Econometrica, Vol. 74, No. 3 (May, 2006), 715–752

OPTIMAL TWO-SIDED INVARIANT SIMILAR TESTS FOR INSTRUMENTAL VARIABLES REGRESSION

BY DONALD W. K. ANDREWS, MARCELO J. MOREIRA, AND JAMES H. STOCK[1]

This paper considers tests of the parameter on an endogenous variable 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 a two-sided power envelope for invariant similar tests. This allows us to assess and compare the power properties of tests such as the conditional likelihood ratio (CLR), the Lagrange multiplier, and the Anderson–Rubin tests. We find that the CLR test is quite close to being uniformly most powerful invariant among a class of two-sided tests.

The finite-sample results of the paper are extended to the case of unknown error covariance matrix and possibly nonnormal 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: Average power, instrumental variables regression, invariant tests, optimal tests, power envelope, similar tests, two-sided tests, weak instruments.

1. INTRODUCTION

IN INSTRUMENTAL VARIABLES (IVs) regression with a single included endogenous regressor, instruments are said to be weak when the partial correlation between the IVs and the included endogenous regressor is small, given the included exogenous regressors. The effect of weak IVs 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 IVs are weak. Except for the important early contribution by Anderson and Rubin (1949) (AR), most of this literature is recent. It includes the papers by Staiger and Stock (1997), Zivot, Startz, and Nelson (1998), Wang and Zivot (1998), Dufour and Jasiak (2001), Moreira (2001, 2003), Kleibergen (2002,

[1]Andrews, Moreira, and Stock gratefully acknowledge the research support of the National Science Foundation via Grant numbers SES-00-01706 and SES-04-17911, SES-04-18268, and SBR-02-14131, respectively. The authors thank three referees, a co-editor, Tom Rothenberg, Jean-Marie Dufour, Grant Hillier, Anna Mikusheva, and seminar and conference participants at Harvard/MIT, Michigan, Michigan State, Queen's, UCLA/USC, UCSD, the Yale Statistics Department, the 2003 NBER/NSF Conference on Weak Instruments at MIT, the 2004 Far-Eastern Econometric Society Meetings in Seoul, and the 2004 Canadian Econometrics Study Group Meetings in Toronto for helpful comments.

2004), Dufour and Taamouti (2005), Guggenberger and Smith (2005, 2006), and Otsu (2006). None of these contributions develops a satisfactory theory of optimal inference in the presence of potentially weak IVs.

The purpose of this paper is to develop a theory of optimal hypothesis testing when IVs might be weak, and to use this theory to develop practical valid hypothesis tests that are nearly optimal whether the IVs are weak or strong. We adopt the natural invariance condition that inferences are unchanged if IVs are transformed by an orthogonal matrix, e.g., changing the order in which the IVs 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. Some results for multiple endogenous variables are provided by Andrews, Moreira, and Stock (2004) (hereafter denoted AMS04).

We show that there does not exist a uniformly most powerful invariant (UMPI) two-sided similar test of $H_0: \beta = \beta_0$ when the model is overidentified, although there is one when the model is just identified. Our numerical results for the overidentified case, 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 two-sided UMPI among similar tests when the model is overidentified and is exactly so when the model is just identified. We recommend the use of the CLR test in empirical practice.

On the other hand, the power of the Lagrange multiplier (LM) test of Kleibergen (2002) and Moreira (2001) is never above that of the CLR test, and in some cases is far below (when the model is overidentified). Hence, the CLR test dominates the LM test in terms of power and we do not recommend the LM test for practical use.

An important use of tests concerning β is the construction of confidence intervals (or sets) obtained by inverting the tests. (Specifically, the set of β_0 values for which $H_0: \beta = \beta_0$ cannot be rejected at level α yields a $100(1-\alpha)\%$ confidence interval for the true β value.) The near optimality of the CLR test yields a corresponding near optimality of the CLR-based confidence set. The latter (nearly) minimizes, among $100(1-\alpha)\%$ confidence sets, the probability of incorrectly including a given β value, call it β_0, in the confidence interval when the true value is an arbitrary value to the left of β_0, say β^*, averaged with the probability of incorrectly including β_0 when the true value is some particular value to the right of β_0, say β_2^* (which depends on β^*).

The optimality results are developed for strictly exogenous IVs, linear structural and reduced-form equations, and homoskedastic Gaussian errors with a known covariance matrix. For this model, we obtain sufficient statistics, a maximal invariant (under orthogonal transformations of the IVs), and the distribution of the maximal invariant. We determine necessary and sufficient conditions for invariant tests to be similar.

We construct a two-sided power envelope for invariant similar tests. There are different ways to do so depending on how one imposes two-sidedness. Here, we impose two-sidedness by comparing tests based on their average power for two parameter values—one greater than the null value β_0 and the other less than β_0. The power envelope is mapped out by a class of two-point optimal invariant similar (POIS2) tests. The choice of which parameter values to pair with each other is determined such that the resultant POIS2 tests are asymptotically efficient (AE) under strong IV asymptotics. In consequence, we refer to this power envelope as the AE two-sided power envelope for invariant similar tests.

The foregoing results are developed by treating the reduced-form error covariance matrix as known. In practice, this matrix is unknown and must be estimated. Using Staiger and 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 nonnormal errors. We show that the finite-sample power envelope derived with known covariance matrix is also the asymptotic Gaussian power envelope with unknown covariance matrix, under weak-IV asymptotics. In a Monte Carlo study reported in AMS04, we find that, for normal errors and unknown covariance matrix Ω, sample sizes of 100–200 observations are sufficient for (i) the sizes of the CLR, LM, and AR tests with estimated covariance matrices 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.

Finally, we obtain asymptotic properties of the tests considered in this paper when the IVs are strong. These results are essential for determining the class of POIS2 tests that are asymptotically efficient under strong IVs, which lies behind the construction of the two-sided power envelope. The CLR and LM tests are shown to be asymptotically efficient with strong IVs against local alternatives, although (as is known) the AR test is not. In AMS04, the CLR, LM, AR, and POIS2 tests are shown to be consistent against fixed alternatives under strong IVs.

In addition to similar tests, AMS04 considers optimal nonsimilar tests using the least-favorable distribution approach described, e.g., by Lehmann (1986). Although the nonsimilar and similar tests differ in theory, AMS04 finds that the power envelopes of invariant similar and nonsimilar tests are numerically very close.

Numerous additional numerical results that supplement those given in Section 5 are provided in Andrews, Moreira, and Stock (2006b) (denoted AMS06b), which also provides detailed tables of conditional critical values for the CLR test. Extensions and results related to this paper, including optimal one-sided tests and versions of the CLR, LM, and AR test statistics that are robust to heteroskedasticity and/or autocorrelation, are provided in AMS04.

Other papers that consider optimal testing in the exact Gaussian IV regression model are papers by 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 when the model is overidentified. Chamberlain (2003) considers minimax decision procedures and his results for tests show that the imposition of the invariance condition considered here does not affect the minimax decision problem.

The remainder of this paper is organized as follows. Section 2 introduces the model and determines sufficient statistics for the 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 introduces POIS2 tests and determines a two-sided power envelope for normal errors and known error covariance matrix Ω. Section 5 presents numerical results that show that the CLR test has power essentially on the power envelope, whereas the LM and AR tests have power that is sometimes on, and sometimes well below, the power envelope. Section 6 analyzes the asymptotic properties of the POIS2 tests under weak IVs, possibly nonnormal errors, and unknown Ω. These results are used to determine a weak-IV asymptotic two-sided power envelope for the case of independent and identically distributed (i.i.d.) normal errors and unknown Ω. Section 7 establishes the asymptotic properties of CLR and POIS2 tests under strong IVs when Ω is unknown and the errors may be nonnormal. 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 IVs, and normal errors with known covariance matrix. In later sections, we allow for nonnormal errors with unknown covariance matrix.

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

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

where $y_1, y_2 \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, and $\widetilde{Z} \in \mathbb{R}^{n\times k}$ are observed variables; $u, v_2 \in \mathbb{R}^n$ are unobserved errors; and $\beta \in \mathbb{R}$, $\pi \in \mathbb{R}^k$, and $\gamma_1, \xi_1 \in \mathbb{R}^p$ are unknown parameters. The exogenous variable matrix X and the IV matrix $\widetilde{Z}$ are fixed (i.e., nonstochastic), and $[X:\widetilde{Z}]$ has full column rank $p+k$. The $n\times 2$ matrix of errors $[u:v_2]$ is i.i.d. across rows, with each row having a mean zero bivariate normal distribution.

Our interest is in the null and alternative hypotheses

$$(2.2) \qquad H_0: \beta = \beta_0 \quad \text{and} \quad H_1: \beta \neq \beta_0.$$

We transform $\widetilde{Z}$ so that the transformed IV matrix, Z, is orthogonal to X:

$$(2.3)\qquad \begin{aligned} &y_2 = Z\pi + X\xi + v_2, \quad \text{where} \\ &Z = M_X\widetilde{Z}, \quad M_X = I_n - P_X, \quad P_X = X(X'X)^{-1}X', \\ &\xi = \xi_1 + (X'X)^{-1}X'\widetilde{Z}\pi, \quad \text{and} \quad Z'X = 0. \end{aligned}$$

The two reduced-form equations are

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

The reduced-form errors $[v_1 : v_2]$ are i.i.d. across rows, with each row having a mean zero bivariate normal distribution with 2×2 nonsingular covariance matrix Ω. For the purposes of obtaining an exact power envelope, we suppose Ω is known. Below we show that the asymptotic power envelope for unknown Ω and weak IVs is the same as the exact envelope with known Ω.

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

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

The distribution of $Y \in \mathbb{R}^{n\times 2}$ 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 $\mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^p \times \mathbb{R}^p$.

Because the multivariate normal is a member of the exponential family of distributions, low-dimensional sufficient statistics are available.

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 $(\beta, \pi')'$*;
(d) *$Z'Y$ has a multivariate normal distribution that does not depend on $\eta = [\gamma : \xi]$*;
(e) *$Z'Y$ is a sufficient statistic for $(\beta, \pi')'$.*

For tests concerning β, there is no loss (in terms of attainable power functions) in considering tests that are based on the sufficient statistic $Z'Y$ for $(\beta, \pi')'$. This eliminates the nuisance parameters $\eta = [\gamma : \xi]$ from the problem. The nuisance parameter π remains. As in Moreira (2003), we consider a one-to-one transformation of $Z'Y$:

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

and $A^{-1/2}$ denotes the symmetric square root of a positive semi-definite matrix A.[2]

The means of S and T depend on the quantities

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

The distributions of S and T 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) $T \sim N(d_\beta \mu_\pi, I_k)$;
(c) *S and T are independent.*

COMMENTS: (i) Lemma 2 holds under H_0 and H_1. Under H_0, S has mean zero.

(ii) The constant d_β that appears in the mean of T can be rewritten as

$$
(2.8)\qquad d_\beta = b'\Omega b_0 \cdot (b_0'\Omega b_0)^{-1/2} (\det(\Omega))^{-1/2}, \quad \text{where} \quad b = (1, -\beta)'.
$$

(iii) The proofs of Lemmas 1 and 2 are standard; see AMS06b for details.

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. Rotations of the coordinate system are equivalent to rotations of the k IVs. Hence, we consider statistics that are invariant to orthonormal transformations of the IVs. We note that Hillier (1984) and Chamberlain (2003) considered similar invariance conditions.

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

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

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

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 a distribution with parameters $\overline{g}_F(\beta, \pi)$, by Lehmann (1986, p. 283). (The second element of $\overline{g}_F$ is determined by $F\mu_\pi = \mu_{\overline{g}_F(\pi)}$, which holds when $\overline{g}_F(\pi) = (Z'Z)^{-1/2}F(Z'Z)^{1/2}\pi$.) Furthermore, the problem of testing H_0 versus H_1 remains invariant under $g_F \in G$ 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 $j = 0, 1$). Invariance under the transformation group G ensures that tests of H_0 are unaffected by changing the units of Z or by respecifying binary units as contrasts.

