



Tests with correct size when instruments can be arbitrarily weak

Marcelo J. Moreira*

Department of Economics, Columbia University, New York, NY 10027, USA
 FGV/EPGE, Rio de Janeiro, RJ 22250, Brazil

ARTICLE INFO

Article history:
 Available online 31 January 2009

JEL classification:
 C12
 C31

Keywords:
 Instrumental variables regression
 Curved exponential family
 Weak instruments
 Pre-testing

ABSTRACT

This paper applies classical exponential-family statistical theory to develop a unifying framework for testing structural parameters in the simultaneous equations model under the assumption of normal errors with known reduced-form variance matrix. The results can be divided into the limited-information and full-information categories. In the limited-information model, it is possible to characterize the entire class of similar tests in a model with only one endogenous explanatory variable. In the full-information framework, this paper proposes a family of similar tests for subsets of endogenous variables' coefficients. For both limited- and full-information models, there exist power upper bounds for unbiased tests. When the model is just-identified, the Anderson–Rubin, score, and (pseudo) conditional likelihood ratio tests are optimal. When the model is over-identified, the (pseudo) conditional likelihood ratio test has power close to the power envelope when identification is strong.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Applied researchers are often interested in making inferences about the parameters of endogenous variables in a structural equation. Identification is achieved by assuming the existence of instrumental variables uncorrelated with the structural error but correlated with the endogenous regressors. If the instruments are strongly correlated with the regressors, standard asymptotic theory can be employed to develop reliable inference methods. However, as emphasized in recent work by Nelson and Startz (1990), Bound et al. (1995), Dufour (1997), and Staiger and Stock (1997), these methods are not satisfactory when instruments are only weakly correlated with the regressors. In particular, the usual tests and confidence regions do not have correct size in the weak instrument case.

The main contribution of this paper is to establish a connection between the weak-instrument problem and classical statistical theory on hypothesis testing. This finding allows the construction of tests for endogenous variables' coefficients with correct size even when instruments can be weak. To develop the theory of hypothesis testing, this paper provides a mathematical definition to distinguish limited-information and full-information models.

In the limited-information model with one endogenous variable, there is a necessary and sufficient condition for a test of the endogenous variable's coefficient to be similar. This unifies the

theory of similar tests of Anderson and Rubin (1949), Dufour and Jasiak (2001), Kleibergen (2002), and Moreira (2002, 2003). The class of similar tests is large and includes all unbiased tests. In the just-identified model, the Anderson–Rubin, score, and conditional likelihood ratio (CLR) tests are optimal among the class of unbiased tests. In the over-identified model, there exists a power upper bound for unbiased tests. No test can uniformly achieve this power envelope.

Monte Carlo simulations show that the CLR test for the endogenous variable's coefficient has good power overall in over-identified models. It dominates the Anderson–Rubin and score tests, and has power close to the power envelope for unbiased tests when instruments are strong. This finding provides a refinement over the first-order asymptotics, which asserts that the score and CLR tests are optimal under local alternatives and are equivalent to the Anderson–Rubin test with fixed alternatives.

In the full-information model with more than one endogenous variable, this paper proposes a class of similar tests for subsets of the endogenous variables' coefficients. Available procedures either rely on strong partial identification or are biased. Within this class of similar tests, there are three tests based on the Anderson–Rubin, score, and CLR approaches for an endogenous variable's coefficient in the full-information model. Previous Monte Carlo results carry over to the full-information model: the (pseudo) CLR test has overall good power and, in particular, reaches a power bound for unbiased tests.

The remainder of this paper is organized as follows. Section 2 presents the simultaneous equations model and introduces some notation. Section 3 derives results for the one endogenous variable's coefficient in the limited-information model. Section 4 obtains tests for subsets of endogenous variables' coefficients

* Corresponding address: Department of Economics, Columbia University, New York, NY 10027, USA.

E-mail address: mjmoreira@columbia.edu.

in the full-information model. Section 5 obtains asymptotic results based on finite-sample theory. Section 6 provides power comparisons for the tests proposed in Sections 3 and 4. Section 7 concludes and gives direction for future research. All proofs are given in Appendix B.

2. The simultaneous equations model

Consider the structural equation

$$y_1 = y_2\beta + X\gamma + u, \quad (1)$$

where y_1 is an n -dimensional vector, y_2 is an $n \times l$ matrix, X is an $n \times m$ matrix of exogenous variables, and u is an $n \times 1$ unobserved error vector. This equation is assumed to be part of a larger linear simultaneous equations model, in which y_2 is allowed to be correlated with u . The complete system contains exogenous variables which can be used as instruments for conducting inference on β . The restrictions on the reduced-form regression coefficients are implied by the identifying assumption that there exist exogenous variables which do not appear in (1). Specifically, it is assumed that

$$y_2 = X\tilde{\Gamma} + \tilde{Z}\Pi + v_2, \quad (2)$$

where \tilde{Z} is an $n \times k$ matrix of exogenous variables having full column rank, Π is a $k \times l$ matrix, and $\tilde{\Gamma}$ is an $m \times l$ matrix. For convenience, transform the matrix \tilde{Z} so that the transformed matrix Z and the exogenous regressor matrix X are orthogonal: $Z'X = 0$. For any matrix Q having full column rank, let $N_Q = Q(Q'Q)^{-1}Q'$ and $M_Q = I - N_Q$. Then, the underlying stochastic equation for y_2 is given by

$$y_2 = X\Gamma + Z\Pi + v_2, \quad (3)$$

where $Z = M_X\tilde{Z}$, and $\Gamma = (X'X)^{-1}X'\tilde{Z}\Pi + \tilde{\Gamma}$. The reduced-form model is

$$y_1 = X(\Gamma\beta + \gamma) + Z\Pi\beta + v_1 \quad (4)$$

$$y_2 = X\Gamma + Z\Pi + v_2.$$

The reduced-form model for $Y = [y_1, y_2]$ can be written concisely as

$$Y = X(\Gamma a' + \gamma e_1') + Z\Pi a' + V,$$

where $a = [\beta, I_l]'$ and $e_1 = [1, 0_l]'$. The n rows of the reduced-form error matrix $V = [v_1, v_2]$ are assumed to be i.i.d. normal with mean zero and known $(l+1) \times (l+1)$ variance matrix

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}, \quad (5)$$

which is partitioned conformably to $Y = [y_1, y_2]$. The assumption of known Ω will be relaxed later using the weak-instrument asymptotics of Staiger and Stock (1997).

The goal is to test (subsets of) β , treating Π , γ , and Γ as nuisance parameters. A test is said to be of size α if the probability of rejecting the null hypothesis when it is true does not exceed α . That is,

$$\sup \text{prob}(\text{rejecting } H_0) = \alpha,$$

where the sup is over all values of β , Π , γ , and Γ consistent with the null hypothesis. Since these parameters are unknown, finding a test with correct size is nontrivial. The task is simplified if one can find tests whose null rejection probability does not depend on the nuisance parameters at all. These tests are called *similar tests*. If, for example, one rejects the null if some test statistic \mathcal{T} is greater than a given constant, the test will be similar if the distribution of \mathcal{T} under the null hypothesis does not depend on the nuisance parameters. Such test statistics are said to be *pivotal*. If \mathcal{T} has a null distribution depending on the nuisance parameters but it can be bounded by a pivotal statistic, then \mathcal{T} is said to be *boundedly pivotal*.

In practice, one often uses test statistics that are only asymptotically pivotal:

$$\lim_{n \rightarrow \infty} \text{prob}(\mathcal{T} > c) = G(c),$$

where the approximate distribution function G does not depend on the unknown parameters β , Π , γ , and Γ compatible with the null hypothesis. These tests may be satisfactory when the convergence is uniform and the sup and lim operators can be interchanged. However, if the convergence is not uniform, the actual size of the test may differ substantially from the size based on the asymptotic distribution of \mathcal{T} . In fact, Dufour (1997) extends finite-sample results by Gleser and Hwang (1987) to show that the true levels of the usual Wald-type tests deviate arbitrarily from their nominal levels if $\Pi \in \mathbf{P}$ cannot be bounded away from the origin; that is, $0 \in \bar{\mathbf{P}}$. In this sense, the instruments can be arbitrarily weak. Since weak instruments appear in empirical research, it is desirable to find tests with approximately correct size α even when Π cannot be bounded away from the origin.

