OPTIMAL INVARIANT SIMILAR TESTS FOR INSTRUMENTAL VARIABLES REGRESSION

By

Donald W.K. Andrews, Marcelo J. Moreira and James H. Stock

August 2004

COWLES FOUNDATION DISCUSSION PAPER NO. 1476

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Optimal Invariant Similar Tests for Instrumental Variables Regression

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

Marcelo J. Moreira
Department of Economics
Harvard University

James H. Stock
Deparment of Economics
Harvard University

October 2002
Revised: July 2004

[1] Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0001706. Stock acknowledges support via NSF grant number SBR-0214131.

Abstract

This paper considers tests of the parameter on endogenous variables in an instrumental variables regression model. The focus is on determining tests that have some optimal power properties. We start by considering a model with normally distributed errors and known error covariance matrix. We consider tests that are similar and satisfy a natural rotational invariance condition. We determine tests that maximize weighted average power (WAP) for arbitrary weight functions among invariant similar tests. Such tests include point optimal (PO) invariant similar tests.

The results yield the power envelope for invariant similar tests. This allows one to assess and compare the power properties of existing tests, such as the Anderson-Rubin, Lagrange multiplier (LM), and conditional likelihood ratio (CLR) tests, and new optimal WAP and PO invariant similar tests. We find that the CLR test is quite close to being uniformly most powerful invariant among a class of two-sided tests. A new unconditional test, P^*, also is found to have this property. For one-sided alternatives, no test achieves the invariant power envelope, but a new test—the one-sided CLR test—is found to be fairly close.

The finite sample results of the paper are extended to the case of unknown error covariance matrix and possibly non-normal errors via weak instrument asymptotics. Strong instrument asymptotic results also are provided because we seek tests that perform well under both weak and strong instruments.

Keywords: Instrumental variables regression, invariant tests, optimal tests, similar tests, weak instruments, weighted average power.

JEL Classification Numbers: C12, C30.

1 Introduction

In instrumental variables regression with a single included endogenous regressor, instruments (IV's) are said to be weak when the partial correlation between the IV's and the included endogenous regressor is small, given the included exogenous regressors. The effect of weak IV's is to make the standard asymptotic approximations to the distributions of estimators and test statistics poor. Consequently, hypothesis tests with conventional asymptotic justifications, such as the Wald test based on the two stage least squares estimator, can exhibit large size distortions.

A number of papers have proposed methods for testing hypotheses about the coefficient, β, on the included endogenous regressors that are valid even when IV's are weak. Except for the important early contribution by Anderson and Rubin (1949) (AR), most of this literature is recent. It includes Staiger and Stock (1997), Zivot, Startz, and Nelson (1998), Wang and Zivot (1998), Dufour and Jasiak (2001), Kleibergen (2002), Moreira (2001, 2003), Dufour and Taamouti (2003), and Startz, Zivot, and Nelson (2003). None of these contributions develops a satisfactory theory of optimal inference in the presence of potentially weak IV's. Absent such a theory, comparisons of power to date between competing valid tests are numerical and incomplete.[2]

The purpose of this paper is to develop a theory of optimal hypothesis testing when IV's might be weak, and to use this theory to develop practical valid hypothesis tests that are nearly optimal whether IV's are weak or strong. We adopt the natural invariance condition that inferences are unchanged if the IV's are transformed by an orthogonal matrix, e.g., changing the order in which the IV's appear. The resulting class of invariant tests includes all tests proposed for this problem of which we are aware, except those that entail potentially dropping an IV. We focus on the practically important case of a single endogenous variable (but some results for multiple endogenous variables are provided).

We show that there does not exist a uniformly most powerful invariant (UMPI) one-sided or two-sided test of $H_0 : \beta = \beta_0$. Our numerical results, however, demonstrate that there are tests that are very nearly optimal, in the sense that their power functions are numerically very close to the power envelope uniformly in the parameter space. In particular, the conditional likelihood ratio (CLR) test proposed by Moreira (2003) is numerically nearly UMPI, as is a new (unconditional) test, the P* test introduced below, which is motivated by the theory of point optimal invariant testing. We recommend the use of the CLR or P* test in empirical practice.

The optimality results are developed for strictly exogenous IV's, linear structural and reduced form equations, and homoskedastic Gaussian errors with a known covariance matrix. For this model, we obtain the sufficient statistics, the maximal invariant (under orthogonal transformations of the IV's), and the distribution of the maximal invariant. We determine necessary and sufficient conditions for invariant tests to be similar. For a one-sided alternative hypothesis, we derive optimal weighted average power (WAP) invariant similar tests. This gives, as a special case, the Gaussian

[2] See Stock, Yogo, and Wright (2002), Dufour (2003), and Hahn and Hausman (2003) for surveys of research on weak IV's. Also, see the recent paper by Forchini and Hillier (2003).

power envelope for one-sided point-optimal invariant similar (POIS) tests and the result that there does not exist a UMPI test, either one- or two-sided.

We propose a new one-sided invariant similar test, the one-sided CLR test (CLR1). Although the one-sided POIS tests have non-monotonic power functions and are undesirable for practical use, the CLR1 test is found to have a power function that is typically not too far below the one-sided Gaussian power envelope, making it an attractive choice for one-sided testing.

In addition to similar tests, we consider optimal non-similar tests for one-sided alternatives using the least-favorable distribution approach described, e.g., in Lehmann (1986). Although the nonsimilar and similar tests differ in theory, we find that the power envelopes of invariant similar and nonsimilar tests are numerically very close.

We consider four approaches to developing tractable families of two-sided invariant similar tests and two-sided power envelopes. The first consists of WAP tests that are symmetric in β around the null value β_0. These tests have the undesirable feature of not being consistent against both alternatives when the IV's are strong. So, we do not pursue such tests further. The second approach is to consider WAP tests that are asymptotically efficient two-sided tests when the IV's are strong. This includes a class of POIS tests based on two-point weight functions. The third is to consider tests that satisfy an additional invariance condition that seems natural in the two-sided problem. The fourth is to consider tests that are unbiased against local alternatives.

We prove that the power envelopes for the second and third approaches to two-sided tests are identical. Moreover, as a numerical matter, these power envelopes are very close to those of locally unbiased tests. We refer to the power envelopes based on the second and third approaches as two-sided asymptotically efficient (AE) power envelopes. Although no UMPI test exists among this class, we find that as a numerical matter the power of the CLR test is uniformly very close to the power envelope for this class, and in this sense the CLR test is approximately UMPI. This is one of the major findings of the paper.

It is known that the power function of the Lagrange multiplier (LM) test is not monotonic. Our theoretical results indicate why this is so. Our numerical results indicate that its power is never above that of the CLR test, and in some cases is far below. Hence, the CLR test dominates the LM test in terms of power and we do not recommend the LM test for practical use.

We also consider POI *nonsimilar* two-sided tests (subject to the additional invariance condition as in the third approach). These tests generally have power functions close to the power envelope. So, following King (1988), we examine the performance of some specific POI nonsimilar tests. One such test, which we call the P* test, is found to be approximately as powerful as the CLR test. Because the P* test does not entail using a table of conditional critical values, researchers might find it more convenient than the CLR test in practice.

The foregoing results are developed treating the reduced-form error covariance matrix as known. In practice, this matrix is unknown and must be estimated. Using Staiger-Stock (1997) weak-IV asymptotics, we show that the exact distributional results extend, in large samples, to feasible versions of these statistics using an estimated covariance matrix and possibly non-normal errors. We show that these finite-sample power envelopes derived with known covariance matrix are also the asymptotic Gaussian power envelopes with unknown covariance matrix, under weak-IV asymptotics. In a Monte Carlo study of the LM, CLR, AR, and P* tests with estimated covariance matrices, we find that sample sizes of 100-200 observations are sufficient for (i) the size of these tests to be well controlled using weak-IV asymptotic critical values and (ii) the weak-IV asymptotic power functions to be good approximations to the finite-sample power functions.

This "plug-in" approach to the reduced form covariance matrix makes it possible to introduce versions of the AR, LM, CLR, CLR1, POIS, and WAP test statistics that are robust to heteroskedasticity or, for time series applications, to heteroskedasticity and autocorrelation. We show that the weak-IV asymptotic distributions of these robust statistics in the presence of heteroskedasticity (or heteroskedasticity and autocorrelation) are those derived in the exact Gaussian model with known covariance matrix.

Finally, we obtain asymptotic properties of the tests considered in this paper when the IV's are strong. These results are essential for determining the class of WAP tests that are asymptotically efficient under strong IV's, which lies behind the second approach to determining two-sided tests. The LM and CLR tests are shown to be asymptotically efficient with strong IV's against local alternatives, although (as is known) the AR test is not. Necessary and sufficient conditions are determined for WAP invariant similar tests based on two-point weight functions to be AE under strong IV asymptotics. Such tests determine the two-sided AE power envelope. The LM, CLR, CLR1, 2-sided AE POIS and two-sided AE WAP tests are also shown to be consistent against fixed alternatives under strong IV's. Curiously, one-sided POIS tests are not consistent against fixed alternatives, and can reject with asymptotic probability zero against alternatives on the "correct" side but with probability one against alternatives on the "wrong" side. Theoretical results explain why this occurs.

Numerous additional numerical results that supplement those given in Sections 8 and 12 below are provided in Andrews, Moreira, and Stock (2004) (hereafter denoted AMS-04).

Other papers that consider optimal testing in the exact Gaussian IV regression model are Moreira (2001) and Chamberlain (2003). Moreira (2001) develops a theory of optimal one-sided testing without an invariance condition and uses this to develop one-sided power envelopes. However, without the invariance condition the family of tests is too large to obtain nearly optimal tests.

Chamberlain (2003) considers minimax decision procedures (including estimators and tests) in the normal model with known covariance matrix that is considered in this paper. His results for tests show that the imposition of the invariance condition considered here does not affect the minimax decision problem. Hence, his results provide a formal minimax justification for the restriction to invariant tests that is adopted in this paper. Chamberlain (2003) does not impose a similarity condition and does not consider restrictions to tests with two-sided properties (although one can do so via the choice of prior employed), so the class of tests he considers are analogous to the one-sided non-similar tests considered in Section 7.1 below. Chamberlain (2003) does not explore the properties of the tests in this class, such as their relative power and the distance of their power functions from the power envelope, which are key items of interest here.

The remainder of this paper is organized as follows. Section 2 introduces the model which has one endogenous regressor variable, multiple exogenous regressor variables, multiple IV's, normally distributed errors, and known covariance matrix. This section determines sufficient statistics for this model. Section 3 introduces a natural invariance condition concerning orthogonal rotations of the IV matrix. It also provides necessary and sufficient conditions for invariant tests to be similar. Section 4 specifies a WAP criterion and determines invariant similar tests that maximize WAP. Section 5 determines the power envelope for one-sided tests by determining the class of POIS tests. Section 6 specifies optimal WAP tests for two-sided alternatives. Section 7 determines optimal invariant non-similar WAP tests. Section 8 presents numerical results for the tests considered in earlier sections. Section 9 adjusts the tests considered in Sections 4 and 6 to allow for an estimated error covariance matrix and analyzes the asymptotic properties of these tests under weak IV's and possibly non-normal errors. This Section also introduces versions of these tests, as well as versions of the AR, LM, CLR, and CLR1 tests, that are robust to heteroskedasticity and other versions that are robust to both heteroskedasticity and autocorrelation. Section 10 provides a weak IV asymptotic optimal WAP result for the tests introduced in Section 9 under the assumption of iid normal errors and unknown covariance matrix Ω. Section 11 provides the asymptotic properties of WAP tests under strong IV's when the error covariance matrix is unknown and the errors may be non-normal. Section 12 presents simulation results for the tests introduced in Section 9 for models with an unknown covariance matrix. Section 13 determines tests that maximize WAP in an IV regression model that is the same as in Section 2, but with *multiple* endogenous regressor variables. An Appendix contains proofs of the results.

2 Model and Sufficient Statistics

In this section, we consider a model with one endogenous variable, multiple exogenous variables, multiple IV's, and errors that are normal with known covariance matrix. In latter sections, we allow for non-normal errors with unknown covariance matrix and multiple endogenous variables.

The model consists of a structural equation and a reduced-form equation:

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

where $y_1, y_2 \in R^n$, $X \in R^{n\times p}$, and $\widetilde{Z} \in R^{n\times k}$ are observed variables; $u, v_2 \in R^n$ are unobserved errors; and $\beta \in R$, $\pi \in R^k$, $\gamma_1 \in R^p$, and $\xi_1 \in R^p$ are unknown parameters. The matrices X and $\widetilde{Z}$ are taken to be fixed (i.e., non-stochastic) and $[X : \widetilde{Z}]$ has full column rank $p+k$. The $n \times 2$ matrix of errors $[u{:}v_2]$ is assumed to be iid across rows with each row having a mean zero bivariate normal distribution with nonsingular covariance matrix.

Our interest is in testing the null hypothesis

$$H_0 : \beta = \beta_0. \tag{2.2}$$

The alternative hypothesis of interest may be one-sided, $H_1 : \beta > \beta_0$ or $H_1 : \beta < \beta_0$, or two-sided, $H_1 : \beta \neq \beta_0$.

First, we rewrite the reduced-form equation in such a way that inference on β can be rendered free of the nuisance parameters (γ_1, ξ_1). The idea is to transform the IV matrix $\widetilde{Z}$ so that the transformed IV matrix Z and the exogenous regressor matrix X are orthogonal. We write

$$\begin{aligned} y_2 &= Z\pi + X\xi + v_2, \text{ where} \\ Z &= M_X\widetilde{Z},\ M_X = I_n - P_X,\ P_X = X(X'X)^{-1}X', \text{ and} \\ \xi &= \xi_1 + (X'X)^{-1}X'\widetilde{Z}\pi. \end{aligned} \tag{2.3}$$

Note that $Z'X = 0$.

Next, we consider the two reduced-form equations that correspond to the structural equation in (2.1) and the reduced-form equation in (2.3). In particular, substitution of the latter into the former gives

$$\begin{aligned} y_1 &= Z\pi\beta + X\gamma + v_1 \\ y_2 &= Z\pi + X\xi + v_2, \text{ where} \\ \gamma &= \gamma_1 + \xi\beta \text{ and } v_1 = u + v_2\beta. \end{aligned} \tag{2.4}$$

The reduced-form errors $[v_1{:}v_2]$ are iid across rows with each row having a mean zero bivariate normal distribution with 2×2 nonsingular covariance matrix Ω. For the purposes of obtaining exact optimal tests, the model we study is the two equation reduced-form model given in (2.4) with *known* nonsingular covariance matrix Ω. As shown below, asymptotically valid tests can be obtained by replacing Ω by an estimator when Ω is unknown.

The two equation reduced-form model can be written in matrix notation as

$$\begin{aligned} Y &= Z\pi a' + X\eta + V, \text{ where} \\ Y &= [y_1 : y_2],\ V = [v_1 : v_2], \\ a &= (\beta, 1)', \text{ and } \eta = [\gamma : \xi]. \end{aligned} \tag{2.5}$$

The distribution of $Y \in R^{n\times 2}$ is multivariate normal with mean matrix $Z\pi a' + X\eta$, independence across rows, and covariance matrix Ω for each row. The parameter space for $\theta = (\beta, \pi', \gamma', \xi')'$ is taken to be $R \times R^k \times R^p \times R^p$.

Because the multivariate normal is a member of the exponential family of distributions, low dimensional sufficient statistics are available for the parameter θ and the sub-vector $(\beta, \pi')'$:

Lemma 1 *For the model in* (2.5),
(a) $Z'Y$ *and* $X'Y$ *are sufficient statistics for* θ,
(b) $Z'Y$ *and* $X'Y$ *are independent,*
(c) $X'Y$ *has a multivariate normal distribution that does not depend on* $(\beta, \pi')'$,
(c) $Z'Y$ *has a multivariate normal distribution that does not depend on* $\eta = [\gamma : \xi]$, *and*
(d) $Z'Y$ *is a sufficient statistic for* $(\beta, \pi')'$.

Our interest is in tests of the null hypothesis $H_0 : \beta = \beta_0$. In consequence, there is no loss (in terms of attainable power functions) in considering tests that are based on the sufficient statistic $Z'Y$ for $(\beta, \pi')'$. Note that the nuisance parameters $\eta = [\gamma : \xi]$ are eliminated from the problem when one considers tests based on $Z'Y$. The nuisance parameter π remains.

As shown in Moreira (2003), the $k \times 2$ sufficient statistic $Z'Y$ can be simplified without loss of information by applying a one-to-one transformation that yields (i) the first transformed column to be independent of the nuisance parameter π under the null, (ii) independence of the two transformed columns (under the null and the alternative), and (iii) independence across rows in each column (under the null and the alternative). Condition (i) is achieved by using a linear combination of the columns of Y that has zero mean when $\beta = \beta_0$. Condition (ii) is achieved by taking the second transformed column of $Z'Y$ to be a linear combination of the columns of $Z'Y$ that is uncorrelated with the first transformed column. Condition (iii) is achieved by rotating each of the transformed columns so that their covariance matrices equal I_k. In particular, we consider[3]

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

The means of S and T depend on the following quantities:

$$\begin{aligned} \mu_\pi &= (Z'Z)^{1/2} \pi \in R^k, \\ c_\beta &= (\beta - \beta_0) \cdot (b_0' \Omega b_0)^{-1/2} \in R, \text{ and} \\ d_\beta &= a' \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2} \in R, \text{ where} \\ a &= (\beta, 1)'. \end{aligned} \tag{2.7}$$

The distributions of the sufficient statistics S and T for the parameters (β, π) are given in the following lemma.

Lemma 2 *For the model in* (2.5),
(a) $S \sim N(c_\beta \mu_\pi, I_k)$,
(b) $T \sim N(d_\beta \mu_\pi, I_k)$, *and*
(c) S *and* T *are independent.*

[3] The statistics S and T are denoted $\overline{S}$ and $\overline{T}$, respectively, in Moreira (2003).

Comments: 1. The results of the lemma hold under H_0 and H_1. Under H_0, S has mean zero.

2. The statistic T can be written as $d_{\beta_0}(Z'Z)^{1/2}\widehat{\pi}_0$, where $\widehat{\pi}_0$ denotes the maximum likelihood estimator of π under H_0.

3. Independence of S and T can be established by showing that S and T are jointly multivariate normal with zero covariance. An alternative proof is by applying Basu's Theorem, e.g., see Lehmann (1986, Thm. 5.2, p. 191). Basu's Theorem says that S and T are independent because the distribution of S does not depend on π and T is a boundedly complete sufficient statistic for π.

4. The constant d_β that appears in the mean of T can be rewritten as

$$\begin{aligned} d_\beta &= b'\Omega b_0 \cdot (b_0'\Omega b_0)^{-1/2}(\det(\Omega))^{-1/2}, \text{ where} \\ b &= (1, -\beta)'. \end{aligned} \tag{2.8}$$

This holds because some algebra shows that

$$\begin{aligned} a_0'\Omega^{-1}a_0 &= b_0'\Omega b_0/\det(\Omega) \text{ and} \\ a'\Omega^{-1}a_0 &= b'\Omega b_0/\det(\Omega). \end{aligned} \tag{2.9}$$

Using (2.8), some calculations show that d_{β_0} is proportional to the variance of the structural equation error u_i when $\beta = \beta_0$.

3 Invariant Similar Tests

The sufficient statistics S and T are independent multivariate normal k-vectors with spherical covariance matrices. The coordinate system used to specify the vectors should not affect inference based on them. In consequence, it is reasonable to restrict attention to coordinate-free functions of S and T. That is, we consider statistics that are invariant to rotations of the coordinate system. We note that Hillier (1984) and Chamberlain (2003) consider similar invariance conditions.

We consider the following groups of transformations on the data matrix $[S:T]$ and correspondingly on the parameters (β, π):

$$\begin{aligned} G &= \{g_F : g_F(x) = Fx \text{ for } x \in R^{k\times 2} \text{ for some } k \times k \text{ orthogonal matrix } F\} \text{ and} \\ \overline{G} &= \{\overline{g}_F : \overline{g}_F(\beta, \pi) = (\beta,\ (Z'Z)^{-1/2}F'(Z'Z)^{1/2}\pi) \text{ for some } k \times k \text{ orthogonal} \\ &\qquad \text{matrix } F\}. \end{aligned} \tag{3.1}$$

The transformations are one-to-one and are such that if $[S{:}T]$ has a distribution with parameters (β, π), then $g_F([S:T])$ has distribution with parameters $\overline{g}_F(\beta, \pi)$, as in Lehmann (1986, p. 283). Furthermore, the problem of testing $H_0 : \beta = \beta_0$ versus the alternative hypothesis H_1 (for any of the alternative hypotheses H_1 considered above) remains invariant under each transformation $g_F \in G$ because H_0 and H_1 are preserved under $\overline{g}_F$ (i.e., $\overline{g}_F(\beta, \pi)$ is in H_j if and only if (β, π) is in H_j for $j = 0, 1$).

An *invariant* test, $\phi(S, T)$, under the group G is one for which $\phi(FS, FT) = \phi(S, T)$ for all $k \times k$ orthogonal matrices F. By definition, a *maximal invariant* is a function of $[S{:}T]$ that is invariant and takes different values on different *orbits* of G.[4] Every invariant test can be written as a function of a maximal invariant, see Thm. 6.1 of Lehmann (1986, p. 285). Hence, it suffices to restrict attention to the class of tests that depend only on a maximal invariant.

Let

$$Q = [S{:}T]'[S{:}T] = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix} = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} \text{ and}$$
$$Q_1 = \left(S'S, S'T\right)' = (Q_S, Q_{ST})'. \tag{3.2}$$

The subscript 1 on Q_1 reflects the fact that Q_1 is the first column of Q.

For convenience, we use Q and (Q_1, Q_T) interchangeably. For example, if we define a function $h(Q)$, then $h(Q_1, Q_T)$ is presumed to be defined such that $h(Q_1, Q_T) = h(Q)$. Although this involves some abuse of notation, it is justified by the one-to-one transformation from Q to (Q_1, Q_T).

Theorem 1 *The 2×2 matrix Q is a maximal invariant for the transformations G.*

Comments: 1. Equivalently, (Q_1, Q_T) is a maximal invariant.

2. The statistic Q is invariant to nonsingular linear transformations of the instruments. Thus, invariance under the transformation group G ensures that tests of $H_0 : \beta = \beta_0$ will be unaffected, for example, by changing the units of Z or by respecifying binary units as contrasts.

3. By definition, the statistic Q has a non-central Wishart distribution because $[S{:}T]$ is a multivariate normal matrix that has independent rows and common covariance matrix across rows. The distribution of Q depends on π only through the scalar $\lambda \geq 0$ defined by

$$\lambda = \pi' Z' Z \pi. \tag{3.3}$$

This occurs for the same reason that a noncentral chi-squared distribution only depends on the mean vector through its length. In consequence, the utilization of invariance has reduced the k-vector nuisance parameter π to a scalar nuisance parameter λ. This is true both under the null and under the alternative.

4. Examples of invariant tests in the literature include the AR test; the standard LR and Wald tests, which use conventional, i.e., strong IV asymptotic, critical values; the LM test of Kleibergen (2002) and Moreira (2001); and the CLR and CW tests of Moreira (2003), which depend on the standard LR and Wald test statistics coupled with "conditional" critical values that depends on Q_T (where for each of the previous tests an estimator of the unknown Ω matrix that appears in the test statistic is replaced by the known matrix Ω because Ω is assumed to be known here). The AR,

[4] An orbit of G is an equivalence class of $k \times 2$ matrices, where $x_1 \sim x_2 \pmod G$ if there exists an orthogonal matrix F such that $x_2 = F x_1$.

LM, and LR test statistics depend on Q or (S, T) in the following ways:

$$\begin{aligned} AR &= Q_S \ = S'S, \\ LM &= Q_{ST}^2/Q_T \ = (S'T)^2/T'T, \text{ and} \\ LR &= \frac{1}{2}\left(Q_S - Q_T + \sqrt{(Q_S+Q_T)^2 - 4(Q_S Q_T - Q_{ST}^2)}\right). \end{aligned} \tag{3.4}$$

The Wald test statistic is a more complicated function of Q. For brevity, we do not give it. The only tests in the IV literature that we are aware of that are not invariant to G are tests that involve preliminary decisions to include or exclude a specific instrument, cf., Donald and Newey (2001).

A test based on the maximal invariant Q is *similar* if its null rejection rate does not depend on π. The parameter π determines the strength of the instrumental variables Z. The finite sample performance of some invariant tests, such as a t test based on the two-stage least squares estimator, varies greatly with π. In consequence, such tests often exhibit substantial size distortion when conventional (strong IV) asymptotic critical values are employed. By definition, invariant similar tests do not suffer from this problem. For this reason, it is important to characterize the class of invariant similar tests. We do so by adding one simple step to the argument Moreira (2001) used to characterize the class of similar tests.

Let the $[0, 1]$-valued statistic $\phi(Q)$ denote a (possibly randomized) test that depends on the maximal invariant Q.

Invariant similar tests are characterized as follows:

Theorem 2 *An invariant test $\phi(Q)$ is similar with significance level α if and only if $E_{\beta_0}(\phi(Q)|Q_T = q_T) = \alpha$ for almost all q_T, where $E_{\beta_0}(\cdot|Q_T = q_T)$ denotes conditional expectation given $Q_T = q_T$ when $\beta = \beta_0$ (which does not depend on π).*

Comments. 1. The Theorem suggests that a method of determining an invariant test with optimal power properties is to find an optimal invariant test conditional on $Q_T = q_T$ for each $q_T > 0$.

2. The AR and LM statistics are invariant statistics whose distributions under the null are independent of Q_T (by Lemma 3(f) below). Hence, the AR and LM tests that reject the null when the corresponding test statistics exceed given constants are invariant similar tests by Theorem 2. (This is not a new result.)

3. The LR and Wald statistics are invariant statistics whose distributions under the null depend on Q_T. Hence, the standard LR and Wald tests that use conventional (strong IV asymptotic) critical values are not invariant similar tests. To obtain similar tests based on the LR and Wald statistics, one must use critical values that depend on Q_T, as in Moreira (2003). The CLR test rejects the null hypothesis when

$$LR > \kappa_{CLR}(Q_T), \tag{3.5}$$

where $\kappa_{CLR}(Q_T)$ is defined to satisfy $P_{\beta_0}(LR > \kappa_{CLR}(Q_T)|Q_T = q_T) = \alpha$ and the conditional distribution of Q_1 given Q_T is specified in Lemma 3(c) below. See Table I of Moreira (2003) for critical values for the CLR test (where his τ corresponds to our q_T). A GAUSS program for p-values of the CLR test is given in AMS-04b. Similarly, the critical value function for the conditional Wald test, $\kappa_{CW}(Q_T)$, depends on Q_T.

4. Theorem 2 states that invariant tests are similar if and only if they have *Neyman structure* with respect to Q_T (e.g., as defined in Lehmann (1986, pp. 141-2)).

5. The proof of Theorem 2 is succinct, so we provide it here. Sufficiency follows immediately from the law of iterated expectations. Necessity uses the fact that S is ancillary under H_0 and the family of distributions of T under H_0 is a k-parameter exponential family indexed by π with parameter space that contains a k-dimensional rectangle. In consequence, T is a complete sufficient statistic for π under H_0 by Thm. 4.1 of Lehmann (1986, p. 142). The statistic Q_T is complete under H_0 because a function of a complete statistic is complete by the definition of completeness. (This is the step added to Moreira's (2001) argument.) In consequence, any function of Q_T whose expectation does not depend on π is equal to a constant with Q_T probability one. In particular, for a invariant similar test $\phi(Q)$, $E_{\beta_0}(\phi(Q)|Q_T)$ is a function of Q_T whose expectation equals α for all π. Hence, by completeness of Q_T, $E_{\beta_0}(\phi(Q)|Q_T = q_T)$ must equal α for almost all q_T. Note that $E_{\beta_0}(\phi(Q)|Q_T)$ does not depend on π by Lemma 3(c) below.

We now introduce a new invariant similar test for the null hypothesis $H_0 : \beta = \beta_0$ and the one-sided alternative $H_1 : \beta > \beta_0$. (The adjustment for $H_1 : \beta < \beta_0$ is straightforward.) The test is the one-sided version of the CLR test, which we refer to as the *one-sided CLR test* and denote by $CLR1$. The test statistic is based on the standard LR statistic (i.e., -2 times the logarithm of the likelihood ratio) for these hypotheses for the model of (2.4) with Ω known. We denote this test statistic by $LR1$.

Define

$$R(\beta) = \frac{b'Y'P_ZYb}{b'\Omega b}, \text{ where } P_Z = Z(Z'Z)^{-1}Z'. \tag{3.6}$$

The LIML-k estimator (i.e., the LIML estimator for the model with known covariance matrix), denoted $\widehat{\beta}_{LIML-k}$, minimizes $R(\beta)$ over $\beta \in R$. An expression for $\widehat{\beta}_{LIML-k}$ is given in the Appendix. As shown in the Appendix, we have

$$\begin{aligned} LR1 &= R(\beta_0) - \inf_{\beta \geq \beta_0} R(\beta) \\ &= \begin{cases} LR & \text{if } \widehat{\beta}_{LIML-k} \geq \beta_0 \\ 0 & \text{if } \widehat{\beta}_{LIML-k} < \beta_0 \text{ and } R(\beta_0) \leq R(\infty) \\ R(\beta_0) - R(\infty) & \text{if } \widehat{\beta}_{LIML-k} < \beta_0 \text{ and } R(\beta_0) > R(\infty), \end{cases} \end{aligned} \tag{3.7}$$

where $R(\infty) = \lim_{\beta \to \infty} R(\beta)$ and, hence, $R(\infty)$ equals $R(\beta)$ with b replaced by $(0, -1)'$. (For $H_1 : \beta < \beta_0$, $R(\infty)$ is replaced by $R(-\infty)$, which equals $R(\beta)$ with b replaced by $(0, 1)'$.) In the Appendix, we show that $R(\beta)$ and $LR1$ depend on the observations only through Q.

The $CLR1$ test rejects the null hypothesis when the statistic $LR1$ exceeds a critical value that depends on Q_T and is such that the null rejection probability is the desired value $1 - \alpha$, as in (3.5). Such critical values can be determined by simulation.

4 Optimal Tests for Weighted Average Power

4.1 Weighted Average Power

The AR, LM, and CLR tests are invariant similar tests and, hence, have good size properties even under weak IV's. These tests are somewhat ad hoc, however, in the sense that they have no known optimal power properties under weak IV's except in the just-identified case, i.e., when $k = 1$. In this case, the AR, LM, and LR tests are equivalent tests and Moreira (2001) shows that this test is uniformly most powerful unbiased for two-sided alternatives.

In this section, we address the question of optimal invariant similar tests when the IV's may be weak. We determine the invariant similar test that has maximum weighted average power (WAP) with respect to (wrt) a given weight function W over the parameter values in the alternative. The two motivations for considering WAP tests are that (i) such tests yield one- and two-sided power envelopes, which are important for evaluating the performance of any test, and (ii) such tests provide a class of tests with the potential for good overall power properties. The use of sufficiency and invariance reduces the dimension of the alternative parameters that need to be considered from $1 + k + 2p$ for $\theta = (\beta, \pi', \xi', \gamma')'$ to just 2 for $(\beta, \lambda)'$. In consequence, it is relatively easy to specify weight functions W of interest.

Let $W(\beta, \lambda)$ be a probability distribution function on $R \times R^+$. Weighted average power of a test $\phi(Q)$ with respect to W is given by the Lebesgue integral

$$K(\phi, W) = \int E_{\beta,\lambda}\phi(Q)dW(\beta, \lambda), \tag{4.1}$$

where $E_{\beta,\lambda}$ denotes expectation when the true parameters are $(\beta, \lambda)'$.

Let

$$g_W(q_1, q_T) = \int_{R\times R^+} f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)dW(\beta, \lambda), \tag{4.2}$$

where $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ denotes the joint density of (Q_1, Q_T) at (q_1, q_T). Let $q_1 = (q_S, q_{ST})'$. WAP can be written as power against the single density $g_W(q_1, q_T)$:

$$\begin{aligned} & K(\phi, W) \\ = & \int_{R\times R^+} \left[\int_{R^+\times R\times R^+} \phi(q_S, q_{ST}, q_T) f_{Q_1,Q_T}(q_S, q_{ST}, q_T; \beta, \lambda) dq_S dq_{ST} dq_T \right] dW(\beta, \lambda) \\ = & \int_{R^+\times R\times R^+} \phi(q_1, q_T) g_W(q_1, q_T) dq_1 dq_T \end{aligned} \tag{4.3}$$

using the Tonelli-Fubini Theorem, e.g., see Dudley (1989, Thm. 4.4.5, p. 104).

For example, suppose the weight function W corresponds to a point mass distribution at (β^*, λ^*). That is,

$$W_{\beta^*,\lambda^*}(\beta,\lambda) = \begin{cases} 1 & \text{if } (\beta,\lambda) \geq (\beta^*,\lambda^*) \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Then, the test that maximizes WAP among invariant similar tests with significance level α is the *point-optimal invariant* (POI) similar test of level α against (β^*, λ^*).

Most existing tests in the literature are two-sided tests. Examples include the tests in (3.4). To obtain optimal two-sided tests one can specify W to give weight to β values both less than and greater than β_0, see Section 6 below.

4.2 Optimal Invariant Similar Tests for Weighted Average Power

We want to find a test that maximizes WAP for weight function W among all level α invariant similar tests. By Theorem 2, invariant similar tests must be similar conditional on $Q_T = q_T$ for almost all q_T. In addition, by (4.3), WAP for weight function W equals unconditional power against the single density $g_W(q_1, q_T)$. In turn, the latter equals expected conditional power given Q_T. Hence, it suffices to determine the test that maximizes conditional power given $Q_T = q_T$ among tests that are invariant and are similar conditional on $Q_T = q_T$, for each q_T.

Conditional power given $Q_T = q_T$ is

$$K(\phi, W|Q_T = q_T) = \int_{R^+\times R} \phi(q_1, q_T) g_W(q_1|q_T) dq_1, \tag{4.5}$$

where $g_W(q_1|q_T)$ denotes the conditional density at q_1 of Q_1 given $Q_T = q_T$. We have

$$g_W(q_1|q_T) = \frac{g_W(q_1, q_T)}{g_W(q_T)}, \tag{4.6}$$

where $g_W(q_1, q_T)$ is defined in (4.2),

$$\begin{aligned} g_W(q_T) &= \int_{R^+\times R} g_W(q_1, q_T) dq_1 \\ &= \int_{R\times R^+} \int_{R^+\times R} f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) dq_1 dW(\beta, \lambda) \\ &= \int_{R\times R^+} f_{Q_T}(q_T; \beta, \lambda) dW(\beta, \lambda), \end{aligned} \tag{4.7}$$

and $f_{Q_T}(q_T; \beta, \lambda)$ denotes the density of Q_T at q_T.

Next, we consider the conditional density of Q_1 given $Q_T = q_T$ under the null hypothesis. Because Q_T is a sufficient statistic for π under H_0, this conditional density does not depend on π or λ. Hence, we denote the conditional density of Q_1 given $Q_T = q_T$ under the null hypothesis by $f_{Q_1|Q_T}(q_1|q_T; \beta_0)$.

For any invariant test $\phi(Q_1, Q_T)$, conditional on $Q_T = q_T$, the null hypothesis is simple because $f_{Q_1|Q_T}(q_1|q_T; \beta_0)$ does not depend on π or λ. Given the WAP criterion function $K(\phi, W)$, the alternative hypothesis of concern also is simple. In particular, conditional on $Q_T = q_T$, the alternative density of interest is $g_W(q_1|q_T)$. In consequence, 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. In particular, the conditional WAP-LR test statistic is

$$LR_W(Q_1, q_T) = \frac{g_W(Q_1|q_T)}{f_{Q_1|Q_T}(Q_1|q_T;\beta_0)} = \frac{g_W(Q_1, q_T)}{g_W(q_T) f_{Q_1|Q_T}(Q_1|q_T;\beta_0)}. \tag{4.8}$$

In order to provide an explicit expression for $LR_W(Q_1, Q_T)$, we now determine the densities $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$, $f_{Q_T}(q_T; \beta, \lambda)$, and $f_{Q_1|Q_T}(q_1|q_T; \beta_0)$ that arise in (4.2), (4.7), and (4.8). These densities and the tests considered below depend on the following quantity:

$$\begin{aligned} \xi_\beta(q) &= h_\beta' q h_\beta \\ &= c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T, \text{ where} \\ h_\beta &= (c_\beta, d_\beta)'. \end{aligned} \tag{4.9}$$

Note that $\xi_\beta(q) \geq 0$ because q is positive semi-definite a.s.

Lemma 3 (a) *The density of* (Q_1, Q_T) *is*

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

where $q_1 = (q_S, q_{ST})' \in R^+ \times R$, $q_T \in R^+$, $q = \begin{bmatrix} q_S & q_{ST} \\ q_{ST} & q_T \end{bmatrix}$,

$$K_1^{-1} = 2^{(k+2)/2} pi^{1/2} \Gamma((k-1)/2),$$

$I_\nu(\cdot)$ *denotes the modified Bessel function of the first kind of order* ν, $pi = 3.1415...$, *and* $\Gamma(\cdot)$ *is the gamma function.*

(b) *The density of* Q_T *is a non-central chi-squared density with* k *degrees of freedom and noncentrality parameter* $d_\beta^2 \lambda$:

$$f_{Q_T}(q_T; \beta, \lambda) = K_2 \exp\left(-\lambda d_\beta^2/2\right) q_T^{(k-2)/2} \exp\left(-q_T/2\right)$$

$$\times \left(\lambda d_\beta^2 q_T\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda d_\beta^2 q_T}\right)$$

for $q_T > 0$, *where* $K_2^{-1} = 2$.

