Econometrica, Vol. 59, No. 4 (July, 1991), 981–995

THE LOCAL NATURE OF HYPOTHESIS TESTS INVOLVING INEQUALITY CONSTRAINTS IN NONLINEAR MODELS[1]

BY FRANK A. WOLAK

This paper derives several properties unique to nonlinear model hypothesis testing problems involving linear or nonlinear inequality constraints in the null or alternative hypothesis. The paper is organized around a lemma which characterizes the set containing the least favorable parameter value for a nonlinear model inequality constraints hypothesis test. We then present two examples which illustrate several implications of this lemma. We also discuss the impact of these properties on the empirical implementation and interpretation of these test procedures.

KEYWORDS: Hypothesis testing, multivariate nonlinear one-sided tests, nonlinear inequality constraints tests.

1. INTRODUCTION

ECONOMETRICIANS HAVE RECENTLY BECOME INTERESTED in hypothesis testing problems involving inequality constraints in nonlinear models. Gourieroux, Holly, and Monfort (1980), hereafter GHM, consider a nonlinear generalization of the multivariate one-sided test: H: $h(\theta)=0$ versus K: $h(\theta)\geqslant 0$, where $h(\theta)$ is a nonlinear vector-valued function of $\theta\in R^k$.[2] Kodde and Palm (1986), henceforth K&P, derive a distance test approach to hypothesis tests of the form H: $h_1(\theta)=0, h_2(\theta)=0$ versus K: $h_1(\theta)\neq 0, h_2(\theta)\geqslant 0$ and H: $h_1(\theta)=0, h_2(\theta)\geqslant 0$ versus K: $\theta\in R^k$, where $h_1(\theta)$ and $h_2(\theta)$ are subvectors of $h(\theta)$. We refer to all nonlinear model hypothesis tests involving nonlinear or linear inequality constraints by the name nonlinear one-sided (NOS) hypothesis tests.

The purpose of this paper is to derive several properties unique to NOS hypothesis tests which clarify the empirical utility of these procedures. These properties arise because of the functional dependence of $V(\theta_0)$, the asymptotic covariance matrix of $n^{1/2}(\tilde{\theta}-\theta_0)$, on θ_0, the true value of θ, where $\tilde{\theta}$ is a consistent, asymptotically normal estimate of θ_0. The addition of nonlinear inequality constraints compounds this problem by adding another source of dependence on θ_0 of the asymptotic covariance matrix relevant to the hypothesis testing problem. By the same logic used to derive the asymptotic distribution for nonlinear equality constraints tests, we show that NOS tests asymptotically reduce to linear inequality constraints tests on a linearized version of the model local to the assumed true parameter value θ_0. This large-sample reduction of a

[1] I would like to thank the referees for helpful comments on earlier drafts that lead to significant improvements in the paper's content and clarity. Lars Peter Hansen provided very insightful comments and useful editorial suggestions. Franz C. Palm made several useful suggestions on a previous version. Bart Hamilton expertly prepared the figures. The final version of this paper was prepared while I was a National Fellow of the Hoover Institution. Its financial support is gratefully acknowledged. Partial financial support was provided by National Science Foundation Grant SES-90-57511.

[2] These authors derive results for the test H: $\theta_1=\theta_1^{\#}$ versus K: $\theta_1\geqslant\theta_1^{\#}$, where θ_1 and $\theta_1^{\#}$ are subvectors of θ and $\theta^{\#}$ (a known vector), but in an appendix outline how an extension to nonlinear constraints is possible.

nonlinear problem to a linearized version at an assumed true parameter value (what we term *the local nature of hypothesis tests involving inequality constraints in nonlinear models*) interacts with the functional dependence of $V(\theta_0)$ on θ_0 to produce all of the properties studied here.

The three main properties treated in the paper are summarized below. The one which limits the empirical utility of these testing techniques most is the lack of an empirically implementable procedure for computing an asymptotically exact size critical value. For general NOS hypothesis tests, we show that the least favorable null asymptotic distribution can only be limited to a set of possible distributions, not to a unique distribution, as is the case for linear inequality constraints tests in linear models. A second property, which also impacts on the empirical content of these techniques, is the absence of tight upper and lower bounds on the asymptotic distribution of the test statistics similar to those available for linear model, linear inequality constraints analogues of these procedures. The third property is the surprising result that even in the case of simple linear inequality constraints on the parameters of a nonlinear model, the least favorable null asymptotic distribution may not occur at the unique parameter value satisfying all of the inequality constraints with equality.

We also discuss several implications of these properties. The first concerns the relative merits of alternative techniques for empirically approximating the true asymptotic critical value. The second implication of these properties is that the bounds on the null asymptotic distribution become more and more slack as the dimension of the vector of nonlinear inequality constraints involved in the hypothesis test gets larger. We also discuss previous work on NOS hypothesis tests in light of these properties.

Another purpose of this paper is to highlight the difference between nonlinear model testing problems involving inequality constraints and those only involving equality constraints. To perform an asymptotically exact test of either of these composite null hypotheses, we must find the parameter value in the set defining the null hypothesis yielding the largest critical value for a fixed size asymptotic test. The major difference between these two testing frameworks is that different true parameter values in the set defining the null hypothesis result in different distributions for the NOS test statistics. The geometry of the constraints set local to the assumed true parameter value determines the null asymptotic distribution for NOS tests. In contrast, for nonlinear equality constraints tests, the same asymptotic distribution obtains for all parameter values satisfying the null hypothesis.

The remainder of the paper proceeds as follows. Section 2 first derives a lemma which specifies a subset of the set defining the null hypothesis which must contain the least favorable null parameter value for a nonlinear inequality constraints test. Then this section presents an example which illustrates the cause of the potential multiplicity of null asymptotic distributions. We then discuss a simple linear inequality constraints test in a nonlinear model where the least favorable parameter value does not satisfy all of the inequalities with equality. Section 3 explains why the bounds on the asymptotic distribution, which are tight for linear models and simple linear inequality constraints, are slack for NOS tests. Section 4 makes specific recommendations for applying and interpreting these test procedures.