3. One endogenous variable

When $l = 1$ and $m > 0$, the reduced-form model is given by

$$y_1 = X(\Gamma\beta + \gamma) + Z\Pi\beta + v_1 \quad (6)$$

$$y_2 = X\Gamma + Z\Pi + v_2,$$

where β is a scalar, Π is a $k \times 1$ vector, and γ and Γ are $m \times 1$ vectors. The focus here is to construct tests with correct size for the null hypothesis $H_\beta : \beta = \beta_0$.

Under the normality assumption, the probability model is a member of the curved exponential family. The sufficient statistics for (γ, Γ) and (β, Π) are given by $X'Y$ and $Z'Y$, respectively. The nuisance parameters γ and Γ can be eliminated by requiring the test to be invariant to linear transformations of X . Any invariant test can be written as a function of a maximal invariant statistic; see Theorem 6.2.1 of Lehmann (1986, p. 285). For the group \mathcal{G} of transformations that preserves H_β , $g(Y) = Y + XF$ for arbitrary conformable matrices F , the maximal invariant in terms of the sufficient statistic is $Z'Y$. For any non-singular 2×2 matrix D , $Z'YD$ is also a maximal invariant. A convenient choice is $D = [b_0, \Omega^{-1}a_0]$, where $b_0 = (1, -\beta_0)'$ is orthogonal to $a_0 = (\beta_0, 1)'$. This yields the pair

$$S_\beta = Z'Yb_0 = Z'u_0 \quad \text{and} \quad T_\beta = Z'Y\Omega^{-1}a_0, \quad (7)$$

where $u_0 = y_1 - y_2\beta_0$.

The vectors S_β and T_β are independent and normally distributed under both the null and alternative hypotheses. Specifically,

$$S_\beta \sim N(Z'Z\Pi(\beta - \beta_0), Z'Z \cdot b_0'\Omega b_0) \quad \text{and}$$

$$T_\beta \sim N(Z'Z\Pi \cdot a_0'\Omega^{-1}a_0, Z'Z \cdot a_0'\Omega^{-1}a_0).$$

Although the null distribution of S_β does not depend on the nuisance parameter Π , the null distribution of T_β is very sensitive to the value of Π . A little algebra shows that

$$T_\beta = a_0'\Omega^{-1}a_0 \cdot Z'Z\hat{\Pi},$$

where $\hat{\Pi}$ is the maximum likelihood estimator of Π when β is constrained to take the null value β_0 . The unknown parameter Π is assumed to change freely, at least over a large enough set. Assumption LI gives a mathematical meaning to the notion of limited-information model.

Assumption LI (Limited Information). The set \mathbf{P} in which Π lies contains a k -dimensional rectangle.

All tests invariant to the group \mathcal{G} can be written as (possibly randomized) functions of S_β and T_β . Specifically, let ϕ be a critical function such that $0 \leq \phi \leq 1$. For each S_β and T_β the test rejects or accepts the null with probabilities $\phi(S_\beta, T_\beta)$ and $1 - \phi(S_\beta, T_\beta)$, respectively; the dependence of ϕ on Z , β_0 and Ω is omitted out of convenience. For example, a nonrandomized test that rejects

the null hypothesis when the test statistic $\mathcal{T}(S_\beta, T_\beta)$ is larger than c is given by $\phi(S_\beta, T_\beta) = I(\mathcal{T}(S_\beta, T_\beta) > c)$. Let E_0 represent expectation over the null marginal distribution of S_β and let \mathcal{P}^{T_β} be the family of distributions of T_β under the null hypothesis. An event holds a.e. \mathcal{P}^{T_β} if it is true except on a set with probability zero for all null distributions of T_β . The following result characterizes all similar tests in terms of the marginal distribution of S_β .

Theorem 1 (Similarity Condition). *If Assumption LI holds, then any test invariant to the group \mathcal{G} is similar at size α if and only if it can be written as $\phi(S_\beta, T_\beta)$ such that $E_0\phi(S_\beta, t) = \alpha$, a.e. \mathcal{P}^{T_β} .*

Comments: 1. The power function of any test is continuous; see Lemma 2.7.2 of Lehmann (1986, p. 48). Hence, any unbiased test for $H_\beta : \beta = \beta_0$ must necessarily be similar.

2. The theorem states that any similar test must be conditionally similar for almost every realization of T_β . Because S_β is pivotal under H_β , it is possible to find similar tests by looking at the marginal distribution of the statistic S_β under the null hypothesis.

3. If Assumption LI does not hold, it may be possible to find a similar test even if $E_0\phi(S_\beta, t) \neq \alpha$. For example, consider the extreme case in which Π is known. Because $\mathbf{P} = \{\Pi\}$, the null hypothesis $H_\beta : \beta = \beta_0$ is simple, and the null distribution of any statistic $\psi(S_\beta, T_\beta)$ is known. A test that rejects H_β if $\psi(S_\beta, T_\beta)$ is larger than its $(1 - \alpha)$ -quantile is trivially similar, but it is not conditionally similar if the $(1 - \alpha)$ -quantile of $\psi(S_\beta, t)$ depends on t .

4. The focus here is on tests invariant to the group \mathcal{G} , but an analogous result holds without appealing to the invariance principle. Suppose that γ and Γ can take any value over a m -dimensional rectangle. Then any test is similar at size α if and only if it can be written as $\phi(S_\beta, T_\beta, X'Y)$ such that $E_0\phi(S_\beta, t, r) = \alpha$ for almost all values of $t \in \mathbb{R}^k$ and $r \in \mathbb{R}^{m \times 2}$.

Three examples of similar tests are the Anderson and Rubin (1949) test, the score test independently proposed by Kleibergen (2002) and Moreira (2002) for the weak-instrument setting, and the conditional likelihood ratio test of Moreira (2003). It is convenient to write these tests in terms of the standardized statistics

$$\begin{aligned} \bar{S}_\beta &= (Z'Z)^{-1/2} S_\beta \cdot (b'_0 \Omega b_0)^{-1/2}, \\ \bar{T}_\beta &= (Z'Z)^{-1/2} T_\beta \cdot (a'_0 \Omega^{-1} a_0)^{-1/2}. \end{aligned}$$

Example A. Reject H_β if the Anderson–Rubin statistic for known Ω is larger than the $(1 - \alpha)$ -quantile of a chi-square- k distribution:

$$AR_\beta = AR(\bar{S}_\beta, \bar{T}_\beta) = \bar{S}'_\beta \bar{S}_\beta > q_\alpha(k).$$

The Similarity Condition states that the null rejection probability of any similar test does not depend on T_β . This, however, does not imply that the test itself does not depend on T_β , as the following example shows.

Example B. Reject H_β if the score statistic is larger than the $(1 - \alpha)$ -quantile of a chi-square-one distribution:

$$LM_\beta = LM(\bar{S}_\beta, \bar{T}_\beta) = \bar{S}'_\beta N_{\bar{T}_\beta} \bar{S}_\beta > c_\alpha.$$

Moreira (2002) shows that \bar{T}_β is independent of \bar{S}_β and, consequently, the null distribution of LM_β is chi-square-one. Interestingly, LM_β is just a particular case of a score-type statistic proposed by Breusch and Pagan (1980) in a general framework; see Appendix A.

Theorem 1 indicates that if the unconditional null rejection probability equals α , then the conditional null rejection probability equals α . Consequently, there is no loss of generality in focusing on conditionally similar tests. This finding supports the conditional

approach of Moreira (2003). Example C describes the conditional likelihood ratio (CLR) test of Moreira (2003) (in which the standard chi-square critical value is replaced by $c_\psi(t; \beta_0, \alpha)$, the $(1 - \alpha)$ -quantile of the distribution of $\psi(\bar{S}_\beta, t, \beta_0)$).

Example C. Reject H_β when the likelihood ratio statistic is larger than its conditional $(1 - \alpha)$ -quantile:

$$\begin{aligned} LR_\beta &= LR(\bar{S}_\beta, \bar{T}_\beta) > c_{LR}(\bar{T}_\beta; \beta_0, \alpha), \quad \text{where} \\ LR(s, t) &= \frac{1}{2} \left[s's - t't + \sqrt{(s's - t't)^2 + 4(t't)^2} \right]. \end{aligned}$$

The Similarity Condition is convenient for constructing tests with correct size, but ultimately one wants to find tests with good power properties. When Π is bounded away from zero and the sample size is large, the standard likelihood ratio, Wald, and Lagrange Multiplier tests have approximate power

$$1 - \kappa \left(c_\alpha; \frac{\Pi'Z'Z\Pi(\beta - \beta_0)^2}{\sigma_0^2} \right), \tag{8}$$

where $\sigma_0^2 = b'_0 \Omega b_0$ and $\kappa(\cdot; \mu)$ is the noncentral $\chi^2(1)$ distribution function with noncentrality parameter μ . However, when Π is near the origin, these tests are not generally similar and the power approximation in (8) is unreliable. To assess the finite-sample power functions of similar tests for the two-sided alternative $K_\beta : \beta \neq \beta_0$, Theorem 2 derives power upper bounds for unbiased tests.