Note that orthonormal transformations of the k IVs lead to the transformations in (3.1). In particular, the transformation $Z \to ZF'$ corresponds to $[S:T] \to F[S:T]$.[3]

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 Theorem 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

$$
\begin{aligned}
(3.2) \qquad Q &= [S:T]'[S:T] = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix} = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix}, \\
Q_1 &= (S'S, S'T)' = (Q_S, Q_{ST})'.
\end{aligned}
$$

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.

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

[3]This holds because $(FZ'ZF')^{-1/2} = (FB\Lambda B'F')^{-1/2} = FB\Lambda^{-1/2}B'F' = F(Z'Z)^{-1/2}F'$, where $Z'Z = B\Lambda B'$ for an orthogonal $k \times k$ matrix B and a diagonal $k \times k$ matrix Λ.

[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 $x_2 = Fx_1$.

COMMENTS: (i) The statistic Q has a noncentral 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

$$\lambda = \pi' Z' Z \pi \geq 0. \tag{3.3}$$

Thus, the utilization of invariance has reduced the k-vector nuisance parameter π to a scalar nuisance parameter λ.

(ii) Examples of invariant tests in the literature include the AR test; the standard likelihood ratio (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 conditional Wald tests of Moreira (2003), which depend on the standard LR and Wald test statistics coupled with "conditional" critical values that depend on Q_T. The LR, LM, and AR test statistics depend on Q or (S, T) in the following ways:

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

(The above expression for LR is simpler than, but equivalent to, the expression given by Moreira (2003).) 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) and Wald tests based on the Chamberlain and Imbens (2004) many IV estimator.

A test based on the maximal invariant Q is *similar* if its null rejection rate does not depend on the parameter π that determines the strength of the IVs Z. (See Lehmann (1986) for a general discussion of similarity.) 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. Invariant similar tests do not suffer from this problem. Using the argument of Moreira (2001), we characterize the class of invariant similar tests.

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

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 $E_{\beta_0}(\cdot|Q_T = q_T)$ de-*

notes conditional expectation given $Q_T = q_T$ *when* $\beta = \beta_0$ (*which does not depend on* π).

COMMENTS: (i) 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 $q_T > 0$.

(ii) 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

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

where $\kappa_{\text{LR},\alpha}(Q_T)$ is defined to satisfy $P_{\beta_0}(\text{LR} > \kappa_{\text{LR},\alpha}(Q_T)|Q_T = q_T) = \alpha$ and the conditional distribution of Q_1 given Q_T is specified in Lemma 3(c) below. See AMS06b for tables of conditional critical values for the CLR test. A GAUSS program for p-values of the CLR test is described by Andrews, Moreira, and Stock (2006a) and is available at James Stock's webpage.

4. TWO-SIDED POWER ENVELOPE

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

We address the question of optimal invariant similar tests when the IVs may be weak. We construct a power envelope for two-sided tests and show numerically that the CLR test essentially lies on the power envelope and, hence, is essentially an optimal two-sided invariant similar test.

There are several ways to construct a two-sided power envelope, depending on how one imposes the two-sidedness condition. Three approaches are to (i) consider average power (AP) for β values less than and greater than the null value β_0, (ii) impose a sign invariance condition, and (iii) impose a necessary condition for unbiasedness. We develop approach (i) in detail here and briefly comment on approaches (ii) and (iii) at the end of this section (the details of which can be found in AMS06b). It turns out that approaches (i) and (ii) yield exactly the same power envelope, and approach (iii) yields a power envelope that is found numerically to be essentially the same as that of approaches (i) and (ii); see AMS06b.

Approach (i) is based on determining the highest possible average power against a point $(\beta, \lambda) = (\beta^*, \lambda^*)$ and some other point, say (β_2^*, λ_2^*), for which β_2^* lies on the other side of the null value β_0 from β^*. (The power envelope then is a function of $(\beta, \lambda) = (\beta^*, \lambda^*)$.) The naive "symmetric alternative" choice $(\beta_2^*, \lambda_2^*) = (2\beta_0 - \beta^*, \lambda^*)$ that yields $|\beta^* - \beta_0| = |\beta_2^* - \beta_0|$ is found to be a poor choice because the testing problem is not correspondingly "symmetric." In fact, the test that maximizes average power against these two points turns out to be a *one-sided* LM test asymptotically under strong-IV asymptotics for any choice of (β^*, λ^*) (see comment (iii) to Theorem 8). This indicates that the symmetric alternative choice of (β_2^*, λ_2^*) is not a good choice for generating two-sided tests.

How then should (β_2^*, λ_2^*) be defined? We are interested in tests that have good all-around two-sided power properties. This includes high power when the IVs are strong. In consequence, given a point (β^*, λ^*), we consider the point (β_2^*, λ_2^*) that has the property that the test that maximizes average power against these two points is *asymptotically efficient* under strong-IV asymptotics. As shown in Section 7, this point is unique. Furthermore, the power of the test that maximizes average power against these two points is the same for each of the two points. This choice also has the desirable properties that (a) β_2^* is on the other side of the null value β_0 from β^*, (b) the marginal distributions of Q_S, Q_{ST}, and Q_T under (β_2^*, λ_2^*) are the same as under (β^*, λ^*), and (c) 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^*.

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

$$(4.1) \qquad (\lambda_2^*)^{1/2} c_{\beta_2^*} = -(\lambda^*)^{1/2} c_{\beta^*} \ (\neq 0) \quad \text{and} \quad (\lambda_2^*)^{1/2} d_{\beta_2^*} = (\lambda^*)^{1/2} d_{\beta^*}.$$

This follows from Lemmas 2 and 3(a) below and $\lambda = \mu_\pi' \mu_\pi$. Note that c_β is proportional to $\beta - \beta_0$ and d_β is linear in β. We denote by β_{AR} the point β at which $d_\beta = 0$.[5] Provided $\beta^* \neq \beta_{\mathrm{AR}}$, the solutions to the two equations in (4.1) are

$$(4.2) \qquad \beta_2^* = \beta_0 - \frac{d_{\beta_0}(\beta^* - \beta_0)}{d_{\beta_0} + 2r(\beta^* - \beta_0)} \quad \text{and}$$

$$\lambda_2^* = \lambda^* \frac{(d_{\beta_0} + 2r(\beta^* - \beta_0))^2}{d_{\beta_0}^2}, \quad \text{where}$$

$$r = e_1' \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2} \quad \text{and} \quad e_1 = (1, 0)'.$$

[5]Surprisingly, the one-sided point-optimal invariant similar test against β_{AR} is the (two-sided) AR test, see AMS04a. Some calculations yield $\beta_{\mathrm{AR}} = (\omega_{11} - \omega_{12}\beta_0)/(\omega_{12} - \omega_{22}\beta_0)$, provided $\omega_{12} - \omega_{22}\beta_0 \neq 0$, where ω_{ij} denotes the (i, j) element of Ω.

(If $\beta^* = \beta_{\mathrm{AR}}$, there is no solution to (4.1) with β_2^* on the other side of β_0 from β^*.)

We refer to the power envelope based on maximizing average power against (β^*, λ^*) and (β_2^*, λ_2^*) with (β_2^*, λ_2^*) as in (4.1) as the *asymptotically efficient* (AE) two-sided power envelope for invariant similar tests.

The average power of a test $\phi(Q)$ against the two points (β^*, λ^*) and (β_2^*, λ_2^*) is given by

$$(4.3) \qquad K(\phi; \beta^*, \lambda^*) = \frac{1}{2}\left[E_{\beta^*,\lambda^*}\phi(Q) + E_{\beta_2^*,\lambda_2^*}\phi(Q)\right] = E^*_{\beta^*,\lambda^*}\phi(Q),$$

where $E_{\beta,\lambda}$ denotes expectation with respect to the density $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$, which is the joint density of (Q_1, Q_T) at (q_1, q_T) when (β, λ) are the true parameters, and $E^*_{\beta^*,\lambda^*}$ denotes expectation with respect to the density

$$(4.4) \qquad f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*) = \frac{1}{2}\left[f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*) + f_{Q_1,Q_T}(q_1, q_T; \beta_2^*, \lambda_2^*)\right].$$

Hence, the average power of $\phi(Q)$ against (β^*, λ^*) and (β_2^*, λ_2^*) can be written as the power against the single density $f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)$.

We want to find the test that maximizes average power against the alternatives (β^*, λ^*) and (β_2^*, λ_2^*) 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), average power against (β^*, λ^*) and (β_2^*, λ_2^*) equals unconditional power against the single density $f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)$. In turn, the latter equals expected conditional power given Q_T against $f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)$. Hence, it suffices to determine the test that maximizes conditional average 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

$$(4.5) \qquad K(\phi | Q_T = q_T; \beta^*, \lambda^*) = \int_{\mathbb{R}^+\times\mathbb{R}} \phi(q_1, q_T) f^*_{Q_1|Q_T}(q_1|q_T; \beta^*, \lambda^*)\, dq_1, \quad \text{where}$$

$$f^*_{Q_1|Q_T}(q_1|q_T; \beta^*, \lambda^*) = \frac{f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)}{f^*_{Q_T}(q_T; \beta^*, \lambda^*)},$$

$$f^*_{Q_T}(q_T; \beta^*, \lambda^*) = \frac{1}{2}\left[f_{Q_T}(q_T; \beta^*, \lambda^*) + f_{Q_T}(q_T; \beta_2^*, \lambda_2^*)\right],$$

and $f_{Q_T}(q_T; \beta, \lambda)$ is the density of Q_T at q_T when the true parameters are (β, λ).

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 λ under H_0, this conditional density does not depend on 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 λ. Given the average power criterion function $K(\phi;\beta^*,\lambda^*)$, the alternative hypothesis of concern is also simple. In particular, conditional on $Q_T = q_T$, the alternative density of interest is $f^*_{Q_1|Q_T}(q_1|q_T;\beta^*,\lambda^*)$. 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 form and rejects H_0 when the LR is sufficiently large. In particular, the point-optimal invariant similar two-sided (POIS2) test statistic is

$$\text{(4.6)}\qquad \mathrm{LR}^*(Q_1, q_T;\beta^*,\lambda^*) = \frac{f^*_{Q_1|Q_T}(Q_1|q_T;\beta^*,\lambda^*)}{f_{Q_1|Q_T}(Q_1|q_T;\beta_0)} = \frac{f^*_{Q_1,Q_T}(Q_1, q_T;\beta^*,\lambda^*)}{f^*_{Q_T}(q_T;\beta^*,\lambda^*) f_{Q_1|Q_T}(Q_1|q_T;\beta_0)}.$$

To provide an explicit expression for $\mathrm{LR}^*(Q_1, q_T;\beta^*,\lambda^*)$, 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 (4.4)–(4.6). These densities depend on the quantity

$$\text{(4.7)}\qquad \xi_\beta(q) = h_\beta' q h_\beta = c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T, \quad \text{where} \quad h_\beta = (c_\beta, d_\beta)'$$

and $q_1 = (q_S, q_{ST})'$. Note that $\xi_\beta(q) \geq 0$ because q is positive semidefinite almost surely.

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

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

where

$$q_1 = (q_S, q_{ST})' \in \mathbb{R}^+ \times \mathbb{R}, \quad q_T \in \mathbb{R}^+,$$

$$q = \begin{bmatrix} q_S & q_{ST} \\ q_{ST} & q_T \end{bmatrix}, \quad K_1^{-1} = 2^{(k+2)/2} pi^{1/2} \Gamma((k-1)/2),$$

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

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

$$f_{Q_T}(q_T;\beta,\lambda) = K_2 \exp\left(-\frac{\lambda d_\beta^2}{2}\right) q_T^{(k-2)/2} \exp\left(-\frac{q_T}{2}\right)(\lambda d_\beta^2 q_T)^{-(k-2)/4} \times I_{(k-2)/2}\left(\sqrt{\lambda d_\beta^2 q_T}\right)$$

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: (i) 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.[6] 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 i.i.d. 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.

