Econometric Theory, **5**, 1989, 1–35. Printed in the United States of America.

LOCAL AND GLOBAL TESTING OF LINEAR AND NONLINEAR INEQUALITY CONSTRAINTS IN NONLINEAR ECONOMETRIC MODELS

FRANK A. WOLAK
Stanford University

This paper considers a general nonlinear econometric model framework that contains a large class of estimators defined as solutions to optimization problems. For this framework we derive several asymptotically equivalent forms of a test statistic for the local (in a way made precise in the paper) multivariate nonlinear inequality constraints test H: $h(\beta) \geq 0$ versus K: $\beta \in R^K$. We extend these results to consider local hypotheses tests of the form H: $h_1(\beta) \geq 0$ and $h_2(\beta) = 0$ versus K: $\beta \in R^K$. For each test we derive the asymptotic distribution for any size test as a weighted sum of χ^2-distributions. We contrast local as opposed to global inequality constraints testing and give conditions on the model and constraints when each is possible. This paper also extends the well-known duality results in testing multivariate equality constraints to the case of nonlinear multivariate inequality constraints and combinations of nonlinear inequality and equality constraints.

1. INTRODUCTION

This paper develops three local (in a way to be made precise) asymptotic tests for a set of nonlinear inequality restrictions on the parameters of nonlinear econometric models from the general class of models considered by Burguete, Gallant, and Souza [10] and Gallant [19], henceforth abbreviated as the BGS class of estimators. Models contained in this class are all of the least mean distance estimators and method of moments estimators. See Gallant [19, chap. 3] for a listing of all of the estimators in this class. The results are extended to devising local large-sample tests for combinations of multivariate nonlinear inequality and equality constraints. For the sake of expositional ease and brevity, we present our results for one member of this class: the maximum likelihood (ML) model. Modifications necessary for these procedures to apply to the BGS class of models are stated later in the paper.

I would like to thank the referees and associate editor for helpful comments on an earlier draft. Peter C.B. Phillips provided very useful advice on content and style, but he is not to blame for any infelicities that remain. This paper has also benefited from the comments of econometrics seminar participants at Berkeley and Stanford. This research was partially supported by NSF grant SES-84-20262 to the Department of Economics at Harvard University.

 0266-4666/89 $5.00 + .00

This section continues with a summary of the nonlinear inequality constraints testing framework and includes a short discussion of the local nature of these hypothesis tests. Next we chronicle the history of this type of work in mathematical statistics. A discussion of recent related work in the theoretical econometrics literature follows. Next we describe potential uses of this hypothesis testing framework in current applied econometric research. Finally, the remainder of the paper is outlined.

For β, the parameter vector from a model in the BGS class of estimators, we would like to perform the hypothesis test

$$H\colon h(\beta) \geq 0 \text{ versus } K\colon \beta \in R^K. \tag{1.1}$$

An asymptotically exact size test of the null hypothesis that $\beta \in C \equiv \{x \mid h(x) \geq 0, x \in R^K\}$ is not in general possible for reasons discussed later in the paper. An asymptotically exact test for general nonlinear inequality constraints is the local test:

$$H\colon h(\beta) \geq 0,\ \beta \in N_{\delta_n}(\beta^0) \text{ versus } K\colon \beta \in R^K, \text{ for all } n, \tag{1.2}$$

where $N_{\delta_n}(\beta^0)$ is a δ_n-neighborhood of β^0, $h(\beta^0) = 0$, $\delta_n = O(n^{-1/2})$, and n indexes the sample size. Asymptotically, Eq. (1.2) reduces to a test of whether or not β, the mean of a $N(\beta, H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)')$ random variable, is contained in the *cone of tangents of C at* β^0, where C is the set defined above. (The appendix contains the definition of the *cone of tangents of S at* x_0 for any arbitrary set S.) In contrast, despite the nonlinearity of the model in the parameters β, we can perform an asymptotically exact test of the form:

$$H\colon \beta \geq 0 \text{ versus } K\colon \beta \in R^K. \tag{1.3}$$

This follows because the *cone of tangents to C at* β^0 for this problem is the positive orthant: the set defining the null hypothesis. More importantly, $\beta^0 = 0$ is the unique value for the entire vector β which satisfies all of the inequalities as equalities. In general, a global inequality constraints test is possible only for testing the entire parameter vector for as many linear inequalities as there are elements of the vector, because the values assumed for the parameters not being tested will in general affect the distribution of the inequality constraints test statistic. A detailed discussion of these issues is given in Section 4.

The existence of large sample test results for Eq. (1.3) implies that the complications which arise in deriving a framework for testing nonlinear inequality constraints are primarily caused by the nonlinearity of the parameters in the inequality constraints. Although nonlinearity in the model alone is also problematic, the greater difficulties caused by nonlinear constraints differ from the case of testing multivariate equality constraints where both nonlinearity of the model in the parameters or constraints only allows the computation of local power functions because of the degenerate nonnull distribution for fixed alternative hypotheses. In this vein, Stroud [38] discusses the general lack of a large-sample approximation to the power function for fixed alternative hypotheses for nonlinear multivariate equality constraints tests. Stroud [37] presents general conditions under which the null and local nonnull distributions of the multivariate equality constraints test statistic exist for a large class of asymptotically normal estimators.