Theorem 2. Consider testing $H_\beta : \beta = \beta_0$ against $K_\beta : \beta \neq \beta_0$, with $\Pi \neq 0$.

(a) If the model is just-identified, the uniformly most powerful unbiased test (invariant to the group \mathcal{G}) has a power function given by

$$P_{\beta, \Pi}(AR_\beta > c_\alpha) = 1 - \kappa \left(c_\alpha; \frac{\Pi'Z'Z\Pi(\beta - \beta_0)^2}{\sigma_0^2} \right). \tag{9}$$

(b) If Π is known, the uniformly most powerful unbiased test invariant to group \mathcal{G} has a power function given by

$$P_{\beta, \Pi}(\mathcal{R} > c_\alpha) = 1 - \kappa \left(c_\alpha; \frac{\Pi'Z'Z\Pi(\beta - \beta_0)^2}{\omega_{11} - \omega_{12}^2/\omega_{22}} \right), \tag{10}$$

where \mathcal{R} is defined in Eq. (23) in Appendix B.

(c) If Π is unknown and Assumption LI holds, the power envelope for the class of unbiased tests invariant to the group \mathcal{G} is given by

$$P_{\beta, \Pi} \left(\frac{(\Pi'S_\beta)^2}{\sigma_0^2 \Pi'Z'Z\Pi} > c_\alpha \right) = 1 - \kappa \left(c_\alpha; \frac{\Pi'Z'Z\Pi(\beta - \beta_0)^2}{\sigma_0^2} \right). \tag{11}$$

Comments: 1. If $\Pi = 0$, the parameter β is unidentified and the power envelope for unbiased tests is $\alpha = 1 - \kappa(c_\alpha; 0)$. Hence, the power upper bounds derived under the assumption that $\Pi \neq 0$ are not only valid, but are also continuous at $\Pi = 0$.

2. The power envelope in (10) is an upper bound for the power (11). Since $\omega_{11} - \omega_{12}^2/\omega_{22}$ is no larger than σ_0^2 , insisting on similarity/unbiasedness lowers the attainable power of the test. Alternatively, the optimal test for known Π can be understood as the optimal unbiased test when the nuisance-parameter set \mathbf{P} contains only one element; the loss in power is then due to an increase in the nuisance parameter space (in the sense of \mathbf{P} containing a k -dimensional rectangle).

3. Only in the case in which $k = 1$ and the model is exactly identified does an optimal result exist (which does not necessarily

require invariance to the group \mathcal{G}). Then, the Anderson–Rubin AR_β test is the Uniformly Most Powerful Unbiased (UMPU) test and has exact power function given by (9). The power function in (9) can also be seen as a result of the power envelope in (11) for the special case of a just-identified model.

3. When $k > 1$ and the model is over-identified, there exists no optimal test. Interestingly, the point-optimal test does not depend on the alternative β or on the quality of the instruments (if Π is multiplied by a constant, the power increases, but the point-optimal unbiased test remains the same). However, the point-optimal test depends on the direction of Π (when thought of as a vector).

4. Since the relative importance of the instruments (the direction of Π) is often unknown, Andrews et al. (2006) focus on tests that are invariant to orthogonal transformations of the instruments. In practice, these invariant tests depend on the $2k$ -dimensional sufficient statistics S_β and T_β only through the three-dimensional statistics $\bar{S}'_\beta \bar{S}_\beta$, $\bar{S}'_\beta \bar{T}_\beta$, and $\bar{T}'_\beta \bar{T}_\beta$. Applying Theorem 1 to these invariant tests, they show that the CLR test is (nearly) optimal within the class of invariant unbiased tests. Thus, the CLR test should be the test used in practice, as long as the applied researcher cannot distinguish the relative importance of the instruments.

3.1. Pre-testing

Although pre-tests are commonly used in econometrics, the fact that the first step typically affects the size of the second-step test is usually ignored.¹ As a positive result, the Similarity Condition yields the following results:

Proposition 1. Let \mathcal{A}_i , $i \in N$, be a partition of sets, each one belonging to $\sigma(T_\beta)$, the σ -algebra generated by T_β , and let $\phi_i(S_\beta, T_\beta)$, $i \in N$, be a sequence of α -similar tests invariant to the group of transformations \mathcal{G} . Finally, let

$$\phi = \sum_{i \in N} I_{\mathcal{A}_i} \phi_i$$

where $I_{\mathcal{A}}$ is the indicator function taking the value one if the outcome of the experiment ω is in \mathcal{A} , and zero otherwise. Then ϕ is also a similar test invariant to the group \mathcal{G} at level α .

Corollary 1. Let $h(T_\beta)$ be a measurable real-valued function and let $\phi_1(S_\beta, T_\beta)$ and $\phi_2(S_\beta, T_\beta)$ be two similar tests invariant to the group \mathcal{G} at level α . Finally, let

$$\phi_3 = I[h(T_\beta) > c] \phi_1 + I[h(T_\beta) \leq c] \phi_2,$$

where I is the indicator function taking the value one if the argument is true and zero otherwise. Then ϕ_3 is also a similar test at level α .

Comments: 1. The proposition asserts that choosing among similar tests leads to a similar test as long as pre-testing is based on the information associated with T_β . Of course, pre-testing based on $\hat{\Pi}$ does not create size distortions either, since $\hat{\Pi}$ is a one-to-one function of T_β .

2. One possible application of the proposition is instrument selection, where the researcher chooses among a countable number of similar tests (each one representing a different combination of the instruments). Finding some linear combination between the instruments or choosing the number of instruments based on $\hat{\Pi}$, it is possible to improve power without creating size distortions.

¹ For example, testing whether the covariance between the structural disturbances u and v_2 equals zero, $H_{\sigma_{u2}} : \sigma_{u2} = 0$, is equivalent to testing $H_\beta : \beta = \beta_0$, where $\beta_0 = \omega_{12}/\omega_{11}$. This implies that doing pre-testing on the parameter σ_{u2} can have serious consequences when making inference on the structural parameter β .

3. The corollary has some implications for applied research. For example, the score test is known to have poor power when $\|\hat{\Pi}\|$ is small. Thus, one might decide to use the Anderson–Rubin test if $\hat{\Pi}$ is near the origin and the score test if $\hat{\Pi}$ is far from the origin. If the decision is based on the reduced-form “F-statistic” $a'_0 \Omega^{-1} a_0 \cdot \hat{\Pi}' Z' Z \hat{\Pi}$, the pre-testing procedure is valid.

4. The corollary also connects and builds on previous work. For example, Zivot et al. (1998) use the first-stage F-statistic to select between testing procedures, but do not take into consideration the effect of pre-testing. On the contrary, pre-testing (or instrument selection) based on the constrained maximum likelihood estimator $\hat{\Pi}$ does not cause any difficulties with similar tests.

4. Multiple endogenous variables

The theory developed in the previous section for testing

$$H_\beta : \beta = \beta_0 \quad \text{vs.} \quad K_\beta : \beta \neq \beta_0$$

can easily be extended to accommodate more than two endogenous variables, as long as inference is conducted simultaneously on the coefficients of all endogenous variables. The sufficient statistic is once more given by $Z'Y$ and $X'Y$. The second statistic can again be eliminated by restricting attention to tests that are invariant to the group of linear transformations \mathcal{G} . As in the scalar case, the maximal invariant is $Z'Y$. However, for any known nonsingular, non-random $(l + 1) \times (l + 1)$ matrix D , $Z'YD$ is also a maximal invariant sufficient statistic. A convenient choice is the matrix

$$D = [b_0, \Omega^{-1} a_0],$$

where b_0 is the $(l + 1) \times 1$ vector $[1, -\beta'_0]'$, and a_0 is the $(l + 1) \times l$ matrix $[\beta_0, I_l]'$. By construction, every column of a_0 is orthogonal to b_0 . Then the maximal invariant sufficient statistic can be represented by the pair

$$S_\beta = Z'Yb_0 = Z'(y_1 - y_2\beta_0) \quad \text{and} \quad T_\beta = Z'Y\Omega^{-1}a_0. \quad (12)$$

As in the scalar case, the statistic S_β is pivotal and independent of the statistic T_β . This $k \times l$ -dimensional statistic T_β is a one-to-one function of the constrained maximum likelihood estimator $\hat{\Pi}$,

$$\hat{\Pi} = (Z'Z)^{-1} T_\beta (a'_0 \Omega^{-1} a_0)^{-1},$$

and is sufficient under the null hypothesis for Π . Here it is assumed that there is limited information about Π :

Assumption LI2 (Limited Information). The vector $\text{vec}(\Pi)$ can take arbitrary values over a $k \cdot l$ -dimensional rectangle.

