Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors

Humberto Moreira
Marcelo J. Moreira

Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors

Humberto Moreira and Marcelo J. Moreira [1]
FGV/EPGE

This version: June 10, 2016

Abstract

This paper considers two-sided tests for the parameter of an endogenous variable in an instrumental variable (IV) model with heteroskedastic and autocorrelated errors. We develop the finite-sample theory of weighted-average power (WAP) tests with normal errors and a known long-run variance. We introduce two weights which are invariant to orthogonal transformations of the instruments; e.g., changing the order in which the instruments appear. While tests using the MM1 weight can be severely biased, optimal tests based on the MM2 weight are naturally two-sided when errors are homoskedastic.

We propose two boundary conditions that yield two-sided tests whether errors are homoskedastic or not. The locally unbiased (LU) condition is related to the power around the null hypothesis and is a weaker requirement than unbiasedness. The strongly unbiased (SU) condition is more restrictive than LU, but the associated WAP tests are easier to implement. Several tests are SU in finite samples or asymptotically, including tests robust to weak IV (such as the Anderson-Rubin, score, conditional quasi-likelihood ratio, and I. Andrews' (2015) PI-CLC tests) and two-sided tests which are optimal when the sample size is large and instruments are strong.

We refer to the WAP-SU tests based on our weights as MM1-SU and MM2-SU tests. Dropping the restrictive assumptions of normality and known variance, the theory is shown to remain valid at the cost of asymptotic approximations. The MM2-SU test is optimal under the strong IV asymptotics, and outperforms other existing tests under the weak IV asymptotics.

1 Introduction

In an instrumental variable (IV) model, researchers often rely on asymptotic approximations when making inference on the structural coefficients. These approximations, however, can be poor when instruments are weakly correlated with the endogenous regressors as explained by Nelson and Startz (1990), Bound, Jaeger, and Baker (1995), Dufour (1997), and Staiger and Stock (1997). The goal is to find reliable econometric methods regardless of how strong the instruments are.

There has been some progress in the IV model with one endogenous variable and k instruments when errors are homoskedastic. Anderson and Rubin (1949) propose a test statistic which has an asymptotic chi-square-k distribution regardless of how weak the instruments are. Moreira (2001, 2009) shows that the Anderson-Rubin statistic is optimal in the just-identified model, but points out potential power gains when there exists more than one instrument. Kleibergen (2002) and Moreira (2002) show that a score (LM) test statistic has a standard chi-square-one distribution whether the instruments are weak or not. Moreira (2003) proposes to replace the critical value number by conditional quantiles of test statistics. These conditional tests are similar by construction, hence have correct size. He applies the conditional method to the likelihood ratio (LR) statistic and the two-sided Wald statistic. Andrews, Moreira, and Stock (2006a) (hereinafter, AMS06) show that the conditional likelihood ratio (CLR) test satisfies natural orthogonal invariance conditions and is nearly optimal. Andrews, Moreira, and Stock (2007) find that conditional Wald (CW) tests, however, have poor behavior and object to their use in empirical work. Mills, Moreira, and Vilela (2014a) show that the bad performance of CW tests is due to the asymmetric distribution of one-sided Wald statistics when instruments are weak. By extending Moreira's (2003) conditional approach, they find approximately unbiased Wald tests whose power is comparable to the CLR test.

While use of the IV model with homoskedastic errors was important to advance the literature on weak identification, the IV model with heteroskedastic and autocorrelated (HAC) errors is considerably more relevant for applied researchers. Some of the theoretical findings for homoskedastic errors are easily extended for more complicated stochastic processes, whereas others are not. Important work by Stock and Wright (2000), Guggenberger and Smith (2005), Kleibergen (2006), Otsu (2006), and Andrews and Mikusheva (2015), among others, extends the tests conceived for the simple homoskedastic IV model to the generalized method of moments (GMM) and generalized empirical likelihood (GEL) frameworks. Their tests are of course applicable to the HAC-IV model, but it is unknown whether these adaptations are optimal. The purpose of this paper is exactly this: to develop a theory of optimal two-sided tests for the HAC-IV model.

We are able to find a statistic that is pivotal and independent of a second statistic, which is sufficient and complete for the instruments' coefficients under the null. We show that the invariance argument of AMS06 for homoskedastic errors is only appli-

cable if a (long-run) variance has a Kronecker product structure. This limitation has profound consequences for the behavior of weighted-average power (WAP) tests. We choose two priors for the structural parameter and the instruments' coefficients and denote the associated test statistics MM1 and MM2. The priors are chosen to illustrate the effect of a poor weight choice on the power of WAP tests. Although priors vanish asymptotically as in the Bernstein-von Mises theorem, the associated tests can behave quite differently in finite samples (or under the weak-instrument asymptotics). When a variance matrix has a Kronecker product structure, both test statistics are orthogonally invariant, but only MM2 satisfies an additional *sign* invariance argument that preserves the two-sided hypothesis testing problem. As a consequence, a WAP similar test based on the MM1 statistic can behave as a one-sided test and have poor power even with homoskedastic errors (this problem is analogous to the conditional Wald tests documented by Andrews, Moreira, and Stock (2007)) while the WAP similar test using the MM2 statistic has overall good power with a Kronecker-product variance matrix. Other weight choices face the same difficulties as the MM1 statistic for the HAC-IV model, including the recently proposed WAP similar test by Olea (2015), denoted ECS (HAC-IV).

When the (long-run) variance matrix does not have a Kronecker product representation and the model is identified, the Anderson-Rubin test (among other equivalent tests) is the uniformly most powerful unbiased test. In the over-identified model, we show theoretically that it is possible to find a weight so that the test is approximately unbiased and admissible. The lack of invariance, however, makes it harder to construct such weights. In practice, we endogeneize this search by imposing in the WAP maximization problem a boundary condition based on the local power around the null hypothesis. This locally unbiased (LU) condition is a weaker requirement than unbiasedness, so it does not rule out admissibility. The WAP-LU tests are found with non-linear algorithms, which makes it difficult to implement them. We then propose a stronger requirement than LU, denoted the strongly unbiased (SU) condition. The resulting class of tests includes several two-sided tests robust to weak IV, including the Anderson-Rubin, score, (pseudo) likelihood ratio tests by Kleibergen (2006) and Andrews and Guggenberger (2014b), and I. Andrews' (2015) PI-CLC tests. Two-sided optimal tests also satisfy the SU condition asymptotically when the sample size is large and instruments are strong. The WAP-SU tests have power close to the WAP-LU tests based on the MM1 and MM2 weights, with the advantage being that the WAP-SU tests are easy to implement with a standard linear programming software package. We refer to the WAP-SU tests based on our weights as MM1-SU and MM2-SU tests.

We follow I. Andrews (2015) and implement numerical simulations based on Yogo (2004). We choose, however, Yogo's (2004) design where the endogenous variable is the real stock return and the instruments are genuinely weak. We find that, as our theory predicts, the WAP similar tests can be quite erratic. In some designs, they behave as usual two-sided tests and have good power. In other designs they behave

as one-sided tests and have power near zero. We do not recommend the MM1 and MM2 similar tests for empirical researchers. The MM2-SU test, however, outperforms other tests (including the MM1-SU test) and when it occasionally has less power than competing tests, the power loss is small. We recommend the use of the MM2-SU test in empirical work. Our asymptotic analysis is quite general and encompasses all WAP similar and WAP-SU tests whose weight does not depend strongly on the sample size.

The remainder of this paper is organized as follows. Section 2 introduces the HAC-IV model and presents the test statistics, including the MM1 and MM2 statistics. Sections 3 and 4 discuss the power maximization problem and the WAP-LU and WAP-SU tests. Section 5 presents power curves and the role of LU and SU conditions in obtaining WAP tests with overall good power. Section 6 develops an asymptotic framework that encompasses the weak IV and strong IV asymptotics. Section 7 revisits the work of I. Andrews (2015) and Yogo (2004) on testing the intertemporal rate of substitution, with one important modification. Section 8 contains concluding remarks. All proofs are given in the appendices.

2 The IV Model and Statistics

Consider the instrumental variable model

$$
\begin{aligned}
y_1 &= y_2\beta + u \\
y_2 &= Z\pi + v_2,
\end{aligned}
$$

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 having full column rank, and u and v_2 are $n \times 1$ unobserved disturbance vectors having 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 we note that can be easily handled by the usual projection arguments; see AMS06.

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 $(Z'Z)^{-1/2} Z'V$ is normally distributed with known variance matrix Σ (this assumption will be relaxed later at the cost of asymptotic approximations). The statistic $P_2'Y$ is ancillary and we do not have previous knowledge about the correlation structure on V. In consequence, we consider

3

tests based on $R = P_1'Y$:

$$R = \mu a' + (Z'Z)^{-1/2} Z'V,$$

where $\mu = (Z'Z)^{1/2} \pi$.

It is convenient to find the one-to-one transformation of R given by the pair

$$
\begin{aligned}
S &= \left[(b_0' \otimes I_k) \Sigma (b_0 \otimes I_k)\right]^{-1/2} (b_0' \otimes I_k) \overline{R} \text{ and} \\
T &= \left[(a_0' \otimes I_k) \Sigma^{-1} (a_0 \otimes I_k)\right]^{-1/2} (a_0' \otimes I_k) \Sigma^{-1} \overline{R},
\end{aligned}
\tag{2.2}
$$

where $\overline{R} = vec\left[(Z'Z)^{-1/2} Z'Y\right]$, $a_0 = (\beta_0, 1)'$ and $b_0 = (1, -\beta_0)'$. The pair S and T have three important properties: *(i)* they are independent; *(ii)* S is pivotal; and *(iii)* T is complete and sufficient for μ under the null. More specifically, the statistics S and T have distribution

$$
\begin{aligned}
S &\sim N\left((\beta - \beta_0) C_{\beta_0}\mu, I_k\right) \text{ and } T \sim N\left(D_\beta\mu, I_k\right), \text{ where} \\
C_{\beta_0} &= \left[(b_0' \otimes I_k) \Sigma (b_0 \otimes I_k)\right]^{-1/2} \text{ and} \\
D_\beta &= \left[(a_0' \otimes I_k) \Sigma^{-1} (a_0 \otimes I_k)\right]^{-1/2} (a_0' \otimes I_k) \Sigma^{-1} (a \otimes I_k).
\end{aligned}
\tag{2.3}
$$

The joint density $f_{\beta,\mu}(s,t)$ is given by

$$
\begin{aligned}
f_{\beta,\mu}(s,t) &= (2pi)^{-k/2} \exp\left(-\frac{\|s - (\beta - \beta_0) C_{\beta_0}\mu\|^2}{2}\right) \times (2pi)^{-k/2} \exp\left(-\frac{\|t - D_\beta\mu\|^2}{2}\right) \\
&= f_{\beta,\mu}^S(s) \times f_{\beta,\mu}^T(t),
\end{aligned}
$$

where $pi = 3.1415...$ and $f_{\beta,\mu}^S(s)$ and $f_{\beta,\mu}^T(t)$ are the marginal densities for S and T.

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 (LR) statistics. Anderson and Rubin (1949) propose to use a pivotal statistic. In our model the Anderson-Rubin statistic is given by

$$AR = S'S. \tag{2.4}$$

In Appendix A, we derive the LM and LR statistics under that the assumption the errors are normal. For any full column rank matrix X, let $N_X = X(X'X)^{-1} X'$ and $M_X = I - N_X$. Then the LM statistic simplifies to

$$LM = S' N_{C_{\beta_0} D_{\beta_0}^{-1} T} S. \tag{2.5}$$

The likelihood ratio statistic is given by

$$LR = \max_a \overline{R}' \Sigma^{-1/2} N_{\Sigma^{-1/2}(a \otimes I_k)} \Sigma^{-1/2} \overline{R} - T'T. \tag{2.6}$$

4

The LR statistic is apparently not a simple function of S and T (which makes it difficult to implement the test coupled with conditional critical values). Kleibergen (2006) instead adapts the formula for the likelihood ratio statistic derived by Moreira (2003) in the homoskedastic IV model to the GMM framework. For the HAC-IV model, this quasi likelihood ratio statistic becomes

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

where AR and LM are defined in (2.4) and (2.5), and $r(T) = T'T$. Andrews and Guggenberger (2014b) use a Kronecker product $\Omega \otimes \Phi$ (where Ω and Φ are positive-definite matrices respectively with dimensions 2×2 and $k \times k$) approximation to the variance Σ; see Van Loan and Ptsianis (1993) for more details on Kronecker product approximations.

We now present two novel WAP statistics based on the weighted-average density

$$h_\Lambda(s,t) = \int f_{\beta,\mu}(s,t) \ d\Lambda(\beta,\mu). \tag{2.8}$$

These weight functions use the Kronecker product $\Omega \otimes \Phi$ approximation to Σ with the Frobenius norm (i.e., the norm of a matrix X is given by $\|X\| = \sqrt{tr(X'X)}$). For the MM1 statistic $h_1(s,t)$, we choose $\Lambda(\beta,\mu)$ to be $N(\beta_0, 1) \times N(0, \sigma^2 \Phi)$. For the MM2 statistic $h_2(s,t)$, we first define the identity $\tan(\theta) \equiv d_{\beta(\theta)}/c_{\beta(\theta)}$, where

$$c_\beta = (\beta - \beta_0) \cdot (b_0' \Omega b_0)^{-1/2} \text{ and } d_\beta = a' \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2}. \tag{2.9}$$

We choose $\Lambda(\beta,\mu)$ so that the prior for θ and μ are $Unif[-pi, pi] \times N\left(0, \|l_{\beta(\theta)}\|^{-2} \zeta \cdot \Phi\right)$, where $l_\beta = (c_\beta, d_\beta)'$.