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

$$I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{(x^2/4)^j}{j!\Gamma(\nu+j+1)} \tag{4.8}$$

[6]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, Eq. (6)) and are not consistent with those of Johnson and Kotz (1970, Eq. (5), p. 133; 1972, Eq. (50), p. 176). Sawa (1969, footnote 6) notes that Anderson's (1946) Equation (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 $\Sigma = I_k$.

for $x \geq 0$, e.g., see Lebedev (1965, p. 108). 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 $\mathbb{R}^+$; 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, $I_{-1/2}(x) = x^{-1/2}(2/pi)^{1/2}(\exp(x) + \exp(-x))/2$ (which arises when $k = 1$) and $I_{1/2}(x) = x^{-1/2}(2/pi)^{1/2}(\exp(x) - \exp(-x))/2$ (which arises when $k = 3$).

Equations (4.4)–(4.6) and Lemma 3 combine to give the following result for the POIS2 test statistic.

COROLLARY 1: *The optimal average-power test statistic against* (β^*, λ^*) *and* (β_2^*, λ_2^*), *where* (β_2^*, λ_2^*) *satisfies* (4.1), *is*

$$\begin{aligned} \mathrm{LR}^*(q_1, q_T; \beta^*, \lambda^*) &= \frac{f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)}{f^*_{Q_T}(q_T; \beta^*, \lambda^*) f_{Q_1|Q_T}(q_1|q_T; \beta_0)} \\ &= \frac{\psi(q_1, q_T; \beta^*, \lambda^*) + \psi(q_1, q_T; \beta_2^*, \lambda_2^*)}{\psi_2(q_T; \beta^*, \lambda^*) + \psi_2(q_T; \beta_2^*, \lambda_2^*)}, \end{aligned}$$

where

$$\begin{aligned} &\psi(q_1, q_T; \beta, \lambda) \\ &\quad = \exp\left(-\frac{\lambda(c_\beta^2 + d_\beta^2)}{2}\right)(\lambda \xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda \xi_\beta(q)}\right), \\ &\psi_2(q_T; \beta, \lambda) = \exp\left(-\frac{\lambda d_\beta^2}{2}\right)(\lambda d_\beta^2 q_T)^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\lambda d_\beta^2 q_T}\right), \end{aligned}$$

and c_β, d_β, *and* $\xi_\beta(q)$ *are defined in* (2.7) *and* (4.7).

COMMENTS: (i) Computation of the integrands of $\psi(q_1, q_T; \beta, \lambda)$ and $\psi_2(q_T; \beta, \lambda)$ in Corollary 1 are easy and extremely fast using GAUSS or Matlab functions to compute the modified Bessel function of the first kind. Hence, calculation of the test statistic $\mathrm{LR}^*(Q_1, Q_T; \beta^*, \lambda^*)$ is very fast.

(ii) When $k = 1$, some calculations using the expression for $I_{-1/2}(x)$ given in comment (ii) to Lemma 3 show that the numerator of the right-hand side expression for $\mathrm{LR}^*(q_1, q_T; \beta^*, \lambda^*)$ in Corollary 1 is increasing in S^2 (see AMS06b). Hence, when $k = 1$, the AR test maximizes average power against (β^*, λ^*) and (β_2^*, λ_2^*) for all (β^*, λ^*) in the class of invariant similar tests. That is, the AR test is a uniformly most powerful (UMP) two-sided invariant similar test. When $k = 1$, LR = LM = kAR, so the same optimality property holds for the CLR and LM tests. In addition, Moreira (2001) shows that these tests are UMP unbiased when $k = 1$. The remainder of this paper focuses on the case $k > 1$.

The POIS2 test with significance level α rejects H_0 if

$$\text{LR}^*(Q_1, Q_T; \beta^*, \lambda^*) > \kappa_\alpha(Q_T; \beta^*, \lambda^*), \tag{4.9}$$

where $\kappa_\alpha(Q_T; \beta^*, \lambda^*)$ is defined by

$$P_{\beta_0}\big(\text{LR}^*(Q_1, q_T; \beta^*, \lambda^*) > \kappa_\alpha(q_T; \beta^*, \lambda^*)|Q_T = q_T\big) = \alpha. \tag{4.10}$$

Here, $P_{\beta_0}(\cdot|Q_T = q_T)$ denotes conditional probability given $Q_T = q_T$ under the null, which can be calculated using the density in Lemma 3(c). Note that $\kappa_\alpha(\cdot; \beta^*, \lambda^*)$ 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 $e_1 = (1, 0, \ldots, 0)' \in \mathbb{R}^k$. 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 $\kappa_\alpha(Q_T; \beta^*, \lambda^*)$ can be approximated by simulating n_{MC} i.i.d. random vectors $S_i \sim N(0, I_k)$ for $i = 1, \ldots, n_{\text{MC}}$, where n_{MC} is large, computing $Q_1(i) = (S_i'S_i, S_i'e_1 \cdot Q_T^{1/2})$ for $i = 1, \ldots, n_{\text{MC}}$, and taking $\ln(\kappa_\alpha(Q_T; \beta^*, \lambda^*))$ to be the $1 - \alpha$ sample quantile of $\{\ln(\text{LR}^*(Q_1(i), Q_T; \beta^*, \lambda^*)) : i = 1, \ldots, n_{\text{MC}}\}$.

The following theorem summarizes the results of this section. The power of the POIS2 tests in the theorem maps out the AE two-sided power envelope for invariant similar tests as (β^*, λ^*) is varied.

THEOREM 3: *The* POIS2 *test that rejects* H_0 *when* $\text{LR}^*(Q_1, Q_T; \beta^*, \lambda^*) > \kappa_\alpha(Q_T; \beta^*, \lambda^*)$ *maximizes average power against the alternatives* (β^*, λ^*) *and* (β_2^*, λ_2^*), *where* (β_2^*, λ_2^*) *satisfies* (4.1), *over all level* α *invariant similar tests.*

Approach (ii) to the construction of a two-sided power envelope uses the additional invariance condition to that in (3.1) given by

$$[S : T] \to [-S : T]. \tag{4.11}$$

The corresponding transformation in the parameter space is $(\beta^*, \lambda^*) \to (\beta_2^*, \lambda_2^*)$, where (β_2^*, λ_2^*) satisfies (4.1). This parameter transformation preserves the null hypothesis and the two-sided alternative (but not a one-sided alternative). The sign-invariance condition in (4.11) 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 above. The maximal invariant under this sign invariance condition (plus the invariance conditions in (3.1)) is $(S'S, |S'T|, T'T) = (Q_S, |Q_{ST}|, Q_T)$. The CLR, LM, and AR test statistics all depend on the data only through this maximal invariant and, hence, satisfy the sign-invariance condition. AMS06b shows that the power envelope for the class of invariant similar tests under the invariance conditions of (3.1) and (4.11) equals the AE two-sided power envelope.

Approach (iii) to the construction of a two-sided power envelope for invariant similar tests is based on a necessary condition for unbiasedness. AMS06b shows that an invariant test $\phi(Q)$ is unbiased with size α only if

$$(4.12)\qquad E_{\beta_0}(\phi(Q)|Q_T=q_T)=\alpha \quad \text{and} \quad E_{\beta_0}\big(\phi(Q)Q_{ST}|Q_T=q_T\big)=0$$

for almost all q_T. A test that satisfies (4.12) is said to be locally unbiased (LU) (although we recognize that the conditions in (4.12) are only first-order conditions, not sufficient conditions, for a test's power function to have a local minimum at the null hypothesis). The first condition in (4.12) implies that all unbiased invariant tests are similar. The second condition is the requirement that the power function of an unbiased invariant test has zero derivative under H_0. AMS06b also shows that any similar level α test that depends on the observations through $(Q_S, |Q_{ST}|, Q_T)$ satisfies the LU conditions in (4.12). In consequence, the CLR, LM, and AR tests are LU and the class of LU invariant similar tests is larger than the class of sign-invariant similar tests and the class of unbiased invariant tests.

The test that maximizes power against (β, λ) among LU invariant tests with significance level α rejects H_0 if

$$(4.13)\qquad \mathrm{LR}(Q_1, Q_T; \beta, \lambda) = \frac{\psi(q_1, q_T; \beta, \lambda)}{\psi_2(q_T; \beta, \lambda)} > \kappa_{1\alpha}(Q_T; \beta, \lambda) + Q_{ST}\kappa_{2\alpha}(Q_T; \beta, \lambda),$$

where $\kappa_{1\alpha}(Q_T; \beta, \lambda)$ and $\kappa_{2\alpha}(Q_T; \beta, \lambda)$ are chosen such that the two conditions in (4.12) hold (cf. Lehmann (1986, Theorem 3.5)). The power of the tests in (4.13) for different (β, λ) maps out the power envelope for LU invariant tests. This power envelope is found numerically to be essentially the same as the AE two-sided power envelope; see AMS06b.

5. NUMERICAL RESULTS

This section reports numerical results for the AE two-sided power envelope developed in Section 4 and the CLR, LM, and AR tests for the case of known Ω and normal errors. The model considered is given in (2.4) or (2.5) with Ω specified by $\omega_{11} = \omega_{22} = 1$ and $\omega_{12} = \rho$.[7] Without loss of generality, no X matrix is included. The parameters that characterize the distribution of the tests are λ $(= \pi'Z'Z\pi)$, the number of IVs k, the correlation between the reduced form

[7]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 positive definite $\widetilde{\Omega}$ 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 $\pi = \widetilde{\omega}_{22}^{-1/2}\widetilde{\pi}$.

errors ρ, and the parameter β. Throughout, we focus on tests with significance level 5% and on the case where the null value is $\beta_0 = 0$.[8] 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 AMS06b.

Conditional critical values for the CLR test were computed by numerical integration based on the distributional results in Lemma 3. All results reported here are based on 5,000 Monte Carlo simulations. Details of the numerical methods are given in AMS06b.

The results are presented as plots of power envelopes and power functions against various alternative values of β and λ. (For the AE two-sided power envelope, $(\beta, \lambda) = (\beta^*, \lambda^*)$.) 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 $1/n^{1/2}$, because λ measures the effective sample size.

Figure 1 plots the power functions of the CLR, LM, and AR tests, along with the AE two-sided power envelope. The striking finding is that the power function of the CLR test effectively achieves the power envelope for AE invariant similar tests. Figure 1 documents other results as well. The power function of the AR test is generally below the AE two-sided power envelope, except at its point of tangency at $\beta = \beta_{AR}$. Also, as is known from previous simulation work (e.g., Moreira (2001) and Stock, Wright, and Yogo (2002)), the power function of the LM statistic is not monotonic. This is due to the switch of the sign of d_β as β moves through the value β_{AR}.

In sum, the results of Figure 1 (and further results documented in AMS06b) show that the CLR test dominates the LM and AR tests and, in a numerical sense, attains the two-sided power envelope of Section 4.

Figure 2 shows how the power results change with k. Figure 2 gives the power envelope of Theorem 3 and the power functions of the CLR, LM, and AR tests for $k = 2$ (Figures 2(a) and 2(b)) and for $k = 10$ (Figures 2(c) and 2(d)).

Two findings of these results (and related results reported in AMS06b) are noteworthy. First, 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.