(c) *Under the null hypothesis, the conditional density of* Q_1 *given* $Q_T = q_T$ *is*

$$f_{Q_1|Q_T}(q_1, q_T; \beta_0) = K_1 K_2^{-1} \exp(-q_S/2) \det(q)^{(k-3)/2} q_T^{-(k-2)/2}.$$

(d) *Under the null hypothesis, the density of* Q_S *is a* (*central*) *chi-squared density with* k *degrees of freedom*:

$$f_{Q_S}(q_S) = K_3 q_S^{(k-2)/2} \exp(-q_S/2)$$

for $q_S > 0$, *where* $K_3^{-1} = 2^{k/2}\Gamma(k/2)$.

(e) *Under the null hypothesis, the density of* $\mathcal{S}_2 = Q_{ST}/(||S|| \cdot ||T||)$ *at* s_2 *is*

$$f_{\mathcal{S}_2}(s_2) = K_4(1 - s_2^2)^{(k-3)/2}$$

for $s_2 \in [-1, 1]$, *where* $K_4^{-1} = pi^{1/2}\Gamma((k-1)/2)/\Gamma(k/2)$.

(f) *Under the null hypothesis,* Q_S, $\mathcal{S}_2$, *and* T *are mutually independent and, hence,* Q_S, $\mathcal{S}_2$, *and* Q_T *also are mutually independent.*

Comments: 1. The joint density $f_{Q_1,Q_T}(q_S, q_T; \beta, \lambda)$ given in part (a) of the Lemma is a noncentral Wishart density.[5] The null density of $\mathcal{S}_2$ given in part (e) of the Lemma is the same as that of the sample correlation coefficient from an iid sample of k observations from a bivariate normal distribution with means zero and covariance matrix I_2 when the means of the random variables are not estimated.

2. Parts (d)-(f) of the Lemma are used below to simplify the calculation of critical values for optimal WAP tests.

3. The modified Bessel function of the first kind that appears in the densities in parts (a) and (b) of the Lemma is defined by

$$I_\nu(x) = (x/2)^\nu \sum_{j=0}^{\infty} \frac{(x^2/4)^j}{j!\Gamma(\nu + j + 1)}, \tag{4.10}$$

for $x \geq 0$, e.g., see Lebedev (1965, p. 108). Sometimes the function $I_\nu(x)$ is referred to as a Bessel function of the first kind *with imaginary argument*. For $|x|$ small, $I_\nu(x) \sim (x/2)^\nu/\Gamma(\nu + 1)$; for $|x|$ large, $I_\nu(x) \sim e^x/\sqrt{2pi \cdot x}$; and for $\nu \geq 0$ (which holds in the expression for $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ whenever $k \geq 2$), $I_\nu(\cdot)$ is monotonically increasing on R^+, see Lebedev (1965, p. 136). Expressions for $I_\nu(x)$ in terms of elementary functions are available whenever ν is a half-integer (which corresponds to k being an odd integer). For example, $I_{-1/2}(x) = (2/pi)^{1/2}(\exp(x) + \exp(-x))/2$ (which arises when $k = 1$) and $I_{1/2}(x) = (2/pi)^{1/2}(\exp(x) - \exp(-x))/2$ (which arises when $k = 3$).

4. Both GAUSS and Matlab have built-in functions for computing the modified Bessel function of the first kind. These functions are extremely fast. Hence, the density $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ can be computed very quickly.

[5] In Johnson and Kotz (1970, 1972), a standard reference for probability densities, the formulae for the noncentral Wishart and chi-squared distributions in terms of $I_{(k-2)/2}(\cdot)$ contain several typographical errors. Hence, the densities in Lemma 3(a) and (b) are based on Anderson (1946, eqn. (6)) and are not consistent with those of Johnson and Kotz (1970, eqn. (5), p. 133; 1972, eqn. (50), p. 176). Sawa (1969, footnote 6) notes that Anderson's (1946) eqn. (6) contains a slight error in that the covariance matrix Σ is missing in one place in the formula. This does not affect our use of Anderson's formula, however, because we apply it with $\Sigma = I_k$.

5. Independence of $\mathcal{S}_2$ and Q_T under H_0 can be established directly using the spherical symmetry of the distribution of $\mathcal{S}_2$. Or, it can be established using (i) the bounded completeness of Q_T for λ pointed out in Comment 4 to Theorem 2, (ii) the fact that the distribution of $\mathcal{S}_2$ does not depend on λ by part (e) of the Lemma, and (iii) Basu's Theorem (e.g., see Lehmann (1986, p. 191)).

Equations (4.2), (4.7), and (4.8) and Lemma 3 combine to give the following result.

Corollary 1 *The optimal* WAP *test statistic for weight function W is*

$$LR_W(q_1, q_T) = \frac{\int f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) dW(\beta, \lambda)}{\int f_{Q_T}(q_T; \beta, \lambda) dW(\beta, \lambda) f_{Q_1|Q_T}(q_1|q_T; \beta_0, \lambda)} = \frac{\psi_W(q_1, q_T)}{\psi_{2,W}(q_T)},$$

where

$$\psi_W(q_1, q_T) = \int \exp(-\lambda(c_\beta^2 + d_\beta^2)/2)(\lambda\xi_\beta(q))^{-(k-2)/4} I_{\frac{k-2}{2}}\left(\sqrt{\lambda\xi_\beta(q)}\right) dW(\beta, \lambda),$$

$$\psi_{2,W}(q_T) = \int \exp\left(-\lambda d_\beta^2/2\right)\left(\lambda d_\beta^2 q_T\right)^{-(k-2)/4} I_{\frac{k-2}{2}}\left(\sqrt{\lambda d_\beta^2 q_T}\right) dW(\beta, \lambda),$$

the integrals are over $(\beta, \lambda) \in R \times R^+$, and c_β, d_β, and $\xi_\beta(q)$ are defined in (2.7) *and Lemma* 3(a)*.*

Comment: Note that $\psi_W(q_1, q_T)$ does **not** equal $\int f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) dW(\beta, \lambda)$ and likewise with $\psi_{2,W}(q_T)$. This is because numerous cancellations occur in the second expression in the first line of the Corollary 1, including the constants K_1-K_4 (because $K_1 = K_2 K_3 K_4$) and the terms that depend on q_1 in the denominator.

Because $\psi_{2,W}(q_T)$ does not depend on q_1, it could be absorbed into the conditional critical value given $Q_T = q_T$. Thus, the test based on $LR_W(q_1, q_T)$ is equivalent to a test based on $\psi_W(q_1, q_T)$. For reasons of numerical stability, however, we recommend constructing critical values using $\ln(LR_W(q_1, q_T))$.

Computation of the integrands of $\psi_W(q_1, q_T)$ and $\psi_{2,W}(q_T)$ in Corollary 1 are easy and extremely fast using GAUSS or Matlab functions for computing the modified Bessel function of the first kind. Hence, calculation of the test statistic $LR_W(Q_1, Q_T)$ is very fast unless the weight function W is ill-behaved. Of course, ill-behaved weight functions can be avoided because the user selects the weight function.

The test that maximizes WAP among invariant similar tests with significance level α rejects H_0 if

$$LR_W(Q_1, Q_T) > \kappa_\alpha(Q_T), \tag{4.11}$$

where $\kappa_\alpha(Q_T)$ is defined such that the test is similar. That is, $\kappa_\alpha(q_T)$ is defined by

$$P_{\beta_0}(LR_W(Q_1, q_T) > \kappa_\alpha(q_T)|Q_T = q_T) = \alpha, \tag{4.12}$$

where $P_{\beta_0}(\cdot|Q_T = q_T)$ denotes conditional probability given $Q_T = q_T$ under the null, which can be calculated using the density in Lemma 3(c). Note that $\kappa_\alpha(\cdot)$ does not depend on Ω, Z, X, or the sample size n.

By Lemma 3(d)-(f), under H_0, (i) Q_S, $\mathcal{S}_2 = Q_{ST}/(||S|| \cdot ||T||)$, and Q_T are independent, (ii) $Q_S \sim \chi^2_k$, and (iii) $\mathcal{S}_2$ has density $f_{\mathcal{S}_2}$. The null distribution of $(Q_S, \mathcal{S}_2)$ can be simulated by simulating $S \sim N(0, I_k)$ and taking $(Q_S, \mathcal{S}_2) = (S'S, S'e_1/||S||)$ for $e_1 = (1, 0, ..., 0)' \in R^k$. Hence, the null distribution of $Q_1 = (S'S, S'T)$ conditional on $Q_T = q_T$ can be simulated easily and quickly by simulating $S \sim N(0, I_k)$ and taking $Q_1 = (S'S, S'e_1 \cdot q_T)$.

The critical value $\kappa_\alpha(Q_T)$ can be approximated by simulating n_{MC} iid random vectors $S_i \sim N(0, I_k)$ for $i = 1, ..., n_{MC}$, where n_{MC} is large, computing $Q_1(i) = (S_i'S_i, S_i'e_1 \cdot Q_T)$ for $i = 1, ..., n_{MC}$, and taking $\ln(\kappa_\alpha(Q_T))$ to be the $1 - \alpha$ sample quantile of $\{\ln(LR_W(Q_1(i), Q_T)) : i = 1, ..., n_{MC}\}$. The p-value for the test based on $LR_W(Q_1, Q_T)$ can be approximated by the fraction of values in $\{\ln(LR_W(Q_1(i), Q_T)) : i = 1, ..., n_{MC}\}$ that exceed the actual value of the statistic computed using the original sample Y.

The following theorem summarizes the results of this section:

Theorem 3 *The test that rejects H_0 when $LR_W(Q_1, Q_T) > \kappa_\alpha(Q_T)$ maximizes WAP for the weight function W over all level α invariant similar tests.*

Comment: The optimal WAP test statistic $LR_W(Q_1, Q_T)$ depends on $S'S$, $S'T$, and $T'T$ in general. In contrast, the AR statistic depends only on $S'S$ and the LM statistic depends on $S'T$ and $T'T$, but not on $S'S$. Hence, power improvements from optimal WAP tests compared to these two tests can be attributed to optimal exploitation of information about β that is contained in all three statistics $S'S$, $S'T$, and $T'T$.

5 Point Optimal Invariant Similar Tests and the One-sided Power Envelope

In this section, we determine the one-sided power envelope for invariant similar tests by considering the point optimal invariant similar (POIS) tests for arbitrary values (β^*, λ^*). In particular, we show that such tests do not depend on λ^*, so that the POIS tests are of a relatively simple form.

Using the definition of $I_\nu(x)$ in (4.10), $\psi_W(q_1, q_T)$ can be written as

$$\begin{aligned} \psi_W(q_1, q_T) &= 2^{-(k-2)/2} \int \exp(-\lambda(c_\beta^2 + d_\beta^2)/2) \sum_{j=0}^{\infty} \frac{(\lambda \xi_\beta(q_1, q_T)/4)^j}{j!\Gamma((k-2)/2 + j + 1)} dW(\beta, \lambda) \\ &= 2^{-(k-2)/2} \sum_{j=0}^{\infty} \frac{\int \exp(-\lambda(c_\beta^2 + d_\beta^2)/2)(\lambda \xi_\beta(q_1, q_T)/4)^j dW(\beta, \lambda)}{j!\Gamma((k-2)/2 + j + 1)}. \quad (5.13) \end{aligned}$$

Obviously, $\psi_{2,W}(q_T)$ can be written analogously.

The integrand in the first line of (5.13) is increasing in $\xi_\beta(q_1, q_T)$ because $\xi_\beta(q_1, q_T) \geq 0$. In consequence, for a fixed value of β, say β^* $(\neq \beta_0)$, the test that rejects H_0 when $\xi_{\beta^*}(Q_1, Q_T)$ is large maximizes weighted average power for all weight functions over λ values. That is, the optimal test for fixed alternative β^* rejects H_0 when

$$\begin{aligned} \xi_{\beta^*}(Q_1, Q_T) &> \kappa_{\beta^*,\alpha}(Q_T), \text{ where} \\ P_{\beta_0}(\xi_{\beta^*}(Q_1, q_T) &> \kappa_{\beta^*,\alpha}(q_T)|Q_T = q_T) = \alpha \end{aligned} \tag{5.14}$$

for all q_T. This test is a *one-sided* test because it directs power at a single point β^* that is either greater than or less than the null value β_0.

Corollary 2 *The level α 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 $\{(\beta^*, \lambda) : \lambda > 0\}$.*

Comments: 1. The one-sided power envelope for invariant similar tests is given by the test in (5.14) by varying the value β^*. Although the form of the test in (5.14) does not depend on λ^*, its power depends on the true value of λ. Hence, the power envelope depends on both parameters β and λ.

2. The test based on $\xi_{\beta^*}(Q_1, Q_T)$ is equivalent to a test based on $Q_S + 2(d_{\beta^*}/c_{\beta^*})Q_{ST}$. Hence, the test statistic is a linear combination of Q_S and Q_{ST}. The coefficients of the linear combination depend on Ω, β_0, and β^*.

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

$$\begin{aligned} POIS1_\delta &= \frac{Q_S + \delta \mathcal{S}_2 \sqrt{Q_S} - k}{\sqrt{2k + \delta^2}} > \overline{\kappa}_{\delta,\alpha}(Q_T), \text{ where} \\ \delta &= (2d_{\beta^*}/c_{\beta^*})\sqrt{Q_T} \text{ and} \\ P_{\beta_0}(POIS1_\delta &> \overline{\kappa}_{\delta,\alpha}(Q_T)|Q_T = q_T) = \alpha. \end{aligned} \tag{5.15}$$

This formulation of the test is convenient because Q_S, $\mathcal{S}_2$, and Q_T are independent under H_0 by Lemma 3(f), which simplifies calculation of critical values.

4. Corollary 2 shows that no UMPI one-sided test exists because $\xi_{\beta^*}(Q_1, Q_T)$ depends on β^*.

5. The quantity d_{β^*} is a linear function of β^* and equals zero iff $\beta^* = \beta_{AR}$, where

$$\beta_{AR} = \frac{\omega_{11} - \omega_{12}\beta_0}{\omega_{12} - \omega_{22}\beta_0} \tag{5.16}$$

and ω_{ij} denotes the (i, j) element of Ω (provided $\omega_{12} - \omega_{22}\beta_0 \neq 0$). In this case, $\delta = 0$ and $POIS1_0$ reduces to $Q_S/\sqrt{2k}$, which is the AR statistic rescaled. Hence, the AR test is one-sided POIS against the alternative $\beta = \beta_{AR}$.

6. The sign of δ in (5.15) 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^*. For example, if $\beta_0 = 0$ and $\omega_{12} > 0$, then $\beta_{AR} = \omega_{22}/\omega_{12} > 0$, $sgn(\delta) = sgn(d_{\beta^*}) > 0$ for $0 < \beta^* < \beta_{AR}$, and $sgn(\delta) = sgn(d_{\beta^*}) < 0$ for $\beta^* > \beta_{AR} > 0$. In consequence, the form of the statistic $POIS_\delta$ 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 has adverse consequences for the overall one-sided power properties of POIS1 tests, see Section 8 below.

7. The optimal one-sided test for β^* local to β_0 with $\beta^* > \beta_0$ and arbitrary weight functions over λ values (i.e., the LMPI test) is the one-sided LM test that rejects H_0 if

$$Q_{ST}/Q_T^{1/2} > \kappa_{\phi,\alpha}, \tag{5.17}$$

where $\kappa_{\phi,\alpha}$ 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 $-Q_{ST}/Q_T^{1/2} > \kappa_{\phi,\alpha}$. (See the Appendix for the proof.)

8. The power function of the one-sided LM test exhibits unusual behavior as β changes from values less than β_{AR} to values greater than β_{AR}. By Lemma 2, the conditional distribution of $Q_{ST}/Q_T^{1/2}$ given T is $N((c_\beta(T_0 + d_\beta\mu_\pi)/||T_0 + d_\beta\mu_\pi||)'\mu_\pi, 1)$, where $T = T_0 + d_\beta\mu_\pi$ and $T_0 \sim N(0, I_k)$. Hence, $Q_{ST}/Q_T^{1/2}$ has a mixed normal distribution. For simplicity, consider the case $\beta_0 = 0$. Then, for $\beta > 0$, c_β is positive and linearly increasing in β. But, d_β is proportional to $\omega_{11} - \omega_{12}\beta$, is linearly decreasing in β when $\omega_{12} > 0$, and is negative for $\beta > \beta_{AR}$ when $\omega_{12} > 0$. In this case, the mean of the mixing distribution switches sign at β_{AR} and this has a dramatic effect on the power of the one-sided LM test. For large enough values of β its power drops to zero because $((T_0 + d_\beta\mu_\pi)/||T_0 + d_\beta\mu_\pi||)'\mu_\pi$ is negative with probability close to one. When $\beta_0 = 0$ and $\omega_{12} < 0$, the one-sided LM test against positive β values does not exhibit this unusual behavior, but the one-sided LM test against negative β values does. The non-monotonic behavior of the one-sided LM test also arises when $\beta_0 \neq 0$.

9. The optimal one-sided test for β^* arbitrarily large and any weight function over λ values is of the form: reject H_0 if

$$Q_S + 2(\det(\Omega))^{-1/2}(\beta_0\omega_{22} - \omega_{12})Q_{ST} > \kappa_{\infty,\alpha}(Q_T) \tag{5.18}$$

for some $\kappa_{\infty,\alpha}(\cdot)$, where ω_{ij} denotes the (i,j) element of Ω. The same test is the optimal one-sided test for β^* negative and arbitrarily large in absolute value for any weight functions over λ. In consequence, the optimal two-sided test for $|\beta^* - \beta_0|$ arbitrarily large and any weight function over λ values is the test in (5.18).

For the common case where the null hypothesis specifies that $\beta_0 = 0$, the optimal test for $|\beta^* - \beta_0|$ large rejects H_0 if

$$Q_S - 2\frac{\rho}{(1-\rho^2)^{1/2}}Q_{ST} > \kappa_{\infty,\alpha}(Q_T), \tag{5.19}$$

where $\rho = \omega_{12}/(\omega_{11}\omega_{22})^{1/2}$ is the correlation between the errors v_1 and v_2 in (2.4). (See the Appendix for the proof.)

6 Two-Sided Tests and Power Envelope

In this section, we discuss tests and the power envelope for the two-sided alternative hypothesis $H_1 : \beta \neq \beta_0$. As described in the following four subsections, there are several methods of constructing WAP tests and the power envelope for two-sided alternatives. The first method we consider is simple, but is found to have significant drawbacks and, hence, is not recommended. The second through fourth methods are recommended and are found to yield closely related results.

6.1 Symmetric-Alternative WAP Tests and Power Envelope

First, we consider invariant similar tests that maximize WAP for a weight function W that depends on β only through $|\beta - \beta_0|$. We call such tests optimal WAP tests for *symmetric alternatives.* A corresponding power envelope can be constructed by considering two-point weight functions for the points $(\beta_0 - \delta, \lambda)$ and $(\beta_0 + \delta, \lambda)$ for $\delta > 0$ and $\lambda > 0$.

Although simple, weight functions for symmetric alternatives have some serious drawbacks. These drawbacks stem from the fact that the underlying testing problem is not symmetric for the parameter vectors $(\beta_0 - \delta, \lambda)$ and $(\beta_0 + \delta, \lambda)$. The distribution of Q_T is noncentral χ_k^2 with non-centrality parameter $d_\beta^2 \lambda$, see Lemma 3(b). This noncentrality parameter takes on different values for the parameter vectors $(\beta_0 - \delta, \lambda)$ and $(\beta_0 + \delta, \lambda)$.[6] In consequence, the problems of testing against these two alternative parameter vectors are not equally difficult testing problems. This has undesirable consequences for the power of WAP tests for symmetric alternatives under strong IV asymptotics. In particular, calculations in Section 11 below show that such tests are not asymptotically efficient under strong IV asymptotics according to the usual criterion for asymptotic efficiency of two-sided tests in regular models.[7] Given this, we do not recommend WAP tests for symmetric alternatives.

The power envelope generated by two-point symmetric alternative weight functions does not represent a proper *two-sided* power envelope because the point optimal tests that generate the envelope are asymptotically one-sided tests under strong instrument asymptotics, see Section 11. Hence, we do not consider this power envelope any further.

6.2 Asymptotically Efficient WAP Tests and Power Envelope

We are interested in tests that have good all-around two-sided power properties. This includes high power when the IV's are strong. The appropriate benchmark power envelope for such tests is a power envelope based on two-point optimal invariant similar (POIS2) tests that are asymptotically efficient under strong IV asymptotics. We consider this power envelope here.

We consider "two-point weight functions," which are distribution functions that

[6] This is true except in the special case in which $\beta_0 = \omega_{12}/\omega_{22}$, where ω_{12} is the off-diagonal element of Ω and ω_{22} is the $(2, 2)$ element of Ω.

[7] The usual criterion is that of Wald (1943), who considers weighted average power over certain ellipses in the parameter space. Lack of asymptotic efficiency for a test does not mean that the test is asymptotically inadmissible under strong IV asymptotics. Rather, it means that the test does not possess the standard two-sided asymptotic optimality properties that LR, LM, and Wald tests possess in regular models.

place equal mass on two points (β^*, λ^*) and (β_2^*, λ_2^*). They are of the form

$$W_{2P}(\beta, \lambda) = \frac{1}{2}1(\beta \geq \beta^*, \lambda \geq \lambda^*) + \frac{1}{2}1(\beta \geq \beta_2^*, \lambda \geq \lambda_2^*), \tag{6.1}$$

where β^* and β_2^* lie on opposite sides of β_0. Different types of two-point weight functions arise depending on how (β_2^*, λ_2^*) is selected given (β^*, λ^*).

Here, we consider two-point weight functions that have a number of desirable features. First, as shown in Section 11 below, they are the only two-point weight functions that lead to POIS2 tests that are *asymptotically efficient* under strong IV asymptotics. Second, they lead to POIS2 tests that have the same power against each of the two points. Third, given (β^*, λ^*), the second point (β_2^*, λ_2^*) satisfies: (i) β_2^* is on the other side of the null value β_0 from β^*, (ii) the marginal distributions of Q_S, $Q_{ST,}$ and Q_T under (β_2^*, λ_2^*) are the same as under (β^*, λ^*), (iii) the joint distribution of $(Q_S, Q_{ST,} Q_T)$ under (β_2^*, λ_2^*) equals that of $(Q_S, -Q_{ST,} Q_T)$ under (β^*, λ^*), which corresponds to β_2^* being on the other side of the null from β^*, and (iv) the distribution of $[-S : T]$ under (β_2^*, λ_2^*) equals that of $[S : T]$ under (β^*, λ^*).

Given (β^*, λ^*), the point (β_2^*, λ_2^*) that has these properties solves

$$(\lambda_2^*)^{1/2} c_{\beta_2^*} = -(\lambda^*)^{1/2} c_{\beta^*} \; (\neq 0) \text{ and } (\lambda_2^*)^{1/2} d_{\beta_2^*} = (\lambda^*)^{1/2} d_{\beta^*}. \tag{6.2}$$

This follows from Lemma 2, Lemma 3(a), and $\lambda = \mu_\pi' \mu_\pi$. Note that c_β is proportional to $\beta - \beta_0$ and d_β is linear in β. Some calculations show that provided $\beta^* \neq \beta_{AR}$, the solution to the two equations in (6.2) are

$$\begin{aligned} \beta_2^* &= \beta_0 - \frac{d_{\beta_0}(\beta^* - \beta_0)}{d_{\beta_0} + 2g(\beta^* - \beta_0)} \text{ and} \\ \lambda_2^* &= \lambda^* \frac{(d_{\beta_0} + 2g(\beta^* - \beta_0))^2}{d_{\beta_0}^2}, \text{ where} \\ g &= e_1' \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2} \text{ and } e_1 = (1, 0)'. \end{aligned} \tag{6.3}$$

(If $\beta^* = \beta_{AR}$, then there is no solution to (6.2) with β_2^* on the other side of the null from β^*.)

We refer to the power envelope based on POIS2 tests with W as in (6.1) and (β_2^*, λ_2^*) as in (6.3) as the *asymptotically efficient two-sided* (AE-2S) power envelope.

Next, we consider WAP tests that are designed to have good all-around two-sided power properties. We consider a class of weight functions that generalize the AE two-point weight functions of (6.1) and (6.3) to more than two points. These weight functions deliver WAP tests that are asymptotically efficient under strong IV asymptotics, see Section 11 below, and, hence, are called AE weight functions. These weight functions are of the form

$$W_{AE}(\beta, \lambda) = \frac{1}{2} W_*(\beta, \lambda) + \frac{1}{2} W_*(\beta_2, \lambda_2), \tag{6.4}$$

where (a) W_* is a distribution function with finite support and (b) given (β, λ), (β_2, λ_2) is defined as (β_2^*, λ_2^*) is defined in (6.3) but with (β, λ) in place of (β^*, λ^*).

Such weight functions place equal weight on (β,λ) and (β_2,λ_2). The parameter vector (β_2,λ_2) is the appropriate "other-sided" parameter vector to (β,λ) in the sense described in (i)-(iv) above. We call tests based on weight functions of the form (6.4) *asymptotically efficient* (AE) WAP tests.

We refer to $LR_{W_{AE}}(q_1,q_T)$ as an AE-WAP test statistic. It can be written conveniently without explicit dependence on (β_2,λ_2) as follows:

$$\begin{aligned}\psi_{W_{AE}}(q_1,q_T) = &\frac{1}{2}\int \exp(-\lambda(c_\beta^2+d_\beta^2)/2)(\lambda\xi_\beta(q))^{-(k-2)/4} I_{\frac{k-2}{2}}\left(\sqrt{\lambda\xi_\beta(q)}\right) dW_*(\beta,\lambda)\\ &+\frac{1}{2}\int \exp(-\lambda(c_\beta^2+d_\beta^2)/2)(\lambda\xi_\beta^*(q))^{-(k-2)/4} I_{\frac{k-2}{2}}\left(\sqrt{\lambda\xi_\beta^*(q)}\right) dW_*(\beta,\lambda),\end{aligned} \tag{6.5}$$

and likewise for $\psi_{2,W_{AE}}(q_T)$, where

$$\xi_\beta^*(q) = c_\beta^2 q_S - 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T. \tag{6.6}$$

This holds because the equations in (6.2) imply that $\lambda_2(c_{\beta_2}^2+d_{\beta_2}^2)=\lambda(c_\beta^2+d_\beta^2)$ and $\lambda_2\xi_{\beta_2}(q)=\lambda\xi_\beta^*(q)$.

Note that the POIS2 test statistics $LR_{W_{2P}}(q_1,q_T)$ with (β_2^*,λ_2^*) as in (6.3), which are used to construct the AE-2S power envelope, can be written as in (6.5) with $W_*(\beta,\lambda)=(1/2)1(\beta\geq\beta^*,\lambda\geq\lambda^*)$. The dependence of such tests on (β^*,λ^*) shows that no UMPI two-sided test exists.

6.3 Sign Invariant WAP Test and Power Envelope

Next, we consider tests that satisfy an additional invariance condition to that in (3.1):

$$[S:T]\rightarrow[-S:T]. \tag{6.7}$$

The corresponding transformation in the parameter space is $(\beta,\lambda)\rightarrow(\beta_2,\lambda_2)$, where (β_2,λ_2) is defined in (b) following (6.4). This sign invariance condition is a natural condition to impose to obtain two-sided tests because the parameter vector (β_2,λ_2) is the appropriate "other-sided" parameter vector to (β,λ) for the reasons stated in the first paragraph of the previous subsection. The maximal invariant under this sign invariance condition (plus the invariance conditions in (3.1)) is

$$(S'S,|S'T|,T'T)=(Q_S,|Q_{ST}|,Q_T). \tag{6.8}$$

The AR, LM, and LR test statistics all depend on the data only through this maximal invariant and, hence, satisfy the sign invariance condition (6.7).

The density of the maximal invariant $(Q_S,|Q_{ST}|,Q_T)$ at (q_S,q_{ST},q_T) for $q_{ST}\geq 0$ is given by

$$\frac{1}{2}f_{Q_1,Q_T}(q_S,q_{ST},q_T)+\frac{1}{2}f_{Q_1,Q_T}(q_S,-q_{ST},q_T), \tag{6.9}$$

where Lemma 3 provides an expression for $f_{Q_1,Q_T}(q_S, q_{ST}, q_T)$. Hence, following the same argument as in Section 4.2, given a weight function $W_*(\beta,\lambda)$, the optimal WAP test statistic, call it $LR^*_{W_*}(q_1, q_T)$, can be shown to satisfy[8]

$$LR^*_{W_*}(q_1, q_T) = LR_{W_{AE}}(q_1, q_T). \tag{6.10}$$

Thus, the class of WAP tests that are invariant to (3.1) and (6.7) and that have weight functions W_* equals the class of WAP tests that are invariant to (3.1) and that have weight functions W_{AE}. Furthermore, the power envelope for the class of invariant similar tests under the invariance conditions of (3.1) and (6.7) equals the AE-2S power envelope.

6.4 Locally-Unbiased WAP Tests

A fourth approach to constructing tests and a power envelope designed for two-sided alternatives is to impose an unbiasedness or a local (to the null) unbiasedness condition. This approach has a long tradition in the statistics literature and is a standard way to derive optimal tests for two-sided alternatives. In exponential families, UMP two-sided tests exist among the class of unbiased tests, see Lehmann (1986, Thm. 4.3, p. 147). This is not the case in the curved exponential family testing problem considered here. Nevertheless, one can develop optimal WAP tests among the class of locally unbiased (LU) invariant tests.

We start by determining two necessary conditions for an invariant test (under the invariance condition of (3.1)) to be unbiased. The first condition is similarity and the second condition is *local unbiasedness.* Local unbiasedness requires that the power function has zero derivative at the null hypothesis. Otherwise, the power function would dip below the size of the test for some alternatives close to the null. We show that the AR, LM, and CLR tests are LU.

Next, we determine the test that maximizes WAP, as defined in (4.1), among the class of LU invariant similar tests. We do so using the same argument as in Section 4.2, but using the generalized Neyman-Pearson Lemma (see Lehmann (1986, Thm. 3.5, pp. 96-7)) in place of the Neyman-Pearson Lemma. The form of the optimal WAP test statistic is the same as in Section 4.2, only the critical value function differs.

Theorem 4 *An invariant test $\phi(Q)$ is unbiased with significance level α only if $E_{\beta_0}(\phi(Q)|Q_T = q_T) = \alpha$ and $E_{\beta_0}(\phi(Q)Q_{ST}|Q_T = q_T) = 0$ for almost all q_T.*

Comments. 1. The first condition establishes that all unbiased invariant tests must be similar. The second establishes that the power function must have zero derivative under H_0. The second condition is the *local* unbiasedness condition.

2. The two conditions in Theorem 4 are closely related to the conditions used for two-sided alternatives in the classical hypothesis testing theory for exponential families, see Lehmann (1986, Ch. 4).

[8] The proof of this relies on the fact that q_{ST} enters the densities only through q_{ST}^2 in each place except in the modified Bessel function. In consequence, the cancellations that occur in the middle expression of the first line of Corollary 1 still hold.

3. The second condition of Theorem 4 is equivalent to

$$E_{\beta_0,\lambda}(\phi(Q)Q_{ST}/Q_T^{1/2}) = 0 \text{ for all } \lambda \geq 0. \tag{6.11}$$

That is, any unbiased invariant test statistic $\phi(Q)$ must be uncorrelated with the pivotal statistic $Q_{ST}/Q_T^{1/2}$ under H_0.[9]

The AR, LM, and LR test statistics depend on the data through (Q_S, Q_{ST}^2, Q_T). The following result shows that these tests satisfy the second condition of Theorem 4.

Corollary 3 *Any similar level α test that depends on the observations through (Q_S, Q_{ST}^2, Q_T) satisfies the local unbiasedness condition of Theorem* 4.

Comment. Corollary 3 shows that the class of AE-2S invariant similar tests considered in Section 6.2 is contained in the class of LU invariant similar tests considered in this section.

The next result uses local unbiasedness to specify an optimal WAP test for two-sided alternatives.

Theorem 5 *The test that maximizes WAP among LU invariant similar tests with significance level α rejects H_0 if*

$$LR_W(Q_1, Q_T) > \kappa_{1\alpha}(Q_T) + Q_{ST}\kappa_{2\alpha}(Q_T),$$

where $\kappa_{1\alpha}(Q_T)$ and $\kappa_{2\alpha}(Q_T)$ are chosen such that the two conditions in Theorem 4 *hold.*

Comment. *Point optimal* LU invariant similar tests are obtained by taking the weight function W to give point mass at a given alternative parameter (β, λ) of interest. The power of these tests maps out the power envelope for locally-unbiased invariant similar (LUIS) tests, which we refer to as the LUIS power envelope.

7 Point Optimal Invariant Non-similar Tests

7.1 One-sided Alternatives

Non-similar tests have null rejection probability below the significance level for some values of the nuisance parameter, in this case, λ. Due to the continuity of the power function, for such values of λ, the power of a non-similar test will be 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 $\beta = \beta_0$ against a point alternative.

[9] The second condition of Theorem 4 clearly implies (6.11). The converse holds by the completeness of Q_T because by iterated expectations the left-hand side in (6.11) can be written as $E_{\beta_0,\lambda}h(Q_T)$, where $h(Q_T) = E_{\beta_0}(\phi(Q)Q_{ST}|Q_T = q_T)/Q_T^{1/2}$.

Our construction of POI non-similar tests follows Lehmann (1997, Sec. 3.8). Consider the composite null hypothesis

$$H_0 : (\beta, \lambda) \in \{(\beta_0, \lambda) : 0 \leq \lambda < \infty\}, \tag{7.1}$$

and the point alternative

$$H_1 : (\beta, \lambda) = (\beta^*, \lambda^*). \tag{7.2}$$

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

$$h_\Lambda(q) = \int f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) d\Lambda(\lambda), \tag{7.3}$$

where $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ is given in Lemma 3(a). 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

$$NP_\Lambda(q) = \frac{f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)}{h_\Lambda(q)} > d_{\Lambda,\alpha}, \tag{7.4}$$

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 (7.1), i.e.,

$$\sup_{0 \leq \lambda < \infty} P_{\beta,\lambda}(NP_\Lambda(Q) > d_{\Lambda,\alpha}) = \alpha, \tag{7.5}$$

then the test ϕ_Λ is most powerful for testing H_0 against H_1, and the distribution Λ is least favorable; cf. Lehmann (1986, Sec. 3.8, Thm. 7, and Cor. 5).

Given a distribution Λ, condition (7.5) is easily checked numerically. What proves more computationally difficult, however, is finding the distribution that satisfies (7.5). In the numerical work we consider distributions Λ that put point mass on some point λ_0. In this case, we have

$$\begin{aligned} NP_\Lambda &= \frac{f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)}{f_{Q_1,Q_T}(q_1, q_T; \beta_0, \lambda_0)} \\ &= \frac{\exp(-\lambda^*(c_{\beta^*}^2 + d_{\beta^*}^2)/2) \left(\lambda^* \xi_{\beta^*}(q)\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda^* \xi_{\beta^*}(q)}\right)}{\exp(-\lambda_0 d_{\beta_0}^2/2) \left(\lambda_0 \xi_{\beta_0}(q)\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda_0 \xi_{\beta_0}(q)}\right)}, \end{aligned} \tag{7.6}$$

where the second equality follows from Lemma 3(a).

Let $R(\beta_0, \lambda_0, \beta^*, \lambda^* | \beta, \lambda)$ be the rejection rate of the test based on the statistic given by (7.6) 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

$$R(\beta_0, \lambda_0^{LF}, \beta^*, \lambda^* | \beta_0, \lambda_0^{LF}) = \alpha \text{ and}$$
$$\sup_{0 \leq \lambda < \infty} R(\beta_0, \lambda_0^{LF}, \beta^*, \lambda^* | \beta_0, \lambda) \leq \alpha. \tag{7.7}$$

If there is a $\lambda_0^{LF}(\beta_0, \beta^*, \lambda^*)$ that satisfies (7.7), then the test based on $NP_{\lambda_0^{LF}}$ is the POI non-similar test. The power envelope for invariant non-similar tests is $R(\beta_0, \lambda_0^{LF}(\beta_0, \beta^*, \lambda^*), \beta^*, \lambda^* | \beta^*, \lambda^*)$.

7.2 Two-sided Alternatives

Next, we consider invariant non-similar tests that have maximum WAP for the two-point AE weight functions defined in (6.1) and that satisfy (6.2). The optimal test is of the same form as in (7.4), but with the numerator replaced by $\int f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) dW_{2P}(\beta, \lambda)$. The numerator can be written in terms of (β^*, λ^*) alone (and not (β_2^*, λ_2^*)) by the same argument as used to get (6.5).

As in the one-sided case, we consider distributions Λ that place point mass on some point λ_0, so that the test statistic is

$$NP_\Lambda = \exp(-\lambda^*(c_{\beta^*}^2 + d_{\beta^*}^2)/2) \left[\left(\lambda^* \xi_{\beta^*}(q)\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda^* \xi_{\beta^*}(q)}\right) + \right.$$
$$\left. \left(\lambda^* \xi_{\beta^*}^*(q)\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda^* \xi_{\beta^*}^*(q)}\right) \right] \times$$
$$\left[\exp(-\lambda_0 d_{\beta_0}^2/2) \left(\lambda_0 d_{\beta_0}^2 q_T\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda_0 d_{\beta_0}^2 q_T}\right) \right]^{-1}, \tag{7.8}$$

where $\xi_\beta^*(q)$ is defined in (6.6). Algebra and the strong-IV asymptotic results of Section 11 reveal that if a one point least favorable distribution exists it must satisfy

$$\lambda_0^{LF} \geq \underline{\lambda}_0, \text{ where } \underline{\lambda}_0 = \lambda_1 d_{\beta_1}^2 / d_{\beta_0}^2. \tag{7.9}$$

Otherwise, the critical value of $NP_{\Lambda^{LF}}$ is unbounded. The numerical problem is to find $\lambda_0 = \lambda_0^{LF}$ that solves (7.7) and the additional restriction (7.9).