In Appendix A, we show that the MM1 and MM2 statistics are

$$h_1(s,t) = (2pi)^{-k-1/2} \int |\Psi_{\beta,\sigma^2}|^{-1/2} \exp\left(-\frac{(s',t') \Psi_{\beta,\sigma^2}^{-1} (s',t')' + (\beta - \beta_0)^2}{2}\right) d\beta \tag{2.10}$$

$$h_2(s,t) = (2pi)^{-(k+1)} \int_{-pi}^{pi} \left|\Psi_{\beta(\theta), \|l_{\beta(\theta)}\|^{-2}\zeta}\right|^{-1/2} \exp\left(-\frac{(s',t') \Psi_{\beta(\theta), \|l_{\beta(\theta)}\|^{-2}\zeta}^{-1} (s',t')'}{2}\right) d\theta,$$

where the matrix Ψ_{β,σ^2} is given by

$$\Psi_{\beta,\sigma^2} = I_2 \otimes I_k + \sigma^2 \begin{bmatrix} (\beta - \beta_0)^2 C_{\beta_0} \Phi C_{\beta_0} & (\beta - \beta_0) C_{\beta_0} \Phi D_\beta' \\ (\beta - \beta_0) D_\beta \Phi C_{\beta_0} & D_\beta \Phi D_\beta' \end{bmatrix}. \tag{2.11}$$

5

2.1 Kronecker Variance Matrix

We consider here the special case where $\Sigma = \Omega \otimes \Phi$ exactly. This framework is particularly interesting for two reasons. First, it encompasses the homoskedastic case by taking Φ to be the identity matrix. We will show that the S and T statistics for general error structure simplify to the original statistics of Moreira (2001, 2009) for the homoskedastic model. Second, the model where Σ has a Kronecker product structure enjoys natural invariance properties. Some statistics are invariant but others are not. This has profound consequences for testing procedures based on these statistics. Indeed, typical tests based on noninvariant statistics (such as those using a constant or Moreira's (2003) conditional critical value function) behave as one-sided tests for parts of the parameter space. We will illustrate this problem numerically in Section 5.

When $\Sigma = \Omega \otimes \Phi$, the statistics S and T defined in (2.2) simplify to

$$
\begin{aligned}
S &= \Phi^{-1/2}(Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2} \text{ and} \\
T &= \Phi^{-1/2}(Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}.
\end{aligned}
\tag{2.12}
$$

Their distribution is given by

$$
S \sim N\left(c_\beta\Phi^{-1/2}\mu, I_k\right) \text{ and } T \sim N\left(d_\beta\Phi^{-1/2}\mu, I_k\right).
\tag{2.13}
$$

AMS06 use invariance arguments for the special case $\Phi = I_k$. However, 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 $g \in \mathcal{O}_n$ and consider the transformation in the sample space

$$
g \circ (S, T) = (gS, gT).
$$

The induced transformation in the parameter space is

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

Invariant tests depend on the data only through

$$
Q = \left[\begin{array}{cc} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{array} \right] = \left[\begin{array}{cc} S'S & S'T \\ S'T & T'T \end{array} \right].
\tag{2.14}
$$

The density of Q at q for the parameters β and $\lambda = \pi'(Z'Z)^{1/2}\Phi^{-1}(Z'Z)^{1/2}\pi$ is given by

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

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

$$\xi_\beta(q) = c_\beta^2 q_S + 2c_\beta d_\beta q_{ST} + d_\beta^2 q_T. \tag{2.15}$$

The following proposition shows that the WAP densities $h_1(s,t)$ and $h_2(s,t)$ are invariant when the covariance matrix is a Kronecker product. Indeed, the Kronecker product approximation $\Omega \otimes \Phi$ to Σ in the definition of the weights was chosen exactly to guarantee the test statistics are *orthogonal* invariant.

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: $\bar{g} \in \{-1, 1\}$. The group transformation in the sample is

$$\bar{g} \circ (Q_S, Q_{ST}, Q_T) = (Q_S, \bar{g} \cdot Q_{ST}, Q_T),$$

whose maximal invariant is Q_S, $|Q_{ST}|$, and Q_T. This group yields a transformation in the parameter space. For $\bar{g} = -1$, AMS06 show that this transformation is

$$\bar{g} \circ (\beta, \lambda) = \left(\beta_0 - \frac{d_{\beta_0}(\beta - \beta_0)}{d_{\beta_0} + 2j_{\beta_0}(\beta - \beta_0)}, \lambda \frac{(d_{\beta_0} + 2j_{\beta_0}(\beta - \beta_0))^2}{d_{\beta_0}^2} \right), \text{ where}$$

$$j_{\beta_0} = \frac{e_1' \Omega^{-1} a_0}{(a_0' \Omega^{-1} a_0)^{-1/2}} \text{ and } e_1 = (1, 0)'. \tag{2.16}$$

(by the definition of a group, the parameter remains unaltered at $\bar{g} = 1$). The transformation in (2.16) flips the sign of $\beta - \beta_0$ for $\beta \neq \beta_{AR}$ defined as

$$\beta_{AR} = \frac{\omega_{11} - \omega_{12}\beta_0}{\omega_{12} - \omega_{22}\beta_0} \text{ where } \Omega = [\omega_{i,l}]. \tag{2.17}$$

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 $H_1 : \beta > \beta_0$.

Proposition 1. *The following holds when $\Sigma = \Omega \otimes \Phi$:*
(i) *The weighted-average densities $h_1(s,t)$ and $h_2(s,t)$ are invariant to orthogonal transformations. That is, they depend on the data only through Q; and*
(ii) *The weighted-average density $h_2(s,t)$ is invariant to sign transformations. It depends on the data only through Q_S, $|Q_{ST}|$, and Q_T.*

The MM1 statistic is not *sign* invariant. We can create a weighted-average statistic that is sign invariant by replacing the weight in $h_1 = \int f_{\beta_0, \lambda}(q_S, q_{ST}, q_T) \, d\Lambda_1(\beta, \lambda)$ by

$$\Lambda(\beta, \lambda) = \frac{\Lambda_1(\beta, \lambda) + \Lambda_1(\bar{g} \circ (\beta, \lambda))}{2}, \tag{2.18}$$

for $\bar{g} = -1$. We note that

$$\int f_{\beta,\lambda}(q_S, q_{ST}, q_T) \, d\Lambda(\beta, \lambda) = \int \int f_{\beta,\lambda}(q_S, q_{ST}, q_T) \, d\Lambda_1(\bar{g} \circ (\beta, \lambda)) \, \nu(d\bar{g}),$$

where ν is the Haar probability measure on the group $\mathcal{O}_1$: $\nu(\{1\}) = \nu(\{-1\}) = 1/2$. Because

$$\begin{aligned}
\int f_{\beta,\lambda}(q_S, -q_{ST}, q_T) \, d\Lambda(\beta, \lambda) &= \int f_{(-1)\circ(\beta,\lambda)}(q_S, q_{ST}, q_T) \, d\Lambda(\beta, \lambda) \\
&= \int f_{\beta,\lambda}(q_S, q_{ST}, q_T) \, d\Lambda(\beta, \lambda),
\end{aligned}$$

the weighted-average statistic based on (2.18) only depends on $q_S, |q_{ST}|, q_T$. But the MM2 statistic is already sign invariant for having chosen a clever prior for β and μ. In fact, the MM2 prior was chosen so that the final statistic is sign invariant. Tests based on $h_2(s, t)$ are naturally two-sided tests for the null $H_0 : \beta = \beta_0$ against the alternative $H_1 : \beta \neq \beta_0$ when $\Sigma = \Omega \otimes \Phi$. This important property does not hold for standard tests based on $h_1(s, t)$. The WAP test (denoted ECS-HACIV) proposed recently by Olea (2015) is not *sign* invariant either. Sections 5 and 7 present numerical simulations showing that all these WAP similar tests can behave like one-sided tests for some parameter values. In the next section, we will discuss ways to circumvent this problem whether Σ has a Kronecker product structure or not.

3   Weighted-Average Power Tests

So far, we have only described test statistics. Coupled with critical values, we obtain the test procedures commonly used in the literature. The Anderson-Rubin test rejects the null when $AR > c(k)$, where $c(d)$ is the $1 - \alpha$ quantile of a chi-square distribution with d degrees of freedom. The LM test rejects the null when $LM > c(1)$. The conditional tests reject the null when each test statistic $\psi(S, T) > \kappa(T)$. Each critical value function $\kappa(T)$ is the null conditional quantile of ψ given $T = t$; see Moreira (2003) for details (we omit the dependence of the critical value function on the statistic ψ when there is no ambiguity). For example, the CQLR test rejects the null when the QLR statistic defined in (2.7) is larger than the conditional critical value.

Our goal in this section is to find optimal tests. Specifically, a test is defined to be a measurable function $\phi(s, t)$ that is bounded by 0 and 1. For a given outcome, the test rejects the null with probability $\phi(s, t)$ and accepts the null with probability $1 - \phi(s, t)$, 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 only takes values 0 and 1; otherwise, it is called a randomized test. We note that

$$E_{\beta,\mu}\phi(S, T) \equiv \int \phi(s, t) f_{\beta,\mu}(s, t) \, d(s, t)$$

8

is the probability of rejecting the null when the parameters are β and μ. The object $E_{\beta,\mu}\phi(S,T)$ taken as a function of β and μ gives the power curve for the test ϕ. In particular, $E_{\beta_0,\mu}\phi(S,T)$ gives the null rejection probability. By Tonelli's theorem, we can write

$$E_\Lambda \phi(S,T) = \int E_{\beta,\mu}\phi(s,t)\, d\Lambda(\beta,\mu) = \int \phi(s,t)\, h_\Lambda(s,t)\ d(s,t), \qquad (3.19)$$

where $h_\Lambda(s,t)$ is defined in (2.8). Hence, $E_\Lambda\phi(S,T)$ is the weighted-average power for the measure $\Lambda(\beta,\mu)$.

A natural first step is to find tests that maximize WAP and have size no larger than α. That is,

$$\max_{0\le\phi\le1} E_\Lambda\phi(S,T),\text{ where } E_{\beta_0,\mu}\phi(S,T) \le \alpha, \forall\mu. \qquad (3.20)$$

Since the parameter μ is unknown, finding a WAP test with correct size is nontrivial. The task entails finding a least favorable distribution Λ_0 to construct the WAP test as described in Section 3.8 of Lehmann and Romano (2005). This test rejects the null when the likelihood ratio is large:

$$\frac{h_\Lambda(s,t)}{\int f_{\beta_0,\mu}^T(t)\ d\overline{\Lambda}(\mu)} > \kappa, \qquad (3.21)$$

where $\kappa\cdot\overline{\Lambda}$ is really a Lagrange multiplier in an infinite-dimensional space; see Lemma 3 of Moreira and Moreira (2010) for details[1]. For a parameter μ of small dimension, we can apply numerical algorithms to approximate the WAP test (such as the one by Elliott, Mueller, and Watson (2015) or the linear programming algorithm of Moreira and Moreira (2013)).

The task of finding tests with correct size is simplified if we can find optimal similar tests:

$$\max_{0\le\phi\le1} E_\Lambda\phi(S,T),\text{ where } E_{\beta_0,\mu}\phi(S,T) = \alpha, \forall\mu. \qquad (3.22)$$

Because the statistic T is sufficient and complete under the null, any similar test is conditionally similar (for almost all levels $T=t$). Hence, we can solve

$$\max_{0\le\phi\le1} E_\Lambda\phi(S,t),\text{ where } E_{\beta_0}\phi(S,t) = \alpha.$$

The WAP similar test rejects the null when

$$\frac{h_\Lambda(s,t)}{f_{\beta_0}^S(s)\cdot h_\Lambda^T(t)} > \kappa(t), \qquad (3.23)$$

---

[1] Also available as Lemma 2 in the most recent version, Moreira and Moreira (2013). Both versions are available on Marcelo Moreira's website: http://www.fgv.br/professor/mjmoreira/

where $\kappa(t)$ is a conditional critical value function and $h_\Lambda^T(t) = \int h_\Lambda(s,t)\, ds$. By Tonelli's theorem,

$$
\begin{aligned}
h_\Lambda^T(t) &= \int \int f_{\beta,\mu}(s,t)\, d\Lambda(\beta,\mu)\, ds \\
&= \int \int f_{\beta,\mu}(s,t)\, ds\, d\Lambda(\beta,\mu) \\
&= \int f_{\beta,\mu}^T(t)\, d\Lambda(\beta,\mu).
\end{aligned}
$$

For arbitrary weights Λ, neither the WAP test with correct size nor the WAP similar test is guaranteed to have overall good power in finite samples[2]. Take for a moment the case where $\Sigma = \Omega \otimes \Phi$. The WAP tests based on $h_1(s,t)$ can have very low power for some parameter values. Because the WAP test with correct size and the WAP similar test based on the MM1 weight are not sign invariant, they can actually behave like one-sided tests for parts of the parameter space.

This issue is analogous to the problem with conditional Wald tests found by Andrews, Moreira, and Stock (2007) which leads them to give a very specific recommendation: "*The evident conclusion for applied work is that researchers choosing among these tests (including conditional Wald) should use the CLR test. The strong asymptotic bias and often low power of the conditional Wald tests indicate that they can yield misleading inferences and are not useful, even as robustness checks.*" For our purposes we can of course circumvent this problem by replacing $h_1(s,t)$ by a sign invariant weight given by (2.18) or by the density $h_2(s,t)$. However, this solution relies on model symmetries (i.e., sign invariance) and only works for Kronecker covariance matrices.

On the other hand, Mills, Moreira, and Vilela (2014a) find approximately unbiased Wald tests which have overall good power. Their procedure only works for the model with homoskedastic errors, but it does hint that imposing additional constraints can actually help to obtain optimal tests with overall good power for general Σ.

