COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 800

Power in Econometric Applications

Donald W. K. Andrews

August 1986

ABSTRACT

This paper is concerned with the use of power properties of tests in econometric applications. Power radius and inverse power functions are defined. These functions are designed to yield summary measures of power that facilitate the interpretation of test results in practice. Simple approximations are introduced for the power radius and inverse power functions of Wald, likelihood ratio, Lagrange multiplier, and Hausman tests. These approximations readily convey the general qualitative features of the power of a test. Examples are provided to illustrate their usefulness in interpreting test results.

1. INTRODUCTION

The purpose of this paper is to provide summary measures of the power of a test for use when interpreting the result of the test in practice. The measures suggested are particularly useful for assessing the evidence provided by a test when it fails to reject the null hypothesis. They also may be useful in determining an appropriate tradeoff between significance level and power, if so desired. The summary measures can be applied with tests of substantive economic hypotheses, model selection, and model specification. The results given incorporate Wald, likelihood ratio, Lagrange multiplier, and Hausman tests. They apply to tests where the null hypothesis is defined by nonlinear restrictions (say, $h(\theta) = \underset{\sim}{0}$).

The summary measures of power introduced here give the range of deviations from the restrictions (i.e., deviations of $h(\theta)$ from $\underset{\sim}{0}$) that correspond to alternative parameter values that have low (high) power, as defined by some chosen probability p_1 (p_2). Approximations to these exact summary measures are given, based on the following idea: Suppose the test involves q restrictions, a significance level α is specified, and the test statistic can be approximated by a non-central chi-square distribution under the alternative hypothesis. Then, there is a unique non-centrality parameter (NCP) that corresponds to the specified probability p_1 or p_2. Since the NCP is a quadratic form in the deviation vector $h(\theta)$, one is able to recover the values of the deviations that are consistent with the NCP. With a single restriction this is simple, since the NCP factors into the product of the squared deviations and the reciprocal of a variance term. For multiple restrictions, one can specify one or more directions η of deviations of interest, and solve for the deviation vector that is consistent with the NCP. Furthermore, the resultant approximate summary measures can be given rigorous asymptotic justifications, see below. These justifications do not restrict consideration to local alternatives, but are closely related to the local power function.

A central tenet underlying the motivation for this paper is that few econometricians or economists make use of power properties of hypothesis tests in applications. To see this, one only needs to peruse those journals where applied econometric research is published. Rarely is power mentioned. Zellner (1980) and McCloskey (1985a, b) substantiate this casual finding in their surveys of the use of power in applied econometrics papers published in leading journals.

The reason that power has been neglected in econometric applications appears to be due to difficulties involved in (i) calculating power, (ii) determining its qualitative features, (iii) summarizing these features, and (iv) relating them to a particular testing problem at hand. Power functions are multi-dimensional objects. In a parametric hypothesis test, the power surface is defined on a space whose dimension equals that of the parameter--including all nuisance parameters. In econometric applications especially, this can lead to a very high dimensional object, whose qualitative characteristics may be difficult to ascertain. In turn, this makes it difficult to use power functions to help interpret the results of a test, or to help select an appropriate trade off between significance level and power, if desired.

The summary measures developed in this paper are designed to overcome these problems. They provide a solution to the problem of high dimensionality, and relate straightforwardly to the interpretation of test results.

We now outline the importance of using power properties of tests in applications. A consequence of the standard framework for testing is that a failure to reject the null hypothesis cannot provide evidence that all deviations from the restrictions (i.e., deviations of $h(\theta)$ from $\underset{\sim}{0}$) are inconsistent with the data. By "inconsistent with the data," we mean that the *a priori* probability of not rejecting the null is "small" for the parameter value in question. At best, a failure to reject the null can provide evidence that *certain* deviations from the restrictions are inconsistent with the data. Thus, when the null is not rejected, one would like to know the range of such deviations. In addition, one may find it useful to know which deviations from the restrictions the test finds hard to distinguish from the null hypothesis, in the sense that their being true yields acceptance of the null to be more likely than rejection.

Can the magnitude of a t-statistic, F-statistic, or chi-square test statistic provide such information? The answer is seen easily to be no. Consider a standard t-test that a coefficient is zero in a linear regression model. A t-statistic equal to 1 can be generated by coefficient and coefficient standard errors of .01 and .01, respectively, as well as by estimates of 10^6 and 10^6. Obviously, the range of coefficient values (or equivalently, deviations of the coefficient values from zero) that are inconsistent with the data is quite different in these two cases, even though the t-statistics are the same. Analogous arguments hold for F-statistics and for any of the multitude of test statistics that are asymptotically chi-square under the null hypothesis.

Thus, if one only has knowledge of the value of a test statistic and its critical value, the failure of the test to reject the null is not very informative. It provides no evidence to suggest that certain deviations from the restrictions are not true. For example, by itself, Hall's (1978) failure to reject the hypothesis that the coefficient on lagged income is zero in his consumption function, provides no support for the conclusion that the true coefficient is zero or close to zero. Similarly, Lillard and Aigner's (1984) failure to reject the hypothesis of exogeneity of air conditioning appliance ownership variables in their demand for electricity equation, by itself, does not justify concluding that the variables are exogenous, and selecting the model that treats them as such. As a third example, it is common to subject a model to various specification tests, and to report the values of the test statistics, and the corresponding critical values for some conventional significance level, e.g., see McAleer, Pagan, and Volker (1985). If the specification tests (or test) fail to reject the null hypotheses, as occurs in the McAleer et al. example, and in countless applications of the Durbin-Watson test, this tells one nothing about the adequacy of the model, unless something is known about the power properties of the tests.

Inverse power approximations are designed to help interpret the results of such tests. Using inverse power summary measures, one sees the range of coefficient values that are likely to be detected by the test, and also the range that have a good chance of going undetected by the test. Hence, one can evaluate the information provided by the test, in the event that it fails to reject.

We note that the present approach can be used to develop summary measures of power for encompassing tests (e.g., see Mizon and Richard (1986)), and various other non-classical tests, in addition to the classical tests discussed below.

The paper is organized as follows: Section 2 defines the exact power radius and inverse power functions. It also describes the approach taken regarding nuisance parameters, and discusses the choice of directions of interest, and the choice of probabilities used to define the regions of "low" and "high" power. Section 3 introduces the approximate inverse power function. The asymptotic justifications of this function are discussed, and its use is exemplified with a test of multiple restrictions taken from a paper by Barro (1978). Section 4 defines the estimated inverse power function and gives its asymptotic justifications. An example is provided from a paper by Lillard and Aigner (1984). Section 5 provides regularity conditions under which the asymptotic justifications hold for the approximations given in Section 3. The Appendix contains additional results regarding asymptotic justifications, and provides proofs of results stated in the paper.

2. EXACT INVERSE POWER FUNCTIONS

2.1. The Restrictions and the Treatment of Nuisance Parameters

We consider null hypotheses that are defined by certain nonlinear restrictions on a parameter θ. The distributions of the data under the maintained hypothesis are indexed by θ, which lies in the parameter space $\Theta \subset R^{\ell}$. The restrictions are given by an R^q-valued function $h(\cdot)$, i.e., under the null, $h(\theta) = \underset{\sim}{0}$. We assume that there exists a parameter space $T \subset R^{\ell}$ and a one-to-one transformation $\nu : \Theta \to T$ such that (i) $h(\theta)$ equals the first q elements of $\tau = \nu(\theta) \in T$, i.e., $h(\theta) = h(\nu^{-1}(\tau)) = \tau_1$ for $\tau = (\tau_1, \tau_2) \in T$, and (ii) given any $(\tau_1, \tau_2) \in T$, $(\underset{\sim}{0}, \tau_2)$ also is in T, where $\tau_1, \underset{\sim}{0} \in R^q$. We call this the <u>nuisance parameter condition</u>. Part (ii) of this condition often is satisfied automatically, since we often have $T = T_1 \times T_2$, where $T_1 \subset R^q$ and $T_2 \subset R^{\ell-q}$.

The ℓ-vector $\nu(\theta) = \tau$ can be partitioned as $\begin{bmatrix} \nu_1(\theta) \\ \nu_2(\theta) \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}$, where $\nu_1(\theta) = \tau_1$ is a q-vector of parameters whose values are restricted under the null hypothesis to be $\underset{\sim}{0}$, and $\nu_2(\theta) = \tau_2$ is an ℓ-q vector of nuisance parameters. The approach we take is to fix the nuisance parameters at some value $b \in R^{\ell-q}$, i.e., set $\nu_2(\theta) = \tau_2 = b$, and analyze the power of a test for fixed b. One usually will choose to consider the power properties of a test when b equals the estimated value of the nuisance parameters. (If one is concerned about this choice of b, it is always possible to carry out a sensitivity analysis by repeating the power calculations using several other appropriate values of b.) The distinct advantage of fixing the nuisance parameter vector at b is that the dimension of the domain of the power function, power radius functions, and inverse power functions (defined below) is reduced considerably.

For fixed b, there is only one parameter vector in T for which the null hypothesis holds, viz., $\tau_b = \begin{bmatrix} \underset{\sim}{0} \\ b \end{bmatrix}$. This vector corresponds to a unique vector θ_b in Θ via $\theta_b = \nu^{-1}(\tau_b)$.

2.2. Power Radius Functions and Inverse Power Functions

When a test fails to reject the null hypothesis, we are interested in what this implies with regard to the restrictions being true or being close to true. Thus, it is desirable from an interpretative standpoint to relate power to deviations from the restrictions (i.e., deviations of $h(\theta)$ from the zero vector). For this reason, we define the inverse power function to be a function whose values correspond to deviations of $h(\theta)$ from $\underset{\sim}{0}$, rather than to parameter values θ.

Consider the deviation space $D \subset R^q$, which consists of values that the vector of restrictions $h(\theta)$ assumes for different $\theta \in \Theta$. The origin represents the null hypothesis in this space. Any unit vector η in R^q defines a direction in the deviation space that corresponds to the ray from the origin through η. For a fixed value b of the nuisance parameters, we assess the power properties of a test in the direction η, for given directions η of interest.

Suppose the test in question is based on a statistic T_n and has exact power function $\gamma_n(\theta)$, for given significance level α. We ask: What is the greatest distance along the ray defined by η such that the power of the test T_n is less than or equal to p for all alternatives that correspond to points on the ray closer to the origin? As a function of p, these distances define the inner power radius function of the test, which is denoted by

$$(2.1) \quad r^n_{b\eta}(p) = r_\eta(p) = \sup\{\|h(\theta)\| : \theta \in \Theta,\ h(\theta) \propto \eta,\ \gamma_n(\theta) \le p,\ \nu_2(\theta) = b\}$$

where "$\propto$" denotes positively proportional to (i.e., $\eta_1 \propto \eta_2$ iff $\eta_1 = c\eta_2$ for some $c \ge 0$).

The points on the ray through η that correspond to the distances $r^n_{b\eta}(p)$ are defined to be the values of the inner inverse power function $iip^n_{b\eta}(p)$ of T_n in direction η. That is, the inner inverse power function in direction η is

$$(2.2) \qquad iip^n_{b\eta}(p) = iip_\eta(p) = r^n_{b\eta}(p)\cdot\eta \ .$$

With knowledge of the inner inverse power function, one can answer questions such as: Which deviations from the restrictions have a good chance of going undetected by the test? The answer is given quite simply using the inner inverse power function by choosing a value p_1 that corresponds to "a good chance" and by choosing several directions η of interest. For any deviation vector in direction η closer to the origin than $\text{iip}_\eta(1 - p_1)$, the test must have acceptance probability greater than or equal to p_1. Thus, if the test fails to reject, it provides "little or no evidence" to suggest that the true deviation of the restrictions is any closer to $\underset{\sim}{0}$ than $\text{iip}_\eta(1 - p_1)$.

The <u>outer power radius function</u>, denoted $s^n_{b\eta}(p)$, and the <u>outer inverse power function</u>, denoted $\text{oip}^n_{b\eta}(p)$, are defined analogously:

$$(2.3)\quad s^n_{b\eta}(p) = s_\eta(p) = \inf\{\|h(\theta)\| : \theta \in \Theta,\ h(\theta) \propto \eta,\ \gamma_n(\theta) \ge p,\ \nu_2(\theta) = b\}\ ,$$

$$(2.4)\qquad \text{oip}^n_{b\eta}(p) = \text{oip}_\eta(p) = s^n_{b\eta}(p)\cdot\eta\ .$$

The outer inverse power function can be used to answer questions such as: Which deviations from the restrictions have a very good chance of being detected by the test? Here, we just need to choose p_2 that corresponds to "a very good chance" and choose several directions η of interest. Any deviation vector in direction η that is farther from the origin than $\text{oip}_\eta(p_2)$ has a very good chance of being detected. Thus, if the test fails to reject, it provides "strong" evidence that such deviations from the restrictions are not true.

We mention that if the power function of T_n is monotone along rays from the origin, then $r_\eta(p) = s_\eta(p)$ and $\text{iip}_\eta(p) = \text{oip}_\eta(p)$. Also, by definition of $\text{iip}_\eta(p)$ and $\text{oip}_\eta(p)$,

$$(2.5)\qquad \sup_{\theta\in A_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) = p \quad\text{and}\quad \inf_{\theta\in B_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) = p\ ,$$

where $\quad A_n(b,\eta,p) = \{\theta \in \Theta : h(\theta) \le \text{iip}^n_{b\eta}(p),\ \nu_2(\theta) = b\}\ ,$

$$B_n(b,\eta,p) = \{\theta \in \Theta : h(\theta) \ge \text{oip}^n_{b\eta}(p),\ \nu_2(\theta) = b\}\ ,$$

"$\le$" is defined for vectors as: $\eta_1 \le \eta_2$ iff $\eta_1 = c\eta_2$ for some c in $[0,1]$, and $k_{q,\alpha}$ is the critical value used for the test when the significance level is α.

It remains to discuss directions η of interest, and appropriate choices of p_1 and p_2. When there is only one restriction ($q = 1$), there are only two directions, $\eta = 1$ and $\eta = -1$, so the choice of η is simple. In practice, the situation is even simpler, because the approximate power radius function (described below) is the same for η and $-\eta$.

When there are multiple restrictions ($q > 1$), there are at least three ways of determining directions of interest. First, the problem itself and previous experience usually suggest directions to consider. For example, if the null hypothesis restricts the coefficients on a variable and each of its lags to be zero, then it is natural to look in the directions where only the contemporaneous coefficient is non-zero, and where the coefficients satisfy plausible lag patterns, such as linear decay to zero. Or, if the null hypothesis restricts several autoregressive parameters to be zero, one might look in the direction corresponding to a first-order autoregressive model.

A second direction that may be of interest is the direction corresponding to the estimated deviation vector, viz., $h(\hat{\theta})$. This direction is of greater interest, the greater is the precision of the estimate $h(\hat{\theta})$.

A third category of directions of interest are those directions that are of interest due to the power properties of the test itself. In particular, it may be useful to determine those directions where the test has greatest power and lowest power, or equivalently, smallest outer power radius and largest inner power radius, respectively.

