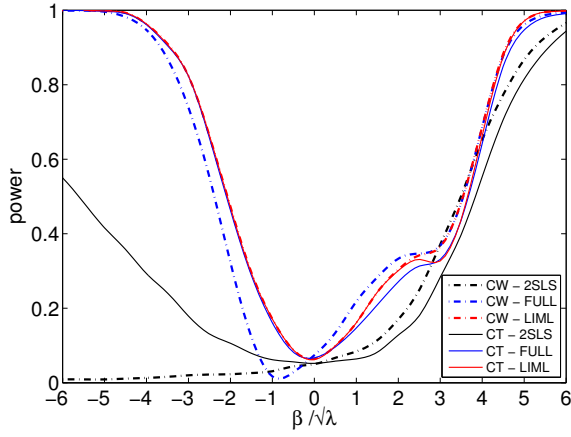


FIGURE 6.2.9B
 $\rho = 0.95, k = 20, \lambda / k = 1$

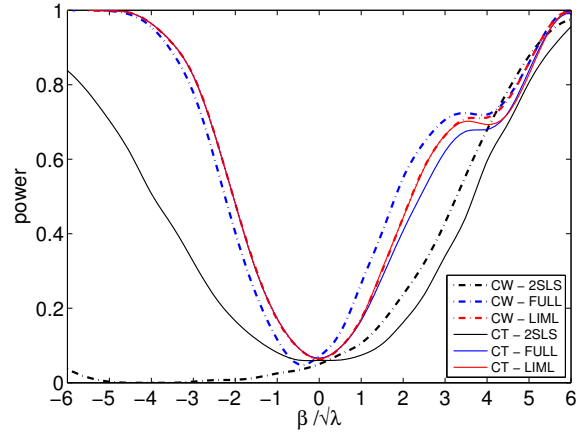


FIGURE 6.2.9C
 $\rho = 0.95, k = 20, \lambda / k = 4$

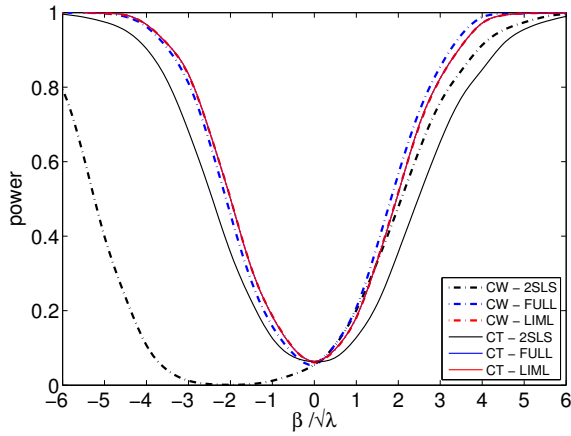
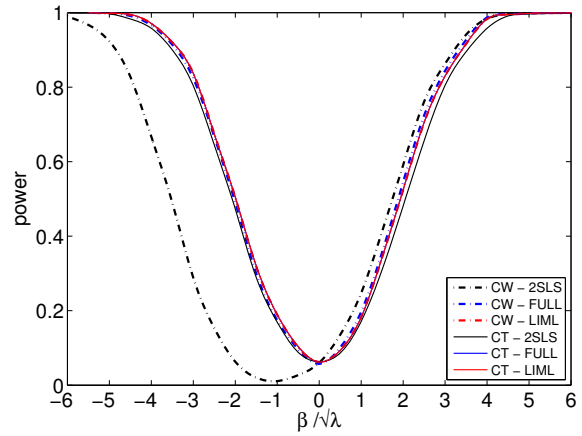