4 Two-Sided Boundary Conditions

The WAP similar test based on $h_2(s,t)$ is a two-sided test in the homoskedastic case precisely because the sign-group of transformations preserves the two-sided testing problem when $\Sigma = \Omega \otimes \Phi$. More specifically, because this test depends only on Q_S, $|Q_{ST}|$, and Q_T it is locally unbiased; see Corollary 1 of Andrews, Moreira, and Stock (2006b). When errors are autocorrelated and heteroskedastic, however, the

---

[2]As the geneticist and statistician Anthony W. F. Edwards (1992, p. 60) remarks, "It is sometimes said, in defence of the Bayesian concept, that the choice of prior distribution is unimportant in practice, because it hardly influences the posterior distribution at all when there are moderate amounts of data. The less said about this 'defence' the better."

covariance Σ typically does not have a Kronecker product structure. In this case, the WAP similar test (or a WAP test with correct size) based on $h_2(s,t)$ may not have good power for parts of the parameter space. Worse yet, when the covariance matrix lacks Kronecker product structure, there is actually no sign invariance argument to accommodate two-sided testing.

Proposition 2. *Assume that we cannot write Σ as $\Omega \otimes \Phi$ for a 2×2 matrix Ω and a $k \times k$ matrix Φ, both symmetric and positive definite. Then for the data group of transformations $[S, T] \to [\pm S, T]$, there exists no group of transformations in the parameter space which preserves the testing problem.*

Proposition 2 asserts that we cannot simplify the two-sided hypothesis testing problem using sign invariance arguments. It is then much more difficult to find a weight so that the test is, loosely speaking, two-sided. An unbiasedness condition instead adjusts the weights automatically (whether Σ has a Kronecker product or not). Hence, we can seek approximately optimal unbiased tests.

An important property of WAP tests is admissibility. Theorem 1 below shows that the WAP unbiased tests are admissible. The proof follows exactly the same steps as the proof for admissibility of WAP similar tests of Moreira and Moreira (2013) (see Comment 1 after their Theorem 4)[3]. For completeness, we provide a proof in the appendix for the following theorem.

Theorem 1. *Let $(\beta, \mu) \in \mathbb{B} \times \mathbb{P}$, where both sets compact. Assume that the weight Λ appearing in (2.8) has full support on $\mathbb{B} \times \mathbb{P}$. Then there exists a sequence of Bayes' tests $\phi_m(s,t)$ which weakly converges (in the weak* topology to the $\mathcal{L}_\infty(\mathbb{R}^{2k})$ space) to the WAP unbiased test. In particular, the WAP unbiased test is admissible.*

Comments: 1. The weak convergence guarantees, for example, that the limiting power function of $\phi_m(s,t)$ is the power function of the WAP unbiased test. See Moreira and Moreira (2013) for details on weak convergence of tests.

2. The theorem assumes the parameter space is compact. It may be possible to drop this assumption with some additional technical conditions; see Lehmann (1952). The compactness assumption, however, may not be overly restrictive in practice. First, one could argue that we can pin down a region large enough in which the parameter lies. Second, the usual mathematical and statistical software packages have limited numerical accuracy, so for all practical purposes the weight Λ in the average density $h_\Lambda(s,t)$ has support in a compact set.

Proposition 2 shows that there is no sign group structure which preserves the null and alternative. This makes the task of finding a weight function $h_\Lambda(s,t)$ which yields

---

[3]Olea (2015) provides an alternative proof that similar tests are admissible by contradiction.

a WAP unbiased test difficult with HAC errors. Instead of seeking a weight function Λ so that the WAP test is approximately unbiased, we can select an arbitrary weight and find the optimal test among unbiased tests; see Moreira and Moreira (2013). In practice, it would be computationally intensive to handle so many constraints of the form $E_{\beta,\mu}\phi(S,T) \geq E_{\beta_0,\mu_0}\phi(S,T)$ for any scalar β and k-dimensional vectors μ and μ_0, especially when k is large. Instead we choose two different restrictions. The first condition is based on the local power around the null hypothesis. It is a weaker condition than unbiasedness, so it does not rule out admissibility. The second condition is a stronger requirement but is easier to implement. Better yet, numerical simulations will show it yields little power reduction compared to the first condition. Both conditions and their associated WAP tests are presented next.

4.1 Locally Unbiased (LU) Condition

If the test is unbiased, the derivative of the power function must be equal to zero under the null. The next proposition uses this fact and completeness of T to provide a necessary condition for a test to be unbiased. This locally unbiased (LU) condition states that the test must be similar and uncorrelated with linear combinations (which depend on the instruments' coefficient μ) of the pivotal statistic S.

Proposition 3. *A test is said to be locally unbiased (LU) if*

$$E_{\beta_0,\mu}\phi(S,T) = \alpha \ \text{ and } E_{\beta_0,\mu}\phi(S,T) S'C_{\beta_0}\mu = 0, \ \forall\mu. \tag{LU}$$

If a test is unbiased, then it is LU.

In the case $k = 1$ where the model is exactly identified, we have an optimality result for any choice of Λ. The Anderson-Rubin test is the uniformly most powerful unbiased (UMPU) test and has power function depending on the noncentrality parameter $(\beta - \beta_0)^2 C_{\beta_0}^2\mu^2$. We can prove this result directly from Theorem 2-(a) of Moreira (2001, 2009) for homoskedastic errors (with the scalar μ and matrix Ω being replaced by μ_Φ and Σ). As this setup resembles the just-identified model with homoskedastic errors, optimality of the Anderson-Rubin test for HAC errors and $k = 1$ follows straightforwardly.

Proposition 4. *If $k = 1$, the Anderson-Rubin test is the uniformly most powerful unbiased test and has a power function given by*

$$P_{\beta,\mu}(AR > c(1)) = 1 - G\left(c(1); \frac{(\beta - \beta_0)^2\mu^2}{b_0'\Sigma b_0}\right),$$

where $G\left(\cdot;\delta^2\right)$ is the noncentral $\chi^2(1)$ distribution function with noncentrality parameter δ^2. Furthermore, the LM and CQLR tests are equivalent to the Anderson-Rubin test, and are also optimal.

Following Proposition 3, the WAP-LU test solves

$$\max_{0 \le \phi \le 1} E_\Lambda \phi\left(S, T\right), \text{ where } E_{\beta_0, \mu} \phi\left(S, T\right) = \alpha \text{ and } E_{\beta_0, \mu} \phi\left(S, T\right) S' C_{\beta_0} \mu = 0, \forall \mu. \quad (4.24)$$

The optimal tests based on $h_1\left(s, t\right)$ and $h_2\left(s, t\right)$ are denoted respectively MM1-LU and MM2-LU tests. In the just-identified model, the MM1-LU test is shown to be the uniformly most powerful unbiased test. The MM2-LU test is equivalent to the MM2 similar test and is also optimal.

Proposition 5. *The following hold when $k = 1$:*
(a) *The MM2-LU and MM2 similar tests are equivalent and uniformly most powerful unbiased tests.*
(b) *Both MM1-LU and MM2-LU tests are uniformly most powerful unbiased tests.*

Comments: 1. The MM2 similar test automatically satisfies the LU condition when $k = 1$. Hence, the MM2-LU and MM2 similar tests are equivalent when the model is exactly identified.

2. The MM1 similar test is not locally unbiased even when $k = 1$. Close inspection of the weighted density $h_1\left(s, t\right)$ shows that d_β / c_β is the relative contribution of the one-sided $S \cdot T$ statistic to the $AR = S^2$ statistic. If Σ is close to being singular (that is, $|\Sigma|$ is near zero), the ratio d_β / c_β can diverge to infinity. The MM1 test can then behave as a one-sided test. We will illustrate this problem numerically in Section 5.

In the case $k > 1$ where the model is overidentified, we no longer have a uniformly most powerful unbiased test. However, we can still find WAP tests which are locally unbiased. Relaxing both constraints in (4.24) assures us the existence of Lagrange multipliers; see Moreira and Moreira (2013). Therefore, we solve the approximated maximization problem:

$$\max_{0 \le \phi \le 1} E_\Lambda \phi\left(S, T\right), \text{ where } \alpha - \epsilon \quad \le \quad E_{\beta_0, \mu} \phi\left(S, T\right) \le \alpha + \epsilon, \forall \mu \quad (4.25)$$
$$\text{and } E_{\beta_0, \mu_l} \phi\left(S, T\right) S' C_{\beta_0} \mu_l \quad = \quad 0, \text{ for } l = 1, ..., m,$$

when ϵ is small and the number of discretizations m is large. The optimal test rejects the null hypothesis when

$$h_\Lambda\left(s, t\right) - s' C_{\beta_0} \sum_{l=1}^{m} c_l^\epsilon \mu_l f_{\beta_0, \mu_l}\left(s, t\right) > \int f_{\beta_0, \mu}\left(s, t\right) \, d\Lambda_\epsilon\left(\mu\right), \quad (4.26)$$

where the measure Λ_ϵ and the scalars c_l^ϵ, $l = 1, ..., m$, are multipliers associated to boundary constraints in the maximization problem (4.25).

We can use $f_{\beta_0, \mu}\left(s, t\right) = f_{\beta_0}^S\left(s\right) \times f_{\beta_0, \mu}^T\left(t\right)$ to write (4.26) as

$$\frac{h_\Lambda\left(s, t\right)}{f_{\beta_0}^S\left(s\right)} - s' C_{\beta_0} \sum_{l=1}^{m} c_l^\epsilon \mu_l f_{\beta_0, \mu_l}^T\left(t\right) > \int f_{\beta_0, \mu}^T\left(t\right) \, d\Lambda_\epsilon\left(\mu\right). \quad (4.27)$$

13

Letting $\epsilon \downarrow 0$, the optimal test rejects the null hypothesis when

$$\frac{h_\Lambda(s,t)}{f_{\beta_0}^S(s)} - s'C_{\beta_0} \sum_{l=1}^m c_l \mu_l f_{\beta_0,\mu_l}^T(t) > \kappa(t), \tag{4.28}$$

where $\kappa(t)$ is the conditional $1 - \alpha$ quantile of

$$\frac{h_\Lambda(S,t)}{f_{\beta_0}^S(S)} - S'C_{\beta_0} \sum_{l=1}^m c_l \mu_l f_{\beta_0,\mu_l}^T(t). \tag{4.29}$$

This representation is very convenient as we can find

$$\kappa(t) = \lim_{\epsilon \downarrow 0} \int f_{\beta_0,\mu}^T(t) \; d\Lambda_\epsilon(\mu) \tag{4.30}$$

by numerical approximations of the conditional distribution instead of searching for an infinite-dimensional multiplier Λ_ϵ. We then search for the values c_l so that

$$E_{\beta_0,\mu_l} \phi(S,T) S' C_{\beta_0} \mu_l = \int \phi(s,t) s' C_{\beta_0} \mu_l f_{\beta_0}^S(s) f_{\beta_0,\mu_l}^T(t) = 0, \tag{4.31}$$

by taking into consideration that $\kappa(t)$ depends on c_l, $l = 1, ..., m$. We can find c_l, $l = 1, ..., m$ with a nonlinear numerical algorithm[4].

As an alternative procedure, we consider a condition stronger than the LU condition which is simpler to implement numerically. This strategy turns out to be useful because it provides a simple way to implement tests with overall good power. We explain this alternate condition next.

4.2   Strongly Unbiased (SU) Condition

The LU condition asserts that the test ϕ is uncorrelated with a linear combination indexed by the instruments' coefficients μ and the pivotal statistic S. We note that the LU condition trivially holds if

$$E_{\beta_0,\mu} \phi(S,T) = \alpha \text{ and } E_{\beta_0,\mu} \phi(S,T) S = 0, \forall \mu. \tag{SU}$$

That is, the test ϕ is uncorrelated with the k-dimensional statistic S itself under the null. This strongly unbiased (SU) condition states that the test $\phi(S,T)$ is uncorrelated with S for all instruments' coefficients μ. The WAP-SU test based on the weight Λ solves

$$\max_{0 \le \phi \le 1} E_\Lambda \phi(S,T), \text{ where } E_{\beta_0,\mu} \phi(S,T) = \alpha \text{ and } E_{\beta_0,\mu} \phi(S,T) S = 0, \forall \mu. \tag{4.32}$$

---

[4]The two-step procedure just described is the usual *substitution method* for a system of equations, but here we have an uncountable number of equations and unknowns.

The optimal tests based on $h_1(s,t)$ and $h_2(s,t)$ are denoted respectively MM1-SU and MM2-SU tests.

When $k = 1$, the LU and SU conditions are equivalent (hence, the MM1-SU and MM2-SU tests are uniformly most powerful unbiased). When $k > 1$, the following lemma proves the LU condition is strictly weaker than the SU condition. Hence, finding WAP similar tests that satisfy the SU instead of the LU condition in theory may entail unnecessary power losses. In practice, numerical simulations in Section 5 indicate that there is little power gain –if any– by using the LU instead of the SU condition (with the MM1-SU and MM2-SU tests having the advantage of being easier to implement).

Lemma 1. *Define the integral*

$$F_\phi(\mu_1, \mu_2) = E_{\beta_0, D_{\beta_0}^{-1}\mu_2}\phi(s,t)\, s'C_{\beta_0}\mu_1 = \int \phi(s,t)\, s'C_{\beta_0}\mu_1 \cdot f_{\beta_0}^S(s)\, f_{\beta_0, D_{\beta_0}^{-1}\mu_2}^T(t)\ d(s,t).$$

For $k > 1$, there exists a test function $\phi : [S, T] \to [0, 1]$ such that $F_\phi(\mu_1, \mu_1) = 0$ for all μ_1, and $F_\phi(\mu_1, \mu_2) \neq 0$, for some μ_1 and μ_2.