By their very nature, choices for p_1 and p_2 are arbitrary. Nevertheless, arguments can be mustered in favor of certain choices. We do this for the choices $p_1 = 1/2$ and $p_2 = 1 - \alpha$, because we feel these values yield suitable summary measures. If some other values seem more appropriate in a given context, then of course, they should be used instead.

The choice of $p_1 = 1/2$ seems natural; due to symmetry, it is an obvious focal point. Even odds or better represent "a good chance" of failing to detect a particular deviation. Furthermore, when the test has less than even odds for detecting a particular deviation, then the toss of a fair coin, with arbitrary choice of rejecting the null on heads or tails, has greater ability to detect this deviation than the test does. In this case, it seems appropriate to classify such deviations as "difficult" for the test to detect. These "arguments" in favor of taking $p_1 = 1/2$ are independent of the significance level α. This makes $p_1 = 1/2$ a convenient general choice for specifying the boundary of a region of low power.

When p_1 is set equal to one-half, the summary measures $r_\eta(1/2)$ and $iip_\eta(1/2)$ can be used to answer the first question raised above. We call these summary measures the even odds (EO) radius in direction η and the even odds (EO) vector in direction η, respectively.

Next, consider the choice of $p_2 = 1 - \alpha$. This corresponds to a type II error of magnitude α. This choice allows one to obtain evidence against certain deviations being true, when the test fails to reject, that is of comparable strength to the evidence against the null obtained when the test does reject. Furthermore, the fact that our choice of error probability $1 - p_2$ is not fixed, but can vary from case to case by varying α, circumvents the same sort of criticism that often is raised against the practice of setting the significance level without regard to power.

Given the choice of $p_2 = 1 - \alpha$, the summary measures $s_\eta(1 - \alpha)$ and $oip_\eta(1 - \alpha)$ can be used to answer the second question posed above. We refer to these summary measures as the power/size (PS) radius in direction η and the power/size (PS) vector in direction η, respectively, where "power/size" refers to power being comparable to size.

3. APPROXIMATE INVERSE POWER FUNCTIONS

3.1. *Definition of the Approximations*

First, we define the Wald test statistic. Suppose $\hat{\theta}$ satisfies $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} Z_\theta \sim N(\underset{\sim}{0}, V(\theta))$ as $n \to \infty$, when θ is true, for any $\theta \in \Theta$, and $H(\theta) = \frac{\partial}{\partial \theta} h(\theta)$ is continuous at all points θ in the null hypothesis. Define $\hat{H} = H(\hat{\theta})$, $\hat{V} = V_n(\hat{\theta})$, and $\hat{\Sigma} = \hat{H}\hat{V}\hat{H}'$, where $V_n(\hat{\theta})$ converges in probability to $V(\theta)$ as $n \to \infty$ when θ is true, for all $\theta \in \Theta$. The Wald test statistic is

$$W_n = nh(\hat{\theta})'(\hat{H}\hat{V}\hat{H}')^{-1}h(\hat{\theta}) = nh(\hat{\theta})'\hat{\Sigma}^{-1}h(\hat{\theta}) \ , \tag{3.1}$$

where $\hat{\Sigma}$ is assumed to be nonsingular with probability that goes to one as $n \to \infty$ (see Andrews (1985a) for the case of singular $\hat{\Sigma}$).

Under the null hypothesis, W_n has an asymptotic chi-square distribution with q degrees of freedom (denoted χ^2_q) . Thus, for a test with significance level α, the critical value $k_{q,\alpha}$ is chosen to satisfy

$$P(X^2_q > k_{q,\alpha}) = \alpha \ , \quad \text{where} \quad X^2_q \sim \chi^2_q \ . \tag{3.2}$$

For the Wald test, we approximate both the inner and outer power radius functions, $r_\eta(p)$ and $s_\eta(p)$, by the approximate power radius function $R^n_{b\eta}(p)$, and both the inner and outer inverse power functions, $iip_\eta(p)$ and $oip_\eta(p)$, by the approximate inverse power function $IP^n_{b\eta}(p)$, where

$$R^n_{b\eta}(p) = R_\eta(p) = \frac{1}{\sqrt{n}}\, \lambda_{q,\alpha}(p)\cdot\left(\eta'\Sigma^{-1}_{nb}\eta\right)^{-1/2}\ , \quad \text{and}$$

(3.3)

$$IP^n_{b\eta}(p) = IP_\eta(p) = R^n_{b\eta}(p)\cdot\eta\ .$$

The non-random quantities $\lambda_{q,\alpha}(p)$ and Σ_{nb} are defined as follows: Let $H_b = H(\theta_b)$, $\bar{V}_{nb} = \bar{V}_n(\theta_b)$, and $\Sigma_{nb} = H_b\bar{V}_{nb}H_b'$, where θ_b is defined in Section 2.1 above, and $\bar{V}_n(\theta)$ is non-random and converges to $V(\theta)$ as $n \to \infty$, $\forall\theta\epsilon\Theta$. For example, if $\hat{\theta}$ is the unrestricted ML estimator, then $\bar{V}_n(\cdot)$ can be taken to be the information matrix of the sample of size n, divided by n, and inverted. Or, if the matrix $V_n(\cdot)$ used in the definition of the Wald statistic is non-random, then $\bar{V}_n(\cdot)$ can be taken to equal $V_n(\cdot)$. Note that Σ_{nb} converges as $n \to\infty$ to the asymptotic covariance matrix of $\sqrt{n}(h(\hat{\theta}) - h(\theta))$ evaluated at the null hypothesis point $\theta = \theta_b$, viz., $\Sigma_b = H_bV_bH_b'$ where $V_b = V(\theta_b)$.

To define $\lambda_{q,\alpha}(p)$, let $\gamma_{q,\alpha}(\delta^2) = P(X^2 > k_{q,\alpha})$, where X^2 has non-central chi-square distribution with q degrees of freedom and non-centrality parameter δ^2 (denoted $\chi^2_q(\delta^2)$) and $k_{q,\alpha}$ is as defined above. The function $\gamma_{q,\alpha}(\cdot) : R^+ \to [0,1]$ is strictly increasing, and hence, has a unique inverse $\gamma^{-1}_{q,\alpha}(\cdot) : [0,1] \to R^+$. By definition, $\lambda_{q,\alpha}(p) = \sqrt{\gamma^{-1}_{q,\alpha}(p)}$. Thus, $\lambda_{q,\alpha}(p)$ is the square root of the inverse of the power function of a level α test based on a test statistic X^2 with $\chi^2_q(\delta^2)$ distribution.

We tabulate $\lambda_{q,\alpha}(p)$ in Tables 1, 2, and 3, for α = .05, .01, .1; p = .1(.1).9, .95, .99; and q = 1(1)30(2)50(25)100. In addition, we note that Haynam, Govindarajulu, and Leone (1962) tabulate $\gamma^{-1}_{q,\alpha}(p)$ for α = .001, .005, .01, .025, .05, .01 ; p = .1(.02).7(.01).99; and q = 1(1)30(2)50(5)100. Also, Yamauti (1972) provides a FORTRAN program that can be used to obtain $\gamma^{-1}_{q,\alpha}(p)$ and $\lambda_{q,\alpha}(p)$ for combinations of α, p, and q that have not been tabulated.

The approximate power radius and inverse power functions, $R_\eta(p)$ and

$IP_\eta(p)$, are extremely simple. It is easy to isolate the affects of different values of p, α, η, b, and n on the approximate inverse power of the test, and hence, to ascertain the general qualitative features of the power of the test.

For fixed power p, the surface in the deviation space D mapped out by $IP_\eta(p)$ for varying η is an ellipse defined by

(3.4) $$\{\xi \in R^q : \xi'(\Sigma_{nb}/n)^{-1}\xi = \lambda^2_{q,\alpha}(p)\} \quad .^2$$

The approximations for LR and LM tests are exactly the same as those given for the Wald test in (3.3), when the Wald test is based on the unrestricted ML estimator. For cases where the Hausman test can be interpreted in the classical framework for parametric hypothesis tests, and is equivalent to the LR and LM tests under the null and local alternatives (see Holly (1982)), the same approximations apply as for the LR and LM tests.

Approximate analogues of the summary measures $iip_\eta(1/2)$ and $oip_\eta(1-\alpha)$ are $IP_\eta(1/2)$ and $IP_\eta(1-\alpha)$, respectively. These vectors are called the <u>approximate even-odds vector</u> in direction η, EO_η, and the <u>approximate power/size vector</u> in direction η, PS_η, respectively. Thus, $EO_\eta = IP_\eta(1/2)$ and $PS_\eta = IP_\eta(1-\alpha)$. Similarly, the <u>approximate even-odds radius</u> in direction η refers to $R_\eta(1/2)$, and the <u>approximate power/size radius</u> in direction η refers to $R_\eta(1-\alpha)$.

Next, we state the asymptotic justifications for the approximations given in (3.3). These justifications are established under a specific set of assumptions in Section 5 below. Let

(3.5)
$$\tilde{A}_n(b,\eta,p) = \tilde{A}_n = \{\theta \in \Theta : h(\theta) \le IP^n_{b\eta}(p),\ \nu_2(\theta) = b\} \ , \quad \text{and}$$
$$\tilde{B}_n(b,\eta,p) = \tilde{B}_n = \{\theta \in \Theta : h(\theta) \ge IP^n_{b\eta}(p),\ \nu_2(\theta) = b\} \ ,$$

(where $\beta_1 \leq \beta_2$ for $\beta_1, \beta_2 \in R^q$ iff $\beta_1 = c\beta_2$ for some c in [0,1]).

The first justification we give is related directly to the intended use of approximate inverse power functions. In practice, we want to be able to say: For any deviation vector in direction η that is closer to (farther from) the origin than $IP_\eta(p)$ (i.e., for any $h(\theta)$ with $\theta \in \tilde{A}_n$ ($\theta \in \tilde{B}_n$)), the probability that the test rejects is less (greater) than or equal to a constant that is approximately equal to p. These statements are justified by the following results:

$$
(3.6) \qquad
\begin{aligned}
&\sup_{\theta \in \tilde{A}_n(b,\eta,p)} P_\theta(W_n > k_{q,\alpha}) \xrightarrow{n\to\infty} p \ , \quad \text{and} \\
&\inf_{\theta \in \tilde{B}_n(b,\eta,p)} P_\theta(W_n > k_{q,\alpha}) \xrightarrow{n\to\infty} p \ ,
\end{aligned}
$$

for any nuisance parameter vector b in the transformed parameter space, any unit direction vector η in R^q, and any p in [0,1] , under the assumptions given in Section 5. Analogous results hold for likelihood ratio, Lagrange multiplier, and Hausman tests, see Section 5.

The second justification of $IP_\eta(p)$ is in terms of its ability to approximate $iip_\eta(p)$ and $oip_\eta(p)$ directly:

$$
(3.7) \qquad
\begin{aligned}
&\sqrt{n}\left[IP^n_{b\eta}(p) - iip^n_{b\eta}(p)\right] \xrightarrow{n\to\infty} \underset{\sim}{0} \ , \quad \text{and} \\
&\sqrt{n}\left[IP^n_{b\eta}(p) - oip^n_{b\eta}(p)\right] \xrightarrow{n\to\infty} \underset{\sim}{0} \ ,
\end{aligned}
$$

for any b, η, and p as above, under the assumptions of Section 5. Additional justifications for $IP^n_{b\eta}(p)$, which explicitly treat the case of estimated nuisance parameter vectors, are given in the Appendix.

3.2. Discussion of the Approximations

Clearly, the inverse power approximations given here are closely related to local power approximations (e.g., see Pitman (1979), Engle (1984), and Rothenberg (1984a)), since both rely on a non-central chi-square approximation of the test statistic. There are several advantages to our approach, however, in contrast to that of reporting the local power for several values of θ, or to that of reporting the chi-square non-centrality parameter as a function of θ. First, our approach reduces the dimensionality of the power function, whose domain is ℓ dimensional, to a function that is q dimensional. Our approach provides a simple method for reducing this function further to just several one dimensional functions. Finally, these one dimensional functions can be evaluated at two points to give convenient and easily interpretable summary measures. In contrast, the local power function allows one to calculate power at several points θ in a high dimensional space, but it is far from clear which points θ might yield useful summary measures.

Second, inverse power approximations provide ranges of deviations from the null hypothesis that correspond to a given power p, rather than providing the value of p for a given parameter vector θ. This approach is desirable in terms of generating summary measures, because it is much easier to choose values of p that are satisfactory to a wide range of people, than it is to choose particular values of θ. The latter are difficult to agree upon because of differing views of which alternatives θ are of interest.

Third, the asymptotic justifications we provide for the approximate inverse power function do not restrict consideration to just local sequences of alternatives. Rather, the nature of the power properties of the tests themselves implies that local alternatives are important asymptotically. Thus, the inverse power function provides an alternative view of local alternatives --one that makes them appear quite natural.

In sum, inverse power approximations and local power approximations are closely related, but the former seem more useful for present purposes.

For sample sizes often encountered in practice, inverse power approximations, like the nominal significance level, may be rather crude. Nevertheless, these approximations provide some information on power, crude or otherwise, in a form that is readily employable in a wide variety of econometric testing situations. Even crude information on power is quite useful, since it establishes the orders of magnitude involved.

If more accurate information on power is required than that provided by inverse power approximations, one can refine the approximations via higher order expansions (e.g., see Rothenberg (1984a,b) and Sargan (1980)), or make use of exact series expansions for power if they are available (e.g., see Phillips (1983)).

Next we make two comments on the use of confidence intervals or regions as an alternative means of interpreting test results. First, confidence regions can be associated naturally with Wald tests, but not with likelihood ratio, Lagrange multiplier, or Hausman tests. Thus, their application is somewhat restricted for our purposes. Furthermore, while confidence <u>intervals</u> are easy to interpret, multidimensional confidence regions are not, even though they are easy to define. This is evidenced by their scarce use in econometric practice.

Second, to interpret test results one needs to investigate the probabilities of the different errors associated with the tests, viz., errors of type I and type II. Inverse power approximations provide summary measures for type II error probabilities. They rely on probability calculations that hold under alternative parameter values. Confidence intervals or regions, on the other hand, do not assess type II error probabilities. They do not rely on calculations of rejection probabilities under alternative parameter values, and hence, cannot address the question: For which deviations from the null hypothesis does the test have high or low power? As an alternative form of inference to testing, however, confidence regions are quite valuable, especially when one is dealing with a single parameter.

3.3. Calculation of η_*, η^*, and Σ_{nb}

Determination of the directions that minimize and maximize the approximate power radius function $R_\eta(p)$ is simple. Since $R_\eta(p)$ depends on η only through $(\eta'\Sigma_{nb}^{-1}\eta)^{-1/2}$, $R_\eta(p)$ is minimized for any p by the eigenvector η_* of Σ_{nb}^{-1} that corresponds to the largest eigenvalue μ_* of Σ_{nb}^{-1}, and $R_\eta(p)$ is maximized for any p by the eigenvector η^* of Σ_{nb}^{-1} that corresponds to the smallest eigenvalue μ^* of Σ_{nb}^{-1}. Call η_* the (approximate) <u>direction of smallest power radius</u>, and η^* the (approximate) <u>direction of largest power radius</u>. The relative magnitude of the smallest and largest power radii tells one the extent to which power varies with the direction chosen.