FIGURE 6.2.9D
 $\rho = 0.95, k = 20, \lambda / k = 16$



6.3 Power Curves for the Conditional t-Test Compared with the AR, LM, and CLR

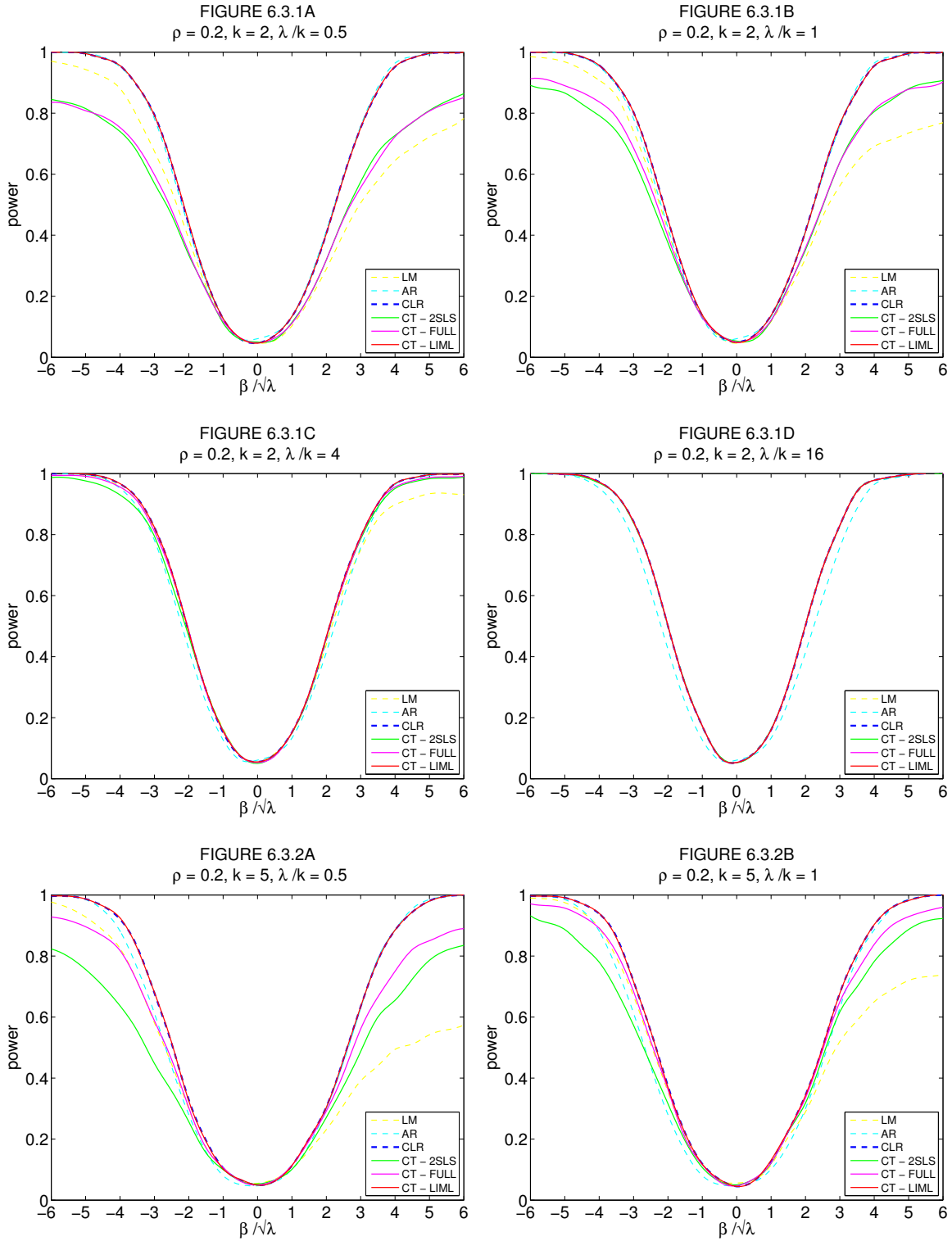


FIGURE 6.3.2C
 $\rho = 0.2, k = 5, \lambda / k = 4$

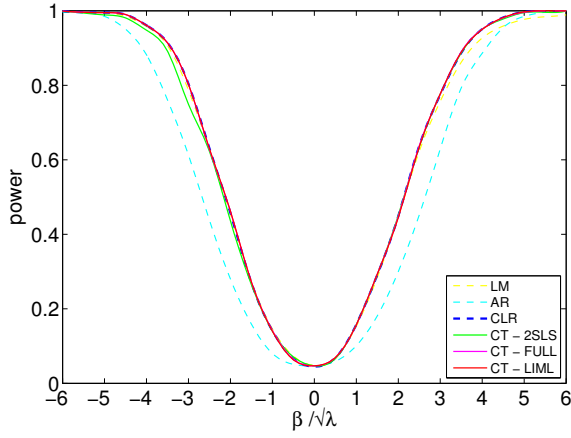


FIGURE 6.3.2D
 $\rho = 0.2, k = 5, \lambda / k = 16$

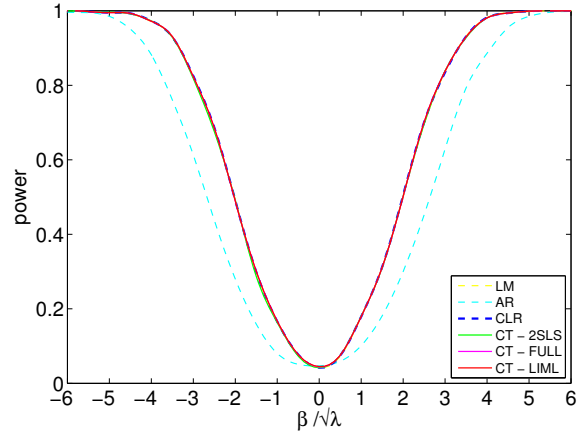


FIGURE 6.3.3A
 $\rho = 0.2, k = 20, \lambda / k = 0.5$

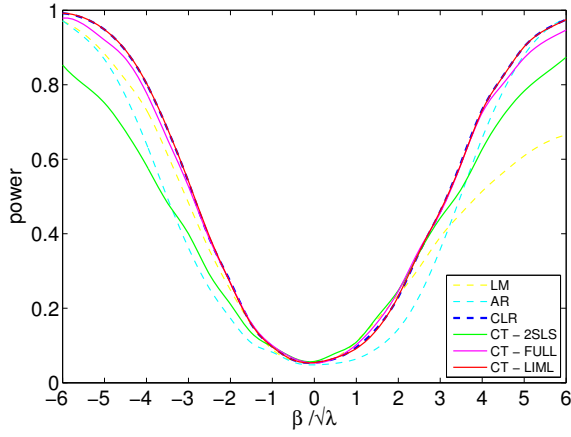


FIGURE 6.3.3B
 $\rho = 0.2, k = 20, \lambda / k = 1$

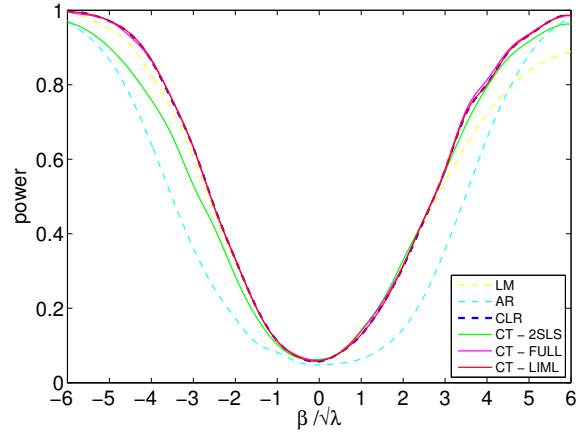


FIGURE 6.3.3C
 $\rho = 0.2, k = 20, \lambda / k = 4$

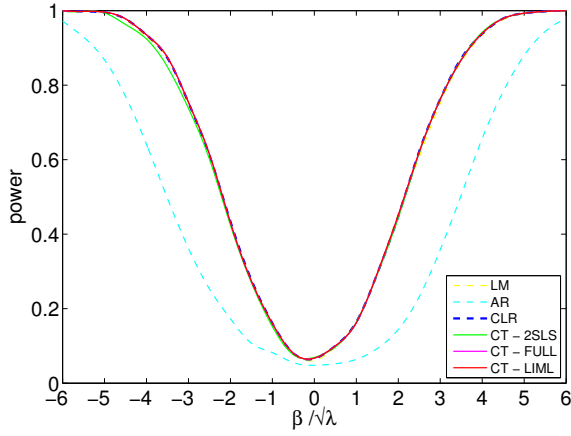


FIGURE 6.3.3D
 $\rho = 0.2, k = 20, \lambda / k = 16$

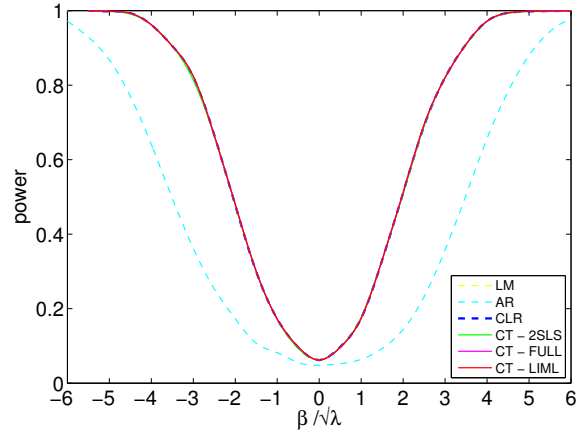


FIGURE 6.3.4A
 $\rho = 0.5, k = 2, \lambda / k = 0.5$

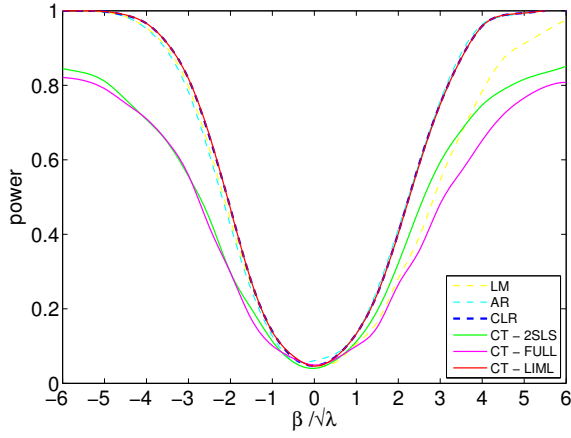


FIGURE 6.3.4B
 $\rho = 0.5, k = 2, \lambda / k = 1$

