Optimal Invariant Similar Tests for Instrumental Variables Regression

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

Marcelo J. Moreira
Department of Economics
Harvard University

James H. Stock
Deparment of Economics
Harvard University

This paper is preliminary and incomplete.
October 2002
Revised: April 2004

Abstract

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

The results yield the power envelope for invariant similar tests. This allows one to assess and compare the power properties of existing tests, such as the Anderson-Rubin, LM, and conditional LR tests, and the new optimal WAP and PO invariant similar tests.

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

JEL Classification Numbers: C12, C30.

1 Introduction

TO BE ADDED.

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

2 Model and Sufficient Statistics

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

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

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

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

Our interest is in testing the null hypothesis

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

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

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

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

Note that $Z'X = 0$.

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

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

The reduced-form errors $[v_1 {:} v_2]$ are iid across rows with each row having a mean zero bivariate normal distribution with 2×2 nonsingular covariance matrix Ω. In order to obtain exact optimal tests, we assume in this section that Ω is known. As shown below, asymptotically valid tests can be obtained by replacing Ω by an estimator when Ω is unknown.

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

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

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

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

Lemma 1 *For the model in* (2.5),
(a) $Z'Y$ *and* $X'Y$ *are sufficient statistics for* θ,
(b) $Z'Y$ *and* $X'Y$ *are independent,*

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

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

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

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

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

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

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

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

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

2. The statistic T can be written as $d_\beta (Z'Z)^{1/2} \widehat{\pi}_0$, where $\widehat{\pi}_0$ denotes the maximum likelihood estimator of π under H_0. This follows from Lemma 2(b) because under H_0 minus two times the log-likelihood function for π based on the normal density of T is a constant plus $(T-d_{\beta_0}(Z'Z)^{1/2}\pi)'(T-d_{\beta_0}(Z'Z)^{1/2}\pi)$, which has first-order conditions given by $-(T-d_{\beta_0}(Z'Z)^{1/2}\widehat{\pi}_0)d_{\beta_0}(Z'Z)^{1/2}=0$. Note that $\widehat{\pi}_0=(Z'Z)^{-1}Z'Y\Omega^{-1}a_0\cdot(a_0'\Omega^{-1}a_0)^{-1}$ and $E\widehat{\pi}_0=\pi$ under H_0.

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

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

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

This holds because some algebra shows that

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

3 Similar Tests

A test based on the sufficient statistics (S,T) is *similar* if its null rejection rate does not depend on π. The parameter π determines the strength of the instrumental variables Z. The finite sample performance of some 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 asymptotic critical values are employed. By definition, similar tests do not suffer from this problem. For this reason, it is important to characterize the class of similar tests.

Let the $[0,1]$-valued statistic $\phi(S,T)$ denote a (possibly randomized) test that depends on the sufficient statistics S and T.

The following result is given in Moreira (2001).

Proposition 1 *A test $\phi(S,T)$ is similar with significance level α if and only if $E_{\beta_0}(\phi(S,T)|T=t)=\alpha$ for almost all t, where $E_{\beta_0}(\cdot|T=t)$ denotes conditional expectation when $\beta=\beta_0$ (which does not depend on π).*

Comments: 1. The proof of this result 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. This implies that any function of T whose expectation does not depend on π is equal to a constant with T probability one. In particular, for a similar test $\phi(S,T)$, $E_{\beta_0}(\phi(S,T)|T)$ is a function of T whose expectation equals α for all π. Hence, by completeness, $E_{\beta_0}(\phi(S,T)|T=t)$ must equal α for almost all t. Note that $E_{\beta_0}(\phi(S,T)|T)$ does not depend on π because S is ancillary under H_0.

2. Moreira (2003a) used Proposition 1 to specify conditional likelihood ratio (CLR) and conditional Wald (CW) tests that are similar. In this paper, we seek similar tests that have some optimal power properties.

3. Examples of similar tests (in the model with multivariate normal errors and known error covariance matrix Ω) include the Anderson and Rubin (1949) (AR) test, the LM test of Kleibergen (2002) and Moreira (2001), and the CLR and CW tests of Moreira (2003a) (where for each test an estimator of the unknown Ω matrix that appears in the test statistic is replaced by the known matrix Ω).

4. We use Proposition 1 below to characterize the class of invariant similar tests.

4 Invariant Tests

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

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

$$\begin{aligned} G &= \{g_F : g_F(x) = Fx \text{ for } x \in R^{k\times 2} \text{ for some } k \times k \text{ orthogonal matrix } F\} \text{ and} \\ \overline{G} &= \{\overline{g}_F : \overline{g}_F(\beta,\pi) = (\beta,\ (Z'Z)^{-1/2}F'(Z'Z)^{1/2}\pi) \text{ for some } k \times k \text{ orthogonal} \\ &\qquad \text{matrix } F\}. \end{aligned} \tag{4.1}$$

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

An *invariant* test, $\phi(S, T)$, under the group G is one for which $\phi(FS, FT) = \phi(S, T)$ for all $k \times k$ orthogonal matrices F. By definition, a *maximal invariant* is a function of $[S:T]$ that is invariant and takes different values on different *orbits* of G.[3] Every invariant test can be written as a function of a maximal invariant, see Thm. 6.1 of Lehmann (1986, p. 285). Hence, it suffices to restrict attention to the class of tests that depend only on a maximal invariant.

Let

$$\begin{aligned} Q &= [S:T]'[S:T] = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix} = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} \text{ and} \\ Q_1 &= \left(S'S, S'T\right)' = (Q_S, Q_{ST})'. \end{aligned} \tag{4.2}$$

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

For convenience, we use Q and (Q_1, Q_T) interchangeably. For example, if we define a function $h(Q)$, then $h(Q_1, Q_T)$ is presumed to be defined such that $h(Q_1, Q_T) =$

$h(Q)$. Although this involves some abuse of notation, it is justified by the one-to-one transformation from Q to (Q_1, Q_T).

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

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

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

$$\lambda = \pi' Z' Z \pi. \tag{4.3}$$

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

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

$$\begin{aligned} \psi_{AR}(Q) &= Q_S \ = S'S, \\ \psi_{LM}(Q) &= Q_{ST}^2/Q_T \ = (S'T)^2/T'T, \text{ and} \\ \psi_{CLR}(Q) &= \frac{1}{2}\left(Q_S - Q_T + \sqrt{(Q_S + Q_T)^2 - 4(Q_S Q_T - Q_{ST}^2)}\right). \end{aligned} \tag{4.4}$$

The CW test statistic is a more complicated function of Q. For brevity, we do not give it.

Invariant similar tests are characterized as follows:

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

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

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

3. The CLR and CW statistics are invariant statistics whose distributions under the null depend on Q_T. Hence, for these tests to be similar, their critical values must depend on Q_T. The CLR test rejects the null hypothesis when

$$\phi_{CLR}(Q) > \kappa_{CLR}(Q_T), \tag{4.5}$$

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

4. An equivalent condition to the one stated in the theorem is "$E_{\beta_0}\phi(Q_1, q_T) = \alpha$ for almost all q_T." This holds because Q_1 and Q_T are independent under H_0 by Lemma 3(f) below and, hence, $E_{\beta_0}(\phi(Q_1, Q_T)|Q_T = q_T) = E_{\beta_0}(\phi(Q_1, q_T)|Q_T = q_T) = E_{\beta_0}\phi(Q_1, q_T)$ for almost all q_T.

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

6. The statistic Q_T is complete under H_0 because $T \sim N(d_{\beta_0}\mu_\pi, I_k)$ is complete by Thm. 4.1 of Lehmann (1986, p. 142) and a function of a complete statistic is complete by the definition of completeness.

5 Optimal Tests for Weighted Average Power

5.1 Weighted Average Power

The invariant similar tests in (4.4) are ad hoc in the sense that they do not have any known optimal power properties. (The exception is the AR and LM tests when $k = 1$, which are equivalent tests when $k = 1$. Moreira (2001) shows that this test is uniformly most powerful unbiased.) We now address the question of optimal invariant similar tests. We determine the invariant similar test that has maximum weighted average power (WAP) with respect to (wrt) a given weight function W over the parameter values in the alternative. The use of sufficiency and invariance reduces the dimension of the alternative parameters that need to be considered from $1 + k + 2p$ for $\theta = (\beta, \pi', \xi', \gamma')'$ to just 2 for $(\beta, \lambda)'$. In consequence, it is relatively easy to specify weight functions W of interest.

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

$$K(\phi, W) = \int E_{\beta,\lambda}\phi(Q)dW(\beta, \lambda), \tag{5.1}$$

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

Let

$$g_W(q_1, q_T) = \int_{R\times R^+} f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)dW(\beta, \lambda), \tag{5.2}$$

where $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ denotes the joint density of (Q_1, Q_T) at (q_1, q_T). Let $q_1 =$

$(q_S, q_{ST})'$. WAP can be written as power against the single density $g_W(q_1, q_T)$:

$$\begin{aligned} & K(\phi, W) \\ &= \int_{R\times R^+} \left[\int_{R^+\times R\times R^+} \phi(q_S, q_{ST}, q_T) f_{Q_1,Q_T}(q_S, q_{ST}, q_T; \beta, \lambda) dq_S dq_{ST} dq_T\right] dW(\beta, \lambda) \\ &= \int_{R^+\times R\times R^+} \phi(q_1, q_T) g_W(q_1, q_T) dq_1 dq_T \end{aligned} \tag{5.3}$$

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

For example, suppose one takes the weight function W to be point mass at (β^*, λ^*). That is,

$$W_{\beta^*,\lambda^*}(\beta, \lambda) = \begin{cases} 1 & \text{if } (\beta, \lambda) = (\beta^*, \lambda^*) \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

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

Most existing tests in the literature are two-sided tests. Examples include the tests in (4.4). To obtain optimal two-sided tests one can specify W to give weight to β values both less than and greater than β_0. Examples are given in Section 6 below.

5.2 Optimal Invariant Similar Tests for Weighted Average Power

We want to find a test that maximizes WAP for weight function W among all level α invariant similar tests. By Theorem 2, invariant similar tests must be similar conditional on $Q_T = q_T$ for almost all q_T. In addition, by (5.3), WAP for weight function W equals unconditional power against the single density $g_W(q_1, q_T)$. In turn, the latter equals expected conditional power given Q_T. Hence, it suffices to determine the test that maximizes conditional power given $Q_T = q_T$ among tests that are invariant and are similar conditional on $Q_T = q_T$, for each q_T.

Conditional power given $Q_T = q_T$ is

$$K(\phi, W|Q_T = q_T) = \int_{R^+\times R} \phi(q_1, q_T) g_W(q_1|q_T) dq_1, \tag{5.5}$$

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

$$g_W(q_1|q_T) = \frac{g_W(q_1, q_T)}{g_W(q_T)}, \tag{5.6}$$

where $g_W(q_1, q_T)$ is defined in (5.2),

$$\begin{aligned} g_W(q_T) &= \int_{R^+\times R} g_W(q_1, q_T) dq_1 \\ &= \int_{R\times R^+} \int_{R^+\times R} f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda) dq_1 dW(\beta, \lambda) \\ &= \int_{R\times R^+} f_{Q_T}(q_T; \beta, \lambda) dW(\beta, \lambda), \end{aligned} \tag{5.7}$$

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

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

For any invariant test $\phi(Q_1, Q_T)$, conditional on $Q_T = q_T$, the null hypothesis is simple because $f_{Q_1|Q_T}(q_1|q_T; \beta_0)$ does not depend on π or λ. Given the WAP criterion function $K(\phi, W)$, the alternative hypothesis of concern also is simple. In particular, conditional on $Q_T = q_T$, the alternative density of interest is $g_W(q_1|q_T)$. In consequence, by the Neyman-Pearson Lemma, the test of significance level α that maximizes conditional power given $Q_T = q_T$ is of the likelihood ratio (LR) form and rejects H_0 when the LR is sufficiently large. In particular, the conditional WAP-LR test statistic is

$$LR_W(Q_1, q_T) = \frac{g_W(Q_1|q_T)}{f_{Q_1|Q_T}(Q_1|q_T; \beta_0)} = \frac{g_W(Q_1, q_T)}{g_W(q_T) f_{Q_1|Q_T}(Q_1|q_T; \beta_0)}. \tag{5.8}$$

The unconditional WAP-LR test statistic is $LR_W(Q_1, Q_T)$.

In order to provide an explicit expression for $LR_W(Q_1, Q_T)$, we now determine the densities $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$, $f_{Q_T}(q_T; \beta, \lambda)$, and $f_{Q_1|Q_T}(q_1|q_T; \beta_0)$ that arise in (5.2), (5.7), and (5.8). These densities and the tests considered below depend on the following quantity:

$$\begin{aligned} \xi_\beta(q) &= h_\beta' q h_\beta \\ &= c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T, \text{ where} \\ h_\beta &= (c_\beta, d_\beta)'. \end{aligned} \tag{5.9}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

$$f_{Q_\mathsf{S}}(q_S) = K_3 q_S^{(k-2)/2} \exp\left(-q_S/2\right)$$

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* $K_4^{-1} = \overline{\pi}^{1/2}\Gamma((k-1)/2)/\Gamma(k/2)$.

