Optimal Invariant Tests in an Instrumental Variables Regression With Heteroskedastic and Autocorrelated Errors

Marcelo J. Moreira
FGV

Geert Ridder
USC

April 26, 2017

Preliminary and Incomplete. Do Not Circulate.

Abstract

This paper uses model symmetries in the instrumental variable (IV) regression to derive an invariant test for the causal structural parameter. Contrary to popular belief, we show there exist model symmetries when equation errors are heteroskedastic and autocorrelated (HAC). Our theory is consistent with existing results for the homoskedastic model (Andrews, Moreira, and Stock (2006) and Chamberlain (2007)), but in general uses information on the structural parameter beyond the Anderson-Rubin, score, and rank statistics. This suggests that tests based only the Anderson-Rubin and score statistics discard information on the causal parameter of interest. We apply our theory to construct designs in which these tests indeed have power arbitrarily close to size. Other tests, including other adaptations to the CLR test, do not suffer the same deficiencies. Finally, we use the model symmetries to propose novel weighted-average power tests for the HAC-IV model.

1 Introduction

We propose novel weighted-average power tests for the structural parameter in a linear regression model with an endogenous regressor and one or more instrumental variables. The errors are heteroskedastic and autocorrelated (HAC). What is novel is our use of model symmetries that, contrary to popular belief, exist if the errors are HAC. To show these model symmetries we consider a simple example.

Consider a model where the random variables $X_i \overset{iid}{\sim} N(\theta, \sigma^2)$. We want to test the null hypothesis $H_0 : \theta = \theta_0$ where $H_1 : \theta \neq \theta_0$, treating σ^2 as a nuisance parameter. If σ^2 is unknown, we can appeal to standard symmetry arguments to find a uniformly most powerful invariant (UMPI) test. The sufficient statistic for (θ, σ^2) is the sample mean $\overline{X}_n$ and the variance estimator $S_X^2 = n^{-1} \sum \left(X_i - \overline{X}_n\right)^2$. With the scale transformation $Y_i = g(X_i - \theta_0)$ we have $Y_i \overset{iid}{\sim} N(g(\theta - \theta_0), g^2\sigma^2)$. For any scalar $g \neq 0$, this transformation preserve the null (and so, the alternative) because the mean of Y_i is zero if and only if we are under the null hypothesis. The transformation induces a transformation in the space of sufficient statistics: $\overline{X}_n$ and S_X^2 become $\overline{Y}_n = g(\overline{X}_n - \theta_0)$ and $S_Y^2 = g^2 S_X^2$, respectively. The maximal invariant is then $\overline{Y}_n^2 / S_Y^2 = \left(\overline{X}_n - \theta_0\right)^2 / S_X^2$. Its distribution only depends on $(\theta - \theta_0)^2 / \sigma^2$ and has a monotone likelihood ratio property. As a result, the UMPI test rejects the null when $\left(\overline{X}_n - \theta_0\right)^2 / S_X^2$ is sufficiently large.

Now, let σ^2 be known. The scale transformation above does not preserve the model if we assume σ^2 to be fixed. How can we use the model symmetries to obtain an optimal invariant test? One possibility is to distinguish the assumption of a known variance from the assumption of a fixed variance. The distinction is whether we actually know σ^2 and treat it as fixed even after we transform the data. If an outsider is telling us the value of σ^2 this person would give a different answer if we asked what the variance is after multiplying the data by a nonzero scalar. The person reports a known, but not fixed, variance. If we take the variance as fixed, we cannot use invariance arguments, but we can still get an optimal test if we restrict ourselves to unbiased tests. Because our canonical model belongs to a one-parameter exponential family, we automatically find that the uniformly most powerful unbiased (UMPU) test rejects the null for large values of $(\overline{X}_n - \theta_0)^2 / \sigma^2$.

In this paper, we assume that the variance is known, but not fixed. In this case, we take the variance σ^2 as both part of the data and a parameter. The sufficient statistic is now the pair $\overline{X}_n$ and σ^2, while the parameters are θ and also σ^2. The same scale transformation as above transforms the sufficient statistic to $\overline{Y}_n = g(\overline{X}_n - \theta_0)$ and $g^2\sigma^2$, and induces a change in the mean from θ to $g(\theta - \theta_0)$ and the variance from σ^2 to $g^2\sigma^2$. The maximal invariant is then $\overline{Y}_n^2 / \sigma^2 = \left(\overline{X}_n - \theta_0\right)^2 / \sigma^2$. This statistic has a noncentral chi-square distribution, where the noncentrality parameter $(\theta - \theta_0)^2 / \sigma^2$ is zero if and only if the null is true. Because this distribution also has a monotonic likelihood ratio property, we again obtain a UMPI test.

In this canonical model, the UMPU and UMPI tests coincide. However, this is not a coincidence: if a UMPU test is unique (up to sets of measure zero) and there exists a UMPI test with respect to some group of transformations, then both coincide (up to sets of measure zero). For the IV model, however, there are no uniformly most powerful tests. Hence, these two approaches do not coincide. In perfect analogy to our canonical model, we introduce two papers in this research agenda. In a companion paper, Moreira and Moreira (2015) seek optimal two-sided tests within a restricted class of tests by fixing a long-run reduced-form variance matrix, i.e., they consider the known and fixed case. In this paper, we instead explore model symmetries by taking the reduced-form variance to be known, but not fixed. As in the canonical model above, we prefer not to take a stance on which thought experiment is more suitable. We consider both approaches to be useful in leading to new insights in the IV model.

If the error variance matrix in the instrumental variable regression is considered known, but not fixed, in the sense discussed above, then the model satisfies some natural symmetries. These symmetries imply that the model is invariant under certain transformations of the data. If these transformations preserve the null hypothesis and if the original data are supportive of the null hypothesis, the transformed data should be equally supportive of this hypothesis. Therefore, the test statistic should be the same if computed from the original or from the transformed data; in other words, the test has to be invariant. The main contribution of this paper is that we propose a test that is invariant for the largest transformation of the data that leaves the model and null hypothesis unchanged. The novel test, denoted conditional invariant likelihood (CIL) test, is an invariant weighted-average power (WAP) test. The weights are derived from relatively invariant measures on the parameter space. The weights of the transformed parameters are then proportional to the weights of the original parameters. The test statistic is the ratio of the integrated likelihoods of the parameter space under the null and alternative. As a result, the invariance of the model combined with the proportional effect of the transformation on the weights makes the CIL test invariant to the transformation, as required.

A second result is that we show theoretically and numerically that the score test has power equal to size in regions of the parameter space *even when the Anderson-Rubin test has power near one*. Other existing tests –including some adaptations of the CLR test– can be interpreted as conditional linear combinations of the Anderson-Rubin and score tests; see Andrews (2016). Hence, tests based on these linear combinations are expected to have power lower than the Anderson-Rubin test for these designs as well. These negative results are due to unexploited information beyond the Anderson-Rubin and score statistics in the HAC-IV model. A different adaptation to the CLR test and a novel WAP test uses all the information and are not expected to have low power problems.

The paper is organized as follows. Section 2.1 introduces the IV model, and proposes a family of invariant similar tests robust to heteroskedastic-autocorrelated errors. Section 3 discusses invariance in the case that the variance matrix has a Kronecker product structure. Section 4 derives the model likelihood, and shows the model symmetries if the errors are HAC. Section 5 proposes invariant similar tests, including weighted-average power and likelihood ratio tests. Section 6 considers the effect of estimation of the long-run variance. Section 7 shows that current tests have power equal to size in certain regions of the parameter space, if the errors are HAC.

2 The IV Model and Statistics

2.1 The HAC-IV model

Consider the instrumental variable model

$$y_1 = y_2\beta + u$$
$$y_2 = Z\pi + v_2,$$

where y_1 and y_2 are $n \times 1$ vectors of observations on two endogenous variables, Z is an $n \times k$ matrix of nonrandom exogenous variables with full column rank, and u and v_2 are $n \times 1$ unobserved disturbance vectors with mean zero. The goal here is to test the null hypothesis $H_0 : \beta = \beta_0$ against the alternative hypothesis $H_1 : \beta \neq \beta_0$, treating π as a nuisance parameter. We do not not include covariates in this model, but they can be handled easily by the usual projection arguments; see Andrews, Moreira, and Stock (2006) (abbreviated as AMS06 in the sequel).

We look at the reduced-form model for $Y = [y_1, y_2]$:

$$Y = Z\pi a' + V, \tag{2.1}$$

where $a = (\beta, 1)'$ and $V = [v_1, v_2] = [u + v_2\beta, v_2]$ is the $n \times 2$ matrix of reduced-form errors.