2. DERIVATION OF SET CONTAINING LEAST FAVORABLE PARAMETER VALUE

To establish the necessary background for our subsequent results, we first review the relevant features of the simple linear inequality constraints tests. This test takes the form: H: $\mu \geqslant 0$ versus K: $\mu \in R^p$ based on $\hat{\mu} \sim N(\mu, \Omega)$, where Ω is a known positive definite matrix. Perlman (1969) presents a general framework for analyzing these types of testing problems. Following his logic, the likelihood ratio (LR) statistic is defined as a minimum chi-squared-like distance between $\hat{\mu}$ and the set in which μ lies under the null hypothesis. In this case we have

$$\text{(2.1)} \qquad \text{LR} = \inf_m \left[(\hat{\mu} - m)'\Omega^{-1}(\hat{\mu} - m) \text{ subject to } m \geqslant 0 \right].$$

This null hypothesis does not specify a unique μ. To derive an exact size α critical value for a linear inequality constraints hypothesis test, a least favorable true value of μ in the set defining the null hypothesis must be chosen. Lemma 8.2 of Perlman (1969) implies that for any $c > 0$,

$$\text{(2.2)} \qquad \sup_{\mu \geqslant 0} \text{pr}[\text{LR} \geqslant c | \mu, \Omega] = \text{pr}[\text{LR} \geqslant c | 0, \Omega],$$

where $\text{pr}[\text{LR} \geqslant c | \mu, \Omega]$ is the probability of the event $[\text{LR} \geqslant c]$ given that $\hat{\mu}$ in (2.1) is $N(\mu, \Omega)$. Therefore, the least favorable value of $\mu \geqslant 0$ which determines the exact size null distribution is $\mu = 0$. As stated in Perlman (1969), crucial to the validity of Lemma 8.2 is the lack of functional dependence of Ω on μ, so that variation in μ cannot affect the elements of Ω. Unfortunately, with nonlinear models or constraints, the functional dependence of the asymptotic covariance matrix of the constraints function estimator on the true vector of parameters is intrinsic to the problem.

We now state a lemma which is the closest we can get to Perlman's Lemma 8.2 for the general nonlinear inequality constraints testing framework. For expositional ease, we derive this lemma for the maximum likelihood model, although under suitable regularity conditions the result holds for more general classes of nonlinear models. Let $\theta \in R^k$ denote the parameter vector, $\Theta \subset R^k$ a compact parameter space containing θ, and $L(\theta)$ the log-likelihood function based on n observations. Under the standard regularity conditions,[3] $n^{1/2}(\hat{\theta} - \theta_0) \to N(0, I(\theta_0)^{-1})$ for all $\theta_0 \in \text{interior}(\Theta)$, where $\hat{\theta}$ is the maximum likelihood (ML) estimate of θ_0 and

$$I(\theta) = \lim_{n \to \infty} E_{\theta_0} \left[-\frac{1}{n} \frac{\partial^2 L(\theta)}{\partial \theta \, \partial \theta'} \right]$$

[3] Here and for the remainder of the paper, the standard regularity conditions refer to those in Amemiya (1985, Chapter 4).

is Fisher's information matrix. This matrix can be consistently estimated by

$$\hat{V}(\boldsymbol{\theta}) = -\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial\theta\,\partial\theta'}.$$

Assuming that θ_0 is the true θ, we can show $\hat{V}(\hat{\theta})$ converges in probability to $I(\theta_0)$, so that $\hat{V}(\hat{\theta})^{-1}$ is a consistent estimate of the asymptotic covariance matrix of $n^{1/2}(\theta - \theta_0)$.

Consider the nonlinear inequality constraints (NIC) test

$$(2.3)\qquad H\colon h(\theta) \geqslant 0 \quad \text{versus} \quad K\colon \theta \in R^k$$

for this maximum likelihood model framework.[4] In this case $h(\theta) = (h_1(\theta), h_2(\theta), \ldots, h_p(\theta))'$ is a p-dimensional nonlinear vector-valued function from R^k to R^p $(p \leqslant k)$. Let $H(\theta)$ represent the $(p \times k)$ matrix of partial derivatives $(\partial h_i(\theta)/\partial\theta_j,\ i = 1, \ldots, p$ and $j = 1, \ldots, k)$ of $h(\theta)$ with respect to the elements of θ. From K&P and Wolak (1989), one of the three asymptotically equivalent forms of the test statistic for (2.3) is:

$$W_n = \inf_t \left[n\big(h(\hat{\theta}) - t\big)'\hat{J}(\hat{\theta})^{-1}\big(h(\hat{\theta}) - t\big) \text{ subject to } t \geqslant 0 \right]$$

where $\hat{J}(\hat{\theta}) = [H(\hat{\theta})\hat{V}(\hat{\theta})^{-1}H(\hat{\theta})']$ and $J(\theta_0) = [H(\theta_0)I(\theta_0)^{-1}H(\theta_0)']$. We now state our result in the form of a three-part lemma proven in the Appendix.

First we define the necessary notation. Let $C = \{\theta \mid h(\theta) \geqslant 0\}$, $C^i = \{\theta \mid h(\theta) > 0\}$, and $C^b = C - C^i$. Partition C^b into the sets: $A = \{\theta \mid \theta \in C^b$ and $h_j(\theta) = 0$ for only one $j \in P\}$, where $P = \{1, 2, \ldots, p\}$ is the set of integers from 1 to p, and $B = C^b - A$. For any $\theta_0 \in C^b$ let $h_b(\theta)$ denote a vector of elements of $h(\theta)$ such that $h_j(\theta_0) = 0$ $(j \in P)$, and $m \leqslant p$ denote the dimension of $h_b(\theta)$. Let $H_b(\theta_0)$ represent the matrix of first partial derivatives of $h_b(\theta)$ with respect to θ.

LEMMA 1: *For testing problem* (2.3) *we have the following three results.*

(1) *For all* $\theta_0 \in C^i$,

$$(2.4)\qquad \lim_{n\to\infty} \mathrm{pr}\left[W_n = 0 \mid \theta_0\right] = 1.$$

(2) *For any* $c > 0$ *and* $\theta_0 \in C^b$ *such that* $H_b(\theta_0)$ *is of full row rank*:

$$(2.5)\qquad \lim_{n\to\infty} \mathrm{pr}\left[W_n \geqslant c \mid \theta_0\right] = \sum_{j=0}^{m} \mathrm{pr}\left[\chi_j^2 \geqslant c\right] w\big(m, m-j, J_b(\theta_0)\big)$$

where $J_b(\theta_0) = [H_b(\theta_0)I(\theta_0)^{-1}H_b(\theta_0)']$ *and the* $w(m, m-j, J_b(\theta_0))$ *are the weights functions defined in Kudo* (*1963*).

