H I E R

Harvard Institute of Economic Research

Discussion Paper Number 1992

A General Theory of Hypothesis Testing in The Simultaneous Equations Model

by

Marcelo J. Moreira

January 2003

Harvard University
Cambridge, Massachusetts

A General Theory of Hypothesis Testing in the Simultaneous Equations Model

Marcelo J. Moreira*

Harvard University

Abstract

Classical exponential-family statistical theory is employed to characterize the class of exactly similar tests for a structural coefficient in a simultaneous equations model with normal errors and known reduced-form covariance matrix. We also find a necessary condition for tests to be unbiased and derive their *power envelope*. When the model is just-identified, we show that the Anderson-Rubin, score, and conditional likelihood ratio tests are optimal. When the model is over-identified, there exists no optimal test. Nevertheless, Monte Carlo simulations indicate that the power curve of the conditional likelihood ratio test is reasonably close to the *power envelope*.

*Lengthy discussions with Thomas Rothenberg were extremely important for this work and I am deeply indebted for all his help and support. I also would like to thank the seminar participants at Berkeley, Brown, Chicago, EPGE/FGV, Harvard, Northwestern, Penn, Pittsburgh, Princeton, PUC-Rio, UCLA, UCSD and Yale.

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, Jaeger and Baker (1995), and Dufour (1997), these methods are not satisfactory when instruments are only weakly correlated with the regressors. More recently, there has been a significant effort to find more reliable econometric methods.

In this paper, we develop a general method for characterizing the whole class of exactly similar tests under the assumption of normal errors and known reduced-form covariance matrix. The class of Gaussian similar tests is quite large and includes the Anderson-Rubin, score, and the conditional tests proposed by Moreira (2001b). We also find a necessary condition for tests to be unbiased and derive their *power envelope*. When the model is just-identified, we show that the Anderson-Rubin, score, and conditional likelihood ratio tests are optimal. When the model is over-identified, there exists no optimal test. Nevertheless, Monte Carlo simulations indicate that the power curve of the conditional likelihood ratio test is reasonably close to the *power envelope*.

The assumption of normality with known variance simplifies the simultaneous equations model, but maintains the difficulties arising from weak instruments. When this assumption is dropped, Moreira (2001b) proposes simple modifications of the Gaussian similar tests by replacing the reduced-form variance by a consistent estimator. The resulting tests are then shown to have limiting power equal to the exact power when the errors are normal with known variance under the weak-instrument asymptotics proposed by Staiger and Stock (1997). In particular, these modified tests are asymptotically similar even when the structural parameters are unidentified.

The paper is organized as follows. Section 2 develops exact results for the special case of a two-equation model with normal errors and known reduced-form covariance matrix. Section 3 extends the results to cases with more than two endogenous variables and with additional exogenous variables. Section 4 contains concluding remarks. All proofs are given in an appendix.

2 Testing with Known Covariance Matrix

2.1 The Model

To avoid tedious notation, it is useful to begin with a simple special case. Consider the structural equation

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

where y_1 and y_2 are $n \times 1$ vectors of observations on two endogenous variables, u is an $n \times 1$ unobserved error vector, and β is an unknown scalar parameter.

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. However, 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 the exogenous variables do not appear in (1). Specifically, it is assumed that the reduced form for $Y = [y_1, y_2]$ can be written as

$$\begin{aligned} y_1 &= Z\pi\beta + v_1 \\ y_2 &= Z\pi + v_2, \end{aligned} \tag{2}$$

where Z is an $n \times k$ matrix of nonrandom exogenous variables having full column rank, π is a $k \times 1$ vector, and the n rows of the $n \times 2$ matrix of reduced-form errors $V = [v_1, v_2]$ are i.i.d. normal with mean zero and *known* 2×2 covariance matrix

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix}.$$

The assumption of normality and known variance is supported by the weak-instrument asymptotics developed by Staiger and Stock (1997). This assumption simplifies the simultaneous equations model, but maintains the difficulties arising from weak instruments.

The main goal is to test the hypothesis $H_0 : \beta = \beta_0$, treating π as a nuisance parameter. A test is said to have size α if the probability of rejecting the null hypothesis when it is true does not exceed α. That is, if $\mathbf{P}$

is the subset of the k-dimensional Euclidian space in which π lies,

$$\sup_{\pi \in \mathbf{P}} \text{prob}\,(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) = \alpha.$$

Since π is unknown, finding a test at correct size α is nontrivial. The task is simplified if one can find *similar tests*, since their null rejection probability does not depend on the nuisance parameters at all. If, for example, one rejects H_0 if some test statistic $\mathcal{T}$ is greater than a given constant, the test will be similar if the test statistic $\mathcal{T}$ is *pivotal*.

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

$$\lim_{n \to \infty} \text{prob}\,(\mathcal{T} > c_\alpha) = \alpha.$$

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 may differ substantially from the size based on the asymptotic distribution of $\mathcal{T}$. In fact, following earlier work by Gleser and Hwang (1987), Dufour (1997) shows that the true levels of the usual $Wald$-type tests deviate arbitrarily from their nominal levels if π cannot be bounded away from zero.[1] Since weak instruments are common in empirical research, it is desirable to find tests with correct size α even when π cannot be bounded away from the origin.

