Cowles Foundation Paper #737

Econometrica, Vol 57, No. 5 (September, 1989), 1059–1090

POWER IN ECONOMETRIC APPLICATIONS

BY DONALD W. K. ANDREWS[1]

This paper is concerned with the use of power properties of tests in econometric applications 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 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

KEYWORDS: Estimated inverse power function, hypothesis test, Lagrange multiplier test, likelihood ratio test, power function, Wald test.

1 INTRODUCTION

A COMMON PROBLEM faced in applied econometrics is that of interpreting the results of a hypothesis test when the test fails to reject the null hypothesis. Most practitioners realize that just because a test fails to reject a hypothesis one cannot claim to accept it. Nevertheless, it is common for this to be ignored, since the practitioner is often in a position where he would like the outcome of the test to provide useful inferences whether or not the test rejects. The purpose of this paper is to introduce inverse power (IP) summary measures that enable the practitioner to avoid such errors and make valid inferences when a test fails to reject the null hypothesis. These summary measures are widely applicable, easy to use (especially in the common case of a test concerning a single restriction), and simple to compute.

When a test rejects the null hypothesis, the implication is that the data are inconsistent with each parameter point in the null in the sense that the probability of type I error for each point is small, viz., α or less. Correspondingly, when a test *fails* to reject the null hypothesis an analogous statement is needed regarding the error probabilities for points in the alternative hypothesis. It is not the case that all points in the alternative are inconsistent with the data in the sense that their probability of type II error is small (α or less). It is possible, however, to determine the region S in the alternative parameter space that is inconsistent with the data in this sense. The IP function introduced below evaluated at $p = 1 - \alpha$ defines this region. For example, in a test of $H_0: \theta = 0$ versus $H_1: \theta \neq 0$, this region often is of the form $\{\theta: |\theta| > c\}$ for some $c > 0$.

[1] I would like to thank two referees, a co-editor, and Quong Vuong for comments that lead to substantial improvements in the presentation of the paper I also would like to thank Chris Cavanagh, Rich Clarida, Ian Domowitz, Ray Fair, Barry Ickes, Grayham Mizon, Adrian Pagan, Doug Rivers, Gene Savin, Bob Shiller, and the participants of the econometrics workshops at Northwestern University, the University of Iowa, the Federal Reserve Bank Board of Governors, and the Canadian Econometrics Study Group for their helpful comments, and Chris Monahan and Yoon Whang 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

When a test fails to reject, the evidence obtained against any parameter value in S being true is of comparable strength to the evidence obtained against null parameter values being true when the test does reject. Thus, when a test fails to reject, one can conclude with significance level α that the true parameter value is significantly different from the parameter values in S. In the example, the failure to reject implies that with significance level α, $|\theta|$ is less than c. If c is "close" to zero in a substantive sense, then the test provides evidence that $|\theta|$ is zero or "close" to zero, as desired.

It also is possible to determine a region of high probability of type II error, say probability $\geqslant 1/2$. The inverse power function at $1/2$ defines this region. In the example, this region is often of the form $\{\theta : 0 < |\theta| \leqslant b\}$ for some $b \in (0, c)$. When a test fails to reject the null, no evidence is provided against parameter values in this region, since one has a better chance of failing to reject the null than one has of rejecting it when such parameter values are true. Knowledge of this region of low power elucidates the limits regarding the evidence that can be provided by a test's failure to reject.

In short, the IP summary measures provided in the paper relate directly to the quantities of concern, viz., error probabilities, and they summarize the power properties of a test fairly compactly, especially in the case of a test of a single restriction.

To illustrate the simplicity and tractability of these summary measures, we present the formulae for them for several simple, but common, testing situations. Consider level α, Wald (W), likelihood ratio (LR), or Lagrange multiplier (LM) tests of the hypotheses $H_0: \theta = 0$ versus $H_1: \theta \neq 0$ and $H_0: \theta = 0$ versus $H_1: \theta > 0$, where θ is some element of an unknown parameter vector β. For the two-sided test, asymptotic approximations to the inverse power function at $p = 1/2$ and $p = 1 - \alpha$ are given by $b = \lambda_{1,\alpha}(1/2)\hat{\sigma}_\theta$ and $c = \lambda_{1,\alpha}(1-\alpha)\hat{\sigma}_\theta$, respectively, where $\hat{\sigma}_\theta$ is a consistent standard error estimate for $\hat{\theta}, \hat{\theta}$ is the estimator used to construct the test in the case of a Wald test, $\hat{\theta}$ is the unrestricted maximum likelihood estimator (or any other asymptotically efficient estimator) of θ in the case of the other tests, and $\lambda_{1,\alpha}(p)$ is a constant whose value is given in tables provided below.[2] For example, $\lambda_{1,\alpha}(1/2) = 1.960$ and $\lambda_{1,\alpha}(1-\alpha) = 3.605$ when $\alpha = .05$, and $\lambda_{1,\alpha}(1/2) = 2.576$ and $\lambda_{1,\alpha}(1-\alpha) = 4.902$ when $\alpha = .01$.[3] With these formulae, it is trivial to determine the regions of low and high power discussed above.

For one-sided tests corresponding approximations are $b = (z_\alpha - z_{1/2})\hat{\sigma}_\theta = z_\alpha\hat{\sigma}_\theta$ and $c = (z_\alpha - z_{1-\alpha})\hat{\sigma}_\theta = 2z_\alpha\hat{\sigma}_\theta$, where z_α is the $(1-\alpha)$th quantile of the standard normal distribution. For $\alpha = .05$, $b = 1.645\hat{\sigma}_\theta$ and $c = 3.290\hat{\sigma}_\theta$. For $\alpha = .01$, $b = 2.326\hat{\sigma}_\theta$ and $c = 4.652\hat{\sigma}_\theta$.

[2] The subscript 1 on $\lambda_{1,\alpha}(p)$ refers to the fact that one restriction is being tested in this example Below, the number of restrictions q is allowed to exceed one and constants $\lambda_{q,\alpha}(p)$ are considered for $q = 1, 2, \ldots$

[3] Due to the symmetry of the normal distribution, the value of $\lambda_{1,\alpha}(1/2)$ is very close to, but not exactly equal to, the critical value of a two-sided level α test of $\theta = 0$ in a linear regression model with normal errors and known error variances

As an example, consider Ashenfelter and Johnson's (1972) study of the effect of unions on wages. A Wald test finds the coefficient on their union dummy variable not to be significantly different from zero at the .05 level in their linear wage equation. Nevertheless, the authors are cautious and refrain from accepting the null hypothesis of no effect of unions on wages.

Inverse power summary measures make precise the inferences one can draw from this test. In doing so, they make it clear why caution is needed. The IP function at 1/2 and .95, viz., b and c, is .51 and .94, respectively. These coefficient values correspond to relative wage differentials of 65% and 156%.[4] Thus, the test shows that the wage differential is less than 156% with significance level .05, but provides no evidence that it is less than 65%. Since even 65% is a huge wage differential, it is clear that the test is unable to discriminate between zero wage differentials and those nonzero wage differentials that are of interest from an economic perspective.

For tests of multiple restrictions, say $h(\boldsymbol{\theta})=\boldsymbol{0}$, the summary measures of power introduced in this paper are quite similar in spirit to those given above, but necessarily are more complicated. They give the range of deviations from the restrictions (i.e., deviations of $h(\boldsymbol{\theta})$ from $\boldsymbol{0}$) that correspond to alternative parameter values that have low (1/2) or high $(1-\alpha)$ power.

Approximations to the 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 noncentral chi-square distribution under the alternative hypothesis. Then, there is a unique noncentrality parameter (NCP) that corresponds to the specified probability 1/2 or $1-\alpha$. Since the NCP is a quadratic form in the deviation vector $h(\boldsymbol{\theta})$, 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 deviation 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.

Natural questions that arise are the following: To what extent are IP summary measures needed? Do they have advantages over other available methods for communicating the power properties of tests?

In response to these questions, we note that power calculations currently are not often used in applied econometric research (e.g., see Zellner (1980) and McCloskey (1985a, b)). Many practitioners do not know how to mobilize information on power to help analyze their test results or at least how to do so in a simple fashion. Simple summary measures of power appear to be needed to enable practitioners to mobilize information on power for analyzing their test results. This is the raison d'être for the current paper.