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

Comments: 1. The joint density $f_{Q_1,Q_\mathsf{T}}(q_S, q_T; \beta, \lambda)$ given in part (a) of the lemma is a noncentral Wishart density.[4] The null density of $\mathcal{S}_2$ given in part (e) of the lemma is the same as that of the sample correlation coefficient from an iid sample of k observations from a bivariate normal distribution with means zero and covariance matrix I_2 when the means of the random variables are not estimated.

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

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

$$I_\nu(x) = (x/2)^\nu \sum_{j=0}^{\infty} \frac{(x^2/4)^j}{j!\Gamma(\nu + j + 1)}, \tag{5.10}$$

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

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

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

Equations (5.2), (5.7), and (5.8) and Lemma 3 combine to give the following result.

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

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

where

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

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

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

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

Because $\psi_{2,W}(q_T)$ does not depend on q_1, it can be absorbed into the conditional critical value given $Q_T = q_T$. Thus, the test based on $LR_W(q_1, q_T)$ is equivalent to a test based on $\psi_W(q_1, q_T)$. Because $\psi_W(q_1, q_T)$ is simpler than $LR_W(q_1, q_T)$, we focus on the test statistic $\psi_W(q_1, q_T)$.

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

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

$$\psi_W(Q_1, Q_T) > \kappa_\alpha(Q_T), \tag{5.11}$$

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

$$P_{\beta_0}(\psi_W(Q_1, q_T) > \kappa_\alpha(q_T) | Q_T = q_T) = \alpha, \tag{5.12}$$

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

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

The critical value $\kappa_\alpha(Q_T)$ can be approximated by simulating R iid random vectors $S_r \sim N(0, I_k)$ for $r = 1, ..., R$, where R is large (at least 1,000), computing $Q_1(r) = (S_r'S_r, S_r'\alpha \cdot Q_T)$ for $r = 1, ..., R$, and taking $\kappa_\alpha(Q_T)$ to be the $1 - \alpha$ sample quantile of $\{\psi_W(Q_1(r), Q_T) : r = 1, ..., R\}$. The p-value for the test based on $\psi_W(Q_1, Q_T)$ can be approximated by the fraction of values in $\{\psi_W(Q_1(r), Q_T) : r = 1, ..., R\}$ that exceed $\psi_W(Q_1, Q_T)$, where (Q_1, Q_T) are the values based on the original sample Y.

The following theorem summarizes the results of this section:

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

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

Using the definition of $I_\nu(x)$ in (5.10), $\psi_W(q_1, q_T)$ can be written as

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

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

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

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

Corollary 2 *The level α test based on $\xi_{\beta^*}(Q_1, Q_T)$ is the uniformly most powerful test among invariant similar tests against the alternative distributions indexed by* $\{(\beta^*, \lambda) : \lambda > 0\}$.

Comments: 1. The test based on $\xi_{\beta^*}(Q_1, Q_T)$ is equivalent to a test based on

$$\begin{aligned} Q_S + 2(d_{\beta^*}/c_{\beta^*})Q_{ST} &= Q_S + 2\frac{b^{*\prime}\Omega b_0}{\beta^* - \beta_0} Q_{ST} \\ &= Q_S + 2(\det(\Omega))^{-1/2}\frac{\omega_{11} - (\beta^* + \beta_0)\omega_{12} + \beta^*\beta_0\omega_{22}}{\beta^* - \beta_0} Q_{ST}, \end{aligned} \tag{5.15}$$

where $b^* = (1, -\beta^*)'$ and ω_{jk} denotes the (j,k) element of Ω. Hence, the test statistic is a linear combination of Q_S and Q_{ST}. When the null hypothesis specifies that $\beta_0 = 0$, the statistic in (5.15) reduces to

$$Q_S + 2(\det(\Omega))^{-1/2}\frac{\omega_{11} - \beta^*\omega_{12}}{\beta^* - \beta_0} Q_{ST}. \tag{5.16}$$

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

$$\begin{aligned} \zeta_\delta &= \frac{Q_S + \delta \mathcal{S}_2\sqrt{Q_S} - k}{\sqrt{2k + \delta^2}} > \overline{\kappa}_{\delta,\alpha}(Q_T), \text{ where} \\ \delta &= (2d_{\beta^*}/c_{\beta^*})\sqrt{Q_T} \text{ and} \\ P_{\beta_0}(\zeta_\delta &> \overline{\kappa}_{\delta,\alpha}(Q_T) | Q_T = q_T) = \alpha. \end{aligned} \tag{5.17}$$

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

3. The result of Corollary 2 is related to a result of Moreira (2001, Thm. 2(c) and its proof) regarding the point optimal similar test against (β^*, π^*). The latter test is the same as the point optimal similar test for any $(\widetilde{\beta}, \widetilde{\pi})$ for which sign$(\widetilde{\beta})$ =sign(β^*) and $\widetilde{\pi}/(\widetilde{\pi}'Z'Z\widetilde{\pi})^{1/2} = \pi^*/(\pi^{*\prime}Z'Z\pi^*)^{1/2}$. Thus, if one specifies a given "direction" of the π vector, then a one-sided (wrt β) similar test is available that is UMP wrt the "magnitudes" of β and π. This test rejects when $\pi^{*\prime}S$ is large if $\beta^* > \beta_0$. On the other hand, Corollary 2 shows that the one-sided (wrt β) invariant similar test depends on β^*, but not on π (or λ).

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

$$Q_{ST}/Q_T^{1/2} > \kappa_{\phi,\alpha}, \tag{5.18}$$

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

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

$$Q_S + 2(\det(\Omega))^{-1/2}(\beta_0\omega_{22} - \omega_{12})Q_{ST} > \kappa_\alpha(Q_T) \tag{5.19}$$

for some $\kappa_\alpha(\cdot)$, where ω_{ij} denotes the (i,j) element of Ω. The same test is the optimal one-sided test for β^* negative and arbitrarily large in absolute value for any weight functions over λ. In consequence, the optimal two-sided test for $|\beta^* - \beta_0|$ arbitrarily large and any weight function over λ values is the test in (5.19).

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

$$Q_S - 2\frac{\rho}{(1-\rho^2)^{1/2}}Q_{ST} > \kappa_\alpha(Q_T), \tag{5.20}$$

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

6 Two-Sided Tests

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

6.1 Symmetric-Alternative WAP Tests

The first method we consider is to use an invariant similar test that maximizes WAP for a weight function W that places weight on β values that are both larger and smaller than the null value β_0 and for which the magnitude of the weight depends on β only through $|\beta - \beta_0|$. We call such tests optimal WAP tests for *symmetric alternatives.*

Although simple, weight functions for symmetric alternatives have some serious drawbacks. These drawbacks stem from the fact that the underlying testing problem is not symmetric for the parameter vectors $(\beta_0-\delta, \lambda)$ and $(\beta_0+\delta, \lambda)$. The distribution of Q_T is noncentral χ^2_k with non-centrality parameter $d^2_\beta\lambda$, see Lemma 3(b). This noncentrality parameter takes on different values for the parameter vectors $(\beta_0-\delta, \lambda)$ and $(\beta_0+\delta, \lambda)$.[5] In consequence, the problems of testing against these two alternative parameter vectors are not equally difficult testing problems. This has undesirable consequences for the power of WAP tests for symmetric alternatives under strong IV asymptotics. In particular, calculations in Section 11 below show that such tests are not asymptotically efficient under strong IV asymptotics according to the usual criterion for asymptotic efficiency of two-sided tests in regular models.[6] Given this, we do not recommend WAP tests for symmetric alternatives.

6.2 Asymptotically Efficient WAP Tests

We are most interested in weight functions W that generate tests that have good all-around two-sided power properties. This includes high power when the IV's are strong. Results in Section 11 below show that to obtain tests that are asymptotically efficient under strong IV asymptotics, the weight function must be of the following form:

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

for some weight function $W_*(\cdot,\cdot)$, where β_2 and λ_2 are such that the distribution of $[-S:T]$ under (β_2,λ_2) equals the distribution of $[S:T]$ under (β,λ). Such weight functions place equal weight on (β,λ) and (β_2,λ_2). The parameter vector (β_2,λ_2) is the appropriate "other-sided" parameter vector to (β,λ) in the sense that (i) β_2 is on the other side of the null value β_0 from β, (ii) the marginal distributions of Q_S, Q_{ST}, and Q_T under (β_2,λ_2) are the same as under (β,λ), and (iii) the joint distribution of (Q_S,Q_{ST},Q_T) under (β_2,λ_2) equals that of $(Q_S,-Q_{ST},Q_T)$ under (β,λ), which corresponds to β_2 being on the other side of the null as β. Furthermore, WAP tests with weight functions as in (6.1) have the property that they have the same power against (β_2,λ_2) as against (β,λ) and, hence, are truly symmetric two-sided tests against these two parameter vectors. We call tests *asymptotically efficient* (AE) WAP tests.

For a weight function $W(\beta,\lambda)$ that is not of the form in (6.1), the WAP test is not truly two-sided in large samples. In particular, under strong IV asymptotics, the test behaves like a one-sided LM statistic, see Section 11. This provides strong motivation for considering weight functions of the form in (6.1) when one desires a two-sided test.

Given the distributions of S and T specified in Lemma 2 and $\lambda=\mu_\pi'\mu_\pi$, (β_2,λ_2) solves

$$\lambda_2^{1/2}c_{\beta_2}=-\lambda^{1/2}c_\beta \text{ and } \lambda_2^{1/2}d_{\beta_2}=\lambda^{1/2}d_\beta. \tag{6.2}$$

Note that c_β is proportional to $\beta-\beta_0$ and d_β is linear in β. Some calculations show that the solution to these two equations in (6.2) are

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

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

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

where

$$\xi^*_\beta(q) = c^2_\beta q_S - 2c_\beta d_\beta q_{ST} + d^2_\beta q_T. \tag{6.4}$$

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

Note that $\psi_{W_{\mathsf{AE}}}(q_1, q_T)$ is not the same as what one gets from a "symmetric alternatives" weight function because d_β takes the same value in each summand of $\psi_{W_{\mathsf{AE}}}(q_1, q_T)$, but does not if a "symmetric alternatives" weight function is employed.

A two-sided power envelope is obtained from the AE-WAP tests that have weight functions $W_*(\beta, \lambda)$ that give point mass to different points (β^*, λ^*). This yields weight functions $W(\beta, \lambda)$ that are two-point weight functions against the alternatives (β^*, λ^*) and (β^*_2, λ^*_2):

$$W_{AE-POI}(\beta, \lambda) = \frac{1}{2}1(\beta = \beta^*, \lambda = \lambda^*) + \frac{1}{2}1(\beta = \beta^*_2, \lambda = \lambda^*_2), \tag{6.5}$$

where (β^*_2, λ^*_2) is defined as (β_2, λ_2) is defined in (6.3) but with (β^*, λ^*) in place of (β, λ). We refer to this power envelope as the *asymptotically efficient two-sided* power envelope.

The class of tests based on weight functions W_{AE} of the form in (6.1) also can be motivated by considering an additional invariance condition to that in (4.1):

$$[S : T] \to [-S : T]. \tag{6.6}$$

The corresponding transformation in the parameter space is $(\beta, \lambda) \to (\beta_2, \lambda_2)$. This sign invariance condition is a natural condition to impose to obtain two-sided tests because the parameter vector (β_2, λ_2) is the appropriate "other-sided" parameter vector to (β, λ) for the reasons stated in the first paragraph of this section. The maximal invariant under this sign invariance condition (plus the invariance conditions in (4.1) is

$$(S'S, |S'T|, T'T) = (Q_S, |Q_{ST}|, Q_T). \tag{6.7}$$

The AR, LM, and CLR test statistics all depend on the data only through this maximal invariant and, hence, satisfy the sign invariance condition (6.6).

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

$$\frac{1}{2}f_{Q_1,Q_{\mathsf{T}}}(q_{S,}q_{ST}, q_T) + \frac{1}{2}f_{Q_1,Q_{\mathsf{T}}}(q_{S,} - q_{ST}, q_T), \tag{6.8}$$

where Lemma 3 provides an expression for $f_{Q_1,Q_{\mathsf{T}}}(q_{S,}q_{ST}, q_T)$. Hence, following the same argument as in Section 5.2, given a weight function $W_*(\beta, \lambda)$, the optimal WAP test statistic, call it $\psi^*_{W_*}(q_1, q_T)$, can be shown to satisfy

$$\psi^*_{W_*}(q_1, q_T) = \psi_{W_{\mathsf{AE}}}(q_1, q_T).^7 \tag{6.9}$$

Thus, the class of WAP tests for weight functions W_* and the invariance conditions of (4.1) and (6.6) equals the class of WAP tests for weight functions W_{AE} and the invariance condition of (4.1).