Second, note that the scale is the same in Figure 2 as in Figure 1 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

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

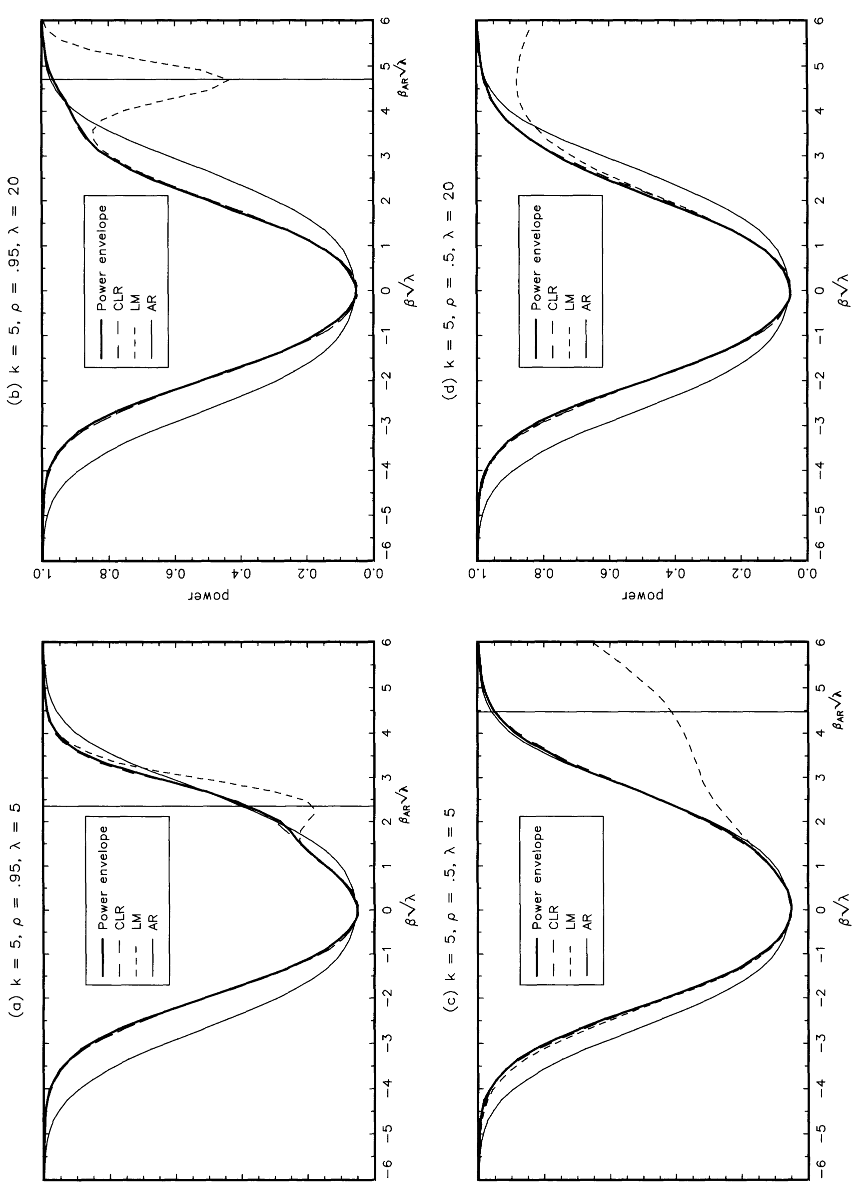

FIGURE 1.—Asymptotically efficient two-sided power envelopes for invariant similar tests and power functions for the two-sided CLR, LM, and AR tests, $k = 5$, $p = 0.95$ and 0.5.

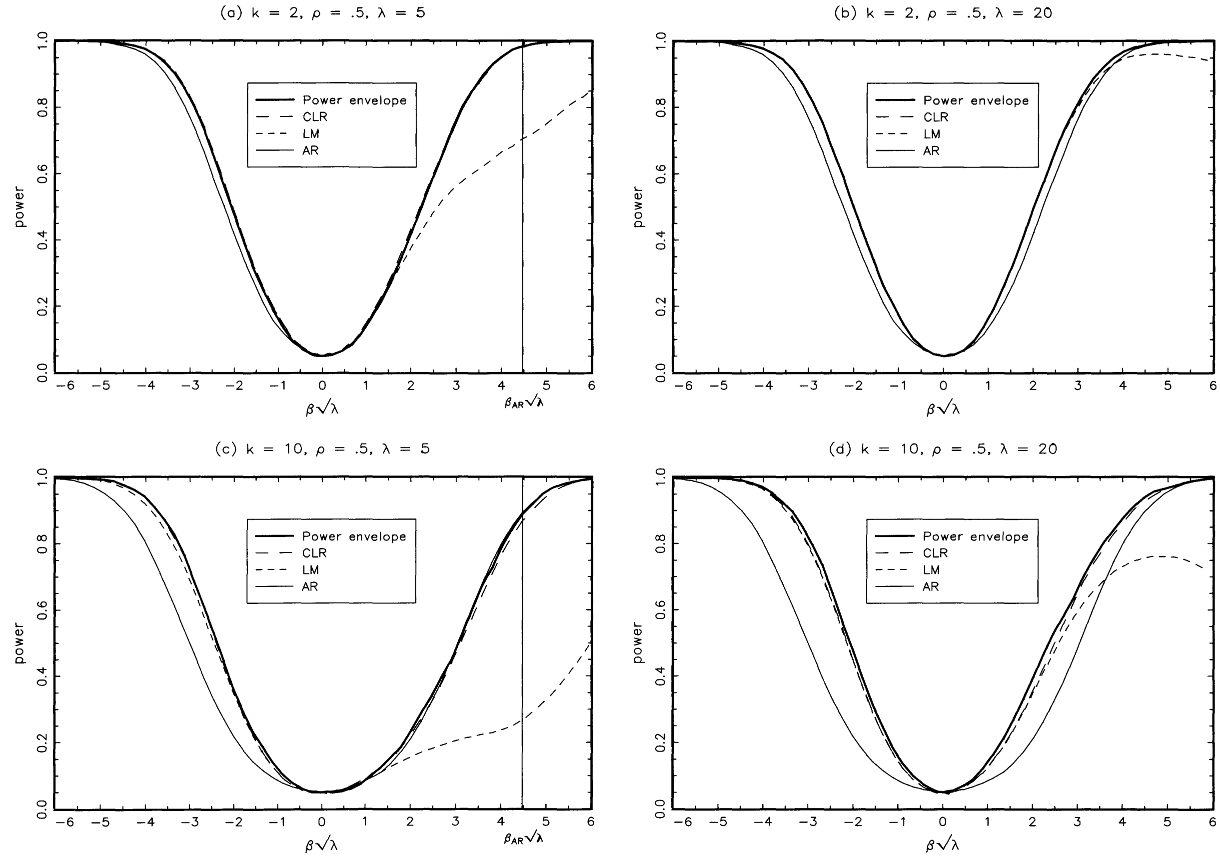

FIGURE 2.—Asymptotically efficient two-sided power envelopes for invariant similar tests and power functions for the two-sided CLR, LM, and AR tests, $k=2$ and 10, $p=0.5$.

on k. In particular, this implies that the two-sided power envelope does not deteriorate significantly with the addition of an irrelevant instrument.

6. WEAK-IV ASYMPTOTICS

In this section, we consider the same model and hypotheses as in Section 2, but with unknown error covariance matrix Ω, (possibly) nonnormal errors, and (possibly) random IVs and/or exogenous variables. We introduce analogues of the CLR, LM, AR, and POIS2 tests that utilize an estimator of Ω. We use weak-IV asymptotics, by Staiger and Stock (1997), to analyze the properties of the tests and to derive a weak-IV asymptotic power envelope that is analogous to the finite-sample AE two-sided power envelope of Section 4.

For clarity of the asymptotics results, throughout this section we write S, T, Q_1, etc. of Sections 2–5, as S_n, T_n, $Q_{1,n}$, etc., respectively, where n is the sample size. All limits are taken as $n \to \infty$. Let $\overline{Z} = [\widetilde{Z} : X]$. Let Y_i, $\widetilde{Z}_i$, X_ι, $\overline{Z}_i$, Z_i, and V_i denote the ith rows of $Y, \widetilde{Z}, X, \overline{Z}, Z$, and V, respectively, written as column vectors of dimensions $2, k, p, k+p, k$, and 2.

6.1. *Assumptions*

We use the same high level assumptions as Staiger and Stock (1997). The parameter π, which determines the strength of the IVs, is local to zero and the alternative parameter β is fixed, not local to the null value β_0. We refer to this as weak IV fixed alternative (WIV-FA) asymptotics. Let p.d. abbreviate "positive definite."

ASSUMPTION WIV-FA: (a) *For some nonstochastic k-vector C, $\pi = C/n^{1/2}$.*
(b) *For all $n \geq 1$, β is a fixed constant.*
(c) *The parameter k is a fixed positive integer that does not depend on n.*

ASSUMPTION 1: *For some p.d. $(k+p) \times (k+p)$ matrix D, $n^{-1}\overline{Z}'\overline{Z} \to_p D$.*

ASSUMPTION 2: *For some p.d. 2×2 matrix Ω, $n^{-1}V'V \to_p \Omega$.*

ASSUMPTION 3: *For some p.d. $2(k+p) \times 2(k+p)$ matrix Φ, $n^{-1/2}\text{vec}(\overline{Z}' \times V) \to_d N(0, \Phi)$, where $\text{vec}(\cdot)$ denotes the column by column vectorization operator.*

ASSUMPTION 4: *There exists $\Phi = \Omega \otimes D$, where Φ is defined in Assumption* 3.

The quantities C, D, and Ω are assumed to be unknown. Primitive sufficient conditions for Assumptions 1–3 are given in AMS04 for i.i.d., independent and non-identically distributed (i.n.i.d.), and stationary sequences with $\{V_i : i \geq 1\}$ being a martingale difference. Given Assumptions 1–3, a sufficient condition for Assumption 4 is homoskedasticity of the errors V_i: $E(V_i V_i' | \overline{Z}_i) = EV_iV_i' = \Omega$ almost surely for all $i \geq 1$.

6.2. *Tests for Unknown Ω and Possibly Nonnormal Errors*

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

$$\widehat{\Omega}_n = (n-k-p)^{-1}\widehat{V}'\widehat{V}, \quad \text{where} \quad \widehat{V} = Y - P_Z Y - P_X Y, \tag{6.1}$$

where k and p are the dimensions of Z_i and X_i, respectively. Let $\widehat{V}_i$ denote the ith row of $\widehat{V}$ written as a column 2-vector. Under Assumptions 1–3, the variance estimator is consistent: $\widehat{\Omega}_n \to_p \Omega$, see Lemma S.1 of AMS06b. The convergence holds uniformly over all true parameters β, C, γ, and ξ no matter what the parameter space is.

We now introduce tests that are suitable for (possibly) nonnormal, homoskedastic, uncorrelated errors and unknown covariance matrix. See AMS04 for tests and results for the case when the errors are not homoskedastic or are correlated.

We define analogues of S_n, T_n, $Q_{1,n}$, and $Q_{T,n}$ with Ω replaced by $\widehat{\Omega}_n$:

$$\begin{aligned} \widehat{S}_n &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2}, \\ \widehat{T}_n &= (Z'Z)^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0 \cdot (a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2}, \\ \widehat{Q}_{1,n} &= (\widehat{Q}_{S,n}, \widehat{Q}_{ST,n})' = (\widehat{S}_n'\widehat{S}_n, \widehat{S}_n'\widehat{T}_n)', \quad \text{and} \quad \widehat{Q}_{T,n} = \widehat{T}_n'\widehat{T}_n. \end{aligned} \tag{6.2}$$

The LR, LM, AR, and POIS2 test statistics for the case of unknown Ω are defined as in (3.4) and Corollary 1, but with Q_S, Q_{ST}, and Q_T replaced by $\widehat{Q}_{S,n}$, $\widehat{Q}_{ST,n}$, and $\widehat{Q}_{T,n}$, respectively. Denote these test statistics by $\widehat{\text{LR}}_n$, $\widehat{\text{LM}}_n$, $\widehat{\text{AR}}_n$, and $\text{LR}^*(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*)$, respectively.