Next we consider the choice of b. We are free to choose any vector b of interest and compute $IP_{b\eta}^n(p)$. It is natural, however, to choose b as given by some estimate B of the true nuisance parameter vector b_0. It is preferable to use an estimator B of b_0 that corresponds to an unrestricted estimator of θ, such as $\hat{\theta}$. The reason is that an estimator of b_0 that corresponds to a restricted estimator of θ may be a poor estimator of b_0 if the null hypothesis is false.

Now we discuss the calculation of Σ_{nb} in certain special cases. Suppose

$h(\theta)$ constrains the first q elements of θ to be $\underset{\sim}{0}$. Then, $\Sigma_{nb} = [\bar{V}_{nb}]_{q\times q}$ and $IP_{\eta}(p) = \lambda_{q,\alpha}(p)\cdot\left(\eta'([\bar{V}_{nb}]_{q\times q}/n)^{-1}\eta\right)^{-1/2}$, where $[\bar{V}_{nb}]_{q\times q}$ is the upper left $q \times q$ submatrix of $\bar{V}_{nb}$. In this case, $[\bar{V}_{nb}]_{q\times q}/n$ is just a particular estimator of the covariance matrix of the first q elements of $\hat{\theta}$. In particular, if $q = 1$, then $IP(p) = \lambda_{q,\alpha}(p)\cdot\left([\bar{V}_{nb}]_{1\times 1}/n\right)^{1/2}$, where $\left([\bar{V}_{nb}]_{1\times 1}/n\right)^{1/2}$ is a particular standard error estimator of the first element of $\hat{\theta}$.

For example, consider a classical linear regression model $y = X\beta+u$ with errors u that have mean $\underset{\sim}{0}$ and variance $\sigma^2 I$ (and are not necessarily normally distributed). Let $\theta = (\beta, \sigma^2)$. Suppose one uses the standard F-statistic to test whether the first q elements of β equal $\underset{\sim}{0}$. In this case, the matrix Σ_{nb}/n equals $[(X'X)^{-1}]_{q\times q}\sigma_b^2$, where σ_b^2 is the error variance that corresponds to the chosen nuisance parameter vector b. If the nuisance parameter vector is estimated using the unrestricted least squares estimators, then σ_b^2 equals the unrestricted estimator of σ^2, and $[(X'X)^{-1}]_{q\times q}\sigma_B^2$ is just the usual covariance matrix of the least squares estimators of the q elements of β that are equal to zero under the null. In particular, if $q = 1$, the approximate inverse power function is just $\lambda_{q,\alpha}(p)$ times the usual standard error estimator of the least squares estimator of the coefficient that is restricted to be zero under the null.

As a second example, consider the Durbin-Watson test that the first-order autoregressive parameter, θ_1, of the errors is zero in a standard linear regression model. From above, $\Sigma_{nb} = [\bar{V}_{nb}]_{1\times 1}$. The approximate inverse power results for the Durbin-Watson test are equal to those for a Wald test based on the sample serial correlation coefficient. For the latter, we find that $V(\tilde{\theta}) = 1$ for all $\tilde{\theta}$ such that $\tilde{\theta}_1 = 0$. Thus, for any b, $IP(p) = \lambda_{1,\alpha}(p)/\sqrt{n}$, $EO = \lambda_{1,\alpha}(1/2)/\sqrt{n}$, and $PS = \lambda_{1,\alpha}(1-\alpha)/\sqrt{n}$. In Table 4, the even-odds and power/size radii are tabulated for several values of n and α. This Table shows that for $\alpha = .01$ fairly large sample sizes are needed to achieve reasonable power.

3.4. *An Example with Multiple Restrictions*

As an example with multiple restrictions, consider Barro's (1978) test of the hypothesis that only unanticipated money growth influences output. His null hypothesis is that the coefficients on a money growth variable and each of its three lags are zero, in a regression equation for output that includes unanticipated money growth variables. He reports the value of a Wald test statistic (viz., .2 as compared to a level .05 critical value of 2.9), and accepts the hypothesis that the money growth coefficients all equal zero.

The information Barro reports is not sufficient to indicate the strength of the evidence that these coefficients are zero or close to zero. Inverse power approximations can be used to provide such information. We compute such approximations with b corresponding to the unrestricted least squares estimators.[3] Thus, $\Sigma_{nB}/n = [(X'X)^{-1}\sigma_B^2]_{4\times4}$, where the latter is the 4×4 sub-matrix of the estimated least squares covariance matrix that corresponds to the four money growth variables. The matrix $(\Sigma_{nB}/n)^{-1}$, that arises in $IP_\eta(p)$, is given in Table 5.

First, we consider inverse power in several directions that represent *a priori* plausible lag structures: $\eta_1 = (2,3,2,1)/\sqrt{18}$, $\eta_2 = (4,3,2,1)/\sqrt{30}$, and $\eta_3 = (1,0,0,0)$. For each of these directions, Table 6 gives the even-odds and power/size vectors, as well as the power radii and sum of the coefficients. Using the matrix $(\Sigma_{nB}/n)^{-1}$ reported in Table 5, and the tabulated values of $\lambda_{4,.05}(p)$ in Table 1, analogous results can be computed for other directions η and other probabilities p.

To interpret the inverse power results, one needs to have an idea of what constitutes "large" and "small" money growth coefficients from an economic perspective. The sum of the coefficients on money growth and its lags gives the percentage increase in output in a year due to a 1% increase in money growth in the same year and in each of the three previous years with other variables (including unanticipated money growth) held constant. A similar interpretation applies to the sum of the coefficients on the unanticipated money growth variable and its three lags. The latter coefficients are estimated to be (.71, .92, .65, .12) with their sum equaling 2.4. Barro interprets these coefficients to be "large" or "significant" from an economic perspective, and finds their lag pattern to be reasonable. Hence, this coefficient vector provides a reasonable standard of comparison for coefficient vectors of the money growth variables.

Consider the approximate inverse power in direction η_1. From Table 6, we have EO_{η_1} = (.09, .14, .09, .05) and PS_{η_1} = (.16, .24, .16, .08) with their coefficient sums being .37 and .64, respectively. Thus, in direction η_1, the test provides strong evidence that the sum of the coefficients is .64 or less, and that each element of the vector is quite "small," in comparison with the estimated coefficients on unanticipated money growth variables (subject to the validity of the maintained hypothesis of correct model specification, of course). Results for directions η_2 - η_4 can be interpreted similarly.

Next, we consider the directions of largest and smallest power radii, given in Table 6 by η^* and η_*, respectively. The considerable difference between the power radii in these two directions indicates that the power of the present test varies substantially with the direction η. In direction η^*, the test has low power (from an economic perspective). The lag pattern corresponding to this direction does not seem plausible, however, because of the large negative coefficient on the first lag (or, in the opposite direction, on the contemporaneous coefficient). In contrast, the direction η_* of smallest power radius is more plausible. The power in this direction is quite high.

In sum, Barro's test has good to very good power in directions corresponding to plausible lag structures. In some less plausible directions, however, the test has much lower power.

4. ESTIMATED INVERSE POWER FUNCTIONS

4.1. Motivation and Definition

This section introduces an approximation to the inverse power function, called the estimated inverse power function, that is easier to calculate than the approximate inverse power function of Section 3. The former uses a random matrix $V_n(\theta)$ in place of the non-random matrix $\bar{V}_n(\theta)$ of the latter. This approximation is useful for three reasons. First, in complicated models, a non-random matrix $\bar{V}_n(\theta)$ may be difficult to compute, whereas a random analogue $V_n(\theta)$ often is readily available. For example, with ML procedures, $\bar{V}_n(\theta)$ requires computation of the "true" information matrix (which involves taking expectations), whereas $V_n(\theta)$ can be given by a "sample" information matrix (which omits these expectations).

Second, if one is using a Wald test and the parameter θ does not specify the distribution of the data completely (e.g., it may leave the distribution of certain errors unspecified), then a non-random matrix $\bar{V}_n(\cdot)$ that converges to $V(\cdot)$ may not exist.

Third, for interpreting test results presented by others, it may not be possible to use the reported information to calculate the approximate inverse power function, but this information may suffice for the estimated inverse power function. If fact, with the very common test of whether a single parameter equals a given constant, the estimated inverse power function usually can be determined from reported information. Thus, it can be used to help one evaluate results presented in numerous applied econometrics papers.

Let b_0 denote the true nuisance parameter vector $\nu_2(\theta_0)$, where the true parameter vector θ_0 may or may not satisfy the restrictions. The approximation we give here to the inner and outer inverse power functions holds with their nuisance parameter vectors equal to b_0, rather than equal to any vector of our choice. Corresponding to b_0 is the null parameter point $\theta_{b_0} = \nu^{-1}(\underline{0}', b_0')$ (see Section 2.1). We consider sequences of local alternatives $\{\theta_n\}$ to θ_{b_0}. In particular, let $\{\theta_n\} = \{\theta_{b_0} + \delta_n/\sqrt{n} :$ $n = 1, 2, \ldots\}$, for some $\delta_n \in R^\ell$ with $\|\delta_n\| \le M$ for some $M < \infty$, for all n.

The estimated inverse power function is defined first for the Wald test. As above, let $\hat{\theta}$ be an unrestricted estimator of θ. The asymptotic covariance matrix of $\sqrt{n}(h(\hat{\theta}) - h(\theta_n))$ under the sequence of local alternatives $\{\theta_n\}$ is $\Sigma_{b_0} = H_{b_0} V_{b_0} H_{b_0}'$. Let $\hat{\Sigma}_n$ be any estimator of this asymptotic covariance matrix that is consistent under these sequences of local alternatives. Since local alternatives generally are contiguous for "regular" models, it is usually the case that any estimator $\hat{\Sigma}_n$ that is consistent for Σ_{b_0} under the null hypothesis point θ_{b_0} also is consistent under the local alternatives $\{\theta_n\}$. In consequence, there exist ample choices for $\hat{\Sigma}_n$. The most obvious choice is the matrix $\hat{\Sigma}$ used in the definition of the Wald statistic. In any case, $\hat{\Sigma}_n$ usually will be chosen to be of the form $H(\tilde{\theta}_n)\tilde{V}_n H(\tilde{\theta}_n)'$, where $\tilde{V}_n$ is some estimator of the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta} - \theta_{b_0})$ that converges in probability to V_{b_0} under θ_{b_0}, and $\tilde{\theta}_n$ is some estimator of θ that converges in probability to θ_{b_0} under θ_{b_0}.

Consider the Wald statistic W_n defined in Section 3.1 above. For the test based on W_n, the estimated power radius function $\hat{R}^n_\eta(p)$ and the estimated inverse power function $\hat{IP}^n_\eta(p)$ in direction η are defined as

$$\hat{R}^n_\eta(p) = \hat{R}_\eta(p) = \frac{1}{\sqrt{n}} \lambda_{q,\alpha}(p)\cdot(\eta'\hat{\Sigma}_n^{-1}\eta)^{-1/2} \left[= \lambda_{q,\alpha}(p)\cdot(\eta'(\hat{\Sigma}_n/n)^{-1}\eta)^{-1/2}\right] ,$$

(4.1) and

$$\hat{IP}^n_\eta(p) = \hat{IP}_\eta(p) = \hat{R}^n_\eta(p)\cdot\eta .$$

The estimated even-odds vector in direction η is $\hat{EO}_\eta = \hat{IP}_\eta(1/2)$, and the estimated power/size vector in direction η is $\hat{PS}_\eta = \hat{IP}_\eta(1 - \alpha)$.

If $\hat{\Sigma}_n$ is of the form $\hat{\Sigma}_n = H(\tilde{\theta}_n)\tilde{V}_n H(\tilde{\theta}_n)'$, then $\hat{\Sigma}_n/n = H(\tilde{\theta}_n)(\tilde{V}_n/n)H(\tilde{\theta}_n)'$, where $\tilde{V}_n/n$ is an estimator of the covariance matrix of $\hat{\theta}$. In particular, if the null hypothesis restricts certain elements of θ to equal constants, then $\hat{\Sigma}_n/n$ is just the submatrix of the estimated covariance matrix of $\hat{\theta}$ that corresponds to the restricted elements. In this case, $\hat{IP}_\eta(p)$ is exceptionally easy to compute, given a direction η of interest. If the null hypothesis restricts a single element θ_1 of θ to equal a given constant, then $(\eta'(\hat{\Sigma}_n/n)^{-1}\eta)^{-1/2}$ equals the estimated standard error, say $\hat{\sigma}_{\theta_1}$, of the corresponding element of $\hat{\theta}$ (since there are only two directions, $\eta = 1$ and $\eta = -1$, when $q = 1$). Here, $\hat{IP}(p)$ is simply $\lambda_{1,\alpha}(p)\cdot\hat{\sigma}_{\theta_1}$, and $\hat{IP}(p)$ can be calculated from information commonly reported in applied papers.

Note that one only needs to report a consistent estimator of the variance-covariance matrix of $\hat{\theta}$ in order for a reader to be able to compute an estimated inverse power function for a Wald, LR, LM, or Hausman test of any nonlinear restriction $h(\theta) = \underset{\sim}{0}$ (where $\hat{\theta}$ is the ML estimator if the test is an

LR, LM, or Hausman test).

Next, we state the asymptotic justifications for the estimated inverse power function $\hat{IP}_\eta(p)$. These justifications are established under specific conditions in the Appendix. Let

$$(4.2)\qquad \begin{aligned} \hat{A}_n(\eta,p) &= \hat{A}_n = \{\theta \in \Theta : h(\theta) \le \hat{IP}^n_\eta(p),\ \nu_2(\theta) = b_0\}\ , \quad \text{and} \\ \hat{B}_n(\eta,p) &= \hat{B}_n = \{\theta \in \Theta : h(\theta) \ge \hat{IP}^n_\eta(p),\ \nu_2(\theta) = b_0\}\ , \end{aligned}$$

(where, "≤" and "≥" are as defined in (3.5)).

Under sequences of local alternatives $\{\theta_n\}$ to θ_{b_0}, the estimated inverse power function of the Wald test satisfies:

$$(4.3)\qquad \begin{aligned} &\text{(a)}\quad \operatorname*{plim}_{n\to\infty} \sup_{\theta \in \hat{A}_n(\eta,p)} P_\theta(W_n > k_{q,\alpha}) = p\ , \\ &\text{(b)}\quad \operatorname*{plim}_{n\to\infty} \inf_{\theta \in \hat{B}_n(\eta,p)} P_\theta(W_n > k_{q,\alpha}) = p\ , \\ &\text{(c)}\quad \sqrt{n}(\hat{IP}^n_\eta(p) - iip^n_{b_0\eta}(p)) \xrightarrow{p} \underset{\sim}{0}\ , \quad \text{and} \\ &\text{(d)}\quad \sqrt{n}(\hat{IP}^n_\eta(p) - oip^n_{b_0\eta}(p)) \xrightarrow{p} \underset{\sim}{0}\ , \end{aligned}$$

for any unit direction vector $\eta \in R^q$ and any $p \in [0,1]$, under the assumptions of the Appendix. These asymptotic justifications of $\hat{IP}_\eta(p)$ are weaker than those of $IP_\eta(p)$. The former hold only under sequences of local alternatives, whereas the latter hold under any fixed θ in Θ (and also can be shown to hold under sequences of local alternatives).