(3) *For all* $\theta_0 \in B$, $\theta_0^* \in A$ *and fixed* $c > 0$,

$$(2.6)\qquad \lim_{n\to\infty} \mathrm{pr}\left[W_n \geqslant c \mid \theta_0\right] \geqslant \lim_{n\to\infty} \mathrm{pr}\left[W_n \geqslant c \mid \theta_0^*\right].$$

[4] All NOS hypothesis tests considered in this paper are under the maintained hypothesis that θ lies in Θ.

This lemma contains several results. Part (1) states that all elements in C^i, the interior of the constraint set, can be removed from consideration as a least favorable value, because W_n converges in probability to zero for these values of θ_0. Part (2) provides a general characterization of the asymptotic distribution of W_n for all values of $\theta_0 \in C^b$. In particular, it shows the impact, on the asymptotic distribution of W_n, of any one of the inequality restrictions not holding as an equality at θ_0. Finally, part (3) eliminates certain elements of C^b from consideration as potential least favorable values of θ_0. Taken together, these three results require the least favorable value of θ_0 to be an element of the set B. Although this lemma specifies that $B \subset C$ must contain the least favorable value of θ_0 under the null hypothesis, in general, it does not yield the solution in θ_0 to

$$(2.7) \qquad \sup_{\theta_0 \in C} \lim_{n \to \infty} \operatorname{pr}\left[W_n \geqslant c \mid \theta_0\right],$$

for a given $c > 0$, which is the optimization problem that must be solved to compute the least favorable $\theta_0 \in C$ for the critical value c.

We now describe an algorithm for solving (2.7) for an arbitrary nonlinear model inequality constraints test. Unfortunately, because of the computational difficulties to be described, this procedure is, in most cases, useful only to demonstrate that the least favorable value exists, not to actually find it. Let $c > 0$ denote an arbitrary critical value. For a fixed value of $\bar{\theta}_0 \in B$, the first step in the process is determining which elements of $h(\theta)$ comprise the vector $h_b(\theta)$ described in part (2) of Lemma 1. Once the vector-valued function $h_b(\theta)$ is specified for $\bar{\theta}_0$, the large-sample probability $\lim_{n \to \infty} \operatorname{pr}[W_n \geqslant c \mid \bar{\theta}_0]$ can be computed from equation (2.5). This process must be repeated for all $\theta_0 \in B$. The value of $\theta_0 \in B$ which maximizes this large sample probability is the least favorable value of θ_0 determining the asymptotically exact size of the test for the critical value c. Because of the functional dependence of this least favorable value of θ_0 on c, we write it as $\theta_0^*(c)$. The critical value for an asymptotically exact size α NIC test is the c that solves $\lim_{n \to \infty} \operatorname{pr}[W_n \geqslant c \mid \theta_0^*(c)] = \alpha$.

Actually implementing this algorithm deserves some comment. Although the $w(p, j, \Sigma)$ $(j = 0, \ldots, p)$ are only available in closed form for $p \leqslant 4$, Kudo (1963) gives expressions for these weights for arbitrary p as sums of products of multivariate normal probabilities determined from submatrices of Σ. Consequently, for each value of $J_b(\theta_0)$, the distribution given in (2.5) can be computed. However, the problem of determining $J_b(\theta_0)$ for each θ_0 still remains. As defined in part (2) of Lemma 1, $J_b(\theta_0)$ is a function of the information matrix, which does not have an analytical form for many nonlinear models. Consequently, for these models the nonexistence of an analytical expression for the matrix $I(\theta)$ precludes computing the asymptotically exact critical value.

Nevertheless, there are two situations where the unique least favorable value of θ_0 exists for the NIC test. If the matrix $I(\theta_0)$ is diagonal for all $\theta_0 \in B$ and the inequality constraints take the form of simple bounds, the least favorable value is the unique θ_0 which satisfies all of the inequalities with equality. For the case of a 1-dimensional nonlinear inequality constraint test, all of the values of θ_0 which satisfy the inequality constraint with equality yield the same asymptotic distribution. This result is a large-sample nonlinear analogue of a one-sided t test.

We now present a two-dimensional example, which conveys the intuition underlying this multiplicity of null asymptotic distributions for NIC hypothesis tests. Consider the testing problem

$$(2.8) \qquad H: h(\theta) \geqslant 0 \quad \text{versus} \quad K: \theta \in R^2$$
$$\text{based on} \qquad Z = \theta + \nu, \qquad \nu \sim N(0, I),$$

where I is a (2×2) identity matrix. Define the elements of $h(\theta) \in R^2$ as follows: $h_1(\theta) = \theta_2 - \theta_1^2$ and $h_2(\theta) = \theta_1 - \theta_2$ for $\theta = (\theta_1, \theta_2)'$. The matrix of first-partial derivatives of $h(\theta)$ takes the form:

$$(2.9) \qquad H(\theta) = \begin{bmatrix} -2\theta_1 & 1 \\ 1 & -1 \end{bmatrix},$$

for all $\theta \in R^2$. Given n observations on Z, the ML estimate of θ_0 is $\hat{\theta} = (1/n)\Sigma_{i=1}^{n} Z_i$, which is $N(\theta_0, (1/n)I)$. Therefore $n^{1/2}(\hat{\theta} - \theta_0) \sim N(0, I)$ and $n^{1/2}(h(\hat{\theta}) - h(\theta_0))$ is, by the delta-method, asymptotically $N(0, J(\theta_0))$, where $J(\theta_0) = H(\theta_0)H(\theta_0)'$. The test statistic for (2.8) is

$$(2.10) \qquad W_n = \inf_t \left[n\left(h(\hat{\theta}) - t\right)' J(\hat{\theta})^{-1} \left(h(\hat{\theta}) - t\right) \text{ subject to } t \geqslant 0 \right],$$