In a limited-information model, it is again possible to characterize the entire class of similar tests for $H_\beta : \beta = \beta_0$.

Theorem 3. If Assumption LI2 holds, then any test invariant to the group of transformations \mathcal{G} is α -similar if and only if $E_0\phi(S_\beta, t) = \alpha$, a.e. \mathcal{P}^{T_β} .

Assumption LI2 is basically the multivariate version of Assumption LI considered when β is scalar. Unlike the scalar case, however, the assumption that $\text{vec}(\Pi)$ can take arbitrary values may be too strong. In applied research, there may be cases with restrictions on how the instruments affect the endogenous variables. It is then possible to use these restrictions to construct similar tests for subsets of the parameter β .

4.1. Testing parameter subsets of the endogenous variables

The interest here is in testing a subset of the parameters of the endogenous variables, β . Without loss of generality, consider the problem of testing the first l_1 elements of β , and treating the last $l_2 = l - l_1$ elements of β as nuisance parameters. That is, for $\beta = [\beta'_1, \beta'_2]'$, the goal is to test

$$H_{\beta_1} : \beta_1 = \beta_{1,0} \quad \text{vs.} \quad K_{\beta_1} : \beta_1 \neq \beta_{1,0}.$$

The structural equation (1) can then be written as

$$y_1 = y_{21}\beta_1 + y_{22}\beta_2 + X\gamma + u, \tag{13}$$

where $y_2 = [y_{21}, y_{22}]$ is partitioned conformably with $\beta = [\beta_1, \beta_2]$.

To construct tests with asymptotically correct size for β_1 , Stock and Wright (2000) propose the assurance of strong identification on β_2 , the parameter of the excluded endogenous variables. Dufour (1997) instead proposes to construct confidence regions for β , and uses projections to obtain valid confidence regions and tests for β_1 . The first approach typically does not lead to tests with correct size if identification is weak on the excluded endogenous variables, whereas the second approach may entail considerable loss of power. A third method can be obtained in the full-information model. Let the matrices Z_1 and Z_2 be respectively the first k_1 columns and the last $k_2 = k - k_1$ columns of the instrument Z . The underlying stochastic equation (3) for y_2 can be written as

$$\begin{aligned} y_{21} &= X\Gamma_1 + Z_1\Pi_{11} + Z_2\Pi_{21} + v_{21} \\ y_{22} &= X\Gamma_2 + Z_1\Pi_{12} + Z_2\Pi_{22} + v_{22}, \end{aligned} \tag{14}$$

where the matrices $\Gamma = [\Gamma_1, \Gamma_2]$ and

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix}$$

are partitioned conformably to $[y_{21}, y_{22}]$ and $[Z_1, Z_2]$. It is assumed that the instruments Z_2 do not affect y_{22} , and that there is limited information on how Z_2 affects y_{21} :

Assumption FI (Full Information). The instruments' coefficient Π_{22} is zero, and the vector $\text{vec}(\Pi_{21})$ can take arbitrary values over a $k_2 \cdot l_1$ -dimensional rectangle.

The zero restriction $\Pi_{22} = 0$ in Assumption FI implies that Assumption LI2 breaks down.² This allows the construction of similar tests for β_1 with non-trivial power. The assumption of limited information on Π_{21} , the effect of Z_2 on y_{21} , is not important to find similar tests. However, this assumption is crucial for obtaining optimality results.

Under Assumption FI, the reduced-form model can be written as

$$y_1 = X(\Gamma_1\beta_1 + \Gamma_2\beta_2 + \gamma) + Z_1\Pi_{11}\beta_1 + Z_2\Pi_{21}\beta_1 + Z_1\Pi_{12}\beta_2 + v_1$$

$$y_{21} = X\Gamma_1 + Z_1\Pi_{11} + Z_2\Pi_{21} + v_{21}$$

$$y_{22} = X\Gamma_2 + Z_1\Pi_{12} + v_{22},$$

where the known variance matrix

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \tag{15}$$

is partitioned conformably with $Y_1 = [y_1, y_{21}]$ and y_{22} . Here, the instrument Z_2 affects $Y_1 = [y_1, y_{21}]$, but not y_{22} . This implies that the sufficient statistic is given by $X'Y$, $Z_1'Y$ and $Z_2'Y_1D$, where $Z_2^* = M_{Z_1}Z_2$. A convenient choice for a non-singular 2×2 matrix is

$$D = [b_{1,0}, \Omega_{11}^{-1}a_{1,0}],$$

where $b_{1,0}$ is the $(l_1 + 1) \times 1$ vector $[1, -\beta'_{1,0}]'$, and $a_{1,0}$ is the $(l_1 + 1) \times l_1$ matrix $[\beta_{1,0}, I_{l_1}]'$. In this case, the sufficient statistic is $X'Y$, $Z_1'Y$, and the pair

$$S_{\beta_1} = Z_2^*Y_1b_{1,0} \quad \text{and} \quad T_{\beta_1} = Z_2^*Y_1\Omega_{11}^{-1}a_{1,0}. \tag{16}$$

These two statistics are independent and normally distributed under both the null and alternative hypotheses. Specifically,

$$S_{\beta_1} \sim N(Z_2^*Z_2^*\Pi_{21}(\beta_1 - \beta_{1,0}), b'_{1,0}\Omega_{11}b_{1,0} \cdot Z_2^*Z_2^*), \quad \text{and}$$

$$T_{\beta_1} \sim N(Z_2^*Z_2^*\Pi_{21} \cdot a'_{1,0}\Omega_{11}^{-1}a_{1,0}, a'_{1,0}\Omega_{11}^{-1}a_{1,0} \otimes Z_2^*Z_2^*),$$

where $a_1 = [\beta_1, I_{l_1}]'$. The following result derives a class of similar tests for testing $H_{\beta_1} : \beta_1 = \beta_{1,0}$.

Theorem 4. *If Assumption FI holds, then a test $\phi(S_{\beta_1}, T_{\beta_1})$ is similar at size α if and only if $E_0\phi(S_{\beta_1}, t) = \alpha$ for almost every value of the $k_2 \times l_1$ matrix t .*

Comment: The theorem establishes the whole class of similar tests depending on S_{β_1} and T_{β_1} . The assumption that these tests do not depend on $X'Y$ seems reasonable, since the focus is on tests invariant to the group \mathcal{G} . Ruling out the possibility that tests may depend on $Z_1'Y$ seems more controversial. However, an easy way to use the information on $Z_1'Y$ to improve power is not readily available.

Examples of similar tests within this class are Examples A–C based on the standardized statistics

$$\bar{S}_{\beta_1} = (Z_2^*Z_2^*)^{-1/2} S_{\beta_1} \cdot (b'_{1,0}\Omega_{11}b_{1,0})^{-1/2} \quad \text{and}$$

$$\bar{T}_{\beta_1} = (Z_2^*Z_2^*)^{-1/2} T_{\beta_1} \cdot (a'_{1,0}\Omega_{11}^{-1}a_{1,0})^{-1/2}.$$

For example, the (pseudo) CLR test for H_{β_1} rejects the null when

$$LR_{\beta_1} > c_{LR}(\bar{T}_{\beta_1}; \beta_{1,0}, \alpha).$$

Finally, Theorem 4 can be used to derive power upper bounds for the important case in which β_1 is a scalar.