6.3 Locally-Unbiased WAP Tests

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

We start by determining two necessary conditions for an invariant test (under the invariance condition of (4.1)) to be unbiased. The first condition is similarity and the second condition is *local unbiasedness.* Local unbiasedness requires that the power function has zero derivative at the null hypothesis. Otherwise, the power function would dip below the size of the test for some alternatives close to the null. We show that the AR, LM, and CLR tests are locally unbiased.

Next, we determine the test that maximizes WAP, as defined in (5.1), among the class of similar locally-unbiased invariant tests. We do so using the same argument as in Section 5.2, but using the generalized Neyman-Pearson Lemma (see Lehmann (1986, Thm. 3.5, pp. 96-7)) in place of the Neyman-Pearson Lemma. The form of the optimal WAP test statistic is the same as in Section 5.2, only the critical value function differs.

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

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

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

3. The second condition of Theorem 4 is equivalent to

$$E_{\beta_0,\lambda}(\phi(Q)Q_{ST}/Q_T^{1/2}) = 0 \text{ for all } \lambda \geq 0.^8 \tag{6.10}$$

That is, any unbiased invariant test statistic $\phi(Q)$ must be uncorrelated with the pivotal statistic $Q_{ST}/Q_T^{1/2}$ under H_0. This condition is a special case of a result of Moreira (2003b, Lemma 1) that establishes that any unbiased test $\phi(S, T)$ must be uncorrelated with the pivotal statistic S under H_0.

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

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

Comment. Corollary 3 shows that the class of asymptotically-efficient two-sided invariant similar tests considered in Section 6.2 is contained in the class of locally-unbiased invariant similar tests considered in this section.

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

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

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

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

Comment. *Point optimal* locally-unbiased invariant similar tests are obtained by taking the weight function W to give point mass at a given alternative parameter (β, λ) of interest.

7 Point Optimal Invariant Non-similar Tests

7.1 One-sided Alternatives

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

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

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

and the point alternative

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

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

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

where $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ is given in Lemma 3(a). The effect of weighting by Λ under the null is to turn the composite null into a point null, so that the most powerful test can be obtained using the Neyman-Pearson Lemma. Specifically, let ϕ_Λ be the most powerful test of h_Λ against $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$, so that ϕ_Λ rejects the null when

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

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

If the test ϕ_Λ has size α for the null hypothesis H_0 in (7.1), i.e.,

$$\sup_{(\beta,\lambda)\in\omega} P_{\beta,\lambda}(NP_\Lambda(Q) > d_{\Lambda,\alpha}) = \alpha, \tag{7.5}$$

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

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

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

where $v = (k-2)/2$; the second expression follows from Lemma 3(a).

Let $R(\beta_0, \lambda_0, \beta^*, \lambda^*|\beta, \lambda)$ be the rejection rate of the test based on the statistic given by (7.6) when the true values are β and λ. The numerical problem is to find the value of λ_0 such that the test has size α. Denote this value of λ_0 by λ_0^{LF}; then λ_0^{LF} solves

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

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

7.2 Two-sided Tests

8 Simulation Results I: Normal Model with Known Covariance Matrix

This section reports numerical results for the one-sided tests developed in Sections 5 and 7. We first compare the power envelope for invariant similar tests with the power envelope for invariant tests that are not necessarily similar. As will be discussed in more detail, these power envelopes are essentially numerically identical. We therefore focus on similar tests, which are more computationally tractable than non-similar tests, and on the performance of point-optimal invariant (POI) similar tests. None of the POI tests have good overall power properties. In particular, their power can decrease as the difference between the true value of β and the hypothesized value β_0 increases. We therefore turn to WAP tests and report preliminary results for specific WAP tests which have power functions that come quite close to the power envelope.

Because the distributions of Q do not depend on the sample size, which is accounted for in λ, the finite-sample power functions are reported as a function of λ, but not of the sample size. In addition, the covariance matrix Ω is consistently estimated even under weak instrument asymptotics. Taken together, these two facts imply that the finite-sample power functions are also the power functions under weak-instrument asymptotics.

The power envelope for the invariant similar tests was computed using Gauss-Legendre quadrature integration of the conditional distribution of ζ_δ in (5.17); all other power functions and power envelopes were computed by Monte Carlo simulation using 10,000 draws. Conditional critical value functions also are computed using 10,000 draws. The solution of the condition in (7.7) for the point optimal non-similar test was simplified because, for nearly all values of (β^*, λ^*), the rejection rate turns out to be monotone in λ_0; once a candidate value of λ_0^{LF} was ascertained, the condition in (7.7) is verified for a grid of values for λ.

Throughout, we focus on tests with significance level 5% and on the case $\beta_0 = 0$. The remaining parameters characterizing the distribution of the tests are λ, k, $\rho = corr(v_1, v_2)$, and the alternative, β. To facilitate comparison of the results across cases, we adopt two transformations of these parameters. Specifically, we consider λ/k rather than λ, as it is a more natural measure of the strength of the instruments, and in addition the power functions are plotted as a function of the rescaled alternative $(\lambda/k)^{1/2}(\beta - \beta_0) \equiv B$. The full set of numerical results have been computed for $\lambda/k = 0.5, 1, 2, 4, 8, 16$, which span the range from very weak instruments to quite strong instruments, and for $\rho = 0.95, 0.50, 0.20, -0.50$. To conserve space, we report only a subset of these results here and consider the case $k = 5$.

Figure 1 presents the power envelopes for the invariant similar tests (solid line) and for the invariant non-similar tests (dashed line). The power envelopes are reported both for negative and positive alternative values of β. Note, however, that the tests (and envelopes) themselves are for one-sided tests. There are three noteworthy features of these results:

1. The power envelopes for the similar and non-similar tests are essentially the same up to numerical accuracy, for all values of λ and ρ. The reason for this is twofold. On the one hand, the conditional critical values for the POI similar test depend on q_T only weakly in the range of q_T that is most likely to occur under the alternative; thus these conditional tests are nearly unconditional. On the other hand, the optimal non-similar tests have rejection rates that are nearly equal to 5%. Thus, these non-

similar tests are nearly similar. Because the optimal non-similar test is nearly similar and the optimal similar test is nearly unconditional, the two tests are virtually the same. The similar tests are more numerically tractable than the non-similar tests, so we focus our attention on the former below.

2. The power functions are neither symmetric nor strongly asymmetric, which suggests that two-sided tests based on equally-weighted symmetric alternatives could perform well.

3. There is a curious blip in some power envelopes. This blip occurs at the value of the alternative for which $\beta = 1/\rho$. The special role of this value is most easily seen by considering the POI similar test based on ζ_δ in (5.17). Straightforward algebra reveals that $\beta = 1/\rho$ is the value of the alternative at which δ in (5.17) changes sign. For values of $\beta < 1/\rho$, large positive values of S_2 lead to rejection; however, for values of $\beta > 1/\rho$, large negative values of S_2 lead to rejection. When $\beta = 1/\rho$, the point optimal test does not depend on S_2 and only depends on Q_S. The blip is therefore associated with this qualitative change in the nature of the conditional POI test.

One approach to testing in the absence of a UMP test is to consider POI tests that have power functions tangent to the power envelope at a certain value, for example at 50% or 70% power. If those power functions remain sufficiently close to the power envelope against alternatives other than that for which the test is point optimal, then that particular POI test provides a good practical choice; cf. King (1988). Here, two issues arise: the choice of λ^* and the choice of the alternative β. To see whether this approach has potential, in Figure 2 we plot the power functions of various POI tests along with the invariant similar power envelope. The individual power functions plotted in Figure 2 are for the locally optimal test given by (5.18), the most distant optimal test given by (5.19), and for several tests with intermediate points of tangency, chosen to be optimal against the rescaled alternatives $B = 0.5, 1.0, 1.5$ (which have power curves tangent to the envelope at powers of approximately 0.3, 0.5, and 0.7). The results in Figure 2 show that the power functions for the POI tests are generally not monotonic. For example, for $\rho = 0.5$ and $\lambda/k = 1$, the power of the locally optimal test and the POI test against $B = 0.5$ initially increase and then decline for $\beta > 1/\rho$, while the opposite is true for the most-distant POI test. Thus, our simulations suggest that no single POI test provides uniformly good performance.

The previous results encourage looking for a conditional WAP test that has a power function uniformly close to the invariant similar power envelope. Figure 3 presents preliminary results for two trial WAP tests. Both weight functions place unit weight on $\lambda^*/k = 1$. For the first WAP test (labeled LRW-P in Figure 3), the weight function is approximately proportional to $2^{-\beta}$. For the second (labeled $LR_W - Q$ in Figure 3), the weight function is the chi-squared density (as a function of β), evaluated on a log scale for β, where the degrees of freedom are chosen by numerical optimization as a function of ρ. For $\rho > 0$, both WAP tests have power functions that are uniformly close to the power envelopes, and when λ/k is large and/or ρ is small, the power functions are essentially on the power envelope. For $\rho = -0.5$, the $LR_W - Q$ test has a power function close to the power envelope, but the $LR_W - P$ test has a non-monotonic power function for distant alternatives.

While preliminary, the results in Figure 3 suggest that WAP tests may provide a feasible way to achieve near-UMP performance in the one-sided testing problem.

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

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

We use weak IV asymptotics, as in Staiger and Stock (1997), to analyze the properties of the procedures considered. We consider three versions of the finite sample tests introduced in Sections 5 and 6. The first version is suitable for the case of uncorrelated errors that exhibit *contemporaneous homoskedasticity.* By this we mean that $E(V_i V_i' | Z_i, X_i)$ is a constant matrix that does not depend on i, where V_i denotes the reduced-form error vector for the i-th observation (i.e., V_i is the i-th row of V written as a column 2-vector). In a time series setting this still allows for the errors to exhibit *temporal conditional heteroskedasticity* with respect to *lagged* values of the errors, IV's, and exogenous variables (i.e., $E(V_i V_i' | Z_{i-1}, X_{i-1}, V_{i-1}, Z_{i-2}, X_{i-2}, V_{i-2}, ...)$ may be random).

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

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

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

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

9.1 Assumptions

We use the following high-level assumptions concerning the IV's, exogenous variables, and errors. The assumptions are quite similar to those of Staiger and Stock (1997), but they allow for the possibility of heteroskedastic and autocorrelated errors because the form of the asymptotic variance matrix Φ in Assumption 4 is not restricted.

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

Assumption 2. $n^{-1}\overline{Z}'\overline{Z} \to_p D$ for some pd $(k+p) \times (k+p)$ matrix D.

Assumption 3. $n^{-1}V'V \rightarrow_p \Omega$ for some pd 2×2 matrix Ω.

Assumption 4. $n^{-1/2}vec(\overline{Z}'V) \rightarrow_d N(0, \Phi)$ for some pd $2(k+p) \times 2(k+p)$ matrix Φ.

In Assumption 4, $vec(\cdot)$ denotes the column by column vec operator.

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

Assumption 1 is the weak IV assumption. Assumptions 2 and 3 hold under suitable conditions by a weak law of large numbers (WLLN), see below. Assumption 4 holds under suitable conditions by a central limit theorem (CLT). Assumptions 1-4 are consistent with non-normal, heteroskedastic, autocorrelated errors and IV's and regressors that may be random or non-random.

For example, Assumptions 2-4 are implied by any one of the following assumptions:

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

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

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

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

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

If the errors are contemporaneously homoskedastic and $\{V_i \otimes \overline{Z}_i : i \geq 1\}$ are uncorrelated, the following key assumption holds. Under this assumption (and Assumptions 1-4), the tests described in Sections 5 and 6 but with Ω replaced by a consistent estimator $\widehat{\Omega}_n$ have asymptotic significance level α, as desired.

Assumption 5. $\Phi = \Omega \otimes D$, where Φ is defined in Assumption 4.

In Section 9.2 below, we impose Assumption 5, but in Sections 9.3 and 9.4, we do not. Assumption 5 is implied by any one of Assumptions IID, INID, and MDS plus the following.

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

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

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

Lemma 4 (a) *Any one of Assumptions* IID, INID, MDS, *and* CORR *implies Assumptions* 2-4.

(b) *Any one of Assumptions* IID, INID, *and* MDS *plus Assumption* HOM *imply Assumption* 5.

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

We estimate Ω ($\in R^{2\times 2}$) (defined in Assumption 3) via

$$\widehat{\Omega}_n = n^{-1}\widehat{V}'\widehat{V}, \text{ where } \widehat{V} = Y - P_Z Y - P_X Y.^9 \tag{9.1}$$

Let $\widehat{V}_i$ denote the i-th row of $\widehat{V}$ written as a column 2-vector.

Under Assumptions 2-4, the variance estimator is consistent.

Lemma 5 *Under Assumptions* 2-4*,* $\widehat{\Omega}_n \to_p \Omega$.

