



Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: [www.elsevier.com/locate/jeconom](http://www.elsevier.com/locate/jeconom)

## 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](mailto: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  $\beta$ . 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,  $\Pi$  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  $\Omega$  will be relaxed later using the weak-instrument asymptotics of Staiger and Stock (1997).

The goal is to test (subsets of)  $\beta$ , treating  $\Pi$ ,  $\gamma$ , and  $\Gamma$  as nuisance parameters. A test is said to be of size  $\alpha$  if the probability of rejecting the null hypothesis when it is true does not exceed  $\alpha$ . That is,

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

where the sup is over all values of  $\beta$ ,  $\Pi$ ,  $\gamma$ , and  $\Gamma$  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  $\beta$ ,  $\Pi$ ,  $\gamma$ , and  $\Gamma$  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  $\alpha$  even when  $\Pi$  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  $\beta$  is a scalar,  $\Pi$  is a  $k \times 1$  vector, and  $\gamma$  and  $\Gamma$  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  $(\gamma, \Gamma)$  and  $(\beta, \Pi)$  are given by  $X'Y$  and  $Z'Y$ , respectively. The nuisance parameters  $\gamma$  and  $\Gamma$  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  $H_\beta$ ,  $g(Y) = Y + XF$  for arbitrary conformable matrices  $F$ , the maximal invariant in terms of the sufficient statistic is  $Z'Y$ . For any non-singular  $2 \times 2$  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  $S_\beta$  and  $T_\beta$  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  $S_\beta$  does not depend on the nuisance parameter  $\Pi$ , the null distribution of  $T_\beta$  is very sensitive to the value of  $\Pi$ . 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  $\Pi$  when  $\beta$  is constrained to take the null value  $\beta_0$ . The unknown parameter  $\Pi$  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  $\Pi$  lies contains a  $k$ -dimensional rectangle.

All tests invariant to the group  $\mathcal{G}$  can be written as (possibly randomized) functions of  $S_\beta$  and  $T_\beta$ . Specifically, let  $\phi$  be a critical function such that  $0 \leq \phi \leq 1$ . For each  $S_\beta$  and  $T_\beta$  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  $\phi$  on  $Z$ ,  $\beta_0$  and  $\Omega$  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  $S_\beta$  and let  $\mathcal{P}^{T_\beta}$  be the family of distributions of  $T_\beta$  under the null hypothesis. An event holds a.e.  $\mathcal{P}^{T_\beta}$  if it is true except on a set with probability zero for all null distributions of  $T_\beta$ . The following result characterizes all similar tests in terms of the marginal distribution of  $S_\beta$ .

**Theorem 1 (Similarity Condition).** *If Assumption LI holds, then any test invariant to the group  $\mathcal{G}$  is similar at size  $\alpha$  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.  $\mathcal{P}^{T_\beta}$ .*

**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  $T_\beta$ . Because  $S_\beta$  is pivotal under  $H_\beta$ , it is possible to find similar tests by looking at the marginal distribution of the statistic  $S_\beta$  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  $\Pi$  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  $H_\beta$  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  $\gamma$  and  $\Gamma$  can take any value over a  $m$ -dimensional rectangle. Then any test is similar at size  $\alpha$  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  $H_\beta$  if the Anderson–Rubin statistic for known  $\Omega$  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  $T_\beta$ . This, however, does not imply that the test itself does not depend on  $T_\beta$ , as the following example shows.

**Example B.** Reject  $H_\beta$  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  $\bar{T}_\beta$  is independent of  $\bar{S}_\beta$  and, consequently, the null distribution of  $LM_\beta$  is chi-square-one. Interestingly,  $LM_\beta$  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  $\alpha$ , then the conditional null rejection probability equals  $\alpha$ . 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  $H_\beta$  when the likelihood ratio statistic is larger than its conditional  $(1 - \alpha)$ -quantile:

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

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

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

where  $\sigma_0^2 = b'_0 \Omega b_0$  and  $\kappa(\cdot; \mu)$  is the noncentral  $\chi^2(1)$  distribution function with noncentrality parameter  $\mu$ . However, when  $\Pi$  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*

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

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

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

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

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

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

**Comments: 1.** If  $\Pi = 0$ , the parameter  $\beta$  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  $\sigma_0^2$ , insisting on similarity/unbiasedness lowers the attainable power of the test. Alternatively, the optimal test for known  $\Pi$  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  $AR_\beta$  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  $\beta$  or on the quality of the instruments (if  $\Pi$  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  $\Pi$  (when thought of as a vector).