Theorem 5. *Consider testing $H_{\beta_1} : \beta_1 = \beta_{1,0}$ against $K_{\beta_1} : \beta_1 \neq \beta_{1,0}$, with $\Pi_{21} \neq 0$, and let $\sigma_{\beta_{1,0}}^2 = b'_{1,0}\Omega_{11}b_{1,0}$.*

(a) *If the model is just-identified, the uniformly most powerful unbiased test based on S_{β_1} and T_{β_1} has a power function given by*

$$P_{\beta, \Pi}(AR_{\beta_1} > c_\alpha) = 1 - \kappa \left(c_\alpha; \frac{\Pi_{21}'Z_2^*Z_2^*\Pi_{21}(\beta_1 - \beta_{1,0})^2}{\sigma_{\beta_{1,0}}^2} \right).$$

(b) *If Π_{21} is unknown and Assumption FI holds, the power envelope for the class of unbiased tests based on S_{β_1} and T_{β_1} is given by*

$$\begin{aligned} P_{\beta, \Pi} \left(\frac{(\Pi_{21}'S_{\beta_1})^2}{\sigma_{1,0}^2 \Pi_{21}'Z_2^*Z_2^*\Pi_{21}} > c_\alpha \right) \\ = 1 - \kappa \left(c_\alpha; \frac{\Pi_{21}'Z_2^*Z_2^*\Pi_{21}(\beta_1 - \beta_{1,0})^2}{\sigma_{1,0}^2} \right). \end{aligned} \tag{17}$$

Comment: This result is parallel to Theorem 2. When the model is just-identified, there exists an optimal test based on S_{β_1} and T_{β_1} . When the model is over-identified, the point-optimal test depends on the alternative β and on the direction of the vector Π_{21} .

5. Weak-instrument asymptotics

The finite-sample results hold under the weak-instrument (WIV) asymptotics of Staiger and Stock (1997) with unknown covariance Ω and possibly nonnormal errors. For example, consider the following high-level assumptions for the simultaneous equations model with one endogenous explanatory variable.

Assumption WIV-LI. (i) $\Pi = C/n^{1/2}$ for some vector C that can take any values in a k -dimensional rectangle; (ii) For all $n \geq 1$, β is a fixed constant; (iii) $Z'Z/n \xrightarrow{p} D_Z > 0$; (iv) $V'V/n \xrightarrow{p} \Omega > 0$; and (v) $Z'V/\sqrt{n} \xrightarrow{d} N(0, \Omega \otimes D_Z)$.

² The assumption that $\Pi_{22} = 0$ may appear too strong at first. However, by transforming the instruments, we can write linear restrictions on Π_{21} and Π_{22} as zero restrictions.

Assumption WIV-LI (i) is the asymptotic version of Assumption LI. Parts (ii)–(iv) are similar to the assumptions of Staiger and Stock (1997), Moreira (2003), and Andrews et al. (2006). Because Ω is unknown, it must be replaced by a consistent estimator, e.g., $\widehat{\Omega}_n = Y'MY/(n - k - m)$

in the construction of each test. This plug-in method yields feasible versions of test $\bar{\phi}(Z'Y, Z'Z, \Omega, \beta_0)$, which are given by $\bar{\phi}(Z'Y, Z'Z, \widehat{\Omega}_n, \beta_0)$. As the next result shows, these feasible tests are asymptotically similar under Assumption WIV-LI.

Lemma 1. Under Assumption WIV-LI,

(a) $n^{-1/2}Z'Y \xrightarrow{d} \mathcal{N}_Z \sim N(D_Z C a', \Omega \otimes D_Z)$ and $(S_\beta, T_\beta) \xrightarrow{d} (\mathcal{S}_\beta, \mathcal{T}_\beta)$, where \mathcal{S}_β and \mathcal{T}_β are independent with marginal distributions given by

$$\mathcal{S}_\beta \sim N(D_Z C(\beta - \beta_0), D_Z \cdot b'_0 \Omega b_0) \quad \text{and}$$

$$\mathcal{T}_\beta \sim N(D_Z C \cdot a' \Omega^{-1} a_0, D_Z \cdot a'_0 \Omega^{-1} a_0).$$

(b) Let $P_{\beta, C, \Omega}(\cdot)$ denote probability when the true parameters are β , C , and Ω . For any function $\bar{\phi}$ that satisfies the homogeneity condition

$$\bar{\phi}(Z'Y, Z'Z, \Omega, \beta_0) = \bar{\phi}(n^{-1/2}Z'Y, n^{-1}Z'Z, \Omega, \beta_0) \quad (18)$$

and the discontinuity condition

$$P_{\beta, C, \Omega}((\mathcal{N}_Z, D_Z, \Omega, \beta_0) \in \mathbf{D}) = 0 \quad (19)$$

where \mathbf{D} is the set of discontinuity points of $\bar{\phi}$, then

$$\bar{\phi}(Z'Y, Z'Z, \widehat{\Omega}_n, \beta_0) \xrightarrow{d} \bar{\phi}(\mathcal{N}_Z, D_Z, \Omega, \beta_0).$$

(c) Let $E_{\beta, C, \Omega}(\cdot)$ denote expectation when the true parameters are β , C , and Ω . For any function $\bar{\phi}$ satisfying (18) and (19),

$$\lim_{n \rightarrow \infty} E_{\beta, \Pi, \Omega} \bar{\phi}(Z'Y, Z'Z, \widehat{\Omega}_n, \beta_0) = E_{\beta, C, \Omega} \bar{\phi}(\mathcal{N}_Z, D_Z, \Omega, \beta_0).$$

Comments: 1. The homogeneity and discontinuity conditions of part (b) hold for the AR, LM, and CLR tests, as well as for the point-optimal unbiased test used to construct the power envelope in Theorem 2(c).

2. Part (c) shows that the asymptotic power function of a feasible test for general error distributions coincides with the finite-sample power function of its feasible counterpart when errors are normally distributed with known variance.

The finite-sample results derived with normal errors with known variance are also helpful in characterizing asymptotically similar tests and deriving asymptotic power envelopes. The following result provides asymptotic versions of the Similarity Condition (Theorem 1) and of the power envelope for unbiased tests (Theorem 2-c).

Define the transformation $h_\Omega(\cdot)$, from \mathcal{N}_Z to $[\mathcal{S}_\beta : \mathcal{T}_\beta]$. This transformation is one-to-one and yields the following equivalence:

$$\begin{aligned} \bar{\phi}(\mathcal{N}_Z, D_Z, \Omega, \beta_0) &= \bar{\phi}(h_\Omega^{-1}(\mathcal{S}_\beta, \mathcal{T}_\beta), D_Z, \Omega, \beta_0) \\ &= \phi(\mathcal{S}_\beta, \mathcal{T}_\beta, D_Z, \Omega, \beta_0). \end{aligned}$$

Let $E_{0, \Omega}$ denote the expectation over the null distribution of S_β .

Theorem 6. Under Assumption WIV-LI,

(a) If $\bar{\phi}(Z'Y, Z'Z, \widehat{\Omega}_n, \beta_0)$ satisfies the homogeneity and discontinuity conditions (18) and (19), then $E_{\beta_0, C, \Omega} \bar{\phi}(\mathcal{S}_\beta, \mathcal{T}_\beta, D_Z, \Omega, \beta_0) = \alpha$ if and only if $E_{0, \Omega} \bar{\phi}(\mathcal{S}_\beta, t, D_Z, \Omega, \beta_0) = \alpha$, a.e. $\mathcal{P}^{\mathcal{T}_\beta}$.

(b) For normal errors, an asymptotic power bound for asymptotically unbiased tests invariant to the group \mathcal{G} is given by $1 - \kappa$

$$\left(c_\alpha; \frac{C'D_Z C(\beta - \beta_0)^2}{\sigma_0^2} \right).$$

Comments: 1. Part (a) does not require normal errors; only Assumption WIV-LI is necessary. It establishes that a test is asymptotically similar if and only if it is asymptotically similar for each level of $\mathcal{T}_\beta = t$.

2. The power upper bound in part (b) is sharp. For β and $\Pi = C/n^{1/2}$, the power envelope is attained by the point-optimal test that rejects the null when

$$\frac{(C'S_\beta)^2}{\widehat{\sigma}_0^2 C'Z'ZC} > c_\alpha,$$

where $\widehat{\sigma}_0^2 = b'_0 \widehat{\Omega}_n b_0$. If the model is over-identified, this test depends on the quantity C that is not consistently estimable. Therefore, no test can uniformly reach the power envelope under the local-to-zero asymptotics in which β is fixed and $\Pi = C/n^{1/2}$.

6. Monte Carlo simulations

This section considers power comparisons when there is only one endogenous variable. Similar results hold for testing one endogenous variable's coefficient in the presence of multiple endogenous variables. A 1000-replication experiment is performed with the hypothesized value β_0 set to zero. The instruments Z are held fixed such that $Z'Z = I_k$, where $k = 5$. The rows of $[u, v_2]$ are i.i.d. normal random vectors with unit variances and correlation ρ . Different values of the instruments' strength ($\lambda = 1, 2, 4, \text{ and } 10$) and different degrees of endogeneity of y_2 ($\rho = 0$ and 0.5) are considered. The rows of $[u, v_2]$ are i.i.d. normal random vectors with unit variances and correlation ρ .