Under well-known general conditions, the Wald, LR, LM, and Hausman test statistics are asymptotically equivalent under sequences of local alternatives. In consequence, the estimated power radius and inverse power functions given above for the Wald test also apply to the LR, LM, and Hausman tests.

4.2. An Example

As an example, consider Lillard and Aigner's (1984) analysis of time-of-day (TOD) electricity demand. One of the foci of their analysis is the question of exogeneity of air conditioning appliance ownership variables. They specify a two equation triangular system in which the first equation explains air conditioning appliance ownership and the second explains TOD electricity demand. The appliance ownership variables enter the second equation as explanatory variables and are exogenous in this equation if the first equation error, ε, is uncorrelated with each of two components, k and r, of the second equation error.

Lillard and Aigner carry out likelihood ratio tests of whether the correlation coefficients between these errors are zero and find them to be jointly and individually insignificant at the 5% level in each case. They then conclude that treating appliance ownership variables as exogenous will not lead to specification error and its corresponding bias. Thereafter, they consider only the estimates based on the constrained model, where $\rho_{k\varepsilon} = \rho_{r\varepsilon} = 0$.

The conclusion that Lillard and Aigner reach does not seem warranted without some investigation of the power properties of their tests. With the information provided in their paper, we are able to compute inverse power approximations for the two univariate tests they report, but not for their joint test. The first univariate test is a test of $\rho_{k\varepsilon} = 0$, and the second is of $\rho_{r\varepsilon} = 0$. We consider Lillard and Aigner's "rate B all customers" results. The tests have high power if $|\rho_{k\varepsilon}| \geq .66$ (i.e., $\hat{PS} = \hat{IP}(.95) = .66$) for the first test, and if $|\rho_{r\varepsilon}| \geq .73$ for the second.[4] Further, the tests have low power if $|\rho_{k\varepsilon}| < .47$ (i.e., $\hat{EO} = \hat{IP}(.5) = .47$) for the first test, and if $|\rho_{r\varepsilon}| < .55$ for the second.

These results show that the two univariate tests are not very powerful. Correlations of magnitude .5 or less have a good chance of going undetected, yet they may introduce biases of some consequence.[5] One needs to investigate the relationship between the size of the correlations and the bias on parameter estimates of interest in the demand equation, in order to give an accurate assessment of the decision to constraint the correlations to be zero (e.g, see Nakamura and Nakamura (1985)).

5. ASYMPTOTIC JUSTIFICATIONS OF APPROXIMATE INVERSE POWER FUNCTIONS

Here we provide conditions under which the asymptotic justifications hold for the approximate inverse power function. See the Appendix for its justification when the nuisance parameter is estimated, and for conditions under which justifications hold for the estimated inverse power function.

Let b be the chosen value of the nuisance parameter vector $\nu_2(\theta)$. Consider the following assumptions:

ASSUMPTION C1: *There exists a parameter space* $T \subset R^{\ell}$ *and a one-to-one transformation* $\nu : \Theta \rightarrow T$ *such that (i)* $h(\theta)$ *equals the first* q *elements of* $\tau = \nu(\theta)$, *(ii) given any* $(\tau_1, \tau_2) \in T$, $(\underset{\sim}{0}, \tau_2)$ *also is in* T, *where* $\tau_1, \underset{\sim}{0} \in R^q$, *and (iii)* $\nu(\cdot)$ *is continuously differentiable in a neighborhood of* θ_b *and has a nonsingular Jacobian at* θ_b.

ASSUMPTION C2: *The test statistic* T_n *satisfies*

$$\sup_{\substack{\theta\in\Theta:\theta=\theta_b+\delta/\sqrt{n} \\ \& \ \|\delta\|\leq M}} |P_\theta(T_n \leq s) - P(X_q^2(\mu_n(\theta)) \leq s)| \xrightarrow{n\to\infty} 0 \ , \ \forall s \in R, \ \forall M < \infty \ ,$$

where $X_q^2(\mu_n(\theta)) \sim \chi_q^2(\mu_n(\theta))$ *and* $\mu_n(\theta) = nh(\theta)'Q_{nb}h(\theta)$ *for some positive* *semi-definite non-random* $q \times q$ *matrices* Q_{nb} *and* Q_b, $n = 1, 2, \ldots,$ *that satisfy* $Q_{nb} \xrightarrow{n\to\infty} Q_b$ *and* Q_b *is positive definite.*

ASSUMPTION C3: *For given* b *and* η, *the set* $\{\|h(\theta)\| : \theta \in \Theta,\ h(\theta) \propto \eta,\ \nu_2(\theta) = b\}$ *contains a neighborhood of zero.*

For the general test statistic T_n, the <u>approximate power radius function</u> is defined by

$$(5.1) \qquad R^n_{b\eta}(p) = R_\eta(p) = \frac{1}{\sqrt{n}} \lambda_{q,\alpha}(p)\cdot(\eta' Q_{nb}\eta)^{-1/2} ,$$

and the <u>approximate inverse power function</u> by

$$(5.2) \qquad IP^n_{b\eta}(p) = IP_\eta(p) = R^n_{b\eta}(p)\cdot\eta .$$

Define the sets $\tilde{A}_n$ and $\tilde{B}_n$, for the test based on T_n, as in (3.5).

To justify the use of $R_\eta(p)$ to approximate the outer inverse power function of T_n, we need an additional assumption. Define $\tilde{B}^a_n = \{\theta \in \tilde{B}_n : \|h(\theta)\| \le R^n_{b\eta}(p) + a/\sqrt{n}\}$, where $a > 0$ is a constant.

ASSUMPTION C4: *Given* b, η, *and* p, *for some* $a > 0$ *and some* n_0,

$$\inf_{\theta\in\tilde{B}_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) = \inf_{\theta\in\tilde{B}^a_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) , \text{ for all } n \ge n_0.$$

Of course, assumption C4 is satisfied if the power of the test based on T_n is non-decreasing along the ray defined by η.

We obtain the following asymptotic justifications:

THEOREM 1: *Suppose C1-C3 hold for given* b, η, *and* p. *Then, the approximate inverse power function* $IP_\eta(p)$ *of the test based on* T_n *is such that:*

(a) $\lim_{n\to\infty} \sup_{\theta \in \tilde{A}_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) = p$,

(b) $\lim_{n\to\infty} \inf_{\theta \in \tilde{B}_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) = p$, *provided* T_n *also satisfies C4,*

(c) $\sqrt{n}(IP^n_{b\eta}(p) - iip^n_{b\eta}(p)) \xrightarrow{n\to\infty} 0$, and

(d) $\sqrt{n}(IP^n_{b\eta}(p) - oip^n_{b\eta}(p)) \xrightarrow{n\to\infty} 0$, *provided* T_n *also satisfies C4.*

See the Appendix for the proof of Theorem 1 and other results below.

Next, we show that the Wald statistic W_n satisfies assumption C2 with $Q_{nb} = \Sigma^{-1}_{nb}$. The unrestricted estimator $\hat{\theta}$ is assumed to satisfy:

ASSUMPTION B1: $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} Z_\theta \sim N(\underset{\sim}{0}, V(\theta))$ *as* $n \to \infty$, *under* P_θ, *uniformly for* $\theta \in C$, *any compact subset of* Θ.[6]

ASSUMPTION B2: *Given any compact subset* C *of* Θ, $\bar{V}_n(\theta) \xrightarrow{n\to\infty} V(\theta)$ *and* $V_n(\hat{\theta}) \xrightarrow{p} V(\theta)$ *as* $n \to \infty$ *under* P_θ, *uniformly for* $\theta \in C$, where $\bar{V}_n(\theta)$ is non-random.

ASSUMPTION B3: *(i)* $h(\theta_b) = \underset{\sim}{0}$, $h(\cdot)$ *is continuously differentiable in a neighborhood of* θ_b, *and* H_b *has full rank* q, *(ii)* $V(\theta)$ *is continuous at* θ_b, *and* V_b *has full rank* q, *and (iii)* θ_b *is an interior point of* Θ, *or the set* $C_{b\varepsilon} = \{\theta \in \Theta : \theta = \theta_b + \delta$, for $\|\delta\| \le \varepsilon\}$ *is compact for some* $\varepsilon > 0$.

Estimators that satisfy B1 are called consistent uniformly asymptotically normal (CUAN) estimators by Rao (1973, pp. 350-351), also see Bickel (1981, p. 10). The CUAN condition is not very restrictive; it is satisfied by most asymptotically normal estimators used in econometrics.

We have:

LEMMA 1: *Under assumptions B1-B3, the Wald statistic* W_n *satisfies*

$$\sup_{\substack{\theta\in\Theta:\theta=\theta_b+\delta/\sqrt{n} \\ \& \ \|\delta\|\le M}} |P_\theta(W_n \le s) - P(X_q^2(\mu_n(\theta)) \le s)| \xrightarrow{n\to\infty} 0 \ , \quad \forall s \in R \ , \ \forall M < \infty \ ,$$

where $X_q^2(\mu_n(\theta)) \sim \chi_q^2(\mu_n(\theta))$, $\mu_n(\theta) = nh(\theta)'\Sigma_{nb}^{-1}h(\theta)$, *and* $\Sigma_{nb}^{-1} \xrightarrow{n\to\infty} \Sigma_b^{-1}$.

Theorem 1 and Lemma 1 combine to give the four results of Section 3.1 for the asymptotic justification $IP_\eta(p)$ for the Wald test:

COROLLARY 1: *Suppose the parametric model, the restrictions* $h(\cdot)$, *and the estimator* $\hat{\theta}$ *are such that assumptions B1-B3, C1, and C3 hold. Then, parts (a) and (c) of Theorem 1 hold with* T_n *equal to the Wald statistic* W_n *and* $\mu_n(\theta) = nh(\theta)'\Sigma_{nb}^{-1}h(\theta)$. *In addition, if* W_n *satisfies C4, then parts (b) and (d) of Theorem 1 hold analogously.*

Under well-known general conditions, the likelihood ratio, Lagrange multiplier, and Hausman test statistics are locally equivalent to the Wald statistic based on the unrestricted ML estimator (e.g., see Silvey (1959; 1970, Ch. 7), Wald (1943), Engle (1984), Holly (1982), and Burguete, Gallant, and Souza (1982)). That is, these test statistics satisfy assumption C2 with $\mu_n(\theta) = nh(\theta)\Sigma_{nb}^{-1}h(\theta)$ (where Σ_{nb} is as defined for the Wald statistic). Thus, the following corollary has wide applicability:

COROLLARY 2: *Suppose the model, the restrictions* $h(\cdot)$, *the unrestricted ML estimator* $\hat{\theta}$, *and the likelihood ratio* (LR_n), *Lagrange multiplier* (LM_n) *and/or Hausman* (m_n) *test statistics are such that assumptions B1-B3 and C1-C3 hold, with* $T_n = LR_n$, $T_n = LM_n$, *and/or* $T_n = m_n$ *in C2 and* $\mu_n(\theta) = nh(\theta)'\Sigma_{nb}^{-1}h(\theta)$ *in C2 (where* Σ_{nb} *is as defined above for the Wald statistic based on the unrestricted ML estimator). Then, parts (a) and (c) of Theorem 1 hold. If assumption C4 also holds, then so do parts (b) and (d) of Theorem 1.*

6. CONCLUSION

This paper provides a method for assessing the evidence provided by a test, when the test fails to reject. Inverse power functions are defined that relate rejection probabilities to deviations from the restrictions that define the null hypothesis. The problem of high dimensionality, which arises with the usual power function, is reduced by specifying a value of a nuisance parameter vector, by considering certain interesting directions of departure from the null hypothesis, and by restricting attention to two probabilities, 1/2 and $1 - \alpha$. Approximations are given to the inverse power function that are simple, and are easy to calculate--especially with the estimated inverse power function. These approximations are helpful in answering questions that are important to address in practical contexts.

APPENDIX

A.1. Asymptotic Justifications for the Case of Estimated Nuisance Parameters

Here we consider asymptotic justifications of the approximate inverse power function of the test based on T_n, when the nuisance parameter vector b is estimated. Let b_0 denote the true nuisance parameter vector $\nu_2(\theta_0)$, where θ_0 denotes the true parameter vector, which may or may not satisfy the restrictions. Let B denote the estimator of b_0. We assume:

ASSUMPTION C1′: *C1 holds with* $b = b_0$.

ASSUMPTION C2′: *The test statistic* T_n *satisfies: For some neighborhood* $N(b_0)$ *of* b_0,

$$\sup_{b \in N(b_0)} \sup_{\substack{\theta\in\Theta:\theta=\theta_b+\delta/\sqrt{n} \\ \& \ \|\delta\|\le M}} \left| P_\theta(T_n \le s) - P(X_q^2(\mu_n(\theta)) \le s) \right| \xrightarrow{n\to\infty} 0 , \quad \forall s \in R, \ \forall M < \infty ,$$

where $X_q^2(\mu_n(\theta)) \sim \chi_q^2(\mu_n(\theta))$ *and* $\mu_n(\theta) = nh(\theta)'Q_{nb}h(\theta)$, *for some positive semi-definite non-random* $q \times q$ *matrices* Q_{nb} *and* Q_b, *for* $n = 1, 2, \ldots,$ *and* $b \in N(b_0)$, *that satisfy* $Q_{nb} \xrightarrow{n\to\infty} Q_b$ *uniformly for* $b \in N(b_0)$, *and* Q_b *is continuous and positive definite for* $b \in N(b_0)$.

ASSUMPTION C3′: *C3 holds for all* b *in a neighborhood of* b_0.

ASSUMPTION C4′: *C4 holds for all* b *in a neighborhood of* b_0 *(with the same constants* a *and* n_0 *for each* b *in the neighborhood).*

Under these assumptions, we have:

THEOREM 2: *For given* η *and* p, *suppose C1'-C3' hold and* $B \xrightarrow{p} b_0$ *as* $n \to \infty$ *under the true parameter* θ_0 *(which may or may not satisfy the restrictions). Then, under* θ_0, *the approximate inverse power function* $IP^n_{B\eta}(p)$ *evaluated at the nuisance parameter estimator* B *satisfies:*

(a) $\sup_{\theta \in \tilde{A}_n(B,\eta,p)} P_\theta(T_n > k_{q,\alpha}) \xrightarrow{p} p$ *as* $n \to \infty$,

(b) $\inf_{\theta \in \tilde{B}_n(B,\eta,p)} P_\theta(T_n > k_{q,\alpha}) \xrightarrow{p} p$ as $n \to \infty$, *provided C4' also holds,*

(c) $\sqrt{n}(IP^n_{B\eta}(p) - iip^n_{b_0\eta}(p)) \xrightarrow{p} \underset{\sim}{0}$ as $n \to \infty$, *and*

(d) $\sqrt{n}(IP^n_{B\eta}(p) - oip^n_{b_0\eta}(p)) \xrightarrow{p} \underset{\sim}{0}$ as $n \to \infty$, *provided C4' also holds.*

Note that the results of Theorem 2 hold under a fixed true parameter vector θ_0 that may be in the null or alternative parameter set. These results do no rely on local sequences of alternatives.