[4] These figures are for the test applied to Ashenfelter and Johnson's equation 23c. See their footnote 2 for the correspondence between equation coefficient values and relative wage differentials

Next consider the question of the advantages of IP summary measures over other available methods. One might ask whether the magnitude of a t statistic, F statistic, or chi-square test statistic can provide the requisite information on power? 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 region of coefficient values that are inconsistent with the data, i.e., the region of high power, is quite different in these two cases even though the t statistics are the same. Analogous arguments hold for other test statistics.

Alternatively, one might ask: What information about power, or equivalently about error probabilities, can be obtained from confidence intervals or standard errors? Do they provide good summary measures of power? By definition of a confidence interval C, the probability of type I error is less than or equal to α if the test corresponding to C rejects. For points in the alternative, however, the probability that C contains the true parameter does not give the probability of type II error or a bound on this probability. At best, it may be possible to *derive* information on type II error probabilities from confidence intervals or standard errors. Such information does not follow directly from their definitions.

Finally, one might ask: Why base summary measures of power on an inverse power function rather than on the power function itself? The answer is as follows. Inverse power summary measures provide ranges of deviations of restrictions from the null hypothesis that correspond to a given level p of power, rather than providing p for a given parameter vector $\boldsymbol{\theta}$. 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 $\boldsymbol{\theta}$. The latter are difficult to agree upon in a given problem because of differing views regarding which alternatives $\boldsymbol{\theta}$ are of interest. Furthermore, the choice of values of $\boldsymbol{\theta}$ is impossible to standardize across different testing problems, since the alternatives of interest necessarily depend on the problem at hand. In contrast, it is possible to standardize the choice of values of p across different testing problems.

The remainder of the paper is organized as follows: Section 2 defines the exact inverse power function. Section 3 introduces the estimated and approximate inverse power functions. Section 4 provides examples. Section 5 gives asymptotic justifications for the estimated and approximate inverse power functions. Section 6 provides a brief conclusion.

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 $\boldsymbol{\theta}$. The distributions of the data under the maintained hypothesis are indexed by $\boldsymbol{\theta}$, which lies in the parameter space $\Theta \subset R^{\ell}$. The restrictions are given by an R^q-valued function $h(\cdot)$. Under the null, $h(\boldsymbol{\theta}) = \boldsymbol{0}$. As is well known, the function $h(\cdot)$ that defines the null hypothesis is not unique. We presume that

$h(\cdot)$ is chosen in the most meaningful way possible for interpreting deviations of the restrictions $h(\boldsymbol{\theta})$ from $\boldsymbol{0}$.

We assume that there exists a parameter space $\Upsilon \subset R^{\ell}$ and a one-to-one transformation $\nu : \Theta \to \Upsilon$ such that (i) $h(\boldsymbol{\theta})$ equals the first q elements of $\tau = \nu(\boldsymbol{\theta}) \in \Upsilon$, i.e., $h(\boldsymbol{\theta}) = h(\nu^{-1}(\tau)) = \tau_1$ for $\tau = (\tau_1, \tau_2) \in \Upsilon$, and (ii) given any $(\tau_1, \tau_2) \in \Upsilon, (\boldsymbol{0}, \tau_2)$ also is in Υ, where $\tau_1, \boldsymbol{0} \in R^q$. We call this the *nuisance parameter condition*. Part (ii) of this condition often is satisfied automatically, since we often have $\Upsilon = \Upsilon_1 \times \Upsilon_2$, where $\Upsilon_1 \subset R^q$ and $\Upsilon_2 \subset R^{\ell - q}$.

The ℓ-vector $\nu(\boldsymbol{\theta}) = \tau$ can be partitioned as

$$\begin{pmatrix} \nu_1(\boldsymbol{\theta}) \\ \nu_2(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix},$$

where $\nu_1(\boldsymbol{\theta}) = \tau_1$ is a q-vector of parameters whose values are restricted under the null hypothesis to be $\boldsymbol{0}$ and $\nu_2(\boldsymbol{\theta}) = \tau_2$ 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 $\nu_2(\boldsymbol{\theta}) = \tau_2 = b$, 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 and inverse power function (defined below) is reduced considerably.

For fixed b, there is only one parameter vector in Υ for which the null hypothesis holds, viz.,

$$\tau_b = \begin{pmatrix} \boldsymbol{0} \\ b \end{pmatrix}.$$

This vector corresponds to a unique vector $\boldsymbol{\theta}_b$ in Θ via $\boldsymbol{\theta}_b = \nu^{-1}(\tau_b)$.

2.2. *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 $h(\boldsymbol{\theta})$ from the zero vector). For this reason, we define the IP function to be a function whose values correspond to deviations of $h(\boldsymbol{\theta})$ from $\boldsymbol{0}$, rather than to parameter values $\boldsymbol{\theta}$.

Consider the deviation space $D \subset R^q$, which consists of values that the vector of restrictions $h(\boldsymbol{\theta})$ assumes for different $\boldsymbol{\theta} \in \Theta$. 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 $\gamma_n(\boldsymbol{\theta})$ for given significance level α. We ask: What is the point on the ray defined by η that is farthest from the origin and is 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 points define the *inner inverse power function* of the test in direction η, which is denoted by

$$\begin{aligned}(2.1)\qquad \pi_i(\eta, p) &= \pi_{ni}(b, \eta, p)\\ &= \sup\left\{\|h(\theta)\| : \theta \in \Theta, h(\theta) \propto \eta, \gamma_n(\theta) \leqslant p, \nu_2(\theta) = b\right\} \cdot \eta,\end{aligned}$$

where " $\propto$ " denotes positively proportional to (i.e., $\eta_1 \propto \eta_2$ iff $\eta_1 = c\eta_2$ for some $c \geqslant 0$) and $\|\cdot\|$ denotes the Euclidean norm.

With knowledge of the inner IP 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 IP 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 $\pi_i(\eta, 1 - p_1)$, 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 $\boldsymbol{0}$ than $\pi_i(\eta, 1 - p_1)$.

The *outer inverse power function* is defined analogously:

$$\begin{aligned}(2.2)\qquad \pi_o(\eta, p) &= \pi_{no}(b, \eta, p)\\ &= \inf\left\{\|h(\theta)\| : \theta \in \Theta, h(\theta) \propto \eta, \gamma_n(\theta) \geqslant p, \nu_2(\theta) = b\right\} \cdot \eta.\end{aligned}$$

The outer IP 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 $\pi_o(\eta, p)$ 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 and continuous along rays from the origin, then $\pi_i(\eta, p) = \pi_o(\eta, p)$. Also, by definition of $\pi_i(\eta, p)$ and $\pi_o(\eta, p)$,

$$(2.3)\qquad \sup_{\theta \in A_n(b, \eta, p)} P_\theta(T_n > k_{q,\alpha}) = p \quad \text{and} \quad \inf_{\theta \in B_n(b, \eta, p)} P_\theta(T_n > k_{q,\alpha}) = p,$$

where

$$\begin{aligned}A_n(b, \eta, p) &= \left\{\theta \in \Theta : h(\theta) \leqslant \pi_{ni}(b, \eta, p), \nu_2(\theta) = b\right\},\\ B_n(b, \eta, p) &= \left\{\theta \in \Theta : h(\theta) \geqslant \pi_{no}(b, \eta, p), \nu_2(\theta) = b\right\},\end{aligned}$$

" $\leqslant$ " is defined for vectors as: $\eta_1 \leqslant \eta_2$ 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 absolute value of the estimated IP 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, it is useful to consider those directions where the test has highest power and lowest power. Second, one can consider 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 $h(\hat{\theta})$.

In addition, the problem itself and previous experience may 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 nonzero and where the coefficients satisfy plausible lag patterns, such as linear decay to zero. Alternatively, if the null hypothesis restricts several autoregressive parameters to be zero, one might look in the direction corresponding to a first-order autoregressive model.

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 that 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.

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.

3 ESTIMATED AND APPROXIMATE INVERSE POWER FUNCTIONS

In this section we introduce two different approximations of the exact IP function.