Because the statistic T is complete, we can carry on power maximization in (4.32) for each level of $T = t$:

$$\max_{0 \leq \phi \leq 1} E_\Lambda \phi(S,t), \text{ where } E_{\beta_0}\phi(S,t) = \alpha \text{ and } E_{\beta_0}\phi(S,t)\, S = 0, \qquad (4.33)$$

where the expectation is taken with respect to S only. The WAP-SU test rejects the null when

$$\frac{h_\Lambda(s,t)}{f_{\beta_0}^S(s) \cdot h_\Lambda^T(t)} > \kappa(s,t),$$

where the function $\kappa(s,t) = \overline{\kappa}_0(t) + s'\overline{\kappa}_1(t)$ is such that the optimal test satisfies the SU condition. The term $h_\Lambda^T(t)$ can be absorbed in the critical value function. For numerical stability, however, we recommend keeping it so that the numerator and denominator are of the same order of magnitude.

In practice, we can find $\overline{\kappa}_0(t)$ and $\overline{\kappa}_1(t)$ using linear programming based on simulations for the statistic S. Consider the approximated problem

$$\max_{0 \leq x^{(j)} \leq 1} \quad J^{-1} \sum_{j=1}^{J} x^{(j)} \frac{h_\Lambda\left(s^{(j)}, t\right)}{h_\Lambda^T(t)} \exp\left(s^{(j)\prime} s^{(j)}/2\right) (2pi)^{k/2}$$

$$\text{s.t.} \quad J^{-1} \sum_{j=1}^{J} x^{(j)} = \alpha \text{ and}$$

$$J^{-1} \sum_{j=1}^{J} x^{(j)} s_l^{(j)} = 0, \text{ for } l = 1, ..., k.$$

15

Each j-th draw of S is iid standard-normal:

$$S^{(j)} = \begin{bmatrix} S_1^{(j)} \\ \vdots \\ S_k^{(j)} \end{bmatrix} \sim N\left(0, I_k\right).$$

We note that for the linear programming, the only term which depends on $T = t$ is $h_\Lambda\left(s^{(j)}, t\right)/h_\Lambda^T(t)$. The multipliers for this linear programming problem are the critical value functions $\overline{\kappa}_0(t)$ and $\overline{\kappa}_1(t)$. To speed up the numerical algorithm, we can use the same sample $S^{(j)}$, $j = 1, ..., J$, for every level $T = t$.

Finally, we use the WAP test found in (4.33) to find a useful *two-sided power envelope*. The next proposition finds the optimal test for any given alternative which satisfies the SU condition.

Proposition 6. *The optimal SU test for a point alternative (β, μ) rejects the null hypothesis when*

$$\frac{\left(s' C_{\beta_0} \mu\right)^2}{\mu C_{\beta_0}^2 \mu} > c(1). \tag{4.34}$$

This test is denoted the Point Optimal Strongly Unbiased (POSU) test and has power given by

$$P_{\beta,\mu}\left(\frac{\left(s' C_{\beta_0} \mu\right)^2}{\mu C_{\beta_0}^2 \mu} > c(1)\right) = 1 - G\left(c(1); (\beta - \beta_0)^2 \mu' C_{\beta_0}^2 \mu\right),$$

where $G\left(\cdot; \delta^2\right)$ is the noncentral $\chi^2(1)$ distribution function with noncentrality parameter δ^2.

Comments: 1. The POSU test does not depend on β but does depend on the direction of the vector $C_{\beta_0}\mu$.

2. When $k = 1$, the Anderson-Rubin and POSU tests are the same.

The power plot of $1 - G\left(c(1); (\beta - \beta_0)^2 \mu' C_{\beta_0}^2 \mu\right)$ as β and μ change yields the two-sided power envelope. This power envelope is the two-sided analogue of the one-sided power envelope among similar tests. This power upper bound, based on the Point Optimal Similar (POS) test for the alternative (β, μ), is given by the plot of $1 - \Phi\left(\sqrt{c(1)} - |\beta - \beta_0| \sqrt{\mu C_{\beta_0}^2 \mu}\right)$, where $\Phi(\cdot)$ is the standard normal distribution.

5   Numerical Evaluation of WAP Tests

In this section, we provide numerical simulations for WAP tests based on the MM statistics. The MM tests are WAP similar tests based on $h_1(s, t)$ and $h_2(s, t)$. The

MM-LU and MM-SU tests also satisfy respectively the locally unbiased and strongly unbiased conditions. The goal in this section is to numerically illustrate the importance of using two-sided conditions to obtain tests with overall good power.

We can write

$$\Omega = \begin{bmatrix} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{bmatrix} P_\Omega \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix} P_\Omega' \begin{bmatrix} \omega_{11}^{1/2} & 0 \\ 0 & \omega_{22}^{1/2} \end{bmatrix},$$

where P_Ω is an orthogonal matrix and $\rho = \omega_{12}/\omega_{11}^{1/2}\omega_{22}^{1/2}$. For the numerical simulations, we specify $\omega_{11} = \omega_{22} = 1$.

We use the decomposition of Ω to perform numerical simulations for a class of covariance matrices:

$$\Sigma = P_\Omega \begin{bmatrix} 1+\rho & 0 \\ 0 & 0 \end{bmatrix} P_\Omega' \otimes diag\left(\varsigma_1\right) + P_\Omega \begin{bmatrix} 0 & 0 \\ 0 & 1-\rho \end{bmatrix} P_\Omega' \otimes diag\left(\varsigma_2\right),$$

where ς_1 and ς_2 are k-dimensional vectors.

We consider two possible choices for ς_1 and ς_2. For the first design, we set $\varsigma_1 = \varsigma_2 = (1/\varepsilon - 1, 1, ..., 1)'$. The covariance matrix then simplifies to a Kronecker product: $\Sigma = \Omega \otimes diag\left(\varsigma_1\right)$. For the non-Kronecker design, we set $\varsigma_1 = (1/\varepsilon - 1, 1, ..., 1)'$ and $\varsigma_2 = (1, ..., 1, 1/\varepsilon - 1)'$. This setup captures the data asymmetry in extracting information about the parameter β from each instrument. For small ε, the angle between ς_1 and ς_2 is nearly 90°. We report numerical simulations for $\varepsilon = (k+1)^{-1}$. As k increases, the vector ς_1 becomes orthogonal to ς_2 in the non-Kronecker design.

We set the parameter $\mu = \left(\lambda^{1/2}/\sqrt{k}\right) 1_k$ for $k = 2, 5, 10, 20$ and $\rho = -0.5, 0.2, 0.5, 0.9$. We choose $\lambda/k = 0.5, 1, 2, 4, 8, 16$, which span the range from weak to strong instruments. We focus on tests with significance level 5% for testing $\beta_0 = 0$. To conserve space, we report here only power plots for $k = 5$, $\rho = 0.9$, and $\lambda/k = 2, 8$. The full set of simulations is available on Marcelo Moreira's website.

We present plots for the power envelope and power functions against various alternative values of β and λ. All results reported here are based on 1,000 Monte Carlo simulations. We plot power as a function of the rescaled alternative $(\beta - \beta_0)\lambda^{1/2}$, which reflects the difficulty in making inference on β for different instruments' strength.

Figure 1 reports numerical results for the Kronecker product design. All four pictures present the power envelope and power curves for two existing tests, the Anderson-Rubin (AR) and score (LM) tests.

The first two graphs plot the power curves for the three WAP tests based on the MM1 statistic with $\sigma^2 = 10$. All three tests reject the null when the $h_1(s, t)$ statistic is larger than an adjusted critical value function. In practice, we approximate these critical value functions with 10,000 replications. The MM1 test sets the critical value function to be the 95% empirical quantile of $h_1(S, t)$. The MM1-SU test uses a conditional linear programming algorithm to find its critical value function. The MM1-LU test uses a nonlinear optimization package.

17

Figure 1: Power Comparison (Kronecker Variance)
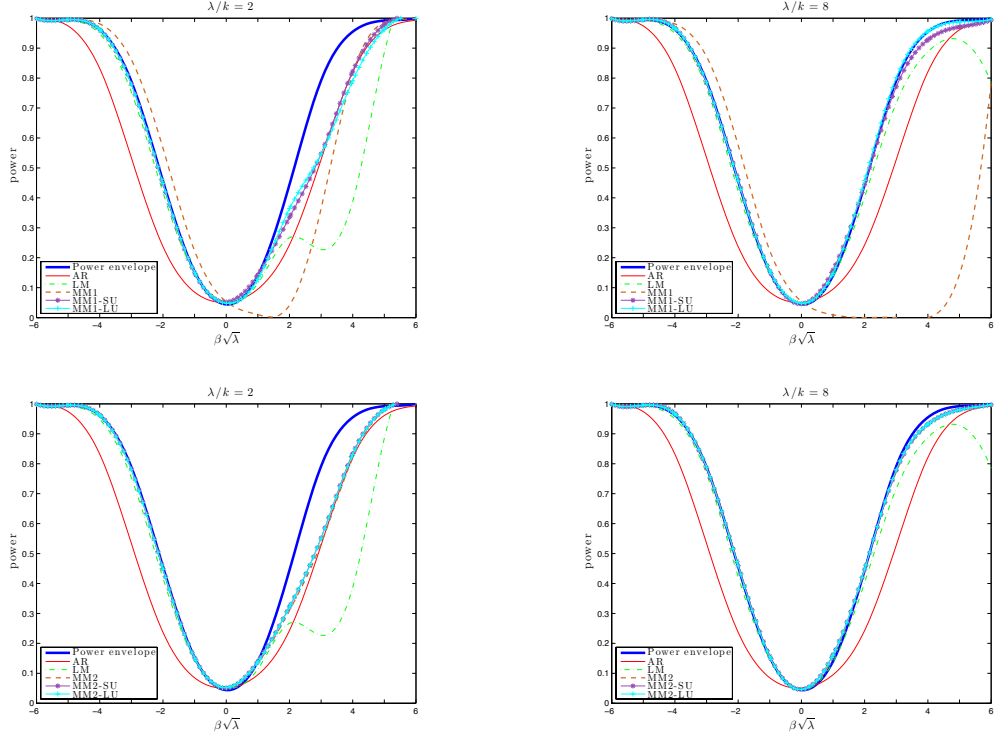
The AR test has power considerably lower than the power envelope when instruments are both weak ($\lambda/k = 2$) and strong ($\lambda/k = 8$). The LM test does not perform well when instruments are weak, and its power function is not monotonic even when instruments are strong. These two facts about the AR and LM tests are well documented in the literature; see Moreira (2003) and AMS06. The figure also reveals some salient findings for the tests based on the MM1 statistic. First, all MM1-based tests have correct size. Second, the MM1 similar test can have large bias to the point that it has zero power for parts of the parameter space. Hence, a naive choice for the density can yield a WAP test which can have overall poor power. We can eliminate this problem by imposing an unbiased condition when selecting an optimal test. The MM1-SU test is easy to implement and has power closer to the power upper bound. When instruments are weak, its power lies moderately below the reported power envelope. This is expected as the number of parameters is too large[5]. When instruments are strong, its power is virtually the same as the power envelope.

To support the use of the MM1-SU test we also consider the MM1-LU test, which imposes a weaker unbiased condition. Close inspection of the graphs show that the derivative of the power function of the MM1 test is different from zero at $\beta = \beta_0$. This

---

[5]The MM1-SU power is nevertheless close to the two-sided power envelope for orthogonally invariant tests as in AMS06 (which is applicable to this design, but not reported here).

observation suggests that the power curve of the WAP test would change considerably if we were to force the power derivative to be zero at $\beta = \beta_0$. Indeed, we implement the MM1-LU test where the locally unbiased condition is true at only one point, the true parameter μ. This parameter is of course unknown to the researcher and this test is not feasible. However, by considering the locally unbiased condition for other values of the instruments' coefficients, the WAP test would be smaller —not larger. The power curves of MM1-LU and MM1-SU tests are very close, which shows that there is not much to be gained by relaxing the strongly unbiased condition.

The last two graphs plot the power curves for the three WAP tests based on the MM2 statistic with $\zeta = 10$. By using the density $h_2(s, t)$, we avoid the pitfalls for the MM1 test. Recall that $h_2(s, t)$ is invariant to those data transformations which preserve the two-sided hypothesis testing problem. Hence, the MM2 similar test is unbiased and has overall good power without imposing any additional unbiased conditions. The graphs illustrate this theoretical finding, as the MM2, MM2-SU, and MM2-LU tests have numerically the same power curves. This conclusion changes dramatically when the covariance matrix is no longer a Kronecker product.