One such test is proposed by Anderson and Rubin (1949). For any matrix Q having full column rank, let $N_Q = Q(Q'Q)^{-1}Q'$ and $M_Q = I - N_Q$. Define $u_0 \equiv y_1 - y_2\beta_0 = Yb_0$ and $\sigma_0^2 = b_0'\Omega b_0$, where $b_0 = (1, -\beta_0)'$. When Ω is

[1] His analysis remains valid in the special case in which Ω is known.

known, the Anderson-Rubin procedure rejects the null hypothesis if

$$AR_0 = u_0' N_Z u_0 / \sigma_0^2 \tag{3}$$

is large. Since $u_0 = Vb_0$ under H_0, the test statistic AR_0 is pivotal with a chi-square-k distribution, regardless of the value of the nuisance parameter π. When $k = 1$, we will show that this test is uniformly most powerful among the class of unbiased tests. However, when $k > 1$, we will show that the Anderson-Rubin test has no particular optimal properties. Indeed, if k is large, the power of the Anderson-Rubin test can be quite low. The score test used by Kleibergen (2002) and Moreira (2001a), and the conditional tests proposed by Moreira (2001b) reflect some attempts to find tests with better power properties and correct size. However, there is still no unifying framework for characterizing similar and unbiased tests in the simultaneous equations model. The goal of this paper is to provide a general theory for testing the structural parameter in the weak-instrument case, using classical statistical theory.

2.2 The Family of Similar Tests

Under the normality assumption, the probability model is a member of the curved exponential family. We can therefore adapt the extensive set of results summarized in Lehmann (1986) for testing H_0. For any non-singular 2×2 matrix D, the two columns of $Z'YD$ are a pair of sufficient statistics for the unknown parameters (β, π). A convenient choice is the pair

$$S = Z'Yb_0 \qquad \text{and} \qquad T = Z'Y\Omega^{-1}a_0, \tag{4}$$

where $a_0 = [\beta_0, 1]'$. Although the null distribution of the statistic $S = Z'u_0$ does not depend on the nuisance parameter π, the null distribution of T is very sensitive to the value of π. Indeed, T is a sufficient statistic for π under the null hypothesis. A little algebra shows that

$$T = a_0'\Omega^{-1}a_0 \cdot Z'Z\hat{\pi},$$

where $\hat{\pi}$ is the maximum likelihood estimate of π when β is constrained to take the null value β_0. The vectors S and T are independent and normally distributed under both the null and alternative hypotheses. Specifically,

$$S \sim N(Z'Z\pi(\beta - \beta_0), Z'Z \cdot b_0'\Omega b_0)$$

$$T \sim N(Z'Z\pi \cdot a'\Omega^{-1}a_0, Z'Z \cdot a_0'\Omega^{-1}a_0),$$

where $a = [\beta, 1]'$.

Because they are sufficient statistics, there is no loss in focusing only on tests that can be written as (possibly randomized) functions of S and T. Specifically, let ϕ be a critical function such that $0 \leq \phi \leq 1$. For each S and T, the test rejects or accepts H_0 with probabilities $\phi(S, T)$ and $1 - \phi(S, T)$, respectively. Here, we omit the dependence of the test on Z, β_0 and Ω out of convenience. Let E_0 represent expectation over the distribution of S under the null $H_0 : \beta = \beta_0$, and suppose that the set $\mathbf{P}$ in which π is known to lie contains a k-dimensional rectangle, so that the statistic T is complete under the null hypothesis. In the appendix, we show that the following holds:

Theorem 1 (Similarity Condition): *A test ϕ is similar at size α if and only if $E_0\phi(S, t) = \alpha$ for almost every t.*

According to Theorem 1, every similar test has the Neyman structure with respect to T. Since every similar test must have the conditional structure, this theorem supports the conditional approach proposed by Moreira (2001b). To review this method, consider the Wald and likelihood ratio statistics.

EXAMPLE 1: The Wald statistic centered around the 2SLS estimator is given by

$$W_0 = \frac{(b_{2SLS} - \beta_0)^2 \, y_2' N_Z y_2}{\hat{\sigma}^2}, \tag{5}$$

where $b_{2SLS} = (y_2' N_Z y_2)^{-1} y_2' N_Z y_1$ and $\hat{\sigma}^2 = [1 - b_{2SLS}]\Omega[1 - b_{2SLS}]'$. Here, the nonstandard structural error variance estimate exploits the fact that Ω is known.

EXAMPLE 2: The likelihood ratio statistic equals

$$LR_0 = \frac{1}{2}\left[\bar{S}'\bar{S} - \bar{T}'\bar{T} + \sqrt{[\bar{S}\bar{S} + \bar{T}'\bar{T}]^2 - 4[\bar{S}'\bar{S} \cdot \bar{T}'\bar{T} - (\bar{S}'\bar{T})^2]}\right],$$

where $\bar{S} = (Z'Z)^{-1/2}\, S/\sqrt{b_0'\Omega b_0}$ and $\bar{T} = (Z'Z)^{-1/2}\, T/\sqrt{a_0'\Omega^{-1}a_0}$.

Although the Wald and likelihood ratio statistics are not pivotal, Moreira (2001b) constructs tests with the Neyman structure by adjusting the critical value, so the null rejection probability remains equal to α for each level $T = t$. The resulting conditional test then rejects H_0 when the test statistic $\psi\,(S, T)$ is larger than the critical value function $c_\psi\,(T; \alpha)$.

2.3 Exogeneity Tests

This conditional approach for finding similar tests can also be applied to other statistical problems involving nuisance parameters, *cf.* Hillier (1987).

Improved inference should be possible whenever $\tilde{T}$, the sufficient statistic under the null hypothesis, is boundedly complete. Furthermore, computation of the critical value function can be simplified if there exists a statistic $\tilde{S}$ that is pivotal (and, consequently, independent of $\tilde{T}$ under the null by Basu's lemma).