6.3. *Weak-IV Asymptotic Distributions of Test Statistics*

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

$$\begin{aligned} &\text{vec}(N_Z) \sim N\big(\text{vec}(D_Z C a'), \Omega \otimes D_Z\big), \\ &S_\infty = D_Z^{-1/2} N_Z b_0 \cdot (b_0'\Omega b_0)^{-1/2} \sim N(c_\beta D_Z^{1/2} C, I_k), \\ &T_\infty = D_Z^{-1/2} N_Z \Omega^{-1} a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \sim N(d_\beta D_Z^{1/2} C, I_k), \quad \text{where} \\ &D_Z = D_{11} - D_{12}D_{22}^{-1}D_{21}, \\ &D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \quad D_{11} \in \mathbb{R}^{k\times k}, \quad D_{12} \in \mathbb{R}^{k\times p}, \quad \text{and} \quad D_{22} \in \mathbb{R}^{p\times p}. \end{aligned} \tag{6.3}$$

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

$$Q_\infty = [S_\infty : T_\infty]'[S_\infty : T_\infty], \tag{6.4}$$

$$Q_{1,\infty} = (S'_\infty S_\infty, S'_\infty T_\infty)', \quad Q_{T,\infty} = T'_\infty T_\infty, \quad Q_{ST,\infty} = S'_\infty T_\infty,$$
$$Q_{S,\infty} = S'_\infty S_\infty, \quad \mathcal{S}_{2,\infty} = S'_\infty T_\infty / (\|S_\infty\| \cdot \|T_\infty\|), \quad \text{and}$$
$$\lambda_\infty = C' D_Z C.$$

By (6.3) and the proof of Lemma 3, we find that 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_∞.

The following results hold under H_0 and fixed (i.e., nonlocal) alternatives.

LEMMA 4: *Under Assumptions* WIV-FA *and* 1–4:
(a) $(S_n, T_n) \rightarrow_d (S_\infty, T_\infty)$;
(b) $(\widehat{S}_n, \widehat{T}_n) - (S_n, T_n) \rightarrow_p 0$;
(c) $(\widehat{S}_n, \widehat{T}_n) \rightarrow_d (S_\infty, T_\infty)$.

COMMENTS: (i) 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 IVs are arbitrarily weak. In consequence, we expect the asymptotic test procedures developed here to perform well in terms of size even for very weak IVs.

(ii) Lemma 4 and the continuous mapping theorem imply that the asymptotic distributions of the $\widehat{\mathrm{LR}}_n$, $\widehat{\mathrm{LM}}_n$, and $\widehat{\mathrm{AR}}_n$ test statistics are given by the distributions of the test statistics in (3.4) with (Q_S, Q_{ST}, Q_T) replaced by $(Q_{S,\infty}, Q_{ST,\infty}, Q_{T,\infty})$. Under H_0, $\widehat{\mathrm{LM}}_n$ and $\widehat{\mathrm{AR}}_n$ have asymptotic χ^2_1 and χ^2_k/k distributions, respectively.

Using Lemma 4, we establish the asymptotic distributions of the $\{\mathrm{LR}^*(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*) : n \geq 1\}$ test statistics and $\{\kappa_\alpha(\widehat{Q}_{T,n}; \beta^*, \lambda^*) : n \geq 1\}$ critical values.

THEOREM 4: *Under Assumptions* WIV-FA *and* 1–4:
(a) $(\mathrm{LR}^*(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*), \ \kappa_\alpha(\widehat{Q}_{T,n}; \beta^*, \lambda^*)) \rightarrow_d (\mathrm{LR}^*(Q_{1,\infty}, Q_{T,\infty}; \beta^*, \lambda^*), \kappa_\alpha(Q_{T,\infty}; \beta^*, \lambda^*))$;
(b) $P(\mathrm{LR}^*(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*) > \kappa_\alpha(\widehat{Q}_{T,n}; \beta^*, \lambda^*)) \rightarrow P(\mathrm{LR}^*(Q_{1,\infty}, Q_{T,\infty}; \beta^*, \lambda^*) > \kappa_\alpha(Q_{T,\infty}; \beta^*, \lambda^*))$;
(c) *under* H_0, $P(\mathrm{LR}^*(Q_{1,\infty}, Q_{T,\infty}; \beta^*, \lambda^*) > \kappa_\alpha(Q_{T,\infty}; \beta^*, \lambda^*)) = \alpha$.

COMMENT: Theorem 4(b) is used below to obtain the weak-IV asymptotic power envelope for the case of an estimated error covariance matrix.

6.4. *Weak-IV Asymptotic Power Envelope*

In this subsection, we show that the POIS2 test based on $\mathrm{LR}(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*)$ exhibits an asymptotic average-power optimality property when the IVs are weak and the errors are i.i.d. normal with *unknown* covariance matrix. These results yield the AE two-sided asymptotic power envelope. It is the same as the finite-sample power envelope of Section 4 when Ω is known.

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, Chap. 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

$$(6.5) \qquad \Omega = \Omega_0 + \Omega_1/n^{1/2} \quad \text{and} \quad \eta = \eta_0 + \eta_1/n^{1/2},$$

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 positive definite.

The least squares estimator of η in the model of (2.5) is $\widehat{\eta}_n = (X'X)^{-1}X'Y$.

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 nonredundant elements of A.

The following basic results hold under H_0 and fixed alternatives $\beta \neq \beta_0$:

LEMMA 5: *Suppose Assumption* WIV-FA *holds, the reduced-form errors* $\{V_i : i \geq 1\}$ *are i.i.d. normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and p.d. variance matrix* Ω, *and* Ω *and* η *are as in* (6.5). *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) $(n^{-1/2}Z'Y,\ n^{1/2}(\widehat{\eta}_n - \eta_0),\ n^{1/2}(\widehat{\Omega}_n - \Omega_0)) \rightarrow_d (N_Z,\ N_X,\ N_\Omega)$, *where* N_Z, N_X, *and* N_Ω *are independent* $k \times 2$, $p \times 2$, *and* 2×2 *normal random matrices, respectively, with* $\text{vec}(N_Z) \sim N(\text{vec}(D_Z C a'), \Omega_0 \otimes D_Z)$, $\text{vec}(N_X) \sim N(\text{vec}(\eta_1), \Omega_0 \otimes D_{22}^{-1})$, N_Ω *is symmetric, and* $\text{vech}(N_\Omega) \sim N(\Omega_1, E(\zeta - E\zeta) \times (\zeta - E\zeta)')$, *where* $\zeta = \text{vech}(v_0 v_0')$, $v_0 \in \mathbb{R}^2$, *and* $v_0 \sim N(0, \Omega_0)$, *provided Assumption* 1 *also holds*.

Given the result of part (a) of Lemma 5, 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 $\phi(\cdot, \cdot, \cdot)$,

$$(6.6) \qquad \phi_n\big(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0)\big) \rightarrow_d \phi(N_Z, N_X, N_\Omega),$$
$$P_{\beta, C, \Omega_0, \eta_0}\big(\phi(N_Z, N_X, N_\Omega) = 1\big) = \alpha$$

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 CLR, LM, AR, and POIS2 tests. Standard Wald and LR tests are not asymptotically similar.

The transformation, call it $h_\Omega(\cdot)$, from N_Z to $[S_\infty : T_\infty]$ in (6.3) is one-to-one. Hence, for some function $\overline{\phi}$, we have

$$\phi(N_Z, N_X, N_\Omega) = \phi\big(h_\Omega^{-1}(S_\infty, T_\infty), N_X, N_\Omega\big) = \overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega). \tag{6.7}$$

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 (6.4)).

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 (6.6) holds and the distribution of $\overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega)$ depends on (S_∞, T_∞) only through Q_∞, i.e.,

$$\overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega) \sim \phi^*(Q_\infty, N_X, N_\Omega) \tag{6.8}$$

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 CLR, LM, AR, and POIS2 tests.

We now establish an upper bound on two-point average asymptotic power.

THEOREM 5: *Suppose Assumptions* WIV-FA *and* 1 *hold, the reduced-form errors* $\{V_i : i \geq 1\}$ *are i.i.d. normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and p.d. variance matrix* Ω, Ω *and* η *are as in* (6.5), *and* (β^*, λ^*) *and* (β_2^*, λ_2^*) *satisfy* (4.1). *For any convergent sequence of asymptotically invariant and asymptotically similar tests* $\{\phi_n : n \geq 1\}$, *we have*

$$\begin{aligned}
&\lim_{n\to\infty} P^*_{\beta^*,C,\Omega,\eta}\big(\phi_n\big(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0)\big) = 1\big) \\
&\quad = P^*_{\beta^*,C,\Omega_0,\eta_0}\big(\phi^*(Q_\infty, N_X, N_\Omega) = 1\big) \\
&\quad \leq P^*_{\beta^*,C,\Omega_0,\eta_0}\big(\mathrm{LR}^*(Q_{1,\infty}, Q_{T,\infty}; \beta^*, \lambda^*) > \kappa_\alpha(Q_{T,\infty}; \beta^*, \lambda^*)\big),
\end{aligned}$$

where $P^*_{\beta^*,C,\Omega,\eta}(\cdot) = (1/2)[P_{\beta^*,C,\Omega,\eta}(\cdot) + P_{\beta_2^*,C_2,\Omega,\eta}(\cdot)]$, $P_{\beta,C,\Omega,\eta}(\cdot)$ *denotes probability when the true parameters are* (β, C, Ω, η), C *satisfies* $C'D_Z C = \lambda^*$, *and* C_2 *satisfies* $C_2' D_Z C_2 = \lambda_2^*$.

Combining Theorem 5 with Theorem 4(b) shows that POIS2 tests attain the asymptotic upper bound on average power and, hence, their power maps out the asymptotic average-power envelope as (β^*, λ^*) vary.

COROLLARY 2: *Under the conditions of Theorem* 5, *the POIS*2 *tests of Section* 6 *are convergent sequences of asymptotically invariant and asymptotically similar tests that attain the upper bound on asymptotic average power given in Theorem* 5.

COMMENTS: (i) The asymptotic power envelope depends only on (β^*, λ^*). It is the same as the finite-sample power envelope for known Ω of Section 4.

(ii) In Theorem 5 and Corollary 2, the assumption that the reduced-form errors $\{V_i : i \geq 1\}$ are i.i.d. normal, independent of $\{\overline{Z}_i : i \geq 1\}$, with mean zero and p.d. variance matrix Ω, can be replaced by Assumptions 2–4. Thus, the asymptotic power envelope and its near attainability by the CLR test still hold with nonnormal errors. However, with this replacement, Lemma 5(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 $(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0))$.

(iii) Theorem 4(b) holds under (6.5) by the same argument as when Ω and η are constants.

7. STRONG-IV ASYMPTOTICS

In this section, we analyze the strong-IV–local alternative asymptotic properties of the tests considered above for the case of unknown covariance matrix and nonnormal errors. The results provided here are essential for the specification above of the AE two-sided power envelope. For strong IV–fixed alternative results, i.e., consistency results, see AMS04. As in Section 6, we denote $S = S_n$, $Q = Q_n$, etc.

We make the following assumption:

ASSUMPTION SIV-LA: (a) *For some constant* $B \in \mathbb{R}$, $\beta = \beta_0 + B/n^{1/2}$.

(b) *For all* $n \geq 1$, π *is a fixed nonzero* k*-vector.*

(c) *The parameter* 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*

$$\begin{aligned} (7.1) \qquad & S_{B\infty} \sim N(\alpha_S, I_k), \\ & \alpha_S = D_Z^{1/2}\pi B(b_0'\Omega b_0)^{-1/2}, \\ & \alpha_T = D_Z^{1/2}\pi(a_0'\Omega^{-1}a_0)^{1/2}. \end{aligned}$$

Using these definitions, we obtain the following results.

LEMMA 6: *Under Assumptions* SIV-LA *and* 1–4, (a) $(S_n, T_n/n^{1/2}) \to_d (S_{B\infty}, \alpha_T)$, (b) $(\widehat{S}_n, \widehat{T}_n/n^{1/2}) = (S_n, T_n/n^{1/2}) + o_p(1)$, *and* (c) $(\widehat{Q}_{S,n}, \widehat{Q}_{ST,n}/n^{1/2}, \widehat{Q}_{T,n}/n) \to_d (S_{B\infty}'S_{B\infty}, \alpha_T'S_{B\infty}, \alpha_T'\alpha_T)$ *as* $n \to \infty$.

Using Lemma 6, we determine the asymptotic distributions of the AR, LM, and LR test statistics under SIV-LA asymptotics.