Next, we show that the Wald statistic satisfies assumption C2'.

ASSUMPTION B3': *B3 holds for all* b *in some neighborhood of* b_0 .

LEMMA 2: *Under assumptions B1, B2, and B3', the Wald statistic* W_n *satisfies:*

$$\sup_{b \in N(b_0)} \sup_{\substack{\theta \in \Theta : \theta = \theta_b + \delta/\sqrt{n} \\ \& \ \|\delta\| \le M}} |P_\theta(W_n \le s) - P(X^2_q(\mu_n(\theta)) \le s)| \xrightarrow{n\to\infty} 0 ,$$

for all $s \in R$, *and all* $M < \infty$, *where* $N(b_0)$ *is some neighborhood of* b_0, $X^2_q(\mu_n(\theta)) \sim \chi^2_q(\mu_n(\theta))$, *and* $\mu_n(\theta) = nh(\theta)'\Sigma^{-1}_{nb}h(\theta)$. *Further,* $\Sigma^{-1}_{nb} \xrightarrow{n\to\infty} \Sigma^{-1}_b$ *uniformly over* $b \in N(b_0)$, *and* Σ^{-1}_b *is continuous at* b_0.

Theorem 2 and Lemma 2 combine to give the following justifications for the approximate inverse power functions $IP^n_{B\eta}(p)$ of the Wald, LR, LM, and

Hausman tests, when the nuisance parameter vector b is estimated:

COROLLARY 3: *(i) Suppose the parametric model, the restrictions* $h(\cdot)$, *and the estimator* $\hat{\theta}$ *are such that assumptions B1, B2, B3′, C1′, and C3′ hold and* $B \xrightarrow{p} b_0$ *as* $n \to \infty$ *under the true parameter vector* θ_0. *Then, parts (a) and (c) of Theorem 2 hold with* T_n *equal to the Wald test statistic* W_n. *In addition, if* W_n *satisfies C4′, then parts (b) and (d) of Theorem 2 hold. (ii) Suppose the estimator* $\hat{\theta}$ *is the unrestricted ML estimator of* θ_0. *Assume the model, the restrictions* $h(\cdot)$, *the estimator* $\hat{\theta}$, *and the likelihood ratio* (LR_n), *the Lagrange multiplier* (LM_n), *and/or the Hausman* (m_n) *test statistics are such that* $B \xrightarrow{p} b_0$ *as* $n \to \infty$ *under* θ_0 *and assumptions B1, B2, B3′, and C1′-C3′ hold with* $T_n = LR_n$, $T_n = LM_n$, *and/or* $T_n = m_n$ *in C2′ and* $\mu_n(\theta) = nh(\theta)'\Sigma_{nb}^{-1}h(\theta)$ *in C2′. Then, parts (a) and (c) of Theorem 2 hold for these choices of* T_n. *If assumption C4′ also holds, then so do parts (b) and (d) of Theorem 2.*

A.2. Asymptotic Justifications for Estimated Inverse Power Functions

Next, we give asymptotic justifications for the estimated inverse power function $\hat{IP}_\eta^n(p)$ of a test based on a statistic T_n. We assume C1 and C3 hold, but assumption C2 is replaced by a weaker assumption C2" that allows for random weight matrices in the non-centrality parameter $\mu_n(\theta)$:

ASSUMPTION C2": *The test statistic* T_n *satisfies*

$$\sup_{\substack{\theta\in\Theta:\theta=\theta_{b_0}+\delta/\sqrt{n} \\ \& \ \|\delta\|\le M}} \left|P_\theta(T_n \le s) - P(X_q^2(\mu_n(\theta)) \le s)\right| \xrightarrow{p} 0 \quad as \quad n \to \infty ,$$

under any sequence of local alternatives (θ_n) *to* θ_{b_0}, $\forall s \in R$ *and* $\forall M < \infty$, *where (i) conditional on the value of the random non-centrality parameter* $\mu_n(\theta)$, $X_q^2(\mu_n(\theta)) \sim \chi_q^2(\mu_n(\theta))$, *and (ii)* $\mu_n(\theta) = nh(\theta)'Q_nh(\theta)$ *for some positive* *semi-definite random* $q \times q$ *matrices* Q_n, $n = 1, 2, \ldots$, *and some positive definite non-random matrix* Q_{b_0}, *that satisfy* $Q_n \xrightarrow{p} Q_{b_0}$ *as* $n \to \infty$ *under* $\{\theta_n\}$.

For the general test statistic T_n, the <u>estimated power radius function</u> $\hat{R}^n_\eta(p)$ and the <u>estimated inverse power function</u> $\hat{IP}^n_\eta(p)$ are defined as

$$\hat{R}^n_\eta(p) = \hat{R}_\eta(p) = \frac{1}{\sqrt{n}}\lambda_{q,\alpha}(p)\cdot(\eta' Q_n \eta)^{-1/2}\ , \quad \text{and}$$

(A.1)

$$\hat{IP}^n_\eta(p) = \hat{IP}_\eta(p) = \hat{R}^n_\eta(p)\cdot\eta\ .$$

For T_n, the random sets $\hat{A}_n$ and $\hat{B}_n$ are defined as in (4.2).

We introduce a condition C4" that is weaker than C4. Let $\hat{B}^a_n(\eta,p) = \{\theta\epsilon\hat{B}_n : \|h(\theta)\| \le \hat{R}^n_{b\eta}(p) + a/\sqrt{n}\}$, where $a > 0$ is a constant.

ASSUMPTION C4": Given η *and* p, *for some* $a > 0$,

$$\inf_{\theta\in\hat{B}_n(\eta,p)} P_\theta(T_n > k_{q,\alpha}) = \inf_{\theta\in\hat{B}^a_n(\eta,p)} P_\theta(T_n > k_{q,\alpha})$$

with probability that converges to one as $n \to \infty$ *under any sequence of local alternatives* $\{\theta_n\}$ *to* θ_{b_0}.

With the above assumptions, we give an analogy to Theorem 1:

THEOREM 3: *Given* η *and* p, *suppose C1, C2", and C3 hold with* $b = b_0$. *Then, the estimated inverse power function* $\hat{IP}^n_\eta(p)$ *is such that under any sequence of local alternatives* $\{\theta_n\}$ *to* θ_{b_0}, *we have*

(a) $\operatorname{plim}_{n\to\infty} \sup_{\theta\in\hat{A}_n(\eta,p)} P_\theta(T_n > k_{q,\alpha}) = p$,

(b) $\operatorname{plim}_{n\to\infty} \inf_{\theta\in\hat{B}_n(\eta,p)} P_\theta(T_n > k_{q,\alpha}) = p$, *provided* T_n *also satisfies C4"*,

(c) $\sqrt{n}(\hat{IP}^n_\eta(p) - iip^n_{b\ \eta}(p)) \xrightarrow{p} \underset{\sim}{0}$ as $n \to \infty$, *and*

(d) $\sqrt{n}(\hat{IP}^n_\eta(p) - oip^n_{b_0\eta}(p)) \xrightarrow{p} \underset{\sim}{0}$ as $n \to \infty$, *provided* T_n *also satisfies C4"*.

To make Theorem 3 operational one needs conditions under which C2" is satisfied. Consider the Wald statistic W_n. The asymptotic covariance matrix of $\sqrt{n}(h(\hat{\theta}) - h(\theta))$ is Σ_{b_0}, under a sequence of local alternatives $\{\theta_n\}$ to θ_{b_0}. Let $\hat{\Sigma}_n$ be any estimator of Σ_{b_0} such that:

ASSUMPTION B4: $\hat{\Sigma}_n \xrightarrow{p} \Sigma_{b_0}$ *as* $n \to \infty$, *under any sequence of local alternatives* $\{\theta_n\}$ *to* θ_{b_0}.

It is not difficult to see that under assumptions B1-B3, the estimator $\hat{\Sigma} = \hat{H}\hat{V}\hat{H}'$, which is used to construct the Wald statistic, satisfies condition B4. Hence, one choice for $\hat{\Sigma}_n$ is $\hat{\Sigma}$.

LEMMA 3: *Under assumptions B1-B4, the Wald statistic* W_n *satisfies*

$$\sup_{\substack{\theta\in\Theta:\theta=\theta_{b_0}+\delta/\sqrt{n} \\ \&\|\delta\|\leq M}} |P_\theta(W_n \leq s) - P(X^2_q(\mu_n(\theta)) \leq s)| \xrightarrow{p} 0 \text{ as } n \to \infty,$$

under any sequence of local alternatives $\{\theta_n\}$ *to* θ_{b_0}, $\forall s \in R$ *and* $\forall M < \infty$, *where (i) conditional on the value of the random non-centrality parameter* $\mu_n(\theta)$, $X^2_q(\mu_n(\theta)) \sim \chi^2_q(\mu_n(\theta))$ *and (ii)* $\mu_n(\theta) = nh(\theta)'\hat{\Sigma}^{-1}_n h(\theta)$.

Lemma 3 and Theorem 3 combine to give the asymptotic justifications for $\hat{IP}_\eta(p)$ stated in Section 4 for the Wald statistic. Furthermore, since the Wald, LR, LM, and Hausman test statistics are locally equivalent under well-known general conditions, the same covariance matrix estimators $\hat{\Sigma}_n$ can be used in the estimated inverse power functions of the LR, LM, and Hausman tests, as are used with the Wald test (based on the unrestricted ML estimator). Thus, we have:

COROLLARY 4: *(i) Suppose the parametric model, the restrictions* $h(\cdot)$, *the estimator* $\hat{\theta}$, *and the random matrix* $\hat{\Sigma}_n$ *are such that assumptions B1-B4, C1, and C3 hold. Then, parts (a) and (c) of Theorem 3 hold with* T_n *equal to the Wald statistic* W_n *and* $\mu_n(\theta) = nh(\theta)'\hat{\Sigma}_n^{-1}h(\theta)$. *In addition, if* W_n *satisfies C4", then parts (b) and (d) of Theorem 3 hold analogously. (ii) Suppose the model, the restrictions* $h(\cdot)$, *the unrestricted ML estimator* $\hat{\theta}$, *the random matrix* $\hat{\Sigma}_n$, *and the likelihood ratio* (LR_n), *Lagrange multiplier* (LM_n), *and/or Hausman* (m_n) *test statistics are such that B1-B4, C1, C2", and C3 hold with* $T_n = LR_n$, $T_n = LM_n$, *and/or* $T_n = m_n$ *in C2" and* $\mu_n(\theta) = nh(\theta)'\hat{\Sigma}_n^{-1}h(\theta)$ *in C2". Then, parts (a) and (c) of Theorem 3 hold. If assumption C4" also holds, then so do parts (b) and (d) of Theorem 3.*

A.3. Proofs

PROOF OF THEOREM 1: First we show that for some constant $M < \infty$ that does not depend on n, given any n and any $\theta_n \in \tilde{A}_n$, θ_n can be written as

$$\text{(A.2)} \qquad \theta_n = \theta_b + \delta/\sqrt{n}$$

for some vector $\delta \in R^{\ell}$ with $\|\delta\| \leq M$. Define $\tilde{C}_n = \nu(\tilde{A}_n)$. Let $\tau_n = (\tau_{n1}, b)$ denote an arbitrary element of $\tilde{C}_n$. Then, corresponding to τ_n there exists $\theta_n = \nu^{-1}(\tau_n) \in \tilde{A}_n$, and $h(\theta_n) = (1/\sqrt{n})c_n\eta$, where $0 \leq c_n \leq \lambda_{q,\alpha}(p)\cdot(\eta' Q_{nb}\eta)^{-1/2} \leq c$, for some $c < \infty$ that does not depend on n. Thus, we have

$$\text{(A.3)} \qquad \sqrt{n}\,\tau_{n1} = \sqrt{n}\,h(\theta_n) = c_n\cdot\eta\ , \quad \text{where} \quad 0 \leq c_n \leq c < \infty\ , \quad \forall n\ .$$

Since $\nu(\cdot)$ is continuously differentiable in a neighborhood of θ_b and its Jacobian at θ_b is non-zero, the inverse mapping theorem implies that $\nu^{-1}(\cdot)$ is continuously differentiable in a neighborhood of τ_b. Thus, for $\theta_n \in \tilde{A}_n$, we have by a mean-value expansion of the j^{th} element of $\nu^{-1}(\tau_n)$:

$$\theta_{nj} = \nu^{-1}(\tau_n)_j = \nu^{-1}(\tau_b)_j + \left[\frac{\partial}{\partial\tau}\, \nu^{-1}(\bar{\tau}_j)_j\right]'(\tau_n - \tau_b) ,$$

$$\text{(A.4)} \qquad \sqrt{n}(\theta_{nj} - \theta_{bj}) = \left[\frac{\partial}{\partial\tau_1}\, \nu^{-1}(\bar{\tau}_j)_j\right]'\sqrt{n}\, \tau_{n1} , \quad \text{and}$$

$$\sqrt{n}|\theta_{nj} - \theta_{bj}| \le \left\|\frac{\partial}{\partial\tau_1}\, \nu^{-1}(\bar{\tau}_j)_j\right\|\cdot\sqrt{n}\|\tau_{n1}\| \le M_1 ,$$

for all n large, $\forall j = 1, \ldots, \ell$, and for some constant $M_1 < \infty$, where $\bar{\tau}_j$ is on the line segment joining τ_n and τ_b. Whence θ_n can be written as in (A.2).