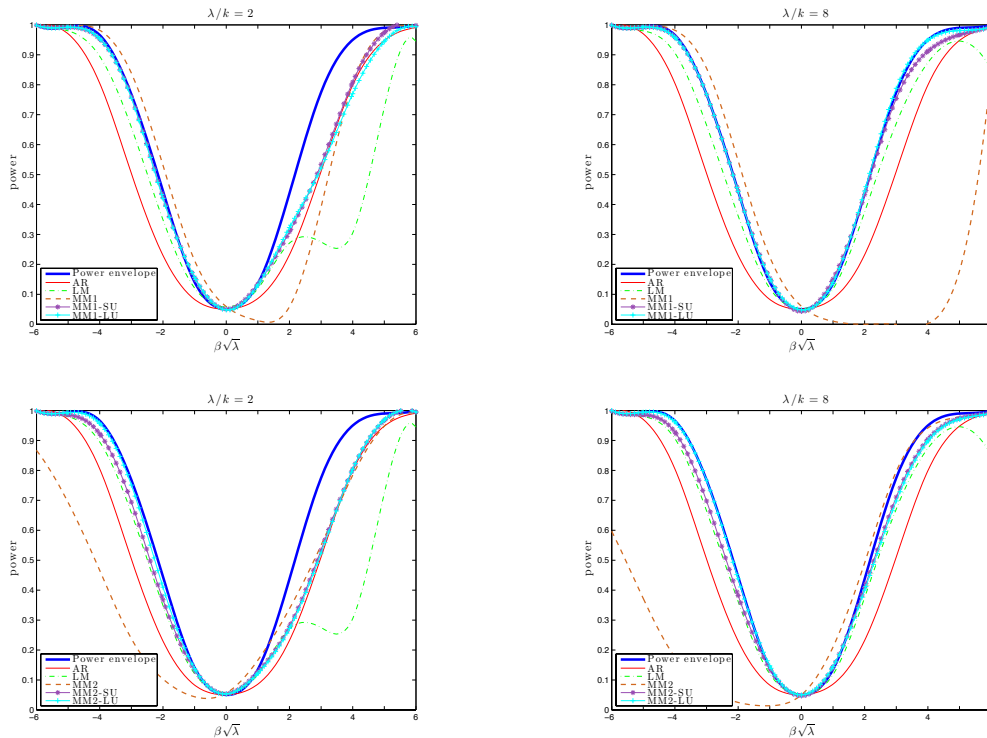Figure 2: Power Comparison (Non-Kronecker Variance)

Figure 2 presents the power curves for all reported tests for the non-Kronecker design. Both MM1 and MM2 tests are severely biased and have overall bad power. For each design, we can make the tests approximately unbiased by choosing the σ^2

and ζ parameters large enough. However, this unbiasedness control is pointwise in the parameter space. We can always find a design such that each test behaves as a one-sided test and has very low power in parts of the parameter space. Hence, the strong asymptotic bias and often-low power of the conditional Wald tests found by Andrews, Moreira, and Stock (2007) also hold for the MM1 (even for the homoskedastic IV model) and MM2 similar tests (only for the HAC-IV model). These WAP similar tests are highly biased with power equal to zero in some parts of the parameter space. Therefore, just as Andrews, Moreira, and Stock (2007) object to the use of conditional Wald tests, we do not recommend the MM1 and MM2 similar tests for empirical researchers.

Proposition 2 shows that we cannot find a group of data transformations which preserve the two-sided testing problem with heteroskedastic-autocorrelated errors. Hence, a choice for the density for the WAP test based on symmetry considerations is not obvious. The correct density choice can be particularly difficult due to the large parameter-dimension (the coefficients μ and covariance Σ). Instead, we can endogenize the weight choice so that the WAP test will be automatically unbiased. This is done by the MM1-LU and MM2-LU tests. These two tests perform as well as the MM1-SU and MM2-SU tests. Because the latter two tests are easy to implement, we recommend their use in empirical practice.

6   Asymptotic Theory

All theoretical and numerical results so far do not rely on the sample size n at all as we have assumed the statistics S and T to be exactly normally distributed with known variance Σ. In this section we relax this assumption at the cost of asymptotic approximations.

Let z_i and v_i denote the i-th row of Z and V, respectively, written as column vectors of dimensions k and 2. We make the following two assumptions as the sample size n grows.

Assumption 1. $n^{-1}Z'Z = n^{-1}\sum_{i=1}^{n} z_i z_i' \to_p D_Z$ for some positive definite $k \times k$ matrix D_Z.

Assumption 2. $n^{-1/2}\sum_{i=1}^{n}(v_i \otimes z_i) \to_d N(0, \Sigma_\infty)$ for some positive definite $2k \times 2k$ matrix Σ_∞.

Assumption 1 holds under Birkhoff's Ergodic Theorem. Assumption 2 holds under suitable conditions by a central limit theorem (CLT). It also assumes that the long-run covariance matrix of Σ_∞ is positive definite, as is usual in the literature. We no longer omit the dependence of Σ on the sample size n and, hereinafter, write Σ_n. Assumption 2 asserts that Σ_∞ is the limit of Σ_n as n grows. Let $\widehat{\Sigma}_n$ be a consistent estimator of Σ_∞ based on $\{(\widehat{v}_i \otimes z_i) : i \leq n\}$, where $\widehat{v}_i$ are reduced-form

residuals. There are many HAC estimators in the literature that can be used for this purpose; see, e.g., Newey and West (1987) and Andrews (1991). For brevity, we do not provide an explicit set of conditions under which one or more of these HAC estimators is consistent; see Jansson (2002) for details. We note, however, that the presence of weak instruments does not complicate standard proofs of the consistency of HAC estimators. Indeed, the convergence for most estimators holds uniformly over all true parameters β and π.

We now introduce feasible versions of S_n and T_n with the variance Σ_n replaced by the estimator $\widehat{\Sigma}_n$:

$$
\begin{aligned}
\widehat{S}_n &= \left[(b_0' \otimes I_k) \, \widehat{\Sigma}_n \, (b_0 \otimes I_k) \right]^{-1/2} (b_0' \otimes I_k) \, \overline{R}_n \text{ and} && (6.35) \\
\widehat{T}_n &= \left[(a_0' \otimes I_k) \, \widehat{\Sigma}_n^{-1} \, (a_0 \otimes I_k) \right]^{-1/2} (a_0' \otimes I_k) \, \widehat{\Sigma}_n^{-1} \overline{R}_n,
\end{aligned}
$$

where $\overline{R}_n = vec \left[(Z'Z)^{-1/2} Z'Y \right]$. Likewise, we define the feasible statistic $\widehat{\psi}_n$ as $\psi(S, T, \Sigma, D_Z)$ with the arguments being replaced by their sample analogues:

$$
\widehat{\psi}_n = \psi(\widehat{S}_n, \widehat{T}_n, \widehat{\Sigma}_n, \widehat{D}_Z), \text{ where } \widehat{D}_Z = n^{-1} Z'Z. \tag{6.36}
$$

Assumption 3. The prior distribution for (β, π) is absolutely continuous to the Lebesgue measure in $\mathbb{R}^{k+1}$. Its density

$$
w(\beta, \pi, \widehat{D}_Z) = w_1(\pi | \beta, \widehat{D}_Z) \cdot w_2(\beta, \widehat{D}_Z)
$$

has full support and is a continuous function of π and β.

Assumption 3 allows the density $w(\beta, \pi, \widehat{D}_Z)$ to depend on the data through $\widehat{D}_Z$. This generalization allows us to cover all tests considered here and asymptotically behaves as $w(\beta, \pi, D_Z)$ (and so we will omit the dependence of the weights on $\widehat{D}_Z$ out of convenience). Although the conditional density $w_1(\pi | \beta)$ does not depend on β for the MM1 tests, it does depend on β for the MM2 tests. Assumption 3 also guarantees that the priors for β and π are not dogmatic and will vanish asymptotically as in the Bernstein-von Mises theorem. If we set the prior on μ, then the associated prior on $\pi = (Z'Z)^{1/2} \mu$ depends on the sample size. For example, the MM statistics introduced in (2.10) use the prior $\mu \sim N(0, \sigma^2 \Phi)$. For the associated prior on $\pi \sim N\left(0, (\sigma^2/n) \widehat{D}_Z^{-1/2} \Phi \widehat{D}_Z^{-1/2}\right)$ not to be sensitive to the sample size, the parameters σ^2 and ζ present in the MM1 and MM2 statistics must eventually grow at the rate n. We make the dependence of $\Lambda(\beta, \mu)$ on the sample size n explicit and, hereinafter, use the notation Λ_n.

We now analyze the asymptotic behavior of the WAP similar and WAP-SU tests. Recall that both of these types of tests depend on the test statistic

$$
\frac{h_{\Lambda_n}(s, t)}{f_{\beta_0}^S(s) \cdot h_{\Lambda_n}^T(t)}. \tag{6.37}
$$

When instruments are weak, the numerator and denominator have the same order of magnitude. When instruments are strong, the integrands in the weighted densities $h_{\Lambda_n}(s,t)$ and $h_{\Lambda_n}^T(t)$ grow exponentially fast and we can apply the Laplace approximation. Because both densities involve $k+1$ integrals, the test statistic in (6.37) is again well-behaved. The caveat is that a simple, closed-form approximation for $h_{\Lambda_n}^T(t)$ does not seem available under strong instruments. The WAP similar and WAP-SU tests, however, remain the same if we standardize (6.37) by any function of t. We replace $h_{\Lambda_n}^T(t)$ by $(1+\|t\|)^{-1} h_{\Lambda_{\beta_0,n}}^T(t)$, where

$$h_{\Lambda_{\beta_0,n}}^T(t) = \int f_{\beta_0,(Z'Z)^{1/2}\pi}^T(t) \, w(\beta_0,\pi) \, d\pi. \tag{6.38}$$

The WAP similar and WAP-SU tests reject the null when

$$WAP = \frac{h_{\Lambda_n}(S,T)}{f_{\beta_0}^S(S) \cdot (1+\|T\|)^{-1} h_{\Lambda_{\beta_0,n}}^T(T)} \tag{6.39}$$

is larger than $\kappa_n(t)$ and $\kappa_n(s,t)$, respectively[6].

Whether the instruments are weak or strong, we are able to obtain an approximation to (6.39). Define

$$
\begin{aligned}
n \cdot Q_n(\beta,\pi) &= \frac{1}{2} \left\| \Sigma^{-1/2}\left(\overline{R} - (a \otimes (Z'Z)^{1/2}\pi)\right)\right\|^2 \\
&= \frac{1}{2} \left\| [S:T] - [(\beta-\beta_0)C_{\beta_0} : D_\beta](I_2 \otimes (Z'Z)^{1/2}\pi)\right\|^2.
\end{aligned}
$$

Lemma **??** in Appendix B shows that the WAP statistic is asymptotically equivalent to

$$\frac{\int \exp\left(-n \cdot Q_n\left(\beta,\pi(\beta)\right)\right) w\left(\beta,\pi(\beta)\right) \left|\left(a' \otimes \widehat{D}_Z^{1/2}\right)\Sigma_n^{-1}\left(a \otimes \widehat{D}_Z^{1/2}\right)\right|^{-1/2} d\beta}{\exp\left(-\frac{S'S}{2}\right)[1+\|T\|]^{-1} w\left(\beta_0,\pi(\beta_0)\right)\left|\left(a_0' \otimes \widehat{D}_Z^{1/2}\right)\Sigma_n^{-1}\left(a_0 \otimes \widehat{D}_Z^{1/2}\right)\right|^{-1/2}}, \tag{6.40}$$

where the constrained maximum likelihood estimator (MLE) for π is

$$
\begin{aligned}
\pi(\beta) &= (Z'Z)^{-1/2}\left[(a' \otimes I_k)\Sigma_n^{-1}(a \otimes I_k)\right]^{-1}(a' \otimes I_k)\Sigma_n^{-1}\overline{R} \text{ and} \tag{6.41} \\
\overline{R} &= \Sigma_n^{1/2}\left[\begin{array}{c} [(b_0' \otimes I_k)\Sigma_n(b_0 \otimes I_k)]^{-1/2}(b_0' \otimes I_k)\Sigma_n^{1/2} \\ [(a_0' \otimes I_k)\Sigma_n^{-1}(a_0 \otimes I_k)]^{-1/2}(a_0' \otimes I_k)\Sigma_n^{-1/2}\end{array}\right]'\left[\begin{array}{c} S \\ T \end{array}\right].
\end{aligned}
$$

---

[6]The use of a Laplace approximation of the ratio of weighted average under the alternative and the null is standard under the usual asymptotics. What is perhaps not standard is the additional term to absorb different rates and unify nonstandard asymptotics. Indeed, if we were to replace $h_\Lambda^T(t)$ only by $h_{\Lambda_{0,n}}^T(t)$, the numerator and denominator in (6.37) would have different orders of magnitude under strong instruments.

The same approximation (6.40) holds for the $\widehat{WAP}$ statistic where we replace S, T, and Σ by their feasible versions given in (6.35). The resulting approximation to the $\widehat{WAP}$ statistic is a function of $\widehat{S}_n$, $\widehat{T}_n$, Σ_n, and $\widehat{D}_Z$. The critical values for the WAP conditional tests and WAP-SU tests, respectively $\kappa_n(t)$ and $\kappa_n(s,t)$, are taken under the assumption that the k-dimensional vector $\widehat{S}_n$ has a standard normal distribution (in practice, these critical values are also functions of the consistent estimators $\widehat{\Sigma}_n$ and $\widehat{D}_Z$ as well, but we omit this dependence out of convenience). For example, for a given weight density $w(\beta, \pi)$, the critical function $\kappa_n(t)$ is simply the $1 - \alpha$ quantile of (6.40) given $T = t$.

We now find the asymptotic distribution for the WAP tests under the WIV asymptotics. We make the following assumption.

Assumption WIV-FA. (a) $\pi = C/n^{1/2}$ for some non-stochastic vector C.
(b) β is a fixed constant for all $n \geq 1$.
(c) k is a fixed positive integer that does not depend on n.

