Journal of Econometrics 152 (2009) 131–140

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: 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.

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.

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.01.012

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

$$y_1 = y_2\beta + X\gamma + u, \tag{1}$$

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

$$y_2 = X\widetilde{\Gamma} + \widetilde{Z}\Pi + v_2, \tag{2}$$

where $\widetilde{Z}$ is an $n \times k$ matrix of exogenous variables having full column rank, Π is a $k \times l$ matrix, and $\widetilde{\Gamma}$ is an $m \times l$ matrix. For convenience, transform the matrix $\widetilde{Z}$ 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

$$y_2 = X\Gamma + Z\Pi + v_2, \tag{3}$$

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

$$\begin{aligned} y_1 &= X\left(\Gamma\beta + \gamma\right) + Z\Pi\beta + v_1 \\ y_2 &= X\Gamma + Z\Pi + v_2. \end{aligned} \tag{4}$$

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

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

where $a = [\beta, I_l]'$ and $e_1 = \left[1, 0_l'\right]'$. 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

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

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:

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

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, $0 \in \overline{\mathbf{P}}$. 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

$$\begin{aligned} y_1 &= X\left(\Gamma\beta + \gamma\right) + Z\Pi\beta + v_1 \\ y_2 &= X\Gamma + Z\Pi + v_2, \end{aligned} \tag{6}$$

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, $g\left(Y\right) = Y + XF$ 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 $D = \left[b_0, \Omega^{-1}a_0\right]$, where $b_0 = (1, -\beta_0)'$ is orthogonal to $a_0 = (\beta_0, 1)'$. This yields the pair

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

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,

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

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

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

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

where $\widehat{\Pi}$ 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 $\phi\left(S_\beta, T_\beta\right)$ and $1 - \phi\left(S_\beta, T_\beta\right)$, respectively; the dependence of ϕ on Z, β_0 and Ω is omitted out of convenience. For example, a nonrandomized test that rejects