We allow the errors to be heteroskedastic and autocorrelated. Let $P_1 = Z\,(Z'Z)^{-1/2}$ and let $[P_1, P_2] \in \mathcal{O}_n$, the group of $n \times n$ orthogonal matrices. Pre-multiplying the reduced-form model (2.1) by $[P_1, P_2]'$, we obtain the pair of statistics $P_1'Y$ and $P_2'Y$. In this section, we assume that the vec of $\widetilde{V} = (Z'Z)^{-1/2} Z'V$ is normally distributed with a known variance matrix Σ (this assumption can be relaxed at the cost of asymptotic approximations; see section 6). The statistic $P_2'Y$ is ancillary and we do not have prior knowledge about the correlation structure of V. In consequence, we consider tests based on $R = P_1'Y$:

$$R = \mu a' + \widetilde{V},$$

where $vec(\widetilde{V}) \sim N(0, \Sigma)$ and $\mu = (Z'Z)^{1/2}\pi$. For our testing problem, it is convenient to use the data transformation

$$R_0 = RB_0, \text{ where } B_0 = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 1 \end{pmatrix}, \tag{2.2}$$

so that the mean of the first column of $R_0 = [R_1 : R_2]$ is zero under the null. The distribution of R_0 is

$$R_0 \sim N\left(\mu a_\Delta', \Sigma_0\right),$$

where $a_\Delta' = (\Delta, 1)$, $\Delta = \beta - \beta_0$, and $\Sigma_0 = (B_0' \otimes I_k)\,\Sigma\,(B_0 \otimes I_k)$.

As we will show, the IV model, even with heteroskedastic and autocorrelated errors, satisfies some natural symmetries. These symmetries imply that the model is invariant under certain transformations of the data. If these transformations preserve the null hypothesis, and if the original data are supportive of the null hypothesis then the transformed data should be equally supportive of this hypothesis. Therefore the test statistic should be the same whether it is computed from the original or from the transformed data. In Section 5.3, we choose weights that are invariant to model symmetries. This yields an invariant WAP test that circumvents the undesirably low power of WAP tests based on generic weights.

2.2 Similar Tests

It is convenient to use the one-to-one transformation of R to S, T given by

$$S = \left[(b_0' \otimes I_k)\, \Sigma\, (b_0 \otimes I_k)\right]^{-1/2} (b_0' \otimes I_k)\, vec\,(R) \text{ and} \tag{2.3}$$
$$T = \left[(a_0' \otimes I_k)\, \Sigma^{-1} (a_0 \otimes I_k)\right]^{-1/2} (a_0' \otimes I_k)\, \Sigma^{-1} vec\,(R)\,,$$

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

The statistics S and T are independent, and have the distribution

$$S \sim N\left((\beta - \beta_0)\, C_{\beta_0}\mu, I_k\right) \text{ and } T \sim N\left(D_\beta \mu, I_k\right), \text{ where} \tag{2.4}$$
$$C_{\beta_0} = \left[(b_0' \otimes I_k)\, \Sigma\, (b_0 \otimes I_k)\right]^{-1/2} \text{ and}$$
$$D_\beta = \left[(a_0' \otimes I_k)\, \Sigma^{-1} (a_0 \otimes I_k)\right]^{-1/2} (a_0' \otimes I_k)\, \Sigma^{-1} (a \otimes I_k)$$

under the assumption that the errors are normal with a HAC variance matrix.

Examples of test statistics based on S and T are the Anderson-Rubin (AR), the score or Lagrange multiplier (LM), and the quasi likelihood ratio (QLR) statistics. Anderson and Rubin (1949) propose a pivotal test statistic. In our model the Anderson-Rubin statistic is given by

$$AR = S'S. \tag{2.5}$$

Moreira and Moreira (2015) derive the LM and statistic under the same distributional assumption that we make here. For any full column rank matrix X, define the projection matrices $N_X = X\,(X'X)^{-1} X'$. The two-sided LM statistic is

$$LM = S' N_{C_{\beta_0} D_{\beta_0}^{-1} T} S. \tag{2.6}$$

Kleibergen (2005) adapts the likelihood ratio statistic for homoskedastic errors to HAC errors. The quasi likelihood ratio statistic is

$$QLR = \frac{AR - r\,(T) + \sqrt{(AR - r\,(T))^2 + 4LM \cdot r\,(T)}}{2}, \tag{2.7}$$

where AR and LM are defined in (2.5) and (2.6), and $r\,(T) = T'T$. Andrews and Guggenberger (2014) propose tests that are robust to singularity of the covariance matrix.

Andrews (2016) proposes PI-CLC (plug-in conditional linear combination tests) based on the following combination:

$$CLC = m\,(T)\,.J + (1 - m\,(T))\,.AR,$$

where $J = AR - LM$ and $0 \leq m\,(T) \leq 1$. Theorem 2 of Andrews (2016) shows that this class includes the QLR statistic.

All tests reject the null hypothesis when the test statistics ψ are larger than $\kappa\,(t, \Sigma_0)$, the null $1 - \alpha$ quantile conditional on $T = t$. For example, the conditional test based on the QLR statistic rejects the null when this statistic is larger than its null conditional quantile.

We note that all these statistics depend on S only through the AR and LM statistics. In section 7, we show that there is more information in the statistic S beyond the Anderson-Rubin and score statistics when the covariance matrix does not have a Kronecker product structure. For that reason, we recommend the use of conditional tests based on either a likelihood ratio statistic or a weighted-average power statistic. These tests take advantage of information beyond the Anderson-Rubin and score statistics.

The test statistics that are the main contribution of this paper are introduced here and derived in steps in the rest of this paper. In Appendix A, we show that the likelihood ratio statistic based on R is

$$LR = \max_{a_\Delta} vec(R_0)'\Sigma_0^{-1/2} N_{\Sigma_0^{-1/2}(a_\Delta \otimes I_k)} \Sigma_0^{-1/2} vec(R_0) - T'T, \quad (2.8)$$

where $a_\Delta' = (\Delta, 1)$, and LR is written in terms of the pivotal statistic S and the complete statistic T; see also Moreira and Moreira (2015) and Andrews and Mikusheva (2015).

In practice, the LR statistic involves maximization, which does not have a closed-form solution in general. This makes it difficult to implement the test with conditional critical values. Alternatively, the theory we develop here justifies the use of a specific WAP test.

For $k = 1$, the variance matrix Σ trivially has a Kronecker structure. Hence, AMS06 is directly applicable. In particular, the Anderson-Rubin test is the UMPI test in the just identified model ($k = 1$); see Comment 2 following Corollary 1 of AMS06[1].

For $k > 1$, we recommend a novel WAP test. The IL statistic is

$$IL = \int \left|(a_\Delta' \otimes I_k)\,\Sigma_0^{-1}\,(a_\Delta \otimes I_k)\right|^{-1/2} .e^{-\frac{1}{2}\left[vec(r_0)'\Sigma_0^{-1/2} N_{\Sigma_0^{-1/2}(a_\Delta \otimes I_k)} \Sigma_0^{-1/2} vec(r_0) - T'T\right]} |\Delta|^{k-2}\, d\Delta.$$

The remainder of the paper develops the theory that justifies the use of the conditional LR and IL tests. Hereinafter, we assume that the model is over-identified. This theory explores model symmetries in the case that the variance matrix is known, but not fixed, so that it changes if we transform the data.

3 Kronecker Variance Matrix

We first consider the special case where $\Sigma = \Omega \otimes \Phi$ with Ω a 2×2 and Φ a $k \times k$ matrix. We standardize the determinant of Φ equal to one, as in the homoskedastic model of AMS06. The Kronecker product framework is particularly interesting, for two reasons. First, the S and T statistics in (2.3) simplify to the original statistics of Moreira (2001, 2009) for the homoskedastic model. Second, we show that invariance, taking into consideration a transformation of Ω, yields the same maximal invariant as that obtained by AMS06 under the assumption that Ω is known and fixed. This result is striking as *the AMS06 approach does not hold* for general Σ, but *ours does.*

[1] AMS06's optimality result for invariant tests when $k = 1$ can be seen from the perspective of unbiased tests. Moreira (2001, 2009) shows that the Anderson-Rubin test is UMPU. If there is a UMPI test, then the Anderson-Rubin test must be the one; see Theorem 6.6.1 of Lehmann and Romano (2005).