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

9.2 Homoskedastic Uncorrelated Errors

We now introduce tests that are suitable for (possibly) non-normal, homoskedastic, uncorrelated errors and unknown covariance matrix. That is, the tests are suitable when Assumptions 1-5 hold.

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

$$\begin{aligned}\widehat{S}_n &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2},\\ \widehat{T}_n &= (Z'Z)^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0 \cdot (a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2},\\ \widehat{Q}_{1,n} &= \left(\widehat{S}_n'\widehat{S}_n, \widehat{S}_n'\widehat{T}_n\right)', \text{ and } \widehat{Q}_{T,n} = \widehat{T}_n'\widehat{T}_n.\end{aligned} \tag{9.2}$$

The AR, LM, and CLR test statistics for the case of unknown Ω are defined as in (4.4), but with S and T replaced by $\widehat{S}_n$ and $\widehat{T}_n$.

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

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

where $\psi_W(\cdot,\cdot)$ is defined in Corollary 1 and $\kappa_\alpha(\cdot)$ is defined in (5.12) (and can be calculated by simulation using the method described there).

Next, we establish the asymptotic distributions of $\widehat{S}_n$ and $\widehat{T}_n$. Let S_∞ and T_∞ be independent random k-vectors with

$$\begin{aligned} S_\infty &\sim N(c_\beta L^{1/2}C, I_k) \text{ and } T_\infty \sim N(d_\beta L^{1/2}C, I_k), \text{ where} \\ L &= D_{11} - D_{12}D_{22}^{-1}D_{21}, \\ D &= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \text{ and } D_{j\ell} \in R^{k\times k} \text{ for } j,\ell = 1,2. \end{aligned} \tag{9.4}$$

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

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

Lemma 6 *Under Assumptions* 1-5,
(a) $(S_n, T_n) \to_d (S_\infty, T_\infty)$,
(b) $(\widehat{S}_n, \widehat{T}_n) - (S_n, T_n) \to_p 0$, *and*
(c) $(\widehat{S}_n, \widehat{T}_n) \to_d (S_\infty, T_\infty)$.

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

2. Lemma 6 and the continuous mapping theorem imply that the asymptotic distributions of the AR, LM, and CLR test statistics are given by the distributions of the test statistics in (4.4) with (S,T) replaced by (S_∞, T_∞). In particular, under the null hypothesis, the AR and LM statistics have asymptotic χ^2_k and χ^2_1 distributions, respectively.

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

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

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

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

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

Theorem 6 *Under Assumptions* 1-5,
(a) $(\psi_W(Q_{1,n}, Q_{T,n}),\ \kappa_\alpha(Q_{T,n})) \to_d (\psi_W(Q_{1,\infty}, Q_{T,\infty}),\ \kappa_\alpha(Q_{T,\infty}))$,
(b) $(\psi_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}),\ \kappa_\alpha(\widehat{Q}_{T,n})) - (\psi_W(Q_{1,n}, Q_{T,n}),\ \kappa_\alpha(Q_{T,n})) \to_p 0$, *and*
(c) $(\psi_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}),\ \kappa_\alpha(\widehat{Q}_{T,n})) \to_d (\psi_W(Q_{1,\infty}, Q_{T,\infty}),\ \kappa_\alpha(Q_{T,\infty}))$.

Theorem 6 leads to the following results.

Corollary 4 *Under Assumptions* 1-5,
(a) $1(\psi_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) > \kappa_\alpha(\widehat{Q}_{T,n})) - 1(\psi_W(Q_{1,n}, Q_{T,n}) > \kappa_\alpha(Q_{T,n})) \to_p 0$,
(b) $P(\psi_W(Q_{1,n}, Q_{T,n}) > \kappa_\alpha(Q_{T,n})) \to P(\psi_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty}))$,
(c) $P(\psi_W(\widehat{Q}_{1,n}, \widehat{Q}_{T,n}) > \kappa_\alpha(\widehat{Q}_{T,n})) \to P(\psi_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty}))$, *and*
(d) *under the null hypothesis,* $P(\psi_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty})) = \alpha$.

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

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

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

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

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

The finite sample optimality properties of the test based on $(\psi_W(Q_{1,n}, Q_{T,n}), \kappa_\alpha(Q_{T,n}))$, see Theorem 3, lead to the following asymptotic WAP optimality results.

Corollary 5 *Suppose Assumptions* 1 *and* 2 *hold and the reduced-form errors* $\{V_i : i \geq 1\}$ *are iid normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and pd variance matrix* Ω. *Let* $\{\phi_n : n \geq 1\}$ *be a sequence of level* α *invariant similar tests for known* Ω. *Then, the level* α *invariant similar tests based on* $\{(\psi_W(Q_{1,n}, Q_{T,n}), \kappa_\alpha(Q_{T,n})) : n \geq 1\}$ *have asymptotic power that satisfies*

$$\begin{aligned} &\overline{\lim}_{n\to\infty} \int P_{\beta,\lambda}(\phi_n \text{ rejects } H_0) dW(\beta, \lambda) \\ &\leq \lim_{n\to\infty} \int P_{\beta,\lambda}(\psi_W(Q_{1,n}, Q_{T,n}) > \kappa_\alpha(Q_{T,n})) dW(\beta, \lambda) \\ &= \int P_{\beta,\lambda}(\psi_W(Q_{1,\infty}, Q_{T,\infty}) > \kappa_\alpha(Q_{T,\infty})) dW(\beta, \lambda), \end{aligned}$$

where $P_{\beta,\lambda}(\cdot)$ *denotes probability when the true parameters are* β *and* π *for arbitrary* π *such that* $\pi' Z' Z \pi = \lambda$.

Comment. The Corollary shows that the asymptotic power of the sequence of optimal WAP tests for known Ω exceeds that of any sequence of level α invariant similar tests. Furthermore, the same inequality holds for the asymptotic power of the sequence of HOM-WAP tests (by Corollary 4(c)), which provides an asymptotic optimality result for these tests. This asymptotic optimality result has the disadvantage that these tests are not similar for each n, but are only asymptotically similar. But, it has the advantage that it applies to tests that do not require knowledge of Ω, which is rarely known in practice. For a stronger asymptotic optimality result for HOM-WAP tests, see Section 10.

9.3 Heteroskedasticity-Robust Tests

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

$$\begin{aligned}
\widetilde{S}_n &= \widetilde{\Sigma}_{S,n}^{-1/2} n^{-1/2} Z'Yb_0 \text{ and} \qquad (9.7)\\
\widetilde{T}_n &= \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0 - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1/2}\widetilde{S}_n\right), \text{ where}\\
\widetilde{\Sigma}_{S,n} &= n^{-1}\sum_{i=1}^{n}\left(\widehat{V}_i'b_0Z_i\right)\left(\widehat{V}_i'b_0Z_i\right)', \widetilde{\Sigma}_{TS,n} = n^{-1}\sum_{i=1}^{n}\left(\widehat{V}_i'\widehat{\Omega}_n^{-1}a_0Z_i\right)\left(\widehat{V}_i'b_0Z_i\right)',\\
\widetilde{\Sigma}_{T,n} &= \widetilde{\Sigma}_{T,n}^{*} - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1}\widetilde{\Sigma}_{TS,n}', \ \widetilde{\Sigma}_{T,n}^{*} = n^{-1}\sum_{i=1}^{n}\left(\widehat{V}_i'\widehat{\Omega}_n^{-1}a_0Z_i\right)\left(\widehat{V}_i'\widehat{\Omega}_n^{-1}a_0Z_i\right)',
\end{aligned}$$

and $\widehat{\Omega}_n$ and $\widehat{V}_i$ are defined in (9.1).[10]

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

Heteroskedasticity-robust AR, LM, and CLR test statistics, denoted HR-AR, HR-LM, and HR-CLR, are defined as in (4.4), but with (S,T) replaced by $(\widetilde{S}_n, \widetilde{T}_n)$. The appropriate critical values for these test statistics are the same as in the homoskedastic case. Thus, the critical values for the AR and LM tests are from χ^2_k and χ^2_1 distributions, respectively. The critical value function for the HR-CLR test is the same as in the homoskedastic error case and is given in Table I of Moreira (2003a).

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

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

and $\kappa_\alpha(\cdot)$ is defined in (5.12) (and can be calculated by the method following (5.12)). Note that the critical value function $\kappa_\alpha(\cdot)$ for the HR-WAP test is the same as for the HOM-WAP test.

We now analyze the asymptotic properties of HR-AR, HR-LM, HR-CLR, and HR-WAP tests. Define

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

The estimators $\widetilde{\Sigma}_{S,n}$, $\widetilde{\Sigma}_{TS,n}$, and $\widetilde{\Sigma}_{T,n}$ converge in probability to $\widetilde{\Sigma}_S$, $\widetilde{\Sigma}_{TS}$, and $\widetilde{\Sigma}_T$, respectively, when Assumptions 1-4 and the following assumptions hold.

Assumption 6. $n^{-1}\sum_{i=1}^n (V_i \otimes \overline{Z}_i)(V_i \otimes \overline{Z}_i)' \rightarrow_p \Phi$.

Assumption 7. $n^{-1}\sum_{i=1}^n (||\overline{Z}_i||^4 + ||\overline{Z}_i||^3||V_i||) = O_p(1)$.

Any one of Assumptions IID, INID, or MDS is sufficient for Assumption 6.

Assumption 7 holds under Assumption IID or MDS plus the following assumption.

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

Assumption 7 holds under Assumption INID plus the following assumption.

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

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

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

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

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

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

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

2. Theorem 7(b) and the continuous mapping theorem imply that the asymptotic distributions of the HR-AR, HR-LM, and HR-CLR test statistics under Assumptions *1-4, 6,* and *7* are given by the distributions of the test statistics in (4.4) with (S, T)

replaced by $(\widetilde{S}_\infty, \widetilde{T}_\infty)$. Hence, the HR-AR and HR-LM statistics have asymptotic null χ^2_k and χ^2_1 distributions, respectively. In addition, the critical value function of the HR-CLR test is the same as that in the homoskedastic case and is determined by the density in Lemma 3(c). If Assumption HOM holds, then the asymptotic power functions of the HR-AR, HR-LM, and HR-CLR tests are the same as the non-heteroskedasticity-robust versions of these tests.

3. Under the assumptions of Corollary 5, Assumption HOM holds and HR-WAP tests have the same asymptotic power function as HOM-WAP tests. Hence, Corollary 5 provides a weak asymptotic optimality result for HR-WAP tests. See Section 10 below for a asymptotic optimality stronger result.

9.4 Heteroskedasticity and Autocorrelation Robust Tests

Tests that are robust to heteroskedasticity and autocorrelation in the reduced-form errors $\{V_i : i \geq 1\}$ are obtained by using the tests introduced in the previous subsection but with different estimators in place of $\widetilde{\Sigma}_{S,n}$, $\widetilde{\Sigma}_{TS,n}$, $\widetilde{\Sigma}_{T,n}$, and $\widetilde{\Sigma}^*_{T,n}$. These are the only changes that are needed. In place of these estimators, one uses estimators of $\Sigma_{S,\infty}$, $\Sigma_{TS,\infty}$, $\Sigma_{T,\infty}$, and $\Sigma^*_{T,\infty}$, respectively, that are consistent (at least under the null hypothesis), where

$$\Sigma_\infty = \begin{bmatrix} \Sigma_{S,\infty} & \Sigma'_{TS,\infty} \\ \Sigma_{TS,\infty} & \Sigma^*_{T,\infty} \end{bmatrix} = \lim_{n\to\infty} var\left(n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} V_i' b_0 Z_i \\ V_i' \Omega^{-1} a_0 Z_i \end{pmatrix} \right) \text{ and}$$
$$\Sigma_{T,\infty} = \Sigma^*_{T,\infty} - \Sigma_{TS,\infty} \Sigma^{-1}_{S,\infty} \Sigma'_{TS,\infty}. \tag{9.11}$$

Let

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

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

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

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

Let $\overline{S}_n$, $\overline{T}_n$, $\overline{Q}_{1,n}$, and $\overline{Q}_{T,n}$ denote $\widetilde{S}_n$, $\widetilde{T}_n$, $\widetilde{Q}_{1,n}$, and $\widetilde{Q}_{T,n}$, respectively, with these changes. Heteroskedasticity and autocorrelation-robust AR, LM, and CLR test statistics, denoted HR-AR, HR-LM, and HR-CLR, respectively, are defined as in (4.4), but with (S, T) replaced by $(\overline{S}_n, \overline{T}_n)$. The appropriate critical values for these test statistics are the same as in the homoskedastic case.

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

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

where $\kappa_\alpha(\cdot)$ is defined in (5.12).