THEOREM 6: *Under Assumptions* SIV-LA *and* 1–4, (a) $\widehat{\text{AR}}_n = \text{AR}_n + o_p(1) \to_d S'_{B\infty}S_{B\infty}/k \sim \chi^2_k(\alpha'_S\alpha_S)/k$, (b) $\widehat{\text{LM}}_n = \text{LM}_n + o_p(1) \to_d (\alpha'_T S_{B\infty})^2/\|\alpha_T\|^2 \sim \chi^2_1((\alpha'_T\alpha_S)^2/\|\alpha_T\|^2)$, *and* (c) $\widehat{\text{LR}}_n = \text{LR}_n + o_p(1) = \text{LM}_n + o_p(1) \to_d \alpha'_T S_{B\infty}/\|\alpha_T\|$.

COMMENTS: (i) Part (c) of Theorem 6 shows that the LR and LM test statistics are asymptotically equivalent under SIV-LA asymptotics for any value of k (the number of IVs). (When $k = 1$, the LR, LM, and AR test statistics are the same, so the tests are trivially asymptotically equivalent.)

(ii) The critical values for the LM and AR tests are nonrandom. However, the critical value for the CLR test is a function of $Q_{T,n}$ or $\widehat{Q}_{T,n}$. Hence, for the CLR and LM tests to be asymptotically equivalent, the CLR critical value, call it $\kappa_{\text{LR},\alpha}(\widehat{Q}_{T,n})$, must converge in probability to a constant as $n \to \infty$. Under strong-IV asymptotics, $\widehat{Q}_{T,n} \to_p \infty$. In consequence, asymptotic equivalence holds if $\kappa_{\text{LR},\alpha}(q_T)$ converges to a finite constant as q_T diverges to infinity. Moreira (2003) shows that $\lim_{q_T\to\infty} \kappa_{\text{LR},\alpha}(q_T)$ equals the $1-\alpha$ quantile of the χ^2_1 distribution. Hence, the CLR and LM tests are indeed asymptotically equivalent under SIV-LA asymptotics.

(iii) Theorem 6(a) and (b) are not new results, but part (c) is new. Moreira (2003) does not provide the SIV-LA asymptotic distribution of $\widehat{\text{LR}}_n$.

Under SIV-LA asymptotics and i.i.d. 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, LR, and LM 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* tests under SIV-LA asymptotics and i.i.d. normal errors.

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

THEOREM 7: *Suppose Assumptions* SIV-LA *and* 1 *hold, and the reduced-form errors* $\{V_i : i \geq 1\}$ *are i.i.d. normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and p.d. variance matrix* Ω *that may be known or unknown. Then the CLR test based on* $\widehat{\text{LR}}_n$ *and the LM test based on* $\widehat{\text{LM}}_n$ *are asymptotically efficient under strong-IV asymptotics.*

COMMENT: The AR test based on $\widehat{\text{AR}}_n$ is not AE under SIV-LA asymptotics and i.i.d. normal errors unless $k = 1$. This holds because its asymptotic distribution under SIV-LA asymptotics differs from that of $\widehat{\text{LM}}_n$ when $k > 1$ by Theorem 6.

Next, we provide results for POIS2 tests. We allow for the case where the second point (β_2^*, λ_2^*) satisfies (4.1) and for the case where it does not. The form of a POIS2 test is that given in Corollary 1 whether or not the second point (β_2^*, λ_2^*) satisfies (4.1). Our results show that, under i.i.d. normal errors, a POIS2 test is asymptotically efficient under SIV-LA asymptotics if and only if (β_2^*, λ_2^*) satisfies (4.1).

THEOREM 8: *Under Assumptions* SIV-LA *and* 1–4, (a) $\mathrm{LR}^*(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*) = \mathrm{LR}^*(Q_{1,n}, Q_{T,n}; \beta^*, \lambda^*) + o_p(1)$, (b) *if* (β_2^*, λ_2^*) *satisfies* (4.1), *then* $\mathrm{LR}^*(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*) = \exp((-(\tau^*)^2)/2)\cosh(\tau^*\mathrm{LM}_n^{1/2}) + o_p(1)$, *where* $\tau^* = (\lambda^*)^{1/2}c_{\beta^*}$, *which is a strictly increasing continuous function of* LM_n, *and* (c) *if* (β_2^*, λ_2^*) *does not satisfy* (4.1), *then* $LR^*(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}; \beta^*, \lambda^*) = \eta_2(Q_{ST,n}/Q_{T,n}^{1/2}) + o_p(1)$ *for a continuous function* $\eta_2(\cdot)$ *that is not even.*

COMMENTS: (i) The critical values for the POIS2 tests converge in probability to constants as $n \to \infty$ under strong-IV asymptotics. (See the Appendix for a proof.) Hence, Theorems 7 and 8(b) and (c) imply that a POIS2 test is AE under SIV-LA asymptotics and i.i.d. normal reduced-form errors if and only if (β_2^*, λ_2^*) satisfies (4.1).

(ii) Theorem 8(a) shows that, under SIV-LA asymptotics and the homoskedastic errors assumptions (which do not require normality), a POIS2 test with estimated error variance matrix Ω is asymptotically equivalent to the corresponding POIS2 test with known Ω. Under the same assumptions, Theorem 8(b) shows that a POIS2 test is asymptotically equivalent to the two-sided LM test with known Ω when (4.1) holds. Under the same assumptions, Theorem 8(c) shows that a POIS2 test 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 (4.1) fails to hold.

(iii) The proof of Theorem 8(c) shows that if the second condition of (4.1) fails to hold, then $\eta_2(\cdot)$ is a monotone function and, hence, the POIS2 test is asymptotically equivalent to one of the one-sided LM tests based on $\pm Q_{ST,n}/Q_{T,n}^{1/2}$. The proof shows that if the second condition of (4.1) holds and the first condition fails, then the POIS2 test 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. Thus, if either condition of (4.1) fails, the POIS2 test is not asymptotically equivalent to the two-sided LM test and, hence, is not asymptotically efficient.

Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Ave., Box 208281, New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu,

Dept. of Economics, Harvard University, Littauer Center M-6, 1875 Cambridge St., Cambridge, MA 02138, U.S.A.; moreira@fas.harvard.edu, *and*

Dept. of Economics, Harvard University, Littauer Center M-6, 1875 Cambridge St., Cambridge, MA 02138, U.S.A.; James_Stock@harvard.edu.

Manuscript received July, 2004; final revision received January, 2006.

APPENDIX A: PROOFS

A.1. *Proofs of Results Stated in Sections* 2–4

PROOF OF LEMMA 1: The proof is standard using normality of Y and zero covariance between $Z'Y$ and $X'Y$; see AMS06b for details. *Q.E.D.*

PROOF OF LEMMA 2: The proof is straightforward; see AMS06b for details. *Q.E.D.*

PROOF OF THEOREM 1: Let $M(S,T) = [S:T]'[S:T] = Q$. The $M(S,T)$ 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, Eq. (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 $\mu_3, \ldots, \mu_k$ such that $\{\mu_1, \ldots, \mu_k\}$ span $\mathbb{R}^k$. Applying the Gram–Schmidt procedure to $\{\mu_1, \ldots, \mu_k\}$, 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 $f_1 = \mu_1/\|\mu_1\|$, $f_2 = M_{\mu_1}\mu_2/\|M_{\mu_1}\mu_2\|, \ldots,$ $f_k = M_{[\mu_1:\cdots:\mu_{k-1}]}\mu_k/\|M_{[\mu_1:\cdots:\mu_{k-1}]}\mu_k\|$. Define $F = [f_1:\cdots:f_k]'$. We have

$$\text{(A.1)} \qquad F\mu_1 = (f_1'\mu_1, \ldots, f_k'\mu_1)' = (\|\mu_1\|, 0, \ldots, 0)',$$
$$F\mu_2 = \big(\mu_1'\mu_2/\|\mu_1\|, \mu_2'M_{\mu_1}\mu_2/\|M_{\mu_1}\mu_2\|, 0, \ldots, 0\big)'.$$

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 $\{\widetilde{\mu}_1, \ldots, \widetilde{\mu}_k\}$ in place of $\{\mu_1, \ldots, \mu_k\}$. 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 $F\mu_1 = (\|\mu_1\|, 0, \ldots, 0)' = (\|\widetilde{\mu}_1\|, 0, \ldots, 0)' = \widetilde{F}\widetilde{\mu}_1$ 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 $\{\mu_2^*, \dots, \mu_k^*\}$ for which $\{\mu_1, \mu_2^*, \dots, \mu_k^*\}$ span $\mathbb{R}^k$. 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. *Q.E.D.*

PROOF OF THEOREM 2: 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 Theorem 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 an added step 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 1. In particular, for an 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). *Q.E.D.*

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 1, and all correlations equal to 0. Hence, $Q = [S:T]'[S:T]$ has a noncentral Wishart distribution with mean matrix of rank 1 and identity covariance matrix. By (6) of Anderson (1946), the density of Q at q is

$$\text{(A.2)} \qquad K_1 \exp\left(-\frac{\operatorname{tr}(M'M)}{2}\right)|q|^{(k-3)/2} \exp\left(-\frac{\operatorname{tr}(q)}{2}\right) \times (\operatorname{tr}(M'Mq))^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{\operatorname{tr}(M'Mq)}\right).$$

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

Part (b) holds because Q_T has a noncentral chi-squared distribution with noncentrality parameter $d_\beta^2\lambda$ by Lemma 2(b) and (3.3). The stated form of the density is given by Anderson (1946, Eq. (6)). Part (c) holds by taking the ratio of the densities given in parts (a) and (b) 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 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 $\alpha \in \mathbb{R}^k$ 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 $\alpha = (1, 0, \dots, 0)'$, $(k-1)^{1/2}A/(1-A^2)^{1/2} =$

$(k-1)^{1/2}S_1/(\sum_{j=2}^k S_j^2)^{1/2} \sim t_{k-1}$ by definition of the t_{k-1} distribution; (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, proof of Theorem 1.5.7(i), pp. 38–39; Eq. (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 $\alpha \in \mathbb{R}^k$ with $\alpha'\alpha = 1$ does not depend on α by spherical symmetry of S. Thus, 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. *Q.E.D.*

A.2. *Proofs of Results Stated in Section* 6

PROOF OF LEMMA 4: To establish part (a), we have

$$n^{-1}Z'Z = n^{-1}\widetilde{Z}'\widetilde{Z} - n^{-1}\widetilde{Z}'P_X\widetilde{Z} \to_p D_{11} - D_{12}D_{22}^{-1}D_{21} = D_Z \tag{A.3}$$

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

$$\begin{aligned} & n^{-1/2}Z'Vb_0 \\ &\quad = n^{-1/2}(\widetilde{Z} - P_X\widetilde{Z})'Vb_0 = n^{-1/2}(\widetilde{Z} - XD_{22}^{-1}D_{21})'Vb_0 + o_p(1) \\ &\quad = [I_k : -D_{12}D_{22}^{-1}]n^{-1/2}\overline{Z}'Vb_0 + o_p(1) \to_d [I_k : -D_{12}D_{22}^{-1}]N^*b_0 \\ &\quad = [I_k : -D_{12}D_{22}^{-1}](b_0' \otimes I_{k+p})\text{vec}(N^*). \end{aligned} \tag{A.4}$$

Hence, we have

$$\begin{aligned} S_n &= (n^{-1}Z'Z)^{-1/2}(n^{-1/2}Z'Vb_0 + n^{-1}Z'ZCa'b_0) \\ &\quad \times (b_0'\Omega b_0)^{-1/2} \to_d H, \quad \text{where} \\ H &= D_Z^{-1/2}\big([I_k : -D_{12}D_{22}^{-1}](b_0' \otimes I_{k+p})\text{vec}(N^*) + D_Z Ca'b_0\big) \\ &\quad \times (b_0'\Omega b_0)^{-1/2} \end{aligned} \tag{A.5}$$

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

$$\begin{aligned} EH &= D_Z^{1/2}Ca'b_0 \cdot (b_0'\Omega b_0)^{-1/2} = c_\beta D_Z^{1/2}C, \\ \text{var}(H) &= D_Z^{-1/2}[I_k : -D_{12}D_{22}^{-1}](b_0' \otimes I_{k+p})(\Omega \otimes D)(b_0 \otimes I_{k+p}) \\ &\qquad \times [I_k : -D_{12}D_{22}^{-1}]'D_Z^{-1/2} \cdot (b_0'\Omega b_0)^{-1} \\ &= D_Z^{-1/2}[I_k : -D_{12}D_{22}^{-1}]D[I_k : -D_{12}D_{22}^{-1}]'D_Z^{-1/2} = I_k, \end{aligned} \tag{A.6}$$

which completes the proof for S_n.