Figs. 1a and 1b plot the rejection probability of the Anderson–Rubin, score, and CLR tests as functions of the true value β .³ In light of results of Section 5, these power plots can be understood as numerically approximating asymptotic power functions under the weak-instrument asymptotics.

In each figure, the power curves are at the 5% level when β equals β_0 . This reflects the fact that these tests are similar. The power envelope, computed from the rejection probability of point-optimal tests, is also included. Although point-optimal tests depend on the direction of Π , the power envelope (11) depends only on the “population” first-stage F-statistic. Hence, power comparisons in Figs. 1a and 1b are quite general, independent of the direction of Π that generated the data.

For small values of λ , the CLR test's power is lower than the power envelope. This reflects the fact that the power envelope uses a point-optimal test that depends on a particular value of Π . However, the direction of Π cannot be estimated accurately when instruments are weak. To avoid this problem, Andrews et al. (2006) propose to look at a smaller class of tests, invariant to orthogonal transformations of the instruments. They show that the CLR test is nearly optimal for this class of tests, regardless of the degree of identification.

For large values of λ , the CLR test has power essentially equal to the power envelope given in (11), but the Anderson–Rubin and score tests do not. This result provides a refinement over the first-order asymptotics. The CLR test not only has optimality results for local-to-null alternatives but also dominates the Anderson–Rubin and score tests for fixed alternatives under strong identification.

Therefore, the CLR test is optimal in two cases: (i) when λ is large; and (ii) when λ is small and the instruments should be treated similarly. The Anderson–Rubin and score tests are not optimal in either case. Only when instruments are weak and there is some previous knowledge of the relative strength of the instruments (the direction of Π), may there be power improvements beyond the CLR test. This can be done, for example, by using the results of Section 3.1.

7. Conclusions and extensions

For a fixed value of Π different from zero and a large enough sample size n , the standard first-order asymptotics provide a

³ As β varies, ω_{11} and ω_{12} change to keep the structural error variance and the correlation between u and v_2 constant.

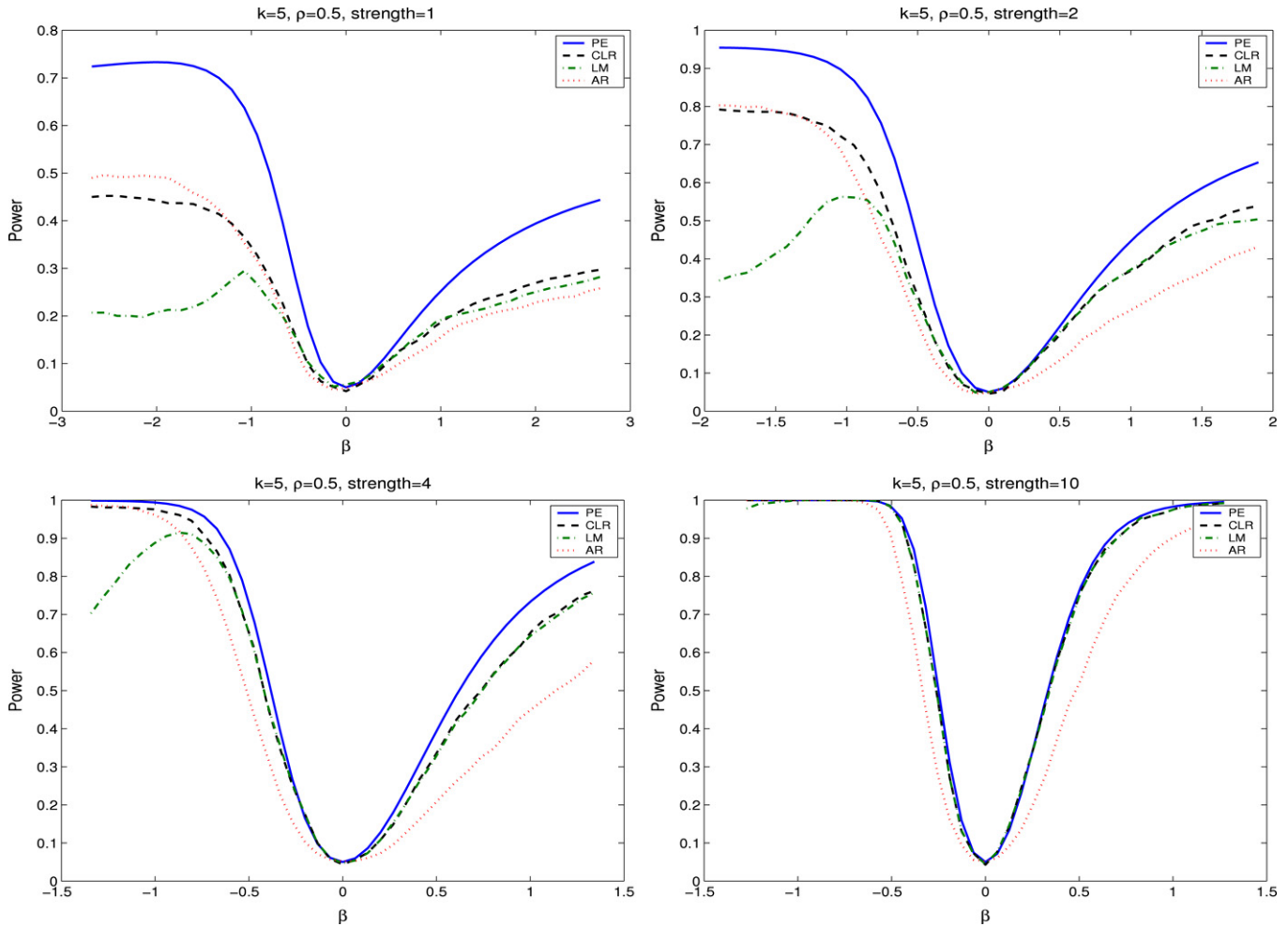


Fig. 1b. Power curves of tests for H_β when $\rho = 0.5$.

is a special case of a testing procedure proposed by Breusch and Pagan (1980). Among many examples discussed by Breusch and Pagan (1980), there is the non-linear regression (p. 242-3) and seemingly unrelated equations (p. 251, footnote 3). But the reduced-form model given in (4) is just a very special case of SUR models, in which there is a restriction between the coefficients in the equations.

Given this restriction, the model can be written by stacking the matrices as follows:

$$\text{vec}(Y) = Z(\beta)\Pi + \text{vec}(V),$$

where $Z(\beta) = [Z'\beta, Z']'$ is a $2n \times k$ matrix (for simplicity, the effect of X on the endogenous variables is omitted here). Following the notation adopted by Breusch and Pagan (1980), there is a vector of residuals,

$$\tilde{e} = \text{vec}(Y) - Z(\beta_0)\hat{\Pi},$$

and a $2n \times (k + 1)$ matrix of the derivative of the mean $Z(\beta)\Pi$ with respect to the scalar β and the k -dimensional vector Π ,

$$\tilde{G} = \begin{bmatrix} Z\beta_0 & Z\hat{\Pi} \\ Z & 0 \end{bmatrix}.$$

One of the many score tests of Breusch and Pagan (1980) is given by

$$\tilde{e}' [\Omega^{-1} \otimes I_n] \tilde{G} [\tilde{G}' [\Omega^{-1} \otimes I_n] \tilde{G}]^{-1} \tilde{G}' [\Omega^{-1} \otimes I_n] \tilde{e}.$$

The score is the $(k + 1)$ -dimensional vector

$$\tilde{G}' [\Omega^{-1} \otimes I_n] \tilde{e},$$

where the first k rows equal zero when evaluated at $(\beta_0, \hat{\Pi})$. The $(k + 1)$ -th row equals $-\hat{\Pi}'S/\sigma_0^2$. The lower-right quadrant of $[\tilde{G}' [\Omega^{-1} \otimes I_n] \tilde{G}]^{-1}$ simplifies to $\hat{\Pi}'Z'Z\hat{\Pi}/\sigma_0^2$. Therefore, one particular score test (which uses $\tilde{G}' [\Omega^{-1} \otimes I_n] \tilde{G}$ as an estimator for the asymptotic variance of the score) proposed by Breusch and Pagan (1980) simplifies to LM_β in the simultaneous equations model.

Appendix B. Proofs

Some proofs below are based on the following result. A maximal invariant in terms of the sufficient statistic $(X'Y, Z'Y)$ for the group of transformations \mathcal{g} is simply $Z'Y$. Recall that any invariant test is a function of the maximal invariant. Therefore, any test invariant to \mathcal{g} depends on $Z'Y$.