For example, consider the problem of testing whether the explanatory variable y_2 is endogenous; that is, if the covariance between the disturbances u and v_2 equals zero ($\sigma_{u2} = 0$). This problem can be shown to be equivalent to testing $\tilde{H}_0 : \xi = 0$, where $\xi = \beta - \omega_{12}/\omega_{11}$. Again, when Ω is known, we can partition the sufficient statistic into two k-dimensional statistics,

$$\tilde{S} = Z'Y\Omega^{-1}e_1 \qquad \text{and} \qquad \tilde{T} = Z'Ye_2 = Z'y_2, \tag{6}$$

where $\tilde{S}$ is pivotal, and $\tilde{T}$ is sufficient and boundedly complete under the null. Consequently, we can proceed as in Moreira (2001b) by constructing a similar test based on a nonpivotal test statistic $\psi(\tilde{S}, \tilde{T})$ by adjusting its critical value for each level $\tilde{T} = t$.

The particular problem of testing the exogeneity of the explanatory variable can be further analyzed by noting that testing $\tilde{H}_0 : \xi = 0$ is equivalent to testing $H_0 : \beta = \omega_{12}/\omega_{22}$. In fact, for the hypothesized value $\beta_0 = \omega_{12}/\omega_{22}$, the statistic $[S, T]$ is proportional to the statistic $[\tilde{S}, \tilde{T}]$, and structural-parameter tests can then be adapted as exogeneity tests. To my knowledge, the recent literature on weak instruments has ignored this connection between tests for structural parameter and tests for exogeneity of the explanatory variable. The *Similarity Condition* applied to exogeneity tests is thus related to the work by Hillier (1987), but with the advantage of having an easily interpretable condition that characterizes exogeneity tests with correct size (due to the assumption of known variance). Theorem 1 is also particularly helpful in finding pre-testing procedures and in deriving the power envelope for similar tests.

2.4 Pre-Testing Procedures

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. Pre-testing on the *constrained* maximum likelihood estimator $\hat{\pi}$, however, does not cause any difficulties with similar tests. More generally, we have the following implication of Theorem 1:

COROLLARY 1: *Let $h(T)$ be a measurable real-valued function and let $\phi_1(S,T)$ and $\phi_2(S,T)$ be two similar tests at level α. Finally, let $\phi_3 = I\left[h(T) > c\right]\phi_1 + I\left[h(T) \leq c\right]\phi_2$ 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.*

This result has important implications for improving inference. 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. Corollary 1 can also be applied to similar tests that are not easily written as a (randomized) function of the sufficient statistics S and T.

Another possible application is instrument selection. Applying the Monotone Convergence Theorem, we can extend Corollary 1 to pre-tests based on $\hat{\pi}$ that select among a countable number of similar tests. Thus, by using $\hat{\pi}$ to choose the number of instruments or to find some linear combination between the instruments, we can improve power without creating size distortions.

2.5 Power Envelope for Similar Tests

When π is far from the origin and the sample size is large, the standard likelihood ratio, Wald, and Lagrange multiplier tests have approximate power

$$1 - G\left(c_\alpha \, ; \frac{\pi' Z' Z \pi \left(\beta - \beta_0\right)^2}{\sigma_0^2}\right). \tag{7}$$

Here, c_α is the $1 - \alpha$ quantile of a chi-square-one distribution, and $G(\cdot; \mu)$ is the noncentral $\chi^2(1)$ distribution function with noncentrality parameter μ. However, these tests are not generally similar, and the power approximation in (7) is unreliable when π is near the origin. Only in the case that $k = 1$, in which the model is exactly identified, do we have an exact optimality result. Then, as shown in the appendix, the Anderson-Rubin AR_0 test is uniformly most powerful unbiased and has exact power function given by (7). Therefore, it is not surprising why Monte Carlo simulations run by Wang and Zivot (1998) and Zivot, Startz and Nelson (1998) suggest that no test dominates the one proposed by Anderson and Rubin (1949) when $k = 1$.

Its power is very close to the power of the Anderson-Rubin AR_0 test, which is itself the optimal test when Ω is known.

When $k > 1$, there exists no uniformly most powerful test. To assess the power properties of similar tests, it is useful to find the *power envelope*, the upper bound of the rejection probability for each alternative. We can also find a useful power envelope for two-sided alternatives by considering the class of unbiased tests. The following lemma states a necessary condition for a test to be unbiased:

LEMMA 1 (UNBIASEDNESS CONDITION): *For testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, a test ϕ is unbiased only if: (i) the Similarity Condition holds; and, (ii) $E_0\phi(S,t)\,S = \mathbf{0}$ for almost every t.*

The first condition in Lemma 1 arises from the fact that any unbiased test must be similar (in the frontier of the null hypothesis). The second condition refers to the *local* behavior of unbiased tests around the null hypothesis. Using the fact that T is complete under the null, condition (ii) in Lemma 1 is equivalent to $E_{\beta_0,\pi}\phi(S,T)\,S = \mathbf{0}$. Hence, if the test ϕ is unbiased, the covariance between ϕ and S is zero under the null hypothesis. Many tests satisfy the *Unbiasedness Condition*, among them the score test used by Kleibergen (2002) and Moreira (2001a):

EXAMPLE 3: The score test rejects the null hypothesis if and only if

$$LM_0 = \left(\bar{T}'\bar{S}\right)^2 / \bar{T}'\bar{T} > c_\alpha.$$

The resulting score test, $\phi_{LM}(S,T)$, satisfies the *Unbiasedness Condition*, since the LM_0 statistic is pivotal and S is symmetric under H_0; see Lehmann (1986, p. 136-7).