When $\Sigma = \Omega \otimes \Phi$, the statistics S and T defined in (2.3) simplify to

$$\begin{aligned} S &= \Phi^{-1/2}(Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2} \text{ and} \\ T &= \Phi^{-1/2}(Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}. \end{aligned} \tag{3.9}$$

Their distribution is given by

$$S \sim N\left(c_\beta \Phi^{-1/2}\mu, I_k\right) \text{ and } T \sim N\left(d_\beta \Phi^{-1/2}\mu, I_k\right) \tag{3.10}$$

with $c_\beta = (\beta - \beta_0) \cdot (b_0'\Omega b_0)^{-1/2}$ and $d_\beta = a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{1/2}$. AMS06 derive an upper bound for the power of invariant tests for the special case $\Phi = I_k$ treating Ω as known and fixed. Even if Φ is known, the parameter $\mu_\Phi = \Phi^{-1/2}\mu$ is unknown because μ is unknown. Hence, AMS06's invariance argument applies to the new parameter $\mu_\Phi = \Phi^{-1/2}\mu$. Specifically, let $h_1 \in \mathcal{O}_n$, the group of orthogonal matrices with matrix multiplication as the group operation. The corresponding transformation in the sample space is

$$h_1 \circ [S : T] = h_1.\,[S : T]\,.$$

The associated transformation in the parameter space is

$$h_1 \circ (\beta, \mu_\Phi) = (\beta, h_1.\mu_\Phi)\,.$$

The *maximal invariant* statistic is

$$Q = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix} = \begin{bmatrix} S'S & S'T \\ S'T & T'T \end{bmatrix}. \tag{3.11}$$

That is, any invariant test depends on the data only through Q. The density of Q at q for the parameters β and $\lambda = \mu_\Phi'\mu_\Phi$ is given by

$$\begin{aligned} f_{\beta,\lambda}(q_S, q_{ST}, q_T) = {} & K_0 \exp(-\lambda(c_\beta^2 + d_\beta^2)/2)\,|q|^{(k-3)/2} \\ & \times \exp(-(q_S + q_T)/2)(\lambda\xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}(\sqrt{\lambda\xi_\beta(q)}), \end{aligned}$$

where $K_0^{-1} = 2^{(k+2)/2}pi^{1/2}\Gamma_{(k-1)/2}$, $\Gamma_{(\cdot)}$ is the gamma function, $I_{(k-2)/2}(\cdot)$ denotes the modified Bessel function of the first kind, and

$$\xi_\beta(q) = c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T. \tag{3.12}$$

AMS06 show there also exists a *sign* transformation that preserves the two-sided hypothesis testing problem. Consider the group $\mathcal{O}_1$, which contains only two elements: $h_2 \in \{-1, 1\}$. The group transformation in the sample is

$$h_2 \circ [S : T] = [-S : T]\,,$$

which yields a transformation in the maximal invariant space for h_1:

$$h_2 \circ (Q_S, Q_{ST}, Q_T) = (Q_S, h_2.Q_{ST}, Q_T)\,.$$

The maximal invariant is the vector with components Q_S, Q_{ST}^2, and Q_T. This group yields a transformation in the parameter space. It is convenient to look at the transformed parameters $\left(c_\beta\lambda^{1/2}, d_\beta\lambda^{1/2}\right)$. For $h_2 = -1$, AMS06 show that the transformation is

$$h_2 \circ \left(c_\beta\lambda^{1/2}, d_\beta\lambda^{1/2}\right) = \left(-c_\beta\lambda^{1/2}, d_\beta\lambda^{1/2}\right).$$

The induced transformation for the original parameters is

$$\begin{aligned} h_2 \circ (\beta, \lambda) &= \left(\beta_0 - \frac{d_{\beta_0}(\beta-\beta_0)}{d_{\beta_0} + 2j_{\beta_0}(\beta-\beta_0)}, \lambda\frac{(d_{\beta_0} + 2j_{\beta_0}(\beta-\beta_0))^2}{d_{\beta_0}^2}\right), \text{ where} \\ j_{\beta_0} &= \frac{e_1'\Omega^{-1}a_0}{(a_0'\Omega^{-1}a_0)^{-1/2}} \text{ and } e_1 = (1,0)', \end{aligned} \tag{3.13}$$

for $\beta \neq \beta_{AR}$ defined as

$$\beta_{AR} = \frac{\omega_{11} - \omega_{12}\beta_0}{\omega_{12} - \omega_{22}\beta_0} \tag{3.14}$$

(by the definition of a group, the parameter remains unaltered at $h_2 = 1$). The transformation in (3.13) flips the sign of $\beta - \beta_0$. So the *sign* transformation preserves the two-sided hypothesis testing problem $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, but not the one-sided, e.g., testing $H_0 : \beta \leq \beta_0$ against $H_1 : \beta > \beta_0$.

3.1 Instrument Transformation

If we had not standardized the statistics S and T by pre-multiplying the corresponding statistics S, T in AMS06 by $\Phi^{-1/2}$, then the orthogonal transformation argument of AMS06 would not have worked. The reason is that the distribution of $\Phi^{1/2}S$ and $\Phi^{1/2}T$ (which are the original statistics in AMS06) would have variance Φ. To apply the orthogonal transformation argument of AMS06, Φ has to be known and fixed.[2] The situation changes if the matrix Φ is not fixed, but changes if we transform the data, i.e. Φ is known, but not fixed, and both part of the data and a parameter. For example, take the special case in which Φ is a diagonal matrix. If we were to permute the entries of S and T jointly, perhaps we should allow the permutation of the diagonal entries of Φ as well.

We first introduce the transformations that leave the model unchanged. Let $R_0 = RB_0$ as in (2.2).[3] The distribution of R_0 is given by

$$R_0 \sim N\left(\mu.\left(\Delta, 1\right), \Omega_0 \otimes \Phi\right),$$

where $\Delta = \beta - \beta_0$, $\Sigma_0 = \Omega_0 \otimes \Phi$, and $\Omega_0 = B_0'\Omega B_0$. The multiplication of R by the matrix B_0 leads to a reparameterization that guarantees that the mean of the first column of R_0 is zero under the null hypothesis.

[2]We could look at $g \in Gl_k$ such that $g\Phi g' = \Phi$. This yields $g = \Phi^{1/2}h_1\Phi^{-1/2}$. Alternatively, we could look at the transformed model $R_\Phi = \Phi^{-1/2}R$ and apply the orthogonal transformations.

[3]An alternative parameterization is to express the mean as the product of the k-dimensional vector $\mu\left(1+\Delta^2\right)^{1/2}$ and the row vector $[\sin\vartheta : \cos\vartheta] = [\Delta : 1] / \left(1+\Delta^2\right)^{1/2}$; see Chamberlain (2007). Testing $\Delta = 0$ is the same as testing $\vartheta = 0$.

The data are the realizations (R_0, Σ_0) and the parameters are (Δ, μ, Σ_0). The variance matrix Σ_0 is assumed to be known, but not fixed. Thus, Σ_0 is a parameter and part of the data simultaneously.

Consider the action on the sample space

$$g_1 \circ (R_0, \Omega_0, \Phi) = (g_1 R_0, \Omega_0, g_1 \Phi g_1'), \tag{3.15}$$

where $g_1 \in Sl_k$, the group of all $k \times k$ nonsingular matrices whose determinant is one and with matrix multiplication as the group operation. The special linear group Sl_k is a subgroup of the general linear group $\mathcal{G}l_k$ which contains all invertible matrices.

We note that

$$g_1 R_0 \sim N\left(g_1 \mu . \left(\Delta, 1\right), \Omega_0 \otimes g_1 \Phi g_1'\right),$$

so the corresponding action on the parameter space is

$$g_1 \circ (\Delta, \mu, \Omega_0, \Phi) = (\Delta, g_1 \mu, \Omega_0, g_1 \Phi g_1'). \tag{3.16}$$

We now show that the matrix Q

$$Q = [S : T]' [S : T] = \begin{bmatrix} S'S & S'T \\ S'T & T'T \end{bmatrix},$$

together with Ω_0 itself, is the maximal invariant statistic. That is, any other invariant statistic can be written as a function of (Q, Ω_0). The distribution of the maximal invariant depends only on the concentration parameter

$$\lambda = \mu' \Phi^{-1} \mu,$$

on the parameter of interest β, and on Ω_0 itself.

Theorem 1. *For the group actions in (3.15) and (3.16):*
(i) *The maximal invariant in the sample space is given by* (Q, Ω_0)*;*
(ii) *The maximal invariant in the parameter space is given by* $\left(c_\beta^2 \lambda, c_\beta d_\beta \lambda, d_\beta^2 \lambda, \Omega_0\right)$*.*

Comments: 1. The data $([S : T], \Omega_0, \Phi)$ is a one-to-one transformation from the primitive data (R, Ω_0, Φ). Hence, there is no loss of generality in using the *pivotal* statistic S and the *complete* statistic T instead of using R (or R_0).

2. There is a one-to-one mapping between Ω_0 and Ω. Hence, (Q, Ω) is a maximal invariant as well. We continue to use Ω_0 because it will be useful to find a maximal invariant for the two-sided transformations considered later.

3. The statistic Q is the maximal invariant based on the compact orthogonal group on $[S : T]$, which is a straightforward application of AMS06. We instead allow the much larger, noncompact group of nonsingular matrices with unitary determinant. The data also contains the variance components given by Ω_0 and Φ. Because the group Sl_k is not amenable, the Hunt-Stein theorem is not applicable and we do not necessarily obtain a minimax result. This is in contrast to Chamberlain (2007), who builds on the fact that the orthogonal group is compact.

4. The component Φ completely vanishes as the noncompact group $\mathcal{G}l_k$ acts *transitively* on Φ. Hence, the matrix Φ is not part of the maximal invariant.