Some results use the following two lemmas proved in Lehmann (1986):

Lemma 2. Let \mathcal{X} be a random vector with probability distribution

$$dP_\theta(x) = C(\theta) \exp \left[\sum_{j=1}^J \nu_j T_j(x) \right] d\mu(x)$$

and let \mathcal{P}^T be the family of distributions of $T = (T_1(X), \dots, T_J(X))$ as $\mathcal{V} = (\mathcal{V}_1, \dots, \mathcal{V}_J)$ ranges over the set W . Then \mathcal{P}^T is complete provided W contains a J -dimensional rectangle.

Lemma 3. Suppose that the distribution of X is given by

$$dP_{\theta, \mathcal{V}}(x) = C(\theta, \mathcal{V}) \exp \left[\theta R(x) + \sum_{j=1}^J \mathcal{V}_j T_j(x) \right] d\mu(x)$$

where the V_j are the nuisance parameters and μ is absolutely continuous with respect to the Lebesgue measure. Suppose that $S = h(R, T)$ is independent of the random vector T when $\theta = \theta_0$ and that $h(r, t) = a(t)r + b(t)$ with $a(t) > 0$.

Then the Uniformly Most Powerful Unbiased (UMPU) test ϕ for $H_0 : \theta = \theta_0$ against $K_\theta : \theta \neq \theta_0$ is given by

$$\phi(s) = \begin{cases} 1 & \text{if } s < C_1 \text{ or } s > C_2 \\ 0 & \text{otherwise} \end{cases}$$

where C_1 and C_2 are determined by $E_0\phi(S) = \alpha$ and $E_0\phi(S)S = \alpha E_0S$.

Proof of Theorem 1. Since randomization is allowed, without loss of generality any test can be written as $\phi(S_\beta, T_\beta)$. Since the test is similar at size α , it must be true that

$$E_{\beta_0, \Pi} \phi(S_\beta, T_\beta) = \alpha, \quad \forall \Pi \in \mathbf{P}. \tag{20}$$

where $E_{\beta, \Pi}$ represents expectation with the parameters β and Π . By Lemma 2, the family of distributions of T_β when the null hypothesis is true, $\mathcal{P}^{T_\beta} = \{P_{\beta_0, \Pi}^{T_\beta}; \Pi \in \mathbf{P}\}$, is complete. Consequently,

$$E_{\beta_0, \Pi} \{ \phi(S_\beta, T_\beta) | T_\beta = t \} = \alpha, \quad a.e. \mathcal{P}^{T_\beta}. \tag{21}$$

Note that the distribution of S_β does not depend on Π under the null hypothesis. By Basu's lemma (e.g., Lehmann (1986)), S_β is independent of T_β under the null hypothesis (in fact, S_β is also independent of T_β under the alternative hypothesis). Using the fact that $\phi(S_\beta, T_\beta)$ is integrable:

$$E_0\phi(S_\beta, t) = \alpha, \quad a.e. \mathcal{P}^{T_\beta}. \tag{22}$$

Conversely, if the test is such that (21) holds, then (20) is trivially true. Therefore, the test is similar at size α . \square

Proof of Theorem 2. The following is true:

a. For some measure μ , the probability distribution of Y can be written as

$$dP_{\theta, \Pi}(y) = C(\theta, \Pi) \exp \left[\theta R_\beta(y) + \Pi T_\beta(y) + \text{tr} \left((\Gamma a' + \gamma e_1)' X' Y \Omega^{-1} \right) \right] d\mu(y)$$

where $R_\beta(Y)$ is the first column of $Z'Y\Omega^{-1}$ and $\theta = \Pi(\beta - \beta_0)$. Since \mathbf{P} does not contain the origin and the model is just identified, testing $H_\beta : \beta = \beta_0$ against $K_\beta : \beta \neq \beta_0$ is equivalent to testing $H_\theta : \theta = \theta_0$ against $K_\theta : \theta \neq \theta_0$. Notice that $\bar{S}_\beta = \delta_1 R_\beta + \delta_2 T_\beta$ where

$$\delta_1 = \sigma_0 (Z'Z)^{-1/2} \quad \text{and} \quad \delta_2 = \frac{-\omega_{22}\beta_0 + \omega_{12}}{\sigma_0} (Z'Z)^{-1/2}.$$

Now Lemma 3 can be applied with $T = (T_\beta, \text{vec}(X'Y)')$. Since $\bar{S}_\beta \sim N(0, 1)$ under the null and, in particular, it is symmetric around zero, it is straightforward to show that the optimal test rejects the null if $AR_\beta > c_\alpha$. Under the alternative β ,

$$AR_\beta \sim \chi^2 \left(1, \frac{\Pi'Z'Z\Pi(\beta - \beta_0)^2}{\sigma_0^2} \right).$$

Consequently, the power of the UMPU test is given by (9).

The proof of UMPU existence for invariant tests is analogous.
b. Because Π is known, the likelihood function of $Z'Y$ can be written as

$$L(Z'Y; \beta) = C(\beta, \Pi) \exp \left[\beta \cdot \Pi' R_\beta \right] d\mu(Z'Y)$$

for some measure μ . This distribution is a one-parameter exponential family, and the UMPU test rejects the null hypothesis if

$$\mathcal{R} = \frac{\left\{ \Pi'Z' \left[(y_1 - Z\Pi\beta_0) - \omega_{12}\omega_{22}^{-1}(y_2 - Z\Pi) \right] \right\}^2}{(\omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{12}) \Pi'Z'Z\Pi} \tag{23}$$

is larger than c_α . Under the alternative β ,

$$\mathcal{R} \sim \chi^2 \left(1, \frac{\Pi'Z'Z\Pi}{(\omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{12})} (\beta - \beta_0)^2 \right).$$

Consequently, the power of the optimal test is given by (10).

c. The power of the test ϕ is given by $E_{\beta, \Pi} \phi(S_\beta, T_\beta)$. Any unbiased test must be locally unbiased:

$$\left. \frac{\partial E_{\beta, \Pi} \phi(S_\beta, T_\beta)}{\beta} \right|_{\beta=\beta_0} = 0, \quad \forall \Pi \in \mathbf{P}.$$

This condition simplifies to

$$\Pi'E_{\beta_0, \Pi} R_\beta \phi(S_\beta, T_\beta) - \Pi'E_{\beta_0, \Pi} R_\beta \alpha = 0, \quad \forall \Pi \in \mathbf{P}. \tag{24}$$

Because any unbiased test must be similar,

$$E_{\beta_0, \Pi} T_\beta \phi(S_\beta, T_\beta) = E_{\beta_0, \Pi} T_\beta E_{\beta_0, \Pi} \{ \phi(S_\beta, T_\beta) | T_\beta \} = E_{\beta_0, \Pi} T_\beta \alpha.$$

Because S_β is symmetric under the null and R_β is a linear combination of S_β and T_β , (24) simplifies to

$$\Pi'E_{\beta_0, \Pi} S_\beta \phi(S_\beta, T_\beta) = 0, \quad \forall \Pi \in \mathbf{P}.$$

Note that

$$E_{\beta_0, \Pi} S_\beta \phi(S_\beta, T_\beta) = E_0 S_\beta E_{\beta_0, \Pi} \{ \phi(S_\beta, T_\beta) | S_\beta \} = E_0 S_\beta \varphi(S_\beta, \Pi),$$

where $\varphi(s, \Pi) = E_{\beta_0, \Pi} \phi(s, T_\beta)$ is an analytic function of Π bounded by 0 and 1; see Theorem 2.7.1 of Lehmann (1986, p. 49). As a result,

$$E_{\beta_0, \Pi} S_\beta \phi(S_\beta, T_\beta) = 0, \quad \forall \Pi \in \mathbf{P}.$$