The HAR-AR, HAR-LM, HAR-CLR, and HAR-WAP tests have correct asymptotic significance level under Assumptions 1-4 plus the additional conditions that are needed to obtain consistency of $\overline{\Sigma}_n$ for Σ_∞. Furthermore, these tests have the same asymptotic power functions as the corresponding AR, LM, CLR, and HOM-WAP tests when Assumptions 1-5 hold and the same asymptotic power functions as the HR-AR, HR-LM, HR-CLR, and HR-WAP tests when Assumptions 1-4, 6, and 7 hold.

10 Asymptotic Optimality with Weak IV's, IID Normal Errors, and Unknown Covariance Matrix

In this section, we show that the tests HOM-WAP, HR-WAP, and HAR-WAP exhibit certain asymptotic WAP optimality properties when the IV's are weak and the errors are iid normal with *unknown* covariance matrix.

For the asymptotic optimality results, we set up a sequence of models (or *experiments*) with the parameters renormalized such that no parameter can be estimated asymptotically without error, as is standard in the asymptotic efficiency literature, e.g., see van der Vaart (1998, Ch. 9). For the parameters β and C, no renormalization is required given Assumption 1 because neither can be consistently estimated in the weak IV asymptotic setup. For the parameters Ω and η, renormalizations are required. We take the true parameters Ω and η to satisfy

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

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

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

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

The following basic finite sample and asymptotic results hold.

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

symmetric, and $vech(N_\Omega) \sim N(\Omega_1, E(\zeta - E\zeta)(\zeta - E\zeta)')$*, where* $\zeta = vech(v_0 v_0')$*,* $v_0 \in R^2$*, and* $v_0 \sim N(0, \Omega_0)$*, provided Assumption 2 also holds.*

Comment. The results of the Lemma hold under the null hypothesis $\beta = \beta_0$ and fixed alternatives $\beta \neq \beta_0$.

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

$$\phi_n(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0)) \rightarrow_d \phi(N_Z, N_X, N_\Omega) \qquad (10.2)$$

for some function $\phi(\cdot, \cdot, \cdot)$. Given Lemma 8(b), there are an abundance of convergent sequences of tests.

A convergent sequence of tests $\{\phi_n : n \geq 1\}$ is said to be *asymptotically similar* if

$$P_{\beta,C,\Omega_0,\eta_0}(\phi(N_Z, N_X, N_\Omega) = 1) = \alpha \qquad (10.3)$$

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

Define

$$\begin{aligned} S_\infty &= L^{-1/2} N_Z b_0 \cdot (b_0 \Omega b_0)^{-1/2} \sim N(c_\beta L^{1/2} C, I_k), \\ T_\infty &= L^{-1/2} N_Z \Omega^{-1} a_0 \cdot (a_0 \Omega^{-1} a_0)^{-1/2} \sim N(d_\beta L^{1/2} C, I_k), \text{ and} \\ Q_\infty &= [S_\infty : T_\infty]'[S_\infty : T_\infty]. \end{aligned} \qquad (10.4)$$

Note that S_∞ and T_∞ are independent and the current definition of S_∞ and T_∞ is consistent with that in (9.4).

The transformation $h_\Omega(\cdot)$ from N_Z to $[S_\infty : T_\infty]$ is one-to-one. Hence, we have

$$\phi(N_Z, N_X, N_\Omega) = \phi(h_\Omega^{-1}(S_\infty, T_\infty), N_Z, N_\Omega) = \overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega) \qquad (10.5)$$

for some function $\overline{\phi}$.

As in Section 4, we consider the group of transformations given in (4.1) but with $\overline{g}_F(\beta, \pi)$ replaced by $\overline{g}_F(\beta, C) = (\beta,\ L^{-1/2} F' L^{1/2} C)$ acting on the parameters (β, C). The maximal invariant is Q_∞.

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

$$\overline{\phi}(S_\infty, T_\infty, N_X, N_\Omega) \sim \phi^*(Q_\infty, N_X, N_\Omega) \qquad (10.6)$$

for some function ϕ^*, where $\sim$ denotes "has the same distribution as." Examples of convergent sequences of asymptotically invariant and asymptotically similar tests include sequences of AR, LM, CLR, HOM-WAP, HR-WAP, and HAR-WAP tests.

We now establish an upper bound on asymptotic WAP.

Theorem 8 *Suppose Assumptions* 1 *and* 2 *hold, the reduced-form errors* $\{V_i : i \geq 1\}$ *are iid normal, independent of* $\{\overline{Z}_i : i \geq 1\}$, *with mean zero and pd variance matrix* Ω, *and* Ω *and* η *are as in* (10.1). *For any convergent sequence of asymptotically invariant and asymptotically similar tests* $\{\phi_n : n \geq 1\}$, *we have*

$$\begin{aligned}&\lim_{n\to\infty} \int P_{\beta,\lambda,\Omega,\eta_0}(\phi_n(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0)) = 1)dW(\beta,\lambda)\\ &= \int P_{\beta,\lambda,\Omega_0,\eta_0}(\phi^*(Q_\infty, N_X, N_\Omega) = 1)dW(\beta,\lambda)\\ &\leq \int P_{\beta,\lambda,\Omega_0,\eta_0}(\psi_W(Q_1, Q_\infty) > \kappa_\alpha(Q_\infty))dW(\beta,\lambda),\end{aligned}$$

where $P_{\beta,\lambda,\Omega,\eta_0}(\cdot)$ *denotes probability when the true parameters are* $(\beta, C, \Omega, \eta_0)$ *for some* C *such that* $CLC' = \lambda$.

Comments. 1. Under $H_0 : \beta = \beta_0$, the left- and right-hand sides of the inequality in the Theorem equal α.

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

Corollary 6 *Under the conditions of Theorem* 8, *the* HOM-WAP, HR-WAP, *and* HAR-WAP *tests of Section* 9 *are convergent sequences of asymptotically invariant and asymptotically similar tests that attain the upper bound on asymptotic* WAP *given in Theorem* 8.

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

TO BE ADDED.

12 Simulation Results II: Non-normal Model with Unknown Covariance Matrix

TO BE ADDED.

13 Normal Model with Multiple Endogenous Variables and Known Covariance Matrix

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

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

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

The null hypothesis is

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

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

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

Lemma 9 *For the model in* (2.5) *generalized as in* (13.1)*,*
(a) *$Z'Y$ and $X'Y$ are sufficient statistics for θ,*
(b) *$Z'Y$ and $X'Y$ are independent,*
(c) *$X'Y$ has a multivariate normal distribution that does not depend on $(\beta', vec(\pi)')'$,*
(c) *$Z'Y$ has a multivariate normal distribution that does not depend on $\eta = [\gamma : \xi]$, and*
(d) *$Z'Y$ is a sufficient statistic for $(\beta', vec(\pi)')'$.*

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

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

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

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

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

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

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

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

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

Lemma 10 *For the model in* (2.5) *generalized as in* (13.1),
(a) $S\sim N(\mu_\pi(\beta-\beta_0)\cdot(b_0'\Omega b_0)^{-1/2},\ I_k)$,
(b) $T_j\sim N(\mu_\pi a'\Omega^{-1}\alpha_{0,j},\ I_k)$ *for* $j=1,...,m,$ *and*
(c) $S,\ T_1,...,T_m$ *are mutually independent.*

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

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

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

Consequently, the T statistic can be written as $(Z'Z)^{1/2}\widehat{\pi}_0a_0'\Omega^{-1}\alpha_0$, where $\widehat{\pi}_0$ denotes the maximum likelihood estimator of π under H_0 and $a_0=[\beta_0:I_m]\in R^{m\times(m+1)}$, $\widetilde{\pi}=(Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_0$, where $a_0=[\beta_0:I_m]$.

Next, we present a similarity result analogous to that of Proposition 1. Let the $[0,1]$-valued statistic $\phi(S,T)$ denote a (possibly randomized) test that depends on the sufficient statistics S and T.

Theorem 9 *A test $\phi(S,T)$ is similar with significance level α if and only if $E_{\beta_0,\pi}(\phi(S,T)|T=t)=\alpha$ for almost all t, where $E_{\beta_0}(\cdot|T=t)$ denotes conditional expectation given $T=t$ when $\beta=\beta_0$ (which does not depend on π).*

Comment: Examples of similar tests in the model with multiple endogenous variables, multivariate normal errors, and known error covariance matrix Ω include the AR test and the LM test of Kleibergen (2002).

We consider the same groups of transformations G and $\overline{G}$ defined in (4.1) when $m \geq 2$ as when $m=1$ (except that $x \in R^{k \times (m+1)}$ in the definition of G rather than $x \in R^{k \times 2}$). An *invariant* test, $\phi(S,T)$, under the group G is one for which $\phi(FS,FT)=\phi(S,T)$ for all $k \times k$ orthogonal matrices F. It suffices to restrict attention to the class of tests that depend only on a maximal invariant.

Define Q, Q_S, Q_{ST}, Q_T, and Q_1 as in (4.2), but with $T=[T_1 : \cdots : T_m]$. Hence, $Q=[S:T]'[S:T] \in R^{(m+1)(m+1)}$, $Q_S=S'S \in R$, $Q_{ST}=S'T \in R^m$, $Q_T=T'T \in R^{m \times m}$, and $Q_1=(S'S,S'T)' \in R^{m+1}$.

Theorem 10 *The $(m+1)\times(m+1)$ matrix Q is a maximal invariant for the transformations G.*

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

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

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

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

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

Invariant similar tests are characterized as follows:

Theorem 11 *An invariant test $\phi(Q)$ is similar with significance level α if and only if $E_{\beta_0}(\phi(Q)|Q_T=q_T)=\alpha$ for almost all q_T, where $E_{\beta_0}(\cdot|Q_T=q_T)$ denotes conditional expectation given $Q_T=q_T$ when $\beta=\beta_0$ (which does not depend on π).*

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

Let W be a weight function over (β, λ) values. That is, W is a probability distribution on the product of R^m and the space of pd $m \times m$ matrices, call it $R_{pd}^{m\times m}$. Weighted average power of a test $\phi(Q)$ with respect to W is given by (5.1). The expressions in (5.2)-(5.8) hold when $m \geq 2$ just as when $m = 1$, provided one adjusts the range of integration suitably. In particular, the integral over (β, λ) values is over $R^m \times R_{pd}^{m\times m}$, rather than $R \times R^+$, and the integral over (q_1, q_T) values is over $(R^+ \times R^m) \times R_{pd}^{m\times m}$, rather than $(R^+ \times R) \times R^+$. In particular, the optimal WAP LR statistic $LR_W(Q_1, Q_T)$ is as given in (5.8).

As in Section 5.2, in order to provide an explicit expression for the optimal WAP LR statistic $LR_W(Q_1, Q_T)$, we determine the densities $f_Q(q; \beta, \lambda)$, $f_{Q_T}(q_T; \beta, \lambda)$, and $f_{Q_1|Q_T}(q_1, q_T; \beta_0)$ that arise in (5.2), (5.7), and (5.8). Let

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

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

Lemma 11 (a) *The density of Q at $q \in R_{pd}^{(m+1)\times(m+1)}$ is a non-central Wishart density with k degrees of freedom, covariance matrix I_{m+1}, and non-centrality matrix (i.e., means sigma matrix) $\Delta_\beta'\lambda\Delta_\beta$:*

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

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

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

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

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

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

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

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

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

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

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

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

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

Equations (5.2), (5.7), and (5.8) and Lemma 11 combine to give the following result.

Corollary 7 *The optimal WAP test statistic for weight function W is given by*

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

where

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

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

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

Comments: 1. As when $m = 1$, $\psi_W(q_1, q_T)$ does **not** equal $\int f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$ $dW(\beta, \lambda)$ and likewise with $\psi_{2,W}(q_T)$. This is because numerous cancellations occur in the second expression in the first line of the Corollary 7, including the constants $K_{1,m}$ and $K_{2,m}$.

2. When $m = 2$, the density formulae given in Comment 3 following Lemma 11 yield alternative expressions for $\psi_W(q_1, q_T)$ and $\psi_{2,W}(q_T)$ that are easier to compute.

Because $\psi_{2,W}(q_T)$ does not depend on q_1, it can be absorbed into the conditional critical value given $Q_T = q_T$. Thus, the test based on $LR_W(q_1, q_T)$ is equivalent to a test based on $\psi_W(q_1, q_T)$. Because $\psi_W(q_1, q_T)$ is simpler than $LR_W(q_1, q_T)$, we focus on the test statistic $\psi_W(q_1, q_T)$.

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

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

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

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

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

The results of this section are summarized as follows:

Theorem 12 *The test that rejects H_0 when $\psi_W(Q_1, Q_T) > \kappa_\alpha(Q_T)$ maximizes WAP for the weight function W over all level α invariant similar tests.*

14 Appendix of Proofs

14.1 Proofs of Results Stated in Sections 2 and 3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Part (e) of the lemma follows from parts (b)-(d). □

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

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

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

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

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

The k-vector T is multivariate normal with mean

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

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

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

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

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

For completeness, we include a proof of Proposition 1.

Proof of Proposition 1. Sufficiency of the stated condition for similarity with significance level α holds by iterated expectations.