Under WIV, $\pi(\beta)$ is $o_p(1)$ and the WAP statistics behave the same as if the weights were simply $w(\beta, 0)$. As $n \to \infty$, the finite-sample critical value functions $\kappa_n(t)$ and $\kappa_n(s,t)$ respectively converge to their asymptotic counterparts $\kappa_\infty(t)$ and $\kappa_\infty(s,t)$, which are based on (6.40) with $w(\beta, \pi(\beta))$ replaced by $w(\beta, 0)$. We then obtain the following convergence by the continuous mapping theorem and the joint distribution

$$
\begin{bmatrix} S_\infty \\ T_\infty \end{bmatrix} \sim N\left(\begin{bmatrix} (\beta - \beta_0) C_{\beta_0,\infty} \\ D_{\beta_0,\infty} \end{bmatrix} (D_Z)^{1/2} C, I_{2k}\right), \text{ where} \tag{6.42}
$$
$$
C_{\beta_0,\infty} = [(b_0' \otimes I_k) \Sigma_\infty (b_0 \otimes I_k)]^{-1/2} \text{ and}
$$
$$
D_{\beta_0,\infty} = [(a_0' \otimes I_k) \Sigma_\infty^{-1} (a_0 \otimes I_k)]^{-1/2} (a_0' \otimes I_k) \Sigma_\infty^{-1} (a \otimes I_k).
$$

Theorem 2. *Under Assumptions WIV-FA and 1-3:*
(i) $\left(\widehat{S}_n, \widehat{T}_n\right) \to_d (S_\infty, T_\infty)$;
(ii) $P\left(WAP\left(\widehat{S}_n, \widehat{T}_n\right) > \kappa_n\left(\widehat{T}_n\right)\right) \to P\left(WAP\left(S_\infty, T_\infty\right) > \kappa_\infty\left(T_\infty\right)\right)$; *and*
(iii) $P\left(WAP\left(\widehat{S}_n, \widehat{T}_n\right) > \kappa_n\left(\widehat{S}_n, \widehat{T}_n\right)\right) \to P\left(WAP\left(S_\infty, T_\infty\right) > \kappa_\infty\left(S_\infty, T_\infty\right)\right)$.

Both WAP conditional and WAP-SU tests have asymptotic null rejection probabilities being equal to α. The asymptotic power of the WAP tests has a complicated form under WIV asymptotics. We can, of course, rely on numerical simulations to compare their performance with other available tests. In Section 7, we present power plots for testing the intertemporal elasticity of substitution based on the designs of Yogo (2004).

For strong instruments with local alternatives (SIV-LA), we consider the Pitman drift where β is local to the null value β_0 as $n \to \infty$.

Assumption SIV-LA. (a) $\beta = \beta_0 + B/n^{1/2}$ for some constant $B \in \mathbb{R}$.
 (b) π is a fixed non-zero k-vector for all $n \geq 1$.
 (c) k is a fixed positive integer that does not depend on n.

Under the SIV-LA asymptotics, the WAP statistics are shown to be increasing transformations of the LR statistic. This result is general and holds for any prior which satisfies Assumption 3.

Theorem 3. *Suppose Assumptions SIV-LA and 1-3 hold. The long-run variance Σ_∞ is known, or unknown but consistently estimable by $\widehat{\Sigma}_n$. Then the WAP similar and WAP-SU tests are asymptotically equivalent to the LR test given in (2.6).*

Comment. 1. In the proof, we apply the Laplace approximation twice, first with respect to the integral for π and then for β. For the MM1 and MM2 statistics, we can alternatively find a simple expression after integrating out the prior for the instruments' coefficients (expression (9.45) in Appendix A) with σ^2 or ζ growing at rate n and then applying the Laplace approximation for β. Both approaches coincide.
2. The SIV-LA behavior of the ECS (HAC-IV) test appears to be just a special case of our theory using Laplace approximations.
3. For higher-order expansions, we can use Watson's lemma; for references, we recommend Olver (1997) for deterministic functions and Onatski, Moreira, and Hallin (2014a, 2014b) for random functions.
4. Because $T_n/n^{1/2} \to_p D_{\beta_0} D_Z^{1/2} \pi$ under SIV-LA, $\|T_n\|$ diverges to infinity w.p.1 (with probability approaching one). The critical value functions for both the WAP conditional and WAP-SU tests collapse then to the $1 - \alpha$ asymptotic (unconditional) quantile. As a result, the WAP conditional and WAP-SU tests are asymptotically similar and efficient under the SIV asymptotics.

The null rejection probability of WAP tests is α under WIV and SIV asymptotics. Pointwise convergence of the null rejection probability, of course, does not necessarily imply the size is asymptotically α (in a uniform sense). Moreira (2003, p. 1037) suggests to use Parzen (1954) and Andrews (1986) to assure size is uniformly controlled. A series of papers, including Andrews, Cheng, and Guggenberger (2011) and Andrews and Guggenberger (2014a), develop several powerful methods to check uniform size control and have been applied to many econometric models; see Andrews and Guggenberger (2010), Andrews and Guggenberger (2014a), and Mills, Moreira, and Vilela (2014b), among others. Conceivably, we can apply those methods to the WAP statistics coupled with the critical value functions $\kappa_n(t)$ and $\kappa_n(s, t)$. This line of research will be considered in a separate paper.

We can also analyze the WAP tests under strong instruments with fixed alternatives (SIV-FA). We follow Mills, Moreira, and Vilela (2014a) and make the following assumption.

Assumption SIV-FA. (a) $\beta = \beta_0 + B$ for some nonzero $B \in \mathbb{R}$.
   (b) π is a fixed non-zero k-vector for all $n \geq 1$.
   (c) k is a fixed positive integer that does not depend on n.

It is natural to expect that the power converges to one if the parameter β is fixed. However, not all tests have this property even in the IV model with homoskedastic errors; see Andrews, Moreira, and Stock (2004) and Mills, Moreira, and Vilela (2014a) for examples. Hence, it is important to establish consistency for the WAP tests.

If the parameter β is fixed, the WAP statistics are proportional to the exponential of LR. Because LR/n converges to a non-zero constant, the WAP tests are consistent. The next theorem formalizes this result.

Theorem 4. *Suppose Assumptions SIV-FA and 1-3 hold. The long-run variance Σ_∞ is known, or unknown but consistently estimable by $\widehat{\Sigma}_n$. Then the following hold:*
(i) $2.\left(\log \widehat{WAP} \right)/n = \widehat{LR}/n + o_p\left(1\right)$; *and*
(ii) $\widehat{LR}/n = LR/n + o_p\left(1\right) \to \gamma > 0.$

Comment: If $D_\beta \neq 0$, the functions $\kappa_n\left(t\right)$ and $\kappa_n\left(s,t\right)$ converge to a constant obtained under SIV-FA. If $D_\beta = 0$, the critical functions do not converge. However, they are bounded, and so WAP tests are consistent.

7   Power Comparison

In this section, we follow I. Andrews (2015) who calibrates designs for power comparison based on the work of Yogo (2004) on the elasticity of intertemporal substitution in eleven developed countries.

Yogo (2004) tests the effect of interest rates on the level of aggregate demand in an IV model. He considers a linear regression in which asset return affects consumption growth, and the reverse form of this regression. In both equations, the endogenous variable (consumption or asset return) can be correlated with the error (innovation). To remedy this problem, he chooses four instruments: lagged values of nominal interest rate, inflation, consumption growth, and log dividend-price ratio.

I. Andrews (2015) selects the real interest rate (rf in Yogo's (2004) notation) as the endogenous variable. Several tests perform well in his design, including MM2-SU, PI-CLC, and (WAP similar) ECS tests. In fact, only in a few countries do these tests have slightly different performance; see Section 7.2.1 of I. Andrews (2015). The difficulty in assessing the relative performance of each test arises because the instruments are not

particularly weak in this design. Indeed, the first-stage F-statistic reported by Yogo (2004) (see his Table I) is below 10 in only four countries (Japan, Switzerland, United Kingdom, and the United States). We instead join de Castro (2015) in choosing the real stock return (*re* in Yogo's (2004) notation) as the endogenous variable. The instruments are considerably weaker in this design: the F-statistic is smaller than 4.18 in all countries, and always less than the F-statistic for interest rate. Our decision to use stock returns aims to highlight the differences between the tests proposed for the HAC-IV model. Apart from using stock returns instead of interest rates, our design is akin to that of I. Andrews (2015). We use the Newey-West estimator with three lags, and the resulting power curves are based on 5,000 Monte Carlo simulations. In parallel to our asymptotic theory, we choose the ratio of the tuning parameters σ^2 and ζ to the sample size to be one-tenth for the MM1 and MM2 statistics, respectively.

Figure 3 plots power curves for the two-sided power envelope, Anderson-Rubin (AR), score (LM), WAP similar MM1, WAP similar MM2, and ECS (HAC-IV) tests. Although the AR and LM tests are unbiased, the MM1, MM2, and ECS tests perform unreliably. To illustrate the problem, we mention three countries. For Australia, the MM1 and ECS tests have low power for parts of the parameter space, while the MM2 test behaves more like a two-sided test. For France, the ECS test performs well, while both MM1 and MM2 tests can have low power. For the USA, the ECS test has power near zero and behaves more as a one-sided test while the MM1 and MM2 tests are nearly unbiased. In some countries, these three tests have power even lower than the Anderson-Rubin test (e.g., the ECS test for Germany and Italy).

Figure 3: Power Comparison (WAP similar tests)
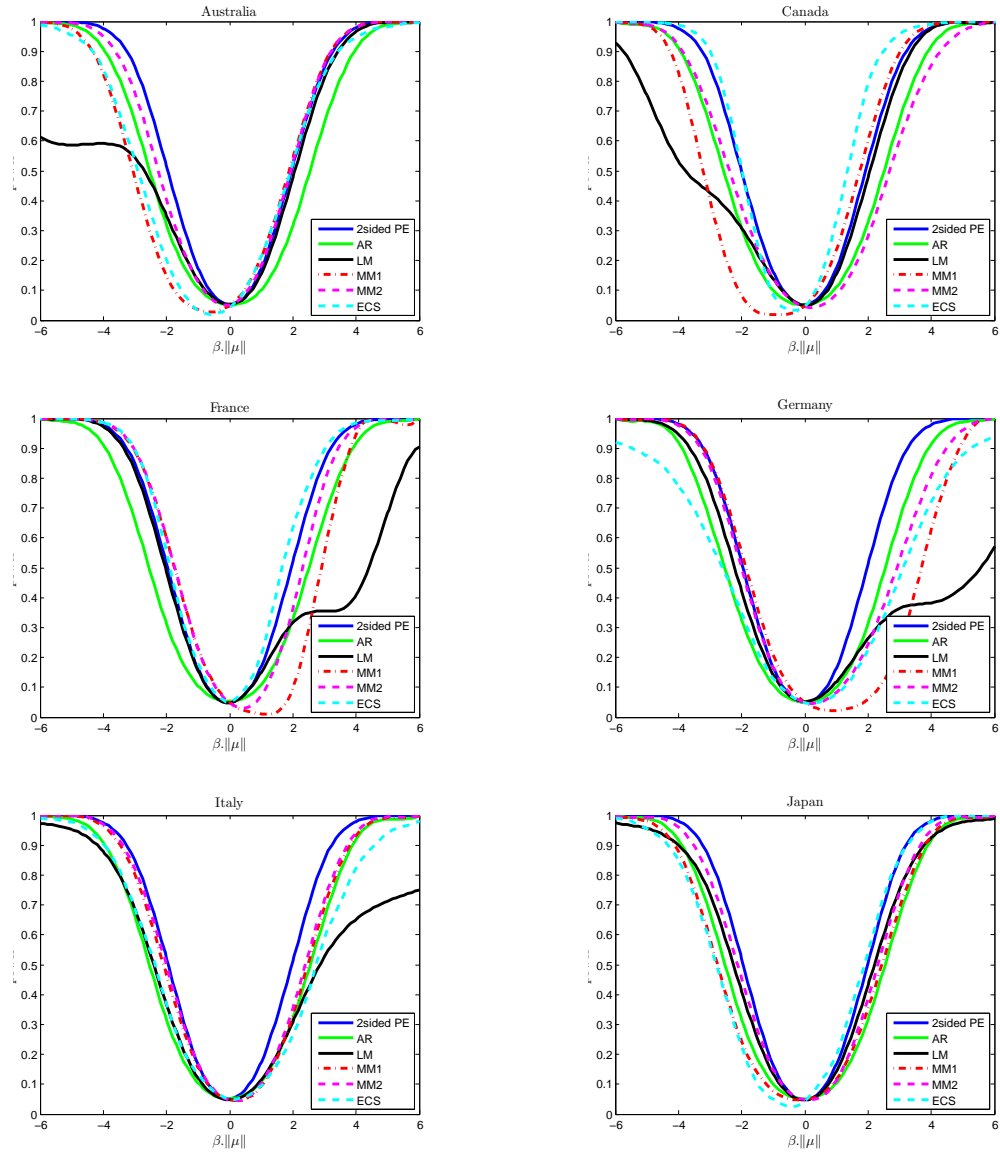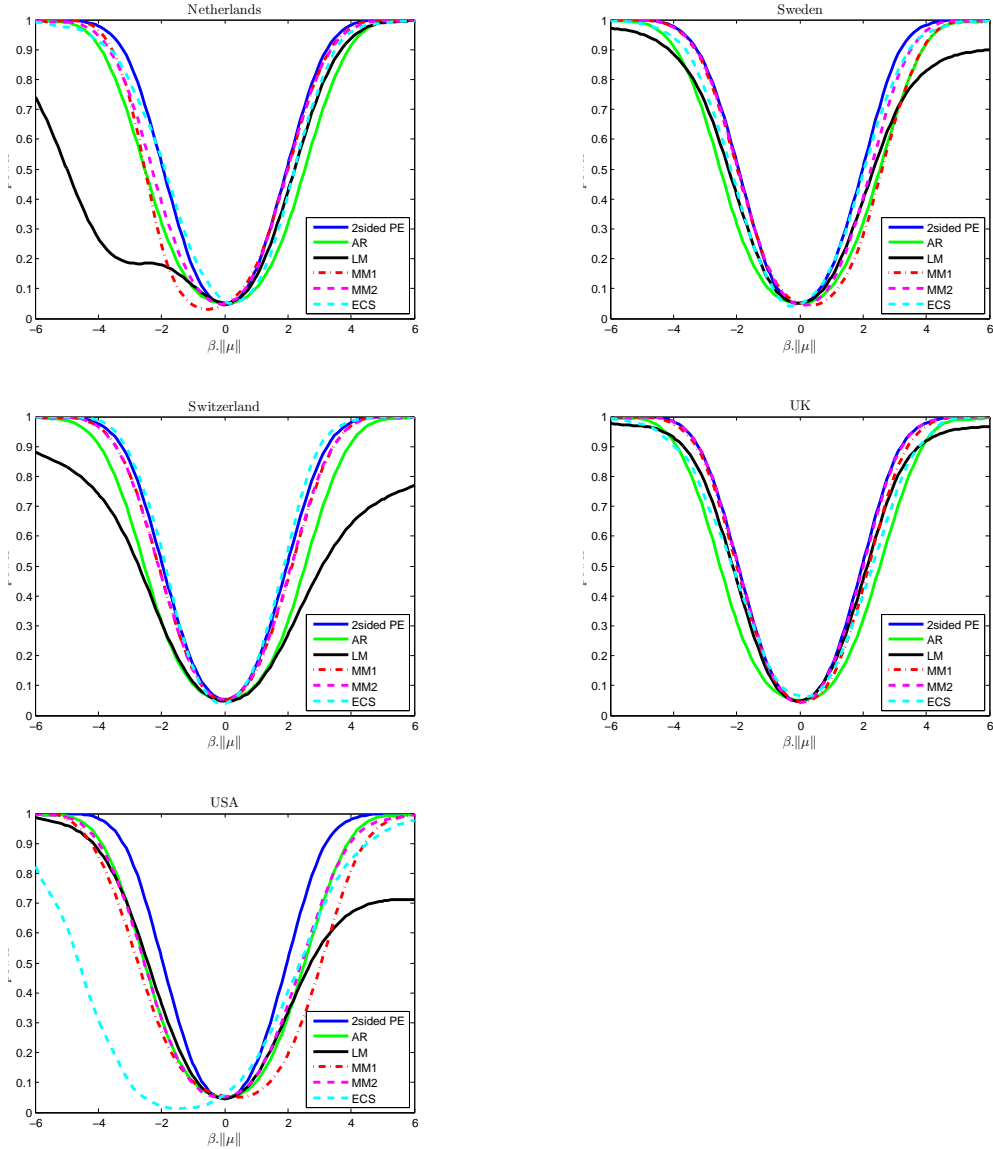
We then compare power among two-sided tests which have arguably better performance. Figure 4 plots power curves for the two-sided power envelope, MM1-SU, MM2-SU, CQLR, CQLR-kron, and PI-CLC tests. All tests are adequate for two-sided hypothesis testing. The PI-CLC and CQLR-kron test show some improvements over the CQLR test for some, but not all, countries. The MM1-SU test behaves near the MM2-SU test for several countries, but it has considerably lower power for Japan and the United States[7]. The MM2-SU test outperforms these tests and when it occa-