8 Numerical Results I: Model with Known Covariance Matrix

This section reports numerical results for power envelopes and comparative powers of tests developed in Sections 4-7 for the case of known Ω and normal errors. The model considered is given in (2.4) with Ω specified by $\omega_{11} = \omega_{22} = 1$ and $\omega_{12} = \rho$.[10] Without loss of generality, no X matrix is included. The parameters characterizing

[10] There is no loss of generality in taking $\omega_{11} = \omega_{22} = 1$ because the distribution of the maximal invariant Q under $(\widetilde{\beta}, \widetilde{\pi}, \widetilde{\Omega})$ for arbitrary pd $\widetilde{\Omega}$ with elements $\widetilde{\omega}_{jk}$ equals its distribution under (β, π, Ω), where $\omega_{11} = \omega_{22} = 1$, $\beta = (\widetilde{\omega}_{22}/\widetilde{\omega}_{11})^{1/2}\widetilde{\beta}$, and $\pi = \widetilde{\omega}_{22}^{-1/2}\widetilde{\pi}$.

the distribution of the tests are λ ($= \pi' Z' Z \pi$), the number of IV's k, the correlation between the reduced form errors ρ, and the parameter β. Throughout, we focus on tests with significance level 5% and on the case where the null value is $\beta_0 = 0$.[11] Numerical results have been computed for $\lambda/k = 0.5$, 1, 2, 4, 8, 16, which span the range from weak to strong instruments, $\rho = 0.95$, 0.50, and 0.20, and $k = 2$, 5, 10, 20. To conserve space, we report only a subset of these results here. The full set of results is available in AMS-04.

Conditional critical values for the $POIS1_\delta$ test and the (two-sided) CLR test were computed by numerical integration based on the distributional results in Lemma 3. Conditional critical values for other statistics were computed by Monte Carlo simulation on a grid of 150 values of q_T, yielding lookup tables that were then interpolated. Least favorable distributions were approximated using a single-point distribution. All results reported here are based on 5,000 Monte Carlo simulations. Details of the numerical methods are given in the supplement to this paper, AMS-04.

The results 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 rescaled alternative $(\beta - \beta_0)\lambda^{1/2}$. These can be thought of as local power plots, where the local neighborhood is $1/\lambda^{1/2}$ instead of the usual $1/n^{1/2}$, since λ measures the effective sample size.

8.1 One-sided tests

Figure 1 presents the power envelope for invariant similar (conditional) tests (POIS1) based on Corollary 2, the power envelope for invariant non-similar (unconditional) tests based on Section 7.1, and the power function of the CLR1 test of the one-sided hypothesis $H_1 : \beta > 0$. The single-point approximation of the least favorable distribution was found to work well in the sense that (7.5) is satisfied within Monte Carlo accuracy.

Inspection of Figure 1 reveals three salient findings. First, the power envelopes for the similar and non-similar tests are essentially the same up to numerical accuracy. This is true not just for the values reported here but for all values of λ, β, ρ, and k considered. The reason for this is twofold. On one hand, the conditional critical values for the POIS1 tests depend on q_T only weakly in the range of q_T that is most likely to occur under the alternative. Thus, the POIS1 tests are very nearly unconditional. On the other hand, the POI non-similar tests have null rejection rates that are very nearly equal to 5% 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.

Second, there is a curious blip in some power envelopes. This blip occurs at the value of the alternative for which $\beta = 1/\rho$. This is the value β_{AR} defined in (5.16).

[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 $\widetilde{\beta} = \beta - \beta_0$.

The blip occurs because the sign of δ changes as β^* changes from values less than $1/\rho$ to values greater than $1/\rho$.

Third, the CLR1 test has power that is close to the power envelope for local and distant alternatives, but deviates from the power envelope for alternatives near the point $\beta = \beta_{AR}$. For smaller values of ρ and larger values of λ/k, the power of the CLR1 test is closer to the power envelope than in Figure 1, see AMS-04.

One approach to testing in the absence of a UMPI test is to consider POI tests that have power functions tangent to the power envelope at a certain value; cf. King (1988). Accordingly, Figure 2 graphs power functions of various POIS1 tests along with the invariant similar power envelope for $\lambda = 5$ and $\lambda = 20$. The individual power functions plotted in Figure 2 are for (i) the local-to-β_0 POIS1 test given by (5.17), (ii) the most distant POIS1 test given by (5.19), and (iii) several tests with intermediate points of tangency (at powers of approximately 0.25, 0.5, and 0.75). Clearly, the power functions of the POIS1 tests are not monotonic. This is a consequence of the flip in sign of δ in the POIS1 test statistic. For $\beta < \beta_{AR}$, POIS1 tests have higher power with positive δ, but for $\beta > \beta_{AR}$ POIS1 tests have higher power with negative δ. Hence, tests designed for $\beta < \beta_{AR}$ perform poorly when $\beta > \beta_{AR}$ and vice versa. In consequence, no single POIS1 test provides good overall performance.

Experiments with optimal WAP tests for various one-sided weight functions lead to some tests whose power was similar to, but not better than, the power of the CLR1 test, see AMS-04. Hence, the best test in terms of overall one-sided power that we found is the CLR1 test. Given that there is some difference between the power of the CLR1 test and the one-sided power envelope, it may be possible to find a one-sided test that performs better, but we were not able to do so. On the other hand, no UMP one-sided test exists, so the CLR1 test may be as good a test as possible in an overall sense.

8.2 Two-sided tests

Figure 3 presents power envelopes for the asymptotically efficient (AE) and locally unbiased (LU) families of invariant similar tests. There is no discernible difference between these two power envelopes. Hence, the power envelopes obtained from two quite different criteria for imposing two-sidedness, viz., local-unbiasedness and two-sided asymptotic efficiency under strong IV asymptotics, yield essentially the same power envelope. This is a notable and very convenient result. The AE invariant similar tests are more tractable numerically than the LU invariant similar tests, so we focus henceforth on the AE family.

Figure 3 also presents the power envelope for the AE family of invariant non-similar tests in the range in which we were able to compute it. According to our computations, if (7.9) is satisfied then there exists a one-point least favorable distribution. This occurs in an interval containing β_0 where the end points of the interval are defined by $\lambda_0^{LF} = \underline{\lambda}_0$. Within this interval, λ_0^{LF} is very close to $\underline{\lambda}_0$—within .02 in all the cases plotted in Figure 3. Outside this interval, we were unable to find a one-point least favorable distribution. Where we could compute it, the power envelope for AE invariant non-similar tests equals that for AE invariant similar tests, within numerical accuracy.[12]

Figure 4 presents the power functions of two POIS2 tests, along with the power envelope for AE invariant similar tests. The first POIS2 tests is point-optimal against $\beta^* = 0.8$ and $\lambda^* = 5$ (for $k = 5$), and the second is point-optimal against $\beta^* = 1.45$ and $\lambda^* = 5$. We denote these as POIS2(.8, 5) and POIS2(1.45, 5), respectively. The POIS2(.8, 5) and POIS2(1.45, 5) tests have power functions that are tangent to the power envelope at powers of approximately 25% and 75% for $\lambda = 5$. Note that these tests depend on ρ but not on the unknown values of β or λ and thus are feasible tests if Ω is known. Unlike the case for the one-sided POIS1 tests, the power functions of the POIS2 tests effectively lie on the two-sided AE power envelope. Evidently, both these tests are numerically nearly UMP among AE invariant similar tests.

Figure 5 plots the power functions of the two-sided CLR, LM, and AR tests, along with the power envelope for AE invariant similar tests. The striking new finding based on this work is that the power function of the CLR test effectively achieves the power envelope for AE invariant similar tests, even more closely than the POIS2(.8, 5) and POIS2(1.45, 5) tests. Figure 5 documents other results as well. The power function of the AR test is generally below the AE power envelope, except at its point of tangency at $\beta = \beta_{AR}$. Also, as is known from previous simulation work (e.g. Moreira (2003) and Stock, Wright, and Yogo (2002)), the power function of the LM statistic is not monotonic. This is due to the switching of the sign of d_β as β moves through the value β_{AR}. The adverse effect of the sign switching of d_β is considerably muted for the two-sided LM test compared to the one-sided LM test because, roughly speaking, power lost on one side of the null is partially picked up by power on the other side.

The numerical equality of the similar and non-similar AE invariant power envelopes and the good overall performance of the POIS2 tests suggests that an unconditional point optimal invariant non-similar test might also exhibit good power properties. We restrict attention to non-similar tests that are asymptotically efficient if the IV's are strong, which requires the α quantile of their null distribution (given λ) to be maximal in the limit $\lambda \to \infty$, which in turn cannot occur if $\lambda_0 < \underline{\lambda}_0$. We therefore consider tests of the form (7.8) with $\lambda_0 = \underline{\lambda}_0$, with upward sloping rejection functions for λ large, and, as a simplification, with $d^2_{\beta_1} = c^2_{\beta_1}$. This leads to test statistics of the form

$$P^* = \frac{\frac{1}{2}\exp(-\kappa/2)\left[\left(\kappa\overline{\xi}\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\kappa\overline{\xi}}\right) + \left(\kappa\overline{\xi}^*\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\kappa\overline{\xi}^*}\right)\right]}{\left(\kappa q_T\right)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\kappa q_T}\right)}, \tag{8.1}$$

where $\overline{\xi} = q_S + 2q_{ST} + q_T$, $\overline{\xi}^* = q_S - 2q_{ST} + q_T$, and $\kappa = \lambda_1 d^2_{\beta_1} = \lambda_1 c^2_{\beta_1} = \lambda_0 d^2_{\beta_0}$. Note that with these simplifications the parameters β_1, β_0, ρ, λ_1, and λ_0 enter only through κ.

Figure 5 plots the power function of the P^* test with $\kappa = 3.25$, which corresponds approximately to the POI non-similar test that is tangent to the AE non-similar

[12] We thank Anna Mikoucheva for research assistance in the computation of the AE invariant non-similar power envelopes.

power envelope at $\beta_1 = 0.792$ and $\lambda_1 = 5.18$ (the 5% critical value for this test is 3.37). The power of this test falls slightly below the power envelope near β_{AR}, but otherwise numerically achieves the power envelope.

In sum, the results of Figure 5 (and further results documented in AMS-04) show that the CLR test dominates the LM and AR tests and is, in a numerical sense, UMP among AE invariant similar tests and among locally-unbiased invariant similar tests. Also, the P^* test nearly achieves this power envelope and thus is, in the same sense, approximately UMPI.

Figure 6 shows how the power results change with k. Figure 6 gives the power envelopes for AE invariant similar tests and the power functions of the two-sided CLR, LM, AR, and P^* tests for $k = 2$ (Figure 6(a) and 6(b)) and for $k = 10$ (Figure 6(c) and 6(d)) (for $k = 2$, $\kappa = 2$ and the 5% critical value is 2.95 for the P^* test, and for $k = 10$, $\kappa = 4.25$ and the 5% critical value is 3.40). Three findings of these (and related results reported in AMS-04) are noteworthy. First, note that the scale is the same in Figure 6 as in Figure 5, and, aside from the location of the blip, the power envelopes are numerically close in each panel in the two figures. This confirms that the appropriate measure of information for optimal invariant testing is $\lambda^{1/2}$, and this scaling does not depend on k. In particular, this implies that the AE power envelope does not deteriorate significantly with the addition of an irrelevant instrument.

Second, the power of the CLR test is numerically essentially the same as the power envelope, confirming the finding above for $k = 5$ that the CLR test is nearly UMP among invariant similar tests of the AE family.

Third, for $k = 2$, the power function of the P^* test is effectively on the power envelope. For $k = 10$, the power of the P^* test drops slightly below the power envelope, suggesting that for values of $k > 5$ it would be of interest to investigate different unconditional POI tests as alternatives to the P^* test.

9 Weak IV Asymptotics for Case of Unknown Covariance Matrix and Non-normal Errors

In this section, we consider the same model and hypotheses as in Section 2, but with unknown error covariance matrix, (possibly) non-normal, heteroskedastic, and/or autocorrelated errors, and (possibly) random IV's and/or exogenous variables. The latter allows for lagged dependent and endogenous variables as regressors or IV's.

We use weak IV asymptotics, as in Staiger and Stock (1997), to analyze the properties of the procedures considered. We consider three versions of the finite sample tests introduced in Sections 4 and 6. The first version is suitable for the case of uncorrelated errors that exhibit *contemporaneous homoskedasticity.* By this we mean that $E(V_i V_i' | Z_i, X_i)$ is a constant matrix that does not depend on i, where V_i denotes the reduced-form error vector for the i-th observation (i.e., V_i is the i-th row of V written as a column 2-vector). In a time series setting this still allows for the errors to exhibit *temporal conditional heteroskedasticity* with respect to *lagged* values of the errors, IV's, and exogenous variables (i.e., $E(V_i V_i' | Z_{i-1}, X_{i-1}, V_{i-1}, Z_{i-2}, X_{i-2}, V_{i-2}, ...)$ may be random).

The second version of the tests that are introduced here is designed for uncorrelated errors that may exhibit contemporaneous heteroskedasticity (i.e., $E(V_iV_i'|Z_i, X_i)$ may be random or depend on i). This version adjusts the statistics (S, T) to obtain robustness to heteroskedasticity. Note that most procedures in the literature, including the AR, LM, CLR, and Staiger and Stock (1997) procedures, are not robust to heteroskedasticity.

The third version of the tests is designed to be robust to both contemporaneous heteroskedasticity and autocorrelation in the reduced-form errors.

For clarity of the asymptotics results, throughout this section we write S, T, Q_1, Q_T, Q_S, Q_{ST}, $\mathcal{S}_2$, λ, AR, LM, LR, and $LR1$ of Sections 2-8, as S_n, T_n, $Q_{1,n}$, $Q_{T,n}$, $Q_{S,n}$, $Q_{ST,n}$, $\mathcal{S}_{2,n}$, λ_n, AR_n, LM_n, LR_n, and $LR1_n$, respectively, where n is the sample size. All limits are taken as $n \to \infty$.

Let $\overline{Z} = [Z : X]$. Let Y_i, Z_i, X_i, $\overline{Z}_i$, and V_i denote the i-th rows of Y, Z, X, $\overline{Z}$, and V, respectively, written as column vectors of dimensions 2, k, p, $k+p$, and 2.

9.1 Assumptions

We use the following high-level assumptions concerning the IV's, exogenous variables, and errors. The assumptions are quite similar to those of Staiger and Stock (1997), but they allow for the possibility of heteroskedastic and autocorrelated errors because the form of the asymptotic variance matrix Φ in Assumption 3 is not restricted. The parameter π which determines the strength of the IV's is local to zero and the alternative parameter β is 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.

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

Assumption 2. $n^{-1}V'V \to_p \Omega$ for some pd 2×2 matrix Ω.

Assumption 3. $n^{-1/2}vec(\overline{Z}'V) \to_d N(0, \Phi)$ for some pd $2(k+p)\times 2(k+p)$ matrix Φ.

In Assumption 3, $vec(\cdot)$ denotes the column by column vec operator.

The quantities C, D, Ω, and Φ are assumed to be unknown.

Assumption WIV-FA is the "weak IV's with fixed alternative" assumption. Assumptions 1 and 2 hold under suitable conditions by a weak law of large numbers (WLLN), see below. Assumption 3 holds under suitable conditions by a central limit theorem (CLT). Assumptions 1-3 are consistent with non-normal, heteroskedastic, autocorrelated errors and IV's and regressors that may be random or non-random.

For example, Assumptions 1-3 are implied by any one of the following assumptions:

Assumption IID. $\{(V_i, \overline{Z}_i) : i \geq 1\}$ are iid, $E(V_i \otimes \overline{Z}_i) = 0$, $E||V_i||^2 + E||\overline{Z}_i||^2 + E||V_i \otimes \overline{Z}_i||^2 < \infty$, $\Omega = EV_iV_i'$ is pd, and $\Phi = E(V_i \otimes \overline{Z}_i)(V_i \otimes \overline{Z}_i)'$ is pd.

Assumption INID. $\{(V_i, \overline{Z}_i) : i \geq 1\}$ are independent, $E(V_i \otimes \overline{Z}_i) = 0$ for all $i \geq 1$, $\sup_{i \geq 1}(E||V_i||^{2+\delta} + E||\overline{Z}_i||^{2+\delta} + E||V_i \otimes \overline{Z}_i||^{2+\delta}) < \infty$ for some $\delta > 0$, $n^{-1}\sum_{i=1}^n EV_iV_i'$ $\to \Omega$ for some pd 2×2 matrix Ω, and $n^{-1}\sum_{i=1}^n E(V_i \otimes \overline{Z}_i)(V_i \otimes \overline{Z}_i)' \to \Phi$ for some pd $2(k+p) \times 2(k+p)$ matrix Φ.

Assumption MDS. $\{(V_i \otimes \overline{Z}_i, \mathcal{F}_i) : i \geq 1\}$ is a martingale difference sequence, where $\mathcal{F}_i = \sigma(V_i, \overline{Z}_i, V_{i-1}, \overline{Z}_{i-1}, ...)$, $\{(V_i, \overline{Z}_i) : i \geq 1\}$ is a stationary and ergodic sequence, $E||V_i||^2 + E||\overline{Z}_i||^2 + E||V_i \otimes \overline{Z}_i||^2 < \infty$, $\Omega = EV_iV_i'$ is pd, and $\Phi = E(V_i \otimes \overline{Z}_i)(V_i \otimes \overline{Z}_i)'$ is pd.

Assumption CORR. $\{(V_i, \overline{Z}_i) : i = ..., 0, 1, ...\}$ is a doubly infinite stationary and ergodic sequence with $E(V_i \otimes \overline{Z}_i) = 0$, $E||V_i||^2 + E||\overline{Z}_i||^2 + E||V_i \otimes \overline{Z}_i||^2 < \infty$, $\sum_{j=1}^{\infty}(E||E(V_i \otimes \overline{Z}_i|\mathcal{F}_{i-j})||^2)^{1/2} < \infty$, where $\mathcal{F}_i = \sigma(V_i, \overline{Z}_i, V_{i-1}, \overline{Z}_{i-1}, ...)$, $\Omega = EV_iV_i'$ is pd, and $\Phi = \sum_{j=-\infty}^{\infty} E(V_i \otimes \overline{Z}_i)(V_{i-j}' \otimes \overline{Z}_{i-j}')$ is pd.

The random vectors $\{V_i \otimes \overline{Z}_i : i \geq 1\}$ are uncorrelated under Assumption IID, INID, or MDS, but are (possibly) correlated under Assumption CORR.

If the errors are contemporaneously homoskedastic and $\{V_i \otimes \overline{Z}_i : i \geq 1\}$ are uncorrelated, the following key assumption holds. Under this assumption (and Assumptions WIV-FA and 1-3), the tests described in Sections 4 and 6 but with Ω replaced by a consistent estimator $\widehat{\Omega}_n$ have asymptotic significance level α, as desired.

Assumption 4. $\Phi = \Omega \otimes D$, where Φ is defined in Assumption 3.

In Section 9.2 below, we impose Assumption 4, but in Sections 9.3 and 9.4, we do not. Assumption 4 is implied by any one of Assumptions IID, INID, and MDS plus the following.

Assumption HOM. $E((V_iV_i') \otimes (\overline{Z}_i\overline{Z}_i')) = \Omega \otimes D$ for all $i \geq 1$.

By iterated expectations, a sufficient condition for Assumption HOM is $E(V_iV_i'|\overline{Z}_i) = EV_iV_i' = \Omega$ $a.s.$ for all $i \geq 1$.

Note that Assumptions MDS and CORR allow for intertemporal conditional heteroskedasticity even when Assumption HOM holds.

Lemma 4 (a) *Any one of Assumptions* IID, INID, MDS, *and* CORR *implies Assumptions* 1-3.

(b) *Any one of Assumptions* IID, INID, *and* MDS *plus Assumption* HOM *imply Assumption* 4.

The asymptotic results stated below hold for any true parameter values β, C, γ, ξ, and Ω, provided Ω is positive definite. Hence, we do not need to be specific regarding the parameter space. Of course, for the testing problem to be well defined, the parameter space should include the null value β_0 and at least one other value of β. In addition, for tests to exist that have non-trivial power, it is necessary for the parameter space to include at least one non-zero vector C.

We estimate Ω ($\in R^{2\times 2}$) (defined in Assumption 2) via

$$\widehat{\Omega}_n = (n-k-p)^{-1}\widehat{V}'\widehat{V}, \text{ where } \widehat{V} = Y - P_ZY - P_XY, \tag{9.1}$$

where k and p are the dimensions of Z_i and X_i, respectively.[13] Let $\widehat{V}_i$ denote the i-th row of $\widehat{V}$ written as a column 2-vector.

Under Assumptions 1-3, the variance estimator is consistent.

[13] 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 off.

Lemma 5 *Under Assumptions* 1-3, $\widehat{\Omega}_n \to_p \Omega$.

Comment. The convergence in the Lemma occurs uniformly over all true parameters β, C, γ, and ξ no matter what the parameter space is. This can be seen by inspection of the proof of the Lemma.

9.2 Homoskedastic Uncorrelated Errors

We now introduce tests that are suitable for (possibly) non-normal, homoskedastic, uncorrelated errors and unknown covariance matrix. That is, the tests are suitable when Assumptions 1-4 hold.

We define analogues of S_n, T_n, $Q_{1,n}$, and $Q_{T,n}$ that replace the unknown matrix Ω with $\widehat{\Omega}_n$:

$$\begin{aligned} \widehat{S}_n &= (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0' \widehat{\Omega}_n b_0)^{-1/2}, \\ \widehat{T}_n &= (Z'Z)^{-1/2} Z'Y \widehat{\Omega}_n^{-1} a_0 \cdot (a_0' \widehat{\Omega}_n^{-1} a_0)^{-1/2}, \\ \widehat{Q}_{1,n} &= \left(\widehat{Q}_{T,n}, \widehat{Q}_{ST,n} \right)' = \left(\widehat{S}_n' \widehat{S}_n, \widehat{S}_n' \widehat{T}_n \right)', \text{ and } \widehat{Q}_{T,n} = \widehat{T}_n' \widehat{T}_n. \end{aligned} \tag{9.2}$$

The AR, LM, LR, LR1, and POIS_δ test statistics for the case of unknown Ω are defined as in (3.4), (3.7), and (5.15), but with Q_S, Q_{ST}, and Q_T replaced by $\widehat{Q}_{S,n}$, $\widehat{Q}_{ST,n}$, and $\widehat{Q}_{T,n}$. Denote these test statistics by $\widehat{AR}_n$, $\widehat{LM}_n$, $\widehat{LR}_n$, $\widehat{LR1}_n$, and $\widehat{POIS1}_{\widehat{\delta}}$, respectively.

A homoskedastic optimal WAP test, referred to as an HOM-WAP test, rejects the null hypothesis $H_0 : \beta = \beta_0$ when

$$LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) > \kappa_\alpha(\widehat{Q}_{T,n}), \tag{9.3}$$

where $LR_W(\cdot,\cdot)$ is defined in Corollary 1 and $\kappa_\alpha(\cdot)$ is defined in (4.12) (and can be calculated by simulation using the method described there). An LU version of the HOM-WAP test is defined using the statistic $LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n})$ combined with the critical value given in Theorem 5 with Q_T and Q_{ST} replaced by $\widehat{Q}_{T,n}$ and $\widehat{Q}_{ST,n}$, respectively.

Next, we show that $\widehat{S}_n$ and $\widehat{T}_n$ converge in distribution to independent k-vectors S_∞ and T_∞, respectively, which are defined as follows. Let N_Z be a $k \times 2$ normal matrix. Let

$$\begin{aligned} vec(N_Z) &\sim N(vec(D_Z C a'), \Omega_0 \otimes D_Z), \\ S_\infty &= D_Z^{-1/2} N_Z b_0 \cdot (b_0 \Omega b_0)^{-1/2} \sim N(c_\beta D_Z^{1/2} C, I_k), \\ T_\infty &= D_Z^{-1/2} N_Z \Omega^{-1} a_0 \cdot (a_0 \Omega^{-1} a_0)^{-1/2} \sim N(d_\beta D_Z^{1/2} C, I_k), \text{ where} \\ D_Z &= D_{11} - D_{12} D_{22}^{-1} D_{21}, \\ D &= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \text{ and } D_{j\ell} \in R^{k \times k} \text{ for } j, \ell = 1, 2. \end{aligned} \tag{9.4}$$

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

$$\begin{aligned} Q_\infty &= [S_\infty : T_\infty]'[S_\infty : T_\infty], \\ Q_{1,\infty} &= \left(S'_\infty S_\infty, S'_\infty T_\infty\right)', \ Q_{T,\infty} = T'_\infty T_\infty, \ Q_{ST,\infty} = S'_\infty T_\infty, \\ Q_{S,\infty} &= S'_\infty S_\infty, \ \mathcal{S}_{2,\infty} = S'_\infty T_\infty/(||S_\infty|| \cdot ||T_\infty||), \text{ and} \\ \lambda_\infty &= C' D_Z C. \end{aligned} \tag{9.5}$$

The following result holds under the null hypothesis and fixed (i.e., non-local) alternative hypotheses.

Lemma 6 *Under Assumptions* WIV-FA *and* 1-4,
(a) $(S_n, T_n) \to_d (S_\infty, T_\infty)$,
(b) $(\widehat{S}_n, \widehat{T}_n) - (S_n, T_n) \to_p 0$, *and*
(c) $(\widehat{S}_n, \widehat{T}_n) \to_d (S_\infty, T_\infty)$.

Comments. 1. Inspection of the proof of the Lemma shows that the results of the Lemma hold uniformly over compact sets of true β and C values and over arbitrary sets of true γ and ξ values. In particular, the results hold uniformly over vectors C that include the zero vector. Hence, the asymptotic results hold uniformly over cases in which the IV's are arbitrarily weak. In consequence, we expect the asymptotic test procedures developed here to perform well in terms of size even for very weak IV's. Note that it is precisely these cases in which the t, Wald, and LR tests based on standard asymptotics perform poorly in terms of size.

2. Lemma 6 and the continuous mapping theorem imply that the asymptotic distributions of the $\widehat{AR}_n$, $\widehat{LM}_n$, $\widehat{LR}_n$, $\widehat{LR1}_n$, and $\widehat{POIS1}_{\widehat{\delta}}$ test statistics are given by the distributions of the test statistics in (3.4), (3.7), and (5.15) with (Q_S, Q_{ST}, Q_T) replaced by $(Q_{S,\infty}, Q_{ST,\infty}, Q_{T,\infty})$. In particular, under the null hypothesis, $\widehat{AR}_n$ and $\widehat{LM}_n$ have asymptotic χ^2_k and χ^2_1 distributions, respectively. The one-sided POI similar test against $\beta^* = \beta_{AR}$ is the AR test, see Comment 4 to Corollary 2, and, hence, the AR test for the case of unknown nonsingular Ω is asymptotically point optimal under weak IV asymptotics.

Using Lemma 6, we establish the asymptotic distributions of the $\{LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) : n \geq 1\}$ test statistics and $\{\kappa_\alpha(\widehat{Q}_{T,n}) : n \geq 1\}$ critical values.

Lemma 7 *The density, conditional density, and independence results of Lemma* 3 *for* $(Q_{1,n}, Q_{T,n})$, $Q_{T,n}$, $Q_{S,n}$, *and* $\mathcal{S}_{2,n}$ *also hold for* $(Q_{1,\infty}, Q_{T,\infty})$, $Q_{T,\infty}$, $Q_{S,\infty}$, *and* $\mathcal{S}_{2,\infty}$ *with* λ_n *replaced by* λ_∞.

Comment. Lemma 7 holds by (9.4) and the proof of Lemma 3.

As above, the following results hold under the null and fixed alternatives.

Theorem 6 *Under Assumptions* WIV-FA *and* 1-4,
(a) $(LR_W(Q_{1,n}, Q_{T,n}),\ \kappa_\alpha(Q_{T,n})) \to_d (LR_W(Q_{1,\infty}, Q_{T,\infty}),\ \kappa_\alpha(Q_{T,\infty}))$,
(b) $(LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}),\ \kappa_\alpha(\widehat{Q}_{T,n})) - (LR_W(Q_{1,n}, Q_{T,n}),\ \kappa_\alpha(Q_{T,n})) \to_p 0$, *and*
(c) $(LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}),\ \kappa_\alpha(\widehat{Q}_{T,n})) \to_d (LR_W(Q_{1,\infty}, Q_{T,\infty}),\ \kappa_\alpha(Q_{T,\infty}))$.

Theorem 6 leads to the following results.

Corollary 4 *Under Assumptions* WIV-FA *and* 1-4,
(a) $1(LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) > \kappa_\alpha(\widehat{Q}_{T,n})) - 1(LR_W(Q_{1,n}, Q_{T,n}) > \kappa_\alpha(Q_{T,n})) \to_p 0$,
(b) $P(LR_W(Q_{1,n}, Q_{T,n}) > \kappa_\alpha(Q_{T,n})) \to P(LR_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty}))$,
(c) $P(LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) > \kappa_\alpha(\widehat{Q}_{T,n})) \to P(LR_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty}))$, *and*
(d) *under the null hypothesis,* $P(LR_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty})) = \alpha$.

Comments. 1. Corollary 4(a) shows that the critical regions of the tests with known and unknown error covariance matrix differ with probability that converges to zero as $n \to \infty$. Hence, estimation of the error covariance matrix has no effect asymptotically.

2. Corollary 4(b) and (c) provide the asymptotic power functions of the tests based on known and unknown error covariance matrix. Consistent with the result of Corollary 4(a) , the asymptotic power functions are the same. The asymptotic power function depends only on β, C, and D_Z. It can be written as:

$$\begin{aligned} Pow_W(\beta, C, D_Z) &= P(LR_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty})) \\ &= \int 1(LR_W(q_1, q_T) > \kappa_\alpha(q_T)) f_{Q_1,Q_T}(q_1, q_T; \beta, C, D_Z) dq_1 dq_T, \end{aligned} \tag{9.6}$$

where $f_{Q_1,Q_T}(q_1, q_T; \beta, C, D_Z)$ is the density given in Lemma 3(a) with $\lambda = C' D_Z C$.

3. Combining Corollary 4(b) and (c) with Corollary 4(d) implies that the tests based on $LR_W(Q_{1,n}, Q_{T,n})$ and $LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n})$ both have asymptotic null rejection rates of α, as desired.

4. For the LU version of the HOM-WAP test, analogous results to those of Theorem 6 and Corollary 4 hold with $\kappa_\alpha(\widehat{Q}_{T,n})$ and $\kappa_\alpha(Q_{T,\infty})$ replaced by $\kappa_{1\alpha}(\widehat{Q}_{T,n}) + \widehat{Q}_{ST,n}\kappa_{s\alpha}(\widehat{Q}_{T,n})$ and $\kappa_{1\alpha}(Q_{T,\infty}) + Q_{ST,\infty}\kappa_{s\alpha}(Q_{T,\infty})$, respectively.

9.3 Heteroskedasticity-Robust Tests

We now introduce alternatives to the statistics $(\widehat{S}_n, \widehat{T}_n)$ that are adjusted to achieve robustness to heteroskedasticity. Define

$$\begin{aligned} \widetilde{S}_n &= \widetilde{\Sigma}_{S,n}^{-1/2} n^{-1/2} Z'Y b_0 \text{ and} \\ \widetilde{T}_n &= \widetilde{\Sigma}_{T,n}^{-1/2} \left(n^{-1/2} Z'Y \widehat{\Omega}_n^{-1} a_0 - \widetilde{\Sigma}_{TS,n} \widetilde{\Sigma}_{S,n}^{-1/2} \widetilde{S}_n \right), \text{ where} \\ \widetilde{\Sigma}_{S,n} &= n_{k,p}^{-1} \sum_{i=1}^n \left(\widehat{V}_i' b_0 Z_i \right) \left(\widehat{V}_i' b_0 Z_i \right)', \widetilde{\Sigma}_{TS,n} = n_{k,p}^{-1} \sum_{i=1}^n \left(\widehat{V}_i' \widehat{\Omega}_n^{-1} a_0 Z_i \right) \left(\widehat{V}_i' b_0 Z_i \right)', \\ \widetilde{\Sigma}_{T,n} &= \widetilde{\Sigma}_{T,n}^* - \widetilde{\Sigma}_{TS,n} \widetilde{\Sigma}_{S,n}^{-1} \widetilde{\Sigma}_{TS,n}', \ \widetilde{\Sigma}_{T,n}^* = n_{k,p}^{-1} \sum_{i=1}^n \left(\widehat{V}_i' \widehat{\Omega}_n^{-1} a_0 Z_i \right) \left(\widehat{V}_i' \widehat{\Omega}_n^{-1} a_0 Z_i \right)', \end{aligned} \tag{9.7}$$

$n_{k,p} = n - k - p$, and $\widehat{\Omega}_n$ and $\widehat{V}_i$ are defined in (9.1).[14]

[14] There is no need to recenter $\{\widehat{V}_i' b_0 Z_i : i \leq n\}$ by subtracting off its sample mean, $n^{-1} \sum_{j=1}^n \widehat{V}_j' b_0 Z_j$, in the definition of $\widetilde{\Sigma}_{S,n}$ because its sample mean is identically zero. The same holds for $\widetilde{\Sigma}_{TS,n}$ and $\widetilde{\Sigma}_{T,n}^*$.

The statistic $\widetilde{S}_n$ is based on $n^{-1/2}Z'Yb_0$, just as $\widehat{S}_n$ is, but is normalized by $\widetilde{\Sigma}_{S,n}^{-1/2}$, which is a consistent estimator of the square root of the asymptotic variance matrix of $n^{-1/2}Z'Yb_0$ even in the presence of heteroskedasticity. The statistic $\widetilde{T}_n$ is based on $n^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0$, as $\widehat{T}_n$ is, but is adjusted by subtracting off $\widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1/2}\widetilde{S}_n$ to achieve zero asymptotic covariance with $\widetilde{S}_n$ even in the presence of heteroskedasticity and is normalized by $\widetilde{\Sigma}_{T,n}^{-1/2}$ to achieve identity asymptotic covariance matrix even in the presence of heteroskedasticity. In the case of homoskedasticity, $\widetilde{\Sigma}_{TS,n} \to_p 0$ and the $\widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1/2}\widetilde{S}_n$ adjustment has no effect asymptotically.

Heteroskedasticity-robust AR, LM, CLR, CLR1, and POIS1 tests, denoted HR-AR, HR-LM, HR-CLR, HR-CLR1, and HR-POIS1, respectively, are defined as follows. The heteroskedasticity-robust test statistics, denoted $\widetilde{AR}_n$, $\widetilde{LM}_n$, $\widetilde{CR}_n$, $\widetilde{LR1}_n$, and $\widetilde{POIS1}_{\widetilde{\delta}}$, respectively, are defined as in (3.4), (3.7), and (5.15), but with (Q_S, Q_{ST}, Q_T) replaced by ($\widetilde{Q}_{S,n}, \widetilde{Q}_{ST,n}$, $\widetilde{Q}_{T,n}$). The appropriate critical values for these test statistics are the same as in the homoskedastic case. Thus, the critical values for the HR-AR and HR-LM tests are from χ^2_k and χ^2_1 distributions, respectively. The critical value functions for the HR-CLR, HR-CLR1, and HR-POIS1 tests are the same as in the homoskedastic error case. For the CLR test, see Table I of Moreira (2003).

A heteroskedasticity-robust optimal WAP test, referred to as an HR-WAP test, rejects $H_0 : \beta = \beta_0$ when

$$\begin{aligned} &LR_W(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n}) > \kappa_\alpha(\widetilde{Q}_{T,n}), \text{ where} \\ &\widetilde{Q}_{1,n} = (\widetilde{S}_n'\widetilde{S}_n, \widetilde{S}_n'\widetilde{T}_n)', \ \widetilde{Q}_{T,n} = \widetilde{T}_n'\widetilde{T}_n, \end{aligned} \tag{9.8}$$

and $\kappa_\alpha(\cdot)$ is defined in (4.12) (and can be calculated by the method following (4.12)). Note that the critical value function $\kappa_\alpha(\cdot)$ for the HR-WAP test is the same as for the HOM-WAP test. An LU version of the HR-WAP test is defined using the statistic $LR_W(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n})$ coupled with the critical value in Theorem 5 with Q_T and Q_{ST} replaced by $\widetilde{Q}_{T,n}$ and $\widetilde{Q}_{ST,n} = \widetilde{S}_n'\widetilde{T}_n$, respectively.

We now analyze the asymptotic properties of the various HR tests. Define

$$\begin{aligned} \widetilde{\Sigma}_S &= MB_0\Phi B_0'M', \ \widetilde{\Sigma}_{TS} = MA_0\Phi B_0'M', \ \widetilde{\Sigma}_T^* = MA_0\Phi A_0'M', \text{ and} \\ \widetilde{\Sigma}_T &= \widetilde{\Sigma}_T^* - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}\widetilde{\Sigma}_{TS}', \text{ where} \\ M &= \left[I_k : -D_{12}D_{22}^{-1}\right], \ B_0 = (b_0' \otimes I_{k+p}), \text{ and } A_0 = (\Omega^{-1}a_0)' \otimes I_{k+p}. \end{aligned} \tag{9.9}$$

The estimators $\widetilde{\Sigma}_{S,n}$, $\widetilde{\Sigma}_{TS,n}$, and $\widetilde{\Sigma}_{T,n}$ converge in probability to $\widetilde{\Sigma}_S$, $\widetilde{\Sigma}_{TS}$, and $\widetilde{\Sigma}_T$, respectively, when Assumptions WIV-FA and 1-3 and the following assumptions hold.

Assumption 5. $n^{-1}\sum_{i=1}^n (V_i \otimes \overline{Z}_i)(V_i \otimes \overline{Z}_i)' \to_p \Phi$.