To show necessity, suppose $\phi(S, T)$ is similar with significance level α. Under the null hypothesis, $S \sim N(0, I_k)$, $T \sim N(d_{\beta_0} \mu_\pi, I_k)$, and S and T are independent by Lemma 2. Hence, S is ancillary for the parameter π. The density of T is

$$(2\pi)^{-k/2} \exp(-T'T/2 + T'(Z'Z)^{1/2} \pi d_{\beta_0} - d_{\beta_0}^2 \mu_\pi' \mu_\pi / 2)). \tag{14.10}$$

For $\pi \in R^k$ this forms an exponential family. Hence, by Thm. 4.1 of Lehmann (1986, p. 142), T is a complete sufficient statistic for π under H_0. By completeness, for any function h such that $E_{\beta_0, \pi} h(T) = 0$ for all $\pi \in R^k$, $h(t) = 0$ for almost all t.

Given $\pi \in R^k$, take $h(t) = E_{\beta_0,\pi}(\phi(S,T)|T = t) - \alpha$. We have

$$E_{\beta_0,\pi}h(T) = E_{\beta_0,\pi}E_{\beta_0,\pi}(\phi(S,T)|T) - \alpha = E_{\beta_0,\pi}\phi(S,T) - \alpha = 0, \qquad (14.11)$$

where the second equality holds by iterated expectations and the third equality holds by similarity of $\phi(S,T)$ with significance level α. Hence, by completeness, $h(t) = E_{\beta_0,\pi}(\phi(S,T)|T = t) - \alpha = 0$ for almost all t. The same argument applies for all $\pi \in R^k$, which completes the proof of necessity.

The distribution of Q given T does not depend on π when $\beta = \beta_0$ because T is sufficient for π in this case. In consequence, the conditional expectation $E_{\beta_0}(\cdot|Q_T = q_T)$ does not depend on π. □

14.2 Proofs of Results Stated in Section 4

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

First, suppose μ_1 and μ_2 are linearly independent (which implies that $k \geq 2$). Then, there exist linearly independent k-vectors $\mu_3, ..., \mu_k$ such that $\{\mu_1, ..., \mu_k\}$ span R^k. Applying the Gram-Schmidt procedure to $\{\mu_1, ..., \mu_k\}$, we now construct an orthogonal matrix F such that $F\mu_1$ and $F\mu_2$ depend on (μ_1,μ_2) only through $\mu_1'\mu_1$, $\mu_1'\mu_2$, and $\mu_2'\mu_2$. For a full column rank $k\times\ell$ matrix A, let $M_A = I_k - A(A'A)^{-1}A'$. We take $f_1 = \mu_1/||\mu_1||$, $f_2 = M_{\mu_1}\mu_2/||M_{\mu_1}\mu_2||, ..., f_k = M_{[\mu_1:\cdots:\mu_{k-1}]}\mu_k/||M_{[\mu_1:\cdots:\mu_{k-1}]}\mu_k||$. Define $F = [f_1{:}\cdots{:}f_k]'$. We have

$$\begin{aligned} F\mu_1 &= (f_1'\mu_1, ..., f_k'\mu_1)' = (||\mu_1||, 0, ..., 0)' \text{ and} \\ F\mu_2 &= (\mu_1'\mu_2/||\mu_1||, \mu_2'M_{\mu_1}\mu_2/||M_{\mu_1}\mu_2||, 0, ..., 0)'. \end{aligned} \qquad (14.12)$$

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

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

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

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

Proof of Theorem 2. Sufficiency of the stated condition for similarity holds by iterated expectations.

To show necessity of the condition, suppose $\phi(Q_1, Q_T)$ is similar with significance level α. Because $\phi(Q_1, Q_T) = \phi^*(S, T)$ for some function ϕ^*, Proposition 1 implies that

$$E_{\beta_0,\pi}(\phi(Q_1, Q_T)|\sigma(T)) = \alpha \ a.s. \tag{14.13}$$

for all $\pi \in R^k$, where $\sigma(T)$ denotes the σ-field generated by T. Because $\sigma(Q_T) \subset \sigma(T)$, the law of iterated expectations and (14.13) give

$$E_{\beta_0,\pi}(\phi(Q_1, Q_T)|\sigma(Q_T)) = E_{\beta_0,\pi}(E_{\beta_0,\pi}(\phi(Q_1, Q_T)|\sigma(T))|\sigma(Q_T)) = \alpha \ a.s. \tag{14.14}$$

for all $\pi \in R^k$. This is the desired result because the distribution of (Q_1, Q_T) only depends on π through λ, so that $E_{\beta_0,\pi}$ is equivalent to $E_{\beta_0,\lambda}$ in the left-hand side of (14.14). □

14.3 Proofs of Results Stated in Section 5

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

$$\begin{aligned} &K_1 \exp(-tr(M'M)/2)|q|^{(k-3)/2} \exp(-tr(q)/2) \\ &\times (tr(M'Mq))^{-(k-2)/4} I_{(k-2)/2}\left(\sqrt{tr(M'Mq)}\right). \end{aligned} \tag{14.15}$$

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

Part (b) holds because the distribution of Q_T is a noncentral chi-squared distribution with non-centrality parameter $d_\beta^2\lambda$ by Lemma 2(b) and (4.3). The stated form of the density is given in Anderson (1946, eqn. (6)).

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

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

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

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

Proof of Comment 4 to Corollary 2. The optimal test against β^* rejects if $\xi_{\beta^*}(Q_1, Q_T)$ is large and we have

$$\begin{aligned}
&\lim_{\beta^* \to \beta_0} \left(\xi_{\beta^*}(q_1, q_T) - d_{\beta^*}^2 q_T \right) / (\beta^* - \beta_0) \\
&= \lim_{\beta^* \to \beta_0} \left((\beta^* - \beta_0)(b_0'\Omega b_0)^{-1} q_S + 2b^{*\prime}\Omega b_0 (b_0'\Omega b_0)^{-1} (\det(\Omega))^{-1/2} q_{ST} \right) \\
&= 2(\det(\Omega))^{-1/2} q_{ST},
\end{aligned} \tag{14.16}$$

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

Proof of Comment 5 to Corollary 2. Comment 5 holds because (i) the optimal test against β^* rejects if $\xi_{\beta^*}(Q_1, Q_T)$ is large, (ii) we have

$$\begin{aligned}
&\lim_{\beta^* \to \infty} \left(\xi_{\beta^*}(q_1, q_T) - d_{\beta^*}^2 q_T \right) / c_{\beta^*}^2 \\
&= \lim_{\beta^* \to \infty} \left(q_S + 2(d_{\beta^*}/c_{\beta^*}) q_{ST} \right) \\
&= q_S + 2(\det(\Omega))^{-1/2} (\beta_0 \omega_{22} - \omega_{12}) q_{ST},
\end{aligned} \tag{14.17}$$

and (iii) the limit as $\beta^* \to -\infty$ in (14.17) is the same as when $\beta^* \to \infty$. The second equality in (14.17) holds because

$$\begin{aligned}
(\det(\Omega))^{1/2} d_{\beta^*}/c_{\beta^*} &= \frac{b^{*\prime}\Omega b_0}{\beta^* - \beta_0} = \frac{\omega_{11} - (\beta^* + \beta_0)\omega_{12} + \beta^*\beta_0\omega_{22}}{\beta^* - \beta_0} \text{ and so} \\
\lim_{\beta^* \to \infty} d_{\beta^*}/c_{\beta^*} &= (\det(\Omega))^{-1/2} (\beta_0\omega_{22} - \omega_{12}) \text{ and} \\
\lim_{\beta^* \to -\infty} d_{\beta^*}/c_{\beta^*} &= (\det(\Omega))^{-1/2} (\beta_0\omega_{22} - \omega_{12}). \ \square
\end{aligned} \tag{14.18}$$

14.4 Proofs of Results Stated in Section 6

THE FOLLOWING PROOF CAN BE ALTERED/SHORTENED BY JUST REFERENCING Moreira (2003b, Lemma 1) AND USING JUST THE FIRST AND LAST PARAGRAPHS OF THE PROOF.

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

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

$$\begin{aligned} I_1 &= \int\int \phi(q_1,q_T)\frac{\partial f_{Q_1|Q_T}(q_1,q_T;\beta_0,\lambda)}{\partial\beta}dq_1 f_{Q_T}(q_T;\beta_0,\lambda)dq_T \text{ and} \\ I_2 &= \int\int \phi(q_1,q_T)f_{Q_1|Q_T}(q_1,q_T;\beta_0)dq_1\frac{\partial f_{Q_T}(q_T;\beta_0,\lambda)}{\partial\beta}dq_T \\ &= \int \alpha\frac{\partial f_{Q_T}(q_T;\beta_0,\lambda)}{\partial\beta}dq_T = 0, \end{aligned} \tag{14.19}$$

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

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

$$\begin{aligned} & f_{Q_1|Q_T}(q_1,q_T;\beta,\lambda) = K_1K_2^{-1}\exp(-q_S/2)\det(q)^{(k-3)/2}q_T^{-(k-2)/2}\times \\ & \sum_{j=0}^{\infty}\frac{(\lambda\xi_\beta(q)/4)^j}{j!\Gamma((k-2)/2+j+1)}\Bigg/\sum_{j=0}^{\infty}\frac{(\lambda d_\beta^2 q_T/4)^j}{j!\Gamma((k-2/2)+j+1)} \end{aligned} \tag{14.20}$$

using Lemma 3(a) and (b) and (5.10).

Tedious algebraic manipulations show that

$$\begin{aligned} \frac{\partial f_{Q_1|Q_T}(q_1,q_T;\beta_0,\lambda)}{\partial\beta} &= \frac{\lambda}{2}f_{Q_1|Q_T}(q_1,q_T;\beta_0)q_{ST}(\det(\Omega))^{-1/2}\times \\ & I_{k/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T})/I_{(k-2)/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T}). \end{aligned} \tag{14.21}$$

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

$$\frac{\partial}{\partial\beta}\sum_{j=0}^{\infty}\frac{(\lambda\xi_\beta(q)/4)^j}{j!\Gamma((k-2)/2+j+1)} = \frac{\lambda}{4}\frac{\partial\xi_\beta(q)}{\partial\beta}\sum_{s=0}^{\infty}\frac{(\lambda\xi_\beta(q)/4)^s}{s!\Gamma(k/2+s+1)} \tag{14.22}$$

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

The necessary condition for unbiasedness, (14.19), and (14.21) give

$$\begin{aligned} 0 &= \int h(q_T)f_{Q_T}(q_T;\beta_0,\lambda)dq_T\frac{I_{k/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T})}{I_{(k-2)/2}(\sqrt{\lambda a_0'\Omega^{-1}a_0q_T})}, \text{ where} \\ h(q_T) &= \int \phi(q_1,q_T)q_{ST}f_{Q_1|Q_T}(q_1,q_T;\beta_0)dq_1. \end{aligned} \tag{14.23}$$

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

Alternatively, the second condition of the Theorem can be derived as a special case of a necessary condition given in Moreira (2003b, Lemma 1) for any test $\phi^*(S,T)$ to be unbiased: $E_{\beta_0}(\phi^*(S,T)S|T) = 0$ a.s. This condition implies that $E_{\beta_0}(\phi^*(S,T)S'T|T) = 0$ a.s. Because $\phi(Q_1, Q_T) = \phi^*(S,T)$ for some function ϕ^*, any unbiased invariant test $\phi(Q_1, Q_T)$ must satisfy: $E_{\beta_0,\pi}(\phi(Q_1, Q_T)Q_{ST}|T) = 0$ a.s. for all $\pi \in R^k$. Because Q_T is determined by T, the law of iterated expectations gives $E_{\beta_0,\pi}(\phi(Q_1, Q_T)Q_{ST}|Q_T) = 0$ for all $\pi \in R^k$. This is the desired result because the distribution of (Q_1, Q_T) only depends on π through λ, so that $E_{\beta_0,\pi}$ is equivalent to $E_{\beta_0,\lambda}$ here. $\square$

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

$$\begin{aligned} & E_{\beta_0}(\phi(Q_S, \mathcal{S}_2^2, Q_T)Q_{ST}|Q_T = q_T) = E_{\beta_0}(\phi(Q_S, \mathcal{S}_2^2, q_T)\mathcal{S}_2 Q_S^{1/2})q_T^{1/2} \\ = & \int E_{\beta_0}(\phi(q_S, \mathcal{S}_2^2, q_T)\mathcal{S}_2)q_S^{1/2} f_{Q_S}(q_S)dq_S \cdot q_T^{1/2} = 0 \end{aligned} \tag{14.24}$$

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

Proof of Theorem 5. By the same argument as in Section 5.2, it suffices to find the test that maximizes power against the single alternative density $g_W(q_1|q_T)$ conditional on $Q_T = q_T$. Given the restriction to locally-unbiased tests, we apply the generalized Neyman-Pearson (GNP) Lemma, see Lehmann (1986, Thm. 3.5, pp. 96-7), rather than the Neyman-Pearson Lemma. The GNP Lemma implies that the optimal (conditional) test rejects when $LR_W(Q_1, q_T) > \widetilde{\kappa}_{1\alpha}(q_T) + \widetilde{\kappa}_{2\alpha}(q_T)Q_{ST}$ for some $\widetilde{\kappa}_{1\alpha}(q_T)$ and $\widetilde{\kappa}_{2\alpha}(q_T)$ that are chosen such that the two conditions of Theorem 4 hold. As in Corollary 1, $LR_W(Q_1, q_T) = \psi_W(Q_1, q_T)/\psi_{2W}(Q_1, q_T)$ and $\psi_{2W}(Q_1, q_T)$ can be absorbed into the critical value functions. This yields the form of the test stated in the Theorem.