The result (A.2) and assumption C2 imply

$$\text{(A.5)} \qquad \sup_{\theta\in\tilde{A}_n} |P_\theta(T_n > k_{q,\alpha}) - P(X_q^2(\mu_n(\theta)) > k_{q,\alpha})| \xrightarrow{n\to\infty} 0 .$$

Let $\varepsilon_n(\theta) = P(X_q^2(\mu_n(\theta)) > k_{q,\alpha}) - P_\theta(T_n > k_{q,\alpha})$. Then, by (A.5) we get

$$|\sup_{\theta\in\tilde{A}_n} P_\theta(T_n > k_{q,\alpha}) - \sup_{\theta\in\tilde{A}_n} P(X_q^2(\mu_n(\theta)) > k_{q,\alpha})|$$

$$\text{(A.6)} \qquad = |\sup_{\theta\in\tilde{A}_n} P_\theta(T_n > k_{q,\alpha}) - \sup_{\theta\in\tilde{A}_n} [P_\theta(T_n > k_{q,\alpha}) + \varepsilon_n(\theta)]|$$

$$\le \sup_{\theta\in\tilde{A}_n} |\varepsilon_n(\theta)| \xrightarrow{n\to\infty} 0 .$$

Now, by definition of $\tilde{A}_n$ and assumption C3, for n sufficiently large, there exists $\theta \in \tilde{A}_n$ for which $h(\theta) = (1/\sqrt{n})\tilde{c}_n\eta$, where $\tilde{c}_n = \lambda_{q,\alpha}(p)\cdot(\eta' Q_{nb}\eta)^{-1/2}$. And for all $\theta \in \tilde{A}_n$, $h(\theta) = (1/\sqrt{n})c_n\eta$, where $c_n \le \tilde{c}_n$, by definition of $\tilde{A}_n$. Hence, for n large, all $\theta \in \tilde{A}_n$ satisfy

$$\text{(A.7)} \qquad \mu_n(\theta) = nh(\theta)' Q_{nb} h(\theta) = c_n^2\eta' Q_{nb}\eta \le \lambda_{q,\alpha}(p)^2 ,$$

and for some $\theta \in \tilde{A}_n$, $\mu_n(\theta) = \lambda_{q,\alpha}(p)^2$. This gives

$$\text{(A.8)} \qquad \sup_{\theta\in\tilde{A}_n} P(X_q^2(\mu_n(\theta)) > k_{q,\alpha}) = P(X_q^2(\lambda_{q,\alpha}(p)^2) > k_{q,\alpha}) = p ,$$

for all n large, since $X_q^2(\mu_n(\theta))$ is stochastically increasing in $\mu_n(\theta)$. (The second equality of (A.8) holds by definition of $\lambda_{q,\alpha}(p)$.) Equations (A.8) and (A.6) combine to give part (a) of the Theorem.

To establish part (b), consider the subset $\tilde{B}_n^a$ of $\tilde{B}_n$ given in assumption C4. The same argument as used to establish (A.2) implies that every θ_n in $\tilde{B}_n^a$ can be written as $\theta_n = \theta_b + \delta/\sqrt{n}$, for some $\delta \in R^\ell$ with $\|\delta\| \le M < \infty$. Thus, by arguments analogous to those given above, equations (A.5), (A.6), (A.7), and (A.8) can be established with (i) "$\tilde{A}_n$" replaced by "$\tilde{B}_n^a$" in (A.5), (ii) "$\sup_{\theta\in\tilde{A}_n}$" replaced by "$\inf_{\theta\in\tilde{B}_n^a}$" where it appears the first four times in (A.6), and by "$\sup_{\theta\in\tilde{B}_n^a}$" where it last appears in (A.6), (iii) the inequality "≤" replaced by "≥" in (A.7), and (iv) "$\sup_{\theta\in\tilde{A}_n}$" replaced by "$\inf_{\theta\in\tilde{B}_n^a}$" in (A.8). The analogues of (A.6) and (A.8), and assumption C4 yield

$$\text{(A.9)} \qquad \lim_{n\to\infty} \inf_{\theta\in\tilde{B}_n} P_\theta(T_n > k_{q,\alpha}) = \lim_{n\to\infty} \inf_{\theta\in\tilde{B}_n^a} P_\theta(T_n > k_{q,\alpha}) = p ,$$

as desired.

To establish part (c), suppose part (c) does not hold. Then, there exists a constant $\varepsilon > 0$ and a subsequence $\{n_m\}$ of $\{n\}$ such that either

(i) $\sqrt{n_m}\, r_{b\eta}^{n_m}(p) < \lambda_{q,\alpha}(p)\cdot(\eta' Q_{n_m b}\eta)^{-1/2} - \varepsilon$, $\forall m = 1, 2, \ldots$, or

(ii) $\sqrt{n_m}\, r_{b\eta}^{n_m}(p) > \lambda_{q,\alpha}(p)\cdot(\eta' Q_{n_m b}\eta)^{-1/2} + \varepsilon$, $\forall m = 1, 2, \ldots$. In case (i), there exists a constant $p_1 < p$ such that $\sqrt{n_m}\, r_{b\eta}^{n_m}(p) < \lambda_{q,\alpha}(p_1)\cdot(\eta' Q_{n_m b}\eta)^{-1/2}$,

$\forall m = 1, 2, \ldots$, using the continuity and monotonicity of $\lambda_{q,\alpha}(p)$ in p. Hence, $A_{n_m}(b,\eta,p) \subset \tilde{A}_{n_m}(b,\eta,p_1)$, and

$$\text{(A.10)} \quad \begin{aligned} p &= \lim_{m\to\infty} \sup_{\theta\in A_{n_m}(b,\eta,p)} P_\theta(T_{n_m} > k_{q,\alpha}) \\ &\le \lim_{m\to\infty} \sup_{\theta\in \tilde{A}_{n_m}(b,\eta,p_1)} P_\theta(T_{n_m} > k_{q,\alpha}) = p_1 < p , \end{aligned}$$

where the second equality follows from part (a). This contradiction implies that case (i) cannot occur. An analogous argument shows that case (ii) also cannot hold, which establishes part (c).

The proof of part (d) is identical to that of part (c), except that "$s_{b\eta}^{n_m}(p)$" replaces "$r_{b\eta}^{n_m}(p)$" and "B" replaces "A" throughout, and reference is made to part (b) rather than to part (a). *Q.E.D.*

PROOF OF LEMMA 1: Let $G_{bM}^n = \{\theta \in \Theta : \theta = \theta_b + \delta/\sqrt{n}$, for some $\delta \in R^\ell$ with $\|\delta\| \le M\}$. A compact set $C_{b\varepsilon}$ (as defined in B3(iii)) exists whether or not θ_b is an interior point of Θ. And for any $M < \infty$, $G_{bM}^n \subset C_{b\varepsilon}$ for n large. Thus, without loss of generality, assume $G_{bM}^n \subset C_{b\varepsilon}$, for all n.

Then, using assumption B1, we have:

$$\text{(A.11)} \quad \sup_{\theta\in G_{bM}^n} |P_\theta(\sqrt{n}(\hat{\theta} - \theta) \le y) - P(Z_\theta \le y)| \xrightarrow{n\to\infty} 0 , \quad \forall y \in R^\ell .$$

By the same Taylor expansion argument as used to establish the δ-method (e.g., see Bishop, Fienberg, and Holland (1975, Ch. 14, pp. 492-497)), (A.11) and assumption B3(i) imply:

$$\text{(A.12)} \quad \sup_{\theta\in G_{bM}^n} |P_\theta(\sqrt{n}(h(\hat{\theta}) - h(\theta)) \le y) - P(H(\theta)Z_\theta \le y)| \xrightarrow{n\to\infty} 0 , \quad \forall y \in R^q .$$

Let $\{\theta_n\}$ denote an arbitrary sequence of parameter vectors such that $\theta_n \in G_{bM}^n$ for all n. Under assumptions B2 and B3(ii), $V_n(\hat{\theta}) \xrightarrow{p} V_b$ and $V_n(\hat{\theta}) - \bar{V}_{nb} \xrightarrow{p} \underset{\sim}{0}$ as $n \to \infty$ uniformly over sequences of distributions indexed by $\{\theta_n\}$ (i.e., $\forall\delta > 0$, $\sup_{\{\theta_n\}} P_{\theta_m}(\|V_m(\hat{\theta}) - V_b\| > \delta) \xrightarrow{m\to\infty} 0$). Also, under assumptions B1 and B3(i), $\hat{H} \xrightarrow{p} H_b$ as $n \to \infty$ uniformly over $\{\theta_n\}$. These results, and the fact that V_b and H_b are full rank q by B3, yield $\hat{\Sigma}^{-1/2} \xrightarrow{p} \Sigma_b^{-1/2}$ as $n \to \infty$ uniformly over $\{\theta_n\}$. In addition, $\sqrt{n}\|h(\hat{\theta}) - h(\theta_n)\| = O_p(1)$ and $\|Z_{\theta_n}\| = O_p(1)$ as $n \to \infty$ uniformly over $\{\theta_n\}$. Hence, using (A.12), we get

$$\text{(A.13)}\quad \sup_{\theta\in G_{bM}^n} |P_\theta(\sqrt{n}\,\hat{\Sigma}^{-1/2}(h(\hat{\theta})-h(\theta))\le y) - P(\Sigma_b^{-1/2}H_bZ_\theta\le y)| \xrightarrow{n\to\infty} 0 \;,\; \forall y \in R^q \;.$$

Using assumption B3(i) and the mean value theorem, we have

$$\text{(A.14)}\quad h(\theta_n) = h(\theta_b+\delta_n/\sqrt{n}) = h(\theta_b) + H(\bar{\theta}_n)\delta_n/\sqrt{n} = H_b\delta_n/\sqrt{n} + o(n^{-1/2}) \;,$$

as $n \to \infty$, where $o(n^{-1/2})$ holds uniformly over $\{\theta_n\}$, $\|\delta_n\| \le M$, and each row of $H(\bar{\theta}_n)$ is evaluated at some point $\bar{\theta}_n$ on the line segment joining θ_b and θ_n. Equation (A.14) and previous results for $\hat{\Sigma}^{-1/2}$ yield: $\sqrt{n}\,\hat{\Sigma}^{-1/2}h(\theta_n) = \sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta_n) + o_p(1)$, where $o_p(1)$ holds uniformly over $\{\theta_n\}$. Thus, from (A.13) we get

$$\text{(A.15)}\quad \sup_{\theta\in G_{bM}^n} |P_\theta(\sqrt{n}\,\hat{\Sigma}^{-1/2}h(\hat{\theta})-\sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta)\le y) - P(Z\le y)| \xrightarrow{n\to\infty} 0 \;,\quad \forall y \in R^q \;,$$

where $Z \sim N(\underset{\sim}{0}, I_q)$.

Since the distribution of Z is absolutely continuous, the convergence in

(A.15) occurs uniformly over $y \in R^q$ (see Billingsley and Topsoe (1967, Theorem 2)). Thus, non-random location shifts that vary with n are justified, and we get:

$$\text{(A.16)}\quad \sup_{\theta\in G^n_{bM}} |P_\theta(\sqrt{n}\,\hat{\Sigma}^{-1/2}h(\hat{\theta})\le y) - P(Z+\sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta)\le y)| \xrightarrow{n\to\infty} 0 \;,\quad \forall y \in R^q \;.$$

Since $W_n = (\sqrt{n}\,\hat{\Sigma}^{-1/2}h(\hat{\theta}))'(\sqrt{n}\,\hat{\Sigma}^{-1/2}h(\hat{\theta}))$, and $(Z + \sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta))'(Z + \sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta)) \sim \chi^2_q(\mu_n(\theta))$, (A.16) yields the desired result.

The convergence of Σ_{nb}^{-1} to Σ_b^{-1} follows immediately from assumptions B2 and B3.

Q.E.D.

PROOF OF THEOREM 2: We claim that for some neighborhood $N(b_0)$ of b_0,

$$\text{(A.17)}\qquad \lim_{n\to\infty} \sup_{b\in N(b_0)} \Big| \sup_{\theta\in\tilde{A}_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) - p\Big| = 0 \;.$$

This result and the assumption $B \xrightarrow{p} b_0$ under θ_0 combine to give part (a) of the Theorem, since $P_{\theta_0}(B \in N(b_0)) \xrightarrow{n\to\infty} 1$.

To prove (A.17), we adjust the proof of Theorem 1. Take $N(b_0)$ in (A.17) to be a bounded subset of the neighborhoods of b_0 given in C2′, C3′, and C4′. Equation (A.3) holds for all $b \in N(b_0)$, where c_n depends on b, but c can be taken independent of b (since $\sup_n \sup_{b\in N(b_0)} \|Q_{nb}\| < \infty$, using C2′). By C1′, (A.4) holds for all $b \in N_1(b_0)$, some neighborhood of b_0, where M_1 does not depend on b. Without loss of generality, assume $N(b_0) \subset N_1(b_0)$. By C2′, (A.5) and (A.6) hold uniformly over $b \in N(b_0)$. Finally, by C3′, (A.7) and (A.8) hold for all $b \in N(b_0)$. Equations (A.8) and (A.6) combine to give (A.17).

Part (b) of Theorem 2 is proved by showing that (A.17) holds with " $\sup_{\theta\in\tilde{A}_n(b,\eta,p)}$ " replaced by " $\inf_{\theta\in\tilde{B}_n(b,\eta,p)}$ ". This is done by adjusting the proof of Theorem 1 part (b). Adjustments analogous to those given above establish the second equality of (A.9) with the convergence as $n \to \infty$ holding uniformly over $b \in N(b_0)$. The desired result then follows using C4'.

The proof of parts (c) and (d) of the Theorem is straightforward. By Theorem 1 parts (c) and (d), it suffices to show:

$$\text{(A.18)} \qquad Y_n \equiv \sqrt{n}(IP^n_{B\eta}(p) - IP^n_{b_0\eta}(p)) \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty , \quad \text{under} \quad P_{\theta_0} .$$

Using C2', we have $Q_{nB} \xrightarrow{p} Q_{b_0}$ as $n \to \infty$ under P_{θ_0}, $Q_{nb_0} \xrightarrow{n\to\infty} Q_{b_0}$, and Q_{b_0} is positive definite. Thus,

$$\text{(A.19)} \quad \|Y_n\| = \sqrt{n}\left\|\frac{1}{\sqrt{n}}\, \lambda_{q,\alpha}(p)\cdot[(\eta' Q_{nB}\eta)^{-1/2} - (\eta' Q_{nb_0}\eta)^{-1/2}]\right\| \xrightarrow{p} 0 \quad \text{as} \quad n\to\infty ,$$

under P_{θ_0}, and the proof is complete. *Q.E.D.*

PROOF OF LEMMA 2: The proof of Lemma 2 proceeds by adjusting that of Lemma 1. Without loss of generality we can assume that the neighborhood $N(b_0)$ in the statement of Lemma 2 is contained in the neighborhood of b_0 given in B3' and is such that $\{\theta_b : b \in N(b_0)\}$ is strictly contained in $C_{b_0\epsilon}$ (as defined in B3). This implies that $G^n_{bM} \subset C_{b_0\epsilon}$ for all $b \in N(b_0)$, for n large. Thus, (A.11) holds uniformly over $b \in N(b_0)$. Using this result, assumption B3'(i), and a δ-method argument, we see that (A.12) holds uniformly over $b \in N(b_0)$.