3.1. *The Estimated Inverse Power Function*

The estimated IP function is defined first for the Wald test. Let $\hat{\theta}$ be an unrestricted estimator of θ that satisfies $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} Z_\theta \sim N(\mathbf{0}, V(\theta))$ as $n \to \infty$ when θ is true, for any $\theta \in \Theta$. Suppose $H(\theta) = \partial h(\theta)/\partial\theta$ is continuous and

$H(\theta)V(\theta)H(\theta)'$ is nonsingular at all points θ in the null hypothesis (see Andrews (1988) for the case of a singular $H(\theta)V(\theta)H(\theta)'$ matrix). 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

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

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

$$(3.2) \qquad P\left(X_q^2 > k_{q,\alpha}\right) = \alpha, \qquad \text{where} \qquad X_q^2 \sim \chi_q^2.$$

For the Wald test, we approximate both the inner and outer IP functions by the *estimated inverse power function* $\Pi(\eta, p)$ defined by

$$(3.3) \qquad \Pi(\eta, p) = \Pi_n(\eta, p) = \frac{1}{\sqrt{n}}\lambda_{q,\alpha}(p)\left(\eta'\hat{\Sigma}^{-1}\eta\right)^{-1/2} \cdot \eta$$
$$\left(= \lambda_{q,\alpha}(p)\left(\eta'(\hat{\Sigma}/n)^{-1}\eta\right)^{-1/2} \cdot \eta\right).$$

Using the estimated IP function, approximate analogues of the summary measures $\pi_i(\eta, 1/2)$ and $\pi_o(\eta, 1-\alpha)$ are $\Pi(\eta, 1/2)$ and $\Pi(\eta, 1-\alpha)$, respectively.

The nonrandom quantity $\lambda_{q,\alpha}(p)$ is defined as follows: Let $\gamma_{q,\alpha}(\delta^2) = P(X^2 > k_{q,\alpha})$, where X^2 has noncentral chi-square distribution with q degrees of freedom and noncentrality parameter δ^2 (denoted $\chi_q^2(\delta^2)$). The function $\gamma_{q,\alpha}(\cdot) \cdot R^+ \to [\alpha, 1)$ is strictly increasing, and hence, has a unique inverse $\gamma_{q,\alpha}^{-1}(\cdot) : [\alpha, 1) \to R^+$. By definition, $\lambda_{q,\alpha}(p) = \sqrt{\gamma_{q,\alpha}^{-1}(p)}$. 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 $\chi_q^2(\delta^2)$ distribution. Note that $\lambda_{q,\alpha}(\alpha) = 0$.

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

The term $\hat{\Sigma}/n$, which appears in the definition of $\Pi(\eta, p)$, can be written as $\hat{H}(\hat{V}/n)\hat{H}'$, where $\hat{V}/n$ 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 $\hat{\Sigma}/n$ is just the submatrix of the estimated covariance matrix of $\hat{\theta}$ that corresponds to the restricted elements. In this case, $\Pi(\eta, p)$ 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 $(\eta'(\hat{\Sigma}/n)^{-1}\eta)^{-1/2}$ 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, $\Pi(p)$ $(= \Pi(1, p))$ is simply $\lambda_{1,\alpha}(p) \cdot \hat{\sigma}_{\theta_1}$ and $\Pi(p)$ can be calculated from information commonly reported in applied papers.

TABLE I

VALUES OF $\lambda_{q,\alpha}(p)$ FOR $\alpha = .05$[a]

q \ p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.095	.099
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	4.166	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

[a] This table is derived from Table II of Haynam *et al* (1970)

Note that one only needs to report a consistent estimator of the covariance matrix of $\hat{\theta}$ in order for a reader to be able to compute an estimated IP function for a Wald test of any nonlinear restriction $h(\theta) = 0$.

The estimated IP function is extremely simple. It is easy to isolate the effects of different values of p, α, η, and n on the estimated inverse power of the test, and hence, to ascertain the general qualitative features of the power of the test.

TABLE II

VALUES OF $\lambda_{q,\alpha}(p)$ FOR $\alpha =$ 01[a]

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

[a] This table is derived from Table II of Haynam *et al* (1962) and Yamauti (1972, p 347)

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

$$\left\{\xi \in R^q : \xi'(\hat{\Sigma}/n)^{-1}\xi = \lambda^2_{q,\alpha}(p)\right\}.^5 \tag{3.4}$$

[5] 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).

TABLE III

VALUES OF $\lambda_{q\ \alpha}(p)$ FOR $\alpha = 1$[a]

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

[a] This table is derived from Table II of Haynam *et al* (1970)

These ellipses are of the same shape as confidence regions for the vector $h(\theta)$, but they are centered at $\boldsymbol{0}$ rather than at $h(\hat{\theta})$.

It is common to use the F distribution to obtain small sample adjustments to approximations based on the χ^2 distribution. Such adjustments can be made with the estimated IP function by replacing the constant $\lambda_{q,\alpha}(p)$ by a constant $\lambda_{q,d,\alpha}(p)$ which is based on the noncentral F distribution with numerator and denominator degrees of freedom given by q and d, respectively. As above, q is

TABLE IV

VALUES OF $\lambda_{q,d,\alpha}(p)$ FOR $\alpha = 05$[a]

$q \backslash d$	$p = 5$: 20	30	60	∞	$p = 95$: 20	30	60	∞
1	2.1	2.0	2 0	2 0	3 8	3 7	3 7	3.6
2	2 4	2 3	2 3	2.2	4 2	4.1	4.0	3 9
3	2 6	2.6	2 5	2.4	4 6	4 4	4.3	4 1
4	2 8	2 7	2 6	2 5	4 9	4 7	4 5	4 3
5	3 0	2 9	2 8	2 6	5.1	4 9	4 7	4 4
6	3 2	3 0	2.9	2 7	5 3	5 1	4 8	4 6
7	3 3	3 2	3.0	2 8	5.5	5 2	5 0	4 7
8	3.5	3.3	3 1	2 9	5 7	5 4	5.1	4 8
9	3.6	3 4	3 2	3.0	5 9	5 6	5 2	4.9
10	3 7	3 5	3 3	3 0	6 1	5.7	5 3	4 9
12	3 9	3 7	3 4	3 1	6.4	6 0	5 5	5 1
15	4.2	3 9	3.6	3 3	6 9	6 3	5 8	5 3
20	4.7	4 3	3 9	3.5	7 6	6 9	6 3	5 5
24	5 0	4.6	4 2	3 6	8 1	7.3	6 5	5 7
30	5 5	5 0	4 5	3 8	8.8	7 9	7 0	6 0
40	6.2	5.6	4 9	4 1	9 8	8.7	7.6	6.3

[a] This table was constructed via Monte Carlo simulation using one half a million repetitions and two IMSL normal random number generators (viz. the inverse cdf method-RNNOR and the acceptance/rejection method-RNNOA) The results were cross-checked using the charts of Pearson and Hartley (1951) and Fox (1956), and Table II of Haynam *et al* (1970)

the number of restrictions being tested. d is the number of degrees of freedom of the estimated model, i.e., the sample size minus the number of estimated parameters. More specifically, $\lambda_{q,d,\alpha}(p)$ is defined exactly as $\lambda_{q,\alpha}(p)$ is defined except that X^2 is assumed to have a noncentral F distribution rather than a noncentral χ^2 distribution and the critical value is based on the F distribution rather than the χ^2 distribution.

Tables IV and V tabulate values of $\lambda_{q,d,\alpha}(p)$ for $q = 1(1)10, 12, 15, 20, 24, 30, 40$; $d = 20, 30, 60, \infty$; $\alpha = 0.5, .01$; and $p = .5, 1 - \alpha$. Note that when $d = \infty$, $\lambda_{q,d,\alpha}(p)$ equals $\lambda_{q,\alpha}(p)$, and hence, Tables IV and V illustrate clearly the magnitude of the adjustment made by using the F distribution rather than the χ^2 distribution. The adjustment is significant if q is large and/or d is small.

As discussed in Section 2, two directions η of interest are the estimated directions of highest and lowest power. To determine these directions one merely has to find the directions that minimize and maximize the length of the estimated IP function. Since the length of $\Pi(\eta, p)$ depends on η only through $(\eta'\hat{\Sigma}^{-1}\eta)^{-1/2}$, its length is minimized for any p by the eigenvector η^* of $\hat{\Sigma}^{-1}$ that corresponds to the largest eigenvalue μ^* of $\hat{\Sigma}^{-1}$ and its length is maximized for any p by the eigenvector η_* of $\hat{\Sigma}^{-1}$ that corresponds to the smallest eigenvalue μ_* of $\hat{\Sigma}^{-1}$. Call η^* the (estimated) *direction of highest power* and η_* the (estimated) *direction of lowest power*. The relative magnitude of the lengths of the IP function in the directions of highest and lowest power tells one the extent to which power varies with the direction chosen.