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

$$\left(E\left(\phi(Q_1, Q_T)|Q_T = q_T\right), E\left(\phi(Q_1, Q_T)Q_{ST}|Q_T = q_T\right)\right) \tag{14.25}$$

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

The set M is convex because the conditional expectation operator is linear. Moreover, M contains $(\alpha, 0)$ by considering the LM test. It also contains points (α, u_α^+) with $u_\alpha^+ > 0$ by considering the one-sided LM test which rejects H_0 when

$Q_{ST}/Q_T^{1/2} > c_\alpha$. This follows because the derivative of the conditional power function of this test is an increasing linear transformation of

$$\int 1\left(q_{ST}/q_T^{1/2} > c_\alpha\right) q_{ST} f_{Q_1|Q_T}(q_1, q_T; \beta_0) dq_1, \tag{14.26}$$

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

14.5 Proofs of Results Stated in Section 9

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

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

$$n^{-1}\widehat{V}'\widehat{V} - \Omega = (n^{-1}V'V - \Omega) - n^{-1}V'P_Z V - n^{-1}V'P_X V. \tag{14.27}$$

The first summand on the right-hand side of (14.27) converges in probability to zero by Assumption 3. The second summand satisfies

$$0 \le n^{-1}V'P_Z V \le n^{-1}V'P_{\widetilde{Z}} V = n^{-1}(n^{-1/2}V'\widetilde{Z})(n^{-1}\widetilde{Z}'\widetilde{Z})^{-1}(n^{-1/2}\widetilde{Z}'V) \to_p 0, \tag{14.28}$$

where the second inequality holds because the span of Z is contained in the span of $\widetilde{Z}$ and the convergence to zero holds by Assumptions 2 and 4. The third summand of (14.27) converges in probability to zero by an analogous argument. □

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

$$n^{-1}Z'Z = n^{-1}\widetilde{Z}'\widetilde{Z} - n^{-1}\widetilde{Z}'P_X\widetilde{Z} \to_p D_{11} - D_{12}D_{22}^{-1}D_{21} = L \tag{14.29}$$

using Assumption 2. Let N^* be a $(k + p) \times 2$ random matrix with $vec(N^*) \sim N(0, \Omega \otimes D)$. Using Assumptions 2 and 4, we obtain

$$\begin{aligned} n^{-1/2}Z'Vb_0 &= n^{-1/2}(\widetilde{Z} - P_X\widetilde{Z})'Vb_0 = n^{-1/2}(\widetilde{Z} - XD_{22}^{-1}D_{21})'Vb_0 + o_p(1) \\ &= \left[I_k : -D_{12}D_{22}^{-1}\right] n^{-1/2}\overline{Z}'Vb_0 + o_p(1) \to_d \left[I_k : -D_{12}D_{22}^{-1}\right] N^*b_0 \\ &= \left[I_k : -D_{12}D_{22}^{-1}\right] (b_0' \otimes I_{k+p})vec(N^*). \end{aligned} \tag{14.30}$$

Hence, we have

$$\begin{aligned} S_n &= (n^{-1}Z'Z)^{-1/2}(n^{-1/2}Z'Vb_0 + n^{-1}Z'ZCa'b_0) \cdot (b_0'\Omega b_0)^{-1/2} \to_d H, \text{ where} \\ H &= L^{-1/2}\left(\left[I_k : -D_{12}D_{22}^{-1}\right](b_0' \otimes I_{k+p})vec(N^*) + LCa'b_0\right) \cdot (b_0'\Omega b_0)^{-1/2} \end{aligned} \tag{14.31}$$

and the first equality holds by Assumption 1 and $Z'X = 0$. Using Assumption 5, the random vector H has a normal distribution with

$$\begin{aligned} EH &= L^{1/2}Ca'b_0 \cdot (b_0'\Omega b_0)^{-1/2} = c_\beta L^{1/2}C \text{ and} \\ var(H) &= L^{-1/2}\left[I_k : -D_{12}D_{22}^{-1}\right](b_0' \otimes I_{k+p})(\Omega \otimes D)(b_0 \otimes I_{k+p}) \\ &\quad \times \left[I_k : -D_{12}D_{22}^{-1}\right]' L^{-1/2} \cdot (b_0'\Omega b_0)^{-1} \\ &= L^{-1/2}\left[I_k : -D_{12}D_{22}^{-1}\right] D \left[I_k : -D_{12}D_{22}^{-1}\right]' L^{-1/2} = I_k, \end{aligned} \tag{14.32}$$

which completes the proof for S_n.

Analogously to (14.30), we have

$$n^{-1/2}Z'V\Omega^{-1}a_0 \to_d \left[I_k : -D_{12}D_{22}^{-1}\right]((a_0'\Omega^{-1}) \otimes I_{k+p})vec(N^*). \tag{14.33}$$

Using this, we obtain

$$\begin{aligned} T_n &= (n^{-1}Z'Z)^{-1/2}\left(n^{-1/2}Z'V\Omega^{-1}a_0 + n^{-1}Z'ZCa'\Omega^{-1}a_0\right) \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \to_d J, \text{ for} \\ J &= L^{-1/2}\left(\left[I_k : -D_{12}D_{22}^{-1}\right]((a_0'\Omega^{-1}) \otimes I_{k+p})vec(N^*) + LCa'\Omega^{-1}a_0\right) \cdot (a_0'\Omega^{-1}a_0)^{-1/2}. \end{aligned} \tag{14.34}$$

Analogously to (14.32), J has a normal distribution with $EJ = d_\beta L^{1/2}C$ and $var(J) = I_k$, which completes the proof for T_n.

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

$$\begin{aligned} &E(b_0' \otimes I_{k+p})vec(N^*)vec(N^*)'((\Omega^{-1}a_0) \otimes I_{k+p}) \\ &= E(b_0' \otimes I_{k+p})(\Omega \otimes D)((\Omega^{-1}a_0) \otimes I_{k+p}) = (b_0'a_0) \otimes D = 0. \end{aligned} \tag{14.35}$$

This completes the proof of part (a).

Part (b) holds by the definitions of $\widehat{S}_n$, $\widehat{T}_n$, S_n, and T_n because (i) $(Z'Z)^{-1/2}Z'Y = O_p(1)$ by the same sort of argument as in (14.29) and (14.30), (ii) $\widehat{\Omega}_n \to_p \Omega$ by Lemma 5, and (iii) Ω is pd by Assumption 3.

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

Proof of Theorem 6. The function $\psi_W(\cdot,\cdot)$ is continuous and does not depend on n, see its definition in Corollary 1. The same is true of the critical value function $\kappa_\alpha(\cdot)$ because the conditional distribution of $Q_{1,n}$ given $Q_{T,n}$ is absolutely continuous with a density that is a smooth function of q_T and does not depend on n, see Lemma 3(c) and the definition of $\kappa_\alpha(\cdot)$ in (5.12). In consequence, the result of the Theorem follows from Lemma 6, (9.5), and the continuous mapping theorem. □

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

$$P(|\widehat{\Psi}_n - \Psi_n| > \varepsilon) \to 0 \text{ for all } \varepsilon > 0. \tag{14.36}$$

We have

$$
\begin{aligned}
& P(|1(\widehat{\Psi}_n > 0) - 1(\Psi_n > 0)| > \varepsilon) \\
\le\; & P(\widehat{\Psi}_n > 0 \;\&\; \Psi_n \le 0) + P(\widehat{\Psi}_n \le 0 \;\&\; \Psi_n > 0).
\end{aligned} \tag{14.37}
$$

The first summand on the right-hand side of (14.37) satisfies

$$
P(\widehat{\Psi}_n > 0 \;\&\; \Psi_n \le 0) \le P(0 < \widehat{\Psi}_n \le \varepsilon) + o(1) \to P(0 < \Psi \le \varepsilon), \tag{14.38}
$$

where the inequality holds by (14.36) and the convergence holds by Theorem 6(c). The right-hand side of (14.38) converges to zero as $\varepsilon \to 0$ because Ψ has an absolutely continuously distribution by Lemma 3(a). Hence, the left-hand side of (14.38) converges to zero as $n \to \infty$.

By an analogous argument, the second summand on the right-hand side of (14.37) converges to zero as $n \to \infty$, which completes the proof of part (a).

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

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

Proof of Corollary 5. The inequality holds because it holds with the limits deleted, conditional on $\overline{Z}$, for each n, by Theorem 3.

The equality holds by Corollary 4(b) (which applies because Assumption 3 holds by a LLN for iid square-integrable random vectors and Assumptions 4 and 5 hold because $n^{-1/2}\overline{Z}'V \sim N(,\Omega \otimes (n^{-1}\overline{Z}'\overline{Z}))$ conditional on $n^{-1}\overline{Z}'\overline{Z}$ and $n^{-1}\overline{Z}'\overline{Z} \to_p D$, which implies that $n^{-1/2}\overline{Z}'V \to_d N(,\Omega \otimes D)$. $\square$

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

$$
\widehat{V}_j' b_0 = V_j' b_0 - Z_j'(Z'Z)^{-1}Z'Vb_0 - X_j'(X'X)^{-1}X'Vb_0 \tag{14.39}
$$

because $\widehat{V} = V - P_Z V - P_X V$. Using (14.39), some manipulations, and Assumption 7, we obtain

$$
n^{-1}\sum_{j=1}^{n}(\widehat{V}_j' b_0)^2 Z_j Z_j' - n^{-1}\sum_{j=1}^{n}(V_j' b_0)^2 Z_j Z_j' \to_p 0. \tag{14.40}
$$

In addition, we have

$$
\begin{aligned}
& n^{-1}\sum_{j=1}^{n}(V_j'b_0)^2 Z_j Z_j' \\
= & \ n^{-1}\sum_{j=1}^{n}(V_j'b_0)^2(\widetilde{Z}_j - D_{12}D_{22}^{-1}X_j)(\widetilde{Z}_j - D_{12}D_{22}^{-1}X_j)' + o_p(1) \\
= & \ n^{-1}\sum_{j=1}^{n} MB(V_i \otimes \overline{Z}_i)(V_i' \otimes \overline{Z}_i')B'M' + o_p(1) \\
\to_p & \ MB\Phi B'M', \qquad (14.41)
\end{aligned}
$$

where the first equality holds using Assumption 2 via some manipulations, the second equality holds by linear algebra, and convergence holds by Assumption 6. Combining (14.40) and (14.41), gives $\widetilde{\Sigma}_{S,n} \to_p \widetilde{\Sigma}_S$.

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

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

$$
\begin{aligned}
\widetilde{S}_n &= \widetilde{\Sigma}_{S,n}^{-1/2}\left(n^{-1/2}Z'Vb_0 + n^{-1}Z'ZCa'b_0\right) \\
&\to_d \widetilde{\Sigma}_S^{-1/2}\left[I_k : -D_{12}D_{22}^{-1}\right](b_0' \otimes I_{k+p})vec(N^*) + \widetilde{\Sigma}_S^{-1/2}LCa'b_0 \\
&\sim N(\widetilde{\Sigma}_S^{-1/2}LCa'b_0, I_k), \qquad (14.42)
\end{aligned}
$$

where $vec(N^*) \sim N(0, \Phi)$, the equality uses (9.7) and Assumption 1, and the convergence holds by part (a), (14.29), and (14.30).