where $J(\hat{\theta}) = H(\hat{\theta})H(\hat{\theta})'$. By an application of Lemma 1 and because $h(\theta) \in R^2$, the least favorable value of θ_0 under the null hypothesis must be a member of the set $T \equiv \{\theta | h(\theta) = 0\}$. For this simple problem there are two $\theta_0 \in T$: $\theta_0^{00} = (0,0)'$ and $\theta_0^{11} = (1,1)'$. The two sets of inequalities $H(\theta_0^{00})\theta \geqslant (0,0)'$ and $H(\theta_0^{11})\theta \geqslant (-1,0)'$ are linearized versions of the inequality constraints $h(\theta) \geqslant 0$ at these two points. Each set of linear inequality constraints defines a cone of tangents to $C = \{\theta | h(\theta) \geqslant 0, \theta \in R^2\}$ at the value of θ satisfying all of the linear inequalities with equality. Because $\hat{\theta}$ converges in probability to θ_0, we know that $J(\hat{\theta})$ converges in probability to $J(\theta_0)$ for all θ_0 in the parameter space, so that we have:

$$J(\theta_0^{00}) = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \quad \text{and} \quad J(\theta_0^{11}) = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}.$$

Because the asymptotic distribution of (2.10) depends on θ_0 through $J(\theta_0)$, each value of θ_0 implies a different null asymptotic distribution. From Wolak (1989), the asymptotic distribution of W_n, given $\theta_0 \in T$, is

$$(2.11) \qquad \lim_{n \to \infty} \operatorname{pr}[W_n \geqslant c] = \sum_{j=0}^{2} \operatorname{pr}\left(\chi_j^2 \geqslant c\right) w(2, 2-j, J(\theta_0)).$$

This asymptotic distribution is functionally dependent on θ_0 via $J(\theta_0)$ through the weights, $w(2, 2-j, J(\theta_0))$. The weights for $J(\theta_0^{00})$ and $J(\theta_0^{11})$ can be calculated from the closed form expressions given in Wolak (1987). These weights

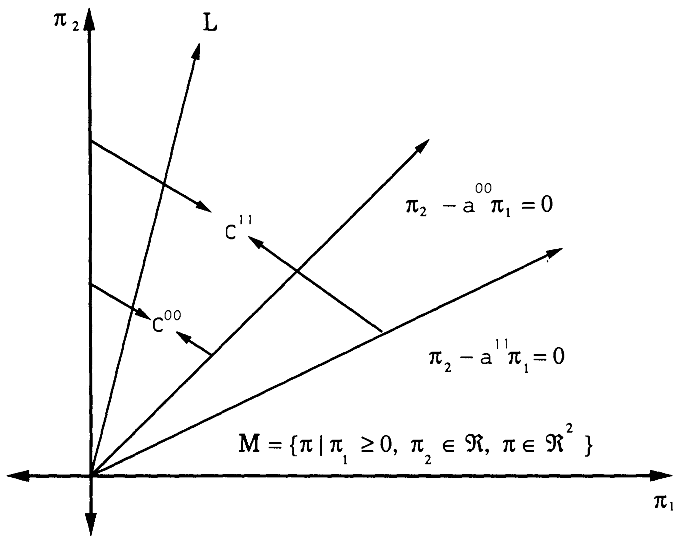

FIGURE 1.—Graphical representation of canonical linear inequality constraint sets for hypothesis test (2.8).

yield two potential null asymptotic distributions which result in two different critical values for the test.

Reducing each local linear inequality constraints test in terms of θ to its canonical form most clearly illustrates the effect of varying θ_0 on the resulting null asymptotic distribution. As discussed in GHM (1982) and Wolak (1988), for any linear inequality constraint framework $R\beta \geqslant r$ with R of full row rank and test statistic based on $\hat{\beta} = \beta + \varepsilon$, $\varepsilon \sim N(0, \Sigma)$, there exists an invertible affine transformation to the framework $A\pi \geqslant 0$ (A is lower triangular with 1's along the diagonal) and test statistic based on $\hat{\pi} = \pi + \eta$, $\eta \sim N(0, I)$. We call this inequality constraint framework in terms of π the canonical linear inequality constraints set. This form has the attractive feature of defining all test statistics in terms of Euclidian distances, because the covariance matrix of $\hat{\pi}$ is the identity matrix. It also defines all inequality constraints so that they describe polyhedral cones. For our two-dimensional problem, once we specify the $(2,1)$ element of A (a in the notation of Wolak (1988)), the entire canonical linear inequality constraints framework is specified. Applying the invertible affine transformation given in Wolak (1988) to the test at θ_0^{00} implies $a^{00} = 1.0$, and applying it to θ_0^{11} yields $a^{11} = 0.6$. The constraints sets in (π_1, π_2)-space corresponding to these two values of θ_0 are given in Figure 1. The constraint set corresponding to θ_0^{11} contains that corresponding to θ_0^{00}, so that θ_0^{00} is the least favorable null value of θ_0.

The dependence of the null asymptotic distribution of NIC tests on the local geometry of the constraint set relative to θ_0 is not unique to this form of $h(\theta)$. For the case of $\theta \in R^k$ and $h(\theta)$ a nonlinear vector-valued function from R^k to R^p, where $p < k$, there may be an uncountable number of $\theta \in R^k$ such that

$h(\theta)=0$, and an uncountable number of potential null asymptotic distributions. By Lemma 1, for inequality constraints tests with $p>2$ the set of possible least favorable values expands beyond T to B. Nevertheless, the intuition provided by this example continues to hold; that is, the $\theta_0 \in B$ with the local linear inequality constraints test having the largest size α canonical linear inequality constraints critical value is the least favorable value of θ_0 under the nonlinear model inequality constraints null hypothesis.

To explore the implications of Lemma 1 further, we present a simple inequality constraints test where the set of least favorable parameter values expands beyond T to B, and the least favorable parameter value does not satisfy all of the inequalities with equality. In the process, we describe the features of the NIC hypothesis test necessary for this result to occur.