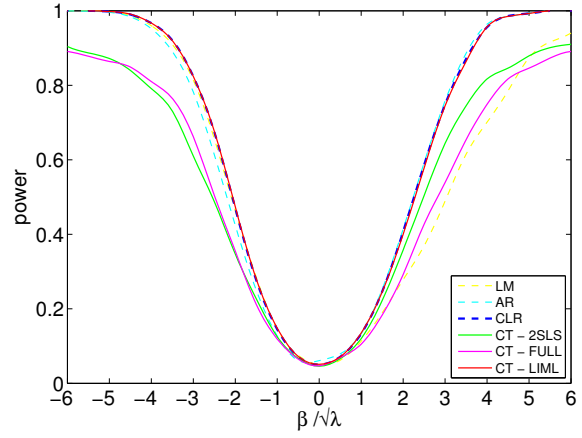


FIGURE 6.3.4C
 $\rho = 0.5, k = 2, \lambda / k = 4$

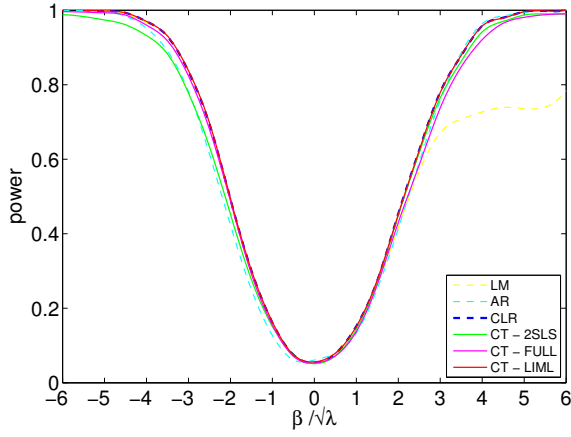


FIGURE 6.3.4D
 $\rho = 0.5, k = 2, \lambda / k = 16$

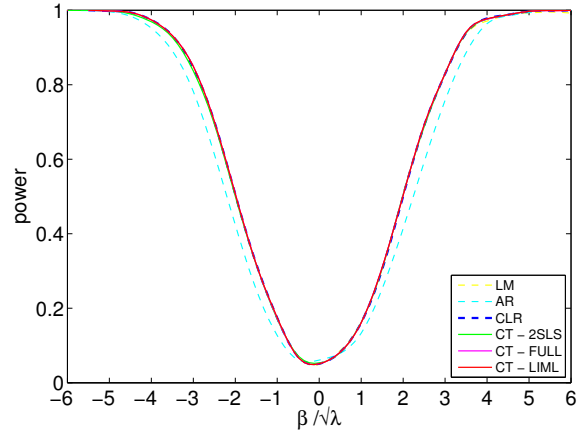


FIGURE 6.3.5A
 $\rho = 0.5, k = 5, \lambda / k = 0.5$

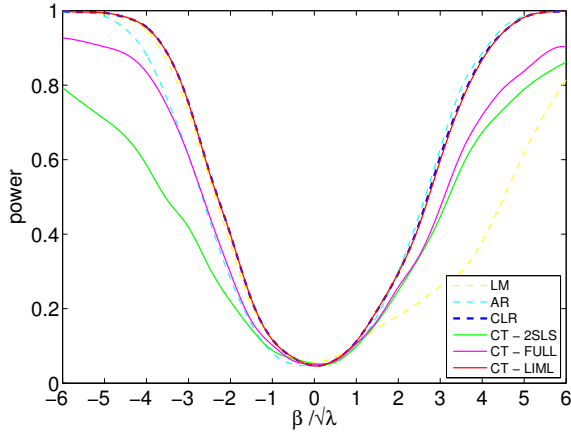


FIGURE 6.3.5B
 $\rho = 0.5, k = 5, \lambda / k = 1$

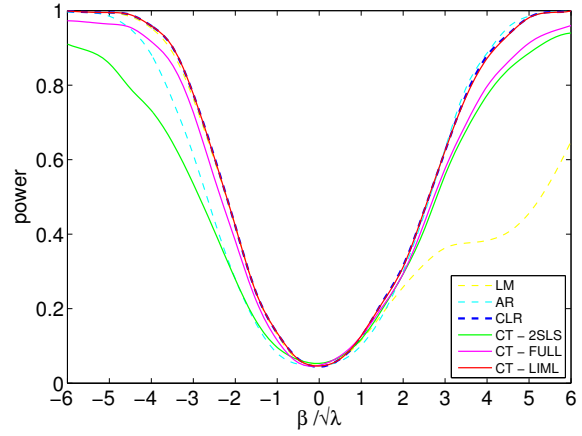


FIGURE 6.3.5C
 $\rho = 0.5, k = 5, \lambda / k = 4$

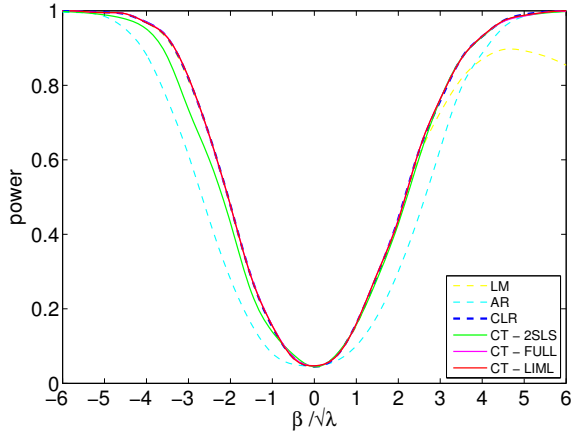


FIGURE 6.3.5D
 $\rho = 0.5, k = 5, \lambda / k = 16$

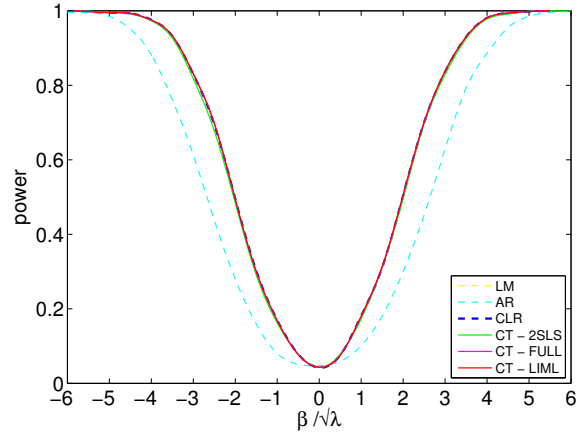


FIGURE 6.3.6A
 $\rho = 0.5, k = 20, \lambda / k = 0.5$

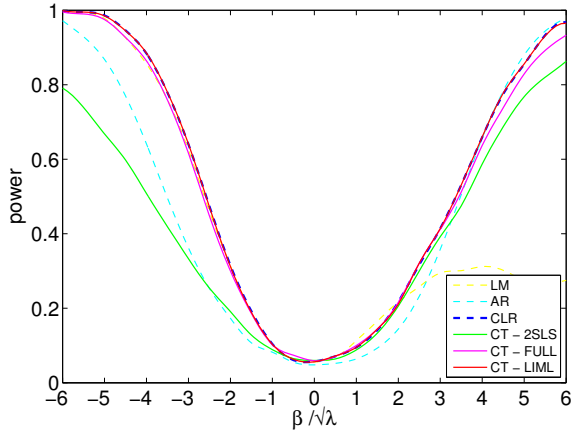


FIGURE 6.3.6B
 $\rho = 0.5, k = 20, \lambda / k = 1$

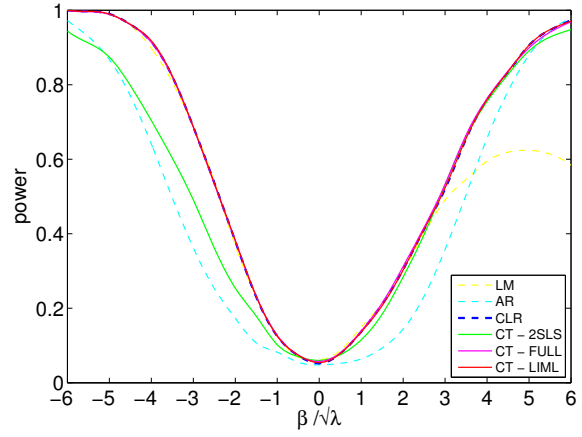


FIGURE 6.3.6C
 $\rho = 0.5, k = 20, \lambda / k = 4$

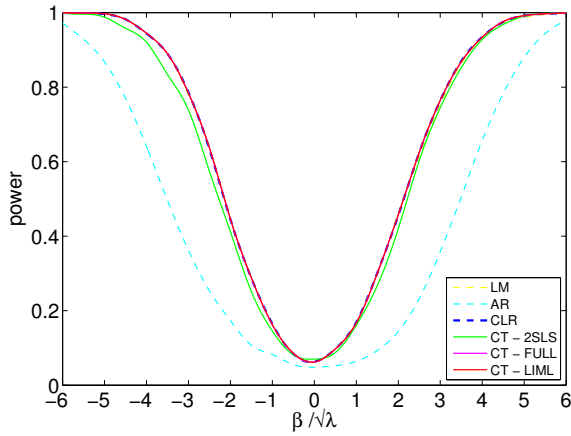


FIGURE 6.3.6D
 $\rho = 0.5, k = 20, \lambda / k = 16$