Assumption 6. $n^{-1}\sum_{i=1}^n (||\overline{Z}_i||^4 + ||\overline{Z}_i||^3||V_i||) = O_p(1)$.

Any one of Assumptions IID, INID, or MDS is sufficient for Assumption 5.

Assumption 6 holds under Assumption IID or MDS plus the following assumption.

Assumption MOM. $E||\overline{Z}_i||^4 + E||\overline{Z}_i||^3||V_i|| < \infty$.

Assumption 6 holds under Assumption INID plus the following assumption.

Assumption MOM2. $E||\overline{Z}_i||^{4+\delta} + E||\overline{Z}_i||^{3+\delta}||V_i||^{1+\delta} < \infty$ for some $\delta > 0$.

Let $\widetilde{S}_\infty$ and $\widetilde{T}_\infty$ be independent random k-vectors with

$$\widetilde{S}_\infty \sim N(\widetilde{\Sigma}_S^{-1/2} D_Z C a' b_0, I_k) \text{ and}$$
$$\widetilde{T}_\infty \sim N\left(\widetilde{\Sigma}_T^{-1/2}\left(D_Z C a' \Omega^{-1} a_0 - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1} D_Z C a' b_0\right), I_k\right). \tag{9.10}$$

Let $\widetilde{Q}_{1,\infty} = \left(\widetilde{S}_\infty'\widetilde{S}_\infty, \widetilde{S}_\infty'\widetilde{T}_\infty\right)'$ and $\widetilde{Q}_{T,\infty} = \widetilde{T}_\infty'\widetilde{T}_\infty$.

The asymptotic properties of tests based on $(\widetilde{S}_n, \widetilde{T}_n)$ are as follows.

Theorem 7 *Under Assumptions* WIV-FA, 1-3, 5, *and* 6,
(a) $\widetilde{\Sigma}_{S,n} \to_p \widetilde{\Sigma}_S$, $\widetilde{\Sigma}_{TS,n} \to_p \widetilde{\Sigma}_{TS}$, *and* $\widetilde{\Sigma}_{T,n} \to_p \widetilde{\Sigma}_T$ *and*
(b) *Lemma* 6(c)*, Theorem* 6(c)*, and Corollary* 4(c) *and* (d) *hold with* $\widehat{S}_n$, $\widehat{T}_n$, S_∞, T_∞, $Q_{1,\infty}$, *and* $Q_{T,\infty}$ *replaced by* $\widetilde{S}_n$, $\widetilde{T}_n$, $\widetilde{S}_\infty$, $\widetilde{T}_\infty$, $\widetilde{Q}_{1,\infty}$, *and* $\widetilde{Q}_{T,\infty}$, *respectively.*

Comments. 1. Part (b) of the Theorem shows that HR-WAP tests have the correct significance level asymptotically whether or not the errors satisfy Assumption HOM. It shows that estimation of Ω, $\widetilde{\Sigma}_S$, $\widetilde{\Sigma}_{TS}$, and $\widetilde{\Sigma}_T$ does not affect the asymptotic distribution of $\{(LR_W(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n}), \kappa_\alpha(\widetilde{Q}_{T,n})) : n \geq 1\}$. It also shows that if the errors satisfy Assumption HOM, then the HR-WAP tests have the same asymptotic power as HOM-WAP tests because $(\widetilde{S}_\infty, \widetilde{T}_\infty)$ and (S_∞, T_∞) have the same distribution in this case.

2. Theorem 7(b) and the continuous mapping theorem imply that the asymptotic distributions of the $\widetilde{AR}_n$, $\widetilde{LM}_n$, $\widetilde{CR}_n$, $\widetilde{LR1}_n$, and $\widetilde{POIS1}_{\widetilde{\delta}}$ test statistics under Assumptions WIV-FA, 1-3, 5, and 6 are given by the distributions of the test statistics in (3.4), (3.7), and (5.15) with (S, T) replaced by $(\widetilde{S}_\infty, \widetilde{T}_\infty)$. Hence, the HR-AR and HR-LM statistics have asymptotic null χ_k^2 and χ_1^2 distributions, respectively. In addition, the critical value functions of the HR-CLR, HR-CLR1, and HR-POIS1 tests are the same as in the homoskedastic case and are determined by the density in Lemma 3(c). If Assumption HOM holds, then the asymptotic power functions of the HR-AR, HR-LM, HR-CLR, HR-CLR1, and HR-POIS1 tests are the same as the non-heteroskedasticity-robust versions of these tests.

3. For the LU version of the HR-WAP test, analogues of Theorem 6(c) and Corollary 4(c) and (d) hold under the assumptions of Theorem 7 with $\widehat{Q}_{1,n}, \widehat{Q}_{T,n}, Q_{1,\infty}$, $Q_{T,\infty}$, $\kappa_\alpha(\widehat{Q}_{T,n})$, and $\kappa_\alpha(Q_{T,\infty})$ replaced by $\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n}, \widetilde{Q}_{1,\infty}$, $\widetilde{Q}_{T,\infty}$, $\kappa_{1\alpha}(\widetilde{Q}_{T,n}) + \widetilde{Q}_{ST,n}\kappa_{2\alpha}(\widetilde{Q}_{T,n})$, and $\kappa_{1\alpha}(\widetilde{Q}_{T,\infty}) + \widetilde{Q}_{ST,\infty}\kappa_{2\alpha}(\widetilde{Q}_{T,\infty})$, respectively, where $\widetilde{Q}_{ST,\infty} = \widetilde{S}_\infty'\widetilde{T}_\infty$.

9.4 Heteroskedasticity and Autocorrelation Robust Tests

Tests that are robust to heteroskedasticity and autocorrelation in the reduced-form errors $\{V_i : i \geq 1\}$ are obtained by using the tests introduced in the previous subsection but with different estimators in place of $\widetilde{\Sigma}_{S,n}$, $\widetilde{\Sigma}_{TS,n}$, $\widetilde{\Sigma}_{T,n}$, and $\widetilde{\Sigma}_{T,n}^*$.

These are the only changes that are needed. In place of these estimators, one uses estimators of $\Sigma_{S,\infty}$, $\Sigma_{TS,\infty}$, $\Sigma_{T,\infty}$, and $\Sigma^*_{T,\infty}$, respectively, that are consistent (at least under the null hypothesis), where

$$\Sigma_\infty = \begin{bmatrix} \Sigma_{S,\infty} & \Sigma'_{TS,\infty} \\ \Sigma_{TS,\infty} & \Sigma^*_{T,\infty} \end{bmatrix} = \lim_{n\to\infty} var\left(n^{-1/2} \sum_{i=1}^n \begin{pmatrix} V_i' b_0 Z_i \\ V_i'\Omega^{-1} a_0 Z_i \end{pmatrix} \right) \text{ and}$$

$$\Sigma_{T,\infty} = \Sigma^*_{T,\infty} - \Sigma_{TS,\infty}\Sigma^{-1}_{S,\infty}\Sigma'_{TS,\infty}. \tag{9.11}$$

Let

$$\overline{\Sigma}_n = \begin{bmatrix} \overline{\Sigma}_{S,n} & \overline{\Sigma}'_{TS,n} \\ \overline{\Sigma}_{TS,n} & \overline{\Sigma}^*_{T,n} \end{bmatrix} \tag{9.12}$$

be a consistent estimator of Σ_∞ based on $\{(\widehat{V}_i' b_0 Z_i', \widehat{V}_i'\widehat{\Omega}_n^{-1} a_0 Z_i')' : i \le n\}$. There are many HAC estimators in the literature that can be used for this purpose, e.g., see Newey and West (1987), Andrews (1991), and Andrews and Monahan (1992). For brevity, we do not provide an explicit set of conditions under which one or more of these HAC estimators is consistent. We note, however, that the presence of weak IV's does not complicate standard proofs of the consistency of HAC estimators.

Given the estimator $\overline{\Sigma}_n$, the estimators $\widetilde{\Sigma}_{S,n}$, $\widetilde{\Sigma}_{TS,n}$, $\widetilde{\Sigma}_{T,n}$, and $\widetilde{\Sigma}^*_{T,n}$ are replaced in (9.7) by $\overline{\Sigma}_{S,n}$, $\overline{\Sigma}_{TS,n}$, $\overline{\Sigma}_{T,n}$, and $\overline{\Sigma}^*_{T,n}$, respectively, where

$$\overline{\Sigma}_{T,n} = \overline{\Sigma}^*_{T,n} - \overline{\Sigma}_{TS,n}\overline{\Sigma}^{-1}_{S,n}\overline{\Sigma}'_{TS,n}. \tag{9.13}$$

Let $\overline{S}_n$, $\overline{T}_n$, $\overline{Q}_{1,n}$, $\overline{Q}_{S,n}$, $\overline{Q}_{ST,n}$, and $\overline{Q}_{T,n}$ denote $\widetilde{S}_n$, $\widetilde{T}_n$, $\widetilde{Q}_{1,n}$, $\widetilde{Q}_{S,n}$, $\widetilde{Q}_{ST,n}$, and $\widetilde{Q}_{T,n}$, respectively, with these changes. Heteroskedasticity and autocorrelation-robust AR, LM, LR, LR1, and POIS1 test statistics, denoted $\overline{AR}_n$, $\overline{LM}_n$, $\overline{LR}_n$, $\overline{LR1}_n$, and $\overline{POIS1}_{\overline{\delta}}$, respectively, are defined as in (3.4), (3.7), and (5.15), but with (Q_S, Q_{ST}, Q_T) replaced by $(\overline{Q}_S, \overline{Q}_{ST}, \overline{Q}_T)$. The corresponding tests are denoted HAR-AR, HAR-LM, HAR-CLR, HAR-CLR1, and HAR-POIS1, respectively. The appropriate critical values for these test statistics are the same as in the homoskedastic case.

A heteroskedasticity and autocorrelation-robust optimal WAP test, referred to as an HAR-WAP test, rejects $H_0 : \beta = \beta_0$ when

$$LR_W(\overline{Q}_{1,n}, \overline{Q}_{T,n}) > \kappa_\alpha(\overline{Q}_{T,n}), \tag{9.14}$$

where $\kappa_\alpha(\cdot)$ is defined in (4.12). An LU version of the HAR-WAP test is defined using the statistic $LR_W(\overline{Q}_{1,n}, \overline{Q}_{T,n})$ combined with the critical value given in Theorem 5 with Q_T and Q_{ST} replaced by $\overline{Q}_{T,n}$ and $\overline{Q}_{ST,n}$, respectively.

The HAR-AR, HAR-LM, HAR-CLR, HAR-CLR1, HAR-POIS1, HAR-WAP, and LU-HAR-WAP tests have correct asymptotic significance level under Assumptions WIV-FA and 1-3 plus the additional conditions that are needed to obtain consistency of $\overline{\Sigma}_n$ for Σ_∞. Furthermore, if Assumption 4 also holds, these tests have the same asymptotic power functions as the corresponding AR, LM, CLR, CLR1, POIS1, HOM-WAP, and LU-HOM-WAP tests, or if Assumptions 5 and 6 also hold, they have the same asymptotic power functions as the HR-AR, HR-LM, HR-CLR, HR-CLR1, HR-POIS1, HR-WAP, and LU-HR-WAP tests.

10 Asymptotic Optimality and Power Envelope with Weak IV's

In this section, we show that the tests HOM-WAP, HR-WAP, and HAR-WAP, and the corresponding LU versions of these tests, exhibit certain asymptotic WAP optimality properties when the IV's are weak and the errors are iid normal with *unknown* covariance matrix. These results immediately provide one- and two-sided asymptotic power envelopes by considering one- and two-point weight functions, see Sections 5 and 6.2. These asymptotic power envelopes are the same as the finite sample power envelopes determined in Sections 5 and 6.2 for the case of iid normal errors with *known* covariance matrix.

For the asymptotic optimality results, we set up a sequence of models (or *experiments*) with the parameters renormalized such that no parameter can be estimated asymptotically without error, as is standard in the asymptotic efficiency literature, e.g., see van der Vaart (1998, Ch. 9). For the parameters β and C, no renormalization is required given Assumption WIV-FA because neither can be consistently estimated in the weak IV asymptotic setup. For the parameters Ω and η, renormalizations are required. We take the true parameters Ω and η to satisfy

$$\Omega = \Omega_0 + \Omega_1/n^{1/2} \text{ and } \eta = \eta_0 + \eta_1/n^{1/2}, \tag{10.1}$$

where Ω_0 and η_0 are taken to be known and the unknown parameters to be estimated are the perturbation parameters η_1 and Ω_1. The matrices Ω_0 and Ω_1 are assumed to be symmetric and pd.

The least squares estimator of η in the model of (2.5) is denoted $\widehat{\eta}_n = (X'X)^{-1}X'Y$. (The form of $\widehat{\eta}_n$ relies on the fact that $Z'X = 0$ by construction of Z.)

For any symmetric $\ell \times \ell$ matrix A, let $vech(A)$ denote the $\ell(\ell+1)/2$-column vector containing the column by column vectorization of the non-redundant elements of A.

The following basic results hold under the null hypothesis $\beta = \beta_0$ and fixed alternatives $\beta \neq \beta_0$:

Lemma 8 *Suppose Assumption* WIV-FA *holds, the reduced-form errors* $\{V_i : i \geq 1\}$ *are iid normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and pd variance matrix* Ω, *and* Ω *and* η *are as in* (10.1)*. Then,*
(a) $(n^{-1/2}Z'Y,\ n^{1/2}(\widehat{\eta}_n - \eta_0),\ n^{1/2}(\widehat{\Omega}_n - \Omega_0))$ *are sufficient statistics for* $(\beta, C, \Omega_1, \eta_1)$,
(b) $(n^{-1/2}Z'Y,\ n^{1/2}(\widehat{\eta}_n - \eta_0),\ n^{1/2}(\widehat{\Omega}_n - \Omega_0)) \to_d (N_Z,\ N_X,\ N_\Omega)$, *where* N_Z, N_X, *and* N_Ω *are independent* $k \times 2$, $p \times 2$, *and* 2×2 *normal random matrices, respectively, with* $vec(N_Z) \sim N(vec(D_Z C a'), \Omega_0 \otimes D_Z)$, $vec(N_X) \sim N(vec(\eta_1), \Omega_0 \otimes D_{22}^{-1})$, N_Ω *is symmetric, and* $vech(N_\Omega) \sim N(\Omega_1, E(\zeta - E\zeta)(\zeta - E\zeta)')$, *where* $\zeta = vech(v_0 v_0')$, $v_0 \in R^2$, *and* $v_0 \sim N(0, \Omega_0)$, *provided Assumption* 1 *also holds.*

Given the result of part (a) of the Lemma, there is no loss in attainable power by considering only tests that depend on the data through $(n^{-1/2}Z'Y,\ n^{1/2}(\widehat{\eta}_n - \eta_0),$ $n^{1/2}(\widehat{\Omega}_n - \Omega_0))$. Let $\phi_n(n^{-1/2}Z'Y,\ n^{1/2}(\widehat{\eta}_n - \eta_0),\ n^{1/2}(\widehat{\Omega}_n - \Omega_0))$ be such a test. The test ϕ_n is $\{0,1\}$-valued and rejects the null hypothesis when $\phi_n = 1$. We say that a sequence of tests $\{\phi_n : n \geq 1\}$ is a *convergent sequence of asymptotically similar* tests if, for some function $\phi(\cdot,\cdot,\cdot)$,

$$\phi_n(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0)) \to_d \phi(N_Z, N_X, N_\Omega) \text{ and } P_{\beta,C,\Omega_0,\eta_0}(\phi(N_Z, N_X, N_\Omega) = 1) = \alpha \quad (10.2)$$

for $\beta = \beta_0$ and all (C, Ω_0, η_0) in the parameter space, where $P_{\beta,C,\Omega_0,\eta_0}(\cdot)$ denotes probability when the true parameters are $(\beta, C, \Omega_0, \eta_0)$. Examples of convergent sequences of asymptotically similar tests include sequences of AR, HR-AR, HAR-AR, LM, HR-LM, HAR-LM, CLR, HR-CLR, HAR-CLR, CLR1, HR-CLR1, HAR-CLR1, HOM-WAP, HR-WAP, and HAR-WAP tests. Standard Wald and LR tests are not asymptotically similar due to the effect of weak IV's.

The transformation, call it $h_\Omega(\cdot)$, from N_Z to $[S_\infty : T_\infty]$ in (9.4) is one-to-one. Hence, for some function $\overline{\phi}$, we have

$$\phi(N_Z, N_X, N_\Omega) = \phi(h_\Omega^{-1}(S_\infty, T_\infty), N_Z, N_\Omega) = \overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega). \quad (10.3)$$

As in Section 3, we consider the group of transformations given in (3.1) but with $\overline{g}_F(\beta, \pi)$ replaced by $\overline{g}_F(\beta, C) = (\beta, D_Z^{-1/2} F' D_Z^{1/2} C)$ acting on the parameters (β, C). The maximal invariant is Q_∞ (defined in (9.5)).

We say that a sequence of tests $\{\phi_n : n \geq 1\}$ is a convergent sequence of *asymptotically invariant* tests if the first condition of (10.2) holds and distribution of $\overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega)$ depends on (S_∞, T_∞) only through Q_∞, i.e.,

$$\overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega) \sim \phi^*(Q_\infty, N_X, N_\Omega) \quad (10.4)$$

for some function ϕ^*, where $\sim$ denotes "has the same distribution as." Examples of convergent sequences of asymptotically invariant and asymptotically similar tests include the tests listed above in the paragraph containing (10.2).

We now establish an upper bound on asymptotic WAP.

Theorem 8 *Suppose Assumptions* WIV-FA *and* 1 *hold, the reduced-form errors* $\{V_i : i \geq 1\}$ *are iid normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and pd variance matrix* Ω, *and* Ω *and* η *are as in* (10.1). *For any convergent sequence of asymptotically invariant and asymptotically similar tests* $\{\phi_n : n \geq 1\}$, *we have*

$$\begin{aligned}
&\lim_{n\to\infty} \int P_{\beta,\lambda,\Omega,\eta}(\phi_n(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0)) = 1) dW(\beta, \lambda) \\
&= \int P_{\beta,\lambda,\Omega_0,\eta_0}(\phi^*(Q_\infty, N_X, N_\Omega) = 1) dW(\beta, \lambda) \\
&\leq \int P_{\beta,\lambda,\Omega_0,\eta_0}(LR_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty})) dW(\beta, \lambda),
\end{aligned}$$

where $P_{\beta,\lambda,\Omega,\eta}(\cdot)$ *denotes probability when the true parameters are* (β, C, Ω, η) *for some* C *such that* $C' D_Z C = \lambda$.

Comment. Under $H_0 : \beta = \beta_0$, the left- and right-hand sides of the inequality in the Theorem equal α.

Combining Theorem 8 with Corollary 4(c), Theorem 7(b), and the results of Section 9.4 gives the following asymptotic optimality property for HOM-WAP, HR-WAP, and HAR-WAP tests.

Corollary 5 *Under the conditions of Theorem* 8*, the* HOM-WAP*,* HR-WAP*, and* HAR-WAP *tests of Section* 9 *are convergent sequences of asymptotically invariant and asymptotically similar tests that attain the upper bound on asymptotic* WAP *given in Theorem* 8*.*

Comments. 1. By considering one- and two-point weight functions, as in Sections 5 and 6.2, Corollary 5 gives the one- and two-sided asymptotic power envelopes for asymptotically invariant and asymptotically similar tests. These asymptotic power envelopes are the same as the finite sample power envelopes for known Ω given in Sections 5 and 6.2 with $\lambda_\infty = C'D_ZC$ in place of $\lambda = \pi'Z'Z\pi$ (where $C'D_ZC = \lim_{n\to\infty} \pi'Z'Z\pi$ under weak IV asymptotics).

2. In Theorem 8 and Corollary 5, the assumption that the reduced-form errors $\{V_i : i \geq 1\}$ are iid normal, independent of $\{\overline{Z}_i : i \geq 1\}$, with mean zero and pd variance matrix Ω, can be replaced by Assumptions 2-4. The latter allow for non-normal errors. But, with this replacement, Lemma 8(a) no longer holds and it is no longer true that there is no loss in attainable power by considering only tests that depend on the data through $(n^{-1/2}Z'Y,\ n^{1/2}(\widehat{\eta}_n - \eta_0),\ n^{1/2}(\widehat{\Omega}_n - \Omega_0))$.

3. We say that a convergent sequence $\{\phi_n : n \geq 1\}$ of asymptotically invariant and asymptotically similar tests is *asymptotically LU* if $E_{\beta,C,\Omega_0,\eta_0}\phi^*(Q_\infty, N_X, N_\Omega)Q_{ST,\infty}/Q_{T,\infty}^{1/2} = 0$ for $\beta = \beta_0$ and for all (C, Ω_0, η_0) in the parameter space, where $Q_{ST,\infty} = S'_\infty T_\infty$ and $Q_{T,\infty} = T'_\infty T_\infty$. An analogue of Theorem 8 holds for sequences of such tests with $\kappa_\alpha(Q_{T,\infty})$ replaced by $\kappa_{1\alpha}(Q_{T,\infty}) + Q_{ST,\infty}\kappa_{2\alpha}(Q_{T,\infty})$ in the upper bound. Similarly, an analogue of Corollary 5 holds for the LU versions of the HOM-WAP, HR-WAP, and HAR-WAP tests with the critical value in the upper bound in Theorem 8 altered as above. Hence, these tests possess some asymptotic optimality properties under weak IV's. By considering one-point weight functions, these results yield the asymptotic power envelope for convergent sequences of asymptotically invariant/similar/LU tests.

11 Strong IV Asymptotics for Case of Unknown Covariance Matrix and Non-normal Errors

In this section, we analyze the strong IV asymptotic properties of the tests considered above for both local alternatives and fixed alternatives. Under strong IV asymptotics, π is a fixed non-zero vector. We utilize the same notation as in Sections 9 and 10. Thus, $S = S_n$, $Q = Q_n$, etc.

11.1 Local Alternatives

For local alternatives, β is local to the null value β_0 as $n \to \infty$. We assume:

Assumption SIV-LA. (a) $\beta = \beta_0 + B/n^{1/2}$ for some constant $B \in R$.

(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-local alternative (SIV-LA) asymptotic behavior of S_n, $\widehat{S}_n$, T_n, and $\widehat{T}_n$ depends on

$$\begin{aligned} \zeta_S &\sim N(\alpha_S, I_k), \\ \alpha_S &= D_Z^{1/2} \pi B (b_0' \Omega b_0)^{-1/2}, \text{ and} \\ \alpha_T &= D_Z^{1/2} \pi (a_0' \Omega^{-1} a_0)^{1/2}. \end{aligned} \quad (11.1)$$

The SIV-LA asymptotic behavior of $\widetilde{S}_n$ and $\widetilde{T}_n$ depends on

$$\begin{aligned} \widetilde{\zeta}_S &\sim N(\widetilde{\alpha}_S, I_k), \\ \widetilde{\alpha}_S &= \widetilde{\Sigma}_S^{-1/2} D_Z \pi B, \text{ and} \\ \widetilde{\alpha}_T &= \widetilde{\Sigma}_T^{-1/2} D_Z \pi a_0' \Omega^{-1} a_0. \end{aligned} \quad (11.2)$$

Using these definitions, we obtain the following results.

Lemma 9 (a) *Under Assumptions* SIV-LA *and* 1-4, (i) $(S_n, T_n/n^{1/2}) \to_d (\zeta_S, \alpha_T)$, (ii) $(\widehat{S}_n, \widehat{T}_n/n^{1/2}) = (S_n, T_n/n^{1/2}) + o_p(1)$, *and* (iii) $(\widehat{Q}_{S,n}, \widehat{Q}_{ST,n}/n^{1/2}, \widehat{Q}_{T,n}/n) \to_d (\zeta_S'\zeta_S, \alpha_T'\zeta_S, \alpha_T'\alpha_T)$ *as* $n \to \infty$.

(b) *Under Assumptions* SIV-LA, 1-3, 5, *and* 6, (i) $\widetilde{\Sigma}_{S,n} \to_p \widetilde{\Sigma}_S$, $\widetilde{\Sigma}_{TS,n} \to_p \widetilde{\Sigma}_{TS}$, *and* $\widetilde{\Sigma}_{T,n} \to_p \widetilde{\Sigma}_T$, (ii) $(\widetilde{S}_n, \widetilde{T}_n/n^{1/2}) \to_d (\widetilde{\zeta}_S, \widetilde{\alpha}_T)$, *and* (iii) $(\widetilde{Q}_{S,n}, \widetilde{Q}_{ST,n}/n^{1/2}, \widetilde{Q}_{T,n}/n) \to_d (\widetilde{\zeta}_S'\widetilde{\zeta}_S, \widetilde{\alpha}_T'\widetilde{\zeta}_S, \widetilde{\alpha}_T'\widetilde{\alpha}_T)$ *as* $n \to \infty$.

Using Lemma 9, we determine the asymptotic distributions of the AR, LM, LR, and LR1 test statistics and their heteroskedasticity-robust versions under SIV-LA asymptotics.

Theorem 9 (a) *Under Assumptions* SIV-LA *and* 1-4, (i) $\widehat{AR}_n = AR_n + o_p(1) \to_d \zeta_S'\zeta_S \sim \chi_k^2(\alpha_S'\alpha_S)$, (ii) $\widehat{LM}_n = LM_n + o_p(1) \to_d (\alpha_T'\zeta_S)^2/||\alpha_T||^2 \sim \chi_1^2((\alpha_T'\alpha_S)^2/||\alpha_T||^2)$, (iii) $\widehat{LR}_n = LR_n + o_p(1) = LM_n + o_p(1) \to_d \alpha_T'\zeta_S/||\alpha_T|| \sim \chi_1^2((\alpha_T'\alpha_S)^2/||\alpha_T||^2)$, *and* (iv) $\widehat{LR1}_n = LR1_n + o_p(1) \to_d LR1_\infty$, *where* $LR1_\infty$ *is defined in the proof.*

(b) *Under Assumptions* SIV-LA, 1-3, 5, *and* 6, (i) $\widetilde{AR}_n \to_d \widetilde{\zeta}_S'\widetilde{\zeta}_S \sim \chi_k^2(\widetilde{\alpha}_S'\widetilde{\alpha}_S)$, (ii) $\widetilde{LM}_n \to_d (\widetilde{\alpha}_T'\widetilde{\zeta}_S)^2/||\widetilde{\alpha}_T||^2 \sim \chi_1^2((\widetilde{\alpha}_T'\widetilde{\alpha}_S)^2/||\widetilde{\alpha}_T||^2)$, (iii) $\widetilde{LR}_n = \widetilde{LM}_n + o_p(1) \to_d (\widetilde{\alpha}_T'\widetilde{\zeta}_S)^2/||\widetilde{\alpha}_T||^2 \sim \chi_1^2((\widetilde{\alpha}_T'\widetilde{\alpha}_S)^2/||\widetilde{\alpha}_T||^2)$, *and* (iv) $\widetilde{LR1}_n \to_d \widetilde{LR1}_\infty$, *where* $\widetilde{LR1}_\infty$ *is defined in the proof.*

Comments. 1. Parts (a)(iii) and (b)(iii) of Theorem 9 show that the LM and LR test statistics are asymptotically equivalent under SIV-LA asymptotics for any value of k (the number of IV's). (When $k = 1$, the AR, LM, and LR test statistics are the same, so the tests are trivially asymptotically equivalent.)

2. The critical values for AR and LM tests are non-random. However, the critical values for CLR tests are functions of $Q_{T,n}$, $\widehat{Q}_{T,n}$, or $\widetilde{Q}_{T,n}$. Hence, for LM and CLR *tests* to be asymptotically equivalent, the CLR critical value, call it $\kappa_\alpha^{CLR}(Q_{T,n})$, must converge in probability to a constant as $n \to \infty$. Under strong IV asymptotics, $Q_{T,n} \to_p \infty$. In consequence, asymptotic equivalence holds if $\kappa_\alpha^{CLR}(q_T)$ converges to a finite constant as q_T diverges to infinity. Moreira (2003) shows that $\lim_{q_T \to \infty} \kappa_\alpha^{CLR}(q_T)$ equals the $1-\alpha$ quantile of the χ_1^2 distribution. Hence, the LM and CLR tests are indeed asymptotically equivalent under SIV-LA asymptotics.

3. When Assumption 4 holds, $\widetilde{\alpha}_S = \alpha_S$, $\widetilde{\alpha}_T = \alpha_T$, $\widetilde{\zeta}_S = \zeta_S$, and the asymptotic distributions of $\widetilde{AR}_n$, $\widetilde{LM}_n$, $\widetilde{LR}_n$, and $\widetilde{LR1}_n$ are the same of those of $\widehat{AR}_n$, $\widehat{LM}_n$, $\widehat{LR}_n$, and $\widehat{LR1}_n$, respectively.

4. Theorem 9(a)(i) and (a)(ii) are not new results, but the rest of Theorem 9 is new. Moreira (2003) does not provide the SIV-LA asymptotic distribution of $\widehat{LR}_n$ and the test statistics $\widehat{LR1}_n$, $\widetilde{AR}_n$, $\widetilde{LM}_n$, $\widetilde{LR}_n$, and $\widetilde{LR1}_n$ are new to this paper.

5. The heteroskedasticity and autocorrelation robust test statistics $\overline{AR}_n$, $\overline{LM}_n$, $\overline{LR}_n$, and $\overline{LR1}_n$ satisfy analogous results to those in part (b) of the Theorem for $\widetilde{AR}_n$, $\widetilde{LM}_n$, $\widetilde{LR}_n$, and $\widetilde{LR1}_n$, but with $\widetilde{\Sigma}_S$ and $\widetilde{\Sigma}_T$ replaced by $\Sigma_{S,\infty}$ and $\Sigma_{T,\infty}$ in definitions of $\widetilde{\alpha}_S$ and $\widetilde{\alpha}_T$. These analogous results hold under assumptions of the Theorem plus the additional conditions that are needed to obtain consistency of $\overline{\Sigma}_n$ for Σ_∞.

Under SIV-LA asymptotics and iid normal errors with unknown covariance matrix Ω, the model for (y_1, y_2) is a "regular" parametric model in the sense of standard likelihood theory. Hence, the usual Wald, likelihood ratio, and Lagrange multiplier tests have standard large sample optimality properties. Such optimality properties include maximizing average asymptotic power over certain ellipses in the parameter space and uniformly maximizing asymptotic power among asymptotically unbiased tests, see Wald (1943). We refer to tests with such properties as *asymptotically efficient* (AE) tests under SIV-LA asymptotics and iid normal errors.

We have the following AE result for LM and CLR tests under SIV-LA asymptotics.

Theorem 10 *Suppose Assumptions SIV-LA and* 1 *hold and the reduced-form errors* $\{V_i : i \geq 1\}$ *are iid normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and pd variance matrix* Ω *which may be known or unknown. Then, the* LM *tests based on* $\widehat{LM}_n$, $\widetilde{LM}_n$, *and* $\overline{LM}_n$ *and the* CLR *tests based on* $\widehat{LR}_n$, $\widetilde{LR}_n$, *and* $\overline{LR}_n$ *are asymptotically efficient under strong IV asymptotics.*

Comment. The AR tests based on $\widehat{AR}_n$, $\widetilde{AR}_n$, and $\overline{AR}_n$ are not AE under SIV-LA asymptotics and iid normal errors unless $k = 1$. This holds because their asymptotic distribution under SIV-LA asymptotics differs from that of $\widehat{LM}_n$ when $k > 1$ by Theorem 9.

We now consider the behavior of one-sided POI similar tests under SIV-LA asymptotics.

Theorem 11 (a) *Under Assumptions SIV-LA and* 1-4, (i) *if* $\beta^* \neq \beta_{AR}$ *and* $\beta^* \neq \beta_0$, *then* $\widehat{POIS}1_{\widehat{\delta}}/\widehat{\delta} = POIS1_\delta/\delta + o_p(1) = sgn(c_{\beta^*}d_{\beta^*})Q_{ST}/Q_T^{1/2} + o_p(1) \rightarrow_d sgn(c_{\beta^*}d_{\beta^*}) \times(\alpha_T'\zeta_S)/||\alpha_T|| \sim N(sgn(c_{\beta^*}d_{\beta^*})(\alpha_T'\alpha_S)/||\alpha_T||, 1)$, *and* (ii) *if* $\beta^* = \beta_{AR}$, $\sqrt{2k}\widehat{POIS}1_{\widehat{\delta}} + k = \sqrt{2k}POIS1_\delta + k + o_p(1) = Q_S \rightarrow_d \zeta_S'\zeta_S \sim \chi_k^2(\alpha_S'\alpha_S)$, *where* β^* *is the alternative against which* $POIS1_\delta$ *is POI and* β_{AR} *is defined in* (5.16).

(b) *Under Assumptions* SIV-LA, 1-3, 5, *and* 6, (i) *if* $\beta^* \neq \beta_{AR}$ *and* $\beta^* \neq \beta_0$, *then* $\widetilde{POIS}1_{\widetilde{\delta}}/\widetilde{\delta} = sgn(c_{\beta^*}d_{\beta^*})\widetilde{Q}_{ST}/\widetilde{Q}_T^{1/2} + o_p(1) \rightarrow_d sgn(c_{\beta^*}d_{\beta^*})(\widetilde{\alpha}_T'\widetilde{\zeta}_S)/||\widetilde{\alpha}_T|| \sim N(sgn(c_{\beta^*}d_{\beta^*})\times$
$(\widetilde{\alpha}_T'\widetilde{\alpha}_S)/||\widetilde{\alpha}_T||, 1)$, *and* (ii) *if* $\beta^* = \beta_{AR}$, *then* $\sqrt{2k}\widetilde{POIS}1_{\widetilde{\delta}} + k = \widetilde{Q}_S \rightarrow_d \widetilde{\zeta}_S'\widetilde{\zeta}_S \sim \chi_k^2(\widetilde{\alpha}_S'\widetilde{\alpha}_S)$.

Comments. 1. The Theorem shows that one-sided POI test statistics are asymptotically equivalent to a one-sided LM statistic whose sign depends on β^* except in the special case in which $\beta^* = \beta_{AR}$. Furthermore, the critical value of a POIS1 test, call it $\kappa_\alpha^{POIS1}(Q_T)$, converges in probability to the $1-\alpha$ quantile of the standard normal distribution as $q_T \rightarrow \infty$. (See the Appendix for a proof.) Hence, one-sided POI *tests* are asymptotically equivalent to one-sided LM *tests* when $\beta \neq \beta_{AR}$.

When $\beta = \beta_{AR}$,, the one-sided POI test statistic is asymptotically equivalent to the (centered and rescaled) AR statistic. In consequence, the AR test is asymptotically POI for a particular one-sided local alternative.

2. It is quite interesting to see that the one-sided LM statistic (to which the one-sided POI test statistic is asymptotically equivalent) can change sign depending upon the magnitude of β^* even for β^* values on the same side of the null hypothesis. This occurs because d_{β^*} can change sign even for β^* values on the same side of the null.

We have $sgn(c_{\beta^*}d_{\beta^*}) = sgn(\beta^* - \beta_0)sgn(d_{\beta^*})$ and d_{β^*} is proportional to $(\omega_{11} - \omega_{12}\beta_0) - \beta^*(\omega_{12} - \omega_{22}\beta_0)$. For example, if $\beta_0 = 0$, then $sgn(c_{\beta^*}d_{\beta^*})$ exhibits the following properties. If $\beta^* > 0$ and $\omega_{12} > 0$, then $slope(d_{\beta^*}) < 0$ (where $slope(d_{\beta^*})$ denotes the slope of d_{β_*} as a function of β^*), $\beta_{AR} > 0$ (where by definition $d_\beta = 0$ for $\beta = \beta_{AR}$), $sgn(c_{\beta^*}d_{\beta^*}) = 1$ for $0 < \beta^* < \beta_{AR}$, and $sgn(c_{\beta^*}d_{\beta^*}) = -1$ for $\beta^* > \beta_{AR}$. If $\beta^* > 0$ and $\omega_{12} < 0$, then $slope(d_{\beta^*}) > 0$, $\beta_{AR} < 0$, $d_{\beta^*} > 0$, and $sgn(c_{\beta^*}d_{\beta^*}) = 1$. If $\beta^* < 0$ and $\omega_{12} < 0$, then $slope(d_{\beta^*}) > 0$, $\beta_{AR} < 0$, $sgn(c_{\beta^*}d_{\beta^*}) = -1$ for $\beta_{AR} < \beta^* < 0$, and $sgn(c_{\beta^*}d_{\beta^*}) = 1$ for $\beta^* < \beta_{AR}$. If $\beta^* < 0$ and $\omega_{12} > 0$, then $slope(d_{\beta^*}) < 0$, $\beta_{AR} > 0$, $d_{\beta^*} > 0$, and $sgn(c_{\beta^*}d_{\beta^*}) = -1$ for all $\beta^* < 0$. Hence, when $\beta_0 = 0$, $c_{\beta^*}d_{\beta^*}$ switches sign for $\beta^* > 0$ when $\omega_{12} > 0$ and $c_{\beta^*}d_{\beta^*}$ switches sign for $\beta^* < 0$ when $\omega_{12} < 0$.

Next, we consider the SIV-LA asymptotic behavior of various WAP tests designed for two-sided alternatives. We consider WAP tests based on two-point weight functions W_{2P}. Our results imply that, under iid normal errors, a test based on W_{2P} is asymptotically efficient under SIV-LA asymptotics if and only if W_{2P} satisfies (6.2). We also consider WAP tests based on AE weight functions W_{AE}. Our results imply that such tests are asymptotically efficient when the errors are iid normal.