3.2 Two-Sided Transformation

Besides the action/transformation g_1, we consider the two-sided transformation in the Kronecker model defined by

$$g_2 \circ (R_0, \Omega_0, \Phi) = (R_0.g_2', g_2.\Omega_0.g_2', \Phi), \tag{3.17}$$

where $g_2 \in \mathcal{G}l_2$, the group of nonsingular 2×2 matrices. We use the transpose of g_2 so that the associated transformation is a *left action*.

The transformation by the matrix

$$g_2' = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

yields a new distribution

$$R_0.g_2' \sim N(\mu.(\Delta.g_{11} + g_{21}, \Delta.g_{12} + g_{22}), (g_2.\Omega_0.g_2') \otimes \Phi). \tag{3.18}$$

The variance matrix in (3.18) matches the transformation in (3.17), hence the model dispersion is preserved. Therefore the action in the parameter space is

$$g_2 \circ (\mu, \Delta, \Omega_0, \Phi) = \left(\mu(\Delta.g_{12} + g_{22}), \frac{\Delta.g_{11} + g_{21}}{\Delta.g_{12} + g_{22}}, g_2.\Omega_0.g_2', \Phi\right). \tag{3.19}$$

The transformed instrument coefficients are $\mu(\Delta.g_{12} + g_{22})$ and the transformed structural parameter is $(\Delta.g_{11} + g_{21}) / (\Delta.g_{12} + g_{22})$. While the model is preserved a.e. (except when $\Delta.g_{12} + g_{22} = 0$), the null hypothesis is not necessarily preserved. The null hypothesis $H_0 : \Delta = 0$ is preserved if and only if $g_{21} = 0$. In this case, the matrix g_2 is an element of $\mathcal{G}_2$, the group of lower (the transposition of g_2' is in the group) triangular matrices. Theorem 2 finds the maximal invariant based on $g_1 \in \mathcal{S}l_k$ and $g_2 \in \mathcal{G}_2$.

Theorem 2. *For the data group actions defined in (3.15) and (3.17) and the parameter actions in (3.16) (3.19), we find*
(i) *The induced group action by g_2 on the space $([S : T], \Omega_0, \Phi)$ is*

$$g_2 \circ ([S : T], \Omega_0, \Phi) = ([sgn(g_{11}).S : sgn(g_{22}).T], g_2\Omega_0 g_2', \Phi);$$

(ii) *The data maximal invariant to $g = (g_1, g_2)$ is*

$$(Q_S, Q_T, Q_{ST}^2);$$

(iii) *The induced group action by g_2 on the parameter functions $(c_\beta, d_\beta, \mu, \Omega_0, \Phi)$ is given by*

$$\begin{aligned} & g_2 \circ (c_\beta \mu, d_\beta \mu, \Omega_0, \Phi) \\ = \; & (sgn(g_{11}) c_\beta \mu, sgn(g_{22}) d_\beta \mu, g_2 \Omega_0 g_2', \Phi); \end{aligned}$$

(iv) *The parameter maximal invariant to $g = (g_1, g_2)$ is*

$$(c_\beta^2 \lambda, d_\beta^2 \lambda, |c_\beta d_\beta| \lambda).$$

Comments: 1. The parameters β and Ω remain unchanged by the action (3.16). Because the parameters c_β and d_β depend only on β and Ω, they are preserved as well. The result now follows trivially because $g_1 \circ (\mu, \Omega, \Phi) = (g_1.\mu, \Omega, g_1 \Phi g_1')$.

2. We note that g_{12} may be different from zero. Hence, the group of transformations is larger than scale multiplication to each entry of the vector $(\Delta, 1)$. In Appendix B, we derive the maximal invariant based on scale transformations, i.e. with $g_{12} = 0$. Tests based on this maximal invariant can behave as one-sided tests. This fact shows how important it is to include the largest possible group.

These actions are defined using the reduced-form matrix Ω. For the homoskedastic model, we could analyze the transformations in the structural-form matrix

$$\Psi = \begin{bmatrix} \sigma_{uu} & \sigma_{u2} \\ \sigma_{u2} & \sigma_{22} \end{bmatrix}.$$

One may wonder if there are actually symmetries in the original model. This turns out to be true, and, in fact, the action in the structural-form variance matrix has a very simple structure.

Proposition 1. *The group action on the reduced-form matrix Ω induces an action on the structural-form matrix Ψ:*

$$\begin{aligned} g_2 \circ (\Delta, \lambda, \Psi) &= \left(\frac{\Delta.g_{11}}{\Delta.g_{12} + g_{22}}, (\Delta.g_{12} + g_{22})^2 \lambda, \Gamma \Psi \Gamma' \right), \text{ where} \\ \Gamma &= \begin{bmatrix} (\Delta.g_{12} + g_{22})^{-1} g_{11} g_{22} & 0 \\ g_{12} & g_{22} \end{bmatrix}. \end{aligned}$$

Comment: Take $\beta_0 = 0$. When $g_{11} = -1$, $g_{12} = 0$, $g_{22} = 1$, we have $g_2 \circ (v_1, v_2) = (-v_1, v_2)$. Therefore, σ_{11} and σ_{22} are preserved while σ_{12} changes sign. Since $\sigma_{12} = \sigma_{u2} + \sigma_{22}\beta$, the new value for the structural-form covariance scalar, $-\sigma_{u2}$, and the new value of the parameter, $-\beta$, is the only transformation that works for any value of σ_{22}.

4 Heteroskedastic and Autocorrelated Errors

We now adapt the group transformation $g = (g_1, g_2)$ to the more general model where the variance matrix does not necessarily have a Kronecker form. The action of $g \in \mathcal{G}l_k \times \mathcal{G}_2$ on the sample space (R_0, Σ_0) is defined as

$$g \circ (R_0, \Sigma_0) = (g_1.R_0.g_2', (g_2 \otimes g_1) \Sigma_0 (g_2' \otimes g_1')) \tag{4.20}$$

(because we introduce a different normalization below, we consider $\mathcal{G}l_k$ instead of $\mathcal{S}l_k$).

The transformed distribution of R_0 is

$$g_1.R_0.g_2' \sim N(g_1.\mu(\Delta, 1) g_2', (g_2 \otimes g_1) \Sigma_0 (g_2' \otimes g_1')),$$

so the induced action on the parameter space is

$$g \circ (\Delta, \mu, \Sigma_0) = \left(\frac{\Delta . g_{11}}{\Delta . g_{12} + g_{22}}, g_1 . \mu \left(\Delta . g_{12} + g_{22} \right), \left(g_2 \otimes g_1 \right) \Sigma_0 \left(g_2' \otimes g_1' \right) \right). \tag{4.21}$$

Recall that the data consist of R_0 and Σ_0, where R_0 has a normal distribution and the distribution of Σ_0 is degenerate. So the density of the data will be the product of two parts. The first part is the normal distribution of R_0, which is absolutely continuous with respect to the Lebesgue measure. The second part is the degenerate distribution of Σ_0 that is absolutely continuous with respect to the counting measure.

The density of $R_0 = [R_1 : R_2]$ evaluated at $r_0 = [r_1 : r_2]$ is given by

$$f_R \left(r_0; \Delta, \mu, \Sigma_0 \right) = \left(2.pi \right)^{-k} . \left| \Sigma_0 \right|^{-1/2} . \exp \left\{ -\frac{1}{2} \left[\begin{array}{c} r_1 - \mu . \Delta \\ r_2 - \mu \end{array} \right]' \Sigma_0^{-1} \left[\begin{array}{c} r_1 - \mu . \Delta \\ r_2 - \mu \end{array} \right] \right\},$$

where $pi = 3.14159...$

As in Proposition 3, we consider the groups of instrument transformations g_1 and two-sided transformations g_2 together, so that we have the joint transformation $g = (g_1, g_2)$ defined above, where we take $g_1 \in \mathcal{G}l_k$ and $g_2 \in \mathcal{G}_2^+$, and their associated transformations $g \circ (\Delta, \mu, \Sigma_0)$ in the parameter space. we consider the group of lower triangular 2×2 matrices with positive diagonal elements.

Basic algebraic manipulations show that

$$f_R \left(g \circ r; g \circ \left(\Delta, \mu, \Sigma_0 \right) \right) = f_R \left(\left[r_1 : r_2 \right]; \Delta, \mu, \Sigma_0 \right) . \left| g_2 \right|^{-k} \left| g_1 \right|^{-2}$$

because

$$\left| \left(g_2 \otimes g_1 \right) \Sigma_0 \left(g_2' \otimes g_1' \right) \right| = |g_2|^{2k} |g_1|^4 |\Sigma_0|.$$

Therefore,

$$f_R \left(r_0; \Delta, \mu, \Sigma_0 \right) = f_R \left(g \circ r_0; g \circ \left(\Delta, \mu, \Sigma_0 \right) \right) \chi \left(g \right),$$

where $\chi(g) = \chi_1(g_1) . \chi_2(g_2)$ for $\chi_1(g_1) = |g_1|^2$ and $\chi_2(g_2) = |g_2|^k$, so that the density of R_0 is invariant with multiplier $\chi(g)$.