Analogously to (A.4), we have

$$n^{-1/2}Z'V\Omega^{-1}a_0 \to_d [I_k : -D_{12}D_{22}^{-1}]\big((a_0'\Omega^{-1}) \otimes I_{k+p}\big)\operatorname{vec}(N^*). \tag{A.7}$$

Using this, we obtain

$$\begin{aligned} T_n &= (n^{-1}Z'Z)^{-1/2}(n^{-1/2}Z'V\Omega^{-1}a_0 + n^{-1}Z'ZCa'\Omega^{-1}a_0) \\ &\quad \times (a_0'\Omega^{-1}a_0)^{-1/2} \to_d J \quad \text{for} \\ J &= D_Z^{-1/2}\big([I_k : -D_{12}D_{22}^{-1}]\big((a_0'\Omega^{-1}) \otimes I_{k+p}\big)\operatorname{vec}(N^*) + D_Z Ca'\Omega^{-1}a_0\big) \\ &\quad \times (a_0'\Omega^{-1}a_0)^{-1/2}. \end{aligned} \tag{A.8}$$

Analogously to (A.6), J has a normal distribution with $EJ = d_\beta D_Z^{1/2}C$ and $\operatorname{var}(J) = I_k$, 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:

$$\begin{aligned} &E(b_0' \otimes I_{k+p})\operatorname{vec}(N^*)\operatorname{vec}(N^*)'\big((\Omega^{-1}a_0) \otimes I_{k+p}\big) \\ &\quad = E(b_0' \otimes I_{k+p})(\Omega \otimes D)\big((\Omega^{-1}a_0) \otimes I_{k+p}\big) = (b_0'a_0) \otimes D = 0. \end{aligned} \tag{A.9}$$

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) $(Z'Z)^{-1/2} \times Z'Y = O_p(1)$ by the same sort of argument as in (A.3) and (A.4), (ii) $\widehat{\Omega}_n \to_p \Omega$ (see AMS06b), and (iii) Ω is p.d. by Assumption 2.

Part (c) follows immediately from parts (a) and (b). *Q.E.D.*

PROOF OF THEOREM 4: The functions $\psi(\cdot, \cdot; \beta, \lambda)$ and $\psi_2(\cdot; \beta, \lambda)$ are continuous and do not depend on n; see their definitions in Corollary 1. The same is true of the critical value function $\kappa_\alpha(\cdot; \beta, \lambda)$ 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 $\kappa_\alpha(\cdot; \beta, \lambda)$ in (4.10). In consequence, the result of part (a) of the theorem follows from Lemma 4, (6.4), and the continuous mapping theorem.

Part (b) follows immediately from part (a).

Part (c) 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 $\kappa_\alpha(\cdot; \beta, \lambda)$, for all constants $q_{T,\infty}$, $P(\mathrm{LR}^*(Q_{1,\infty}, q_{T,\infty}; \beta, \lambda) > \kappa_\alpha(q_{T,\infty}; \beta, \lambda) | Q_{1,\infty} = q_{T,\infty}) = \alpha$. This result and iterated expectations establishes part (c). *Q.E.D.*

PROOF OF LEMMA 5: Part (a) holds because (i) given that Ω_0 and η_0 are known, and Ω_1 and η_1 are unknown, $(Z'Y, X'Y, Y'Y)$ are seen to be sufficient statistics for $(\beta, C, \Omega_1, \eta_1)$ by inspection of the normal density of Y conditional on $[Z:X]$ 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) $\text{vec}(n^{-1/2}Z'V) \sim N(0, \Omega \otimes (n^{-1}Z'Z))$ conditional on $n^{-1}Z'Z$ and $n^{-1}Z'Z \to_p D_Z$ (by (A.4) using Assumption 1) imply that $\text{vec}(n^{-1/2}Z'V) \to_d N(0, \Omega \otimes D_Z)$; (ii) $\text{vec}(n^{-1/2}Z'Z\pi a') = \text{vec}(n^{-1}Z'ZCa') \to_p D_Z Ca'$ 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 $\text{vec}(n^{1/2}(\widehat{\eta}_n - \eta_0)) \to_d N(\eta_1, \Omega \otimes D_{22}^{-1})$; (iv) $n^{1/2}(\widehat{\Omega}_n - \Omega_0) = n^{1/2}(n^{-1}V'V - \Omega_0) - n^{-1/2}V'P_Z V - n^{-1/2}V'P_X V$; (v) $n^{1/2} \times (n^{-1}V'V - \Omega_0) = n^{-1/2}(V'V - EV'V) + \Omega_1$; (vi) $\text{vech}(n^{-1/2}(V'V - EV'V)) \to_d N(0, E(\zeta - E\zeta)(\zeta - E\zeta)')$ by a triangular array CLT for rowwise i.i.d. 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); (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. *Q.E.D.*

PROOF OF THEOREM 5: 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 because it is independent of N_X and N_Ω, and the latter have distributions that do not depend on β or C; (ii) result (i) implies that the average power 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_∞; (iii) Theorem 3 with Q replaced by Q_∞ implies that the average power of the similar test $\widetilde{\phi}(Q_\infty)$ is less than or equal to the upper bound given in Theorem 5. *Q.E.D.*

A.3. *Proofs of Results Stated in Section* 7

PROOF OF LEMMA 6: To prove part (a), we use (2.6), (6.3), and Assumptions SIV-LA, 1, 3, and 4 to obtain

$$
\begin{aligned}
\text{(A.10)}\quad S_n &= c_\beta \mu_\pi + (Z'Z)^{-1/2}Z'Vb_0 \cdot (b_0'\Omega b_0)^{-1/2} \to_d S_{B\infty}, \\
T_n/n^{1/2} &= d_\beta \mu_\pi / n^{1/2} + (Z'Z/n)^{-1/2}(Z'V/n)\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \\
&= d_\beta (Z'Z/n)^{1/2}\pi + o_p(1) = \alpha_T + o_p(1).
\end{aligned}
$$

Part (b) holds because $\widehat{\Omega}_n \to_p \Omega$; see AMS06b. Part (c) holds by part (a), part (b), and the continuous mapping theorem. *Q.E.D.*

PROOF OF THEOREM 6: Theorem 6(a) and (b) follow immediately from Lemma 6.

The first equality of part (c) follows from Lemma 6. The second equality of part (c) of the theorem is established as follows. By Lemma 6, we have

$$\text{(A.11)}\qquad \frac{Q_T}{(Q_T - Q_S)^2} = \frac{Q_T/n}{(Q_T/n - Q_S/n)^2} n^{-1} = \frac{\alpha_T'\alpha_T + o_p(1)}{(\alpha_T'\alpha_T + o_p(1))^2} n^{-1} = o_p(1).$$

By a mean-value expansion, $\sqrt{1+x} = 1 + (1/2)x(1 + o(1))$ as $x \to 0$. This and some algebra give

$$\begin{aligned}
\text{(A.12)}\qquad \mathrm{LR} &= \frac{1}{2}\Big(Q_S - Q_T + \sqrt{(Q_T - Q_S)^2 + 4Q_{ST}^2}\Big) \\
&= \frac{1}{2}\left(Q_S - Q_T + |Q_T - Q_S|\sqrt{1 + \frac{4Q_T}{(Q_T - Q_S)^2}\mathrm{LM}}\right) \\
&= \frac{1}{2}\left(Q_S - Q_T + |Q_T - Q_S|\left(1 + \frac{2Q_T(1 + o_p(1))}{(Q_T - Q_S)^2}\mathrm{LM}\right)\right) \\
&= \frac{Q_T(1 + o_p(1))}{Q_T - Q_S}\mathrm{LM},
\end{aligned}$$

where the fourth equality uses $|Q_T - Q_S| = Q_T - Q_S$ with probability that goes to 1 by the calculation in the denominator of (A.11). As in (A.11), by Lemma 6, we have $Q_T/(Q_T - Q_S) = 1 + o_p(1)$. This and (A.12) combine to give the second equality of part (c). *Q.E.D.*

PROOF OF THEOREM 7: 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 $\mathrm{LM}_n = Q_{ST}^2/Q_T$. By Theorem 6, the LR statistic is 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 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

$$\text{(A.13)}\qquad -\frac{1}{2}(S - c_\beta\mu_\pi)'(S - c_\beta\mu_\pi) - \frac{1}{2}(T - d_\beta\mu_\pi)'(T - d_\beta\mu_\pi).$$

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

$$\text{(A.14)}\quad \left(\frac{d}{d\beta} c_\beta \mu_\pi' S - \frac{1}{2}\frac{d}{d\beta}(c_\beta^2)\mu_\pi'\mu_\pi + \frac{d}{d\beta} d_\beta \mu_\pi' T - \frac{1}{2}\frac{d}{d\beta}(d_\beta^2)\mu_\pi'\mu_\pi\right)\Bigg|_{(\beta,\pi)=(\beta_0,\widehat{\pi}_0)}$$
$$= \frac{d}{d\beta} c_{\beta_0}\mu_{\widehat{\pi}_0}' S + \frac{d}{d\beta} d_{\beta_0}\mu_{\widehat{\pi}_0}' T - d_{\beta_0}\frac{d}{d\beta} d_{\beta_0}\mu_{\widehat{\pi}_0}'\mu_{\widehat{\pi}_0}$$
$$= \frac{d}{d\beta} c_{\beta_0} \cdot d_{\beta_0}^{-1} T'S,$$

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 $\text{plim}_{n\to\infty} T'T/n = \alpha_T'\alpha_T$. Hence, the standard LM statistic is $(T'S)^2/T'T = LM_n$, which completes the proof. *Q.E.D.*

PROOF OF THEOREM 8: Part (a) of the theorem holds by Lemma 6(a) and the continuity of $\psi(q_1, q_T; \beta, \lambda)$ and $\psi_2(q_T; \beta, \lambda)$ in (q_1, q_T).