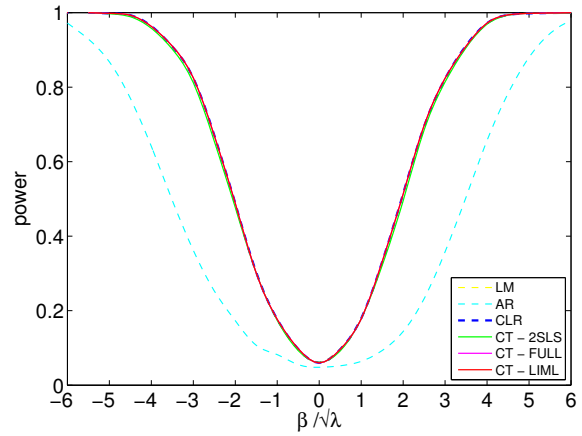


FIGURE 6.3.7A
 $\rho = 0.95, k = 2, \lambda / k = 0.5$

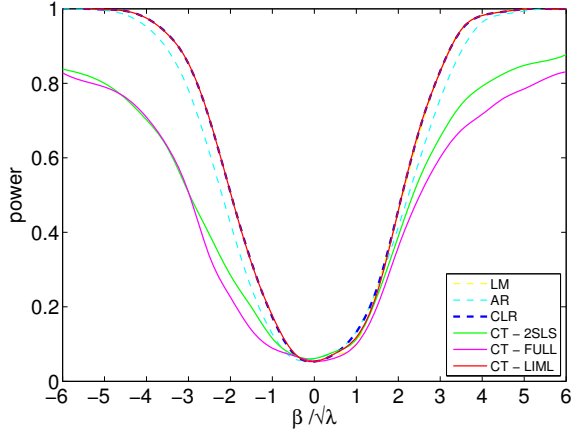


FIGURE 6.3.7B
 $\rho = 0.95, k = 2, \lambda / k = 1$

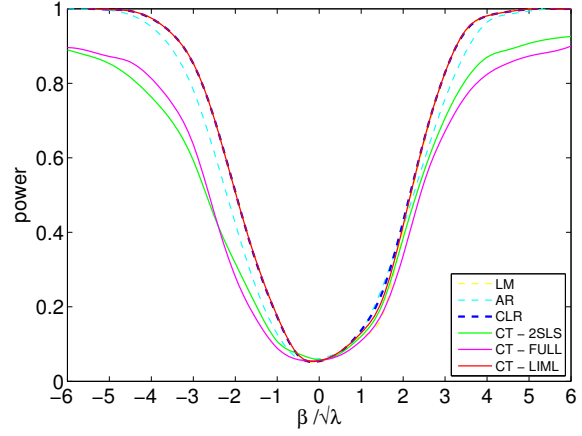


FIGURE 6.3.7C
 $\rho = 0.95, k = 2, \lambda / k = 4$

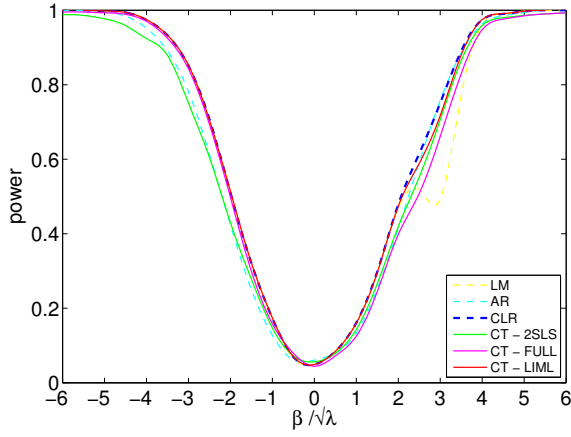


FIGURE 6.3.7D
 $\rho = 0.95, k = 2, \lambda / k = 16$

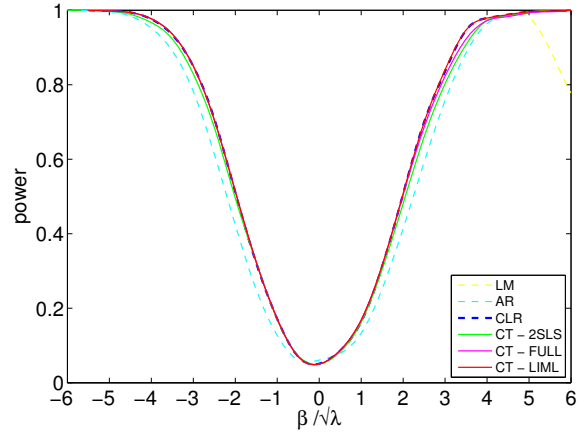


FIGURE 6.3.8A
 $\rho = 0.95, k = 5, \lambda / k = 0.5$

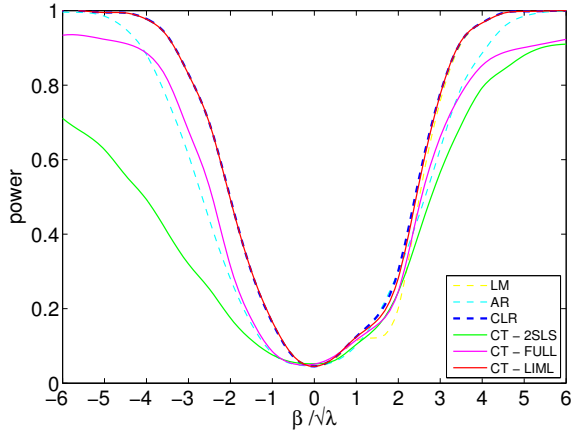


FIGURE 6.3.8B
 $\rho = 0.95, k = 5, \lambda / k = 1$

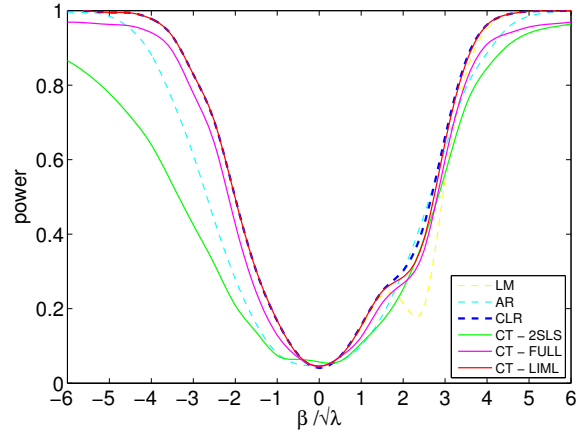


FIGURE 6.3.8C
 $\rho = 0.95, k = 5, \lambda/k = 4$

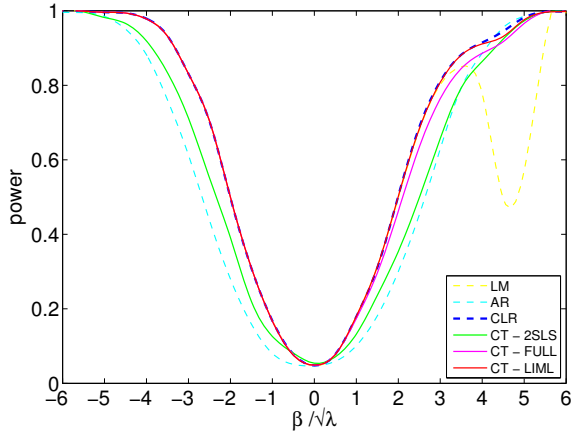


FIGURE 6.3.8D
 $\rho = 0.95, k = 5, \lambda/k = 16$

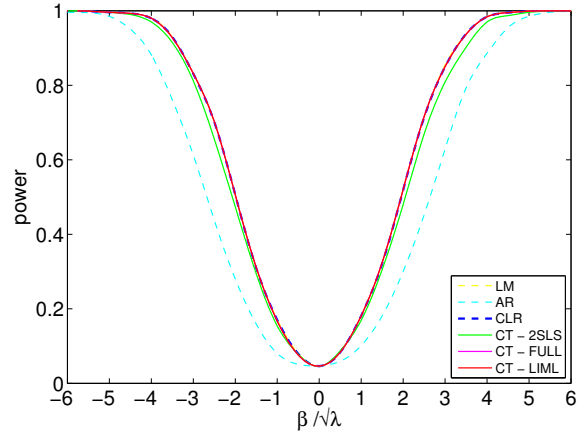


FIGURE 6.3.9A
 $\rho = 0.95, k = 20, \lambda/k = 0.5$

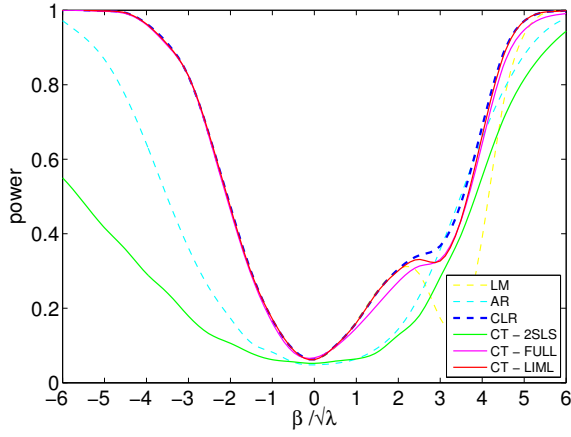