Consider the ML model for the case of independent identically distributed observations from the bivariate normal distribution

$$X_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}\right).$$

For this model, we have $\theta=(\sigma^2,\tau^2,\rho)'$ and $\Theta=\{\theta \mid \theta_1 \geqslant 0, \theta_2 \geqslant 0, -1 \leqslant \theta_3 \leqslant 1\}$. The ML estimate of θ_0, $\hat{\theta}$, is given in Lehmann (1983, pp. 439–440). Under the standard regularity conditions, $n^{1/2}(\hat{\theta}-\theta_0)$ converges in distribution to a $N(0, I(\theta_0)^{-1})$ random vector, where $I(\theta_0)^{-1}$ takes the form given in Lehmann (1983, p. 441). The correlation matrix of $I(\theta)^{-1}$ is equal to:

$$(2.12)\qquad c\left(I(\theta)^{-1}\right)=\begin{bmatrix} 1 & \rho^2 & \rho/\sqrt{2} \\ \rho^2 & 1 & \rho/\sqrt{2} \\ \rho/\sqrt{2} & \rho/\sqrt{2} & 1 \end{bmatrix}.$$

This correlation matrix is the minimum information necessary to compute the null asymptotic distribution for any NIC hypothesis test involving θ because the $w(p,j,\Sigma)$ depend on Σ only through its correlation matrix.

Within the context of this three-parameter model consider the hypothesis test

$$(2.13)\qquad H: \theta \leqslant \theta^V \qquad \text{versus} \qquad K: \theta \in \Theta \subset R^3,$$

where $\theta^V=(1.0, 1.0, 0.95)$. One of the three asymptotically equivalent forms of the test statistic for this hypothesis test is:

$$W_n=\inf_t\left[n(\hat{\theta}-t)'I(\hat{\theta})(\hat{\theta}-t) \text{ subject to } t \leqslant \theta^V\right].$$

For $\theta_0=\theta^V$ and any $c>0$, the exact limiting distribution of W_n is given by the following weighted sum of χ^2-distributions:

$$(2.14)\qquad \lim_{n\to\infty} \operatorname{pr}\left[W_n \geqslant c\right]=0.015263\operatorname{pr}\left[\chi_3^2 \geqslant c\right] + 0.168204\operatorname{pr}\left[\chi_2^2 \geqslant c\right]+0.484737\operatorname{pr}\left[\chi_1^2 \geqslant c\right].$$

The weights in (2.14) were calculated using the correlation matrix given in (2.12) evaluated at $\theta = \theta^V$ and the formulae given in the Appendix of Wolak (1987). In the Appendix we show that the asymptotic distribution of W_n for $\theta_0 = \theta^B = (1.0, 1.0, 0.0)'$ is

$$\lim_{n \to \infty} \operatorname{pr}[W_n \geqslant c] = \tfrac{1}{4} \operatorname{pr}[\chi_2^2 \geqslant c] + \tfrac{1}{2} \operatorname{pr}[\chi_1^2 \geqslant c]. \tag{2.15}$$

We can also show that $\theta_0 = \theta^B$ yields the least favorable null asymptotic distribution for this hypothesis test. Figure 2 contains plots of $\Gamma(c|\theta) \equiv \lim_{n \to \infty} \operatorname{pr}[W_n \geqslant c|\theta]$ for θ^V and θ^B. The plot labeled θ^V in the figure corresponds to $\Gamma(c|\theta^V)$ and the plot labeled θ^B corresponds to $\Gamma(c|\theta^B)$. For all values of c in the diagram, $\Gamma(c|\theta^B) > \Gamma(c|\theta^V)$, despite the fact that θ^V is the value of θ which satisfies all of the inequality constraints with equality. In this case, because of the form of the functional dependence of the asymptotic covariance matrix of $n^{1/2}(\hat{\theta} - \theta_0)$ on θ_0, the least favorable value of $\theta_0 \in C$ only satisfies two of the three inequalities with equality.

This result is particularly relevant to the NIC framework derived by K&P. These authors treat the case in which $J(\theta)$ does not depend on θ, and consequently the least favorable value of θ_0 lies in T. However, as Lemma 1 and this example show, the least favorable value of θ_0 need not lie in T when $J(\theta)$ depends on θ as is often the case in nonlinear models.

We now discuss an alternative procedure for computing critical values for these tests which can run into difficulties for the reasons given in the above example. This procedure involves computing the weights, $w(p, j, J(\theta_0))$, using a consistent estimate of $J(\theta_0)$ evaluated at $\hat{\theta}$, the unrestricted estimate of θ_0. For the ML model and hypothesis test of Lemma 1, the weights are computed using $\hat{J}(\hat{\theta})$ as described earlier, instead of $J(\theta_0)$. By this logic, the critical value $c_\alpha(\hat{\theta})$ solves the following equation in c:

$$\alpha = \sum_{j=0}^{p} \operatorname{pr}[\chi_j^2 \geqslant c] w(p, p-j, \hat{J}(\hat{\theta})) \tag{2.16}$$

where $\hat{\theta}$ is the MLE of θ. We can show that so long as the least favorable value of θ_0 is in T, this procedure will yield an asymptotically exact size α critical value. However, if the least favorable value of θ_0 does not satisfy all the inequalities with equality, as in hypothesis test (2.13), this critical value will not lead to an asymptotically exact size test. In these instances, because of the asymptotic dimension reduction in the nonlinear inequality constraints test for each $h_j(\theta_0) > 0$, the quantity $c_\alpha(\hat{\theta})$ defined in (2.16) does not converge in probability to the asymptotically exact size critical value $c_\alpha^b(\theta_0^*)$, where θ_0^* is the least favorable value of $\theta_0 \in B$. When $\theta_0 = \theta_0^*$ (the true parameter value equals the least favorable parameter value), the critical value $c_\alpha^b(\theta_0^*)$ is asymptotically equivalent to the solution in c to

$$\alpha = \sum_{j=0}^{m} \operatorname{pr}[\chi_j^2 \geqslant c] w(m, m-j, \hat{J}_b(\hat{\theta})), \tag{2.17}$$

where $\hat{J}_b(\theta) = H_b(\hat{\theta})\hat{V}(\hat{\theta})^{-1}H_b(\hat{\theta})'$. Even under the usual assumption for comput-