Although, in principle, we can also check whether the Anderson-Rubin, conditional Wald and conditional likelihood ratio tests are unbiased, the most interesting application of the *Unbiasedness Condition* is to compute an upper bound for the power of unbiased tests. In the appendix, we show:

THEOREM 2: *For testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, with $\pi \neq 0$, we have:*

a. *If the model is just-identified, the uniformly most powerful unbiased test has a power function given by*

$$P_{\beta,\pi}(AR_0 > c_\alpha) = 1 - G\left(c_\alpha\ ; \frac{\pi' Z' Z \pi\, (\beta - \beta_0)^2}{\sigma_0^2}\right). \tag{8}$$

b. *If π is known, the uniformly most powerful unbiased test has a power function given by*

$$P_{\beta,\pi}(\mathcal{R} > c_\alpha) = 1 - G\left(c_\alpha; \frac{\pi' Z' Z \pi\, (\beta - \beta_0)^2}{\omega_{11} - \omega_{12}^2/\omega_{22}}\right), \tag{9}$$

where $\mathcal{R}$ is defined in equation A.7.

c. *If π is unknown and* **P** *contains a k-dimensional rectangle, the power envelope for the class of unbiased tests is given by*

$$P_{\beta,\pi}\left(\frac{(\pi' S)^2}{\sigma_0^2 \pi' Z' Z \pi} > \ c_\alpha\right) = 1 - G\left(c_\alpha; \frac{\pi' Z' Z \pi\, (\beta - \beta_0)^2}{\sigma_0^2}\right). \tag{10}$$

Note that (9) is an upper bound for the power of any two-sided test with correct size.[2] Since $\omega_{11} - \omega_{12}^2/\omega_{22}$ is no larger than σ_0^2, insisting on similarity lowers the attainable power of the test. Alternatively, the optimal test for known π can be understood as the optimal similar 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. Finally, the power function in (8) can be seen as a result of the power envelope in (10) for the special case that the model is just-identified, and the LR_0 and LM_0 statistics collapse to the AR_0 statistic.

2.5.1 Monte Carlo Simulations

Monte Carlo simulations run by Moreira (2001b) indicate that the conditional likelihood ratio test outperforms the Anderson-Rubin, score, and conditional Wald tests. To compare the power of the conditional likelihood-ratio test with the *power envelope* given by (9), we perform here a 1,000 replication experiment based on design I of Staiger and Stock (1997). The hypothesized value β_0 is zero. The elements of the 100×4 matrix Z are drawn as independent standard normal and then held fixed over the replications. Two different values of the π vector are used so that $\lambda'\lambda/k = \pi'Z'Z\pi/(\omega_{22}k)$, the "population" first-stage F-statistic (in the notation of Staiger and Stock), takes the values 1 (weak instruments) and 10 (good instruments). The rows of $[u, v_2]$ are i.i.d. normal random vectors with unit variances and correlation ρ. Here, we report only results for $\rho = 0.50$, although we have considered

[2] In the appendix, we also derive the power envelope for one-sided similar tests.

different degrees of endogeneity of y_2.

Figures 1 and 2 graph, for fixed values of π and ρ, the rejection probability of the conditional likelihood ratio test as a function of the true value β.[3] In each figure, the power curve is at approximately the 5% level when β equals β_0. This reflects the fact that the conditional likelihood ratio test is similar. The power envelope is also included. As expected, both power curves become steeper as the quality of instruments improves.

Figure 2 indicates that the conditional likelihood ratio test has power essentially equal to the *power envelope* when identification is strong. This is expected, since this test is optimal under the usual asymptotics. However, Figure 1 also shows that the conditional likelihood ratio test has power lower than the power envelope. This reflects the fact that the power envelope uses the unknown direction of π, but this direction cannot be well estimated when instruments are weak. This suggests that an improvement on the conditional likelihood ratio test may still be possible, at least when there is some previous knowledge of the direction of π.

[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.

3 Inference on Parameters of Endogenous Variables in the Multivariate Setting

3.1 Inference on All Parameters

The previous theory can easily be extended to a structural equation with more than two endogenous variables and with additional exogenous variables, as long as inference is conducted on the coefficients of all endogenous variables. Consider the structural equation

$$y_1 = Y_2\beta + X\gamma + u, \tag{11}$$

where Y_2 is the $n \times l$ matrix of observations on the l explanatory endogenous variables, and X is the $n \times r$ matrix of observations on r exogenous variables. The complete system contains $k \geq l$ additional exogenous variables, which can be used as instruments for conducting inference on the structural coefficients β. The reduced form for $Y = [y_1, Y_2]$ is

$$\begin{aligned} y_1 &= Z\Pi\beta + X\delta + v_1 \\ Y_2 &= Z\Pi + X\Gamma + V_2, \end{aligned}$$

where $\delta = \Gamma\beta + \gamma$. The rows of $V = [v_1, V_2]$ are i.i.d. with mean zero and covariance matrix Ω. It is assumed that X and Z have full column rank. The problem is to test the joint hypothesis $H_0 : \beta = \beta_0$, treating Π, Γ, and δ as nuisance parameters.

The nuisance parameters δ and Γ can be eliminated by requiring the test to be invariant to linear transformations of X. For the group G of transformations that preserves H_0, $g(Y) = Y + XF$ for arbitrary conformable matrices F, the maximal invariant in terms of the sufficient statistic is $Z'M_XY$. However, for any known nonsingular, nonrandom $(l+1) \times (l+1)$ matrix D, $Z'M_XYD$ is also an invariant sufficient statistic. A convenient choice is the matrix

$$D_0 = [b_0,\ \Omega^{-1}A_0],$$

where b_0 is the $(l+1) \times 1$ vector $[1, -\beta_0']'$, and A_0 is the $(l+1) \times l$ matrix $[\beta_0, I_l]'$. Note that every column of A_0 is orthogonal to b_0. Then the invariant sufficient statistic can be represented by the pair $[S, T]$, where

$$S = Z'M_XYb_0 = Z'(y_1 - Y_2\beta_0) \text{ and } T = Z'M_XY\Omega^{-1}A_0 \tag{12}$$

are independent and normally distributed.