By Lemma 5, part (a), and Assumption 4, the use of $\widehat{\Omega}_n^{-1}$, rather than Ω^{-1}, in the definition of T_n has no effect asymptotically. Hence, we have

$$
\begin{aligned}
\widetilde{T}_n &= \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1/2}Z'Y\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1}n^{-1/2}Z'Yb_0\right) + o_p(1) \\
&= \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1/2}Z'V\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1}n^{-1/2}Z'Vb_0\right) \\
&\quad + \widetilde{\Sigma}_{T,n}^{-1/2}\left(n^{-1}Z'ZCa'\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS,n}\widetilde{\Sigma}_{S,n}^{-1}n^{-1}Z'ZCa'b_0\right) + o_p(1) \\
&\to_d \widetilde{\Sigma}_T^{-1/2}\left(MA\,vec(N^*) - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB\,vec(N^*)\right) \\
&\quad + \widetilde{\Sigma}_T^{-1/2}\left(LCa'\Omega^{-1}a_0 - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}LCa'b_0\right), \qquad (14.43)
\end{aligned}
$$

where M, A, and B are defined in (9.9) and the convergence holds by (14.30) and (14.34). The covariance matrix of the limiting distribution in (14.43) is I_k because

$$
\begin{aligned}
& var\left(MA\,vec(N^*) - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB\,vec(N^*)\right) \\
= & \ MA\Phi A'M' - MA\Phi B'M'\widetilde{\Sigma}_S^{-1}\widetilde{\Sigma}_{TS}' - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB\Phi A'M' \\
& + \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB\Phi B'M'\widetilde{\Sigma}_S^{-1}\widetilde{\Sigma}_{TS}' \\
= & \ \widetilde{\Sigma}_T^* - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}\widetilde{\Sigma}_{TS}' = \widetilde{\Sigma}_T. \qquad (14.44)
\end{aligned}
$$

The convergence in (14.42) and (14.43) is joint and the limit random vectors are independent because

$$\begin{aligned} & cov(MA\,vec(N^*) - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB\,vec(N^*),\ \widetilde{\Sigma}_S^{-1/2}MB\,vec(N^*)) \qquad (14.45) \\ =\ & MA\Phi M'B'\widetilde{\Sigma}_S^{-1/2} - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1}MB\Phi B'M'\widetilde{\Sigma}_S^{-1/2} = \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1/2} - \widetilde{\Sigma}_{TS}\widetilde{\Sigma}_S^{-1/2} = 0. \end{aligned}$$

To complete the proof of part (b), we note that (i) Theorem 6(c) (with the changes indicated in Theorem **??**(b)) follows from (14.42)-(14.45) by the continuous mapping theorem, (ii) Corollary 4(c) follows immediately from Theorem 6(c), and (iii) Corollary 4(d) holds with $(\widetilde{Q}_{1,\infty}, \widetilde{Q}_{T,\infty})$ by the same reason as with $(Q_{1,\infty}, Q_{T,\infty})$. □

14.6 Proofs of Results Stated in Section 10

Proof of Lemma 8. Part (a) holds because (i) conditional on $[Z : X]$, equation (14.2) with (π, Ω, η) replaced by $(C/n^{12}, \Omega_0 + \Omega_1/n^{1/2}, \eta_0 + \eta_1/n^{1/2})$, where Ω_0 and η_0 are known and Ω_1 and η_1 are unknown, implies that $(Z'Y, X'Y, Y'Y)$ are sufficient statistics for $(\beta, C, \Omega_1, \eta_1)$ and (ii) $(n^{-1/2}Z'Y, n^{1/2}(\widehat{\eta}_n - \eta_0), n^{1/2}(\widehat{\Omega}_n - \Omega_0))$ is an equivalent set of sufficient statistics to $(Z'Y, X'Y, Y'Y)$.

Part (b) holds because (i) $vec(n^{-1/2}Z'V) \sim N(0, \Omega \otimes (n^{-1}Z'Z))$ conditional on $n^{-1}Z'Z$ and $n^{-1}Z'Z \to_p L$ (by (14.30) using Assumption 2) imply that $vec(n^{-1/2}Z'V)$ $\to_d N(0, \Omega \otimes L)$, (ii) $vec(n^{-1/2}Z'Z\pi a') = vec(n^{-1}Z'ZCa') \to_p LCa'$ by Assumption 2, (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 2) imply that $vec(n^{1/2}(\widehat{\eta}_n - \eta_0)) \to_d N(\eta_1, \Omega \otimes D_{22}^{-1})$, (iv) $n^{1/2}(\widehat{\Omega}_n - \Omega_0) = n^{1/2}(n^{-1}V'V - \Omega_0) - n^{-1/2}V'P_ZV - n^{-1/2}V'P_XV$ using (14.27), (v) $n^{1/2}(n^{-1}V'V - \Omega_0) = n^{-1/2}(V'V - EV'V) + \Omega_1$, (vii) $vech(n^{-1/2}(V'V - EV'V)) \to_d N(0, E(\zeta - E\zeta)(\zeta - E\zeta)')$ by a triangular array CLT for row-wise iid random vectors, (viii) $n^{-1/2}V'P_ZV = n^{-1/2} \cdot n^{-1/2}V'Z(n^{-1}Z'Z)^{-1}n^{-1/2}Z'V \to_p 0$ using (i), (ix) $n^{-1/2}V'P_XV \to_p 0$ by an analogous argument to (viii), and (x) 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. □

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

14.7 Proofs of Results Stated in Section 13

Proof of Lemma 9. The proof is essentially the same as that of Lemma 1. □

Proof of Lemma 10. The proof is similar to that of Lemma 2. For brevity, we only discuss the aspects of the proof that differ. To show independence of S and T_j, it suffices to show lack of covariance between S and T_j, because S and T_j are jointly multivariate normal. We have

$$cov(Z'Yb_0, Z'Y\Omega^{-1}\alpha_{0,j}) = cov(\sum_{i=1}^{n} Z_iY_i'b_0, \sum_{i=1}^{n} Z_iY_i'\Omega^{-1}\alpha_{0,j})$$
$$= \sum_{i=1}^{n} Z_iZ_i'cov(Y_i'b_0, Y_i'\Omega^{-1}\alpha_{0,j}) = \sum_{i=1}^{n} Z_iZ_i'b_0'\Omega\Omega^{-1}\alpha_{0,j} = 0, \qquad (14.46)$$

because $b_0'\alpha_{0,j} = 0$. By analogous calculations T_j and T_ℓ have zero covariance for $j \neq \ell$ provided $\alpha_{0,j}'\Omega^{-1}\alpha_{0,\ell} = 0$ for all $j \neq \ell$. Lastly, T_j has covariance matrix equal to I_k provided $cov(Z'Yb_0, Z'Y\Omega^{-1}\alpha_{0,j}) = Z'Z$. By analogous calculations to those in (14.46), the latter occurs if $\alpha_{0,j}'\Omega^{-1}\alpha_{0,j} = 1$ for $j = 1, ..., m$. The vectors $\alpha_{0,j}$ are chosen so that the desired conditions $b_0'\alpha_{0,j} = 0$, $\alpha_{0,j}'\Omega^{-1}\alpha_{0,\ell} = 0$, and $\alpha_{0,j}'\Omega^{-1}\alpha_{0,j} = 1$ hold. □

Proof of Theorem 9. The proof is the same as that of Proposition 1 provided the family of distributions of $T = [T_1 : \cdots : T_m]$ under H_0 is a km-parameter exponential family with parameter space that contains a km-dimensional rectangle. The log of the null density of T times minus two is $k\log(2\pi)$ plus

$$\sum_{j=1}^{m}(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j})'(T_j - (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j})$$
$$= tr\left(\sum_{j=1}^{m} T_jT_j'\right) + tr\left(\sum_{j=1}^{m}(Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j}\left((Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j}\right)'\right)$$
$$-2tr\left(\sum_{j=1}^{m}(Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j}T_j'\right), \qquad (14.47)$$

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

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

$$-2tr\left(\sum_{j=1}^{m}\widetilde{\pi}_jT_j'\right) = -2\sum_{j=1}^{m}\sum_{\ell=1}^{k}\widetilde{\pi}_{j,\ell}T_{j,\ell}, \text{ where}$$
$$\widetilde{\pi}_j = (Z'Z)^{1/2}\pi a_0'\Omega^{-1}\alpha_{0,j} \in R^k,$$
$$\widetilde{\pi}_j = (\widetilde{\pi}_{j,1}, ..., \widetilde{\pi}_{j,k})', \text{ and}$$
$$T_j = (T_{j,1}, ..., T_{j,k})'. \qquad (14.48)$$

The parameters $\widetilde{\pi} = [\widetilde{\pi}_1 : \cdots : \widetilde{\pi}_m] \in R^{k\times m}$ are the "natural" parameters of the exponential family. There is a one-to-one transformation from π to $\widetilde{\pi}$ provided

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

Proof of Theorem 10. The proof is the same as that of Theorem 1, but one considers vectors $(\mu_1, ..., \mu_m)$ and $(\widetilde{\mu}_1, ..., \widetilde{\mu}_m)$ instead of (μ_1, μ_2) and $(\widetilde{\mu}_1, \widetilde{\mu}_2)$. □

Proof of Theorem 11. The proof is the same as that of Theorem 2 using Theorem 9 in place of Proposition 1. □

Proof of Lemma 11. First, we establish part (a). The $k \times (m+1)$ matrix $[S{:}T]$ is multivariate normal with mean matrix $M = \mu_\pi \Delta_\beta$, all variances equal to one, and all correlations equal to zero by Lemma 10. Hence, $Q = [S{:}T]'[S{:}T]$ has a noncentral Wishart distribution with k degrees of freedom, covariance matrix I_{m+1}, and matrix of noncentrality parameters $M'M = \Delta_\beta' \lambda \Delta_\beta$, where $\lambda = \mu_\pi' \mu_\pi$. By (10.3.1) of Muirhead (1982), the density of Q at q is as given in part (a) of the lemma.

Part (b) is established as follows. The distribution of Q_T is a noncentral Wishart distribution with k degrees of freedom, covariance matrix I_m, and matrix of noncentrality parameters $\Delta_{T,\beta}' \lambda \Delta_{T,\beta}$ by Lemma 10(b). By (10.3.1) of Muirhead (1982), the density of Q_T at q_T is as given in part (b) of the lemma.

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

$$f_{Q_1|Q_T}(q_1|q_T; \beta_0, \lambda) = K_{1,m} K_{2,m}^{-1} |q|^{(k-m-2)/2} |q_T|^{-(k-m-1)/2} etr(-q_S/2) \times {}_0F_1\left(k/2; \Delta_{\beta_0}' \lambda \Delta_{\beta_0} q/4\right)\right) \left({}_0F_1\left(k/2; \Delta_{T,\beta_0}' \lambda \Delta_{T,\beta_0} q_T/4\right)\right)\right)^{-1}. \quad (14.49)$$

We show below that the conditional distribution of Q_1 given $Q_T = q_T$ does not depend on λ. Hence, we can take $\lambda = 0$ in (14.49). Because ${}_0F_1\ (k/2; 0_{m\times m}) = 1$ for all positive integers m (e.g., see Muirhead (1982) p. 226 for the case $m = 1$ and pp. 227-8 and p. 258 for the case $m \geq 1$), this yields the expression given in part (c) of the lemma.

The conditional distribution of Q_1 given $Q_T = q_T$ does not depend on λ by the following argument. Theorem 11 states that invariant tests are similar if and only if they have *Neyman structure* with respect to Q_T (e.g., as defined in Lehmann (1986, pp. 141-2)). By Theorem 4.2 of Lehmann (1986, p. 144), the latter implies that Q_T is a boundedly complete sufficient statistic under H_0 for the parameter $\lambda > 0$. Sufficiency of Q_T implies the desired result.

An alternative (and more direct) proof that the conditional distribution of Q_1 given $Q_T = q_T$ does not depend on λ is the following: (i) there is a one-to-one transformation from Q_1 to $\widetilde{Q}_1 = (Q_S, S'T_1/||S||, ..., S'T_m/||S||)$, so it suffices to show that the conditional distribution of $\widetilde{Q}_1$ does not depend on λ, (ii) the distribution of $\widetilde{Q}_1$ depends on $T = [T_1 : \cdots : T_m]$ only through $T_j'T_\ell$ for $j, \ell = 1, ..., m$ by the spherical symmetry of the null distribution of S, which is $N(0, I_k)$ by Lemma 10(a), (iii) by (ii) the conditional distribution of $\widetilde{Q}_1$ given $Q_T = T'T$ is the same as the conditional distribution of $\widetilde{Q}_1$ given T, and (iv) the conditional distribution of $\widetilde{Q}_1$ given T is a random function of S only and the null distribution of S is $N(0, I_k)$, which does not depend on λ. $\square$

Footnotes

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

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

[3] 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 $x_2 = Fx_1$.

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

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

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

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

[8] The second condition of Theorem 4 clearly implies (6.10). The converse holds by the completeness of Q_T because by iterated expectations the left-hand side in (6.10) can be written as $E_{\beta_0,\lambda} h(Q_T)$, where $h(Q_T) = E_{\beta_0}(\phi(Q)Q_{ST}|Q_T = q_T)/Q_T^{1/2}$.

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

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

References

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

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

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

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

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

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

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

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

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

King, M.L. (1988), "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169-218.

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

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

Lehmann, E. L. (1986): *Testing Statistical Hypotheses*, second edition. New York: Wiley.

Moreira, M. J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," unpublished manuscript, Department of Economics, University of California, Berkeley.

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

——— (2003b): "A General Theory of Hypothesis Testing in the Simultaneous Equations Model," unpublished manuscript, Department of Economics, Harvard University.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

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

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.