To prove Theorem 8(b) and (c), we establish some preliminary results. Let β_1 and λ_1 be any fixed constants for which $d_{\beta_1} \neq 0$ (i.e., $\beta_1 \neq \beta_{\text{AR}}$). Define $h_{\beta_1} = (c_{\beta_1}, d_{\beta_1})'$. Then (i) $Q_T/n \to_p \alpha_T'\alpha_T > 0$ by Lemma 6(a) and Assumption SIV-LA(b); (ii) $Q_{ST}/\sqrt{Q_T} = O_p(1)$ by (i) and Lemma 6(a); (iii) $Q_S/Q_T = o_p(1)$ and $Q_S/Q_T^{1/2} = o_p(1)$ by (i) and Lemma 6(a); 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

$$\text{(A.15)}\quad \sqrt{h_1'Qh_1} - \sqrt{d_{\beta_1}^2 Q_T}$$
$$= \frac{1}{2} m^{-1/2}\left(2c_{\beta_1}d_{\beta_1}Q_{ST} + c_{\beta_1}^2 Q_S\right)$$
$$= \frac{c_{\beta_1}d_{\beta_1}Q_{ST}}{(d_{\beta_1}^2 Q_T)^{1/2}}\left(\frac{d_{\beta_1}^2 Q_T}{m}\right)^{1/2} + \frac{1}{2}\frac{c_{\beta_1}^2 Q_S}{(d_{\beta_1}^2 Q_T)^{1/2}}\left(\frac{d_{\beta_1}^2 Q_T}{m}\right)^{1/2}$$
$$= \frac{c_{\beta_1}\,\text{sgn}(d_{\beta_1})Q_{ST}}{Q_T^{1/2}} + o_p(1),$$

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

$$\text{(A.16)} \quad I_\nu\big(\sqrt{d_{\beta_1}^2 Q_T}\big)\exp\big(-\sqrt{d_{\beta_1}^2 Q_T}\big)\big(2pi\sqrt{d_{\beta_1}^2 Q_T}\big)^{1/2} = 1 + O_p(n^{-1/2})$$

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

Now, suppose (β_2^*, λ_2^*) does not necessarily satisfy (4.1). It is convenient to make a change of variables from (β, λ) to (τ, δ), where

$$\text{(A.17)} \quad \tau = \lambda^{1/2} c_\beta \quad \text{and} \quad \delta = \lambda^{1/2} d_\beta.$$

Let $\widetilde{h} = (\tau, \delta)'$. Then $\lambda \xi_\beta(Q) = \widetilde{h}' Q \widetilde{h}$ and $\lambda d_\beta^2 Q_T = \delta^2 Q_T$. Let $F_{2P}(\tau, \delta)$ be the two-point distribution on (τ, δ) that puts equal weight on $(\tau^*, \delta^*) = ((\lambda^*)^{1/2} c_{\beta^*}, (\lambda^*)^{1/2} d_{\beta^*})$ and $(\tau_2^*, \delta_2^*) = ((\lambda_2^*)^{1/2} c_{\beta_2^*}, (\lambda_2^*)^{1/2} d_{\beta_2^*})$. Let $\delta_{\max}$ denote the value of δ that maximizes $|\delta|$ over δ in the support of $F_{2P}(\tau, \delta)$; that is, $\delta_{\max} = \max\{|\delta^*|, |\delta_2^*|\}$. Let $\nu = (k-2)/2$.

Using this notation and the definition of LR* in Corollary 1, we have LR* equals

$$\begin{aligned}
\text{(A.18)} \quad & \frac{\int e^{-(\tau^2+\delta^2)/2}(\widetilde{h}' Q \widetilde{h})^{-\nu/2} I_\nu(\sqrt{\widetilde{h}' Q \widetilde{h}})\, dF_{2P}(\tau, \delta)}{\int e^{-\delta^2/2}(\delta^2 Q_T)^{-\nu/2} I_\nu(\sqrt{\delta^2 Q_T})\, dF_{2P}(\tau, \delta)} \\
&= \frac{\int e^{-(\tau^2+\delta^2)/2}(\widetilde{h}' Q \widetilde{h})^{-(\nu+1/2)/2} e^{\sqrt{\widetilde{h}' Q \widetilde{h}}}\, dF_{2P}(\tau, \delta)}{\int e^{-\delta^2/2}(\delta^2 Q_T)^{-(\nu+1/2)/2} e^{\sqrt{\delta^2 Q_T}}\, dF_{2P}(\tau, \delta)}(1 + o_p(1)) \\
&= \bigg\{\int e^{-(\tau^2+\delta^2)/2}\bigg(\frac{\widetilde{h}' Q \widetilde{h}}{\delta^2 Q_T}\bigg)^{-(\nu+1/2)/2}(\delta^2)^{-(\nu+1/2)/2} \\
&\qquad \times e^{(\sqrt{\delta^2} - \sqrt{\delta_{\max}^2})\sqrt{Q_T}} e^{\sqrt{\widetilde{h}' Q \widetilde{h}} - \sqrt{\delta^2 Q_T}}\, dF_{2P}(\tau, \delta)\bigg\} \\
&\quad \times \bigg\{\int e^{-\delta^2/2}(\delta^2)^{-(\nu+1/2)/2} e^{(\sqrt{\delta^2} - \sqrt{\delta_{\max}^2})\sqrt{Q_T}}\, dF_{2P}(\tau, \delta)\bigg\}^{-1} \\
&\quad \times (1 + o_p(1)) \\
&= \frac{\int e^{-(\tau^2+\delta^2)/2}(\delta^2)^{-(\nu+1/2)/2} e^{(\sqrt{\delta^2} - \sqrt{\delta_{\max}^2})\sqrt{Q_T}} e^{\tau\,\mathrm{sgn}(\delta) Q_{ST} Q_T^{-1/2}}\, dF_{2P}(\tau, \delta)}{\int e^{-\delta^2/2}(\delta^2)^{-(\nu+1/2)/2} e^{(\sqrt{\delta^2} - \sqrt{\delta_{\max}^2})\sqrt{Q_T}}\, dF_{2P}(\tau, \delta)} \\
&\quad \times (1 + o_p(1)),
\end{aligned}$$

where the first equality holds by (A.16), the second equality holds by algebra, and the third equality holds by (iv) and (A.15).

If (β_2^*, λ_2^*) satisfies (4.1), then $\tau^* = -\tau_2^*$, $\delta^* = \delta_2^*$, and $\delta_{\max} = |\delta^*| = |\delta_2^*|$. In this case, the terms in the numerator and denominator of the right-hand side of (A.18) that involve $(\sqrt{\delta^2} - \sqrt{\delta_{\max}^2})\sqrt{Q_T}$ equal zero, and the right-hand side of (A.18) without $(1 + o_p(1))$ equals

$$\text{(A.19)} \quad \frac{\frac{1}{2} e^{-((\tau^*)^2 + (\delta^*)^2)/2} ((\delta^*)^2)^{-(\nu+1/2)/2} (e^{\tau^* \operatorname{sgn}(\delta^*) Q_{ST} Q_T^{-1/2}} + e^{-\tau^* \operatorname{sgn}(\delta^*) Q_{ST} Q_T^{-1/2}})}{e^{-(\delta^*)^2/2} ((\delta^*)^2)^{-(\nu+1/2)/2}}$$
$$= e^{-(\tau^*)^2/2} \cosh(\tau^* Q_{ST} Q_T^{-1/2}),$$

using $(\exp(x) + \exp(-x))/2 = \cosh(x)$. The function $\cosh(\cdot)$ is even. Hence, $\cosh(\tau^* Q_{ST} Q_T^{-1/2}) = \cosh(\tau^* \mathrm{LM}_n^{1/2})$. The latter is strictly increasing in LM_n because $\cosh(\cdot)$ is continuous and strictly increasing on $\mathbb{R}^+$. This completes the proof of Theorem 8(b).

We now establish Theorem 8(c). Suppose (β_2^*, λ_2^*) does not satisfy the second condition of (4.1). Then either $\delta_{\max} > |\delta_2^*|$ or $\delta_{\max} > |\delta^*|$. Suppose $\delta_{\max} > |\delta_2^*|$. Then $\exp((\sqrt{(\delta_2^*)^2} - \sqrt{\delta_{\max}^2})\sqrt{Q_T}) = o_p(1)$ using (i), $\delta_{\max} = |\delta^*| > 0$, and the right-hand side of (A.18) without $(1 + o_p(1))$ equals

$$\text{(A.20)} \quad \frac{e^{-((\tau^*)^2 + (\delta^*)^2)/2} ((\delta^*)^2)^{-(\nu+1/2)/2} e^{\tau^* \operatorname{sgn}(\delta^*) Q_{ST} Q_T^{-1/2}} + o_p(1)}{e^{-(\delta^*)^2/2} ((\delta^*)^2)^{-(\nu+1/2)/2} + o_p(1)}$$
$$= e^{-(\tau^*)^2/2} e^{\tau^* \operatorname{sgn}(\delta^*) Q_{ST} Q_T^{-1/2}} + o_p(1),$$

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 $\delta_{\max} > |\delta^*|$.

Note that the case where $\beta^* = \beta_{\mathrm{AR}}$ or $\beta_2^* = \beta_{\mathrm{AR}}$ is subsumed in the case just considered, because in such cases there is no solution to the second equation in (4.1) and, hence, we must have $\delta_{\max} > |\delta^*|$ or $\delta_{\max} > |\delta_2^*|$.

Next, suppose (β_2^*, λ_2^*) satisfies the second condition of (4.1), but not the first condition. Then $\tau^* \neq -\tau_2^*$, $\delta^* = \delta_2^*$, $\delta_{\max} = |\delta^*| = |\delta_2^*| > 0$, and the right-hand side of (A.18) without $(1 + o_p(1))$ equals

$$\text{(A.21)} \quad \frac{1}{2}\left(e^{-(\tau^*)^2/2} e^{\tau^* \operatorname{sgn}(\delta^*) Q_{ST} Q_T^{-1/2}} + e^{-(\tau_2^*)^2/2} e^{\tau_2^* \operatorname{sgn}(\delta^*) Q_{ST} Q_T^{-1/2}}\right),$$

which is a continuous function of $Q_{ST} Q_T^{-1/2}$ that is not even because $\tau^* \neq -\tau_2^*$. This completes the proof of Theorem 8(c). *Q.E.D.*

PROOF OF COMMENT (i) TO THEOREM 8: We write the $\mathrm{LR}^*(Q_1, Q_T; \beta^*, \lambda^*)$ statistic as a function of Q_S, $\mathcal{S}_2^2$, and Q_T, say $\mathrm{LR}^*(Q_S, \mathcal{S}_2^2, Q_T; \beta^*, \lambda^*)$. 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 (A.15)–(A.19) with $(Q_S, \mathcal{S}_2^2)$ held fixed, when (β_2^*, λ_2^*) satisfies (4.1) we have $\lim_{m\to\infty} \mathrm{LR}^*(Q_S, \mathcal{S}_2^2, q_{T,m}; \beta^*, \lambda^*) = \exp(-\frac{1}{2}(\tau^*)^2)\cosh(|\tau^*|(Q_S\mathcal{S}_2^2)^{1/2})$. Because $Q_S\mathcal{S}_2^2 \sim \chi_1^2$, this implies that the conditional critical value function of LR*, viz., $\kappa_\alpha(q_T; \beta^*, \lambda^*)$, 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 $\kappa_\alpha(Q_T; \beta^*, \lambda^*)$ converges in probability to the same constant as $n \to \infty$ because $Q_T/n \to_p \alpha_T'\alpha_T > 0$. *Q.E.D.*

REFERENCES

ANDERSON, T. W. (1946): "The Non-Central Wishart Distribution and Certain Problems of Multivariate Statistics," *The 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," *The Annals of Mathematical Statistics*, 21, 570–582.

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression with Weak Instruments," Discussion Paper 1476, Cowles Foundation, Yale University. Available at *http://cowles.econ.yale.edu*.

——— (2006a): "Performance of Conditional Wald Tests in IV Regressions with Weak Instruments," *Journal of Econometrics*, forthcoming.

——— (2006b): "Supplement to 'Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression'," *Econometrica Supplementary Material*, 74, *http://www.econometricsociety.org/ecta/supmat/5333data.pdf*. Also available at James Stock's website.

CHAMBERLAIN, G. (2003): "Instrumental Variables, Invariance, and Minimax," Unpublished Manuscript, Department of Economics, Harvard University.

CHAMBERLAIN, G., AND G. IMBENS (2004): "Random Effects Estimators with Many Instrumental Variables," *Econometrica*, 72, 295–306.

DONALD, S. G., AND W. K. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

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 (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1366.

GUGGENBERGER, P., AND R. J. SMITH (2005): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," Working Paper, Department of Economics, UCLA.

——— (2006): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory*, 21, 667–709.

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*, Vol. 2. New York: Wiley.

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

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

——— (2004): "Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model," *Review of Economics and Statistics*, 86, 418–423.

LEBEDEV, N. N. (1965): *Special Functions and Their Applications*. Englewood Cliffs, NJ: Prentice–Hall.

LEHMANN, E. L. (1986): *Testing Statistical Hypotheses* (Second Ed.). New York: Wiley.

MOREIRA, M. J. (2001): "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," Working Paper Series 37, Center for Labor Economics, Department of Economics, University of California, Berkeley.

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

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. New York: Wiley.

OTSU, T. (2006): "Generalized Empirical Likelihood under Weak Identification," *Econometric Theory*, 21, forthcoming.

SAWA, T. (1969): "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimator," *Journal of the American Statistical Association*, 64, 923–937.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20, 518–529.

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 E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.