Because T_β is complete, it follows that

$$E_0 S_\beta \phi(S_\beta, t) = 0, \quad a.e. \mathcal{P}^{T_\beta}. \tag{25}$$

Therefore, there are $k + 1$ boundary conditions for unbiased tests: (22) and (25). Theorem 3.6.1 of Lehmann (1986, p. 77) asserts that a sufficient condition for a test to maximize power is the existence of a constant scalar d_1 and a k -dimensional vector d such that

$$\phi = I(e^{s' \Pi(\beta - \beta_0)} > d_1 + d's)$$

satisfies (22) and (25). For $d = d_2 \Pi(\beta - \beta_0)$, ϕ rejects the null when

$$e^{s' \Pi(\beta - \beta_0)} > d_1 + d_2 \cdot s' \Pi(\beta - \beta_0).$$

This region is either one-sided or outside of an interval. If one-sided, the power function is strictly monotone and cannot satisfy (25). As a result, the optimal test rejects the null when $S'_\beta \Pi < c_1$ or $S'_\beta \Pi > c_2$. Standardizing $\Pi'S_\beta$ yields the optimal test ϕ that satisfies (22) and (25). It rejects the null if

$$\frac{(\Pi'S_\beta)^2}{\sigma_0^2 \Pi'Z'Z\Pi} > c_\alpha.$$

Recall that $S_\beta \sim N(Z'Z\Pi^*(\beta - \beta_0), \sigma_0^2 Z'Z)$ where Π^* is the true value of the instruments' coefficients. Therefore, the statistic

$$\frac{(\Pi'S_\beta)^2}{\sigma_0^2 \Pi'Z'Z\Pi}$$

has a noncentral chi-square distribution and ϕ is indeed unbiased. If $\Pi = \Pi^*$, then

$$\frac{(\Pi'S_\beta)^2}{\sigma_0^2 \Pi'Z'Z\Pi} \sim \chi^2 \left(1, \frac{\Pi'Z'Z\Pi (\beta - \beta_0)^2}{\sigma_0^2} \right).$$

This gives the *power envelope* in (11). \square

Proof of Proposition 1. By the Monotone Convergence Theorem,

$$\begin{aligned} E_{\beta_0, \Pi} \phi(S_\beta, T_\beta) &= \sum_{i \in N} E_{\beta_0, \Pi} I_{\mathcal{A}_i} \phi_i(S_\beta, T_\beta) \\ &= \sum_{i \in N} E_{\beta_0, \Pi} \{ I_{\mathcal{A}_i} E_{\beta_0, \Pi} \{ \phi_i(S_\beta, T_\beta) | T_\beta \} \}, \end{aligned}$$

using the Law of Iterated Expectations and the fact that $I_{\mathcal{A}_i}$ is measurable with respect to $\sigma(T_\beta)$. By Theorem 1, ϕ_i 's are such that $E_0 \phi_i(S_\beta, t) = \alpha$. Therefore,

$$E_{\beta_0, \Pi} \phi(S_\beta, T_\beta) = \sum_{i \in N} E_{\beta_0, \Pi} \{ I_{\mathcal{A}_i} \alpha \} = \alpha \cdot \sum_{i \in N} P(\mathcal{A}_i) = \alpha. \quad \square$$

Proof of Corollary 1. Take $\mathcal{A}_1 = \{h(T) > c\}$, $\mathcal{A}_2 = I\{h(T) \leq c\}$. Because \mathcal{A}_1 and \mathcal{A}_2 form a partition and each set belongs to $\sigma(T)$, the result follows from Proposition 1. \square

Proof of Theorem 3. The likelihood function of $Z'Y$ under the null can be written as

$$L(Z'Y; \beta_0, \Pi) = C(\beta, \Pi) \exp[\text{vec}(\Pi)' \text{vec}(Z'Y \Omega^{-1} a_0)] d\mu(Z'Y)$$

for some measure μ . By Lemma 2, the family of distributions of $T_\beta = Z'Y \Omega^{-1} a_0$ when the null hypothesis is true, $\mathcal{P}^{T_\beta} = \{P_{\beta_0, \Pi}^{T_\beta}; \Pi \in \mathbf{P}\}$, is complete. The proof now follows analogously to Theorem 1. \square

Proof of Theorems 4 and 5. Note that

$$\begin{aligned} S_{\beta_1} &\sim N(Z_2^* Z_2^* \Pi_{21} (\beta_1 - \beta_{1,0}), b'_{1,0} \Omega_{11} b_{1,0} \cdot Z_2^* Z_2^*), \quad \text{and} \\ T_{\beta_1} &\sim N(Z_2^* Z_2^* \Pi_{21} \cdot a'_{1,0} \Omega_{11}^{-1} a_{1,0}, a'_{1,0} \Omega_{11}^{-1} a_{1,0} \otimes Z_2^* Z_2^*), \end{aligned}$$

with S_{β_1} being independent of T_{β_1} . The statistic S_{β_1} does not depend on Π_{21} under the null and T_{β_1} is complete under the null hypothesis H_{β_1} . The proof of Theorem 4 follows analogously to that of Theorem 3. When β_1 is a scalar, the derivation of the power envelope follows the same arguments used in Theorem 2. \square

Proof of Lemma 1. Part (a) is standard and proved in Lemma 4 of Andrews et al. (2006). Part (b) follows from the continuous mapping theorem. Part (c) follows from Part (b) and the fact that $0 \leq \hat{\phi} \leq 1$. \square

Proof of Theorem 6. For part (a), the test $\bar{\phi}(Z'Y, Z'Z, \hat{\Omega}_n, \beta_0)$ converges in distribution to $\phi(\delta_\beta, \mathcal{T}_\beta, D_Z, \Omega, \beta_0)$. Recall that $\mathcal{T}_\beta \sim N(D_Z C \cdot a'_0 \Omega^{-1} a_0, D_Z \cdot a'_0 \Omega^{-1} a_0)$ under the null. Because C can take any values over a k -dimensional rectangle, $\mathcal{P}^{\mathcal{T}_\beta}$ is complete. Hence, any similar test $\phi(\delta_\beta, \mathcal{T}_\beta, D_Z, \Omega, \beta_0)$ at level α must satisfy $E_0 \phi(\delta_\beta, t, D_Z, \Omega, \beta_0) = \alpha, a.e. \mathcal{P}^{\mathcal{T}_\beta}$. The converse is trivially true.

For part (b), suppose for now that Ω is known. The distribution of $n^{-1/2} Z'Y \xrightarrow{d} \mathcal{N}_Z \sim N(D_Z C a', \Omega \otimes D_Z)$. This distribution matches the finite-sample distribution of $Z'Y \sim N(Z'Z \Pi a', \Omega \otimes Z'Z)$, where D_Z and C are replaced by $Z'Z$ and Π , respectively. Hence, following the proof of Theorem 2(c), the optimal asymptotically unbiased test at β and C rejects the null when

$$\frac{(C'S_\beta)^2}{\sigma_0^2 C'Z'ZC} > c_\alpha.$$

When Ω is unknown, the sufficient statistic for normal errors with unknown variance is given by the independent statistics $Z'Y$ and $\hat{\Omega}_n$. Note that the class of asymptotically unbiased tests with known Ω contains the class of asymptotically unbiased tests when Ω is unknown. Therefore, the power envelope for asymptotically unbiased tests for unknown Ω must be smaller than the power envelope for asymptotically unbiased tests with known Ω . It remains to show that the asymptotic bound using unknown Ω is sharp. This is indeed the case since

$$\begin{aligned} \frac{(C'S_\beta)^2}{\hat{\sigma}_0^2 C'Z'ZC} &= \frac{(C'S_\beta)^2}{\sigma_0^2 C'Z'ZC} + o_p(1) \quad \text{and} \\ \frac{(C'S_\beta)^2}{\sigma_0^2 C'Z'ZC} &\xrightarrow{d} \frac{(C'\delta_\beta)^2}{\sigma_0^2 C'D_Z C}. \quad \square \end{aligned}$$

Appendix C. Figures

See Figs. 1a and 1b.

References

Anderson, T.W., Rubin, H., 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., Moreira, M.J., Stock, J.H., 2006. Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752.

Bound, J., Jaeger, D.A., Baker, R.M., 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.

Breusch, T.S., Pagan, A.R., 1980. The Lagrange multiplier test and its applications to model specifications in econometrics. *The Review of Economic Studies* 47, 239–253.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1388.

Dufour, J.-M., Jasiak, J., 2001. Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review* 42, 815–843.

Gleser, L.J., Hwang, J.T., 1987. The non-existence of 100(1- α)% confidence sets of finite expected diameter in errors-in-variables and related models. *The Annals of Statistics* 15, 1351–1362.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781–1803.

Lehmann, E.L., 1986. *Testing Statistical Hypothesis*, 2nd ed. In: *Wiley Series in Probability and Mathematical Analysis*.

Moreira, M.J., 2002. Tests with correct size in the simultaneous equations model. Ph.D. thesis. UC Berkeley.

Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.

Nelson, C.R., Startz, R., 1990. Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 58, 967–976.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stock, J.H., Wright, J., 2000. GMM with weak identification. *Econometrica* 68, 1055–1096.

Zivot, E., Startz, R., Nelson, C.R., 1998. Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 39, 1119–1144.