TABLE V

VALUES OF $\lambda_{q,d,\alpha}(p)$ FOR $\alpha = .01$[a]

$q \backslash d$	$p = .5$ 20	30	60	∞	$p = .99$ 20	30	60	∞
1	2.8	2.7	2.7	2.6	5.4	5.2	5.0	4.9
2	3.2	3.1	3.0	2.9	5.9	5.7	5.4	5.2
3	3.5	3.4	3.2	3.1	6.3	6.0	5.7	5.5
4	3.8	3.6	3.4	3.2	6.7	6.3	6.0	5.6
5	4.0	3.8	3.5	3.3	7.0	6.6	6.2	5.8
6	4.2	3.9	3.7	3.4	7.3	6.8	6.3	5.9
7	4.4	4.1	3.8	3.5	7.6	7.1	6.5	6.0
8	4.5	4.2	3.9	3.6	7.8	7.2	6.7	6.1
9	4.7	4.4	4.0	3.7	8.1	7.4	6.8	6.2
10	4.9	4.5	4.1	3.8	8.3	7.6	7.0	6.3
12	5.2	4.7	4.3	3.9	8.8	8.0	7.2	6.5
15	5.6	5.1	4.6	4.1	9.4	8.5	7.6	6.7
20	6.2	5.6	4.9	4.3	10.4	9.2	8.1	7.0
24	6.6	5.9	5.2	4.5	11.0	9.8	8.5	7.2
30	7.2	6.4	5.6	4.7	12.0	10.5	9.0	7.5
40	8.1	7.1	6.1	5.0	13.4	11.6	9.8	7.8

[a] This table was constructed via Monte Carlo simulation using one half a million repetitions and two IMSL normal random number generators (viz., the inverse cdf method-RNNOR and the acceptance/rejection method-RNNOA). The results were cross-checked using the charts of Pearson and Hartley (1951) and Fox (1956), and Table II of Haynam *et al.* (1970).

The estimated IP function can be given two different asymptotic justifications. The first is related directly to its intended use: For any deviation vector in direction η that is closer to (farther from) the origin than $\Pi(\eta, p)$, the probability that the test rejects is less (greater) than or equal to $\hat{p} = p + o_p(1)$ under sequences of local alternatives. The second asymptotic justification of $\Pi(\eta, p)$ is in terms of its ability to approximate $\pi_t(\eta, p)$ and $\pi_o(\eta, p)$ directly: The difference between $\Pi(\eta, p)$ and $\pi_t(\eta, p)$ is $o_p(1/\sqrt{n})$ for any direction vector η and any $p \in [0,1]$ under sequences of local alternatives. The same result holds with $\pi_t(\eta, p)$ replaced by $\pi_o(\eta, p)$. See Section 5 for a detailed account of these asymptotic justifications (including assumptions under which they hold). We note that these justifications hold whether or not $\lambda_{q,\alpha}(p)$ is replaced by $\lambda_{q,d,\alpha}(p)$.

Thus far we have considered only the Wald test. 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.

For the LR, LM, and Hausman tests, one may wish to use a different covariance matrix than $\hat{\Sigma}$ in the definition of $\Pi(\eta, p)$. This can be done, since the only requirement on $\hat{\Sigma}$ is that it is consistent for the asymptotic covariance matrix of $h(\hat{\theta})$ (suitably normalized) under sequences of local alternatives (where

$\hat{\theta}$ is the unrestricted ML estimator). For example, $\hat{\Sigma}$ could be formed using a restricted estimator of θ rather than $\hat{\theta}$. Or, it could be formed using an unrestricted estimator that differs from $\hat{\theta}$.

Note that the approach of this paper can be used to develop summary measures of power for encompassing tests (e.g., see Mizon and Richard (1986)) and other nonclassical tests, in addition to the classical tests discussed above.

Clearly, the estimated IP function defined above is closely related to the local power function (e.g., see Pitman (1979), Engle (1984), and Rothenberg (1984a)), since both rely on a noncentral chi-square approximation of the test statistic. For sample sizes often encountered in practice, the estimated IP function, like the local power function and the nominal significance level, may be rather crude. (This may be true even if $\lambda_{q,d,\alpha}(p)$ is used in place of $\lambda_{q,\alpha}(p)$.) Nevertheless, the approximations provided by the estimated IP function yield 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 the estimated IP function, one can refine the latter 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)).

3.2. *The Approximate Inverse Power Function*

In this section we consider an alternative approximation of the exact IP function, called the approximate IP function. It has the advantage over the estimated IP function that its asymptotic justifications are stronger; they hold for any fixed parameter value θ, as well as for sequences of local alternatives. On the other hand, the approximate IP function may be more difficult to compute than the estimated IP function and may not be defined in cases where the parameter θ does not specify the distribution of the data completely.

Consider the Wald test defined in Section 3.1. For a fixed value b of the nuisance parameter vector $\nu_2(\theta)$, we approximate both the inner and outer IP functions by the *approximate inverse power function* $\pi(\eta, p)$ defined by

$$(3.5) \qquad \pi(\eta, p) = \pi_n(b, \eta, p) = \frac{1}{\sqrt{n}} \lambda_{q,\alpha}(p)\left(\eta' \Sigma_{nb}^{-1} \eta\right)^{-1/2} \cdot \eta .$$

Note that the only difference between the approximate IP function and the estimated IP function is the replacement of $\hat{\Sigma}$ by Σ_{nb}. The latter is 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 *nonrandom* and converges to $V(\theta)$ as $n \to \infty$ $\forall \theta \in \Theta$. 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. Alternatively, if the matrix $V_n(\cdot)$ used in the definition of the Wald statistic is nonrandom, then $\bar{V}_n(\cdot)$ can be taken to equal $V_n(\cdot)$. (Note that $\bar{V}_n(\cdot)$ may depend on exogenous variables if these variables are conditioned on.) The feature that distinguishes Σ_{nb} from $\hat{\Sigma}$ is that, whether or not the null hypothesis is true, Σ_{nb} converges to the asymptotic covariance matrix of $\sqrt{n}(h(\hat{\theta}) - h(\theta))$ evaluated at the null hypothesis point θ_b, viz., $\Sigma_b = H_b V_b H_b'$ where $V_b = V(\theta_b)$. In contrast, $\hat{\Sigma}$ generally only converges to Σ_b when the null hypothesis is true or under sequences of local alternatives.

As in Section 3.1, the noncentral χ^2 approximation used in defining $\pi(\eta, p)$ can be replaced by a noncentral F approximation by replacing $\lambda_{q,\alpha}(p)$ in (3.5) by $\lambda_{q,d,\alpha}(p)$. This substitution is of consequence if q is large and/or d (the degrees of freedom) is small.

Now consider the choice of b. One is free to choose any vector b of interest and compute $\pi_n(b, \eta, p)$. 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.

Next we discuss the calculation of Σ_{nb} in certain special cases. Suppose $h(\theta)$ constrains the first q elements of θ to be $\boldsymbol{0}$. Then, $\Sigma_{nb} = [\bar{V}_{nb}]_{q\times q}$, the upper left $q \times q$ submatrix of $\bar{V}_{nb}$. $[\bar{V}_{nb}]_{q\times q}/n$ is just a particular estimator of the covariance matrix of the first q elements of $\hat{\theta}$.

For example, consider a linear regression model $y = X\beta + u$ with errors u that have mean $\boldsymbol{0}$ 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 $\boldsymbol{0}$. In this case, 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 estimator, 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 estimator of the q elements of β that are equal to zero under the null. In particular, if $q = 1$, the approximate IP 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 linear regression model with intercept. 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, $\pi(p) = \lambda_{1,\alpha}(p)/\sqrt{n}$ for a two-sided test. For the more usual one-sided Durbin-Watson test, the formula given in the Introduction with $\hat{\sigma}_\theta$ set equal to $[\bar{V}_{nb}]_{1\times 1} = 1/\sqrt{n}$ yields $\pi(p) = (z_\alpha - z_p)/\sqrt{n}$ for any b. Hence, for $\alpha = 0.5$, $\pi(1/2) = 1.645/\sqrt{n}$ and $\pi(.95) = 3.290/\sqrt{n}$.