Of course, the action $g \in \mathcal{G}l_k \times G_2$ is not *proper*. We can impose $|g_1| = 1$ (in which case $g_1 \in Sl_k$, as in Section 3) so that $\chi_1(g_1) = 1$. Alternatively, we can use another standardization such as $g_{22} = 1$. We will use Haar measure to obtain invariant tests. It is harder to work with the Haar measure for Sl_k than for Gl_k; see Dedić (1990). On the other hand, it is relatively simple to derive the Haar measure for 2×2 lower triangular matrices whose element $(2, 2)$ equals one. For this reason, we prefer to impose a restriction on G_2 (instead of on $\mathcal{G}l_k$, as in Section 3).

For the second part, the data Σ_0 have a distribution that assigns probability one to the value Σ_0 itself. Therefore, the density at some arbitrary matrix value σ_0 is

$$f_\Sigma \left(\sigma_0; \Sigma_0 \right) = P_{\Sigma_0} \left(\Sigma_0 = \sigma_0 \right) = I \left(\sigma_0 = \Sigma_0 \right). \tag{4.22}$$

Using (4.22), we have

$$f_\Sigma (g \circ \sigma_0; g \circ \Sigma_0) = f_\Sigma \left(\left(g_2 \otimes g_1 \right) \sigma_0 \left(g_2' \otimes g_1' \right); \left(g_2 \otimes g_1 \right) \Sigma_0 \left(g_2' \otimes g_1' \right) \right) = f_\Sigma \left(\sigma_0; \Sigma_0 \right)$$

so that this density is invariant with multiplier 1.

The joint likelihood is then given by

$$f(r_0,\sigma_0;\Delta,\mu,\Sigma_0) = f_R(r_0;\Delta,\mu,\Sigma_0) . f_\Sigma(\sigma_0;\Sigma_0),$$

and presents the following symmetries:

$$f(r_0,\sigma_0;\Delta,\mu,\Sigma_0) = f(g\circ(r_0,\sigma_0); g\circ(\Delta,\mu,\Sigma_0)) . \chi(g). \tag{4.23}$$

i.e. the likelihood is invariant with multiplier $\chi(g)$. Because the Lebesgue measure is relatively left invariant for the group g with multiplier $\chi(g)$, the invariance of the likelihood follows directly.

5 Invariant Tests

5.1 Optimal Tests

Our goal in this section is to find optimal tests. Specifically, a test is defined to be a measurable function $\phi(r_0,\sigma_0)$ that is bounded by 0 and 1. For a given outcome, the test rejects the null with probability $\phi(r_0,\sigma_0)$ and accepts the null with probability $1-\phi(r_0,\sigma_0)$, e.g., the Anderson-Rubin test is simply $I(AR > c(k))$ where $I(\cdot)$ is the indicator function. The test is said to be nonrandomized if ϕ takes only values 0 and 1; otherwise, it is called a randomized test. The rejection probability is given by

$$E_{\Delta,\mu,\Sigma_0}\phi(R_0,\Sigma_0) \equiv \int \phi(r_0,\sigma_0) f(r_0,\sigma_0;\Delta,\mu,\Sigma_0)\ dr_0\ \eta(d\sigma_0), \tag{5.24}$$

where η is the counting measure. The rejection probability (5.24) simplifies to

$$\begin{aligned} E_{\Delta,\mu,\Sigma_0}\phi(R_0,\Sigma_0) &= \int \phi(r_0,\sigma_0) f_R(r_0;\Delta,\mu,\Sigma_0) . f_\Sigma(\sigma_0;\Sigma_0)\ dr_0\ \eta(d\sigma_0) \\ &= \int \phi(r_0,\Sigma_0) f_R(r_0;\Delta,\mu,\Sigma_0)\ dr_0. \end{aligned} \tag{5.25}$$

The rejection probability $E_{\Delta,\mu,\Sigma_0}\phi(R_0,\Sigma_0)$ taken as a function of Δ, μ, and Σ_0 gives the power curve for the test ϕ. In particular, $E_{0,\mu,\Sigma_0}\phi(R_0,\Sigma_0)$ gives the null rejection probability.

Let the parameter space for Δ,μ,σ_0 be denoted by Θ, with σ-field the intersection of Θ and sets in $\mathcal{B}^{k+1}\times\{\Sigma_0\}$. Let w be a measure on that σ-field. We average the power curve over the parameter space to obtain the weighted-average power (WAP) with weights that are given by the measure w. By Tonelli's theorem, the weighted average power is

$$E_w\phi(R_0,\Sigma_0) = \int E_{\Delta,\mu,\Sigma_0}\phi(R_0,\Sigma_0)\, dw(\Delta,\mu,\Sigma_0). \tag{5.26}$$

If the weights are such that for $B\times\{\Sigma_0\}$

$$w(B\times\{\Sigma_0\}) = w_R(B) . w_\Sigma(\{\sigma_0\}), \tag{5.27}$$

where $B \in \mathcal{B}^{k+1}$ and $w_{\Sigma}(\{\sigma_0\})$ has unitary mass on $\{\Sigma_0\}$, then

$$E_w \phi(R_0, \Sigma_0) = \int \phi(r_0, \Sigma_0) f_{w_R}(r_0, \Sigma_0) \; dr_0, \tag{5.28}$$

where $f_{w_R}(r_0, \Sigma_0)$ is defined as

$$f_{w_R}(r_0, \Sigma_0) = \int f_R(r_0; \Delta, \mu, \Sigma_0) \; dw_R(\Delta, \mu).$$

We seek optimal similar tests

$$\max_{0 \leq \phi \leq 1} E_{w_R} \phi(R_0, \Sigma_0), \text{ where } E_{0,\mu,\Sigma_0} \phi(R_0, \Sigma_0) = \alpha, \forall \mu. \tag{5.29}$$

The next proposition finds the WAP test.

Proposition 2. *The optimal test in (5.29) rejects the null when*

$$\frac{f_{w_R}(r_0, \Sigma_0)}{f_S(s)} > \kappa(t, \Sigma_0), \tag{5.30}$$

where $f_S(s) = (2pi)^{-k} e^{-s's/2}$ *is the density of the statistic* S *under the null.*

Because T is sufficient for μ under the null we condition on $T = t$. The dependence of the test statistic on t is absorbed in the critical value of the test.

For arbitrary weights w, the WAP similar test is not guaranteed to have overall good power in finite samples. In particular, Moreira and Moreira (2015) show that the power can be near zero for parts of the parameter space. We circumvent this problem by replacing the weights w by invariant weights. This makes the test statistic invariant which avoids that the test has low power for regions in the parameter space.

5.2 Similar Invariant Tests

Invariance of conditional tests follow from the relative invariance of test statistics. We define

Definition 1. *A statistic* ψ *is relatively (left) invariant to* g *with multiplier* χ_1 *if*

$$\psi(g \circ (s, t, \sigma_0)) = \chi_1(g) . \psi(s, t, \sigma_0),$$

for any (s, t, σ_0).

Proposition 3 establishes the invariance of the conditional test if the test statistic is relatively invariant.

Proposition 3. *Suppose that $\psi(S,t,\Sigma_0)$ is a continuous random variable under $H_0 : \Delta = 0$ for every t. Define $\kappa_\psi(t,\Sigma_0)$ to be the $1-\alpha$ quantile of the null distribution of $\psi(S,t,\Sigma_0)$. Then the following hold:*
(i) *The conditional test $\phi(s,t,\Sigma_0)$ that rejects the null when*

$$\psi(s,t,\Sigma_0) > \kappa_\psi(t,\Sigma_0)$$

is similar at level α;
(ii) *If $\psi(g \circ (s,t,\Sigma_0))$ is relatively invariant under $g \in \mathcal{G}l_k \times G_2$ with multiplier χ_1, then $\kappa_\psi(t,\Sigma_0)$ is itself relatively invariant with multiplier χ_1; and*
(iii) The conditional test $\phi(s,t,\Sigma_0)$ is invariant.

Comments: 1. Close inspection of the proof shows that invariance of the conditional quantile does not depend on the group transformation used. It is also applicable to other models as long as there is a sufficient statistic, e.g. here under the null, that is boundedly complete.

2. The comment above explains why the conditional quantile of the LR statistic depends only on $T'T$ in the homoskedastic case. The LR statistic does not depend on Ω_0 at all, and $T'T$ is the maximal invariant to orthogonal transformations $h_1 \circ T = h_1.T$. This is consistent with the results of Moreira (2003) and AMS06, but with no need to use pivotal statistics and independence.

Before we introduce the conditional WAP test we establish that the AR, LM, LR and QLR statistics are g invariant.