Theorem 12 (a) *Under Assumptions* SIV-LA *and* 1-4*,* (i) $LR_{W_{2P}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) = LR_{W_{2P}}(Q_{1,n}, Q_{T,n}) + o_p(1)$*,* (ii) *if* W_{2P} *satisfies* (6.2)*, then* $LR_{W_{2P}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) = e^{-\frac{1}{2}(\gamma^*)^2}\cosh(\gamma^* LM_n^{1/2}) + o_p(1)$*, where* $\gamma^* = (\lambda^*)^{1/2} c_{\beta^*}$*, which is a strictly-increasing continuous function of* LM_n*,* (iii) *if* W_{2P} *does not satisfy* (6.2)*, then* $LR_{W_{2P}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) = \eta_2(Q_{ST,n}/Q_{T,n}^{1/2}) + o_p(1)$ *for a continuous function* $\eta_2(\cdot)$ *that is not even, and* (iv) *if* W_{AE} *satisfies* (6.4) *and conditions* (a) *and* (b) *following it, then* $LR_{W_{AE}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) = \eta_3(LM_n) + o_p(1)$ *for a strictly-increasing continuous function* $\eta_3(\cdot)$*.*

(b) *Under Assumptions* SIV-LA, 1-3, 5, *and* 6*,* (i) *if* W_{2P} *satisfies* (6.2)*, then* $LR_{W_{2P}}(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n}) = \eta_1(\widetilde{LM}_n) + o_p(1)$ *for* $\eta_1(\cdot)$ *as above,* (ii) *if* W_{2P} *does not satisfy* (6.2)*, then* $LR_{W_{2P}}(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n}) = \eta_2(\widetilde{Q}_{ST,n}/\widetilde{Q}_{T,n}^{1/2}) + o_p(1)$ *for* $\eta_2(\cdot)$ *as above, and* (iii) $LR_{W_{AE}}(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n}) = \eta_3(\widetilde{LM}_n) + o_p(1)$ *for* $\eta_3(\cdot)$ *as above.*

Comments. **1.** The critical values for the $LR_{W_{2P}}$ and $LR_{W_{AE}}$ tests converge in probability to constants as $n \to \infty$ under strong IV asymptotics. (See the Appendix for a proof.) Hence, Theorem 12(a)(ii) and (a)(iii), combined with Theorem 10, imply that a WAP test based on W_{2P} is AE under SIV-LA asymptotics and iid normal reduced-form errors iff W_{2P} satisfies (6.2).

2. Theorem 12(a)(i) shows that, under SIV-LA asymptotics and the homoskedastic errors assumptions (which do not require normality), a WAP test with estimated error variance matrix Ω is asymptotically equivalent to the corresponding WAP test with known Ω. Under the same assumptions, Theorem 12(a)(ii) shows that a WAP test based on W_{2P} is asymptotically equivalent to the two-sided LM test with known Ω when (6.2) holds. Under the same assumptions, Theorem 12(a)(iii) shows that a WAP test based on W_{2P} is asymptotically equivalent to a test based on a continuous function of the two one-sided LM statistics with known Ω, viz., $\pm Q_{ST,n}/Q_{T,n}^{1/2}$, when (6.2) fails to hold. Theorem 12(b)(i)-(iii) establishes analogous results to those of Theorem 12(a)(ii)-(iv) but under assumptions that allow for heteroskedastic errors and for heteroskedasticity-robust WAP and two- and one-sided LM tests.

3. The proof of Theorem 12(a)(iii) shows that if the second condition of (6.2) fails to hold, then $\eta_2(\cdot)$ is a monotone function and, hence, the WAP test based on W_{2P} is asymptotically equivalent to one or the other of the one-sided LM tests based on $\pm Q_{ST,n}/Q_{T,n}^{1/2}$. On the other hand, the proof shows that if second condition of (6.2) holds and the first condition fails, then the WAP test based on W_{2P} is asymptotically equivalent to a function of both one-sided LM statistics $\pm Q_{ST,n}/Q_{T,n}^{1/2}$ that is not invariant to permutations of the two one-sided statistics.

4. As defined, two-point weight functions W_{2P} place equal weight (1/2) on the two points (β^*, λ^*) and (β_2^*, λ_2^*). One also could consider weight functions that place unequal weight on two points. The proof of Theorem 12 shows that any such weight function behaves as in Theorem 12(a)(iii) and, hence, leads to a test that is not AE under iid normal errors.

5. The HAR test statistics $LR_{W_{2P}}(\overline{Q}_{1,n}, \overline{Q}_{T,n})$ and $LR_{W_{AE}}(\overline{Q}_{1,n}, \overline{Q}_{T,n})$ satisfy analogous results to those in part (b) of the Theorem for $LR_{W_{2P}}(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n})$ and $LR_{W_{AE}}(\widetilde{Q}_{1,n}, \widetilde{Q}_{T,n})$, but with $\widetilde{\Sigma}_S$ and $\widetilde{\Sigma}_T$ replaced by $\Sigma_{S,\infty}$ and $\Sigma_{T,\infty}$ in definitions of $\widetilde{\alpha}_S$ and $\widetilde{\alpha}_T$. These analogous results hold under the given assumptions plus the additional conditions that are needed to obtain consistency of $\overline{\Sigma}_n$ for Σ_∞.

11.2 Fixed Alternatives

We now consider strong IV-fixed alternative (SIV-FA) asymptotics. This asymptotic framework determines the *consistency*, or lack thereof of, of tests. For SIV-FA asymptotics, we assume:

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.

Let

$$\lambda_{FA} = \pi' D_Z \pi, \tag{11.3}$$

where D_Z is defined in (9.4). Define

$$\begin{aligned}
\widetilde{\varphi}_S &= \widetilde{\Sigma}_S^{-1/2} D_Z \pi (\beta - \beta_0), \\
\widetilde{\varphi}_T &= \widetilde{\Sigma}_T^{-1/2} \left(D_Z \pi a' \Omega^{-1} a_0 - \widetilde{\Sigma}_{TS} \widetilde{\Sigma}_S^{-1/2} \widetilde{\varphi}_S \right), \text{ and} \\
\varsigma_k &\sim N(0, I_k).
\end{aligned} \tag{11.4}$$

We have the following basic results.

Lemma 10 (a) *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) $(\widehat{S}_n/n^{1/2}, \widehat{T}_n/n^{1/2}) = (S_n/n^{1/2}, T_n/n^{1/2}) + o_p(1)$, (iii) $(\widehat{Q}_{S,n}/n, \widehat{Q}_{ST,n}/n,\ \widehat{Q}_{T,n}/n) \to_p (c_\beta^2 \lambda_{FA}, c_\beta d_\beta \lambda_{FA}, d_\beta^2 \lambda_{FA})$, *and* (iv) *if* $\beta = \beta_{AR}$ *and Assumption* 4 *also holds, then* $T_n \to_d \varsigma_k$, $\widehat{T}_n = T_n + o_p(1)$, *and* $(\widehat{Q}_{S,n}/n, \widehat{Q}_{ST,n}/n^{1/2}, \widehat{Q}_{T,n}) \to_d (c_\beta^2 \lambda_{FA}, c_\beta \pi' D_Z^{1/2} \varsigma_k, \varsigma_k' \varsigma_k)$ *as* $n \to \infty$.

(b) *Under Assumptions SIV-FA* 1-3, 5, *and* 6, (i) $\widetilde{\Sigma}_{S,n} \to_p \widetilde{\Sigma}_S$, $\widetilde{\Sigma}_{TS,n} \to_p \widetilde{\Sigma}_{TS}$, *and* $\widetilde{\Sigma}_{T,n} \to_p \widetilde{\Sigma}_T$, (ii) $(\widetilde{S}_n/n^{1/2}, \widetilde{T}_n/n^{1/2}) \to_p (\widetilde{\varphi}_S, \widetilde{\varphi}_T)$, *and* (iii) $(\widetilde{Q}_{S,n}/n, \widetilde{Q}_{ST,n}/n, \widetilde{Q}_{T,n}/n) \to_d (\widetilde{\varphi}_S' \widetilde{\varphi}_S, \widetilde{\varphi}_S' \widetilde{\varphi}_T, \widetilde{\varphi}_T' \widetilde{\varphi}_T)$ *as* $n \to \infty$.

Using Lemma 10, we determine the asymptotic behavior under SIV-FA asymptotics of the AR, LM, LR, POIS1, POIS2, and WAP test statistics and their HR versions.

Theorem 13 (a) *Under Assumptions SIV-FA and* 1-3, (i) $\widehat{AR}_n/n = AR_n/n + o_p(1) \to_p c_\beta^2 \lambda_{FA} > 0$, (ii) $\widehat{LM}_n/n = LM_n/n + o_p(1) \to_p c_\beta^2 \lambda_{FA} > 0$ *provided* $\beta \neq \beta_{AR}$, (iii) $\widehat{LR}_n/n = LR_n/n + o_p(1) \to_p c_\beta^2 \lambda_{FA} > 0$, (iv) $(\widehat{POIS1}_{\widehat{\delta}} \cdot (2k + \widehat{\delta}^2)^{1/2} + k)/n = (POIS1_\delta \cdot (2k + \delta^2)^{1/2} + k)/n + o_p(1) \to_p \lambda_{FA}(c_\beta^2 + 2(d_{\beta^*}/c_{\beta^*}) c_\beta d_\beta)$, *where* β^* *is the alternative against which* $POIS1_\delta$ *is POI,* (v) *if* $\beta \neq \beta_{AR}$ *and Assumption* 4 *also* *holds, then* $\widehat{LM}_n/n = LM_n/n + o_p(1) \to_d (c_\beta \pi' D_Z^{1/2} \varsigma_k)^2/\varsigma_k' \varsigma_k$ $(\neq 0$ *a.s.), (vi) if* W_{2P} *satisfies* (6.2), *then* $LR_{W_{2P}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) \to_p \infty$, *and* (vii) $LR_{W_{AE}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) \to_p \infty$.

(b) *Under Assumptions SIV-FA,* 1-3, 5, *and* 6, (i) $\widetilde{AR}_n/n \to_p \widetilde{\varphi}_S' \widetilde{\varphi}_S > 0$, (ii) $\widetilde{LM}_n/n \to_p (\widetilde{\varphi}_S' \widetilde{\varphi}_T)^2/\widetilde{\varphi}_T' \widetilde{\varphi}_T > 0$ *provided* $\beta \neq \beta_{AR}$, (iii) $2\widetilde{LR}_n/n \to_p \widetilde{\varphi}_S' \widetilde{\varphi}_S - \widetilde{\varphi}_T' \widetilde{\varphi}_T + ((\widetilde{\varphi}_S' \widetilde{\varphi}_S + \widetilde{\varphi}_T' \widetilde{\varphi}_T)^2 - 4(\widetilde{\varphi}_S' \widetilde{\varphi}_S \widetilde{\varphi}_T' \widetilde{\varphi}_T - (\widetilde{\varphi}_S' \widetilde{\varphi}_T)^2))^{1/2}$, *and* (iv) $(\widetilde{POIS1}_{\widetilde{\delta}} \cdot (2k + \widetilde{\delta}^2)^{1/2} + k)/n \to_p \widetilde{\varphi}_S' \widetilde{\varphi}_S + 2(d_{\beta^*}/c_{\beta^*}) \widetilde{\varphi}_S' \widetilde{\varphi}_T$.

Comments. 1. The results of part (a) of the Theorem establish the consistency against any alternative $\beta \neq \beta_0$ of the tests based on $\widehat{AR}_n$, $\widehat{LM}_n$, $\widehat{LR}_n$, $LR_{W_{2P}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n})$ (provided W_{2P} is AE), and $LR_{W_{AE}}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n})$. (This makes use of the fact that the critical values of these tests are either constants or converge in probability to constants as $n \to \infty$, see comments in Section 11.1 regarding this.)

2. The result of Theorem 13(a)(iv) indicates that the one-sided POIS test, $\widehat{POIS1}_{\widehat{\delta}}$, is consistent against an unusual array of alternatives. First, if the alternative for which the test is designed equals the true value, i.e., $\beta^* = \beta$, then the POIS1 test is consistent because $c_\beta^2 + 2(d_{\beta^*}/c_{\beta^*}) c_\beta d_\beta = c_\beta^2 + 2d_\beta^2 > 0$. Second, the POIS1 test may fail to be consistent against alternatives on the same side of the null as β^* yet be consistent against alternatives on the other ("wrong") side of the null from β^*.

For example, suppose the alternative is $H_1 : \beta > \beta_0$, $\beta^* > \beta_0$, $\beta_0 = 0$, and $\omega_{12} > 0$. Then, $c_\beta > 0$ and $c_{\beta^*} > 0$. In addition, d_β is a linear function of β with $d_\beta > 0$ for all $\beta < \beta_{AR}$, and $d_\beta < 0$ for $\beta > \beta_{AR}$, where $\beta_{AR} = \omega_{22}/\omega_{12} > 0$, and likewise with β^* in place of β. Hence, if $0 < \beta^* < \beta_{AR}$, then the POIS1 test is inconsistent against all alternatives for which β is sufficiently large. This holds because the second term of $c_\beta^2 + 2(d_{\beta^*}/c_{\beta^*}) c_\beta d_\beta$ has negative sign and is arbitrarily large for β large. In addition, if $0 < \beta < \beta_{AR}$, then the POIS1 test is inconsistent for all sufficiently large values of β^*. These results are borne out in the power curves of Figure 3.

On the other hand, POIS1 tests with $\beta^* > \beta_{AR} > 0$ are consistent for all values $\beta < 0$ in the example. For all $\beta < 0$, $c_\beta < 0$ and $d_\beta > 0$. For $\beta^* > \beta_{AR}$, $c_{\beta^*} > 0$ and $d_{\beta^*} < 0$. Hence, $c_\beta^2 + 2(d_{\beta^*}/c_{\beta^*}) c_\beta d_\beta > 0$. That is, POIS1 tests designed for $\beta > \beta_{AR} > 0$ are consistent when $\beta < 0$.

3. The results of part (b) of the Theorem show that the HR tests $\widetilde{AR}_n$ and $\widetilde{LM}_n$ are consistent against any alternative $\beta \neq \beta_0$ and the HR tests $\widetilde{LR}_n$ and $\widetilde{POIS1}_{\widetilde{\delta}}$ are consistent against any alternative for which the limit value given in the Theorem is non-zero.

12 Numerical Results II: Model with Unknown Covariance Matrix

This section summarizes the results of a Monte Carlo study of the finite-sample rejection rates of selected two-sided tests for the case that Ω is unknown—the tests based on $\widehat{LR}_n$, $\widehat{LM}_n$, $\widehat{AR}_n$, and $\widehat{P}_n^*$ (where the $\widehat{P}_n^*$ is the P^* test with Ω replaced by

$\widehat{\Omega}_n$). The model considered is the same as in Section 8 (where as above $\omega_{11} = \omega_{22} = 1$ wlog) except that Ω is unknown and both equations include an intercept (so X is a column of 1's). The IV's are taken to be iid standard normal random variables. All results are for $\beta_0 = 0$ (where as above this choice is wlog). We consider several values of (i) the distance of the true parameter from the null, $\beta\sqrt{\lambda}$, viz., 0.0, -2.0, 2.0, (ii) the strength of IV's, λ, viz., 5, 20 (which corresponds to $\lambda/k = 1.0, 4.0$ when $k = 5$), (iii) the sample size, n, viz., 50, 100, 200, and ∞ (i.e., the weak IV asymptotic limit), and (iv) the number of IV's, k, viz., 2, 5, and 10. All results are based on 5,000 Monte Carlo simulations.

The results are summarized in Table 1. The first column reports $\beta\sqrt{\lambda}$; the second column reports λ; and the third column reports n. The remaining columns report the rejection rates under the specified true values of β, λ, and n. When $\beta = 0$, the entries correspond to the size of the test and for $\beta \neq 0$ the entries are (size-unadjusted) power. To obtain more accurate size-unadjusted power comparisons, we use the asymptotic χ^2 critical values for the $\widehat{AR}_n$ test rather than the exact F critical values.

The results (and additional unreported results) suggest four general conclusions. First, for $n \geq 100$, the size typically is very close to .05 with the largest deviation being .019. Second, the four tests have comparable size distortions for very small n. Third, the size is better controlled when the estimator of Ω is adjusted for degrees of freedom (results with no df adjustment are not reported here).[14] Fourth, for $n \geq 100$, the rejection rates are close to the asymptotic powers of the tests. Taken together, these findings suggest that a sample size of 100 is sufficient for the weak-IV asymptotic results to provide reliable guides to the sampling distribution of these statistics uniformly in λ, both for size and power.

13 Normal Model with Multiple Endogenous Variables and Known Covariance Matrix

In this section, we consider a generalization of the model considered in Sections 2-8 to the case where m endogenous variables appear. We assume that $m \leq k$ (where k is the number of instrumental variables, i.e., the number of columns of Z). In particular, we consider the model as specified in (2.1)-(2.5), but with

$$\begin{aligned} y_2, v_2 &\in R^{n\times m}; \beta \in R^m; \pi \in R^{k\times m}; \xi_1, \xi \in R^{p\times m}; \eta \in R^{p\times(2m)}; \\ \Omega &\in R^{m\times m}; Y, V \in R^{n\times(m+1)}; \\ \theta &= (\beta', vec(\pi)', vec(\gamma)', vec(\xi)')' \in R^{m+km+2pm};\ \text{and} \\ a &= [\beta : I_m] \in R^{m\times(m+1)}. \end{aligned} \tag{13.1}$$

The known $(m+1)\times(m+1)$ covariance matrix Ω is assumed to be nonsingular. The parameter space for $\theta = (\beta, \pi', \gamma', \xi')'$ is taken to be $R^{m+km+2pm}$.

[14] Larger degrees of freedom adjustments than $n - k - p$ appear to improve the size results. This is a topic of ongoing research.

The null hypothesis is

$$H_0 : \beta = \beta_0 \text{ for some } \beta_0 \in R^m. \tag{13.2}$$

The alternative hypothesis can be two-sided $H_1 : \beta \neq \beta$, multivariate one-sided $H_1 : \beta < \beta_0$ or $H_1 : \beta > \beta_0$, , or $H_1 : \beta \in B$ for any subset B of R^m that does not include β_0.

As in the case where $m = 1$, low dimensional sufficient statistics are available for θ and the sub-vector $(\beta, \pi')'$:

Lemma 11 *For the model in* (2.5) *generalized as in* (13.1)*,*
(a) $Z'Y$ *and* $X'Y$ *are sufficient statistics for* θ*,*
(b) $Z'Y$ *and* $X'Y$ *are independent,*
(c) $X'Y$ *has a multivariate normal distribution that does not depend on* $(\beta', vec(\pi)')'$*,*
(c) $Z'Y$ *has a multivariate normal distribution that does not depend on* $\eta = [\gamma : \xi]$*, and*
(d) $Z'Y$ *is a sufficient statistic for* $(\beta', vec(\pi)')'$*.*

As when $m = 1$, given our interest in tests concerning β, we base tests on the sufficient statistic $Z'Y \in R^{k \times m}$ for $(\beta', vec(\pi)')'$. (This is done without loss of attainable power.) We consider a one-to-one transformation of $Z'Y$ that yields (i) the first column to be independent of the nuisance parameter π under H_0; (ii) independence of the m transformed columns under the null and alternative; (iii) independence across rows of each transformed column; and (iv) unit variance for all transformed elements. Define

$$\begin{aligned} S &= (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0' \Omega b_0)^{-1/2} \in R^k \text{ and} \\ T_j &= (Z'Z)^{-1/2} Z'Y \Omega^{-1} \alpha_{0,j} \in R^k, \text{ for } j = 1, ..., m, \\ T &= [T_1 : \cdots : T_m] = (Z'Z)^{-1/2} Z'Y \Omega^{-1} \alpha_0 \in R^{k \times m}, \text{ where} \\ b_0 &= (1, -\beta_0')', \; \alpha_0 = [\alpha_{0,1} : \cdots : \alpha_{0,m}], \end{aligned} \tag{13.3}$$

and $\alpha_{0,1}, ..., \alpha_{0,m}$ are defined as follows. For conditions (ii)-(iv) to hold, it turns out that $\alpha_{0,j}$ must satisfy $b_0' \alpha_{0,j} = 0$ and $\alpha_{0,j}' \Omega^{-1} \alpha_{0,j} = 1$ for all $j = 1, ..., m$ and $\alpha_{0,j}' \Omega^{-1} \alpha_{0,\ell} = 0$ for all $j, \ell = 1, ..., m$ with $j \neq \ell$. These conditions are satisfied by constructing $\{\alpha_{0,j} : j = 1, ..., m\}$ using a Gram-Schmidt-*like* orthogonalization scheme applied to the linearly independent $(m+1)$-vectors $\{b_0, e_2, ..., e_{m+1}\}$, where e_j is the j-th elementary $(m+1)$-vector for $j = 2, ..., m+1$. Let

$$\begin{aligned} \alpha_{0,1} &= M_{b_0} e_2 / ||\Omega^{-1/2} M_{b_0} e_2||, \\ \alpha_{0,2} &= M_{[b_0 : \Omega^{-1}\alpha_{0,1}]} e_3 / ||\Omega^{-1/2} M_{[b_0 : \Omega^{-1}\alpha_{0,1}]} e_3||, \\ &\vdots \\ \alpha_{0,m} &= M_{[b_0 : \Omega^{-1}\alpha_{0,1} : \cdots : \Omega^{-1}\alpha_{0,m-1}]} e_{m+1} / ||\Omega^{-1/2} M_{[b_0 : \Omega^{-1}\alpha_{0,1} : \cdots : \Omega^{-1}\alpha_{0,m-1}]} e_{m+1}||, \end{aligned} \tag{13.4}$$

where as above $M_A = I - A(A'A)^{-1}A'$ for any matrix A.

Some algebra shows that when $m = 1$ we obtain $\alpha_{0,1} = a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}$, where a_0 is defined in (2.6). Thus, T of Section 2 is the same as T defined in (13.4) when $m = 1$.

The means of S and T_j for $j = 1, ..., m$ depend on

$$\mu_\pi = (Z'Z)^{1/2}\pi \in R^{k\times m}. \tag{13.5}$$

The distributions of the sufficient statistics $\{S, T_1, ..., T_m\}$ for the parameters $(\beta', vec(\pi)')'$ are given in the following lemma.

Lemma 12 *For the model in* (2.5) *generalized as in* (13.1),
(a) $S \sim N(\mu_\pi(\beta - \beta_0) \cdot (b_0'\Omega b_0)^{-1/2},\ I_k)$,
(b) $T_j \sim N(\mu_\pi a'\Omega^{-1}\alpha_{0,j},\ I_k)$ *for* $j = 1, ..., m$, *and*
(c) S, $T_1, ..., T_m$ *are mutually independent.*

Comments: 1. Under H_0, S has mean zero.

2. Minus two times the log-likelihood function for π based on the normal density of T is a constant plus

$$\sum_{j=1}^{m}(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j})'(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j})$$

$$= tr\left(\sum_{j=1}^{m}(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j})(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j})'\right).$$

Consequently, the T statistic can be written as $(Z'Z)^{1/2}\widehat{\pi}_0 a_0'\Omega^{-1}\alpha_0$, where $\widehat{\pi}_0$ denotes the maximum likelihood estimator of π under H_0 and $a_0 = [\beta_0 : I_m] \in R^{m\times(m+1)}$, $\widetilde{\pi} = (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_0$, where $a_0 = [\beta_0 : I_m]$.

Next, we consider the same groups of transformations G and $\overline{G}$ defined in (3.1) when $m \geq 2$ as when $m = 1$ (except that $x \in R^{k\times(m+1)}$ in the definition of G rather than $x \in R^{k\times 2}$). An *invariant* test, $\phi(S, T)$, under the group G is one for which $\phi(FS, FT) = \phi(S, T)$ for all $k \times k$ orthogonal matrices F. It suffices to restrict attention to the class of tests that depend only on a maximal invariant.

Define Q, Q_S, Q_{ST}, Q_T, and Q_1 as in (3.2), but with $T = [T_1 : \cdots : T_m]$. Hence, $Q = [S:T]'[S:T] \in R^{(m+1)(m+1)}$, $Q_S = S'S \in R$, $Q_{ST} = S'T \in R^m$, $Q_T = T'T \in R^{m\times m}$, and $Q_1 = (S'S, S'T)' \in R^{m+1}$.

Theorem 14 *The* $(m+1)\times(m+1)$ *matrix* Q *is a maximal invariant for the transformations* G.

Comments: 1. As in the model with one endogenous variable, when $m \geq 2$ the statistic Q has a non-central Wishart distribution because $[S:T]$ is a multivariate normal matrix that has independent rows and common covariance matrix across rows. The distribution of Q depends on π only through the positive definite (pd) matrix λ defined by

$$\lambda = \pi' Z' Z \pi \in R^{m \times m}. \tag{13.6}$$

In consequence, the utilization of invariance has reduced the km dimensional nuisance parameter $vec(\pi)$ to the $m \times m$ symmetric matrix nuisance parameter λ, which has $m(m+1)/2$ non-redundant elements. This is true both under the null and under the alternative. For example, if $k = 5$ and $m = 2$, then the reduction is from 10 nuisance parameters to 3 nuisance parameters.

2. Examples of invariant tests in the literature include the AR test, the LM test of Kleibergen (2002) and Moreira (2001), and the CLR test of Moreira (2003). The AR and LM tests depend on Q or (S, T) in the following ways:

$$\begin{aligned} AR &= Q_S \ = S'S, \\ LM &= Q_{ST} Q_T^{-1} Q_{ST}' = S'T(T'T)^{-1} T'S. \end{aligned} \tag{13.7}$$

Invariant similar tests are characterized as follows:

Theorem 15 *An invariant test $\phi(Q)$ is similar with significance level α if and only if $E_{\beta_0}(\phi(Q)|Q_T = q_T) = \alpha$ for almost all q_T, where $E_{\beta_0}(\cdot|Q_T = q_T)$ denotes conditional expectation given $Q_T = q_T$ when $\beta = \beta_0$ (which does not depend on π).*

Comment: The two tests in (13.7) are invariant similar tests. Hence, they satisfy the property specified in the theorem.

Let W be a weight function over (β, λ) values. That is, W is a probability distribution on the product of R^m and the space of pd $m \times m$ matrices, call it $R_{pd}^{m \times m}$. Weighted average power of a test $\phi(Q)$ with respect to W is given by (4.1). The expressions in (4.2)-(4.8) hold when $m \geq 2$ just as when $m = 1$, provided one adjusts the range of integration suitably. In particular, the integral over (β, λ) values is over $R^m \times R_{pd}^{m \times m}$, rather than $R \times R^+$, and the integral over (q_1, q_T) values is over $(R^+ \times R^m) \times R_{pd}^{m \times m}$, rather than $(R^+ \times R) \times R^+$. In particular, the optimal WAP LR statistic $LR_W(Q_1, Q_T)$ is as given in (4.8).

As in Section 4.2, in order to provide an explicit expression for the optimal WAP LR statistic $LR_W(Q_1, Q_T)$, we determine the densities $f_Q(q; \beta, \lambda)$, $f_{Q_T}(q_T; \beta, \lambda)$, and $f_{Q_1|Q_T}(q_1, q_T; \beta_0)$ that arise in (4.2), (4.7), and (4.8). Let

$$\begin{aligned} \Delta_\beta &= [\beta - \beta_0 : a'\Omega^{-1}\alpha_0] \in R^{m \times (m+1)} \text{ and} \\ \Delta_{T,\beta} &= a'\Omega^{-1}\alpha_0 = [\beta : I_m]\Omega^{-1}\alpha_0 \in R^{m \times m}. \end{aligned} \tag{13.8}$$

Note that $tr(\Delta_{\beta_0}' \lambda \Delta_{\beta_0}) = tr(\Delta_{T,\beta_0}' \lambda \Delta_{T,\beta_0})$. Let $etr(A)$ denote $\exp(tr(A))$ for a matrix A.

Lemma 13 (a) *The density of Q at $q \in R_{pd}^{(m+1)\times(m+1)}$ is a non-central Wishart density with k degrees of freedom, covariance matrix I_{m+1}, and non-centrality matrix (i.e., means sigma matrix) $\Delta_\beta' \lambda \Delta_\beta$:*

$$f_Q(q;\beta,\lambda) = K_{1,m} etr(-\Delta_\beta'\lambda\Delta_\beta/2)|q|^{(k-m-2)/2} etr(-q/2)\ {}_0F_1\left(k/2; \Delta_\beta'\lambda\Delta_\beta q/4)\right),$$

where $q \in R^{(m+1)\times(m+1)}$,

$$K_{1,m}^{-1} = 2^{k(m+1)/2}\Gamma_{m+1}(k/2),$$

${}_0F_1(\cdot;\cdot)$ denotes a hypergeometric function with matrix argument, and $\Gamma_{m+1}(k/2)$ denotes the multivariate gamma function.

(b) *The density of Q_T at $q_T \in R_{pd}^{m\times m}$ is a non-central Wishart density with k degrees of freedom, covariance matrix I_m, and noncentrality parameter $\Delta_{T,\beta}'\lambda\Delta_{T,\beta}$:*

$$f_{Q_T}(q_T;\beta,\lambda) = K_{2,m} etr(-\Delta_{T,\beta}'\lambda\Delta_{T,\beta}/2)|q_T|^{(k-m-1)/2} \times etr(-q_T/2)\ {}_0F_1\left(k/2; \Delta_{T,\beta}'\lambda\Delta_{T,\beta}q_T/4)\right),$$

where $q_T \in R^{m\times m}$ and

$$K_{2,m}^{-1} = 2^{km/2}\Gamma_m(k/2).$$

(c) *Under the null hypothesis, the conditional density of Q_1 given $Q_T = q_T$ is*

$$f_{Q_1|Q_T}(q_1|q_T;\beta_0) = K_{1,m}K_{2,m}^{-1}|q|^{(k-m-2)/2}|q_T|^{-(k-m-1)/2} etr(-q_S/2)$$

Comments: 1. Hypergeometric functions of matrix argument are defined in Muirhead (1982, p. 258). They involve series of zonal polynomials.

2. The multivariate gamma function at $k/2$, $\Gamma_{m+1}(k/2)$, can be written in terms of the ordinary gamma function as follows: $\Gamma_{m+1}(k/2) = pi^{k(k-2)/16}\prod_{j=1}^{k/2}\Gamma((k-j+1)/2)$, e.g., see Muirhead (1982, Thm. 2.1.12, p. 62), where $pi = 3.1415...$ The test statistics considered below do not depend on $\Gamma_{m+1}(k/2)$, however, so computation is not an issue.

3. When $m = 2$ alternative expressions for the densities in parts (a)-(c) of the lemma are available in Anderson (1946, eqn. (7)), which are easier to compute. These expressions are in terms of the modified Bessel function of the first kind.

Equations (4.2), (4.7), and (4.8) and Lemma 13 combine to give the following result.

Corollary 6 *The optimal WAP test statistic for weight function W is given by*

$$LR_W(q_1,q_T) = \frac{\int f_{Q_1,Q_T}(q_1,q_T;\beta,\lambda)dW(\beta,\lambda)}{\int f_{Q_T}(q_T;\beta,\lambda)dW(\beta,\lambda) f_{Q_1|Q_T}(q_1|q_T;\beta_0,\lambda)} = \frac{\psi_W(q_1,q_T)}{\psi_{2,W}(q_T)},$$

where

$$\psi_W(q_1,q_T) = \int etr(-\Delta_\beta'\lambda\Delta_\beta/2)\ {}_0F_1\left(k/2; \Delta_\beta'\lambda\Delta_\beta q/4)\right) dW(\beta,\lambda),$$

$$\psi_{2,W}(q_T) = \int etr(-\Delta'_{T,\beta}\lambda\Delta_{T,\beta}/2)\ {}_0F_1\left(k/2; \Delta'_{T,\beta}\lambda_0\Delta_{T,\beta}q_T/4)\right) dW(\beta,\lambda),$$

the integrals are over $(\beta, \lambda) \in R^m \times R^{m\times m}_{pd}$, *and* Δ_β *and* $\Delta_{T,\beta}$ *are defined in* (13.8).

Comments: 1. As when $m = 1$, $\psi_W(q_1, q_T)$ does **not** equal $\int f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)\, dW(\beta, \lambda)$ and likewise with $\psi_{2,W}(q_T)$. This is because numerous cancellations occur in the second expression in the first line of the Corollary 6, including the constants $K_{1,m}$ and $K_{2,m}$.

2. When $m = 2$, the density formulae given in Comment 3 following Lemma 13 yield alternative expressions for $\psi_W(q_1, q_T)$ and $\psi_{2,W}(q_T)$ that are easier to compute.

Because $\psi_{2,W}(q_T)$ does not depend on q_1, it could be absorbed into the conditional critical value given $Q_T = q_T$. But, as above, for reasons of numerical stability, we recommend obtaining critical values for the equivalent test statistic $\ln(LR_W(q_1, q_T))$.

The test that maximizes WAP among invariant similar tests with significance level α rejects H_0 if

$$LR_W(Q_1, Q_T) > \kappa_\alpha(Q_T), \tag{13.9}$$

where $\kappa_\alpha(Q_T)$ is defined such that the test is similar. That is, $\kappa_\alpha(q_T)$ is defined by

$$P_{\beta_0}(LR_W(Q_1, q_T) > \kappa_\alpha(q_T)|Q_T = q_T) = \alpha, \tag{13.10}$$

where $P_{\beta_0}(\cdot|Q_T = q_T)$ denotes conditional probability given $Q_T = q_T$ under the null, which can be calculated using the density in Lemma 3(c).

The results of this section are summarized as follows:

Theorem 16 *The test that rejects* H_0 *when* $LR_W(Q_1, Q_T) > \kappa_\alpha(Q_T)$ *maximizes WAP for the weight function* W *over all level* α *invariant similar tests.*

14 Appendix of Proofs

14.1 Proofs of Results Stated in Section 2

Proof of Lemma 1. Let $Z = [Z_1 : \cdots : Z_n]'$ and $X = [X_1 : \cdots : X_n]'$. The distribution of Y is multivariate normal with

$$EY = Z\pi a' + X\eta, \tag{14.1}$$

independence across rows, and covariance matrix Ω for each row. Hence, the density of Y evaluated at the $n \times 2$ matrix $y = [y_1 : \cdots : y_n]'$ is

$$\begin{aligned}
&(2\pi)^{-n/2}|\Omega|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (y_i - a\pi' Z_i - \eta' X_i)' \Omega^{-1} (y_i - a\pi' Z_i - \eta' X_i) \right) \\
&= (2\pi)^{-n/2}|\Omega|^{-n/2} \exp\left(-\frac{1}{2} \left[\sum_{i=1}^{n} y_i' \Omega^{-1} y_i - 2\pi' (\sum_{i=1}^{n} Z_i y_i') \Omega^{-1} a \right.\right. \\
&\quad \left.\left. -2tr((\sum_{i=1}^{n} X_i y_i') \Omega^{-1} \eta') + \sum_{i=1}^{n} (a\pi' Z_i - \eta' X_i)' \Omega^{-1} (a\pi' Z_i - \eta' X_i) \right] \right).
\end{aligned} \tag{14.2}$$

If a density can be factorized as $p_\theta(x) = f_\theta(T(x))h(x)$, then $T(X)$ is a sufficient statistic for θ. In consequence, given that Ω is known, Z_i and X_i are fixed and known, $a = (\beta, 1)'$, and $\eta = [\gamma : \xi]$, sufficient statistics for $\theta = (\beta, \pi', \gamma', \xi')'$ are $\sum_{i=1}^{n} Z_i Y_i' = Z'Y$ and $\sum_{i=1}^{n} X_i Y_i' = X'Y$ and part (a) of the lemma holds.

To prove part (b) of the lemma, note that $Z'Y$ and $X'Y$ are (jointly) multivariate normal random matrices and $Z'X = 0$. For any $m_1, m_2 \in R^2$, we have

$$\begin{aligned}
cov(Z'Ym_1, X'Ym_2) &= cov(\sum_{i=1}^{n} Z_i Y_i' m_1, \sum_{i=1}^{n} X_i Y_i' m_2) \\
&= \sum_{i=1}^{n} Z_i X_i' cov(Y_i' m_1, Y_i' m_2) = Z'X \cdot m_1' \Omega m_2 = 0,
\end{aligned} \tag{14.3}$$

where the second equality uses independence across i and the third equality uses the assumption that the covariance matrix Ω of Y_i does not depend on i. Hence, $Z'Y$ and $X'Y$ are independent.

The distribution of $X'Y$ is multivariate normal with variances and covariances that depend on X and Ω, but not on θ, and with mean

$$X'EY = X'(Z\pi a' + X\eta) = X'X\eta \tag{14.4}$$

because $X'Z = 0$. Hence, the distribution of $X'Y$ does not depend on (β, π) and part (c) of the lemma holds.

The distribution of $Z'Y$ is multivariate normal with variances and covariances that depend on Z and Ω, but not on θ, and with mean

$$Z'EY = Z'(Z\pi a' + X\eta) = Z'Z\pi a' \tag{14.5}$$

because $Z'X = 0$. Hence, the distribution of $Z'Y$ does not depend on (γ, ξ) and part (d) of the lemma holds.

Part (e) of the lemma follows from parts (b)-(d). $\square$

Proof of Lemma 2. The k-vector S is multivariate normal with mean

$$\begin{aligned} ES &= (Z'Z)^{-1/2} Z'EY b_0 \cdot (b_0' \Omega b_0)^{-1/2} \\ &= (Z'Z)^{-1/2} Z'(Z\pi a' + X\eta) b_0 \cdot (b_0' \Omega b_0)^{-1/2} = c_\beta \mu_\pi \end{aligned} \tag{14.6}$$

using (14.1), $Z'X = 0$, and $a'\beta_0 = \beta - \beta_0$. We have

$$var(Z'Yb_0) = var(\sum_{i=1}^n Z_i Y_i' b_0) = \sum_{i=1}^n Z_i Z_i' var(Y_i' b_0) = \sum_{i=1}^n Z_i Z_i' b_0' \Omega b_0 = Z'Z b_0' \Omega b_0. \tag{14.7}$$

Hence, from the definition of S, $var(S) = I_k$ and part (a) of the lemma holds.