The asymptotic justifications for the approximate IP function are the same as those described above for the estimated IP function except that the results hold for any fixed parameter θ in the null or in the alternative, rather than just under sequences of local alternatives. See Section 5 for details. As it turns out, the nature of the power properties of the tests themselves implies that local alternatives are important asymptotically. Thus, the IP function provides a different perspective on local alternatives—one that makes them appear quite natural.

Under well-known general conditions, the Wald, LR, LM, and Hausman test statistics are asymptotically equivalent under sequences of local alternatives. In consequence, the approximate IP function given above for the Wald test also applies to the LR, LM, and Hausman tests.

4 EXAMPLES

4.1. *Test of a Single Restriction*

The first example illustrates the use of estimated IP summary measures with a test of a single restriction. 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 exogenity 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 estimated IP summary measures 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 $|\rho_{k\varepsilon}| \geqslant .66$ (i.e., $\Pi(.95) = .66$) for the first test and if $|\rho_{r\varepsilon}| \geqslant .73$ for the second.[6] Further, the tests have low power if $|\rho_{k\varepsilon}| < .47$ (i.e., $\Pi(1/2) = .47$) 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

[6] 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: $\Pi(.95) = \lambda_{1,.05}(.95) \cdot .4635 = 1.67$, which corresponds to a $\rho_{k\varepsilon}$ value of $(2/\pi)\tan^{-1}(1.67) = .66$.

they *may* introduce biases of some consequence.[7] 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 constrain the correlations to be zero (e.g., see Nakamura and Nakamura (1985)).

4.2. *Test of Multiple Restrictions*

The second example illustrates the use of approximate IP summary measures with a test of 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 estimator.[8] Thus, $\Sigma_{nB}/n = [(X'X)^{-1}\sigma_B^2]_{4\times 4}$, where the latter is the 4×4 submatrix of the estimated least squares covariance matrix that corresponds to the four money growth variables. The matrix $(\Sigma_{nB}/n)^{-1}$ that arises in $\pi(\eta, p)$ is given in Table VI.

We consider approximate inverse power in the directions of lowest and highest power and in the estimated direction $h(\hat{\theta})$. These directions are given in Table VII by η_*, η^*, and η_1, respectively. In addition, we consider inverse power in several directions that represent *a priori* plausible lag structures: $\eta_2 = (2,3,2,1)/\sqrt{18}$, $\eta_3 = (4,3,2,1)/\sqrt{30}$, and $\eta_4 = (1,0,0,0)$. For each of these directions, Table VII gives the approximate IP function at $1/2$ and $1-\alpha$, as well as the length of the IP function at these points and sum of their coefficients. Using the matrix $(\Sigma_{nB}/n)^{-1}$ reported in Table VI and the tabulated values of $\lambda_{4,.05}(p)$ in Table I, 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

[7] 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

[8] To facilitate comparison with Barro's results, we ignore Pagan's (1984) observation 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

TABLE VI

THE MATRIX $(\Sigma_{nB}/n)^{-1}$ FOR BARRO'S (1978) TEST THAT ONLY UNANTICIPATED MONEY AFFECTS OUTPUT

$$\left([(X'X)^{-1}\sigma_B^2]_{4\times 4}\right)^{-1} = \begin{pmatrix} 9.93 & 18.08 & 23\,94 & 21.27 \\ 18\,08 & 43\,47 & 60\,60 & 60.95 \\ 23\,94 & 60\,60 & 93.36 & 96.20 \\ 21.27 & 60.95 & 96.20 & 110\,31 \end{pmatrix}$$

TABLE VII

APPROXIMATE INVERSE POWER FOR BARRO'S (1978) TEST FOR SIX DIRECTIONS OF INTEREST

	$\pi(\eta, p)$	$\Vert\pi(\eta, p)\Vert$	Sum of Coefficients
$\pi(\eta_*, 1/2)$	$\pm$(1 86, −1.37, −.05,.45)	2 36	89
$\pi(\eta_*, 95)$	$\pm$(3.18, −2.33, −.08,.76)	4.02	1.53
$\pi(\eta^*, 1/2)$	$\pm$(03,.07,.10, 11)	16	31
$\pi(\eta^*, 95)$	$\pm$(.04, 11,.17, 18)	28	.50
$\pi(\eta_1, 1/2)$	(1.03,.29, −.64,.32)	1.29	1.00
$\pi(\eta_1, .95)$	(1.75, 50, −1.10, 55)	2.19	1 70
$\pi(\eta_2, 1/2)$	(.09,.14,.09,.05)	.20	.37
$\pi(\eta_2, .95)$	(.16, 24,.16,.08)	34	64
$\pi(\eta_3, 1/2)$	(.17,.13,.084,.042)	.23	.43
$\pi(\eta_3, 95)$	(29, 22,.15, 073)	.40	.73
$\pi(\eta_4, 1/2)$	(.8,0,0,0)	.8	8
$\pi(\eta_4, 95)$	(1 35,0,0,0)	1 35	1 35

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.

The considerable difference between the lengths of the IP functions in the directions of lowest and highest power indicates that the power of the present test varies substantially with the direction η. In direction η_*, the test has low power in a substantive economic sense. 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 highest power is more plausible. The power in this direction is quite high in a substantive economic sense.

Next we consider the approximate inverse power in direction η_2. From Table VII, we have $\pi(\eta_2, 1/2) = (.09,.14,.09,.05)$ and $\pi(\eta_2,.95) = (.16,.24,.16,.08)$ with their coefficient sums being .37 and .64, respectively. Thus, in direction η_2, 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_1, η_3, and η_4 can be interpreted similarly.

In sum, Barro's test has power that varies considerably with the direction η. It 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.

5 ASYMPTOTIC JUSTIFICATIONS

This section provides a detailed treatment of the asymptotic justifications of the approximate IP function with fixed and estimated nuisance parameters and of the estimated IP function. It is convenient to treat the approximate IP function first, due to the structure of the proofs.

5.1. *The Approximate IP Function with Fixed Nuisance Parameters*

Consider a test statistic T_n. 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 $\Upsilon \subset R^{\ell}$ and a one-to-one transformation $\nu: \Theta \to \Upsilon$ such that (i) $h(\theta)$ equals the first q elements of $\tau = \nu(\theta)$, (ii) given any $(\tau_1, \tau_2) \in \Upsilon, (\boldsymbol{0}, \tau_2)$ also is in Υ, where τ_1, $\boldsymbol{0} \in R^q$, and (iii) $\nu(\cdot)$ is continuously differentiable in a neighborhood of θ_b and has a nonsingular Jacobian at θ_b (where θ_b is defined at the end of Section* 2.1).

ASSUMPTION C2: *The test statistic T_n satisfies*

$$\sup_{\theta \in \Theta \cdot \theta = \theta_b + \delta/\sqrt{n} \,\&\, \|\delta\| \leqslant M} \left| P_\theta(T_n \leqslant s) - P\left(X_q^2(\mu_n(\theta)) \leqslant s\right) \right| \xrightarrow{n \to \infty} 0,$$

$$\forall s \in R, \qquad \forall M < \infty,$$

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 semidefinite nonrandom $q \times q$ 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*

$$\left\{ \|h(\theta)\| : \theta \in \Theta,\ h(\theta) \propto \eta,\ \nu_2(\theta) = b \right\}$$

contains a neighborhood of zero.

For the test statistic T_n, the *approximate inverse power function* is defined by

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

Define the sets $\tilde{A}_n$ and $\tilde{B}_n$ by

$$(5.2)\qquad \begin{aligned} \tilde{A}_n &= \tilde{A}_n(b, \eta, p) = \{\theta \in \Theta : h(\theta) \leqslant \pi_n(b, \eta, p),\ \nu_2(\theta) = b\} \quad \text{and} \\ \tilde{B}_n &= \tilde{B}_n(b, \eta, p) = \{\theta \in \Theta : h(\theta) \geqslant \pi_n(b, \eta, p),\ \nu_2(\theta) = b\} \end{aligned}$$

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

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

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

$$\inf_{\theta \in \tilde{B}_n(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) = \inf_{\theta \in \tilde{B}_n^a(b,\eta,p)} P_\theta(T_n > k_{q,\alpha}) \qquad \forall n \geqslant n_0.$$

Of course, Assumption C4 is satisfied if the power of the test based on T_n is nondecreasing along the ray defined by η. This assumption is not innocuous, however. It is violated in some applications, e.g., see Krämer (1984) and Nelson and Savin (1988).