Proposition 4. *The AR, LM, LR and QLR statistics are invariant to $g = (g_1, g_2) \in \mathcal{G}l_k \times \mathcal{G}_2$.*

5.3 An Invariant WAP Similar Test

The goal is to obtain a WAP invariant similar test in the over-identified model ($k > 1$). This entails finding weights so that the final test is relatively invariant.

Definition 2. *A measure m is relatively (left) invariant with multiplier χ if*

$$\int F\left(g^{-1} \circ \theta\right) m(d\theta) = \chi_1(g) \int F(\theta)\, m(d\theta)$$

for any real-valued continuous function F with bounded support.

We could apply this result for θ being all the parameters (Δ, μ, Σ_0). However, the parameter Σ_0 is known but changes according to the data transformation. Therefore, it is enough to allow θ to be the parameters (Δ, μ) only.

Lemma 1. *The product measure $|\Delta|^{k-2} d\Delta \times\ d\mu$ is (left) invariant to $g = (g_1, g_2)$ with multiplier $|g_1| .g_{11}^{k-2}$.*

Consider the test (5.30) using the product measure $|\Delta|^{k-2}\, d\Delta \times\, d\mu$ as a weight. The next proposition shows that the conditional test is invariant and can be evaluated with a single (and not multiple) integral.

Theorem 3. *The conditional WAP test based on the test statistic*

$$IL = \int \left|(a'_\Delta \otimes I_k)\, \Sigma_0^{-1}\, (a_\Delta \otimes I_k)\right|^{-1/2} .e^{-\frac{1}{2}\left[vec(r_0)'\Sigma_0^{-1/2} N_{\Sigma_0^{-1/2}(a_\Delta \otimes I_k)} \Sigma_0^{-1/2} vec(r_0) - T'T\right]} |\Delta|^{k-2}\, d\Delta \tag{5.31}$$

is invariant.

Comment: The WAP invariant test uses the non-amenable group $\mathcal{G}l_k$, hence it may not be admissible. This issue is similar to that encountered for the commonly accepted and widely used Hotelling T^2 statistic for testing means of different populations.

6 Unknown Long-Run Variance

An alternative approach is to allow a nondegenerate distribution for the estimator $\widehat{\Sigma}_n$ of the covariance with sample space $\mathbb{S}_{2k}$. Standard results for HAC estimation imply that the data R and $\widehat{\Sigma}_n$ are independent with marginals given by

$$R \sim N\left(\mu. [\beta, 1], \Sigma\right) \text{ and } \sqrt{nh_n}\left(\widehat{\Sigma}_n - \Sigma\right) \sim N\left(0, c.\Sigma\right),$$

where c and h_n depend on the specific kernel we are estimating. Likewise,

$$R_0 \text{ and } \widehat{\Sigma}_{0,n} = \left(B_0 \otimes I_k\right) \widehat{\Sigma}_n \left(B'_0 \otimes I_k\right)$$

are independent with marginals given by

$$R_0 \sim N\left(\mu. (\Delta, 1), \Sigma_0\right) \text{ and } \sqrt{nh_n}\left(\widehat{\Sigma}_{0,n} - \Sigma_0\right) \sim N\left(0, c.\Sigma_0\right).$$

For any $g_1 \in \mathcal{G}l_k$ and $g_2 \in G_2$, the action by $g = (g_1, g_2)$ in the data space is

$$g \circ \left(R_0, \widehat{\Sigma}_{0,n}\right) = \left(g_1.R_0.g'_2, (g_2 \otimes g_1)\, \widehat{\Sigma}_{0,n}\, (g'_2 \otimes h'_1)\right).$$

The associated transformation in the parameter space is given by

$$g \circ (\Delta, \mu, \Sigma_0) = \left(\frac{\Delta.g_{11}}{\Delta.g_{12} + 1}, g_1.\mu\left(\Delta.g_{12} + 1\right), (g_2 \otimes g_1)\, \Sigma_0\, (g_2 \otimes g'_1)\right).$$

In this case, we can replace the variance Σ by the estimator $\hat{\Sigma}$ in the formula of standard frequentist tests. For example, the feasible CLR test would be based on

$$\begin{aligned} LR &= \max_{a_\Delta} vec(R_0)'\widehat{\Sigma}_{0,n}^{-1/2} N_{\widehat{\Sigma}_{0,n}^{-1/2}(a_\Delta \otimes I_k)} \widehat{\Sigma}_{0,n}^{-1/2} vec(R_0) - \widehat{T}'_n \widehat{T}_n, \text{ where} \\ \widehat{T}_n &= \left[(a'_0 \otimes I_k)\, \widehat{\Sigma}_{0,n}^{-1}\, (a_0 \otimes I_k)\right]^{-1/2} (a'_0 \otimes I_k)\, \widehat{\Sigma}_{0,n}^{-1} vec\,(R)\,. \end{aligned} \tag{6.32}$$

As for WAP tests, in principle, we would need to take into consideration that μ changes as the sample size n grows. This change would assure that WAP tests are asymptotically efficient. Define

$$\mu_n = \mu/\sqrt{n},$$

and consider the Lebesgue integral over μ_n. Following the proof of Theorem 3, we find that the WAP test rejects the null when

$$n^{-k/2}. \int \left|(a_\Delta' \otimes I_k)\, \Sigma_0^{-1}\, (a_\Delta \otimes I_k)\right|^{-1/2} .e^{-\frac{1}{2}\left[vec(r_0)'\Sigma_0^{-1/2} N_{\Sigma_0^{-1/2}(a_\Delta \otimes I_k)} \Sigma_0^{-1/2} vec(r_0) - T'T\right]} \left|\Delta\right|^{k-2} d\Delta$$

is larger than its conditional quantile. As the constant $n^{-k/2}$ can be absorbed into the conditional quantile, the WAP invariant test in (5.31) is asymptotically optimal under SIV asymptotics. Unlike the MM1-SU and MM2-SU tests of Moreira and Moreira (2015) (or, typically, most WAP tests), efficiency follows directly because (1) the sample rate is a multiplicative constant in the integral; and (2) it is trivial to get the Laplace transform. For other WAP tests, we truly need to allow for the weight on μ to change with the sample size (see Moreira and Moreira (2015)), for the final conditional WAP test to be asymptotically efficient under SIV asymptotics.

Tests, such as the MM-SU tests or other WAP tests, in general change if μ changes with the sample size. This does not have an effect on the CIL test, because term depending on the rate $1/\sqrt{n}$ is absorbed by the critical value function.

Of course, under nonstandard asymptotic theory (such as WIV asymptotics), the tests are no longer equivalent. Furthermore, the WAP statistics can have very poor power for parts of the parameter space if the weights are not invariant. Hence, our suggestion to use the CIL test (5.31). An advantage of WAP tests over frequentist tests, as the CLR test, is that they do not require likelihood maximization. Instead, we can use standard computational techniques developed for Bayesian statistics (see e.g. Chenozhukov and Hong (2003)).

7 Tests with Low Power for Regions in the Parameter Space

The partitioned inverse of the variance

$$\Sigma_0 = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

is given by

$$\begin{aligned} \Sigma_0^{-1} &= \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}, \text{ where} \\ \Sigma^{11} &= \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)^{-1}, \ \Sigma^{22} = \left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}, \text{ and} \\ \Sigma^{21} &= \left(\Sigma^{12}\right)' = -\Sigma^{22}\Sigma_{21}\Sigma_{11}^{-1} = -\Sigma_{22}^{-1}\Sigma_{21}\Sigma^{11}. \end{aligned}$$

The one-sided LM statistic is

$$LM1 = \frac{S' C_{\beta_0} D_{\beta_0}^{-1} T}{\left(T' D_{\beta_0}^{-1} C_{\beta_0}^2 D_{\beta_0}^{-1} T\right)^{1/2}}.$$

By the definition of S and T, the $LM1$ statististic is given by

$$\frac{\left(\Delta.\Sigma_{11}^{-1/2}\mu + U_S\right)' \Sigma_{11}^{-1/2} \left(\Sigma^{22}\right)^{-1/2} \left(\left(\Sigma^{22}\right)^{1/2} \left(I - \Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu + U_T\right)}{\left(\left(\mu' \left(I - \Delta.\Sigma_{11}^{-1}\Sigma_{12}\right)\left(\Sigma^{22}\right)^{1/2} + U_T\right)\left(\Sigma^{22}\right)^{-1/2} \Sigma_{11}^{-1} \left(\Sigma^{22}\right)^{-1/2} \left(\left(\Sigma^{22}\right)^{1/2} \left(I - \Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu + U_T\right)\right)^{1/2}},$$

where U_S and U_T are independent $N(0, I_k)$.