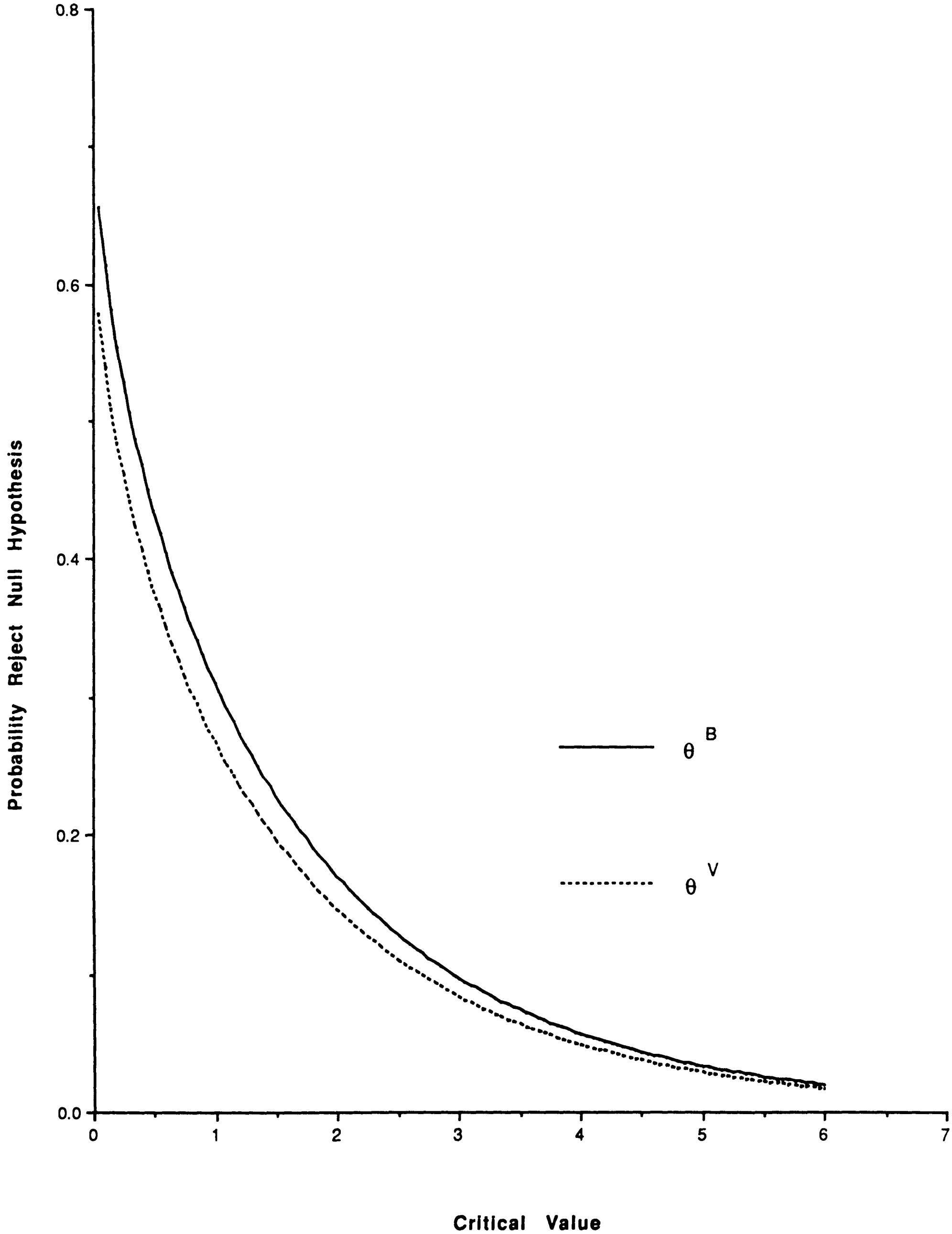

FIGURE 2.—Probability of rejecting null hypothesis for test (2.13) as a function of critical value.

ing critical values[5] that $\theta_0 = \theta_0^*$, if the least favorable value does not satisfy all of the inequalities with equality, solving equation (2.16) will not yield an asymptotically exact critical value. This procedure does not take into account the dimension reduction in the constraints set (i.e., $m < p$) caused by some of the inequalities not holding with equality at θ_0^*. In these cases, solving (2.16) instead of (2.17) can yield too small or large of an asymptotic critical value, so that $\lim_{n \to \infty} \operatorname{pr}[W_n \geqslant c_\alpha(\hat{\theta}) | \theta_0^*] \neq \alpha$.

Kodde and Palm (1986) suggest solving (2.16) to compute critical values for NOS tests. Their procedure yields an asymptotically exact test for all NOS tests where $\theta_0^* \in T$. However, as the above discussion demonstrates, for higher dimension ($p > 2$) NIC tests with $\theta_0^* \in B$, solving (2.16) for $c_\alpha(\hat{\theta})$ to compute critical values can lead to incorrect asymptotic size tests.

3. SLACKNESS OF BOUNDS ON NULL DISTRIBUTION

This section discusses the slackness of the bounds on the exact null asymptotic distribution of the NOS test statistics. Using the canonical linear inequality constraints form of an NOS test, we can derive bounds on the null asymptotic distribution of these test procedures. Consider a multivariate inequality constraints test (2.3) for $\theta \in R^k$ and $h(\theta) \in R^p$ ($p \leqslant k$). As described earlier, we can reduce any large-sample local linear inequality constraint test at θ_0 to its canonical form in terms of π and $A\pi \geqslant 0$ as shown in Figure 1. Each $\theta_0 \in B$ implies a polyhedral cone defining the large-sample canonical linear inequality constraints null hypothesis in π-space. For any polyhedral cone defined by $A\pi \geqslant 0$ there is a half-space in R^p containing it and a half-line in R^p contained in it. In terms of our notation, for any $C^A \equiv \{x | Ax \geqslant 0, x \in R^k\}$, there exists a half-space, $M \supset C^A$ and half-line, $L \subset C^A$. (See Figure 1 for the case of $p = 2$.) Define

$$W_M = \inf_{\phi \in M} (\hat{\pi} - \phi)'(\hat{\pi} - \phi) \quad \text{and} \quad W_L = \inf_{\phi \in L} (\hat{\pi} - \phi)'(\hat{\pi} - \phi).$$

Clearly, $W_M \leqslant W \leqslant W_L$, where W is the large-sample value of W_n. Consequently, we have:

$$\text{(3.1)} \qquad \inf_{\theta_0 \in S} \operatorname{pr}\left[D \geqslant c | \theta_0, J(\theta_0)\right] \geqslant \operatorname{pr}\left[W_M \geqslant c | \pi = 0\right] = \tfrac{1}{2} \operatorname{pr}\left[\chi_1^2 \geqslant c\right];$$

$$\text{(3.2)} \qquad \sup_{\theta_0 \in S} \operatorname{pr}\left[D \geqslant c | \theta_0, J(\theta_0)\right] \leqslant \operatorname{pr}\left[W_L \geqslant c | \pi = 0\right]$$

$$= \tfrac{1}{2} \operatorname{pr}\left[\chi_{p-1}^2 \geqslant c\right] + \tfrac{1}{2} \operatorname{pr}\left[\chi_p^2 \geqslant c\right],$$

where $S = B$ for the NIC test, $S = T$ for the test H: $h(\theta) = 0$ versus K: $h(\theta) \geqslant 0$, and D is the large-sample value of the respective NOS test statistic.