We obtain the following asymptotic justifications:

THEOREM 1: *Suppose C1–C3 hold for given b, η, and p. Then, the approximate inverse power function $\pi(\eta, p)$ 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) $\sqrt{n}(\pi_n(b, \eta, p) - \pi_{ni}(b, \eta, p)) \xrightarrow{n\to\infty} 0$ *and*

(d) $\sqrt{n}(\pi_n(b, \eta, p) - \pi_{no}(b, \eta, p)) \xrightarrow{n\to\infty} 0$ *provided T_n also satisfies C4.*

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

$$(5.3)\qquad \theta_n = \theta_b + \delta/\sqrt{n}$$

for some vector $\delta \in R^\ell$ with $\|\delta\| \leqslant M$. 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 $0 \leqslant c_n \leqslant \lambda_{q,\alpha}(p)\cdot(\eta' Q_{nb}\eta)^{-1/2} \leqslant c$ for some $c < \infty$ that does not depend on n. Thus, we have

$$(5.4)\qquad \sqrt{n}\,\tau_{n1} = \sqrt{n}\,h(\theta_n) = c_n\eta, \qquad \text{where} \qquad 0 \leqslant c_n \leqslant c < \infty, \qquad \forall n.$$

Since $\nu(\cdot)$ is continuously differentiable in a neighborhood of θ_b and its Jacobian at θ_b is nonzero, 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 jth element of $\nu^{-1}(\tau_n)$:

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

$$\sqrt{n}\,(\theta_{nj} - \theta_{bj}) = \left[\frac{\partial}{\partial \tau_1}\nu^{-1}(\bar{\tau}_j)_j\right]'\sqrt{n}\,\tau_{n1}, \quad \text{and} \tag{5.5}$$

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

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 (5.3).

The result (5.3) and Assumption C2 imply

$$\sup_{\theta \in \tilde{A}_n} \left| P_\theta(T_n > k_{q,\alpha}) - P\left(X_q^2(\mu_n(\theta)) > k_{q,\alpha}\right) \right| \xrightarrow{n \to \infty} 0. \tag{5.6}$$

Let $\varepsilon_n(\theta) = P(X_q^2(\mu_n(\theta)) > k_{q,\alpha}) - P_\theta(T_n > k_{q,\alpha})$. Then, by (5.6) we get

$$\begin{aligned} &\left| \sup_{\theta \in \tilde{A}_n} P_\theta(T_n > k_{q,\alpha}) - \sup_{\theta \in \tilde{A}_n} P\left(X_q^2(\mu_n(\theta)) > k_{q,\alpha}\right) \right| \\ &= \left| \sup_{\theta \in \tilde{A}_n} P_\theta(T_n > k_{q,\alpha}) - \sup_{\theta \in \tilde{A}_n} \left[P_\theta(T_n > k_{q,\alpha}) + \varepsilon_n(\theta)\right] \right| \\ &\leqslant \sup_{\theta \in \tilde{A}_n} |\varepsilon_n(\theta)| \xrightarrow{n \to \infty} 0. \end{aligned} \tag{5.7}$$

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 $\tilde{c}_n \equiv \lambda_{q,\alpha}(p)\cdot(\eta' Q_{nb}\eta)^{-1/2}$. And for all $\theta \in \tilde{A}_n$, $h(\theta) = (1/\sqrt{n})c_n\eta$, where $c_n \leqslant \tilde{c}_n$ by definition of $\tilde{A}_n$. Hence, for n large, all $\theta \in \tilde{A}_n$ satisfy

$$\mu_n(\theta) = nh(\theta)' Q_{nb} h(\theta) = c_n^2 \eta' Q_{nb}\eta \leqslant \lambda_{q,\alpha}^2(p) \tag{5.8}$$

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

$$\sup_{\theta \in \tilde{A}_n} P\left(X_q^2(\mu_n(\theta)) > k_{q,\alpha}\right) = P\left(X_q^2\left(\lambda_{q,\alpha}^2(p)\right) > k_{q,\alpha}\right) = p \tag{5.9}$$

for all n large, since $X_q^2(\mu_n(\theta))$ is stochastically increasing in $\mu_n(\theta)$. (The second equality of (5.9) holds by definition of $\lambda_{q,\alpha}(p)$.) Equations (5.7) and (5.9) 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 (5.3) 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 $\|\delta\| \leqslant M < \infty$. Thus, by arguments analogous to those given above, equations (5.6), (5.7), (5.8), and (5.9) can be established with (i) "$\tilde{A}_n$" replaced by "$\tilde{B}_n^a$" in (5.6), (ii) "$\sup_{\theta \in \tilde{A}_n}$" replaced by "$\inf_{\theta \in \tilde{B}_n^a}$" where it appears the first four times in (5.7) and by "$\sup_{\theta \in \tilde{B}_n^a}$" where it last appears in (5.7), (iii) the inequality "$\leqslant$" replaced by "$\geqslant$" in (5.8), and (iv) "$\sup_{\theta \in \tilde{A}_n}$" replaced by "$\inf_{\theta \in \tilde{B}_n^a}$" in (5.9). The analogues of (5.7) and (5.9) and Assumption C4 yield

$$\text{(5.10)} \qquad \lim_{n \to \infty} \inf_{\theta \in \tilde{B}_n} P_\theta(T_n > k_{q,\alpha}) = \lim_{n \to \infty} \inf_{\theta \in \tilde{B}_n^a} P_\theta(T_n > k_{q,\alpha}) = p,$$

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

$$\text{(i)} \qquad \sqrt{n_m}\,\|\pi_{n_m t}(b, \eta, p)\| < \lambda_{q,\alpha}(p) \cdot (\eta' Q_{n_m b} \eta)^{-1/2} - \varepsilon, \qquad \forall m = 1, 2, \ldots,$$

or

$$\text{(ii)} \qquad \sqrt{n_m}\,\|\pi_{n_m t}(b, \eta, p)\| > \lambda_{q,\alpha}(p) \cdot (\eta' Q_{n_m b} \eta)^{-1/2} + \varepsilon, \qquad \forall m = 1, 2, \ldots.$$

In case (i), there exists a constant $p_1 < p$ such that

$$\sqrt{n_m}\,\|\pi_{n_m t}(b, \eta, p)\| < \lambda_{q,\alpha}(p_1) \cdot (\eta' Q_{n_m b} \eta)^{-1/2}, \qquad \forall m = 1, 2, \ldots,$$

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

$$\text{(5.11)} \qquad p = \lim_{m \to \infty} \sup_{\theta \in A_{n_m}(b, \eta, p)} P_\theta(T_{n_m} > k_{q,\alpha})$$
$$\leqslant \lim_{m \to \infty} \sup_{\theta \in \tilde{A}_{n_m}(b, \eta, p_1)} P_\theta(T_{n_m} > k_{q,\alpha}) = p_1 < p,$$

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 $\pi_{n_m o}(b, \eta, p)$ replaces $\pi_{n_m t}(b, \eta, p)$ and B replaces A throughout and reference is made to part (b) rather than to part (a). *Q.E.D.*

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

ASSUMPTION B1: *$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} Z_\theta \sim N(\mathbf{0}, V(\theta))$ as $n \to \infty$ under P_θ, uniformly for $\theta \in C$, any compact subset of $\Theta \subset R^\ell$.*[9]

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 nonrandom.*

ASSUMPTION B3: (i) *$h(\theta_b) = \mathbf{0}$, $h(\cdot)$ 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*

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

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

q, and (iii) *θ_b is an interior point of Θ or the set $C_{b\varepsilon} \equiv \{\theta \in \Theta : \theta = \theta_b + \delta$, for $\|\delta\| \leqslant \varepsilon\}$ 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 the following Lemma:

LEMMA 1: *Under Assumptions B1–B3, the Wald statistic W_n satisfies*

$$\sup_{\theta \in \Theta : \theta = \theta_b + \delta/\sqrt{n} \,\&\, \|\delta\| \leqslant M} \left| P_\theta(W_n \leqslant s) - P\left(X_q^2(\mu_n(\theta)) \leqslant s\right) \right| \xrightarrow{n\to\infty} 0,$$

$$\forall s \in R, \qquad \forall M < \infty,$$

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 Corollary 1:

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.*

PROOF OF LEMMA 1: Let $G_{bM}^n = \{\theta \in \Theta : \theta = \theta_b + \delta/\sqrt{n}$ for some $\delta \in R^\ell$ with $\|\delta\| \leqslant M\}$. 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:

$$\sup_{\theta \in G_{bM}^n} \left| P_\theta\left(\sqrt{n}(\hat{\theta} - \theta) \leqslant y\right) - P(Z_\theta \leqslant y) \right| \xrightarrow{n\to\infty} 0, \qquad \forall y \in R^\ell. \tag{5.12}$$

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)), (5.12) and Assumption B3(i) imply:

$$\sup_{\theta \in G_{bM}^n} \left| P_\theta\left(\sqrt{n}\left(h(\hat{\theta}) - h(\theta)\right) \leqslant y\right) - P\left(H(\theta) Z_\theta \leqslant y\right) \right| \xrightarrow{n\to\infty} 0, \qquad \forall y \in R^q. \tag{5.13}$$

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 $V_n(\hat{\theta}) - \bar{V}_{nb} \xrightarrow{p} 0$ as $n \to \infty$ uniformly over sequences of distributions indexed by $\{\theta_n\}$ (i.e., $\forall \delta > 0$, $\sup_{\theta_n \in G_{bM}^n} P_{\theta_n}(\|V_n(\hat{\theta}) - V_b\| > \delta) \xrightarrow{n\to\infty} 0$). 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 Assumption 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 (5.13), we get

$$(5.14) \quad \sup_{\theta \in G^n_{bm}} \left| P_\theta\left(\sqrt{n}\hat{\Sigma}^{-1/2}\left(h(\hat{\theta})-h(\theta)\right) \leqslant y\right) - P\left(\Sigma_b^{-1/2}H_b Z_\theta \leqslant y\right)\right| \xrightarrow{n\to\infty} 0, \quad \forall y \in R^q.$$

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

$$(5.15) \quad h(\theta_n) = h\left(\theta_b + \delta_n/\sqrt{n}\right) = h(\theta_b) + H\left(\bar{\theta}_n\right)\delta_n/\sqrt{n} = H_b\delta_n/\sqrt{n} + o\left(n^{-1/2}\right)$$

as $n \to \infty$, where $o(n^{-1/2})$ holds uniformly over $\{\theta_n\}$, $\|\delta_n\| \leqslant M$, 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 (5.15) 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 (5.14) we get

$$(5.16) \quad \sup_{\theta \in G^n_{bM}} \left| P_\theta\left(\sqrt{n}\hat{\Sigma}^{-1/2}h(\hat{\theta}) - \sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta) \leqslant y\right) - P(Z \leqslant y)\right| \xrightarrow{n\to\infty} 0, \quad \forall y \in R^q,$$

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

Since Z has absolutely continuous distribution, the convergence in (5.16) occurs uniformly over $y \in R^q$ (see Billingsley and Topsoe (1967, Theorem 2)). Thus, nonrandom location shifts that vary with n are justified and we get:

$$(5.17) \quad \sup_{\theta \in G^n_{bM}} \left| P_\theta\left(\sqrt{n}\hat{\Sigma}^{-1/2}h(\hat{\theta}) \leqslant y\right) - P\left(Z + \sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta) \leqslant y\right)\right| \xrightarrow{n\to\infty} 0, \quad \forall y \in R^q.$$

Since $W_n = (\sqrt{n}\hat{\Sigma}^{-1/2}h(\hat{\theta}))'(\sqrt{n}\hat{\Sigma}^{-1/2}h(\hat{\theta}))$ and $(Z + \sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta))'(Z + \sqrt{n}\,\Sigma_{nb}^{-1/2}h(\theta)) \sim \chi_q^2(\mu_n(\theta))$, (5.17) yields the desired result.

The convergence of Σ_{nb}^{-1} to Σ_b^{-1} follows from Assumptions B2 and B3. *Q.E.D.*

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), Burguete, Gallant, and Souza (1982), and Andrews and Fair (1988)). 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.

5.2. *The Approximate IP Function with Estimated Nuisance Parameters*

Here we consider asymptotic justifications of the approximate IP 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 make the following assumptions:

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

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

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

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 semidefinite nonrandom $q \times q$ 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 the following Theorem:

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 $\pi_n(b, \eta, p)$ 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) $\sqrt{n}(\pi_n(B, \eta, p) - \pi_{nl}(b_0, \eta, p)) \xrightarrow{p} \mathbf{0}$ *as* $n \to \infty$, *and*
(d) $\sqrt{n}(\pi_n(B, \eta, p) - \pi_{no}(b_0, \eta, p)) \xrightarrow{p} \mathbf{0}$ *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 not rely on *local* sequences of alternatives.

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

$$(5.18)\qquad \lim_{n\to\infty}\ \sup_{b\in N(b_0)}\left|\sup_{\theta\in\tilde{A}_n(b,\eta,p)} P_\theta(T_n>k_{q,\alpha})-p\right|=0.$$

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 (5.18), we adjust the proof of Theorem 1. Take $N(b_0)$ in (5.18) to be a bounded subset of the neighborhoods of b_0 given in C2′, C3′, and C4′. Equation (5.4) 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′, (5.5) 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′, (5.6) and (5.7) hold uniformly over $b\in N(b_0)$. Finally, by C3′, (5.8) and (5.9) hold for all $b\in N(b_0)$. Equations (5.7) and (5.9) combine to give (5.18).

Part (b) of Theorem 2 is proved by showing that (5.18) 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 (5.10) 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:

$$(5.19)\qquad Y_n\equiv\sqrt{n}\,(\pi_n(B,\eta,p)-\pi_n(b_0,\eta,p))\xrightarrow{p}\mathbf{0}\qquad\text{as}\qquad n\to\infty,$$

under P_{θ_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,

$$(5.20)\qquad \|Y_n\|=\sqrt{n}\left\|\frac{1}{\sqrt{n}}\lambda_{q,\alpha}(p)\cdot\left[(\eta'Q_{nB}\eta)^{-1/2}-(\eta'Q_{nb_0}\eta)^{-1/2}\right]\right\|\xrightarrow{p}0\qquad\text{as}\qquad n\to\infty$$

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

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*:

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

for all $s\in R$ and all $M<\infty$, where $N(b_0)$ is some neighborhood of b_0, $X_q^2(\mu_n(\theta))$

$\sim \chi_q^2(\mu_n(\theta))$, *and* $\mu_n(\theta) = nh(\theta)'\Sigma_{nb}^{-1}h(\theta)$. *Further*, $\Sigma_{nb}^{-1} \xrightarrow{n\to\infty} \Sigma_b^{-1}$ *uniformly over* $b \in N(b_0)$ *and* Σ_b^{-1} *is continuous at* b_0.

Theorem 2 and Lemma 2 combine to give the following justifications for the approximate IP function $\pi_n(B, \eta, p)$ of the Wald, LR, LM, and Hausman tests, when the nuisance parameter vector b is estimated by B:

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.*

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 $G_{bM}^n \subset C_{b_0\epsilon}$ for all $b \in N(b_0)$ for n large. Thus, (5.12) holds uniformly over $b \in N(b_0)$. Using this result, Assumption B3′(i), and a δ-method argument, we see that (5.13) 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 $\theta_{nb_n} \in G_{b_nM}^n$ for all n. Under Assumption B2, $V_n(\hat{\theta}) - \overline{V}_{nb_n} \xrightarrow{p} \boldsymbol{0}$ 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 $\hat{\Sigma}^{-1/2} - \Sigma_{b_r}^{-1/2} \xrightarrow{p} \boldsymbol{0}$ as $n \to \infty$ uniformly over $\{\theta_{nb_n}\}$. Hence, we see that (5.14) holds uniformly over $b \in N(b_0)$. By similar means (5.16) and (5.17) 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.*