The k-vector T is multivariate normal with mean

$$\begin{aligned} ET &= (Z'Z)^{-1/2} Z'Y \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2} \\ &= (Z'Z)^{-1/2} Z'(Z\pi a' + X\eta) \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2} = d_\beta \mu_\pi. \end{aligned} \tag{14.8}$$

>From (14.7) with b_0 replaced by $\Omega^{-1} a_0$, we have $var(Z'Y\Omega^{-1}a_0) = Z'Za_0'\Omega^{-1}a_0$. Hence, from the definition of T, $var(T) = I_k$ and part (b) of the lemma holds.

The random vectors S and T are independent because they are non-stochastic functions of $Z'Yb_0$ and $Z'Y\Omega^{-1}a_0$, respectively, and the latter are jointly multivariate normal with covariance given by

$$\begin{aligned} & cov(Z'Yb_0, Z'Y\Omega^{-1}a_0) = cov(\sum_{i=1}^n Z_i Y_i' b_0, \sum_{i=1}^n Z_i Y_i' \Omega^{-1} a_0) \\ &= \sum_{i=1}^n Z_i Z_i' cov(Y_i' b_0, Y_i' \Omega^{-1} a_0) = \sum_{i=1}^n Z_i Z_i' b_0' \Omega \Omega^{-1} a_0 = 0, \end{aligned} \tag{14.9}$$

using $b_0' a_0 = 0$. Hence, part (c) of the lemma holds. $\square$

14.2 Proofs of Results Stated in Section 3

Proof of Theorem 1. Let $M(S,T) = [S{:}T]'[S{:}T] = Q$. $M(S,T)$ is a maximal invariant if it is invariant and it takes different values on different *orbits* of G. Obviously, $M(S,T)$ is invariant. The latter condition holds if given any k-vectors $\mu_1, \mu_2, \widetilde{\mu}_1$, and $\widetilde{\mu}_2$ such that $M(\mu_1, \mu_2) = M(\widetilde{\mu}_1, \widetilde{\mu}_2)$ there exists an orthogonal $k \times k$ matrix $\overline{F}$ such that $\widetilde{\mu}_1 = \overline{F}\mu_1$ and $\widetilde{\mu}_2 = \overline{F}\mu_2$, e.g., see Lehmann (1986, eqn. (7), p. 285).

First, suppose μ_1 and μ_2 are linearly independent (which implies that $k \geq 2$). Then, there exist linearly independent k-vectors $\mu_3, ..., \mu_k$ such that $\{\mu_1, ..., \mu_k\}$ span R^k. Applying the Gram-Schmidt procedure to $\{\mu_1, ..., \mu_k\}$, we now construct an orthogonal matrix F such that $F\mu_1$ and $F\mu_2$ depend on (μ_1, μ_2) only through $\mu_1'\mu_1$,

$\mu_1'\mu_2$, and $\mu_2'\mu_2$. For a full column rank $k\times\ell$ matrix A, let $M_A = I_k - A(A'A)^{-1}A'$. We take $f_1 = \mu_1/||\mu_1||$, $f_2 = M_{\mu_1}\mu_2/||M_{\mu_1}\mu_2||, ...,$ $f_k = M_{[\mu_1:\cdots:\mu_{k-1}]}\mu_k/||M_{[\mu_1:\cdots:\mu_{k-1}]}\mu_k||$. Define $F = [f_1:\cdots:f_k]'$. We have

$$\begin{aligned} F\mu_1 &= (f_1'\mu_1, ..., f_k'\mu_1)' = (||\mu_1||, 0, ..., 0)' \text{ and} \\ F\mu_2 &= (\mu_1'\mu_2/||\mu_1||, \mu_2' M_{\mu_1}\mu_2/||M_{\mu_1}\mu_2||, 0, ..., 0)'. \end{aligned} \quad (14.10)$$

Because $\mu_2' M_{\mu_1}\mu_2 = \mu_2'\mu_2 - (\mu_1'\mu_2/||\mu_1||)^2$, we find that $F\mu_1$ and $F\mu_2$ depend on (μ_1, μ_2) only through $\mu_1'\mu_1$, $\mu_1'\mu_2$, and $\mu_2'\mu_2$.

Define $\widetilde{F}$ analogously to F but with $\{\widetilde{\mu}_1, ..., \widetilde{\mu}_k\}$ in place of $\{\mu_1, ..., \mu_k\}$. Then, $\widetilde{F}\widetilde{\mu}_1$ and $\widetilde{F}\widetilde{\mu}_2$ depend on $(\widetilde{\mu}_1, \widetilde{\mu}_2)$ only through $\widetilde{\mu}_1'\widetilde{\mu}_1$, $\widetilde{\mu}_1'\widetilde{\mu}_2$, and $\widetilde{\mu}_2'\widetilde{\mu}_2$.

Now, suppose (μ_1, μ_2) and $(\widetilde{\mu}_1, \widetilde{\mu}_2)$ are such that $M(\mu_1, \mu_2) = M(\widetilde{\mu}_1, \widetilde{\mu}_2)$. That is, $\mu_1'\mu_1 = \widetilde{\mu}_1'\widetilde{\mu}_1$, $\mu_1'\mu_2 = \widetilde{\mu}_1'\widetilde{\mu}_2$, and $\mu_2'\mu_2 = \widetilde{\mu}_2'\widetilde{\mu}_2$. Then, the orthogonal matrices F and $\widetilde{F}$ are such that $F\mu_1 = (||\mu_1||, 0, ..., 0)' = (||\widetilde{\mu}_1||, 0, ..., 0)' = \widetilde{F}\widetilde{\mu}_1$ and $\widetilde{\mu}_1 = \widetilde{F}^{-1}F\mu_1 = \overline{F}\mu_1$, where $\overline{F} = \widetilde{F}^{-1}F$ is an orthogonal matrix. Similarly, $F\mu_2 = \widetilde{F}\widetilde{\mu}_2$ and $\widetilde{\mu}_2 = \widetilde{F}^{-1}F\mu_2 = \overline{F}\mu_2$. This completes the proof for the case where μ_1 and μ_2 are linearly independent.

Next, suppose μ_1 and μ_2 are linearly dependent (as necessarily occurs when $k = 1$). Then, we can ignore μ_2 and proceed as above using just μ_1 and some additional linearly independent vectors $\{\mu_2^*, ..., \mu_k^*\}$ for which $\{\mu_1, \mu_2^*, ..., \mu_k^*\}$ span R^k. The matrix $\overline{F}$ constructed in this way is such that if $M(\mu_1, \mu_2) = M(\widetilde{\mu}_1, \widetilde{\mu}_2)$, then $\widetilde{\mu}_1 = \overline{F}\mu_1$. In addition, because $\mu_2 = \kappa\mu_1$ and $\widetilde{\mu}_2 = \kappa\widetilde{\mu}_1$ for some κ, we obtain $\widetilde{\mu}_2 = \overline{F}\mu_2$. This completes the proof. $\square$

Derivation of the One-sided Likelihood Ratio Statistic

The log-likelihood function for known Ω with all parameters concentrated out except β is

$$\mathcal{L}(Y;\beta) = -\frac{n}{2}\ln\det(\Omega) - \frac{1}{2}\left(tr(\Omega^{-1}Y'Y) + R(\beta)\right), \quad (14.11)$$

e.g., see Moreira (2003, Appendix A). Hence, we have

$$LR1 = \sup_{\beta\geq\beta_0}\mathcal{L}(Y;\beta) - \mathcal{L}(Y;\beta_0) = R(\beta_0) - \inf_{\beta\geq\beta_0} R(\beta). \quad (14.12)$$

We now determine $\inf_{\beta\geq\beta_0} R(\beta)$. By definition, $\widehat{\beta} = \widehat{\beta}_{LIML-k}$ minimizes $\mathcal{L}(Y;\beta)$ over $\beta \in R$. Equivalently, $\widehat{\beta}$ minimizes $R(\beta)$ over $\beta \in R$. If $\widehat{\beta} \geq \beta_0$, then $\inf_{\beta\geq\beta_0} R(\beta) = R(\widehat{\beta}) = \inf_{\beta\in R} R(\beta)$ and $LR1 = R(\beta_0) - \inf_{\beta\in R} R(\beta) = LR$. If $\widehat{\beta} < \beta_0$, 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 pd weight matrices. Hence, the second equality in (3.7) holds.

Next, we show that $LR1$ only depends on the observations through Q. Let

$$J = \left[\frac{\Omega^{1/2}b_0}{\sqrt{b_0'\Omega b_0}} : \frac{\Omega^{-1/2}a_0}{\sqrt{a_0'\Omega^{-1}a_0}}\right]. \quad (14.13)$$

By algebra,

$$\begin{aligned}
[S:T] &= (Z'Z)^{-1/2}Z'Y\Omega^{-1/2}J, \\
Q &= [S:T]'[S:T] = J'\Omega^{-1/2}Y'P_ZY\Omega^{-1/2}J, \text{ and} \\
Y'P_ZY &= \Omega^{1/2}J'^{-1}QJ^{-1}\Omega^{1/2}.
\end{aligned} \tag{14.14}$$

Hence, $R(\beta)$ and $LR1$ only depend on the observations through Q.

It remains to provide an expression for $\widehat{\beta}_{LIML-k}$. The LIML-k estimator maximizes $\mathcal{L}(Y;\beta)$ or minimizes $R(\beta)$ over $\beta \in R$. We have

$$R(\beta) = \frac{\widetilde{b}'\Omega^{-1/2}Y'P_ZY\Omega^{-1/2}\widetilde{b}}{\widetilde{b}'\widetilde{b}} = \frac{\widetilde{b}'J'^{-1}QJ^{-1}\widetilde{b}}{\widetilde{b}'\widetilde{b}}, \text{ where } \widetilde{b} = \Omega^{1/2}b. \tag{14.15}$$

The minimum of the rhs is obtained by the eigenvector $\widetilde{b}^*$ that corresponds to the smallest eigenvalue of $J'^{-1}QJ^{-1}$. Hence,

$$\widehat{\beta}_{LIML-k} = -b_2^*/b_1^*, \text{ where } b^* = (b_1^*, b_2^*)' = \Omega^{-1/2}\widetilde{b}^*. \tag{14.16}$$

14.3 Proofs of Results Stated in Sections 4 and 5

Proof of Lemma 3. First, we prove part (a). The $k \times 2$ matrix $[S{:}T]$ is multivariate normal with mean matrix $M = \mu_\pi h_\beta'$, where $h_\beta = (c_\beta, d_\beta)'$, all variances equal to one, and all correlations equal to zero. Hence, $Q = [S{:}T]'[S{:}T]$ has a noncentral Wishart distribution with mean matrix of rank one and identity covariance matrix. By (6) of Anderson (1946), the density of Q at q is

$$\begin{aligned}
&K_1 \exp(-tr(M'M)/2)|q|^{(k-3)/2}\exp(-tr(q)/2) \\
&\times (tr(M'Mq))^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{tr(M'Mq)}\right).
\end{aligned} \tag{14.17}$$

We have $M'M = \lambda h_\beta h_\beta{}'$, where $\lambda = \mu_\pi'\mu_\pi$, $tr(M'M) = \lambda(c_\beta^2 + d_\beta^2)$, $tr(M'Mq) = \lambda h_\beta' q h_\beta$, and $h_\beta' q h_\beta = \xi_\beta(q)$. Hence, part (a) holds.

Part (b) holds because the distribution of Q_T is a noncentral chi-squared distribution with non-centrality parameter $d_\beta^2\lambda$ by Lemma 2(b) and (3.3). The stated form of the density is given in Anderson (1946, eqn. (6)).

Part (c) holds by calculating the ratio of the densities given in parts (a) and (b) of the lemma each evaluated at $\beta = \beta_0$ and using the fact that $c_{\beta_0} = 0$ and $\xi_{\beta_0}(q) = d_{\beta_0}^2 q_T$.

Part (d) holds because the null distribution of Q_S is a central chi-squared distribution with k degrees of freedom by Lemma 2(a) and $c_{\beta_0} = 0$.

For part (e), the null density of $\mathcal{S}_2$ is derived as follows: (i) $\mathcal{S}_2 = S'T/(||S|| \cdot ||T||)$ has the same distribution as $A = S'\alpha/||S||$ for any $\alpha \in R^k$ with $\alpha'\alpha = 1$ because $S \sim N(0, I_k)$ under the null and S and T are independent using Lemma 2(a) and (c), (ii) for $\alpha = (1, 0, ..., 0)'$, $(k-1)^{1/2}A/(1-A^2)^{1/2} = (k-1)^{1/2}S_1/(\sum_{j=2}^k S_j^2)^{1/2} \sim t_{k-1}$ by definition of the t_{k-1} distribution, and (iii) transformation of $(k-1)^{1/2}A/(1-A^2)^{1/2}$ to A gives the density in part (d), e.g., see Muirhead (1982, pf. of Thm. 1.5.7(i), pp. 38-9; eqn. (5), p. 147).

Next, we prove part (f). Under the null, $S \sim N(0, I_k)$, $T \sim N(d_{\beta_0}\mu_\pi, I_k)$, and S and T are independent by Lemma 2. Hence, $Q_S = S'S$ and T are independent. The distribution of $S'\alpha/||S||$ for $\alpha \in R^k$ with $\alpha'\alpha = 1$ does not depend on α by spherical symmetry of S. In consequence, the conditional distribution of $\mathcal{S}_2 = S'T/(||S|| \cdot ||T||)$ given $T = t$ does not depend on t and $\mathcal{S}_2$ is independent of T. Independence of $Q_S = S'S$ and $S'\alpha/||S||$ is a well-known result that holds by spherical symmetry of S. $\square$

Proof of Comment 7 to Corollary 2. The optimal test against β^* rejects if $\xi_{\beta^*}(Q_1, Q_T)$ is large and we have

$$\begin{aligned}
&\lim_{\beta^* \to \beta_0} \left(\xi_{\beta^*}(q_1, q_T) - d^2_{\beta^*} q_T \right) / (\beta^* - \beta_0) \\
&= \lim_{\beta^* \to \beta_0} \left((\beta^* - \beta_0)(b_0'\Omega b_0)^{-1} q_S + 2b^{*\prime}\Omega b_0 (b_0'\Omega b_0)^{-1} (\det(\Omega))^{-1/2} q_{ST} \right) \\
&= 2(\det(\Omega))^{-1/2} q_{ST}, \qquad (14.18)
\end{aligned}$$

where $b^* = (1, -\beta^*)'$. Hence, if $\beta^* - \beta_0 > 0$, the optimal test rejects when $Q_{ST} = S'T$ is large or, equivalently, when $Q_{ST}/Q_T^{1/2}$ is large since the critical value can depend on Q_T. The null distribution of $Q_{ST}/Q_T^{1/2}$ conditional on T or on Q_T is standard normal by Lemma 2, so the critical value for the test is the $1 - \alpha$ quantile, $\kappa_{\phi,\alpha}$, of the standard normal distribution. $\square$

Proof of Comment 9 to Corollary 2. Comment 9 holds because (i) the optimal test against β^* rejects if $\xi_{\beta^*}(Q_1, Q_T)$ is large, (ii) we have

$$\begin{aligned}
&\lim_{\beta^* \to \infty} \left(\xi_{\beta^*}(q_1, q_T) - d^2_{\beta^*} q_T \right) / c^2_{\beta^*} \\
&= \lim_{\beta^* \to \infty} \left(q_S + 2(d_{\beta^*}/c_{\beta^*}) q_{ST} \right) \\
&= q_S + 2(\det(\Omega))^{-1/2} (\beta_0 \omega_{22} - \omega_{12}) q_{ST}, \qquad (14.19)
\end{aligned}$$

and (iii) the limit as $\beta^* \to -\infty$ in (14.19) is the same as when $\beta^* \to \infty$. The second equality in (14.19) holds because

$$\begin{aligned}
(\det(\Omega))^{1/2} d_{\beta^*}/c_{\beta^*} &= \frac{b^{*\prime}\Omega b_0}{\beta^* - \beta_0} = \frac{\omega_{11} - (\beta^* + \beta_0)\omega_{12} + \beta^*\beta_0\omega_{22}}{\beta^* - \beta_0} \text{ and so} \\
\lim_{\beta^* \to \infty} d_{\beta^*}/c_{\beta^*} &= (\det(\Omega))^{-1/2} \left(\beta_0 \omega_{22} - \omega_{12} \right) \text{ and} \\
\lim_{\beta^* \to -\infty} d_{\beta^*}/c_{\beta^*} &= (\det(\Omega))^{-1/2} \left(\beta_0 \omega_{22} - \omega_{12} \right). \square \qquad (14.20)
\end{aligned}$$

14.4 Proofs of Results Stated in Section 6

Proof of Theorem 4. By continuity of the power function, which holds by Lehmann (1986, Thm. 9, p. 59), any unbiased test $\phi(Q)$ is similar. Hence, the first condition of the Theorem holds by Theorem 2.

Now, for a test to be unbiased, $(\partial/\partial\beta)E_{\beta,\lambda}\phi(Q_1,Q_T)|_{\beta=\beta_0}=0$ for all values of λ. By interchanging derivatives and integrals (which is justified by Lehmann (1989, Thm. 2.9, p. 59)) and the chain rule, the left-hand side of this equality equals I_1+I_2, where

$$\begin{aligned} I_1 &= \int\int \phi(q_1,q_T)\frac{\partial f_{Q_1|Q_T}(q_1,q_T;\beta_0,\lambda)}{\partial\beta}dq_1 f_{Q_T}(q_T;\beta_0,\lambda)dq_T \text{ and} \\ I_2 &= \int\int \phi(q_1,q_T) f_{Q_1|Q_T}(q_1,q_T;\beta_0)dq_1\frac{\partial f_{Q_T}(q_T;\beta_0,\lambda)}{\partial\beta}dq_T \\ &= \int \alpha\frac{\partial f_{Q_T}(q_T;\beta_0,\lambda)}{\partial\beta}dq_T = 0, \end{aligned} \tag{14.21}$$

where the second last equality holds by the condition for similarity and the last equality holds because $\int f_{Q_T}(q_T;\beta,\lambda)dq_T=1$ for all β.

To compute the derivative of the conditional density of Q_1 given $Q_T=q_T$ with respect to β evaluated at β_0, it is convenient to write the conditional density of Q_1 given $Q_T=q_T$ as

$$\begin{aligned} & f_{Q_1|Q_T}(q_1,q_T;\beta,\lambda) = K_1K_2^{-1}\exp(-q_S/2)\det(q)^{(k-3)/2}q_T^{-(k-2)/2}\times \\ & \sum_{j=0}^{\infty}\frac{(\lambda\xi_\beta(q)/4)^j}{j!\Gamma((k-2)/2+j+1)}\Bigg/\sum_{j=0}^{\infty}\frac{(\lambda d_\beta^2 q_T/4)^j}{j!\Gamma((k-2/2)+j+1)} \end{aligned} \tag{14.22}$$

using Lemma 3(a) and (b) and (4.10).

Tedious algebraic manipulations show that

$$\begin{aligned} \frac{\partial f_{Q_1|Q_T}(q_1,q_T;\beta_0,\lambda)}{\partial\beta} &= \frac{\lambda}{2}f_{Q_1|Q_T}(q_1,q_T;\beta_0)q_{ST}(\det(\Omega))^{-1/2}\times \\ & I_{k/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T})/I_{(k-2)/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T}). \end{aligned} \tag{14.23}$$

The function $I_{k/2}(\cdot)$ arises because

$$\frac{\partial}{\partial\beta}\sum_{j=0}^{\infty}\frac{(\lambda\xi_\beta(q)/4)^j}{j!\Gamma((k-2)/2+j+1)} = \frac{\lambda}{4}\frac{\partial\xi_\beta(q)}{\partial\beta}\sum_{s=0}^{\infty}\frac{(\lambda\xi_\beta(q)/4)^s}{s!\Gamma(k/2+s+1)} \tag{14.24}$$

and likewise with $\xi_\beta(q)$ replaced by $(d_\beta^2 q_T)$.

The necessary condition for unbiasedness, (14.21), and (14.23) give

$$\begin{aligned} 0 &= \int h(q_T)f_{Q_T}(q_T;\beta_0,\lambda)dq_T\frac{I_{k/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T})}{I_{(k-2)/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T})}, \text{ where} \\ h(q_T) &= \int \phi(q_1,q_T)q_{ST}f_{Q_1|Q_T}(q_1,q_T;\beta_0)dq_1. \end{aligned} \tag{14.25}$$

By completeness of Q_T under H_0, see Comment 5 following Theorem 2, it must be the case that $h(q_T)$ is zero for almost all q_T and all $\lambda\geq 0$, which yields the second condition of the Theorem. □

Proof of Corollary 3. Any test that depends on (Q_S, Q_{ST}^2, Q_T) can be written as $\phi(Q_S, \mathcal{S}_2^2, Q_T)$, where $\mathcal{S}_2 = Q_{ST}/(Q_S Q_T)^{1/2}$. By Lemma 3(e) and (f), Q_S, $\mathcal{S}_2$, and Q_T are independent under H_0 and $\mathcal{S}_2$ has a distribution that is symmetric about zero. Hence, we have

$$\begin{aligned} & E_{\beta_0}(\phi(Q_S, \mathcal{S}_2^2, Q_T) Q_{ST} | Q_T = q_T) = E_{\beta_0}(\phi(Q_S, \mathcal{S}_2^2, q_T)\mathcal{S}_2 Q_S^{1/2}) q_T^{1/2} \\ = & \int E_{\beta_0}(\phi(q_S, \mathcal{S}_2^2, q_T)\mathcal{S}_2) q_S^{1/2} f_{Q_S}(q_S) dq_S \cdot q_T^{1/2} = 0 \end{aligned} \tag{14.26}$$

for all q_T, where the last equality holds because $\phi(q_S, \mathcal{S}_2^2, q_T)\mathcal{S}_2$ is an odd function of $\mathcal{S}_2$ and $\mathcal{S}_2$ is symmetrically distributed about zero. □

Proof of Theorem 5. By the same argument as in Section 4.2, it suffices to find the test that maximizes power against the single alternative density $g_W(q_1|q_T)$ conditional on $Q_T = q_T$. Given the restriction to locally-unbiased tests, we apply the generalized Neyman-Pearson (GNP) Lemma, see Lehmann (1986, Thm. 3.5, pp. 96-7), rather than the Neyman-Pearson Lemma. The GNP Lemma implies that the optimal (conditional) test rejects when $LR_W(Q_1, q_T) > \widetilde{\kappa}_{1\alpha}(q_T) + \widetilde{\kappa}_{2\alpha}(q_T) Q_{ST}$ for some $\widetilde{\kappa}_{1\alpha}(q_T)$ and $\widetilde{\kappa}_{2\alpha}(q_T)$ that are chosen such that the two conditions of Theorem 4 hold.

It remains to verify the conditions needed to apply the generalized Neyman-Pearson Lemma. Let M be the set of points

$$\left(E\left(\phi(Q_1, Q_T)|Q_T = q_T\right), E\left(\phi(Q_1, Q_T) Q_{ST}|Q_T = q_T\right)\right) \tag{14.27}$$

as ϕ ranges over all possible critical functions. It suffices to show that $(\alpha, 0)$ is an interior point of M, see Lehmann (1986, Thm. 3.5(iv), p. 97).

The set M is convex because the conditional expectation operator is linear. Moreover, M contains $(\alpha, 0)$ by considering the LM test. It also contains points (α, u_α^+) with $u_\alpha^+ > 0$ by considering the one-sided LM test which rejects H_0 when $Q_{ST}/Q_T^{1/2} > c_\alpha$. This follows because the derivative of the conditional power function of this test is an increasing linear transformation of

$$\int 1\left(q_{ST}/q_T^{1/2} > c_\alpha\right) q_{ST} f_{Q_1|Q_T}(q_1, q_T; \beta_0) dq_1, \tag{14.28}$$

which is strictly positive. Likewise, M also contains points (α, u_α^-) with $u_\alpha^- < 0$ by considering the test which rejects H_0 when $-Q_{ST}/Q_T^{1/2} > c_\alpha$ by an analogous argument. This completes the verification that $(\alpha, 0)$ lies in the interior of M. □

14.5 Proofs of Results Stated in Section 9

Proof of Lemma 4. Under Assumptions IID, INID, or MDS, Assumptions 1 and 2 hold by standard LLN's and Assumption 3 holds by a MDS CLT, such as Cor. 3.1 of Hall and Heyde (1980, p. 58). Under Assumption CORR, Assumptions 1 and 2 hold by the ergodic theorem and Assumption 3 holds by the CLT given in the Theorem of Heyde (1975) (of which there is only one). □

Proof of Lemma 5. Using the definition $Y = Z\pi a' + X\eta + V$, we obtain $\widehat{V} = V - P_Z V - P_X V$. This and $P_Z P_X = 0$ gives

$$n^{-1}\widehat{V}'\widehat{V} - \Omega = (n^{-1}V'V - \Omega) - n^{-1}V'P_Z V - n^{-1}V'P_X V. \tag{14.29}$$

The first summand on the right-hand side of (14.29) converges in probability to zero by Assumption 2. The second summand satisfies

$$0 \leq n^{-1}V'P_Z V \leq n^{-1}V'P_{\widetilde{Z}}V = n^{-1}(n^{-1/2}V'\widetilde{Z})(n^{-1}\widetilde{Z}'\widetilde{Z})^{-1}(n^{-1/2}\widetilde{Z}'V) \rightarrow_p 0, \tag{14.30}$$

where the second inequality holds because the span of Z is contained in the span of $\widetilde{Z}$ and the convergence to zero holds by Assumptions 1 and 3. The third summand of (14.29) converges in probability to zero by an analogous argument. □

Proof of Lemma 6. To establish part (a), we have

$$n^{-1}Z'Z = n^{-1}\widetilde{Z}'\widetilde{Z} - n^{-1}\widetilde{Z}'P_X\widetilde{Z} \rightarrow_p D_{11} - D_{12}D_{22}^{-1}D_{21} = D_Z \tag{14.31}$$

using Assumption 1. Let N^* be a $(k+p) \times 2$ random matrix with $vec(N^*) \sim N(0, \Omega \otimes D)$. Using Assumptions 1 and 3, we obtain

$$\begin{aligned} n^{-1/2}Z'Vb_0 &= n^{-1/2}(\widetilde{Z} - P_X\widetilde{Z})'Vb_0 = n^{-1/2}(\widetilde{Z} - XD_{22}^{-1}D_{21})'Vb_0 + o_p(1) \\ &= \left[I_k : -D_{12}D_{22}^{-1}\right] n^{-1/2}\overline{Z}'Vb_0 + o_p(1) \rightarrow_d \left[I_k : -D_{12}D_{22}^{-1}\right] N^*b_0 \\ &= \left[I_k : -D_{12}D_{22}^{-1}\right](b_0' \otimes I_{k+p})vec(N^*). \end{aligned} \tag{14.32}$$

Hence, we have

$$\begin{aligned} S_n &= (n^{-1}Z'Z)^{-1/2}(n^{-1/2}Z'Vb_0 + n^{-1}Z'ZCa'b_0) \cdot (b_0'\Omega b_0)^{-1/2} \rightarrow_d H, \text{ where} \\ H &= D_Z^{-1/2}\left(\left[I_k : -D_{12}D_{22}^{-1}\right](b_0' \otimes I_{k+p})vec(N^*) + D_Z Ca'b_0\right) \cdot (b_0'\Omega b_0)^{-1/2} \end{aligned} \tag{14.33}$$

and the first equality holds by Assumption WIV-FA and $Z'X = 0$. Using Assumption 4, the random vector H has a normal distribution with

$$\begin{aligned} EH &= D_Z^{1/2}Ca'b_0 \cdot (b_0'\Omega b_0)^{-1/2} = c_\beta D_Z^{1/2}C \text{ and} \\ var(H) &= D_Z^{-1/2}\left[I_k : -D_{12}D_{22}^{-1}\right](b_0' \otimes I_{k+p})(\Omega \otimes D)(b_0 \otimes I_{k+p}) \\ &\quad \times \left[I_k : -D_{12}D_{22}^{-1}\right]' D_Z^{-1/2} \cdot (b_0'\Omega b_0)^{-1} \\ &= D_Z^{-1/2}\left[I_k : -D_{12}D_{22}^{-1}\right] D \left[I_k : -D_{12}D_{22}^{-1}\right]' D_Z^{-1/2} = I_k, \end{aligned} \tag{14.34}$$

which completes the proof for S_n.

Analogously to (14.32), we have

$$n^{-1/2}Z'V\Omega^{-1}a_0 \rightarrow_d \left[I_k : -D_{12}D_{22}^{-1}\right]((a_0'\Omega^{-1}) \otimes I_{k+p})vec(N^*). \tag{14.35}$$

Using this, we obtain

$$T_n = (n^{-1}Z'Z)^{-1/2}\Big(n^{-1/2}Z'V\Omega^{-1}a_0 + n^{-1}Z'ZCa'\Omega^{-1}a_0\Big)\cdot(a_0'\Omega^{-1}a_0)^{-1/2} \to_d J, \text{ for}$$
$$J = D_Z^{-1/2}\big(\big[I_k : -D_{12}D_{22}^{-1}\big]((a_0'\Omega^{-1})\otimes I_{k+p})vec(N^*) + D_Z Ca'\Omega^{-1}a_0\big)\cdot(a_0'\Omega^{-1}a_0)^{-1/2}. \tag{14.36}$$

Analogously to (14.34), J has a normal distribution with $EJ = d_\beta D_Z^{1/2}C$ and $var(J) = I_k$, which completes the proof for T_n.

The asymptotic normal distributions of S_n and T_n are independent because the covariance of the random components of H and J is zero:

$$\begin{aligned} & E(b_0' \otimes I_{k+p})vec(N^*)vec(N^*)'((\Omega^{-1}a_0)\otimes I_{k+p}) \\ = \ & E(b_0' \otimes I_{k+p})(\Omega \otimes D)((\Omega^{-1}a_0)\otimes I_{k+p}) = (b_0'a_0)\otimes D = 0. \end{aligned} \tag{14.37}$$

This completes the proof of part (a).

Part (b) holds by the definitions of $\widehat{S}_n$, $\widehat{T}_n$, S_n, and T_n because (i) $(Z'Z)^{-1/2}Z'Y = O_p(1)$ by the same sort of argument as in (14.31) and (14.32), (ii) $\widehat{\Omega}_n \to_p \Omega$ by Lemma 5, and (iii) Ω is pd by Assumption 2.

Part (c) follows immediately from parts (a) and (b). □

Proof of Theorem 6. The functions $\psi_W(\cdot,\cdot)$ and $\psi_{2,W}(\cdot)$ are continuous and do not depend on n, see their definitions in Corollary 1. The same is true of the critical value function $\kappa_\alpha(\cdot)$ because the conditional distribution of $Q_{1,n}$ given $Q_{T,n}$ is absolutely continuous with a density that is a smooth function of q_T and does not depend on n, see Lemma 3(c) and the definition of $\kappa_\alpha(\cdot)$ in (4.12). In consequence, the result of the Theorem follows from Lemma 6, (9.5), and the continuous mapping theorem. □

Proof of Corollary 4. To prove part (a), let $\widehat{\Psi}_n = LR_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) - \kappa_\alpha(\widehat{Q}_{T,n})$, $\Psi_n = LR_W(Q_{1,n}, Q_{T,n}) - \kappa_\alpha(Q_{T,n})$, and $\Psi = LR_W(Q_{1,\infty}, Q_{T,\infty}) - \kappa_\alpha(Q_{T,\infty})$. By Theorem 6(b),

$$P(|\widehat{\Psi}_n - \Psi_n| > \varepsilon) \to 0 \text{ for all } \varepsilon > 0. \tag{14.38}$$

We have

$$\begin{aligned} & P(|1(\widehat{\Psi}_n > 0) - 1(\Psi_n > 0)| > \varepsilon) \\ \le \ & P(\widehat{\Psi}_n > 0 \ \& \ \Psi_n \le 0) + P(\widehat{\Psi}_n \le 0 \ \& \ \Psi_n > 0). \end{aligned} \tag{14.39}$$

The first summand on the right-hand side of (14.39) satisfies

$$P(\widehat{\Psi}_n > 0 \ \& \ \Psi_n \le 0) \le P(0 < \widehat{\Psi}_n \le \varepsilon) + o(1) \to P(0 < \Psi \le \varepsilon), \tag{14.40}$$

where the inequality holds by (14.38) and the convergence holds by Theorem 6(c). The right-hand side of (14.40) converges to zero as $\varepsilon \to 0$ because Ψ has an absolutely continuously distribution by Lemma 3(a). Hence, the left-hand side of (14.40) converges to zero as $n \to \infty$.

By an analogous argument, the second summand on the right-hand side of (14.39) converges to zero as $n \to \infty$, which completes the proof of part (a).

Parts (b) and (c) follow immediately from Theorem 6(a) and (c).

Part (d) holds for the following reasons. The conditional distribution of $Q_{1,\infty}$ given $Q_{T,\infty} = q_T$ is the same as that of $Q_{1,n}$ given $Q_{T,n} = q_T$ because the former distribution does not depend on λ_∞ and the latter does not depend on λ, see Lemma 3(c). Hence, by definition of $\kappa_\alpha(\cdot)$, for all constants $q_{T,\infty}$, $P(LR_W(Q_{1,\infty}, q_{T,\infty}) > \kappa_\alpha(q_{T,\infty})|Q_{1,\infty} = q_{T,\infty})) = \alpha$. This result and iterated expectations establishes part (d). □

Proof of Theorem 7. First, we prove part (a). We have

$$\widehat{V}_j' b_0 = V_j' b_0 - Z_j'(Z'Z)^{-1}Z'Vb_0 - X_j'(X'X)^{-1}X'Vb_0 \tag{14.41}$$

because $\widehat{V} = V - P_Z V - P_X V$. Using (14.41), some manipulations, and Assumption 6, we obtain

$$n^{-1}\sum_{j=1}^{n}(\widehat{V}_j' b_0)^2 Z_j Z_j' - n^{-1}\sum_{j=1}^{n}(V_j' b_0)^2 Z_j Z_j' \to_p 0. \tag{14.42}$$

In addition, we have

$$\begin{aligned}
& n^{-1}\sum_{j=1}^{n}(V_j' b_0)^2 Z_j Z_j' \\
= \ & n^{-1}\sum_{j=1}^{n}(V_j' b_0)^2(\widetilde{Z}_j - D_{12}D_{22}^{-1}X_j)(\widetilde{Z}_j - D_{12}D_{22}^{-1}X_j)' + o_p(1) \\
= \ & n^{-1}\sum_{j=1}^{n} MB_0(V_i \otimes \overline{Z}_i)(V_i' \otimes \overline{Z}_i')B_0'M' + o_p(1) \\
\to_p \ & MB_0\Phi B_0'M',
\end{aligned} \tag{14.43}$$

where the first equality holds using Assumption 1 via some manipulations, the second equality holds by linear algebra, and convergence holds by Assumption 5. Combining (14.42) and (14.43), gives $\widetilde{\Sigma}_{S,n} \to_p \widetilde{\Sigma}_S$.

By similar arguments, $\widetilde{\Sigma}_{TS,n} \to_p \widetilde{\Sigma}_{TS}$ and $\widetilde{\Sigma}_{T,n}^* \to_p \widetilde{\Sigma}_T^*$. (The arguments are somewhat more involved because b_0 is replaced by the random quantity $\widehat{\Omega}_n^{-1}a_0$, but no additional assumptions are needed.) These results combine to give $\widetilde{\Sigma}_{T,n} \to_p \widetilde{\Sigma}_T$.

To establish part (b), we first show that the result of Lemma 6(c) holds. We have

$$\begin{aligned}
\widetilde{S}_n \ &= \ \widetilde{\Sigma}_{S,n}^{-1/2}\left(n^{-1/2}Z'Vb_0 + n^{-1}Z'ZCa'b_0\right) \\
&\to_d \ \widetilde{\Sigma}_S^{-1/2}\left[I_k : -D_{12}D_{22}^{-1}\right](b_0' \otimes I_{k+p})vec(N^*) + \widetilde{\Sigma}_S^{-1/2}D_Z Ca'b_0 \\
&\sim \ N(\widetilde{\Sigma}_S^{-1/2}D_Z Ca'b_0, I_k),
\end{aligned} \tag{14.44}$$

where $vec(N^*) \sim N(0, \Phi)$, the equality uses (9.7) and Assumption 1, and the convergence holds by part (a), (14.31), and (14.32).