Therefore, analogous to the case $l = 1$, we can construct tests with correct size for β by adjusting the critical value based on T. Again, when β is known to equal β_0, T is a sufficient statistic for Π and is a one-to-one function of the constrained maximum likelihood estimator $\hat{\Pi}$:

$$\hat{\Pi} = (Z'Z)^{-1} T \left(A_0'\Omega^{-1}A_0\right)^{-1}.$$

If we are working with a limited-information model (in the sense that the set $\mathbf{P}$ in which $vec(\Pi)$ takes values contains a $k \cdot l$-dimensional rectangle), then T is also boundedly complete. As a consequence, the *Similarity Condition* also holds for the multivariate case. Of course, if there are additional restrictions on how the instruments affect the endogenous explanatory variables, power improvement may be possible by considering tests that do not have Neyman structure with respect to T.

3.2 Inference on a Subset of the Parameters

The last observation also suggests a way to overcome limitations of the available methods when the null hypothesis involves a subset of the parameter β. Kleibergen (2002) proposes to assure strong identification on the parameters of the excluded endogenous variables, and Dufour (1997) proposes a projection technique for the confidence region. The first approach 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 with additional restrictions in the underlying stochastic equation for Y_2.

Consider the partition of $Y_2 = [Y_{2,h}, Y_{2,n}]$ and $\beta = [\beta_h, \beta_n]$, where $Y_{2,h}$ is a $n \times l_h$ matrix and β_h is a l_h-dimensional vector. Here, the l_n-dimensional vector β_n is a nuisance parameter for the problem of testing $H_0 : \beta_h = \beta_{h,0}$. The reduced-form model can be re-written as:

$$\begin{aligned} y_1 &= Z_h \Pi_h \beta_h + Z_n \Pi_n \beta_n + X\delta + v_1 \qquad (13) \\ [Y_{2,h}, Y_{2,n}] &= [Z_h \Pi_h, Z_n \Pi_n] + X\Gamma + [V_{2,h}, V_{2,n}], \end{aligned}$$

where Z_h and Z_n are (not necessarily mutually exclusive) exogenous variables based on Z that affect $Y_{2,h}$ and $Y_{2,n}$, respectively.

If the restrictions in the full-information model guarantee that the column space of Z_n does not contain the column space of Z_h, we can then create non-degenerate statistics based on the data $Y_h = [y_1, Y_{2,h}]$:

$$S_h = Z_h' M_{X,Z_n} Y_h b_{h,0} \text{ and } T_h = Z_h' M_{X,Z_n} Y_h (\Omega_h)^{-1} A_{h,0},$$

where $b_{h,0}$ is the $(l_h + 1) \times 1$ vector $\left[1, -\beta'_{h,0}\right]'$, $A_{0,h}$ is the $(l_h + 1) \times l_h$ matrix $\left[\beta_{h,0}, I_{l_h}\right]'$, and Ω_h is the covariance matrix of each observation of the reduced-form errors $V = (v_1, V_{2,h})$. Again, using the fact that S_h and T_h are independent, we can construct similar tests for β_h by employing the conditional method proposed by Moreira (2001b).

4 Conclusions

Previous authors have noted that the simultaneous equations model with known reduced-form variance has a simpler mathematical structure than the model with unknown variance, but inference procedures for the two models behave very much alike in moderate-sized samples. In fact, Moreira (2001b) shows that replacing the reduced-form covariance matrix with a consistent estimate and relaxing the normality assumption does not have an effect under the weak-instrument asymptotics.

We then apply classical statistical theory to characterize the class of tests with correct size in the simpler model under the assumption of normality. We find two conditions observed in exponential families that also hold in our simultaneous equations model. The *Similarity Condition* guarantees that every similar test has the Neyman structure with respect to the constrained maximum likelihood estimator for the instruments' coefficients. The *Unbiasedness Condition* asserts that any unbiased test must be uncorrelated with the pivotal statistic on which the Anderson-Rubin test is based.

Unlike alternative procedures using the bootstrap or higher-order asymptotics, the testing procedures discussed here behave well even when there is no identification at all. As a negative result, we show that there exists no optimal test in the over-identified case, due to the fact that the probability model is a member of the curved exponential family. Nevertheless, Monte Carlo simulations suggest that the conditional likelihood ratio test has good power overall when instruments are strong. In the weak-instrument case, the power of the conditional likelihood ratio test is below the *power envelope*, which suggests that improvement may still be possible by exploring other tests that satisfy the *Similarity Condition*.

5 References

ANDERSON, T., N. KUNITOMO, AND T. SAWA (1982): "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, 50, 1009-28.

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

ANDREWS, D. AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383-414.

BOUND, J., D. JAEGER AND R. BAKER (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak," *Journal of American*

Statistical Association, 90, 443-50.

BREUSCH, T. AND A. PAGAN (1980): "The Lagrange Multiplier Test and its Applications to Model Specifications in Econometrics," *The Review of Economic Studies*, 47, 239-53.

CRUZ, L. AND M. MOREIRA (2002): "Recipes for Applied Researchers: Inference when Instruments May Be Weak," Manuscript, UC Berkeley.

DONALD, S. AND W. WHITNEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161-92.

DUFOUR, J-M. (1997): "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica*, 65, 1365-88.

GLESER, L., AND J. HWANG (1987): "The non-existence of 100(1-α)% confidence sets of finite expected diameter in errors-in-variables and related models," *Annals of Statistics*, 15, 1351-62.

HILLIER, G. (1987): "Classes of Similar Regions and Their Power Properties for Some Econometric Testing Problems," *Econometric Theory*, 3, 1-44.

KLEIBERGEN, F. (2002): "Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781-803.