Under the strong instrumental variable asymptotics,

$$\mu = \left(Z'Z\right)^{1/2}\pi = \sqrt{n}\left(\frac{Z'Z}{n}\right)^{1/2}\pi = \sqrt{n}.\left[m + o_p(1)\right],$$

where $m = E\left(z_i z_i'\right)$ and $\Delta = \delta/\sqrt{n}$. Therefore,

$$\begin{aligned} \Delta.\Sigma_{11}^{-1/2}\mu + U_S &= \frac{\delta}{\sqrt{n}}.\Sigma_{11}^{-1/2}.\sqrt{n}.\left[m + o_p(1)\right] + U_S \\ &= \delta.\Sigma_{11}^{-1/2}.m + U_S + o_p(1) \end{aligned}$$

and

$$\begin{aligned} \frac{\left(I - \Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu + U_T}{\sqrt{n}} &= \left(I - \frac{\delta}{\sqrt{n}}.\Sigma_{21}\Sigma_{11}^{-1}\right).\left[m + o_p(1)\right] + \frac{U_T}{\sqrt{n}} \\ &= m + o_p(1). \end{aligned}$$

Therefore,

$$LM1 = \frac{\left(\delta.\Sigma_{11}^{-1/2}.m + U_S + o_p(1)\right)' \Sigma_{11}^{-1/2} m}{\left(m'\Sigma_{11}^{-1}m\right)^{1/2}} \to_d N\left(\delta.\left(m'\Sigma_{11}^{-1}m\right)^{1/2}, 1\right).$$

This yields the asymptotic efficiency of the two-sided LM test. The AR statistic has the same noncentrality parameter $\delta^2.m'\Sigma_{11}^{-1}m$ as the LM statistic. However, if the model is over-identified, the Anderson-Rubin test is no longer optimal.

Consider now the weak instrumental variable asymptotics, in which Δ is fixed and μ is a constant. The expectation of the numerator of $LM1$ is

$$\Delta.\mu'\Sigma_{11}^{-1}\left(I - \Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu.$$

It is possible that $\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu$ is equal to zero, in which case the numerator becomes $\Delta.\mu'\Sigma_{11}^{-1}\mu$. Because we can make the denominator arbitrarily large, the power of the LM test can be made arbitrarily close to the level α. In the spirit of Kadane (1971), we formalize this intuition by considering sequences of parameters.

By the definition of S and T, the $LM1$ statistic is

$$\frac{\left(\Delta.\Sigma_{11}^{-1/2}\mu+U_S\right)'\left(\Sigma_{11}^{-1/2}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu+\Sigma_{11}^{-1/2}\left(\Sigma^{22}\right)^{-1/2}U_T\right)}{\left(\left(\Sigma_{11}^{-1/2}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu+\Sigma_{11}^{-1/2}\left(\Sigma^{22}\right)^{-1/2}U_T\right)'\left(\Sigma_{11}^{-1/2}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu+\Sigma_{11}^{-1/2}\left(\Sigma^{22}\right)^{-1/2}U_T\right)\right)^{1/2}}.$$

Under the assumptions (i) $\liminf \mu'\left(I-\Delta.\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu>0$, (ii) $\Sigma_{11}^{-1/2}\left(\Sigma^{22}\right)^{-1/2}\to 0$, (iii) $\Sigma_{11}^{-1}\left(\Sigma^{22}\right)^{-1/2}\to 0$, and (iv) $\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\to 0$, we get for a sequence of Σ_0 for which (i)-(iv) hold

$$LM1=\frac{\left(\Delta.\Sigma_{11}^{-1/2}\mu+U_S\right)'\Sigma_{11}^{-1/2}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu}{\left(\mu'\left(I-\Delta.\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu\right)^{1/2}}+o_p\left(1\right),$$

where the first term has distribution

$$N\left(\frac{\Delta.\mu'\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu}{\left(\mu'\left(I-\Delta.\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu\right)^{1/2}},1\right).$$

If the orthogonality condition

$$\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu=0,$$

holds, then the mean of $LM1$ in the limit is bounded by

$$\begin{aligned}\left|\frac{\Delta.\mu'\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu}{\left(\mu'\left(I-\Delta.\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu\right)^{1/2}}\right| &= \frac{\Delta.\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\left(I-\Delta.\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu\right)^{1/2}}\\ &\leq \frac{\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}}.\end{aligned}$$

This bound is uniform in Δ and μ. However, the normal limit above is only pointwise in Δ and μ. The next proposition shows that we can approximate $LM1$ uniformly in Δ if the orthogonality condition holds. The limit has a normal distribution with a mean that is uniformly bounded by a term that does not depend on Δ. The noncentrality parameter of the AR statistic is $\Delta^2.\mu'\Sigma_{11}^{-1}\mu$ and this does depend on Δ.

Proposition 5. *Assume that (i)* $\liminf \mu'\left(I-\Delta.\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu>0$, *(ii)* $\Sigma_{11}^{-1/2}\left(\Sigma^{22}\right)^{-1/2}\to 0$, *and (iii)* $\Sigma_{11}^{-1}\left(\Sigma^{22}\right)^{-1/2}\to 0$. *If the orthogonality condition*

$$\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu=0$$

holds, then:
(a) LM is approximated uniformly by a random variable whose distribution is

$$N\left(\frac{\Delta.\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\mu+\Delta^2.\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}},1\right);\textit{and}$$

(b) The absolute value of the asymptotic mean of LM is uniformly bounded by

$$\frac{\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}}.$$

Comments: 1. Because of the orthogonality condition,

$$\mu'\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\Sigma_{11}^{-1}\left(I-\Delta.\Sigma_{21}\Sigma_{11}^{-1}\right)\mu=\mu'\Sigma_{11}^{-1}\mu+\Delta^{2}.\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu>0.$$

so that (iv) above is satisfied for all Σ_0 sequences.

2.The bound on the mean of $LM1$ does not depend on Δ. Hence, the (one-sided or two-sided) LM statistic will not have power converging to one if Δ moves away from zero.

3. The limit is uniform in Δ, but not uniform in μ. As a matter of fact, the score test is asymptotically efficient.

4. For the design in which the score test has power equal to size, we expect that the current tests will not do better than the AR test. Recall the characterization in Andrews (2016),

$$PICLC=\eta\left(r\left(T\right)\right).AR+\left(1-\eta\left(r\left(T\right)\right)\right).LM.$$

As the LM statistic is close to being ancillary, a convex combination of the AR statistic and an ancillary statistic is expected to yield a statistic with lower power than the AR statistic itself.

5. To our knowledge, de Castro (2015) is the first to point out power losses of conditional tests based only on the Anderson-Rubin and score statistics. His thesis provides designs in which Anderson-Rubin and score-based tests are dominated by tests which use all data information. However, he does not consider the orthogonality condition above or provide theoretical justification for the power losses he encounters.

7.1 Low Power Design

We now present a design where the LM test has an approximate mean that is close to 0. Define J_k to be the anti-diagonal $k\times k$ matrix with all anti-diagonal elements being equal to one. For example, $J_2=(1_2 1_2'-I_2)$ when $k=2$. Now, let $\Sigma_{11}=c_{11}.I_k$, $\Sigma_{12}=c_{12}J_k$, and $\Sigma_{22}=c_{22}.I_k$, so that

$$\begin{aligned}\Sigma^{22}&=\left(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-1}\\&=\left(c_{22}.I_2-c_{12}.J_2.I_2.c_{12}.J_2\right)^{-1}\\&=\left(c_{22}-\left(c_{12}\right)^2\right)^{-1}.I_2.\end{aligned}$$

The constants c_{11}, c_{12}, and c_{22} are chosen so that the matrix Σ_0 is positive definite. Each one of the eigenvalues of Σ_0,

$$\begin{aligned}\varsigma_1&=\frac{c_{11}+c_{22}+\sqrt{\left(c_{11}-c_{22}\right)^2+4.c_{12}^2}}{2}\text{ and}\\\varsigma_2&=\frac{c_{11}+c_{22}-\sqrt{\left(c_{11}-c_{22}\right)^2+4.c_{12}^2}}{2}\end{aligned}$$

appear with multiplicity k. As long as $c_{11}, c_{22} \geq 0$ and $c_{11}.c_{22} \geq c_{12}^2$, the matrix Σ_0 is semi-positive definite.