FIGURE 6.3.9B
 $\rho = 0.95, k = 20, \lambda/k = 1$

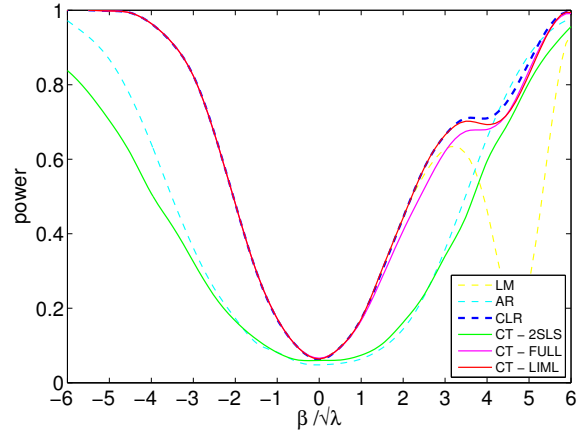


FIGURE 6.3.9C
 $\rho = 0.95, k = 20, \lambda/k = 4$

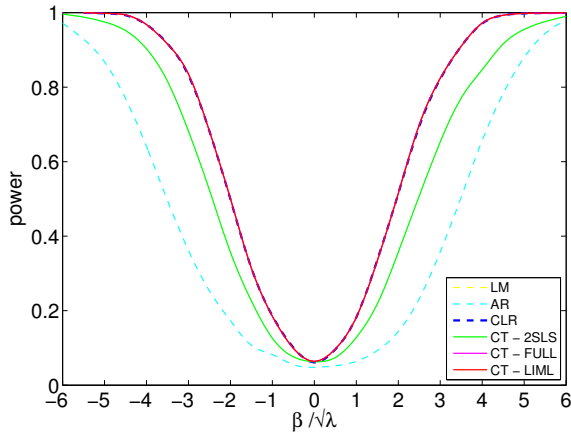
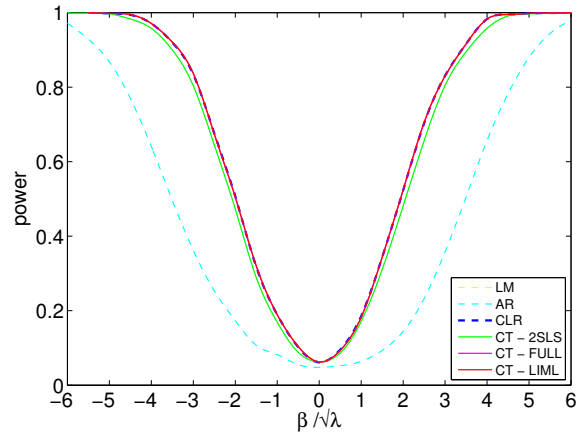


FIGURE 6.3.9D
 $\rho = 0.95, k = 20, \lambda/k = 16$



6.4 Proofs

PROOF OF LEMMA 3.1.1: Since the expectation in the left-hand side of 3.1.2 is a probability, it is always between one and zero, and thus exists. By Rudin [1976], Definition 11.34, under the null, $\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} \in \mathcal{L}^2$ and $LM \in \mathcal{L}^2$. Then by Rudin [1976], Theorem 11.35,

$$\begin{aligned} \left| E_{\beta_0} \left(\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} LM \right) \right| &\leq E_{\beta_0} \left(\left| \mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} LM \right| \right) \\ &\leq \left\| \mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} \right\| \|LM\| \\ &\quad \left(\text{where } \|f\| = \left\{ \int |f|^2 d\mu \right\}^{1/2} \right) \\ &\leq \|LM\| \\ &< \infty \quad \square \end{aligned}$$

PROOF OF LEMMA 3.1.2:

a)

$$\begin{aligned} g(C_1(q_T), C_2(q_T)) &= E_{\beta_0} \left(\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} \right) \\ &= P(C_1(q_T) < t_\kappa(Q_S, LM, q_T) < C_2(q_T)) \\ &= P(t_\kappa(Q_S, LM, q_T) < C_2(q_T)) - P(t_\kappa(Q_S, LM, q_T) < C_1(q_T)) \\ &= P(t_\kappa(Q_S, LM, q_T) \leq C_2(q_T)) - P(t_\kappa(Q_S, LM, q_T) \leq C_1(q_T)) \quad \text{Since the cdf of } t_\kappa(Q_S, LM, q_T) \\ &= F_{t_\kappa(Q_S, LM, q_T)}(C_2(q_T)) - F_{t_\kappa(Q_S, LM, q_T)}(C_1(q_T)) \quad \text{is continuous.} \\ &\quad \text{(a sum of two cdfs.)} \end{aligned}$$

$\therefore g(C_1(q_T), C_2(q_T)) = E_{\beta_0} \left(\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} \right)$ is a continuous function of $C_1(q_T)$ and $C_2(q_T)$. \square

b)

Let p_0 be the probability density under the null with respect to a measure μ and let $(C_1^n(q_T), C_2^n(q_T)) \rightarrow (C_1^*(q_T), C_2^*(q_T))$.

Note that since

$$0 \leq \mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} \leq 1$$

we have

$$\left| \mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} LM \right| \leq |LM|.$$

Define

$$\varphi_{C_1, C_2} = \mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\}.$$

Then

$$\begin{aligned} \lim_{n \rightarrow \infty} g^*(C_1^n(q_T), C_2^n(q_T)) &= \lim_{n \rightarrow \infty} E_{\beta_0}(\mathbb{I}\{C_1^n(q_T) < t_\kappa(Q_S, LM; q_T) < C_2^n(q_T)\} LM) \\ &= \lim_{n \rightarrow \infty} \int (\varphi_{C_1^n, C_2^n}) LM p_0 d\mu \\ &= \int \lim_{n \rightarrow \infty} (\varphi_{C_1^n, C_2^n}) LM p_0 d\mu \quad \text{by dominated convergence.} \\ &= \int (\varphi_{C_1^*, C_2^*}) LM p_0 d\mu \\ &= E_{\beta_0}\{(\varphi_{C_1^*, C_2^*}) LM\} \end{aligned}$$