LEHMANN, E. (1986): *Testing Statistical Hypothesis.* 2nd edition, Wiley Series in Probability and Mathematical Statistics.

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

______________ (2001b): *A Conditional Likelihood Ratio Test for*

Structural Models. Manuscript, UC Berkeley.

______________ (2002): *Tests with Correct Size in the Simultaneous Equations Model*. PhD Thesis, UC Berkeley.

Moreira, M. and B. Poi (2003): "Implementing Tests with Correct Size in the Simultaneous Equations Model," forthcoming, *Stata Journal*.

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

Staiger, D., and J. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557-86.

Stock, J., M. Yogo and J. Wright (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518 – 529

Wang, J., and E. Zivot (1998): "Inference on a structural parameter in instrumental variables regression with weak instruments," *Econometrica*, 66, 1389-404.

Zivot, E., R. Startz, and C. Nelson (1998): "Valid confidence intervals and inference in the presence of weak instruments," *International Economic Review*, 39, 1119-44.

Figure 1

Power Curves: Weak Instruments

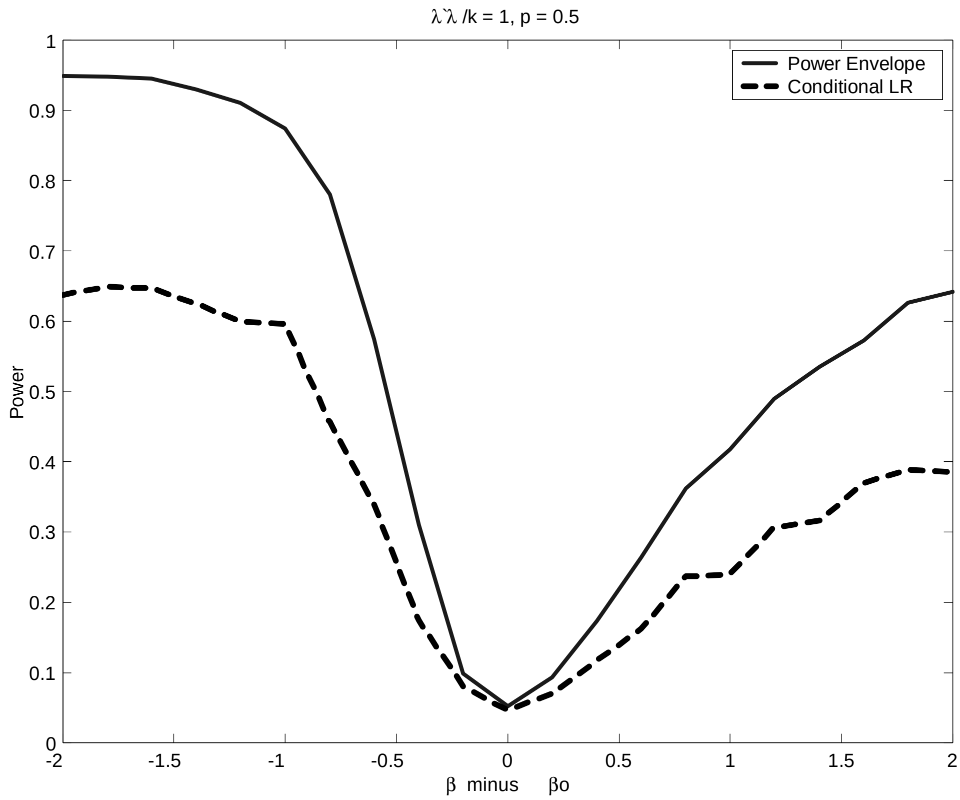
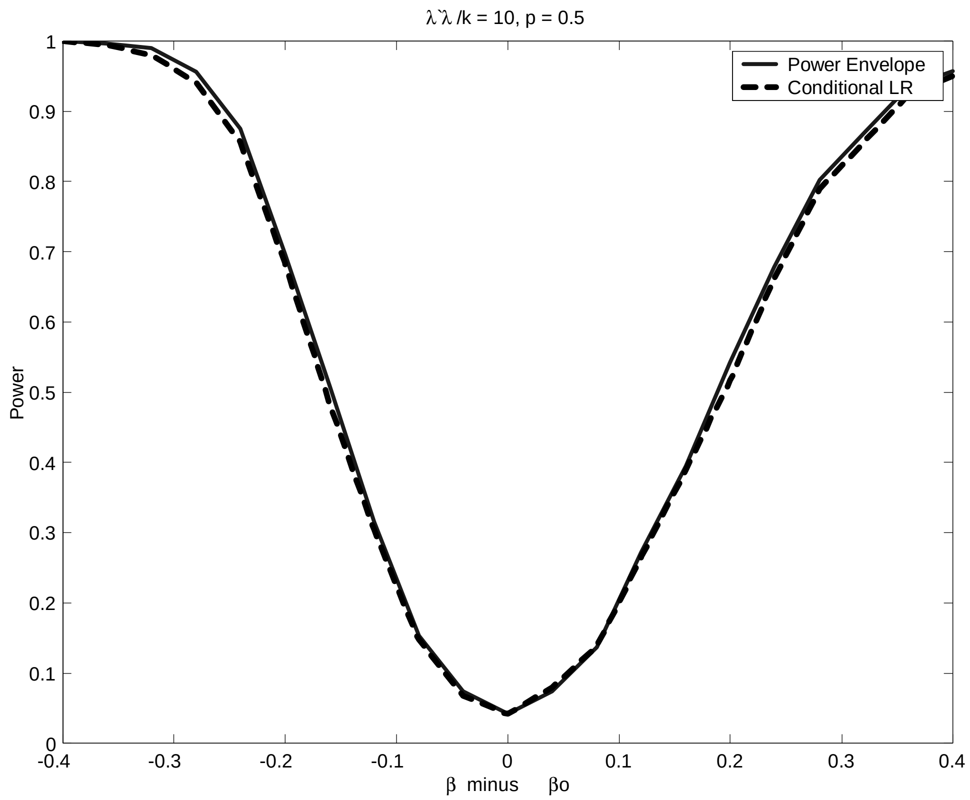