Let $\{\theta_{nb_n}\}$ denote an arbitrary sequence of parameter vectors such that $b_n \in N(b_0)$ and $\theta_{nb_n} \in G^n_{b_nM}$, for all n. Under assumption B2, $V_n(\hat{\theta}) - \bar{V}_{nb_n} \xrightarrow{p} \underline{0}$ as $n \to \infty$ uniformly over sequences of distributions indexed by $\{\theta_{nb_n}\}$, and under B1 and B3'(i), $\hat{H} - H_{b_n} \xrightarrow{p} 0$ as $n \to \infty$ uniformly over $\{\theta_{nb_n}\}$. These results, and the assumption that V_b has determinant bounded away from zero for $b \in N(b_0)$ (for $N(b_0)$ chosen appropriately, using B3') yield $\hat{\Sigma}^{-1/2} - \Sigma_{b_n}^{-1/2} \xrightarrow{p} \underset{\sim}{0}$ as $n \to \infty$ uniformly over $\{\theta_{nb_n}\}$. Hence, we see that (A.13) holds uniformly over $b \in N(b_0)$. By similar means (A.15) and (A.16) can be shown to hold uniformly over $b \in N(b_0)$, which yields the main result of Lemma 2.

The uniform convergence of Σ_{nb}^{-1} to Σ_b^{-1} over $b \in N(b_0)$ as $n \to \infty$, and the continuity of Σ_b^{-1} at b_0 follow from B2 and B3' straightforwardly. *Q.E.D.*

PROOF OF THEOREM 3: To prove parts (a) and (b), consider a sequence of independent random matrices $\tilde{Q}_n$, n = 1, 2, ..., defined on a probability space $(\Omega, \mathcal{B}, P)$ such that $\mathcal{L}_P(\tilde{Q}_n) = \mathcal{L}_{P_{\theta_n}}(Q_n)$, for all n (where $\mathcal{L}_P(\tilde{Q}_n)$ denotes the distribution or law of $\tilde{Q}_n$ under P). (These matrices are introduced so that we only need to work with a single probability measure P, rather than the sequence $\{P_{\theta_n} : n = 1, 2, \ldots\}$.) By C2", $\tilde{Q}_n \xrightarrow{p} Q_{b_0}$ as $n \to \infty$ under P. Hence, every subsequence of $\{n\}$ contains a sub-subsequence $\{n_m\}$ such that $Q_{n_m} \xrightarrow{m\to\infty} Q_{b_0}$ a.s. [P] (e.g., see Lukacs (1968, Theorem 2.4.4, p. 46)). Let ω denote a realization in Ω such that $\tilde{Q}_{n_m\omega} \xrightarrow{m\to\infty} Q_{b_0}$. For this realization, assumption C2 is satisfied, and Theorem 1 implies

$$\sup_{\theta\in\hat{A}_{n_m}(\eta,p)_\omega} P_\theta(T_{n_m} > k_{q,\alpha}) \xrightarrow{m\to\infty} p\ , \quad \text{and}$$

(A.20)

$$\inf_{\theta\in\hat{B}^a_{n_m}(\eta,p)_\omega} P_\theta(T_{n_m} > k_{q,\alpha}) \xrightarrow{m\to\infty} p\ ,$$

where $\hat{A}_{n_m}(\eta,p)_\omega$ denotes the realized value of the random set $\hat{A}_{n_m}(\eta,p)$, defined using $\tilde{Q}_n$ in place of Q_n, when ω occurs. Thus, given any subsequence of $\{n\}$, there exists a sub-subsequence $\{n_m\}$ such that (A.20) holds for all ω in a set with P-probability one. This implies

(A.21)
$$\sup_{\theta\in\hat{A}_n(\eta,p)} P_\theta(T_n > k_{q,\alpha}) \xrightarrow{P} p \ , \quad \text{and}$$
$$\inf_{\theta\in\hat{B}^a_n(\eta,p)} P_\theta(T_n > k_{q,\alpha}) \xrightarrow{P} p \ , \quad \text{as} \quad n \to \infty \ ,$$

under P, where $\hat{A}_n$ and $\hat{B}^a_n$ are defined using $\tilde{Q}_n$ in place of Q_n (e.g., see Lukacs (1968, Theorem 2.4.4, p. 46)). Since $L_P(\tilde{Q}_n) = L_{P_{\theta_n}}(Q_n)$, (A.21) also holds under $\{\theta_n\}$ with $\hat{A}_n$ and $\hat{B}^a_n$ defined using Q_n. This gives part (a), and in view of C4", part (b) as well.

Parts (c) and (d) are proved analogously. *Q.E.D.*

PROOF OF LEMMA 3: Consider a sequence of random matrices $\tilde{\Sigma}_n$, $n = 1, 2, \ldots$, defined on a probability space (Ω, B, P) such that $L_P(\tilde{\Sigma}_n) = L_{P_{\theta_n}}(\hat{\Sigma}_n)$, for all n. By B4, $\tilde{\Sigma}_n \xrightarrow{P} \Sigma_{b_0}$ as $n \to \infty$ under P. Hence, every subsequence of $\{n\}$ has a sub-subsequence such that $\tilde{\Sigma}_{n_m} \xrightarrow{m\to\infty} \Sigma_{b_0}$ a.s. [P]. Let ω be a realization in Ω such that $\tilde{\Sigma}_{n_m\omega} \xrightarrow{m\to\infty} \Sigma_{b_0}$. For this realization, the proof of Lemma 1, with $\Sigma_{n_m b_0}$ replaced by $\tilde{\Sigma}_{n_m\omega}$, implies,

$$\text{(A.22)} \qquad \sup_{\substack{\theta\in\Theta:\theta=\theta_{b_0}+\delta/\sqrt{n} \\ \&\|\delta\|\le M}} \left|P_\theta(W_{n_m} \le s) - P(X^2_q(\mu_{n_m}(\theta)_\omega) \le s)\right| \xrightarrow{m\to\infty} 0 \ ,$$

for all $s \in R$ and all $M < \infty$, where $\mu_{n_m}(\theta)_\omega = n_m h(\theta)'\Sigma_{n_m\omega} h(\overline{\theta})^{-1}$. Hence, given any subsequence of $\{n\}$ there exists a sub-subsequence $\{n_m\}$ such that (A.22) holds for all ω in a set with P-probability one. This yields

$$\text{(A.23)} \qquad \sup_{\substack{\theta\in\Theta:\theta=\theta_{b_0}+\delta/\sqrt{n} \\ \&\|\delta\|\leq M}} \left| P_\theta(W_n \leq s) - P(X_q^2(\mu_n(\theta)) \leq s) \right| \xrightarrow{P} 0 \ , \quad \text{as} \quad n \to \infty \ ,$$

under P, for all $s \in R$ and all $M < \infty$, where $\mu_n(\theta) = nh(\theta)'\tilde{\Sigma}_n^{-1}h(\theta)$. Since $L_P(\tilde{\Sigma}_n) = L_{P_{\theta_n}}(\hat{\Sigma}_n)$, (A.23) holds under $\{\theta_n\}$ with $\mu_n(\theta) = nh(\theta)'\hat{\Sigma}_n^{-1}h(\theta)$. *Q.E.D.*

FOOTNOTES

[1] I would like to thank Chris Cavanagh, Rich Clarida, Ian Domowitz, Ray Fair, Lars Hansen, Barry Ickes, Grayham Mizon, Adrian Pagan, Gene Savin, Bob Shiller, and the participants of the econometrics workshops at Northwestern University, the University of Iowa, and the Federal Reserve Bank Board of Governors for their helpful comments, and Chris Monahan for research assistance. The financial support of the Sloan Foundation, through a Research Fellowship, and the National Science Foundation, through Grant No. SES-8419789, is acknowledged gratefully.

[2] In contrast, or analogy, the exact power contours of the F-test of linear restrictions in a linear regression model are cylinders in R^{ℓ} whose bases are elliptical cones of one nappe (since $\sigma > 0$) in the (q+1)-dimensional space corresponding to the linear restrictions and σ , see Scheffe (1959, pp. 47-48) and Savin (1984).

[3] To facilitate comparison with Barro's results, we ignore Pagan's (1984a) demonstration that the standard errors in Barro's estimated models need to be adjusted to account for the use of estimated, rather than observed, values of the unanticipated money growth variables. That is, we proceed as though these variables actually were observed. If desired, one can recompute straightforwardly inverse power approximations that take account of the estimation of the unanticipated money growth variables.

[4] The inverse power approximations given here are calculated in terms of the parameterization used by Lillard and Aigner, viz., $\alpha_{k\varepsilon} = \tan(\rho_{k\varepsilon}\pi/2)$ and $\alpha_{r\varepsilon} = \tan(\rho_{r\varepsilon}\pi/2)$, and then translated into values in terms of $\rho_{k\varepsilon}$ and $\rho_{r\varepsilon}$. The estimated standard errors of $\hat{\alpha}_{k\varepsilon}$ and $\hat{\alpha}_{r\varepsilon}$ are .4635 and .5966, respectively. For example, for $\alpha_{k\varepsilon}$ we have: $IP(.95) = \lambda_{1,.05}(.95)\cdot.4635 = 1.67$, which corresponds to a $\rho_{k\varepsilon}$ values of $\frac{2}{\pi}\tan^{-1}(1.67) = .66$.

[5] We note that the joint test of $\rho_{k\varepsilon} = \rho_{r\varepsilon} = 0$ may turn out to be more powerful than either of the two univariate tests. This does not affect the main point of this example, however, which is to illustrate that inverse power approximations can be useful in interpreting and justifying model selection tests.

[6] By definition, $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} Z_\theta$ uniformly for $\theta \in C$, if

$$\sup_{\theta \in C}|P_\theta(\sqrt{n}(\hat{\theta} - \theta) \leq x) - P(Z_\theta \leq x)| \xrightarrow{n\to\infty} 0 , \quad \text{for all} \quad x \in R^{\ell} .$$

[7] This table is derived from Table II of Haynam <u>et al</u>. (1970).

[8] This table is derived from Table II of Haynam <u>et al</u>. (1962) and Yamauti (1972, p. 347).

REFERENCES

Andrews, D. W. K. (1985a): "Asymptotic Results for Generalized Wald Tests," Cowles Foundation Discussion Paper #761R, Yale University, New Haven, CT.

Barro, R. J. (1978): "Unanticipated Money, Output, and the Price Level in the United States," *Journal of Political Economy*, 86, 549-580.

Bickel, P. J. (1981): "Quelques Aspects de la Statistique Robuste," in *Ecole d'Ete de Probabilites de Saint-Flour IX-1979*, Springer-Verlag Lecture Notes in Mathematics No. 876, ed. by P. L. Hennequin. New York: Springer-Verlag.

Billingsley, P. and F. Topsoe (1967): "Uniformity in Weak Convergence," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 7, 1-16.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland (1975): *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.

Burguete, J. F., A. R. Gallant, and G. Souza (1982): "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews*, 1, 151-190.

Engle, R. F. (1984): "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics Volume II*, ed. by Z. Griliches and M. D. Intriligator. New York: North-Holland.

Hall, R. E. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971-987.

Haynam, G. E., Z. Govindarajulu, and F. C. Leone (1962): *Tables of the Cumulative Non-central Chi-square Distribution*. Case Statistical Laboratory, Publication No. 104. Reprinted in *Selected Tables in Mathematical Statistics*, ed. by H. L. Harter and D. B. Owens. Chicago: Markham, 1970.

Holly, A. (1982): "A Remark on Hausman's Specification Test," *Econometrica*, 50, 749-759.

Lillard, L. A. and D. J. Aigner (1984): "Time-of-Day Electricity Consumption Response to Temperature and the Ownership of Air Conditioning Appliances," *Journal of Business and Economic Statistics*, 2, 40-53.

Lukacs, E. (1968): *Stochastic Convergence*. Lexington, MA: Heath.

McAleer, M., A. R. Pagan, and P. Volker (1985): "What Will Take the Con Out of Econometrics," *American Economic Review*, 75, 293-306.

McCloskey, D. N. (1985a): "The Loss Function Has Been Mislead: The Rhetoric of Significance Tests," *American Economic Review*, 75, 201-205.

__________ (1985b): *The Rhetoric of Economics*. Madison, WI: University of Wisconsin Press.

Mizon, G. E. and J. F. Richard (1986): "The Encompassing Principle and its Application to Testing Non-nested Hypotheses," *Econometrica*, 54, 657-678.

Nakamura, A. and M. Nakamura (1985): "On the Performance of Tests by Wu and by Hausman for Detecting the Ordinary Least Squares Bias Problem," *Journal of Econometrics*, 29, 213-227.

Pagan, A. R. (1984a): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221-247.

Phillips, P. C. B. (1983): "Exact Small Sample Theory," in *Handbook of Econometrics Volume I*, ed. by Z. Griliches and M. D. Intriligator. New York: North-Holland.

Pitman, E. J. G. (1979): *Some Basic Theory for Statistical Inference*. London: Chapman and Hall.

Rao, C. R. (1973): *Linear Statistical Inference and Its Applications*, Second Edition. New York: Wiley.

Rothenberg, T. J. (1984a): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics Volume II*, ed. by Z. Griliches and M. D. Intriligator. New York: North-Holland.

Sargan, J. D. (1980): "Some Approximations to the Distribution of Econometric Criteria which are Asymptotically Distributed as Chi-Squared," *Econometrica*, 48, 1107-1138.

Savin, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics Volume II*, ed. by Z. Griliches and M. D. Intrilligator. New York: North-Holland.

Scheffe, H. (1959): *The Analysis of Variance*. New York; Wiley, 1959.

Silvey, S. D. (1959): "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics*, 30, 389-407.

__________ (1970): *Statistical Inference*. London: Penguin Books.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426-482.

Yamouti, Z., ed. (1972): *Statistical Tables and Formulas with Computer Applications JSA-1972*. Tokyo: Japanese Standards Association.

Zellner, A. (1980): "Statistical Analysis of Hypotheses in Economics and Statistics," *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 199-203.