By Lemma 5, part (a), and Assumption 3, the use of $\widehat{\Omega}_n^{-1}$, rather than Ω^{-1}, in the definition of T_n has no effect asymptotically. Hence, we have

$$\begin{aligned}
\widetilde{T}_n &= \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1/2}Z'Y\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1}n^{-1/2}Z'Yb_0\right) + o_p(1) \\
&= \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1/2}Z'V\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1}n^{-1/2}Z'Vb_0\right) \\
&\quad + \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1}Z'ZCa'\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1}n^{-1}Z'ZCa'b_0\right) + o_p(1) \\
&\to_d \widetilde{\Sigma}_T^{-1/2}\left(MA_0\, vec(N^*) - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB_0\, vec(N^*)\right) \\
&\quad + \widetilde{\Sigma}_T^{-1/2}\left(D_Z Ca'\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}D_Z Ca'b_0\right), \qquad (14.45)
\end{aligned}$$

where M, A_0, and B_0 are defined in (9.9) and the convergence holds by (14.32) and (14.36). The covariance matrix of the limiting distribution in (14.45) is I_k because

$$\begin{aligned}
& var\left(MA_0\, vec(N^*) - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB_0\, vec(N^*)\right) \\
&= MA_0\Phi A_0'M' - MA_0\Phi B_0'M'\widetilde{\Sigma}_S^{-1}\widetilde{\Sigma}_{TS}' - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB_0\Phi A_0'M' \\
&\quad + \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB_0\Phi B_0'M'\widetilde{\Sigma}_S^{-1}\widetilde{\Sigma}_{TS}' \\
&= \widetilde{\Sigma}_T^* - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}\widetilde{\Sigma}_{TS}' = \widetilde{\Sigma}_T. \qquad (14.46)
\end{aligned}$$

The convergence in (14.44) and (14.45) is joint and the limit random vectors are independent because

$$\begin{aligned}
& cov(MA_0\, vec(N^*) - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB_0\, vec(N^*),\ \widetilde{\Sigma}_S^{-1/2}MB_0\, vec(N^*)) \qquad (14.47) \\
&= MA_0\Phi M'B_0'\widetilde{\Sigma}_S^{-1/2} - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB_0\Phi B_0'M'\widetilde{\Sigma}_S^{-1/2} = \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1/2} - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1/2} = 0.
\end{aligned}$$

To complete the proof of part (b), we note that (i) Theorem 6(c) (with the changes indicated in Theorem 7(b)) follows from (14.44)-(14.47) by the continuous mapping theorem, (ii) Corollary 4(c) follows immediately from Theorem 6(c), and (iii) Corollary 4(d) holds with $(\widetilde{Q}_{1,\infty}, \widetilde{Q}_{T,\infty})$ by the same reason as with $(Q_{1,\infty}, Q_{T,\infty})$. □

14.6 Proofs of Results Stated in Section 10

Proof of Lemma 8. Part (a) holds because (i) conditional on $[Z : X]$, equation (14.2) with (π, Ω, η) replaced by $(C/n^{1/2}, \Omega_0 + \Omega_1/n^{1/2}, \eta_0 + \eta_1/n^{1/2})$, where Ω_0 and η_0 are known and Ω_1 and η_1 are unknown, implies that $(Z'Y, X'Y, Y'Y)$ are sufficient statistics for $(\beta, C, \Omega_1, \eta_1)$ and (ii) $(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0))$ is an equivalent set of sufficient statistics to $(Z'Y, X'Y, Y'Y)$.

Part (b) holds because (i) $vec(n^{-1/2}Z'V) \sim N(0, \Omega \otimes (n^{-1}Z'Z))$ conditional on $n^{-1}Z'Z$ and $n^{-1}Z'Z \to_p D_Z$ (by (14.32) using Assumption 1) imply that $vec(n^{-1/2}Z'V) \to_d N(0, \Omega \otimes D_Z)$, (ii) $vec(n^{-1/2}Z'Z\pi a') = vec(n^{-1}Z'ZCa') \to_p D_Z Ca'$ by Assumption 1, (iii) $n^{1/2}(\widehat{\eta}_n - \eta_0) = (n^{-1}X'X)^{-1}n^{-1/2}X'V + \eta_1 \sim N(\eta_1, \Omega \otimes$

$(n^{-1}X'X)^{-1})$ conditional on $n^{-1}X'X$ and $(n^{-1}X'X)^{-1} \to_p D_{22}^{-1}$ (using Assumption 1) imply that $vec(n^{1/2}(\widehat{\eta}_n - \eta_0)) \to_d N(\eta_1, \Omega \otimes D_{22}^{-1})$, (iv) $n^{1/2}(\widehat{\Omega}_n - \Omega_0) = n^{1/2}(n^{-1}V'V - \Omega_0) - n^{-1/2}V'P_Z V - n^{-1/2}V'P_X V$ using (14.29), (v) $n^{1/2}(n^{-1}V'V - \Omega_0) = n^{-1/2}(V'V - EV'V) + \Omega_1$, (vi) $vech(n^{-1/2}(V'V - EV'V)) \to_d N(0, E(\zeta - E\zeta)(\zeta - E\zeta)')$ by a triangular array CLT for row-wise iid random vectors, (vii) $n^{-1/2}V'P_Z V = n^{-1/2} \cdot n^{-1/2}V'Z(n^{-1}Z'Z)^{-1}n^{-1/2}Z'V \to_p 0$ using (i), (viii) $n^{-1/2}V'P_X V \to_p 0$ by an analogous argument to (vii), and (ix) the three random matrices on the left-hand side of part (b) are asymptotically independent because they are independent in finite samples conditional on $n^{-1}Z'Z$ and $n^{-1}X'X$ and the randomness in $n^{-1}Z'Z$ and $n^{-1}X'X$ is asymptotically negligible. $\square$

Proof of Theorem 8. The equality in the Theorem holds by the definition of a convergent sequence of asymptotically invariant tests. The inequality holds because (i) given the random quantities $(Q_\infty, N_X, N_\Omega)$, Q_∞ is a sufficient statistic for β and C since it is independent of N_X and N_Ω and the latter have distributions that do not depend on β or C, (ii) part (i) implies that the WAP of the similar test $\phi^*(Q_\infty, N_X, N_\Omega)$ is less than or equal to that of some similar test $\widetilde{\phi}(Q_\infty)$ that depends on $(Q_\infty, N_X, N_\Omega)$ only through Q_∞, and (iii) Theorem 3 with Q replaced by Q_∞ implies that the WAP of the similar test $\widetilde{\phi}(Q_\infty)$ is less than or equal to the upper bound given in Theorem 8. $\square$

14.7 Proofs of Results Stated in Section 11

Proof of Lemma 9. To prove part (a)(i), we use (2.6), (9.4), (14.6)-(14.8), and Assumptions SIV-LA, 1, 3, and 4 to obtain

$$\begin{aligned} S_n &= c_\beta \mu_\pi + (Z'Z)^{-1/2}Z'Vb_0 \cdot (b_0'\Omega b_0)^{-1/2} \to_d \zeta_S \text{ and} \\ T_n/n^{1/2} &= d_\beta \mu_\pi / n^{1/2} + (Z'Z/n)^{-1/2}(Z'V/n)\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \\ &= d_\beta (Z'Z/n)^{1/2}\pi + o_p(1) = \alpha_T + o_p(1). \end{aligned} \quad (14.48)$$

Part (a)(ii) holds by Lemma 5. Part (a)(iii) holds by part (a)(i), part (a)(ii), and the continuous mapping theorem.

Part (b)(i) holds by Theorem 7(a) (whose proof does not rely on Assumption WIV-FA). Part (b)(ii) holds for $\widetilde{S}_n$ using part (b)(i) and (14.44) but with $n^{-1}Z'ZCa'b_0$ replaced by $n^{-1/2}Z'Z\pi a'b_0 = n^{-1}Z'Z\pi B = D_Z \pi B + o_p(1)$. Part (b)(ii) holds for $\widetilde{T}_n$ using part (b)(i), the result of part (b)(ii) for $\widetilde{S}_n$, Lemma 5, and (9.4):

$$\begin{aligned} \widetilde{T}_n/n^{1/2} &= \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1}Z'Y\widehat{\Omega}_n^{-1}a_0 - n^{-1/2}\widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1/2}\widetilde{S}_n\right) \\ &= \widetilde{\Sigma}_T^{-1/2}\left(n^{-1}Z'Z\pi a'\Omega^{-1}a_0 + n^{-1}Z'V\Omega^{-1}a_0\right) + o_p(1) \\ &= \widetilde{\Sigma}_T^{-1/2}D_Z \pi a_0'\Omega^{-1}a_0 + o_p(1) = \widetilde{\alpha}_T + o_p(1). \end{aligned} \quad (14.49)$$

Part (b)(iii) holds by part (b)(ii) and the continuous mapping theorem. $\square$

Proof of Theorem 9. Parts (a)(i), (a)(ii), (b)(i), and (b)(ii) of the Theorem follow immediately from Lemma 9(a) and 9(b).

The first equality of part (a)(iii) follows from Lemma 9(a). The second equality of part (a)(iii) of the Theorem is established as follows. For brevity, we drop the subscript n on $Q_{S,n}$, $Q_{T,n}$, and $Q_{ST,n}$. Equation (3.4) and simple algebra yields

$$\begin{aligned} 2 \cdot LR_n &= Q_S - Q_T + (Q_S + Q_T)\sqrt{1-\delta_n/n} \\ &= Q_S\left(1+\sqrt{1-\delta_n/n}\right) - (Q_T/n)n\left(1-\sqrt{1-\delta_n/n}\right), \text{ where} \\ \delta_n &= 4n\left(Q_SQ_T - Q_{ST}^2\right)(Q_S+Q_T)^{-2}. \end{aligned} \quad (14.50)$$

Some algebra and Lemma 9(a) gives

$$\begin{aligned} \delta_n &= 4\left(Q_S - Q_{ST}^2/Q_T\right)\left(Q_SQ_T^{-1}+1\right)^{-2}(Q_T/n)^{-1} \\ &= 4\left(Q_S - Q_{ST}^2/Q_T\right)(\alpha_T'\alpha_T)^{-1} + o_p(1) = O_p(1). \end{aligned} \quad (14.51)$$

A mean-value expansion about $\delta_n/n = 0$ gives

$$n\left(1-\sqrt{1-\delta_n/n}\right) = n\left(1-\left(1-\frac{1}{2}\frac{\delta_n}{n}+O_p(n^{-2})\right)\right) = \frac{1}{2}\delta_n + o_p(1). \quad (14.52)$$

This result, $\sqrt{1-\delta_n/n} \to_p 1$, (14.50), and (14.51) yield

$$LR_n = \frac{1}{2}\left(2Q_S - 2\left(Q_S - Q_{ST}^2/Q_T\right)\right) + o_p(1) = Q_{ST}^2/Q_T + o_p(1), \quad (14.53)$$

which establishes the second equality of part (a)(iii).

By the same argument as in (14.50)-(14.53), but using Lemma 9(b) in place of Lemma 9(a), we obtain $\widetilde{LR}_n = \widetilde{LM}_n + o_p(1)$, which establishes the equality in part (b)(ii).

Next, for part (a)(iv) of the Theorem, define $LR1_\infty$ as $LR1$ is defined in (3.7) but with LR, $R(\beta)$, and $\widehat{\beta}_{LIML-k}$ replaced by $(\alpha_T'\zeta_S)^2/||\alpha_T||^2$, $R_\infty(\beta)$, and $\widehat{\beta}_{LIML-k,\infty}$, respectively, where $R_\infty(\beta)$ is defined as $R(\beta)$ is defined in (14.15) but with Q replaced by its SIV-LA limit given in Lemma 9(a)(iii) and by definition $\widehat{\beta}_{LIML-k,\infty}$ minimizes $R_\infty(\beta)$ over $\beta \in R$. The first equality of part (a)(iv) holds by Lemma 9(a)(ii) and the convergence holds by Lemma 9(a)(iii) and the continuous mapping theorem given the absolute continuity of ζ_S.

For part (b)(iv), define $\widetilde{LR}1_\infty$ analogously but with (ζ_S, α_T) replaced by $(\widetilde{\zeta}_S, \widetilde{\alpha}_T)$ and with Q replaced by its SIV-LA limit given in Lemma 9(b)(iii). The result of part (b)(iv) of the Theorem holds by Lemma 9(b)(iii) and the continuous mapping theorem given the absolute continuity of $\widetilde{\zeta}_S$. $\square$

Proof of Theorem 10. We suppose that Ω is known and determine the standard LM statistic for this case, which is asymptotically efficient by standard results. In particular, we show that the standard LM statistic is $LM_n = Q_{ST}^2/Q_T$. By Theorem 9, all the LM and LR statistics listed in the statement of the present Theorem are asymptotically equivalent to LM_n under the null hypothesis and local alternatives under strong IV asymptotics and the asymptotic behavior of these statistics does not depend on knowledge of Ω. Hence, the tests based on these statistics are asymptotically efficient whether or not Ω is known.

The standard LM statistic is a quadratic form in the derivative with respect to (wrt) β of the log likelihood function of the sufficient statistics (S, T) evaluated at the null restricted maximum likelihood estimator of π, which we denote by $\widehat{\pi}_0$. Under the null hypothesis, $S \sim N(0, I_k)$ is ancillary, $\widehat{\pi}_0$ depends on $T \sim N(d_{\beta_0}\mu_\pi, I_k)$ alone, and $\widehat{\pi}_0$ is easily seen to be $\widehat{\pi}_0 = d_{\beta_0}^{-1}(Z'Z)^{-1/2}T$. The log-likelihood of (S, T) is proportional to

$$-\frac{1}{2}(S - c_\beta\mu_\pi)'(S - c_\beta\mu_\pi) - \frac{1}{2}(T - d_\beta\mu_\pi)'(T - d_\beta\mu_\pi). \tag{14.54}$$

The derivative of this expression wrt β evaluated at $(\beta, \pi) = (\beta_0, \widehat{\pi}_0)$ is

$$\begin{aligned}
&\left(\frac{d}{d\beta}c_\beta\mu_\pi' S - \frac{1}{2}\frac{d}{d\beta}(c_\beta^2)\mu_\pi'\mu_\pi + \frac{d}{d\beta}d_\beta\mu_\pi' T - \frac{1}{2}\frac{d}{d\beta}(d_\beta^2)\mu_\pi'\mu_\pi\right)\Bigg|_{(\beta,\pi)=(\beta_0,\widehat{\pi}_0)} \\
&= \frac{d}{d\beta}c_{\beta_0}\mu_{\widehat{\pi}_0}' S + \frac{d}{d\beta}d_{\beta_0}\mu_{\widehat{\pi}_0}' T - d_{\beta_0}\frac{d}{d\beta}d_{\beta_0}\mu_{\widehat{\pi}_0}'\mu_{\widehat{\pi}_0} \\
&= \frac{d}{d\beta}c_{\beta_0} \cdot d_{\beta_0}^{-1}T'S,
\end{aligned} \tag{14.55}$$

using the facts that $c_{\beta_0} = 0$, $\mu_{\widehat{\pi}_0} = d_{\beta_0}^{-1}T$ and $\mu_{\widehat{\pi}_0}'T = d_{\beta_0}\mu_{\widehat{\pi}_0}'\mu_{\widehat{\pi}_0}$. The asymptotic variance of $T'S/n^{1/2}$ under H_0 is $p\lim_{n\to\infty} T'T/n = \alpha_T'\alpha_T$. Hence, the standard LM statistic is $(T'S)^2/T'T = LM_n$, which completes the proof. $\square$

Proof of Theorem 11. First, we establish Theorem 11(a)(i). The first equality holds by Lemma 9(a)(ii) and the definition of $POIS1_\delta/\delta$ in (5.15). Next, using (5.15), we have

$$POIS1_\delta/\delta = \frac{Q_S/\delta + Q_{ST}/\sqrt{Q_T} - k/\delta}{sgn(\delta)\sqrt{2k/\delta + 1}}. \tag{14.56}$$

By Lemma 9(a)(i),

$$1/\delta = \left((2d_{\beta^*}/c_{\beta^*})\sqrt{Q_T/n}\right)^{-1}/\sqrt{n} = O_p(n^{-1/2}). \tag{14.57}$$

By definition of δ, $sgn(\delta) = sgn(c_{\beta^*}d_{\beta^*})$. Combining this, (14.56), (14.57), and Lemma 9(a)(i) gives $Q_S/\delta = o_p(1)$ and the second equality of part (a)(i) holds. The convergence in part (a)(i) holds by Lemma 9(a)(i).

Theorem 11(a)(ii) holds because $\beta^* = \beta_{AR}$ implies that $d_{\beta^*} = 0$ and $\delta = 0$. In consequence, the second equality of part (a)(ii) holds by the definition of $POIS1_\delta$ and the first equality holds using Lemma 9(a)(ii).

The proof of part (b) is the same as that of part (a) except that Lemma 9(b) is used in place of Lemma 9(a). $\square$

Proof of Comment 1 to Theorem 11. The critical value function of a POIS1 test, $\kappa_\alpha^{POIS1}(q_T)$, converges to the $1 - \alpha$ quantile of the standard normal distribution as $q_T \to \infty$. This holds because Q_T enters $POIS1_\delta$ only through δ and

$\lim_{\delta\to\infty} POIS1_\delta = Q_{ST}/Q_T^{1/2} \sim N(0,1)$. Since $Q_T \to_p \infty$ under SIV-LA asymptotics, this implies that $\kappa_\alpha^{POIS1}(Q_T)$ converges in probability to the $1-\alpha$ quantile of the standard normal distribution as $q_T \to \infty$. $\square$

Proof of Theorem 12. Part (a)(i) of the Theorem holds by Lemma 9(a)(i) and the continuity of $\psi_{W_{2P}}(q_1, q_T)$ and $\psi_{2,W_{2P}}(q_1, q_T)$ in (q_1, q_T).

To prove Theorem 12(a)(ii) and 12(a)(iii), we establish some preliminary results. Let β_1 and λ_1 be any fixed constants for which $d_{\beta_1} \neq 0$ (i.e., $\beta_1 \neq \beta_{AR}$). Define $h_{\beta_1} = (c_{\beta_1}, d_{\beta_1})'$. Then, we have (I) $Q_T/n \to_p \alpha_T'\alpha_T > 0$ by Lemma 9(a)(i) and Assumption SIV-LA(b), (II) $Q_{ST}/\sqrt{Q_T} = O_p(1)$ by (I) and Lemma 9(a)(i), (III) $Q_S/Q_T = o_p(1)$ and $Q_S/Q_T^{1/2} = o_p(1)$ by (I) and Lemma 9(a)(i), and (IV) $h_1'Qh_1/(d_{\beta_1}^2 Q_T) \to_p 1$ by (II) and (III). Next, we apply the mean-value theorem: $(x+a)^{1/2} - x^{1/2} = (1/2)(x^*)^{-1/2}a$, where x^* lies between x and a, with $x = d_{\beta_1}^2 Q_T$ and $a = 2c_{\beta_1}d_{\beta_1}Q_{ST} + c_{\beta_1}^2 Q_S$. This gives

$$\begin{aligned}\sqrt{h_1'Qh_1} - \sqrt{d_{\beta_1}^2 Q_T} &= \frac{1}{2} m^{-1/2}\left(2c_{\beta_1}d_{\beta_1}Q_{ST} + c_{\beta_1}^2 Q_S\right)\\ &= \frac{c_{\beta_1}d_{\beta_1}Q_{ST}}{(d_{\beta_1}^2 Q_T)^{1/2}}\left(\frac{d_{\beta_1}^2 Q_T}{m}\right)^{1/2} + \frac{1}{2}\frac{c_{\beta_1}^2 Q_S}{(d_{\beta_1}^2 Q_T)^{1/2}}\left(\frac{d_{\beta_1}^2 Q_T}{m}\right)^{1/2}\\ &= \frac{c_{\beta_1} sgn(d_{\beta_1}) Q_{ST}}{Q_T^{1/2}} + o_p(1), \qquad (14.58)\end{aligned}$$

where m lies between $h_1'Qh_1$ and $d_{\beta_1}^2 Q_T$ and the third equality holds using (II)-(IV) and the definition of m.

By Lebedev (1965, (5.11.10), p. 123), we have $I_\nu(x) = \exp(x)(2pi\cdot x)^{-1/2}(1 + O(x^{-1}))$ as $x \to \infty$ for any $\nu \in R$. Hence, using (I), we obtain

$$I_\nu\left(\sqrt{d_{\beta_1}^2 Q_T}\right) e^{-\sqrt{d_{\beta_1}^2 Q_T}}\left(2pi\sqrt{d_{\beta_1}^2 Q_T}\right)^{1/2} = 1 + O_p(n^{-1/2}) \qquad (14.59)$$

and likewise with $h_1'Qh_1$ in place of $d_{\beta_1}^2 Q_T$.

We now consider a weight function $W_{2P}(\beta,\lambda)$ (that does not necessarily satisfy (6.2)). It is convenient to make a change of variables from (β,λ) to (γ,μ), where

$$\gamma = \lambda^{1/2}c_\beta \text{ and } \mu = \lambda^{1/2}d_\beta. \qquad (14.60)$$

Let $\widetilde{h} = (\gamma,\mu)'$. Then, $\lambda\xi_\beta(Q) = \widetilde{h}'Q\widetilde{h}$ and $\lambda d_\beta^2 Q_T = \mu^2 Q_T$. Let $F_{2P}(\gamma,\mu)$ be the two-point distribution on (γ,μ) that corresponds to $W_{2P}(\beta,\lambda)$ and, hence, puts equal weight on $(\gamma^*,\mu^*) = ((\lambda^*)^{1/2}c_{\beta^*}, (\lambda^*)^{1/2}d_{\beta^*})$ and $(\gamma_2^*,\mu_2^*) = ((\lambda_2^*)^{1/2}c_{\beta_2^*}, (\lambda_2^*)^{1/2}d_{\beta_2^*})$. Let $\mu_{\max}$ denote the value of μ that maximizes $|\mu|$ over μ in the support of $F_{2P}(\gamma,\mu)$. That is, $\mu_{\max} = \max\{|\mu^*|, |\mu_2^*|\}$. Let $\nu = (k-2)/2$.

Using this notation and the definition of LR_W in Corollary 1, we have $LR_{W_{2P}}$

equals

$$
\begin{aligned}
&\frac{\int e^{-\frac{1}{2}(\gamma^2+\mu^2)}\left(\widetilde{h}'Q\widetilde{h}\right)^{-\frac{1}{2}\nu} I_\nu\left(\sqrt{\widetilde{h}'Q\widetilde{h}}\right) dF_{2p}(\gamma,\mu)}{\int e^{-\frac{1}{2}\mu^2}\left(\mu^2 Q_T\right)^{-\frac{1}{2}\nu} I_\nu\left(\sqrt{\mu^2 Q_T}\right) dF_{2P}(\gamma,\mu)} \\
=\;& \frac{\int e^{-\frac{1}{2}(\gamma^2+\mu^2)}\left(\widetilde{h}'Q\widetilde{h}\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} e^{\sqrt{\widetilde{h}'Q\widetilde{h}}} dF_{2p}(\gamma,\mu)}{\int e^{-\frac{1}{2}\mu^2}\left(\mu^2 Q_T\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} e^{\sqrt{\mu^2 Q_T}} dF_{2P}(\gamma,\mu)}(1+o_p(1)) \\
=\;& \frac{\int e^{-\frac{1}{2}(\gamma^2+\mu^2)}\left(\frac{\widetilde{h}'Q\widetilde{h}}{\mu^2 Q_T}\right)^{-\frac{1}{2}(\nu+\frac{1}{2})}\left(\mu^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} e^{(\sqrt{\mu^2}-\sqrt{\mu^2_{\max}})\sqrt{Q_T}} e^{\sqrt{\widetilde{h}'Q\widetilde{h}}-\sqrt{\mu^2 Q_T}} dF_{2p}(\gamma,\mu)}{\int e^{-\frac{1}{2}\mu^2}\left(\mu^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} e^{(\sqrt{\mu^2}-\sqrt{\mu^2_{\max}})\sqrt{Q_T}} dF_{2P}(\gamma,\mu)} \\
&\times(1+o_p(1)) \qquad (14.61)\\
=\;& \frac{\int e^{-\frac{1}{2}(\gamma^2+\mu^2)}\left(\mu^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} e^{(\sqrt{\mu^2}-\sqrt{\mu^2_{\max}})\sqrt{Q_T}} e^{\gamma sgn(\mu) Q_{ST} Q_T^{-1/2}} dF_{2p}(\gamma,\mu)}{\int e^{-\frac{1}{2}\mu^2}\left(\mu^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} e^{(\sqrt{\mu^2}-\sqrt{\mu^2_{\max}})\sqrt{Q_T}} dF_{2P}(\gamma,\mu)}(1+o_p(1))
\end{aligned}
$$

where the first equality holds by (14.59), the second equality holds by algebra, and the third equality holds by (IV) and (14.58).

If W_{2P} satisfies (6.2), then $\gamma^* = -\gamma_2^*$, $\mu^* = \mu_2^*$, and $\mu_{\max} = |\mu^*| = |\mu_2^*|$. In this case, the terms in the numerator and denominator of the right-hand side (rhs) of (14.61) that involve $(\sqrt{\mu^2} - \sqrt{\mu^2_{\max}})\sqrt{Q_T}$ equal zero and the rhs of (14.61) without $(1 + o_p(1))$ equals

$$
\begin{aligned}
&\frac{\frac{1}{2}e^{-\frac{1}{2}((\gamma^*)^2+(\mu^*)^2)}\left((\mu^*)^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})}\left(e^{\gamma^* sgn(\mu^*) Q_{ST} Q_T^{-1/2}} + e^{-\gamma^* sgn(\mu^*) Q_{ST} Q_T^{-1/2}}\right)}{e^{-\frac{1}{2}(\mu^*)^2}\left((\mu^*)^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})}} \\
&= e^{-\frac{1}{2}(\gamma^*)^2}\cosh(\gamma^* Q_{ST} Q_T^{-1/2}), \qquad (14.62)
\end{aligned}
$$

using $(\exp(x) + \exp(-x))/2 = \cosh(x)$. The function $\cosh(\cdot)$ is even. Hence, $\cosh(\gamma^* Q_{ST} Q_T^{-1/2}) = \cosh(\gamma^* LM_n^{1/2})$. The latter is strictly increasing in LM_n because $\cosh(\cdot)$ is continuous and strictly increasing on R^+. This completes the proof of Theorem 12(a)(ii).

We now establish Theorem 12(a)(iii). Suppose W_{2P} does not satisfy the second condition of (6.2), then either $\mu_{\max} > |\mu_2^*|$ or $\mu_{\max} > |\mu^*|$. Suppose $\mu_{\max} > |\mu_2^*|$, then $\exp((\sqrt{(\mu_2^*)^2} - \sqrt{\mu^2_{\max}})\sqrt{Q_T}) = o_p(1)$ using (I), $\mu_{\max} = |\mu^*| > 0$, and the rhs of (14.61) without $(1 + o_p(1))$ equals

$$
\begin{aligned}
&\frac{e^{-\frac{1}{2}((\gamma^*)^2+(\mu^*)^2)}\left((\mu^*)^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} e^{\gamma^* sgn(\mu^*) Q_{ST} Q_T^{-1/2}} + o_p(1)}{e^{-\frac{1}{2}(\mu^*)^2}\left((\mu^*)^2\right)^{-\frac{1}{2}(\nu+\frac{1}{2})} + o_p(1)} \\
&= e^{-\frac{1}{2}(\gamma^*)^2} e^{\gamma^* sgn(\mu^*) Q_{ST} Q_T^{-1/2}} + o_p(1), \qquad (14.63)
\end{aligned}
$$

which is a strictly monotone, continuous function of $Q_{ST}Q_T^{-1/2}$ and, hence, is not an even function of $Q_{ST}Q_T^{-1/2}$. The same argument applies when $\mu_{\max} > |\mu^*|$.

Note that the case where $\beta^* = \beta_{AR}$ or $\beta_2^* = \beta_{AR}$ is subsumed in the case just considered because in such cases there is no solution to the second equation in (6.2) and, hence, we must have $\mu_{\max} > |\mu^*|$ or $\mu_{\max} > |\mu_2^*|$.

Next, suppose W_{2P} satisfies the second condition of (6.2), but not the first condition. Then, $\gamma^* \neq -\gamma_2^*$, $\mu^* = \mu_2^*$, $\mu_{\max} = |\mu^*| = |\mu_2^*| > 0$, and the rhs of (14.61) without $(1 + o_p(1))$ equals

$$\frac{1}{2}\left(e^{-\frac{1}{2}(\gamma^*)^2} e^{\gamma^* sgn(\mu^*) Q_{ST} Q_T^{-1/2}} + e^{-\frac{1}{2}(\gamma_2^*)^2} e^{\gamma_2^* sgn(\mu^*) Q_{ST} Q_T^{-1/2}}\right), \tag{14.64}$$

which is a continuous function of $Q_{ST} Q_T^{-1/2}$ that is not even because $\gamma^* \neq -\gamma_2^*$. This completes the proof of Theorem 12(a)(iii).

We now establish Theorem 12(a)(iv). We use the same argument as above except with W_{AE} in place of W_{2P}. Let $F_{AE}(\gamma, \mu)$ be the weight function on (γ, μ) that corresponds to $W_{AE}(\beta, \gamma)$ on (β, λ). By (6.4) and the conditions following it, $F_{AE}(\gamma, \mu)$ is of the form

$$F_{AE}(\gamma, \mu) = \frac{1}{2} F_*(\gamma, \mu) + \frac{1}{2} F_*(-\gamma, \mu) \tag{14.65}$$

for some function $F_*(\gamma, \mu)$ with finite support. Then, $LR_{W_{AE}}$ equals the expressions in (14.61) with F_{2P} replaced by F_{AE}.

Let $\mu_{\max}$ denote the value of μ that maximizes $|\mu|$ over μ in the support of $F_*(\gamma, \mu)$. Let $\Theta_{\mu \max}$ denote the set of (γ, μ) values in the support of $F_*(\gamma, \mu)$ for which $|\mu| = \mu_{\max}$. Now, by (14.65), the rhs of (14.61) without $(1 + o_p(1))$ can be written as

$$\begin{aligned} & \frac{\int e^{-\frac{1}{2}(\gamma^2 + \mu^2)} \left(\mu^2\right)^{-\frac{1}{2}(\nu + \frac{1}{2})} e^{(\sqrt{\mu^2} - \sqrt{\mu_{\max}^2})\sqrt{Q_T}} \cosh(\gamma Q_{ST} Q_T^{-1/2}) dF_*(\gamma, \mu)}{\int e^{-\frac{1}{2}\mu^2} \left(\mu^2\right)^{-\frac{1}{2}(\nu + \frac{1}{2})} e^{(\sqrt{\mu^2} - \sqrt{\mu_{\max}^2})\sqrt{Q_T}} dF_*(\gamma, \mu)} \\ = \ & \frac{\sum_{(\gamma,\mu) \in \Theta_{\mu \max}} e^{-\frac{1}{2}\gamma^2} \cosh(\gamma (Q_{ST}^2 Q_T^{-1})^{1/2}) f_*(\gamma, \mu)}{\sum_{(\gamma,\mu) \in \Theta_{\mu \max}} f_*(\gamma, \mu)} + o_p(1), \end{aligned} \tag{14.66}$$

where $f_*(\gamma, \mu)$ denotes the probability mass $F_*(\gamma, \mu)$ puts on (γ, μ) and the equality uses the fact that $\exp((\sqrt{\mu^2} - \sqrt{\mu_{\max}^2})\sqrt{Q_T}) = o_p(1)$ for any μ with $|\mu| < \mu_{\max}$. By the properties of $\cosh(\cdot)$, the rhs is a strictly-increasing, continuous function of $Q_{ST}^2 Q_T^{-1} = LM_n$, which establishes Theorem 12(a)(iv).

Given Lemma 9(b)(iii), the proof of Theorem 12(b)(i)-(iii) is the same as that of Theorem 12(a)(ii)-(iv) with (Q_{ST}, Q_T) replaced by $(\widetilde{Q}_{ST}, \widetilde{Q}_T)$ throughout. □

Proof of Comment 1 to Theorem 12. We write the $LR_{W_{2P}}(Q_1, Q_T)$ statistic as a function of Q_S, $\mathcal{S}_2^2$, and Q_T, say $LR_{W_{2P}}(Q_S, \mathcal{S}_2^2, Q_T)$. The statistics $(Q_S, \mathcal{S}_2^2, Q_T)$ are independent under the null. Hence, we can condition on Q_T without affecting the distribution of $(Q_S, \mathcal{S}_2^2)$. Consider a sequence of constants $\{q_{T,m} : m \geq 1\}$ for which $q_{T,m}/m \to \alpha_T' \alpha_T > 0$. Then, by the argument of (14.58)-(14.66) with $(Q_S, \mathcal{S}_2^2)$ held fixed, when W_{2P} satisfies (6.2) we have $\lim_{m \to \infty} LR_{W_{2P}}(Q_S, \mathcal{S}_2^2, q_{T,m}) =$

$\exp(-\frac{1}{2}(\gamma^*)^2)\cosh(|\gamma^*|(Q_S\mathcal{S}_2^2)^{1/2})$. Because $Q_S\mathcal{S}_2^2 \sim \chi_1^2$, this implies that the conditional critical value function of $LR_{W_{2P}}$, viz., $\kappa_\alpha^{W_{2P}}(q_T)$, converges as $q_T \to \infty$ to a strictly-increasing continuous function of the $1-\alpha$ quantile of χ_1^2. In turn, this implies that $\kappa_\alpha^{W_{2P}}(Q_T)$ converges in probability to the same constant as $n \to \infty$ because $Q_T/n \to_p \alpha_T'\alpha_T > 0$. An analogous argument holds for $LR_{W_{AE}}(Q_1, Q_T)$. □

Proof of Lemma 10. Part (a)(i) of the Lemma is established as follows:

$$\begin{aligned} n^{-1}Z'Yb_0 &= n^{-1}Z'(Z\pi a' + X\eta + V)b_0 \\ &= n^{-1}Z'Z\pi a'b_0 + n^{-1}Z'Vb_0 \to_p D_Z\pi a'b_0 \end{aligned} \quad (14.67)$$

using Assumptions SIV-FA, 1, and 3 and $Z'X = 0$.. Hence, we have

$$\begin{aligned} S_n/n^{1/2} &= (n^{-1}Z'Z)^{-1/2}n^{-1}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2} \\ &\to_p D_Z^{1/2}\pi a'b_0 \cdot (b_0'\Omega b_0)^{-1/2} = D_Z^{1/2}\pi c_\beta. \end{aligned} \quad (14.68)$$

Similarly,

$$\begin{aligned} T_n/n^{1/2} &= (n^{-1}Z'Z)^{-1/2}n^{-1}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \\ &\to_p D_Z^{1/2}\pi a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} = D_Z^{1/2}\pi d_\beta. \end{aligned} \quad (14.69)$$

Part (a)(ii) of the Lemma follows from Lemma 5 and part (a)(i). Part (a)(iii) of the Lemma follows from parts (a)(i) and (a)(ii) and Slutsky's Theorem.

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

$$\begin{aligned} T_n &= (n^{-1}Z'Z)^{-1/2}n^{-1/2}Z'(Z\pi a' + X\eta + V)\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \\ &= (n^{-1}Z'Z)^{-1/2}n^{-1/2}Z'V\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \\ &\to_d \varsigma_k \sim N(0, I_k). \end{aligned} \quad (14.70)$$

The remaining two results of part (a)(iv) follow from Lemma 5, part (a)(i), part (a)(ii), and the continuous mapping theorem.

The proof of part (b) of the Lemma is similar to that of part (a) using the definitions of $\widetilde{S}_n$ and $\widetilde{T}_n$ in (9.7) and Theorem 7(a) (which holds without Assumption WIV-FA). □

Proof of Theorem 13. Parts (a)(i)-(iv) of the Theorem hold by Lemma 10(a)(ii) and 10(a)(iii) and simple calculations. In the case of LM_n, part (a)(ii) only holds if $\beta \neq \beta_{AR}$ (which implies $d_\beta \neq 0$) because d_β appears in the denominator. Part (a)(v) of the Theorem hold by Lemma 10(a)(ii) and 10(a)(iv) and simple calculations.

Next, we prove part (a)(vi) of the Theorem by altering the proof of Theorem 12(a)(ii). Result (I) of this proof still holds under Assumption SIV-FA, but (II)-(IV) and (14.58) do not. We want to make use of (14.61), so we need alternatives to (IV) and (14.58). Our alternative to (14.58) is the following:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\sqrt{\widetilde{h}'\widehat{Q}_n\widetilde{h}}-\sqrt{\mu^2\widehat{Q}_{T,n}}\right)\\
=\ &\sqrt{\gamma^2\widehat{Q}_{T,n}/n+2\gamma\mu\widehat{Q}_{ST,n}/n+\mu^2\widehat{Q}_{T,n}/n}-\sqrt{\mu^2\widehat{Q}_{T,n}/n}\\
\rightarrow_p\ &\lambda_{FA}^{1/2}\left(\sqrt{\gamma^2c_\beta^2+2\gamma\mu c_\beta d_\beta+\mu^2d_\beta^2}-\sqrt{\mu^2d_\beta^2}\right)=g_1(\gamma,\mu,\beta), \qquad (14.71)
\end{aligned}$$

using Lemma 10(a)(iii). Now, the two-point df F_{2P} puts mass on $(\gamma,\mu)=(\gamma^*,\mu^*)$ and $(-\gamma^*,\mu^*)$ because (6.2) is assumed to hold. We claim that $g_1(\gamma^*,\mu^*,\beta)\neq 0$ or $g_1(-\gamma^*,\mu^*,\beta)\neq 0$ or both. The claim holds because $g_1(\gamma^*,\mu^*,\beta)=0$ iff $(\gamma^*)^2c_\beta^2+2\gamma^*\mu^*c_\beta d_\beta=0$ iff $\gamma^*c_\beta+2\mu^*d_\beta=0$ and similarly $g_1(-\gamma^*,\mu^*,\beta)=0$ iff $\gamma^*c_\beta-2\mu^*d_\beta=0$. These two equations are incompatible given that $\gamma^*\neq 0$ (because $\beta^*\neq\beta_0$) and $c_\beta\neq 0$ (because $\beta\neq\beta_0$). The claim implies that the left-hand side of (14.71) multiplied by $\sqrt{n}$ diverges to infinity in probability for $(\gamma,\mu)=(\gamma^*,\mu^*)$ or $(-\gamma^*,\mu^*)$ or both.

Our alternative to (IV) is the following:

$$\frac{\widetilde{h}'\widehat{Q}_n\widetilde{h}}{\mu^2\widehat{Q}_{T,n}}=\frac{\widetilde{h}'(\widehat{Q}_n/n)\widetilde{h}}{\mu^2\widehat{Q}_{T,n}/n}\rightarrow_p\frac{\gamma^2c_\beta^2+2\gamma\mu c_\beta d_\beta+\mu^2d_\beta^2}{\mu^2d_\beta^2}=g_2(\gamma,\mu,\beta) \qquad (14.72)$$

using Lemma 10(a)(iii).