---

[7]Conceivably, this power loss can be due to numerical integration over the whole real line. Power may be improved by transforming the parameter β to the quantity $\theta = \tan^{-1}(d_\beta/c_\beta)$. This improvement is left for future work.

sionally has less power, the power loss is small. This application based on real data supports our theoretical contribution and the use of the MM2-SU test in practice.

Figure 4: Power Comparison (two-sided tests)
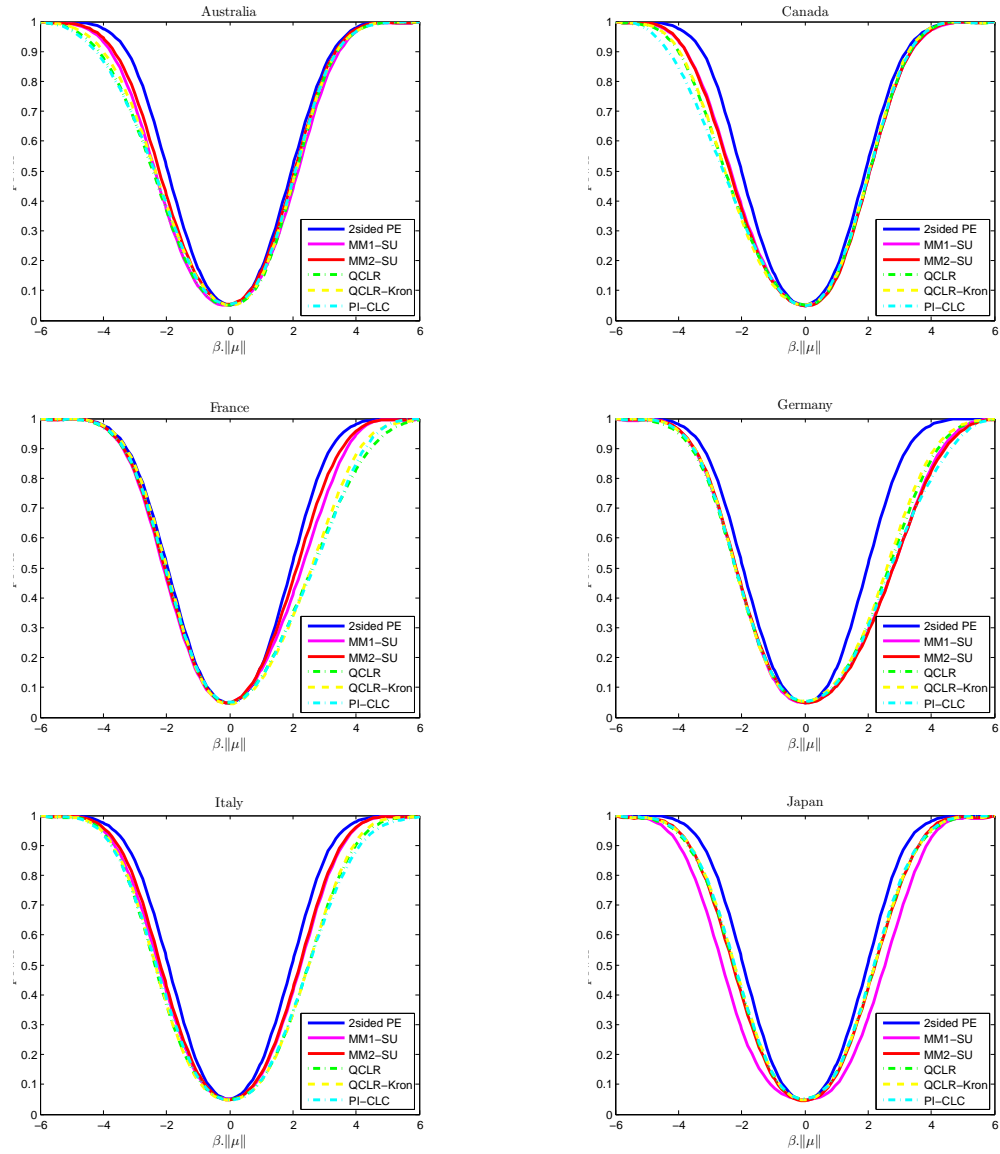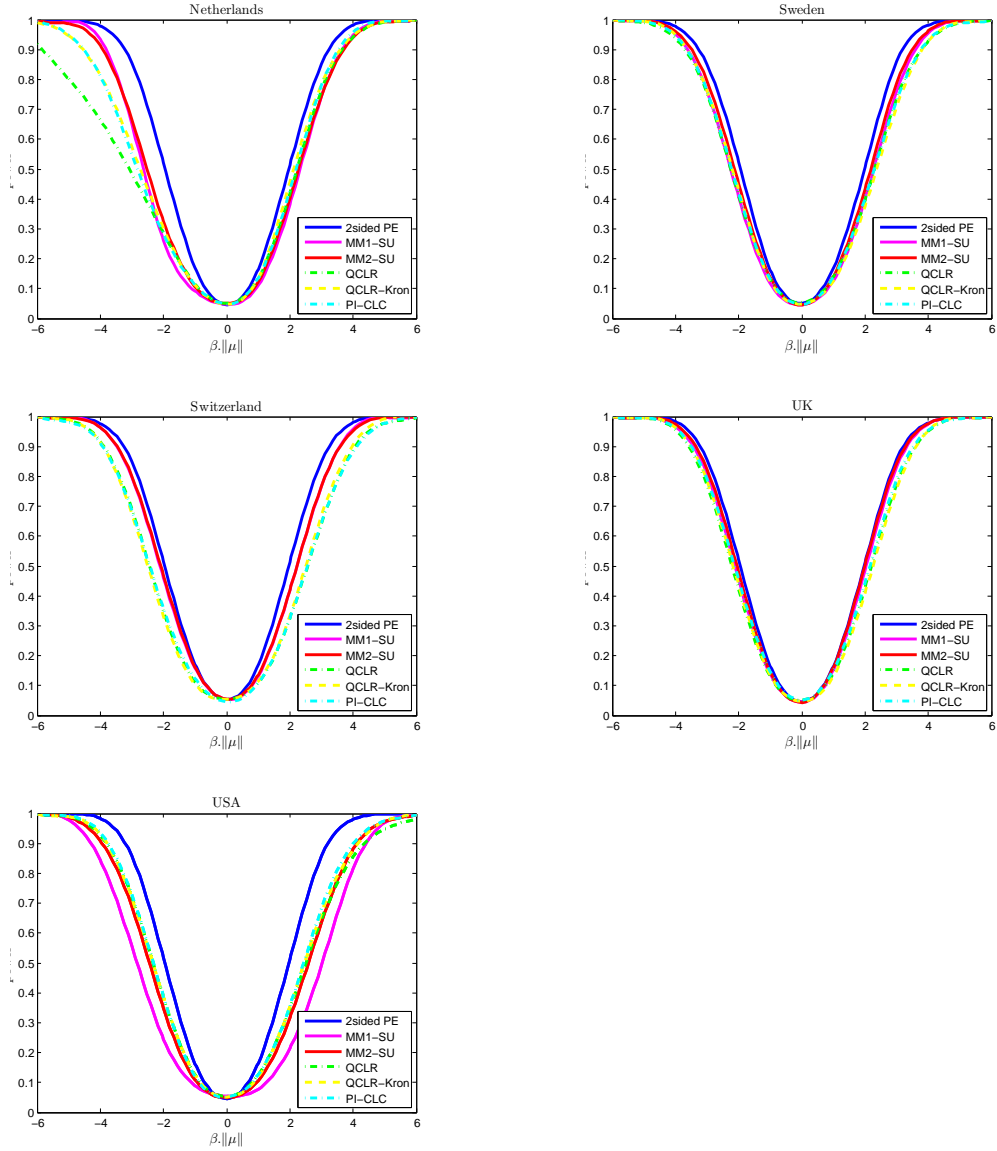
8 Concluding Remarks

In this paper, we study the instrumental variable (IV) model with one endogenous regressor and heteroskedastic and autocorrelated (HAC) errors. The HAC-IV model with a known variance matrix is simpler than the model with an unknown but consistently estimable long-run variance. However, inference in both models is approximately the same whether or not the instruments are weakly correlated with the endogenous variable. This simplification allows us to develop a theory of optimal two-sided tests when the error stochastic process is of unknown form.

We find that a test that has correct size and is optimal under standard asymptotics may still have unacceptably low power in finite samples. This issue appears in several econometric models. For the HAC-IV model, we solve this problem by finding weighted-average power tests satisfying additional two-sided conditions. In this paper, we consider two possibilities: the locally unbiased (LU) and strongly unbiased (SU) conditions. While the local condition yields admissible tests, the stronger condition is easier to implement. Better yet, the MM1-SU and MM2-SU tests have power numerically very close to their LU versions. Numerical simulations also show that the MM2-SU test outperforms other tests proposed for the HAC-IV model.

The only other paper that satisfactorily addresses optimality of two-sided tests in the HAC-IV model is that of I. Andrews (2015). He explores linear combinations of the Anderson-Rubin and score statistics, with weights dependent on the conditioning statistic T. A class of these conditional linear combination (CLC) tests is unbiased and admissible in the conditional problem. By proposing a minimax regret criterion, he delivers a test which plugs in a nuisance-parameter estimator. There is some power gained by broadening the focus beyond those three statistics. On the other hand, we impose k additional constraints which are related to the SU condition. It would be interesting to reduce the required computational time while maintaining the power gains of the MM2-SU test by reducing the number of boundary conditions when finding a WAP test.

Finally, the asymptotic theory based on Laplace approximations, developed in this paper, is easily adaptable to other econometric models. For the HAC-IV model, it relies on priors for the parameters β and π being insensitive to the sample size. For the MM1 and MM2 weights, this implies that the tuning parameters σ^2 and ζ (used in the prior for $\mu = (Z'Z)^{1/2}\pi$) eventually grow at the sample size n. Some power gains with weak instruments may be possible when the tuning parameters are held constant. Another alternative is to find an automatic rate for σ^2 and ζ using a plug-in method. For example, we could let these parameters be proportional to either $\|T\|^2$ or $n \cdot \|\pi(\beta_0)\|^2$. These quantities are stochastically bounded under weak instruments and grow at the rate n under strong instruments (which assures asymptotic optimality). Since the constrained MLE $\pi(\beta_0)$ is a one-to-one transformation of T, these modifications of WAP-SU tests are still similar and uncorrelated with the pivotal statistic S (hence, satisfy the SU Condition)[8]. We will consider this possibility in future work.

9   Appendix A: Derivation of Test Statistics

In this section, we derive the likelihood ratio (LR), Lagrange multiplier (or score), and both MM statistics.