TABLE 1

Values of $\lambda_{q,\alpha}(p)$ for $\alpha = .05$ [7]

q \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
1	0.653	1.114	1.435	1.706	1.960	2.213	2.484	2.802	3.241	3.605	4.286
2	0.790	1.316	1.666	1.958	2.226	2.493	2.775	3.104	3.557	3.930	4.626
3	0.883	1.448	1.817	2.122	2.400	2.675	2.965	3.302	3.764	4.144	4.850
4	0.954	1.550	1.933	2.247	2.534	2.815	3.112	3.455	3.924	4.314	5.024
5	1.013	1.633	2.029	2.351	2.644	2.931	3.233	3.582	4.058	4.447	5.170
6	1.063	1.705	2.111	2.441	2.739	3.031	3.338	3.691	4.174	4.567	5.296
7	1.108	1.769	2.184	2.520	2.823	3.120	3.430	3.788	4.276	4.673	5.408
8	1.148	1.826	2.249	2.591	2.899	3.199	3.514	3.876	4.368	4.769	5.510
9	1.185	1.877	2.309	2.656	2.968	3.272	3.591	3.956	4.453	4.857	5.603
10	1.219	1.925	2.364	2.716	3.032	3.340	3.661	4.030	4.531	4.938	5.689
11	1.250	1.970	2.415	2.771	3.091	3.402	3.727	4.099	4.604	5.014	5.769
12	1.279	2.011	2.463	2.823	3.147	3.461	3.789	4.164	4.673	5.085	5.845
13	1.306	2.050	2.507	2.872	3.199	3.516	3.847	4.225	4.737	5.152	5.916
14	1.332	2.087	2.550	2.919	3.249	3.568	3.901	4.282	4.798	5.216	5.984
15	1.356	2.122	2.590	2.963	3.296	3.618	3.954	4.337	4.856	5.276	6.048
16	1.380	2.155	2.628	3.004	3.341	3.665	4.003	4.390	4.912	5.334	6.110
17	1.402	2.187	2.665	3.045	3.383	3.711	4.051	4.440	4.965	5.389	6.169
18	1.423	2.218	2.700	3.083	3.424	3.754	4.097	4.488	5.016	5.442	6.225
19	1.444	2.247	2.734	3.120	3.464	3.796	4.141	4.534	5.065	5.493	6.280
20	1.464	2.275	2.766	3.155	3.502	3.836	4.183	4.578	5.112	5.542	6.332
21	1.482	2.302	2.798	3.190	3.538	3.875	4.224	4.621	5.158	5.590	6.383
22	1.501	2.328	2.828	3.223	3.574	3.912	4.263	4.663	5.202	5.636	6.432
23	1.518	2.354	2.857	3.255	3.608	3.948	4.301	4.703	5.244	5.681	6.480
24	1.535	2.378	2.885	3.286	3.641	3.983	4.338	4.742	5.286	5.724	6.526
25	1.552	2.402	2.913	3.316	3.673	4.017	4.374	4.780	5.326	5.766	6.571
26	1.568	2.425	2.939	3.345	3.704	4.051	4.409	4.817	5.365	5.807	6.614
27	1.583	2.447	2.965	3.373	3.735	4.083	4.443	4.852	5.403	5.846	6.657
28	1.598	2.469	2.990	3.401	3.765	4.114	4.476	4.887	5.440	5.885	6.698
29	1.613	2.490	3.015	3.428	3.793	4.145	4.508	4.921	5.476	5.923	6.738
30	1.628	2.511	3.039	3.454	3.822	4.175	4.540	4.954	5.512	5.960	6.778
32	1.655	2.551	3.085	3.505	3.876	4.232	4.601	5.019	5.580	6.031	6.854
34	1.682	2.589	3.129	3.553	3.928	4.287	4.659	5.080	5.645	6.099	6.927
36	1.707	2.625	3.172	3.599	3.976	4.340	4.714	5.139	5.708	6.164	6.997
38	1.731	2.661	3.212	3.644	4.025	4.391	4.768	5.195	5.768	6.227	7.065
40	1.754	2.694	3.251	3.687	4.071	4.440	4.819	5.249	5.826	6.288	7.130
42	1.777	2.727	3.289	3.728	4.116	4.487	4.869	5.302	5.882	6.346	7.192
44	1.799	2.758	3.325	3.768	4.159	4.532	4.917	5.353	5.936	6.403	7.253
46	1.820	2.788	3.360	3.807	4.200	4.576	4.964	5.402	5.988	6.458	7.312
48	1.840	2.818	3.394	3.844	4.240	4.619	5.009	5.450	6.039	6.511	7.369
50	1.860	2.846	3.427	3.880	4.279	4.660	5.053	5.496	6.088	6.563	7.424
75	2.066	3.145	3.775	4.264	4.691	5.098	5.515	5.985	6.611	7.110	8.013
100	2.225	3.376	4.046	4.562	5.012	5.439	5.877	6.368	7.021	7.540	8.476

TABLE 2

Values of $\lambda_{q,\alpha}(p)$ for α = .01[8]

q \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
1	1.294	1.734	2.051	2.322	2.576	2.829	3.100	3.417	3.857	4.221	4.902
2	1.516	1.985	2.318	2.600	2.862	3.123	3.401	3.726	4.175	4.544	5.236
3	1.662	2.150	2.494	2.783	3.051	3.318	3.601	3.932	4.387	4.762	5.461
4	1.775	2.278	2.630	2.925	3.199	3.470	3.758	4.093	4.554	4.932	5.639
5	1.868	2.384	2.743	3.044	3.322	3.597	3.888	4.227	4.693	5.076	5.788
6	1.948	2.475	2.841	3.146	3.428	3.706	4.002	4.344	4.815	5.200	5.918
7	2.019	2.556	2.927	3.237	3.522	3.804	4.103	4.448	4.929	5.311	6.034
8	2.083	2.629	3.005	3.319	3.608	3.893	4.194	4.543	5.021	5.412	6.139
9	2.141	2.696	3.077	3.395	3.686	3.974	4.277	4.629	5.111	5.505	6.236
10	2.195	2.757	3.143	3.464	3.758	4.048	4.355	4.709	5.194	5.591	6.326
11	2.245	2.815	3.205	3.529	3.826	4.118	4.427	4.784	5.272	5.671	6.410
12	2.291	2.868	3.263	3.590	3.889	4.184	4.495	4.854	5.346	5.746	6.489
13	2.335	2.919	3.317	3.647	3.949	4.246	4.559	4.920	5.415	5.818	6.564
14	2.377	2.967	3.368	3.701	4.005	4.304	4.620	4.983	5.480	5.885	6.635
15	2.42	3.01	3.417	3.753	4.059	4.360	4.677	5.043	5.543	5.950	6.703
16	2.45	3.06	3.464	3.802	4.110	4.413	4.732	5.100	5.602	6.011	6.768
17	2.49	3.10	3.509	3.849	4.159	4.464	4.785	5.155	5.660	6.070	6.830
18	2.52	3.14	3.551	3.894	4.207	4.513	4.836	5.207	5.714	6.127	6.889
19	2.56	3.17	3.592	3.937	4.252	4.560	4.885	5.258	5.767	6.182	6.947
20	2.59	3.21	3.632	3.979	4.295	4.606	4.932	5.307	5.818	6.234	7.002
21	2.62	3.25	3.670	4.019	4.338	4.649	4.977	5.354	5.867	6.285	7.055
22	2.65	3.28	3.707	4.058	4.378	4.692	5.021	5.399	5.915	6.334	7.107
23	2.68	3.31	3.743	4.096	4.418	4.733	5.063	5.444	5.961	6.382	7.158
24	2.71	3.35	3.778	4.133	4.456	4.772	5.104	5.486	6.006	6.428	7.206
25	2.732	3.377	3.811	4.168	4.493	4.811	5.145	5.528	6.050	6.473	7.254
26	2.758	3.407	3.844	4.203	4.529	4.848	5.183	5.568	6.092	6.517	7.300
27	2.784	3.437	3.876	4.237	4.564	4.885	5.221	5.608	6.133	6.560	7.345
28	2.809	3.466	3.907	4.269	4.598	4.920	5.258	5.646	6.173	6.601	7.389
29	2.833	3.493	3.937	4.301	4.632	4.955	5.294	5.684	6.213	6.642	7.432
30	2.856	3.521	3.966	4.332	4.664	4.989	5.329	5.720	6.251	6.681	7.473
32	2.902	3.573	4.023	4.392	4.727	5.055	5.397	5.791	6.325	6.758	7.554
34	2.945	3.623	4.078	4.450	4.787	5.117	5.462	5.858	6.396	6.831	7.632
36	2.986	3.672	4.130	4.505	4.845	5.177	5.525	5.923	6.464	6.902	7.706
38	3.026	3.718	4.180	4.558	4.901	5.235	5.585	5.986	6.529	6.969	7.778
40	3.064	3.762	4.228	4.609	4.954	5.290	5.642	6.046	6.592	7.035	7.847
42	3.101	3.805	4.274	4.658	5.005	5.344	5.698	6.104	6.653	7.098	7.913
44	3.137	3.847	4.319	4.706	5.055	5.396	5.752	6.160	6.712	7.159	7.978
46	3.171	3.887	4.363	4.752	5.103	5.446	5.804	6.214	6.769	7.218	8.041
48	3.205	3.925	4.405	4.796	5.150	5.495	5.855	6.267	6.825	7.275	8.101
50	3.237	3.963	4.446	4.840	5.196	5.542	5.904	6.318	6.878	7.331	8.160
75	3.578	4.361	4.878	5.297	5.676	6.043	6.425	6.861	7.449	7.923	8.788
100	3.842	4.670	5.214	5.655	6.051	6.435	6.833	7.287	7.899	8.389	9.283

TABLE 3

Values of $\lambda_{q,\alpha}(p)$ for $\alpha = .1$ [7]

q \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
1	0.000	0.775	1.112	1.389	1.643	1.898	2.169	2.486	2.926	3.289	3.971
2	0.000	0.922	1.304	1.609	1.886	2.157	2.444	2.777	3.238	3.609	4.308
3	0.000	1.022	1.432	1.754	2.044	2.327	2.624	2.966	3.435	3.817	4.529
4	0.000	1.099	1.530	1.866	2.166	2.457	2.762	3.112	3.589	3.978	4.670
5	0.000	1.163	1.611	1.958	2.266	2.564	2.875	3.232	3.717	4.112	4.842
6	0.000	1.217	1.680	2.037	2.352	2.657	2.974	3.336	3.828	4.227	4.965
7	0.000	1.266	1.742	2.107	2.429	2.738	3.060	3.428	3.926	4.330	5.074
8	0.000	1.309	1.797	2.169	2.497	2.812	3.138	3.511	4.014	4.422	5.173
9	0.000	1.348	1.847	2.226	2.559	2.879	3.210	3.586	4.095	4.506	5.263
10	0.000	1.385	1.893	2.279	2.617	2.941	3.275	3.656	4.170	4.585	5.346
11	0.000	1.418	1.936	2.328	2.671	2.998	3.336	3.721	4.239	4.657	5.424
12	0.000	1.450	1.977	2.374	2.721	3.052	3.394	3.782	4.304	4.725	5.498
13	0.000	1.480	2.014	2.417	2.768	3.102	3.448	3.839	4.366	4.790	5.567
14	0.000	1.507	2.050	2.457	2.812	3.150	3.499	3.893	4.424	4.851	5.632
15	0.000	1.534	2.084	2.496	2.854	3.196	3.547	3.945	4.479	4.909	5.694
16	0.000	1.559	2.116	2.533	2.895	3.239	3.593	3.994	4.532	4.964	5.754
17	0.000	1.583	2.146	2.568	2.933	3.281	3.637	4.041	4.582	5.017	5.811
18	0.000	1.606	2.176	2.601	2.970	3.320	3.680	4.086	4.630	5.067	5.865
19	0.000	1.628	2.204	2.634	3.005	3.358	3.720	4.129	4.677	5.116	5.918
20	0.000	1.649	2.231	2.665	3.040	3.395	3.760	4.171	4.722	5.163	5.968
21	0.000	1.670	2.258	2.695	3.072	3.430	3.797	4.211	4.765	5.209	6.017
22	0.000	1.689	2.283	2.723	3.104	3.465	3.834	4.251	4.807	5.253	6.065
23	0.000	1.709	2.307	2.751	3.135	3.498	3.869	4.288	4.847	5.295	6.110
24	0.000	1.727	2.331	2.778	3.164	3.530	3.903	4.324	4.887	5.336	6.155
25	0.000	1.745	2.354	2.805	3.193	3.561	3.937	4.360	4.925	5.376	6.198
26	0.000	1.762	2.376	2.830	3.221	3.591	3.969	4.394	4.962	5.415	6.240
27	0.000	1.779	2.398	2.855	3.249	3.620	4.000	4.428	4.998	5.453	6.281
28	0.000	1.796	2.419	2.879	3.275	3.649	4.031	4.460	5.033	5.490	6.321
29	0.000	1.812	2.439	2.903	3.301	3.677	4.061	4.492	5.067	5.526	6.360
30	0.000	1.827	2.459	2.926	3.326	3.704	4.090	4.523	5.100	5.561	6.398
32	0.000	1.857	2.498	2.970	3.375	3.757	4.146	4.583	5.165	5.629	6.471
34	0.000	1.886	2.535	3.012	3.421	3.807	4.200	4.640	5.227	5.694	6.541
36	0.000	1.913	2.570	3.052	3.466	3.855	4.251	4.695	5.286	5.756	6.609
38	0.000	1.940	2.604	3.091	3.509	3.901	4.300	4.748	5.343	5.816	6.674
40	0.000	1.965	2.636	3.129	3.550	3.945	4.348	4.799	5.398	5.874	6.736
42	0.000	1.989	2.668	3.164	3.589	3.988	4.394	4.848	5.451	5.930	6.796
44	0.000	2.013	2.698	3.199	3.628	4.030	4.438	4.896	5.502	5.984	6.855
46	0.000	2.036	2.727	3.233	3.665	4.070	4.481	4.941	5.551	6.036	6.911
48	0.000	2.058	2.756	3.265	3.701	4.108	4.523	4.986	5.599	6.086	6.966
50	0.000	2.079	2.783	3.297	3.735	4.146	4.563	5.029	5.646	6.135	7.019
75	0.000	2.303	3.073	3.630	4.103	4.543	4.989	5.485	6.139	6.656	7.584
100	0.000	2.477	3.297	3.888	4.388	4.853	5.322	5.842	6.525	7.064	8.028

α	n	20	35	50	100	300
.05	EO	.44	.33	.28	.20	.11
	PS	.81	.61	.51	.36	.21
.01	EO	.57	.43	.36	.26	.15
	PS	>1.0	.83	.69	.49	.28

TABLE 4

Even-odds and Power/Size Radii, for the Durbin-Watson Test that the First-order Autoregressive Parameter Equals Zero, as a Function of Sample Size n and Significance Level α

$$\left[[(X'X)^{-1}\sigma_B^2]_{4\times4}\right]^{-1} = \begin{bmatrix} 9.93 & 18.08 & 23.94 & 21.27 \\ 18.08 & 43.47 & 60.60 & 60.95 \\ 23.94 & 60.60 & 93.36 & 96.20 \\ 21.27 & 60.95 & 96.20 & 110.31 \end{bmatrix}$$

TABLE 5

The Matrix $(\Sigma_{nB}/n)^{-1}$ for Barro's (1978) Test that Only Unanticipated Money Affects Output

	$IP_\eta(p)$	Power Radius	Sum of Coefficients
EO_{η_1}	(.09, .14, .09, .05)	.20	.37
PS_{η_1}	(.16, .24, .16, .08)	.34	.64
EO_{η_2}	(.17, .13, .084, .042)	.23	.43
PS_{η_2}	(.29, .22, .15, .073)	.40	.73
EO_{η_3}	(.8, 0, 0, 0)	.8	.8
PS_{η_3}	(1.35, 0, 0, 0)	1.35	1.35
EO_{η_4}	(1.03, .29, −.64, .32)	1.29	1.00
PS_{η_4}	(1.75, .50, −1.10, .55)	2.19	1.70
EO_{η^*}	± (1.86, −1.37, −.05, .45)	2.36	.89
PS_{η^*}	± (3.18, −2.33, −.08, .76)	4.02	1.53
EO_{η_*}	± (.03, .07, .10, .11)	.16	.31
PS_{η_*}	± (.04, .11, .17, .18)	.28	.50

TABLE 6

Even-odds and Power/Size Vectors for Barro's (1978) Test
for Six Directions of Interest