Now, using the above alternatives to (14.58) and (IV), $LR_{W_{2P}}(\widehat{Q}_{1,n},\widehat{Q}_{T,n})$ equals the expressions in (14.61) with the following adjustments to the rhs expression in (14.61): $\exp(\gamma sgn(\mu)Q_{ST}Q_T^{-1/2})$ is replaced by $\exp(\sqrt{n}[g_1(\gamma,\mu,\beta)+o_p(1)])$, $g_2(\gamma,\mu,\beta)^{-\frac{1}{2}(\nu+\frac{1}{2})}$ is added in the numerator, and (Q,Q_T) is replaced by $(\widehat{Q}_n,\widehat{Q}_{T,n})$. The term $\exp(\sqrt{n}[g_1(\gamma,\mu,\beta)+o_p(1)])$ with $(\gamma,\mu)=(\gamma^*,\mu^*)$ or $(-\gamma^*,\mu^*)$ ensures that the rhs of (14.61) diverges to infinity in probability because $g_1(\gamma^*,\mu^*,\beta)\neq 0$ or $g_1(-\gamma^*,\mu^*,\beta)\neq 0$ or both. This completes the proof of part (a)(vi).

The proof of part (a)(vii) of the Theorem is quite similar to that of part (a)(vi).

The proofs of parts (b)(i)-(iv) of the Theorem are analogous to those of parts (a)(i)-(iv) using Lemma 10(b)(ii)-(iv) in place of Lemma 10(a)(ii)-(iv). □

14.8 Proofs of Results Stated in Section 13

Proof of Lemma 11. The proof is essentially the same as that of Lemma 1. □

Proof of Lemma 12. The proof is similar to that of Lemma 2. For brevity, we only discuss the aspects of the proof that differ. To show independence of S and T_j, it suffices to show lack of covariance between S and T_j, because S and T_j are jointly multivariate normal. We have

$$\begin{aligned}
&cov(Z'Yb_0,Z'Y\Omega^{-1}\alpha_{0,j})=cov(\sum_{i=1}^n Z_iY_i'b_0,\sum_{i=1}^n Z_iY_i'\Omega^{-1}\alpha_{0,j})\\
=\ &\sum_{i=1}^n Z_iZ_i'cov(Y_i'b_0,Y_i'\Omega^{-1}\alpha_{0,j})=\sum_{i=1}^n Z_iZ_i'b_0'\Omega\Omega^{-1}\alpha_{0,j}=0, \qquad (14.73)
\end{aligned}$$

because $b_0'\alpha_{0,j} = 0$. By analogous calculations T_j and T_ℓ have zero covariance for $j \neq \ell$ provided $\alpha_{0,j}'\Omega^{-1}\alpha_{0,\ell} = 0$ for all $j \neq \ell$. Lastly, T_j has covariance matrix equal to I_k provided $cov(Z'Yb_0, Z'Y\Omega^{-1}\alpha_{0,j}) = Z'Z$. By analogous calculations to those in (14.73), the latter occurs if $\alpha_{0,j}'\Omega^{-1}\alpha_{0,j} = 1$ for $j = 1, ..., m$. The vectors $\alpha_{0,j}$ are chosen so that the desired conditions $b_0'\alpha_{0,j} = 0$, $\alpha_{0,j}'\Omega^{-1}\alpha_{0,\ell} = 0$, and $\alpha_{0,j}'\Omega^{-1}\alpha_{0,j} = 1$ hold. □

Proof of Theorem 14. The proof is the same as that of Theorem 1, but one considers vectors $(\mu_1, ..., \mu_m)$ and $(\widetilde{\mu}_1, ..., \widetilde{\mu}_m)$ instead of (μ_1, μ_2) and $(\widetilde{\mu}_1, \widetilde{\mu}_2)$. □

Proof of Theorem 15. The proof is the same as that of Theorem 2 provided T is a complete sufficient statistic. The latter holds if the family of distributions of $T = [T_1 : \cdots : T_m]$ under H_0 is a km-parameter exponential family with parameter space that contains a km-dimensional rectangle. The log of the null density of T times minus two is $k\log(2\pi)$ plus

$$\begin{aligned}
&\sum_{j=1}^{m}(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j})'(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j}) \\
= {}& tr\left(\sum_{j=1}^{m} T_j T_j'\right) + tr\left(\sum_{j=1}^{m}(Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j}\left((Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j}\right)'\right) \\
&-2tr\left(\sum_{j=1}^{m}(Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j}T_j'\right), \qquad (14.74)
\end{aligned}$$

where $a_0' = [\beta_0 : I_m] \in R^{m\times(m+1)}$.

The first summand depends on the data, but not the parameters. The second summand depends on the parameters, but not the data. Hence, these two terms are not important. The third term can be written as

$$\begin{aligned}
&-2tr\left(\sum_{j=1}^{m}\widetilde{\pi}_j T_j'\right) = -2\sum_{j=1}^{m}\sum_{\ell=1}^{k}\widetilde{\pi}_{j,\ell}T_{j,\ell}, \text{ where} \\
&\widetilde{\pi}_j = (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j} \in R^k, \\
&\widetilde{\pi}_j = (\widetilde{\pi}_{j,1}, ..., \widetilde{\pi}_{j,k})', \text{ and} \\
&T_j = (T_{j,1}, ..., T_{j,k})'. \qquad (14.75)
\end{aligned}$$

The parameters $\widetilde{\pi} = [\widetilde{\pi}_1 : \cdots : \widetilde{\pi}_m] \in R^{k\times m}$ are the "natural" parameters of the exponential family. There is a one-to-one transformation from π to $\widetilde{\pi}$ provided $Z'Z$ and Ω are nonsingular, which is assumed, $a_0' = [\beta_0 : I_m]$ is full row rank m, which holds by the definition of a_0, and $\alpha_0 = [\alpha_{0,1} : \cdots : \alpha_{0,m}] \in R^{(m+1)\times m}$ is full column rank m. The latter holds because $\Omega^{-1/2}\alpha_{0,1}, ..., \Omega^{-1/2}\alpha_{0,m}$ are orthogonal by construction, so $\Omega^{-1/2}\alpha_0 = [\Omega^{-1/2}\alpha_{0,1} : \cdots : \Omega^{-1/2}\alpha_{0,m}]$ is full column rank m and, in turn, α_0 is full column rank using the fact that Ω is nonsingular. The parameter space for π includes a km-dimensional rectangle. Hence, the same is true for $\widetilde{\pi}$. We conclude that the family of distributions of T under H_0 is a km-parameter exponential family with parameter space that contains a km-dimensional rectangle. □

Proof of Lemma 13. First, we establish part (a). The $k \times (m+1)$ matrix $[S{:}T]$ is multivariate normal with mean matrix $M = \mu_\pi \Delta_\beta$, all variances equal to one, and all correlations equal to zero by Lemma 12. Hence, $Q = [S{:}T]'[S{:}T]$ has a noncentral Wishart distribution with k degrees of freedom, covariance matrix I_{m+1}, and matrix of noncentrality parameters $M'M = \Delta_\beta' \lambda \Delta_\beta$, where $\lambda = \mu_\pi' \mu_\pi$. By (10.3.1) of Muirhead (1982), the density of Q at q is as given in part (a) of the lemma.

Part (b) is established as follows. The distribution of Q_T is a noncentral Wishart distribution with k degrees of freedom, covariance matrix I_m, and matrix of noncentrality parameters $\Delta_{T,\beta}' \lambda \Delta_{T,\beta}$ by Lemma 12(b). By (10.3.1) of Muirhead (1982), the density of Q_T at q_T is as given in part (b) of the lemma.

For part (c), by calculating the ratio of the densities in parts (a) and (b) of the lemma evaluated at $\beta = \beta_0$ and using the fact that $tr(\Delta_{\beta_0}' \lambda \Delta_{\beta_0}) = tr(\Delta_{T,\beta_0}' \lambda \Delta_{T,\beta_0})$, we obtain

$$f_{Q_1|Q_T}(q_1|q_T;\beta_0,\lambda) = K_{1,m} K_{2,m}^{-1} |q|^{(k-m-2)/2} |q_T|^{-(k-m-1)/2} etr(-q_S/2) \quad (14.76)$$
$$\times\ {}_0F_1\Big(k/2; \Delta_{\beta_0}' \lambda \Delta_{\beta_0} q/4\Big)\Big) \Big({}_0F_1\Big(k/2; \Delta_{T,\beta_0}' \lambda \Delta_{T,\beta_0} q_T/4\Big)\Big)\Big)^{-1}.$$

We show below that the conditional distribution of Q_1 given $Q_T = q_T$ does not depend on λ. Hence, we can take $\lambda = 0$ in (14.76). Because ${}_0F_1\,(k/2; 0_{m\times m}) = 1$ for all positive integers m (e.g., see Muirhead (1982) p. 226 for the case $m = 1$ and pp. 227-8 and p. 258 for the case $m \geq 1$), this yields the expression given in part (c) of the lemma.

The conditional distribution of Q_1 given $Q_T = q_T$ does not depend on λ by the following argument. Theorem 15 states that invariant tests are similar if and only if they have *Neyman structure* with respect to Q_T (e.g., as defined in Lehmann (1986, pp. 141-2)). By Theorem 4.2 of Lehmann (1986, p. 144), the latter implies that Q_T is a boundedly complete sufficient statistic under H_0 for the parameter $\lambda > 0$. Sufficiency of Q_T implies the desired result.

An alternative (and more direct) proof that the conditional distribution of Q_1 given $Q_T = q_T$ does not depend on λ is the following: (i) there is a one-to-one transformation from Q_1 to $\widetilde{Q}_1 = (Q_S, S'T_1/||S||, ..., S'T_m/||S||)$, so it suffices to show that the conditional distribution of $\widetilde{Q}_1$ does not depend on λ, (ii) the distribution of $\widetilde{Q}_1$ depends on $T = [T_1 : \cdots : T_m]$ only through $T_j' T_\ell$ for $j, \ell = 1, ..., m$ by the spherical symmetry of the null distribution of S, which is $N(0, I_k)$ by Lemma 12(a), (iii) by (ii) the conditional distribution of $\widetilde{Q}_1$ given $Q_T = T'T$ is the same as the conditional distribution of $\widetilde{Q}_1$ given T, and (iv) the conditional distribution of $\widetilde{Q}_1$ given T is a random function of S only and the null distribution of S is $N(0, I_k)$, which does not depend on λ. □

References

Anderson, T. W. (1946): "The Non-central Wishart Distribution and Certain Problems of Multivariate Statistics," *Annals of Mathematical Statistics,* 17, 409-431.

Anderson, T. W. and H. Rubin (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *Annals of Mathematical Statistics,* 21, 570-582.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Andrews, D. W. K. and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2004): "Supplement to 'Optimal Invariant Similar Tests for Instrumental Variables Regression'," unpublished manuscript, Department of Economics, Harvard University. Available on the web at http://ksghome.harvard.edu/∼.JStock.Academic.Ksg/ams/websupp/index.htm.

Bekker, P. and F. Kleibergen (2003): "Finite-Sample Instrumental Variables Inference Using an Asymptotically Pivotal Statistic," *Econometric Theory,* 19, 744-753.

Chamberlain, G. (2003): "Instrumental Variables, Invariance, and Minimax," unpublished manuscript, Department of Economics, Harvard University.

Donald, S.G. and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica,* 69, 1161-1191.

Dudley, R. M. (1989): *Real Analysis and Probability.* Pacific Grove, CA: Wadsworth and Brooks/Cole.

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

Dufour, J.-M. and J. Jasiak (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42, 815-843.

Dufour, J.-M. and M. Taamouti (2003): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," unpublished manuscript, Department of Economics, University of Montreal.

Forchini, G. and G. Hillier (2003): "Conditional Inference for Possibly Unidentified Structural Equations," *Econometric Theory,* 19, 707-743.

Hahn, J. and J. Hausman (2003): "Weak Instruments: Diagnosis and Cures in Empirical Economics," *American Economic Review*, 93, 118-125.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory.* New York: Academic Press.

Heyde, C. C. (1975): "On the Central Limit Theorem and Iterated Logarithm Law for Stationary Processes," *Bulletin of the Australian Mathematical Society*, 12, 1–8.

Hillier, G. H. (1984): "Hypothesis Testing in a Structural Equation: Part I, Reduced Form Equivalence and Invariant Test Procedures," unpublished manuscript, Department of Econometrics and Operations Research, Monash University.

Johnson, N. L. and S. Kotz (1970): *Distributions in Statistics: Continuous Univariate Distributions-2.* New York: Wiley.

——— (1972): *Distributions in Statistics: Continuous Multivariate Distributions.* New York: Wiley.

King, M.L. (1988), "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169-218.

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

Lebedev, N. N. (1965): *Special Functions and Their Applications.* Englewood Cliffs, NJ: Prentice-Hall.

Lehmann, E. L. (1986): *Testing Statistical Hypotheses*, second edition. New York: Wiley.

Moreira, M. J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," unpublished manuscript, Department of Economics, University of California, Berkeley.

——— (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica,* 71, 1027-1048.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

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

Startz, R., E. Zivot, and C. R. Nelson (2003): "Improved Inference in Weakly Identified Instrumental Variables Regression," unpublished manuscript, Department of Economics, University of Washington.

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, 147-162.

van der Vaart, A. W. (1998): *Asymptotic Statistics.* Cambridge, U. K.: Cambridge University Press.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

Wang, J. and Zivot, E. (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389-1404.

Zivot, E., Startz, R. and Nelson, C. R. (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119-1144.

Figure 1. Power envelopes for one-sided invariant similar and nonsimilar tests and power function of the one-sided CLR (CLR1) test; $\rho = .95$, $k = 5$

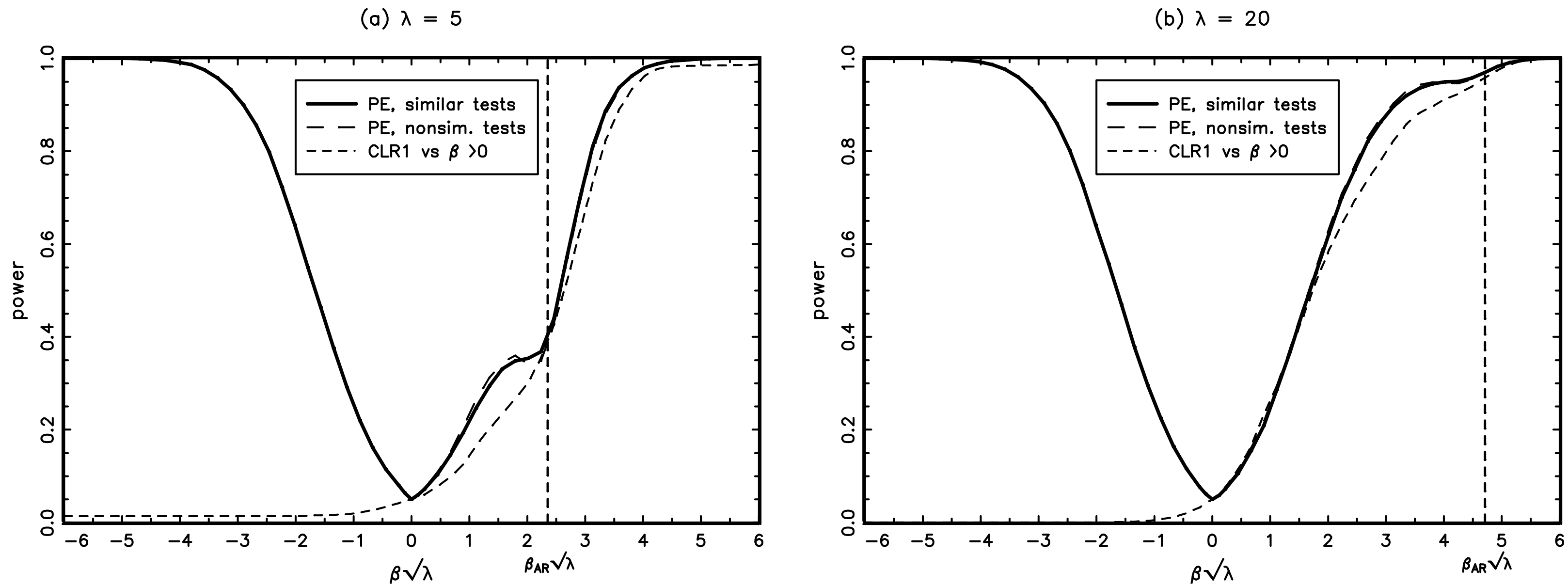

Figure 2. Power envelope for one-sided invariant similar tests and power functions of various POI similar tests: locally most powerful, POI at powers of .25, .5, and .75, and the most-distant most-powerful; $\rho = .95$, $k = 5$

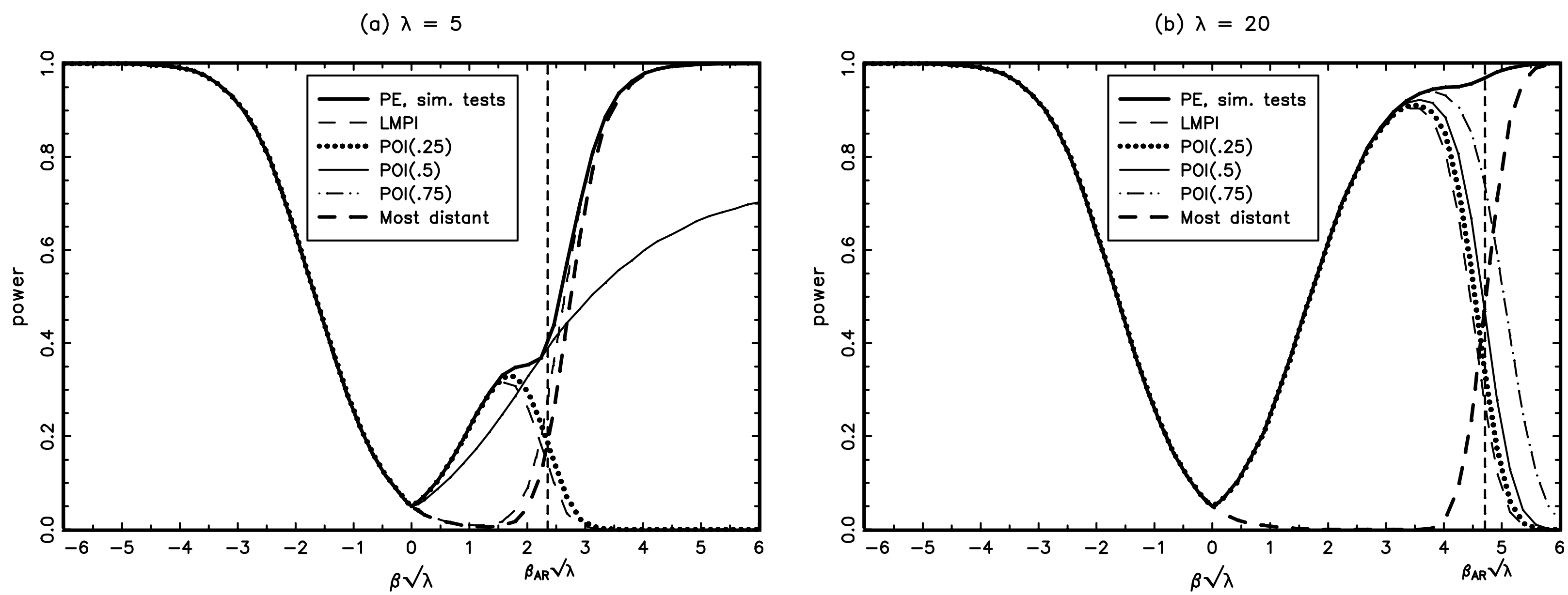

Figure 3. Power envelopes for two-sided asymptotically efficient invariant similar tests, locally unbiased invariant similar tests, and asymptotically efficient invariant nonsimilar tests; $\rho = .95$, $k = 5$

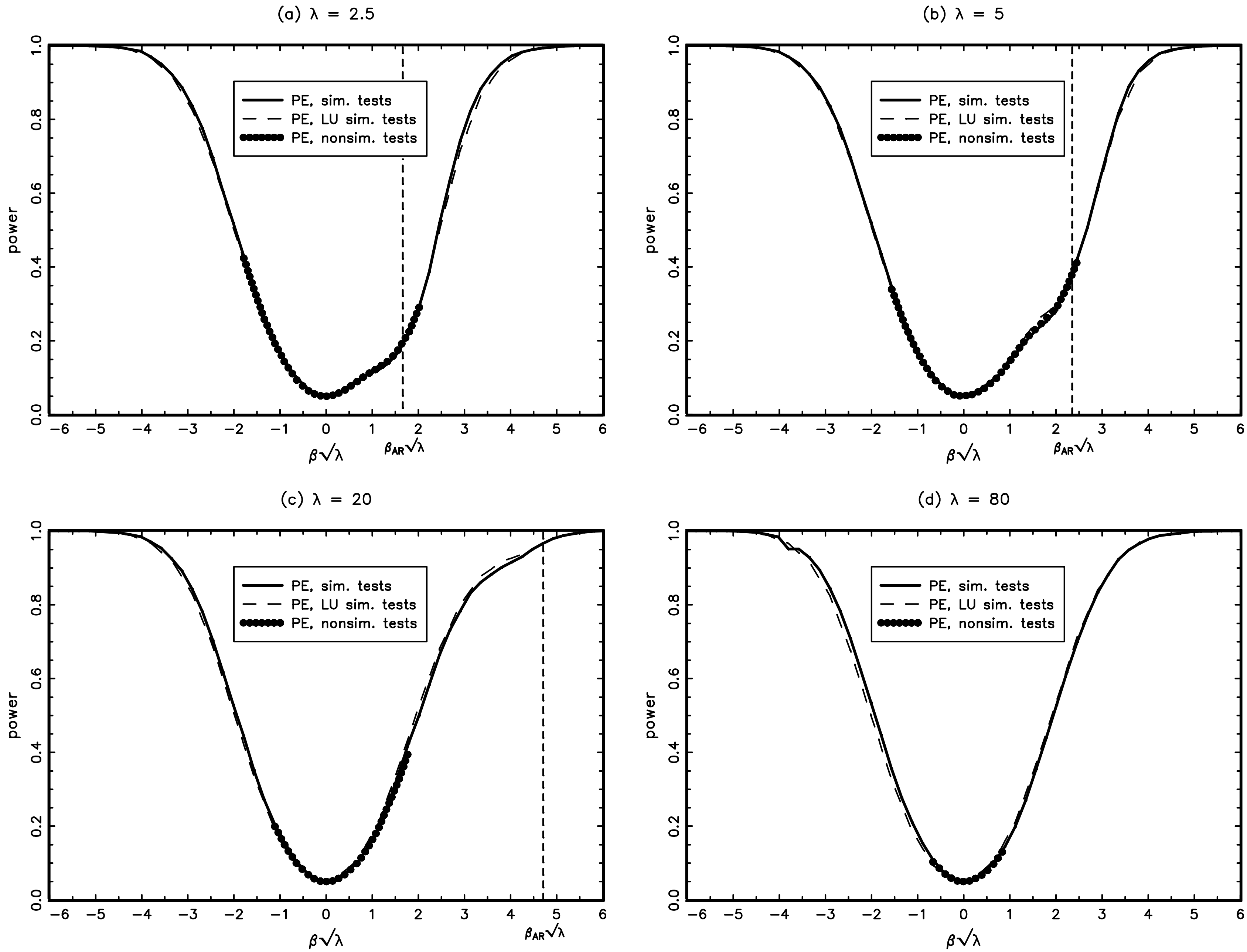

Figure 4. Power envelope for two-sided asymptotically efficient invariant similar tests and power functions of two-sided POI similar tests that are point optimal against $\beta^* = 0.8$, $\lambda^* = 5$ (POIS2(.8,5)) and against $\beta^* = 1.45$, $\lambda^* = 5$ (POIS2(1.45,5)); $\rho = .95$, $k = 5$

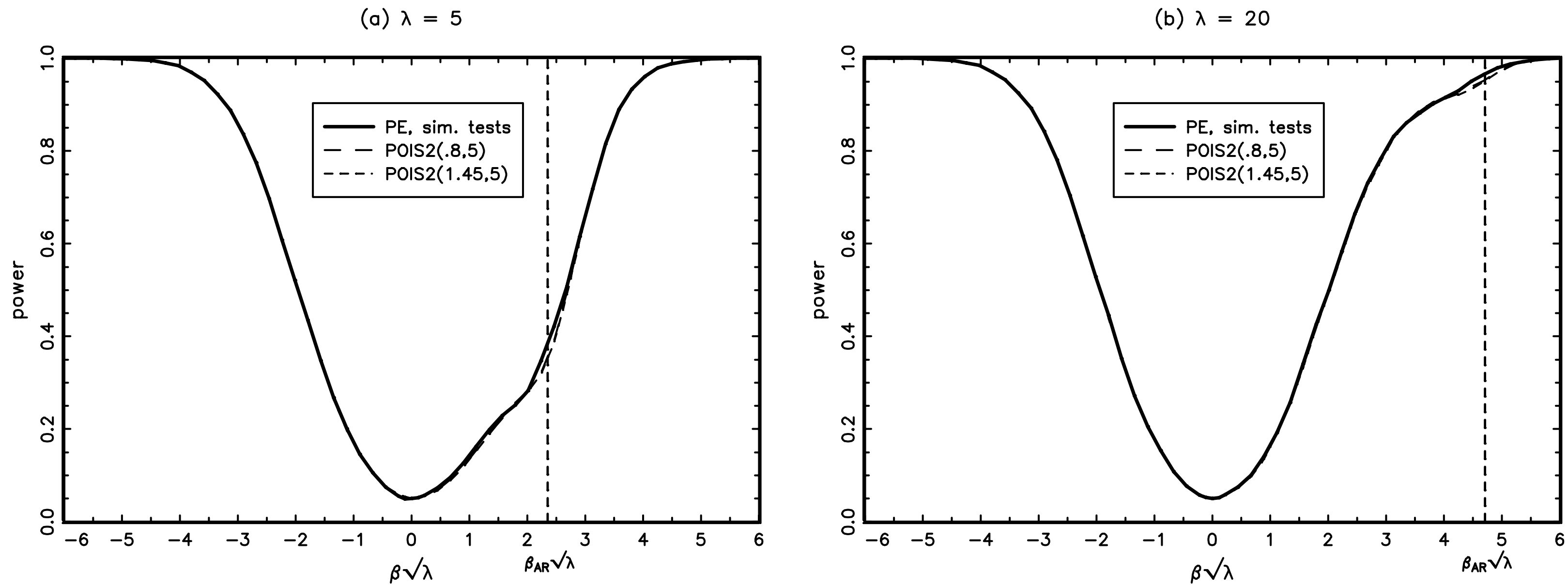

Figure 5. Power envelopes for two-sided asymptotically efficient invariant similar tests and power functions for the two-sided CLR, LM, AR, and P* tests; $\rho = .95$, $k = 5$

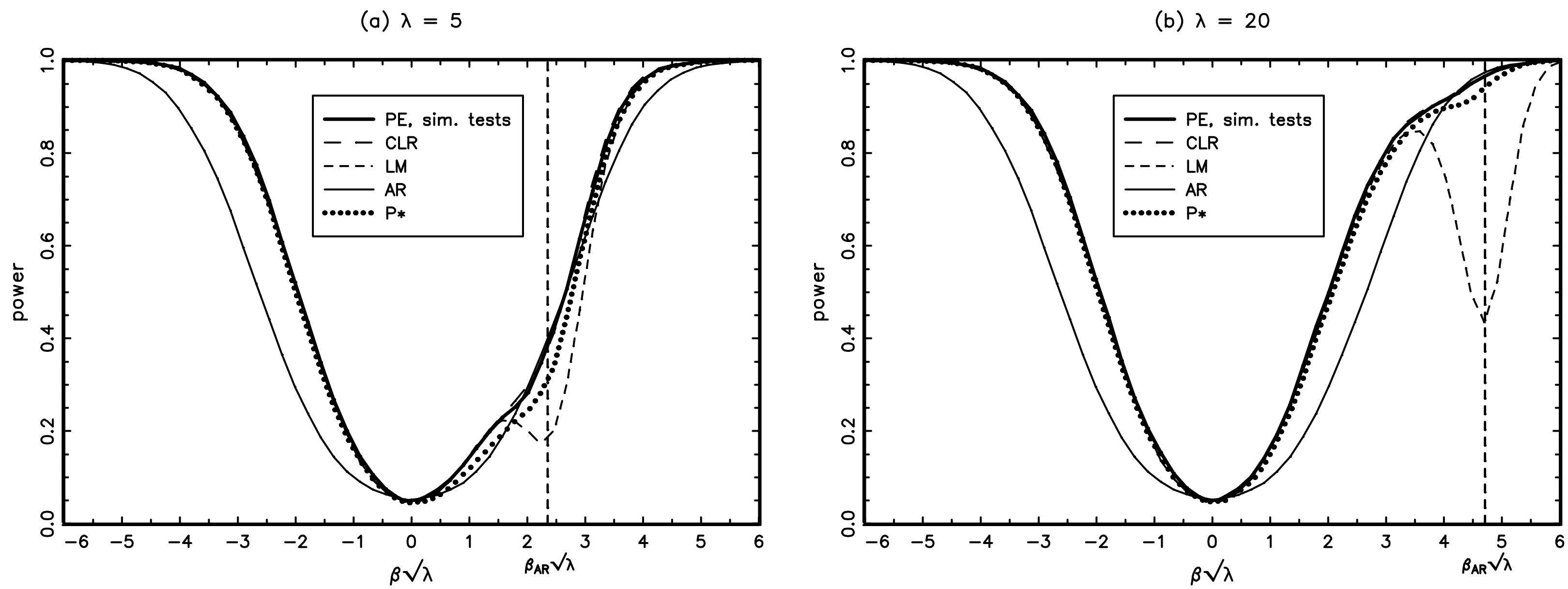

Figure 6. Power envelopes for two-sided asymptotically efficient invariant similar tests and power functions for the two-sided CLR, LM, AR, and P* tests; $\rho = .95$, $k = 2$ and $k = 10$

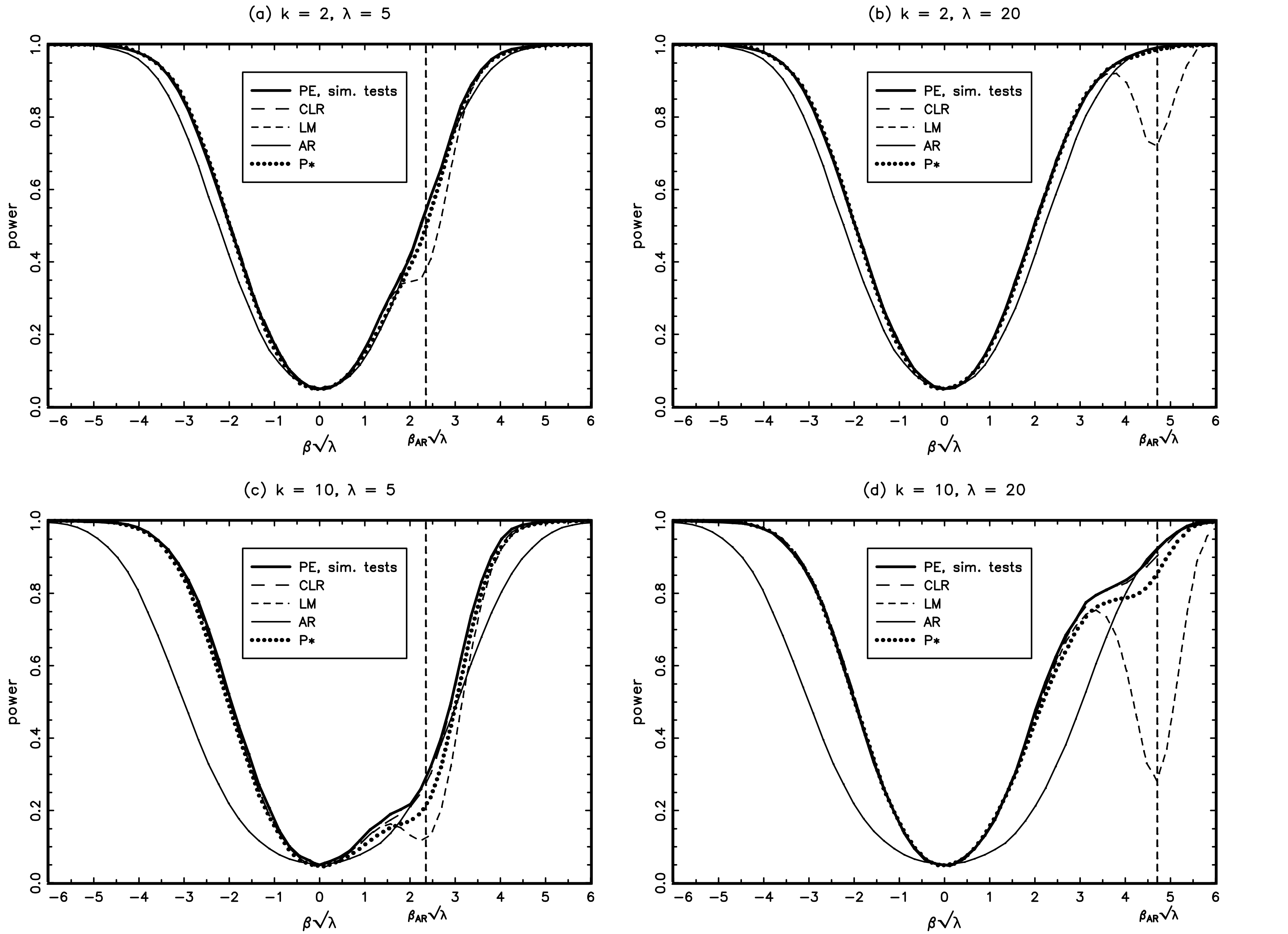

Table 1

Monte Carlo Rejection Rates of 5% 2-sided CLR, LM, AR, and P^* Tests with Ω estimated ($\rho = 0.5$)

$\beta\sqrt{\lambda}$	λ	*n*	*k* = 2				*k* = 5				*k* = 10			
			CLR	LM	AR	*P**	CLR	LM	AR	*P**	CLR	LM	AR	*P**
A. Size														
0	5	50	0.060	0.059	0.061	0.052	0.068	0.060	0.071	0.054	0.090	0.075	0.090	0.075
0	5	100	0.056	0.054	0.056	0.048	0.058	0.055	0.060	0.046	0.069	0.063	0.069	0.055
0	5	200	0.050	0.050	0.051	0.043	0.054	0.052	0.054	0.041	0.059	0.056	0.060	0.046
0	5	∞	0.050	0.050	0.050	0.045	0.050	0.050	0.050	0.038	0.050	0.050	0.050	0.036
0	20	50	0.058	0.058	0.061	0.054	0.061	0.058	0.071	0.055	0.073	0.065	0.090	0.069
0	20	100	0.054	0.054	0.056	0.050	0.055	0.054	0.060	0.048	0.061	0.058	0.069	0.055
0	20	200	0.050	0.050	0.051	0.046	0.052	0.051	0.054	0.045	0.054	0.053	0.060	0.048
0	20	∞	0.050	0.050	0.050	0.047	0.050	0.050	0.050	0.043	0.050	0.050	0.050	0.041
B. Power (size-unadjusted)														
-2.0	5	50	0.481	0.475	0.416	0.460	0.421	0.408	0.310	0.389	0.365	0.342	0.256	0.335
-2.0	5	100	0.488	0.485	0.418	0.465	0.422	0.412	0.298	0.388	0.357	0.341	0.237	0.323
-2.0	5	200	0.490	0.487	0.413	0.465	0.418	0.408	0.293	0.382	0.357	0.345	0.229	0.321
-2.0	5	∞	0.484	0.483	0.417	0.461	0.431	0.423	0.292	0.390	0.352	0.345	0.217	0.314
-2.0	20	50	0.498	0.497	0.416	0.484	0.473	0.470	0.310	0.450	0.436	0.428	0.256	0.411
-2.0	20	100	0.503	0.502	0.418	0.490	0.473	0.473	0.298	0.450	0.437	0.432	0.237	0.410
-2.0	20	200	0.507	0.506	0.413	0.491	0.475	0.472	0.293	0.448	0.444	0.440	0.229	0.411
-2.0	20	∞	0.502	0.500	0.417	0.486	0.488	0.486	0.292	0.455	0.445	0.443	0.217	0.405
2.0	5	50	0.433	0.376	0.415	0.403	0.328	0.250	0.306	0.285	0.280	0.202	0.258	0.241
2.0	5	100	0.430	0.374	0.413	0.398	0.321	0.243	0.303	0.276	0.258	0.182	0.239	0.215
2.0	5	200	0.430	0.372	0.414	0.398	0.319	0.242	0.297	0.273	0.241	0.172	0.227	0.198
2.0	5	∞	0.433	0.377	0.418	0.400	0.311	0.233	0.294	0.260	0.223	0.156	0.210	0.180
2.0	20	50	0.486	0.482	0.415	0.466	0.425	0.416	0.306	0.394	0.378	0.355	0.258	0.346
2.0	20	100	0.487	0.484	0.413	0.465	0.430	0.422	0.303	0.397	0.372	0.356	0.239	0.336
2.0	20	200	0.491	0.489	0.414	0.470	0.434	0.427	0.297	0.399	0.365	0.352	0.227	0.329
2.0	20	∞	0.497	0.494	0.418	0.472	0.429	0.422	0.294	0.398	0.356	0.348	0.210	0.317

Notes: Rows with $n = \infty$ correspond to the weak-instrument asymptotic limit. Conditional p-values for the CLR statistic were computed by numerical integration of the conditional density based on Lemma 3. The P^* tests were computed using $\kappa = 2$ for $k = 2$, $\kappa = 3.25$ for $k = 5$, and $\kappa = 4.25$ for $k = 10$. All rejections are based on the weak-instrument asymptotic critical values or conditional critical values, as appropriate. All calculations are based on 50,000 Monte Carlo simulations.