5.3. *The Estimated Inverse Power Function*

Next, we give asymptotic justifications for the estimated IP function $\Pi(\eta, p)$ of a test based on a statistic T_n. Let b_0 denote the true nuisance parameter vector $\nu_2(\theta_0)$. The estimated IP function approximates the inner and outer IP functions 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 $\theta_{b_0} = \nu^{-1}(0', b_0')$ (see Section 2.1). We consider sequences of local alternatives $\{\theta_n\}$ to θ_{b_0}. In particular, let $\{\theta_n\} = \{\theta_{b_0} + \delta_n/\sqrt{n} : n = 1, 2, \dots\}$ for some $\delta_n \in R^{\ell}$ with $\|\delta_n\| \leqslant M$ for some $M < \infty$ for all 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 noncentrality parameter $\mu_n(\theta)$:

ASSUMPTION C2″: *The test statistic T_n satisfies*

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

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 noncentrality parameter $\mu_n(\theta)$, $X_q^2(\mu_n(\theta)) \sim \chi_q^2(\mu_n(\theta))$ and* (ii) *$\mu_n(\theta) = nh(\theta)'Q_n h(\theta)$ for some positive semidefinite random $q \times q$ matrices Q_n, $n = 1, 2, \dots$, and some positive definite nonrandom matrix Q_{b_0}, that satisfy $Q_n \xrightarrow{p} Q_{b_0}$ as $n \to \infty$ under $\{\theta_n\}$.*

For the test statistic T_n, the *estimated inverse power function* $\Pi(\eta, p)$ is defined by

$$\Pi(\eta, p) = \Pi_n(\eta, p) = \frac{1}{\sqrt{n}} \lambda_{q,\alpha}(p)(\eta' Q_n \eta)^{-1/2} \cdot \eta. \tag{5.21}$$

Define the random sets $\hat{A}_n$ and $\hat{B}_n$ by

$$\begin{aligned} \hat{A}_n &= \hat{A}_n(\eta, p) = \{\theta \in \Theta : h(\theta) \leqslant \Pi_n(\eta, p),\ \nu_2(\theta) = b_0\} \qquad \text{and} \\ \hat{B}_n &= \hat{B}_n(\eta, p) = \{\theta \in \Theta : h(\theta) \geqslant \Pi_n(\eta, p),\ \nu_2(\theta) = b_0\} \end{aligned} \tag{5.22}$$

(where, "$\leqslant$" and "$\geqslant$" are as defined in (5.2)).

We introduce a condition C4″ that is weaker than C4. Let

$$\hat{B}_n^a(\eta, p) = \left\{\theta \in \hat{B}_n : \|h(\theta)\| \leqslant \|\Pi_n(\eta, p)\| + a/\sqrt{n}\right\},$$

where $a > 0$ is a constant.

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

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

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 $\Pi(\eta, p)$ 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) $\sqrt{n}\,(\Pi_n(\eta,p) - \pi_{nl}(b_0,\eta,p)) \xrightarrow{p} 0$ *as* $n\to\infty$, *and*

(d) $\sqrt{n}\,(\Pi_n(\eta,p) - \pi_{no}(b_0,\eta,p)) \xrightarrow{p} 0$ *as* $n\to\infty$

provided T_n *also satisfies C4″*.

Proof of Theorem 3: To prove parts (a) and (b), consider a sequence of independent random matrices $\tilde{Q}_n$, $n = 1,2,\ldots$, defined on a probability space (Ω, B, P) such that $L_P(\tilde{Q}_n) = L_{P_{\theta_n}}(Q_n)$ for all n (where $L_P(\tilde{Q}_n)$ 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

$$
\begin{aligned}
&\sup_{\theta\in\hat{A}_{n_m}(\eta,p)_\omega} P_\theta(T_{n_m} > k_{q,\alpha}) \xrightarrow{m\to\infty} p \qquad \text{and} \\
&\inf_{\theta\in\hat{B}^a_{n_m}(\eta,p)_\omega} P_\theta(T_{n_m} > k_{q,\alpha}) \xrightarrow{m\to\infty} p,
\end{aligned}
\tag{5.23}
$$

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 (5.23) holds for all ω in a set with P-probability one. This implies

$$
\begin{aligned}
&\sup_{\theta\in\hat{A}_n(\eta,p)} P_\theta(T_n > k_{q,\alpha}) \xrightarrow{p} p \qquad \text{and} \\
&\inf_{\theta\in\hat{B}^a_n(\eta,p)} P_\theta(T_n > k_{q,\alpha}) \xrightarrow{p} p \qquad \text{as } n\to\infty
\end{aligned}
\tag{5.24}
$$

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 $L_P(\tilde{Q}_n) = L_{P_{\theta_n}}(Q_n)$, (5.24) 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.*

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 $\Sigma_{b_0}=H_{b_0}V_{b_0}H'_{b_0}$ 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*

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

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 noncentrality parameter $\mu_n(\theta)$ we have $X_q^2(\mu_n(\theta))\sim\chi_q^2(\mu_n(\theta))$ and* (ii) *$\mu_n(\theta)=nh(\theta)'\hat{\Sigma}_n^{-1}h(\theta)$.*

Lemma 3 and Theorem 3 combine to give the asymptotic justifications of $\Pi(\eta,p)$ outlined in Section 3.1 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 IP functions of the LR, LM, and Hausman tests, as are used with the Wald test (when the latter is based on the unrestricted ML estimator). Thus, we have the following Corollary:

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.*

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

$$(5.25) \qquad \sup_{\theta \in \Theta:\, \theta = \theta_{b_0} + \delta/\sqrt{n}\,\&\|\delta\| \leqslant M} \left| P_\theta\left(W_{n_m} \leqslant s\right) - P\left(X_q^2\left(\mu_{n_m}(\theta)_\omega\right) \leqslant s\right)\right| \xrightarrow{m\to\infty} 0$$

for all $s \in R$ and all $M < \infty$, where $\mu_{n_m}(\theta)_\omega = n_m h(\theta)'\tilde{\Sigma}_{n_m\omega}^{-1} h(\theta)$. Hence, given any subsequence of $\{n\}$ there exists a sub-subsequence $\{n_m\}$ such that (5.25) holds for all ω in a set with P-probability one. This yields

$$(5.26) \qquad \sup_{\theta \in \Theta \quad \theta = \theta_{b_0} + \delta/\sqrt{n}\,\&\|\delta\| \leqslant M} \left| P_\theta\left(W_n \leqslant s\right) - P\left(X_q^2\left(\mu_n(\theta)\right) \leqslant s\right)\right| \xrightarrow{p} 0$$

as $n \to \infty$

under P for all $s \in R$ and all $M < \infty$, where $\mu_n(\theta) = nh(\theta)'\tilde{\Sigma}_n^{-1}h(\theta)$. Since $L_P(\tilde{\Sigma}_n) = L_{P_{\theta_n}}(\hat{\Sigma}_n)$, (5.26) holds under $\{\theta_n\}$ with $\mu_n(\theta) = nh(\theta)'\hat{\Sigma}_n^{-1}h(\theta)$. *Q.E.D.*

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 is reduced by specifying or estimating 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. These approximations are helpful in answering questions that are important to address in practical contexts.

Cowles Foundation, Department of Economics, Yale University, New Haven, CT 06520, U.S.A.

Manuscript received September, 1986, final revision received December, 1988

REFERENCES

ANDREWS, D. W K (1988): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.

ANDREWS, D. W. K., AND R. C. FAIR (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615–640.

ASHENFELTER, O., AND G. E. JOHNSON (1972). "Unionism, Relative Wages, and Labor Quality in U.S. Manufacturing Industries," *International Economic Review*, 13, 488–508

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.

Fox, M. (1956): "Charts of the Power of the F-test," *Annals of Mathematical Statistics*, 27, 484–497.

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.

Krämer, W. (1984): "The Power of the Durbin-Watson Test for Regressions without an Intercept," *Journal of Econometrics*, 28, 363–370.

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

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

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.

Nelson, F. D., and N. E. Savin (1988). "The Nonmonotonicity of the Power Function of the Wald Test in Nonlinear Models," Department of Economics Working Paper Series No. 88-7, University of Iowa, Iowa City, IA.

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

Pearson, E. S., and H. O. Hartley (1951): "Charts of the Power Function of the Analysis of Variance Tests, Derived from the Noncentral F-distribution," *Biometrika*, 38, 112–130.

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.

——— (1984b): "Hypothesis Testing in Linear Models When the Error Covariance Matrix Is Nonscalar," *Econometrica*, 52, 827–842.

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. Intriligator. New York: North-Holland.

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

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.

Yamauti, 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.