An alternative way to get the right-hand side of (3.1) is to take the infimum with respect to all positive definite Ω of the rejection probability given in (2.2)

[5] Lehmann (1986, pp. 68–69) discusses this approach to computing critical values for hypothesis tests having a general composite null hypothesis.

holding $\mu = 0$. The supremum of (2.2) over this same set of Ω holding $\mu = 0$ yields to the right side of (3.2). For this linear model linear inequality constraint framework, because Ω, the covariance matrix of μ, is allowed to vary independently of μ, the infimum and supremum of the probability in (2.2) can be taken with respect to all positive definite matrices while maintaining $\mu = 0$. Consequently, for this case, Perlman (1969) shows that the right-hand side of (3.1) and (3.2) are tight bounds on the null distribution. However, for the case of nonlinear models and/or constraints, the asymptotic covariance matrix of the constraints vector $J(\theta_0)$ can only be altered by changing θ_0. Consequently, by varying θ_0 over S, the entire set of positive definite matrices cannot be traced out by $J(\theta_0)$. This implies that the distributions given in (3.1) and (3.2) are slack for most NOS tests. Both of the examples in Section 2 illustrate this point.

Exactly how slack these bounds are depends on the number of inequality constraints under examination, as well as the specific model and constraints set. Because (3.1) is independent of p, the dimension of $h(\theta)$, by inspection of (3.2), as p grows the difference between the upper and lower bounds on the asymptotic distribution will increase. Unfortunately, there is no simple set of conditions on the model or constraint set which will allow a characterization of the instances where the actual asymptotic distribution is closer to the upper bound than the lower bound or vice versa.

The results given (3.1) and (3.2) have implications for the distributional bounds given in K&P. For NOS tests when $J(\theta)$ depends on θ, the upper and lowers bounds on the null asymptotic distribution given in equations (3.1) of K&P are no longer tight. Nevertheless, as demonstrated above, the upper and lower critical values computed in Table I of their paper still yield valid *slack* upper and lower bounds on the asymptotically exact critical value for any NOS hypothesis tests when $J(\theta)$ depends on θ. Consequently, these critical values can still be used to draw asymptotically valid inferences, but the increasing distance between the upper and lower critical values as p grows will make drawing definitive inferences increasingly unlikely.

These results also impact on the nonlinear model linear multivariate one-sided test considered by GHM (1980): H: $\theta_1 = \theta_1^{\#}$ versus K: $\theta_1 \geqslant \theta_1^{\#}$, where θ_1 and $\theta_1^{\#}$ are subvectors of θ and $\theta^{\#}$ (a known vector). Even though this test has a point null hypothesis, the problem of multiple null asymptotic distributions still arises if the asymptotic covariance matrix of the constraints vector depends on θ_2, the elements of θ excluded from the test. The bounds on the null asymptotic distribution given in (3.1) and (3.2) are also slack for this testing framework because, in general, the set of all positive definite matrices cannot be traced out by the asymptotic covariance matrix of $n^{1/2}(\theta_1 - \theta_1^{\#})$ by varying θ_2 over its parameter set.

4. CONCLUSIONS AND IMPLICATIONS FOR NOS TESTS

All of the problems associated with testing problems involving nonlinear inequality constraints arise from the dependence of the asymptotic distribution on the geometry of the inequality constraints local to the assumed true θ_0. A framework which explicitly recognizes these limitations of nonlinear inequality constraint tests in nonlinear models should not fall prey to these complications. Wolak (1989) presents a general local hypothesis testing framework for the case that either the null or alternative hypothesis is specified by a system of nonlinear inequality constraints or combinations of nonlinear inequality and equality constraints. That paper derives asymptotically exact local results for these hypothesis tests. Emphasis is placed on deriving the precise geometry of the set specified by the inequality constraints and consistently estimating the exact null asymptotic distribution implied by this set.

The results presented here are not meant to discourage the empirical implementation of these kinds of NOS procedures, only to encourage their proper use and the correct interpretation of the results. Several recommendations along these lines are possible. The major result of this paper is the difficulty in obtaining an empirically implementable asymptotically exact hypothesis test. This conclusion implies that in most instances an asymptotic bounds test is necessary. The results of Section 3 enter here. They illustrate that, in general, only slack upper and lower bounds on the asymptotic distribution of these test statistics exist. For tests involving higher dimensional inequality constraints, these bounds become very slack, making inconclusive test results more probable. However, economic theory or some other form of *a priori* information often yields a specific point on the boundary of C relative to which the nonlinear inequality constraints test can be performed. In these cases, the researcher should then perform the inequality constraint test local to this point. Proceeding as described in Wolak (1989), an asymptotically exact distribution for the test statistic is available so that a definitive conclusion concerning the hypothesis test can be reached.

Department of Economics, Stanford University, Stanford, CA 94305, U.S.A.

Manuscript received March, 1987; final revision received August, 1990.

APPENDIX

Proof of Lemma 1: First we show that for any set of nonlinear inequality constraints and any $\theta_0 \in C$, as $n \to \infty$, the number of inequality constraints used to compute W_n is reduced by the number of $h_j(\theta)$ such that $h_j(\theta_0) > 0$ for $j \in P$. Given this result, we then derive parts (1), (2), and (3) of Lemma 1.