---

[8]See Moreira (2001, 2009) for selecting among similar tests without creating size distortions; the argument uses completeness of T and is applicable to the SU condition as well.

9.1 The LR statistic

For the statistic $\overline{R} = \left[vec \left(Z'Z \right)^{-1/2} Z'Y \right]$, the log-likelihood is proportional to

$$L_n \left(\beta, \mu \right) = -\frac{1}{2} \left(\overline{R} - (a \otimes I_k)\mu \right)' \Sigma^{-1} \left(\overline{R} - (a \otimes I_k)\mu \right). \tag{9.43}$$

Taking the derivative with respect to μ,

$$\frac{\partial L_n \left(\beta, \mu \right)}{\partial \mu} = (a' \otimes I_k)\Sigma^{-1} \left(\overline{R} - (a \otimes I_k)\mu \right) = 0,$$

yields the constrained maximum likelihood estimator (MLE)

$$\mu \left(\beta \right) = \left[(a' \otimes I_k)\Sigma^{-1}(a \otimes I_k) \right]^{-1} (a' \otimes I_k)\Sigma^{-1}\overline{R}.$$

The concentrated log-likelihood function, $L_n^c \left(\beta \right)$ is

$$\begin{aligned}
L_n \left(\beta, \mu \left(\beta \right) \right) &= -\frac{1}{2}\overline{R}'\Sigma^{-1/2}M_{\Sigma^{-1/2}(a \otimes I_k)}\Sigma^{-1/2}\overline{R} \\
&= -\frac{1}{2}\overline{R}'\Sigma^{-1}\overline{R} + \frac{1}{2}\overline{R}'\Sigma^{-1/2}N_{\Sigma^{-1/2}(a \otimes I_k)}\Sigma^{-1/2}\overline{R}.
\end{aligned} \tag{9.44}$$

The LR statistic follows from

$$\begin{aligned}
LR &= 2 \cdot \left[\max_{\beta} L_n \left(\beta, \mu \left(\beta \right) \right) - L_n \left(\beta_0, \mu \left(\beta_0 \right) \right) \right] \\
&= \max_{\beta} \overline{R}'\Sigma^{-1/2}N_{\Sigma^{-1/2}(a \otimes I_k)}\Sigma^{-1/2}\overline{R} - \overline{R}'\Sigma^{-1/2}N_{\Sigma^{-1/2}(a_0 \otimes I_k)}\Sigma^{-1/2}\overline{R}.
\end{aligned}$$

9.2 The LM statistic

The score is given by

$$\begin{aligned}
\frac{\partial L_n^c \left(\beta \right)}{\partial \beta} &= \overline{R}'\Sigma^{-1}(a \otimes I_k) \left[(a' \otimes I_k)\Sigma^{-1}(a \otimes I_k) \right]^{-1} (e_1' \otimes I_k)\Sigma^{-1}\overline{R} \\
&\quad -\frac{1}{2}\overline{R}'\Sigma^{-1}(a \otimes I_k) \left[(a' \otimes I_k)\Sigma^{-1}(a \otimes I_k) \right]^{-1} \\
&\quad \times \left\{ (e_1' \otimes I_k)\Sigma^{-1}(a \otimes I_k) + (a' \otimes I_k)\Sigma^{-1}(e_1 \otimes I_k) \right\} \\
&\quad \times \left[(a' \otimes I_k)\Sigma^{-1}(a \otimes I_k) \right]^{-1} (a' \otimes I_k)\Sigma^{-1}\overline{R}.
\end{aligned}$$

At $\beta = \beta_0$:

$$\begin{aligned}
\frac{\partial L_n^c \left(\beta_0 \right)}{\partial \beta} &= \overline{R}'\Sigma^{-1}(a_0 \otimes I_k) \left[(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k) \right]^{-1} \\
&\quad \times (e_1' \otimes I_k)\Sigma^{-1/2}M_{\Sigma^{-1/2}(a_0 \otimes I_k)}\Sigma^{-1/2} \left[(e_1, e_2) \otimes I_k \right] \overline{R} \\
&= \overline{R}'\Sigma^{-1}(a_0 \otimes I_k) \left[(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k) \right]^{-1} \\
&\quad \times (e_1' \otimes I_k)\Sigma^{-1/2}M_{\Sigma^{-1/2}(a_0 \otimes I_k)}\Sigma^{-1/2} \left((e_1, a_0 - \beta_0 e_1) \otimes I_k \right) \overline{R}.
\end{aligned}$$

Note that

$$C_{\beta_0}^{-1} = (e_1' \otimes I_k)\Sigma^{-1}(e_1 \otimes I_k) - (e_1' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)$$
$$\times \left[(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)\right]^{-1} (a_0' \otimes I_k)\Sigma^{-1}(e_1 \otimes I_k).$$

Indeed, define

$$[X_1, X_2] = \left[\Sigma^{-1/2}(e_1 \otimes I_k), \Sigma^{-1/2}(a_0 \otimes I_k)\right] = \Sigma^{-1/2}[e_1, a_0] \otimes I_k.$$

Using

$$\left([X_1, X_2]'[X_1, X_2]\right)^{-1} = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} = \begin{pmatrix} X^{11} & X^{12} \\ X^{21} & X^{22} \end{pmatrix},$$

we obtain

$$\begin{aligned} \left([X_1, X_2]'[X_1, X_2]\right)^{-1} &= \left[\left([e_1, a_0]' \otimes I_k\right)\Sigma^{-1}\left([e_1, a_0] \otimes I_k\right)\right]^{-1} \\ &= \left([e_1, a_0]^{-1} \otimes I_k\right)\Sigma\left([e_1, a_0]^{-1'} \otimes I_k\right) \\ &= \left(\begin{pmatrix} 1 & -\beta_0 \\ 0 & 1 \end{pmatrix} \otimes I_k\right)\Sigma\left(\begin{pmatrix} 1 & 0 \\ -\beta_0 & 1 \end{pmatrix} \otimes I_k\right). \end{aligned}$$

Therefore, the top-left submatrix X^{11} of the matrix $\left([X_1, X_2]'[X_1, X_2]\right)^{-1}$ equals C_{β_0}:

$$(e_1' \otimes I_k)\Sigma^{-1/2}M_{\Sigma^{-1/2}(a \otimes I_k)}\Sigma^{-1/2}e_1 = \left[(b_0' \otimes I_k)\Sigma(b_0 \otimes I_k)\right]^{-1}.$$

We obtain

$$\begin{aligned} \frac{\partial L_n^c}{\partial \beta} &= \overline{R}'\Sigma^{-1}(a_0 \otimes I_k)\left[(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)\right]^{-1} \\ &\quad \times \left[(b_0' \otimes I_k)\Sigma(b_0 \otimes I_k)\right]^{-1}(b_0' \otimes I_k)\overline{R} \\ &= S'C_{\beta_0}D_{\beta_0}^{-1}T. \end{aligned}$$

We can standardize it by a consistent estimator of the asymptotic variance. In particular, we can choose $T'D_{\beta_0}^{-1}C_{\beta_0}^2 D_{\beta_0}^{-1}T$ and obtain the LM statistic.

9.3 The MM statistics

The weighting functions are chosen after approximating the covariance matrix Σ by the Kronecker product $\Omega \otimes \Phi$. Let $\|\Sigma\|_F = (tr(\Sigma'\Sigma))^{1/2}$ denote the Frobenius norm of Σ. For a positive-definite covariance matrix Σ, we rely on Van Loan and Ptsianis (1993, p. 14) to find symmetric and positive definite matrices Ω and Φ with dimensions 2×2 and $k \times k$ which minimize $\|\Sigma - \Omega_0 \otimes \Phi_0\|_F$.

For the MM1 statistic $h_1(s, t)$, we choose $\Lambda_1(\beta, \mu)$ to be $N(\beta_0, 1) \times N(0, \sigma^2 \Phi)$. For the prior $\mu \sim N(0, \sigma^2 \Phi)$, the integrated likelihood is

$$(2pi)^{-k} \left| \Psi_{\beta, \sigma^2} \right|^{-1/2} \exp \left(-\frac{(s', t') \Psi_{\beta, \sigma^2}^{-1} (s', t')'}{2} \right) \tag{9.45}$$

where the $2k \times 2k$ covariance matrix is given by (2.11). The integrated likelihood for S and T is then

$$h_1(s, t) = (2pi)^{-k-1/2} \int_{-\infty}^{\infty} \left| \Psi_{\beta, \sigma^2} \right|^{-1/2} \exp \left(-\frac{(s', t') \Psi_{\beta, \sigma^2}^{-1} (s', t')' + (\beta - \beta_0)^2}{2} \right) d\beta.$$

For the MM2 statistic $h_2(s, t)$, we first make a change of variables from β to θ given by $\tan(\theta) = d_\beta / c_\beta$. It is now convenient to change variables:

$$(\cos(\theta), \sin(\theta))' = l_\beta / \|l_\beta\|.$$

Using equation (2.8) of AMS06, we can show that the one-to-one mapping $\beta(\theta)$ is

$$\beta = \beta_0 + \frac{b_0' \Omega b_0}{e_2' \Omega b_0 + \tan(\theta) \cdot |\Omega|^{1/2}}.$$

We then choose the density for θ and μ to be $Unif[-pi, pi] \times N(0, \|l_{\beta(\theta)}\|^{-2} \zeta \cdot \Phi)$, where $l_\beta = (c_\beta, d_\beta)'$. For the prior on μ, the integrated likelihood is

$$(2pi)^{-k} \left| \Psi_{\beta, \|l_\beta\|^{-2} \zeta} \right|^{-1/2} \exp \left(-\frac{(s', t') \Psi_{\beta, \|l_\beta\|^{-2} \zeta}^{-1} (s', t')'}{2} \right).$$

The integrated likelihood for S and T using the prior on both μ and θ yields

$$h_2(s, t) = (2pi)^{-(k+1)} \int_{-pi}^{pi} \left| \Psi_{\beta(\theta), \|l_{\beta(\theta)}\|^{-2} \zeta} \right|^{-1/2} \exp \left(-\frac{(s', t') \Psi_{\beta(\theta), \|l_{\beta(\theta)}\|^{-2} \zeta}^{-1} (s', t')'}{2} \right) d\theta.$$

References

ANDERSON, T. W., AND H. RUBIN (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

ANDREWS, D. W. K. (1986): "Complete Consistency: A Testing Analogue of Estimator Consistency," *The Review of Economic Studies*, 53, 263–269.

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

ANDREWS, D. W. K., X. CHENG, AND P. GUGGENBERGER (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Tests, Cowles Foundation Discussion Papers 1813, Cowles Foundation for Research in Economics, Yale University.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2010): "Applications of Subsampling, Hybrid, and Size-Correction Methods," *Journal of Econometrics*, 158, 285–305.

——— (2014a): "Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models," Working paper, Yale University.

——— (2014b): "Identification- and Singularity-Robust Inference for Moment Condition Models," Working paper, Yale University.

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression," NBER Working Paper t0299.

——— (2006a): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752.

——— (2006b): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, 715–752, Supplement.

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

ANDREWS, I. (2015): "Conditional Linear Combination Tests for Weakly Identified Models," Unpublished Manuscript, MIT.

ANDREWS, I., AND A. MIKUSHEVA (2015): "Conditional Inference with a Functional Nuisance Parameter," Working paper, MIT.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of American Statistical Association*, 90, 443–450.

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

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

EDWARDS, A. W. F. (1992): *Likelihood.* John Hopkins University Press, Baltimore.

ELLIOTT, G., U. MUELLER, AND M. WATSON (2015): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

GUGGENBERGER, P., AND R. SMITH (2005): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak and Strong Identification," *Econometric Theory*, 21, 667–709.

JANSSON, M. (2002): "Consistent Covariance Matrix Estimation for Linear Processes," *Econometric Theory*, 18, 1449–1459.

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

———— (2006): "Testing Parameters in GMM Without Assuming That They Are Identified," *Econometrica*, 73, 1103–1123.

LEHMANN, E. L. (1952): "On the Existence of Least Favorable Distributions," *Annals of Mathematical Statistics*, 23, 408–416.

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 (2014a): "Tests Based on t-Statistics for IV Regression with Weak Instruments," *Journal of Econometrics*, 182, 351–363.

———— (2014b): "Tests Based on t-Statistics for IV Regression with Weak Instruments," *Journal of Econometrics*, 182, 351–363, Supplement.

MOREIRA, H., AND M. J. MOREIRA (2010): "Contributions to the Theory of Similar Tests," Working Paper, FGV/EPGE.

———— (2013): "Contributions to the Theory of Optimal Tests," *Ensaios Economicos*, 747, FGV/EPGE.

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

———— (2002): "Tests with Correct Size in the Simultaneous Equations Model," Ph.D. thesis, 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.

Nelson, C. R., and R. Startz (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

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

Olea, J. L. M. (2015): "Efficient Conditionally Similar Tests: Finite-Sample Theory and Large-Sample Applications," Working paper, NYU.

Olver, F. W. J. (1997): *Asymptotics and Special Functions.* San Diego, Academic Press.

Onatski, A., M. J. Moreira, and M. Hallin (2014a): "Asymptotic Power of Sphericity Tests for High-Dimensional Data," *Annals of Statistics*, 41, 1204–1231.

——— (2014b): "Signal Detection in High Dimension: The Multispiked Case," *Annals of Statistics*, 42, 225–254.

Otsu, T. (2006): "Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models Under Weak Identification," *Econometric Theory*, 22, 513–527.

Parzen, E. (1954): "On Uniform Convergence of Families of Sequences of Random Variables," *University of California Publications in Statistics*, 2, 23–54.

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

Stock, J. H., and J. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

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.

Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Rate of Substitution When Instruments Are Weak," *Review of Economics and Statistics*, 86, 797–810.