For $\mu = \lambda^{1/2} e_1$, we have

$$\mu' \Sigma_{11}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \mu = \lambda . e_k' e_1 = 0$$

so that the orthogonality condition holds. We can choose $c_{11} = 1$, $c_{12} \to +\infty$, and $c_{22} = c_{12}^2 + (c_{12})^{-3}$. In this case,

$$\Sigma^{22} = (c_{12})^3 . I_2.$$

We have

$$\begin{aligned} \Sigma_{11}^{-1} \left(\Sigma^{22}\right)^{-1/2} &= \Sigma_{11}^{-1/2} \left(\Sigma^{22}\right)^{-1/2} = \frac{1}{(c_{12})^{3/2}} . I_2 \to 0, \text{ and} \\ \Sigma_{11}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \left(\Sigma^{22}\right)^{-1/2} &= \Sigma_{21} \left(\Sigma^{22}\right)^{-1/2} = \frac{1}{(c_{12})^{1/2}} . I_2 \to 0. \end{aligned}$$

The intuition behind this choice is that we can make $E(S)' C_{\beta_0} D_{\beta_0}^{-1} E(T)$ independent of Δ, and at the same time we have that $E(T)$ is not zero. In fact, if the orthogonality condition holds, then

$$\begin{aligned} E(S)' C_{\beta_0} D_{\beta_0}^{-1} E(T) &= \mu' \Sigma_{11}^{-1} \mu - \Delta \mu' \Sigma_{11}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \mu = \mu' \Sigma_{11}^{-1} \mu \\ E(T) &= \left(\Sigma^{22}\right)^{1/2} \left(I - \Delta . \Sigma_{21} \Sigma_{11}^{-1}\right) \mu = \lambda^{1/2} \left(\Sigma^{22}\right)^{1/2} (e_1 - \Delta . c_{21} . e_2) . \end{aligned}$$

This is not possible if the variance matrix has a Kronecker product structure, because in that case the Σ_{ij}, $i, j = 1, 2$ are proportional to each other.

As $E(T)$ can be quite different from zero, it may be reasonable to look at other invariant statistics of the form

$$\frac{S' A(\Sigma_0) T}{(T' A(\Sigma_0) T)^{1/2}}$$

(where A can depend on Σ_0), that are $N(0,1)$ under the null and have a noncentrality parameter different from zero. The caveat is that they will not be asymptotically efficient. On the other hand, the (conditional or unconditional) likelihood ratio test and the IL test are asymptotically efficient under SIV asymptotics. They are invariant and are not a function of the AR and the (two-sided) LM statistics (conditional on T) only.

7.2 Model Distances

If we first consider only linear transformations $g \circ vec(R)$, then it can be shown to be given by the group transformation by $g \circ R = g_1 R g_2'$. Consider an affine transformation $(A, G) \in R^{2k} \times R^{2k \times 2k}$ applied to $vec(R)$. The mean is

$$A + \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} . \begin{bmatrix} \Delta . \mu \\ \mu \end{bmatrix} = \begin{bmatrix} A_1 + (G_{11} . \Delta + G_{12}) . \mu \\ A_2 + (G_{21} . \Delta + G_{22}) . \mu \end{bmatrix} .$$

Under the null, the first mean needs to be zero for any value of μ. This forces $A_1 = 0$ and $G_{12} = 0$. The two vectors need to be proportional to each other, which forces $A_2 = 0$. So we need to have

$$G_{11} . \Delta . \mu \propto (G_{21} . \Delta + G_{22}) . \mu$$

for any μ. Hence $G_{11} \propto G_{21} \propto G_{22}$. Therefore, we get

$$G = \begin{bmatrix} g_{11}.g_1 & 0 \\ g_{21}.g_1 & g_{22}.g_1 \end{bmatrix}.$$

This gives the $g = (g_1, g_2)$ transformation $g_1 R g_2'$. In vectorial form,

$$vec\,(g_1 R g_2') = (g_2 \otimes g_1)\, R,$$

as we needed to show.

We have

$$g \circ (R_0, \Sigma_0) = (g_1.R_0.g_2', (g_2 \otimes g_1)\, \Sigma_0\, (g_2' \otimes g_1')) .$$

Consider the Kronecker approximation $\Omega_0 \otimes \Phi$ to Σ_0; see Van Loan and Ptsianis (1993) and Golub and Van Loan (1996). The approximation error is $\Sigma_0 - \Omega_0 \otimes \Phi$.

We then get

$$\begin{aligned} & g \circ (R_0, \Sigma_0 - \Omega_0 \otimes \Phi, \Omega_0, \Phi) \\ = \; & (g_1.R_0.g_2', (g_2 \otimes g_1)\,[\Sigma_0 - \Omega_0 \otimes \Phi]\,(g_2' \otimes g_1')\,, g_2 \Omega_0 g_2', g_1 \Phi g_1') . \end{aligned}$$

Recall that

$$g_1 = h_1.g_1^{+} \text{ and } g_2 = h_2.g_2^{+},$$

where h_1 is an orthogonal matrix and h_2 is a diagonal sign matrix.

The first decomposition is the QR decomposition of g_1. So we consider first

$$\begin{aligned} & g^{+} \circ (R_0, \Sigma_0 - \Omega_0 \otimes \Phi, \Omega_0, \Phi) \\ = \; & \left(g_1^{+}.R_0.g_2^{+\prime}, \left(g_2^{+} \otimes g_1^{+}\right)[\Sigma_0 - \Omega_0 \otimes \Phi]\left(g_2^{+\prime} \otimes g_1^{+\prime}\right), g_2^{+} \Omega_0 g_2^{+\prime}, g_1^{+} \Phi g_1^{+\prime}\right) . \end{aligned}$$

Write the LU decomposition for $\Omega_0 = \Omega_0^{1/2} \Omega_0^{1/2\prime}$ and $\Phi = \Phi^{1/2} \Phi^{1/2\prime}$. The maximal invariant $\left(\overline{R}_0, \overline{\Gamma}_0\right)$ is given by

$$\left(\Phi^{-1/2}.R_0.\Omega_0^{-1/2\prime}, \left(\Omega_0^{-1/2} \otimes \Phi^{-1/2}\right)[\Sigma_0 - \Omega_0 \otimes \Phi]\left(\Omega_0^{-1/2\prime} \otimes \Phi^{-1/2\prime}\right)\right).$$

The action given by h_1 yields

$$h_1 \circ \left(\overline{R}_0, \overline{\Gamma}_0\right) = \left(h_1.\overline{R}_0, (I_2 \otimes w_1)\, \overline{\Gamma}_0\, (I_2 \otimes w_1')\right).$$

Recall that each $k \times k$ sub-matrix $\overline{\Gamma}_{ij}$ of $\overline{\Gamma}_0$ admit a spectral decomposition:

$$\overline{\Gamma}_{ij} = h_{ij} \Lambda_{ij} h_{ij}', \text{ where } h_{ij} \in \mathcal{O}_k$$

(of course $h_{12} = h_{21}'$). Let us assume $\overline{\Gamma}_{11}$ is invertible (so that the decomposition is unique). Then, the maximal invariant is

$$h_{11}^{'} \overline{R}_0,\; h_{11}^{'}\; \overline{\Gamma}_{ij} h_{11}, \text{ and } \Lambda_{11}.$$

If $\overline{\Gamma}_{11}$ is not invertible, then the non-uniqueness yields a group that shortens the maximal invariant. In the extreme case $\overline{\Gamma}_{ij} = 0$, then the maximal invariant is $\overline{R}_0^{'} \overline{R}_0$.

Here, we do not yet analyze the effect of the sign transformation matrix h_2. Our main goal is to show that the maximal invariant will be much larger than the Q matrix, as done in the homoskedastic case. As for the maximal invariant in the parameter space, the results are the same with small adjustments, if we replace $\overline{R}_0$ by (Δ, μ). In particular, the vector μ by itself is important through the quantity $w_{11}^{'} \mu$.

References

Andrews, D. W. K., and P. Guggenberger (2014): "Identification- and Singularity-Robust Inference for Moment Condition Models," Working paper, Yale University.

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

——— (2007): "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 139, 116–132.

Andrews, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, 84, 2155–2182.

Andrews, I., and A. Mikusheva (2015): "Conditional Inference with a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612.

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

de Castro, G. R. (2015): "Invariant Tests in an Instrumental Variables Model with Unknown Data Generating Process," Master's thesis, FGV/EPGE.

Dedić, L. (1990): "On Haar Measure on SL(N,R)," *Publications de L'Institut Mathématique*, 47, 56–60.

Golub, G. H., and C. Van Loan (1996): *Matrix Computations*. Cambridge: Cambridge University Press.

Kadane, J. B. (1971): "Comparison of k-Class Estimators When the Disturbances Are Small," *Econometrica*, 39, 723–737.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming That They Are Identified," *Econometrica*, 73, 1103–1123.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Third edn., Springer Series in Statistics.

Mills, B., M. J. Moreira, and L. P. Vilela (2014): "Tests Based on t-Statistics for IV Regression with Weak Instruments," *Journal of Econometrics*, 182, 351–363.

Moreira, H., and M. J. Moreira (2015): "Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors," arXiv:1505.06644.

Moreira, M. J. (2001): "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," *Center for Labor Economics Working Paper Series*, 37, UC Berkeley.

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

——— (2009): “Tests with Correct Size when Instruments Can Be Arbitrarily Weak,” *Journal of Econometrics*, 152, 131–140.

Van Loan, C., and N. Ptsianis (1993): “Approximation with Kronecker Products,” in *Linear Algebra for Large Scale and Real-Time Applications*, ed. by M. Moonen, and G. Golub, pp. 293–314. Springer, Leuven.