Define $l_n(\theta_0) = n^{1/2}(h(\hat{\theta}) - h(\theta_0))$ for any $\theta_0 \in C$. Under our regularity conditions, $l_n(\theta_0)$ converges in distribution to a $N(0, J(\theta_0))$ random vector for all $\theta_0 \in$ interior(Θ). For a given value of $\theta_0 \in B$, we can rewrite the optimization problem determining W_n as

$$W_n = \inf_t \left[(l_n(\theta_0) - t)'\hat{J}(\hat{\theta})^{-1}(l_n(\theta_0) - t) \text{ subject to } t = x - n^{1/2}h(\theta_0),\ x \geqslant 0 \right]. \tag{A.1}$$

Substituting $t + n^{1/2}h(\theta_0)$ for x in (A.1) yields:

$$W_n = \inf_t \left[(l_n(\theta_0) - t)'\hat{J}(\hat{\theta})^{-1}(l_n(\theta_0) - t) \text{ subject to } t \geqslant -n^{1/2}h(\theta_0) \right]. \tag{A.2}$$

Suppose $\theta_0 \in B - T$. As defined in part (2), $h_b(\theta) \in R^m$ is the subvector of the inequality constraints satisfied with equality, so that $h_b(\theta_0) = 0$. Let $h_s(\theta) \in R^{p-m}$ denote the subset of $h(\theta)$ satisfied with strict inequality, so that $h_s(\theta_0) > 0$. Partitioning t conformably to $h(\theta)$, we have $t = (t_b', t_s')$. Using this notation, we can rewrite (A.2) to obtain:

$$\text{(A.3)} \qquad W_n = \inf_t \left[(l_n(\theta_0) - t)'\hat{J}(\hat{\theta})^{-1}(l_n(\theta_0) - t) \text{ subject to } t_b \geqslant 0, t_s \geqslant -n^{1/2}h_s(\theta_0) \right].$$

Taking the limit of (A.3) as $n \to \infty$ we obtain:

$$\text{(A.4)} \qquad W = \inf_t \left[(z - t)'J(\theta_0)^{-1}(z - t) \text{ subject to } t_b \geqslant 0, t_s \in R^{p-m} \right],$$

where the exact distribution of z is $N(0, J(\theta_0))$. Consequently, for the elements of $h(\theta_0)$ that are strictly greater than zero, in the limit, the corresponding elements of t are unconstrained in the optimization problem determining W, the limiting value of W_n.

To establish part (1) of the lemma note that if θ_0 is such that $h_s(\theta_0) = h(\theta_0)$, all of the inequality constraints are slack, then (A.4) reduces to:

$$\text{(A.5)} \qquad W = \inf_t \left[(z - t)'J(\theta_0)^{-1}(z - t) \text{ subject to } t \in R^p \right].$$

Because t is unrestricted, we have $W = 0$, so that $\lim_{n\to\infty} \text{pr}[W_n = 0] = 1$. Rothenberg (1973, p. 50) establishes a related result which shows that the asymptotic distribution of the inequality restricted estimator is the same as that for the unrestricted estimator when the true parameter value lies in the interior of the constraints set.

To establish part (2) we apply the standard results for linear inequality constraint tests to (A.4). The results from Wolak (1987) imply that the asymptotic distribution of W_n for this θ_0 is:

$$\text{(A.6)} \qquad \lim_{n\to\infty} \text{pr}[W_n \geqslant c] = \text{pr}[W \geqslant c] = \sum_{j=0}^{m} \text{pr}\left[\chi_j^2 \geqslant c\right] w(m, m-j, J_b(\theta_0)),$$

where $J_b(\theta_0)$ is defined in the statement of part (2). We require the condition that $H_b(\theta_0)$ is full row rank to guarantee that $J_b(\theta_0)$ is positive definite. This condition is required to compute weights entering the null asymptotic distribution using the functions given in Kudo (1963). By the continuity of $H(\theta)$, values of θ_0 where $H_b(\theta_0)$ is less than full row rank can be eliminated from consideration as the least favorable parameter value.

To establish part (3) we note that for values of $\theta_0 \in C$ such that $h_j(\theta_0) = 0$ for only a single $j \in P$, by the logic of (A.1)–(A.4) we know that the distribution of W_n asymptotically reduces to the distribution associated with a univariate inequality constraints test. This distribution is $\frac{1}{2}\text{pr}[\chi_1^2 \geqslant c]$ for $c > 0$. From equation (3.1) we can see that this distribution is also the lower bound on the exact null asymptotic distribution of a NIC test of arbitrary dimension, which establishes part (3).

Proof of Equation (2.15): Specializing $h(\theta)$ to the case of test (2.13) we obtain $h(\theta) = \theta^V - \theta$. For the case that $\rho < 0.95$, the correlation matrix of $V_b(\theta_0)$ is the (2×2) submatrix of $c(I(\theta)^{-1})$:

$$\text{(A.7)} \qquad \begin{bmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{bmatrix}.$$

Let V^B denote (A.7) evaluated at the point θ^B. Using the closed form solutions for the weights in Wolak (1987) gives $w(2,2,V^B) = \frac{1}{4}$, $w(2,1,V^B) = \frac{1}{2}$, and $w(2,0,V^B) = \frac{1}{4}$. Applying (A.6) with these weights gives (2.15), because $\text{pr}[\chi_0^2 \geqslant c] = 0$ for $c > 0$.

REFERENCES

Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Gourieroux, C., A. Holly, and A. Monfort (1980): "Kuhn-Tucker, Likelihood Ratio and Wald Tests for Nonlinear Models with Inequality Constraints on the Parameters," Harvard Institute of Economic Research Discussion Paper No. 770, Department of Economics, Harvard University, Cambridge, MA.

——— (1982): "Likelihood Ratio, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, 50, 63–80.

KODDE, D. A., AND F. C. PALM (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 54, 1243–1248.

KUDO, A. (1963): "A Multivariate Analogue of a One-Sided Test," *Biometrika*, 50, 403–418.

LEHMANN, E. L. (1983): *Theory of Point Estimation*. New York: John Wiley & Sons, Inc.

——— (1986): *Testing Statistical Hypotheses* (2nd Edition). New York: John Wiley & Sons, Inc.

PERLMAN, M. D. (1969): "One-Sided Testing Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, 40, 549–567.

ROTHENBERG, T. J. (1973): *Efficient Estimation with A Priori Information*. New Haven: Yale University Press.

WOLAK, F. A. (1987): "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model," *Journal of the American Statistical Association*, 82, 782–793.

——— (1988): "Duality in Testing Multivariate Hypotheses," *Biometrika*, 75, 611–615.

——— (1989): "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models," *Econometric Theory*, 5, 1–35.