4. Since the relative importance of the instruments (the direction of  $\Pi$ ) 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  $S_\beta$  and  $T_\beta$  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.<sup>1</sup> 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  $\sigma$ -algebra generated by  $T_\beta$ , and let  $\phi_i(S_\beta, T_\beta)$ ,  $i \in N$ , be a sequence of  $\alpha$ -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  $\omega$  is in  $\mathcal{A}$ , and zero otherwise. Then  $\phi$  is also a similar test invariant to the group  $\mathcal{G}$  at level  $\alpha$ .

**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  $\alpha$ . 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  $\phi_3$  is also a similar test at level  $\alpha$ .

**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  $T_\beta$ . Of course, pre-testing based on  $\hat{\Pi}$  does not create size distortions either, since  $\hat{\Pi}$  is a one-to-one function of  $T_\beta$ .

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.

<sup>1</sup> 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  $\sigma_{u2}$  can have serious consequences when making inference on the structural parameter  $\beta$ .

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  $S_\beta$  is pivotal and independent of the statistic  $T_\beta$ . This  $k \times l$ -dimensional statistic  $T_\beta$  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  $\Pi$ . Here it is assumed that there is limited information about  $\Pi$ :

**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  $\alpha$ -similar if and only if  $E_0\phi(S_\beta, t) = \alpha$ , a.e.  $\mathcal{P}^{T_\beta}$ .

Assumption LI2 is basically the multivariate version of Assumption LI considered when  $\beta$  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  $\beta$ .

### 4.1. Testing parameter subsets of the endogenous variables

The interest here is in testing a subset of the parameters of the endogenous variables,  $\beta$ . Without loss of generality, consider the problem of testing the first  $l_1$  elements of  $\beta$ , and treating the last  $l_2 = l - l_1$  elements of  $\beta$  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  $\beta_1$ , Stock and Wright (2000) propose the assurance of strong identification on  $\beta_2$ , the parameter of the excluded endogenous variables. Dufour (1997) instead proposes to construct confidence regions for  $\beta$ , and uses projections to obtain valid confidence regions and tests for  $\beta_1$ . 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  $\Pi_{22}$  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.<sup>2</sup> This allows the construction of similar tests for  $\beta_1$  with non-trivial power. The assumption of limited information on  $\Pi_{21}$ , 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  $2 \times 2$  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  $\alpha$  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  $S_{\beta_1}$  and  $T_{\beta_1}$ . 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  $H_{\beta_1}$  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  $\beta_1$  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  $S_{\beta_1}$  and  $T_{\beta_1}$  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  $\Pi_{21}$  is unknown and Assumption FI holds, the power envelope for the class of unbiased tests based on  $S_{\beta_1}$  and  $T_{\beta_1}$  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  $S_{\beta_1}$  and  $T_{\beta_1}$ . When the model is over-identified, the point-optimal test depends on the alternative  $\beta$  and on the direction of the vector  $\Pi_{21}$ .

### 5. Weak-instrument asymptotics

The finite-sample results hold under the weak-instrument (WIV) asymptotics of Staiger and Stock (1997) with unknown covariance  $\Omega$  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  $n \geq 1$ ,  $\beta$  is a fixed constant; (iii)  $Z'Z/n \xrightarrow{p} D_Z > 0$ ; (iv)  $V'V/n \xrightarrow{p} \Omega > 0$ ; and (v)  $Z'V/\sqrt{n} \xrightarrow{d} N(0, \Omega \otimes D_Z)$ .

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