Figure 2

Power Curves: Good Instruments



6 Appendix

The results in Section 2 are based on the following two lemmas proved in Lehmann (1986), pp. 142-3:

LEMMA A.1: *Let X be a random vector with probability distribution*

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

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

LEMMA A.2: *Suppose that the distribution of X is given by*

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

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 T when $\theta = \theta_0$ and that

$$h(r,t) = a(t)r + b(t) \qquad \text{with } a(t) > 0.$$

Then the uniformly most powerful unbiased (UMPU) test ϕ for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ is given by

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

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

PROOF OF THEOREM 1: Since randomization is allowed, any test can be written as $\phi(S,T)$. Since the test is similar at size α, it must be the case that:

$$E_0\phi(S,T) = \alpha, \qquad \forall \pi \in P. \tag{A.1}$$

By Lemma A.1, the family of distributions of T when the null hypothesis is true, $\mathcal{P}^T = \left\{P^T_{\beta_0,\pi}; \pi \in P\right\}$, is complete. Consequently, the following holds:

$$E_0\{\phi(S,T)\,|t\} = \alpha, \qquad a.e.\ \mathcal{P}^T.$$

Note that the distribution of S does not depend on π under the null hypothesis, and that S is independent of T. Therefore, using the fact that ϕ is integrable:

$$E_0\phi(S,t) = \alpha, \qquad a.e.\ \mathcal{P}^T. \tag{A.2}$$

Conversely, if the test is such that (A.2) holds, then (A.1) is trivially true. Therefore, the test is similar at size α.

■

PROOF OF LEMMA 1: Because the power function of exponential family distributions is continuous, any unbiased test must be similar (in the frontier of the null hypothesis). Now, it just remains to show that $E_0\phi(S,T)\,S = \mathbf{0}$ holds. The random variable S has the following distribution:

$$\exp\left\{-\frac{[s-(Z'Z)\,\pi\,(\beta-\beta_0)]'\,(Z'Z)^{-1}\,[s-(Z'Z)\,\pi\,(\beta-\beta_0)]}{2\cdot\sigma_0^2}\right\}, \tag{A.3}$$

up to a multiplication by the term $(2\pi)^{-k/2}\left|\sigma_0^2 \cdot Z'Z\right|^{-1/2}$. Expression (A.3) can then be written as:

$$C\left((\beta-\beta_0)^2 \pi' (Z'Z)\pi/\sigma_0^2\right)\exp\left\{s'\pi(\beta-\beta_0)\right\}v(s).$$

The conditional power function of a test ϕ is given by $\zeta(\beta,\pi) = E_{\beta,\pi}\phi(S,t)$:

$$\zeta(\beta,\pi) = \int \phi(S,t)\, C\left((\beta-\beta_0)^2 \pi' (Z'Z)\pi/\sigma_0^2\right)\exp\left\{s'\pi(\beta-\beta_0)\right\}v(s)\,ds.$$

Taking the derivative of the power function with respect to β, we have:

$$\zeta_\beta(\beta,\pi) = E_{\beta,\pi}\phi(S,t)\left\{S'\pi + \frac{C'(.)}{C(.)}\cdot\left[2(\beta-\beta_0)\pi'(Z'Z)\pi/\sigma_0^2\right]\right\}, \quad \text{(A.4)}$$

using the fact that we can interchange derivatives/limits and integrals; see Lehmann (1986, p. 59). This expression is closely related to the findings in Lehman (1986, p. 136 and p. 148) for exponential families, but with one important distinction. The nuisance parameter π is present in the expectation and in the integrand. This additional complication comes from the fact that our probability model belongs to a curved exponential family.

As in Lehmann (1986), consider expression (A.4) for $\phi \equiv \alpha$. Then we have:

$$\zeta_\beta(\beta,\pi) = 0 = \alpha \cdot E_{\beta,\pi}s'\pi + \alpha\cdot\frac{C'(.)}{C(.)}\cdot\left[2(\beta-\beta_0)\pi'(Z'Z)\pi/\sigma_0^2\right].$$

Thus we can simplify expression (A.4) (for an arbitrary test ϕ) to

$$\zeta_\beta(\beta,\pi) = \left[E_{\beta,\pi}\phi(S,t)\,S - E_{\beta,\pi}\phi(S,t)\cdot E_{\beta,\pi}S\right]'\pi.$$

If the test is unbiased, then the expression above must be equal to zero under the null for *any* value of π. Since the distribution of S does not depend on

π under the null, we have:

$$\left[E_0\phi\left(S,t\right)S - E_0\phi\left(S,t\right)\cdot E_0S\right]'\pi = 0 \text{ for any } \pi. \tag{A.5}$$

Thus, using the fact that $E_0S = 0$, expression (A.5) is equivalent to:

$$E_0\phi\left(S,t\right)S = \mathbf{0}. \tag{A.6}$$

Alternatively, an equivalent condition to (A.6) is:

$$E_{\beta_0,\pi}\phi\left(S,T\right)S = \mathbf{0},$$

using the fact that T is complete under the null hypothesis.