$\therefore g^*(C_1(q_T), C_2(q_T)) = E_{\beta_0}(\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T)\} LM)$ is a continuous function of $C_1(q_T)$ and $C_2(q_T)$. \square

PROOF OF THEOREM 3.1.3:

a) By Lemma 3.1.1, the expectations both exist. Since $\{Q_S^j, LM^j\}$ is i.i.d., the sequences $\left\{\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T)\} LM^j\right\}$ and $\left\{\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T)\}\right\}$ are i.i.d.. The results thus follow as a straightforward application of the Strong Law of Large Numbers. \square

b) Let

$$\hat{Q}_n((C_1(q_T), C_2(q_T))) = \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I}\{C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T)\} LM^j \right|$$

and

$$Q_0((C_1(q_T), C_2(q_T))) = E_{\beta_0}\left(\left|\mathbb{I}\{C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T)\} LM^j\right|\right)$$

Fix $Q_T = q_T$ and let $(C_1^*(q_T), C_2^*(q_T))$ be the solutions to (3.1.4). $(C_1^*(q_T), C_2^*(q_T))$ is a point in the interior of a convex set $\Theta \subset \mathbb{R}^2$. By assumption, $Q_0(C_1(q_T), C_2(q_T))$ is uniquely maximized at $(C_1^*(q_T), C_2^*(q_T))$. By theorem 3.1.3, $\hat{Q}_n((C_1(q_T), C_2(q_T))) \xrightarrow{p} Q_0((C_1(q_T), C_2(q_T)))$ for all $(C_1(q_T), C_2(q_T)) \in \Theta$. Since $\hat{Q}_n((C_1(q_T), C_2(q_T)))$ is a convex function, the result follows from Newey and McFadden [1994] Theorem 2.7. \square

PROOF OF COROLLARY 3.1.5: By definition 3.1.4,

$$\begin{aligned} \left(C_1^J(q_T), C_2^J(q_T) \right) &= \underset{(C_1(q_T), C_2(q_T))}{\operatorname{argmin}} \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} LM^j \right| \\ \text{such that} \quad & \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} = 1 - \alpha \end{aligned}$$

Taking the probability limit,

$$\begin{aligned} \operatorname{plim}_{J \rightarrow \infty} \left(C_1^J(q_T), C_2^J(q_T) \right) &= \operatorname{plim}_{J \rightarrow \infty} \underset{(C_1(q_T), C_2(q_T))}{\operatorname{argmin}} \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} LM^j \right| \\ \text{such that} \quad & \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} = 1 - \alpha . \end{aligned}$$

Then by Theorem 3.1.2b,

$$\begin{aligned} \operatorname{plim}_{J \rightarrow \infty} \left(C_1^J(q_T), C_2^J(q_T) \right) &= \underset{(C_1(q_T), C_2(q_T))}{\operatorname{argmin}} \operatorname{plim}_{J \rightarrow \infty} \left| \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} LM^j \right| \\ \text{such that} \quad & \frac{1}{J} \sum_{j=1}^J \mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S^j, LM^j; q_T) < C_2(q_T) \right\} = 1 - \alpha . \end{aligned}$$

By Theorem 3.1.2a,

$$\begin{aligned} \operatorname{plim}_{J \rightarrow \infty} \left(C_1^J(q_T), C_2^J(q_T) \right) &= \underset{(C_1(q_T), C_2(q_T))}{\operatorname{argmin}} \left| E_{\beta_0} \left(\mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \right\} LM \right) \right| \\ \text{such that} \quad & E_{\beta_0} \left(\mathbb{I} \left\{ C_1(q_T) < t_\kappa(Q_S, LM; q_T) < C_2(q_T) \right\} \right) = 1 - \alpha . \end{aligned}$$

Hence by 3.1.4.,

$$\operatorname{plim}_{J \rightarrow \infty} \left(C_1^J(q_T), C_2^J(q_T) \right) = \left(C_1(q_T), C_2(q_T) \right) . \quad \square$$

6.5 MATLAB Code, Results, and Replication Files

Annotated MATLAB code used for simulations and graphing, power results, and all other files necessary for replication can be found at:

http://www.columbia.edu/~bsm2112/thesis_replication.zip

References

- T. W. Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):pp. 46–63, 1949.
- Donald W. K. Andrews, Marcelo J. Moreira, and James H. Stock. Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica*, 74(3):715–752, 2006a.
- Donald W. K. Andrews, Marcelo J. Moreira, and James H. Stock. Supplement to 'optimal two-sided invariant similar tests for instrumental variables regression'. *Econometrica*, 74(3):pp. 715–752, 2006b.
- Donald W. K. Andrews, Marcelo J. Moreira, and James H. Stock. Performance of conditional wald tests in iv regression with weak instruments. *Journal of Econometrics*, 139(1):116–132, July 2007.
- T. S. Breusch and A. R. Pagan. The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):pp. 239–253, 1980.
- Wayne A Fuller. Some properties of a modification of the limited information estimator. *Econometrica*, 45(4):939–53, 1977.
- Frank Kleibergen. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):pp. 1781–1803, 2002.
- E. L Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer, New York, 3rd ed edition, 2005.
- MJ Moreira. *Tests with correct size in the simultaneous equation model*. PhD thesis, UC Berkeley, 2002.
- MJ Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048, July 2003.
- A. L. Nagar. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica*, 27(4):pp. 575–595, 1959.
- Whitney K. Newey and Daniel McFadden. Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier, 1994.
- Walter Rudin. *Principles of mathematical analysis*. International series in pure and applied mathematics. McGraw-Hill, New York, 3d ed edition, 1976.

D Staiger and JH Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3): 557–586, May 1997.