■

PROOF OF THEOREM 2: The following is true:

a. For some measure $\mu\left(y\right)$, the probability distribution of Y can be written as:

$$dP_{\theta,\pi}\left(y\right) = C\left(\theta,\pi\right)\exp\left[\theta R\left(y\right) + \pi T\left(y\right)\right]d\mu\left(y\right)$$

where $R\left(Y\right)$ is the first column of $Z'Y\Omega^{-1}$ and $\theta = \pi\left(\beta - \beta_0\right)$. Since P does not contain the origin and the model is just identified, testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ is equivalent to testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Let

$$\bar{S} = \frac{\left(Z'Z\right)^{-1/2}Z'\left(y_1 - y_2\beta_0\right)}{\sigma_0}.$$

Notice that $\bar{S} = \delta_1 R + \delta_2 T$ where

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

Now Lemma A.2 can be applied. Since $\bar{S} \sim N(0,1)$ under H_0 and, in particular, it is symmetric around zero, it is straightforward to show that the optimal test rejects the null if $AR_0 > c_\alpha$. Under the alternative β,

$$AR_0 \sim \chi^2 \left(1, \frac{\pi' Z' Z \pi \left(\beta - \beta_0\right)^2}{\sigma_0^2}\right).$$

Consequently, the power of the optimal test is given by (8).

b. Since π is known, for some measure $\mu(y)$, the probability distribution can be written as:

$$dP_{\beta,\pi}(y) = C(\beta,\pi) \exp\left[R(y)' \pi \beta\right] d\mu(y).$$

Since this distribution is a one-parameter exponential family, the $UMPU$ test rejects the null hypothesis if

$$\mathcal{R} = \frac{\left\{\pi' Z' \left[\left(y_1 - Z\pi\beta_0\right) - \omega_{12}\omega_{22}^{-1}\left(y_2 - Z\pi\right)\right]\right\}^2}{\left(\omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{12}\right)\pi' Z' Z \pi} \tag{A.7}$$

is larger than c_α. Under the alternative β,

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

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

c. The power of the test ϕ is given by $E_{\beta,\pi}\phi(S,T)$. Since S and T are independent, then:

$$E_{\beta,\pi}\phi\left(S,T\right)=\int\left[\int\phi\left(s,t\right)f\left(s,\beta,\pi\right)ds\right]g\left(t,\beta,\pi\right)dt,$$

where $f\left(s,\beta,\pi\right)$ and $g\left(t,\beta,\pi\right)$ are the density functions associated to S and T, respectively. Notice that the power conditioned on $T=t$ is

$$\int\phi\left(s,t\right)f\left(s,\beta,\pi\right)ds.$$

Consider the test $\phi^{*}(S)$ that assigns 1 if $f\left(s,\beta,\pi\right)>kf\left(s,\beta_0\right)$ and 0 otherwise, where k is chosen such that $E_{\beta_0,\pi}\phi^{*}(S)=\alpha$. The claim is that the test $\phi^{*}(S)$ is most powerful among all similar tests at the significance level α.

Let S^{+} and S^{-} be the sets in the sample space where $\phi^{*}(s)-\phi\left(s,t\right)>0$ and $\phi^{*}(s)-\phi\left(s,t\right)<0$, respectively. Notice that, if s is in S^{+}, $\phi^{*}(s)=1$ and $f\left(s,\beta,\pi\right)>kf\left(s,\beta_0\right)$. Analogously, if s is in S^{-}, $\phi^{*}(s)=0$ and $f\left(s,\beta,\pi\right)\leq kf\left(s,\beta_0\right)$. Therefore:

$$\int\left[\phi^{*}(s)-\phi\left(s,t\right)\right]\left[f\left(s,\beta,\pi\right)-kf\left(s,\beta_0\right)\right]ds\geq 0.$$

The difference in power satisfies

$$\int\left[\phi^{*}(s)-\phi\left(s,t\right)\right]f\left(s,\beta,\pi\right)dv\geq k\int\left[\phi^{*}(s)-\phi\left(s,t\right)\right]f\left(v,\beta_0\right)ds.$$

By Theorem 1, if the test $\phi\left(S,T\right)$ is similar then $E_0\phi\left(S,t\right)=\alpha$, *a.e.* $\mathcal{P}^{T}$. Without loss of generality, it can be considered that $E_0\phi\left(S,t\right)=\alpha$, $\forall t$. That is:

$$\int \phi(s,t) f(s,\beta,\pi)\, ds = \alpha \qquad , \forall t$$

Therefore, the following holds:

$$\int [\phi^*(s) - \phi(s,t)] f(s,\beta,\pi)\, ds \geq 0.$$

Since the test that maximizes the conditional power does not depend on t, then this test itself maximizes power, as was to be proved. Since S is normally distributed, $f(s,\beta,\pi) > kf(s,\beta_0)$ for some k such that $E\phi^*(S) = \alpha$ if and only if the following holds. If $\beta > \beta_0$ then the test rejects the null if $\pi'S > z_\alpha\sqrt{\sigma_0^2\pi'Z'Z\pi}$. If $\beta < \beta_0$ then the test rejects the null if $\pi'S < -z_\alpha\sqrt{\sigma_0^2\pi'Z'Z\pi}$, where z_α is the critical value of a $\mathcal{N}(0,1)$ distribution for the significance level α.

Alternatively, we could find the point-optimal test (for fixed β and π) which satisfies the *Similarity Condition*. Given this condition is the same for finding one-sided optimal tests in exponential models, we can rely on Lehmann's (1986, p. 145-8) results. Analogously, for a two-sided alternative, we can find the point-optimal test (for fixed β and π) which satisfies the *Unbiasedness Condition*. Again, this condition is the same constraint for finding a UMPU test in the exponential model, and we can rely on Lehmann's results. Therefore, the point-optimal test that satisfies the *Unbiasedness Condition* rejects H_0 if

$$\frac{(\pi'S)^2}{\sigma_0^2\pi'Z'Z\pi} > c_\alpha.$$

Under the alternative β,

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

Consequently, the *power envelope* is given by (10).

■