Our three test statistics resemble the likelihood ratio, Wald and Lagrange multiplier test statistics for multivariate equality constraints given in Burguete, Gallant, and Souza [10] and Gallant [19]. The current paper can be thought of as an extension of this work and the work of Wald [40], Aitchison and Silvey [1], and Silvey [35] to the case of multivariate nonlinear inequality constraints and combinations of nonlinear equality and inequality constraints. Recent work by Kodde and Palm [26] presents a distance test approach to testing multivariate inequality constraints and combinations of multivariate inequality and equality constraints. The present work, derived independently of theirs, integrates the distance test approach with a likelihood ratio-based approach to testing inequality constraints. Wolak [42] points out several complications that arise because their framework does not explicitly take into account the local nature of the nonlinear inequality constraints testing problem. The present paper focuses on precisely this issue and discusses the severe limitations of global inequality constraints tests. This paper rigorously illustrates what is meant by a local nonlinear inequality constraints test. It also states the exact distribution hypothesis test which is asymptotically equivalent to each of the local hypothesis tests involving inequality constraints presented here. Finally, this paper extends the classical large-sample duality result in testing multivariate equality constraints to the nonlinear inequality constraints and combinations of nonlinear equality and inequality constraints testing frameworks.

Although testing nonlinear inequality constraints has not been explicitly dealt with in the statistics literature, work related to this problem has been ongoing for some time. Chernoff [11] examined the asymptotic distribution of the likelihood ratio statistic when the true value of the parameter (θ^0) is a boundary point of both the set defining the null hypothesis (ω_1) and the set defining the alternative hypothesis (ω_2). In Chernoff's framework, the sets defining the null and alternative hypotheses need not be hyperplanes as is the case in standard equality constraints hypothesis testing problems. Feder [15] generalized Chernoff's results to the case where the true value of the parameter (θ_n^0) is near the boundaries of the sets ω_1 and ω_2 in the sense that $\theta_n^0 = \theta^0 + o(1)$, where $\theta^0 \in \bar{\omega}_1 \cap \bar{\omega}_2$ and $\bar{\omega}_1$ denotes the closure of ω_1. Both Chernoff's results and Feder's results for $d(\theta_n^0, \omega_i) = O(n^{-1/2})$, $i = 1,2$ (where $d(\theta, \omega)$ is the Euclidian distance from the point θ to the set ω) are utilized in the derivation of our results.

The multivariate one-sided hypothesis testing literature is related to the work presented here. This literature is concerned with testing H: $\mu = 0$ versus K: $\mu \geq 0$, where μ is the mean of a multivariate normal random vector with a known covariance matrix. Bartholomew [6,7,8] considered a related problem: testing homogeneity of independent normal means versus ordered alternative hypotheses concerning these means. Kudo [27] first considered this multivariate analogue of a one-sided test. Perlman [32] generalized these results to the case of testing H: $\mu \in P_1$ versus K: $\mu \in P_2$ where P_1 and P_2 are positively homogenous sets with $P_1 \subset P_2$. A special case of Perlman's framework is the hypothesis test H: $\mu \in \Delta$ versus K: $\mu \in R^K$, where Δ is a closed, convex cone in R^K. Under certain conditions linear inequality constraints define closed, convex cones in R^K; so that his framework is particularly useful to our purpose.

Recently, econometricians have become interested in the multivariate one-sided test. Gourieroux, Holly, and Monfort [22], hereafter referred to as GHM, have extended the multivariate one-sided hypothesis test to linear econometric models. The same authors [21] also considered the test for the case of nonlinear models. Farebrother [14] derived exact distribution results for the standard linear regression model for combinations of multivariate one-sided and two-sided hypotheses on the elements of the coefficient vector. Rogers [33] took an alternative approach, not based on the likelihood ratio principle to examine multivariate one-sided hypotheses in the ML model. He calls his approach the modified Lagrange multiplier test. Dufour [13] considers tests for these kinds of hypotheses on the coefficients of the linear regression model and derives bounds on the exact null distribution of the test statistics.

Due to the widespread use of inequality constrained estimation in econometrics, there are many possible applications of an inequality constraints testing procedure. Estimation under inequality restrictions in nonlinear models has become especially prevalent in the analysis of producer and consumer behavior. Lau [29] first discussed estimation under inequality restrictions as a way to impose monotonicity, convexity, and quasi-convexity constraints on econometrically estimated production, profit, and utility functions. Jorgenson, Lau, and Stoker [25] imposed inequality restrictions on the parameters of their model of consumer behavior to ensure that the individual indirect utility function is globally quasi-convex in the prices. Gallant and Golub [18] utilized this estimation procedure to impose the curvature restrictions implied by economic theory on the flexible functional forms used in production and demand analysis. Barnett [4], Barnett and Lee [5], Diewert and Wales [12], and Gallant [16,17] either proposed methods to estimate or estimated globally regular flexible functional forms by imposing inequality constraints on the estimated parameters of their econometric models. The widespread use of flexible functional forms in applied econometric work shows a clear need for tests of these hypotheses. These tests provide a way to empirically verify that the parameters of an econometric model satisfy the restrictions implied by economic theory. Possible applications of this testing procedure arise, specifically, whenever a flexible functional form is used in demand or production analysis, or in general, whenever a researcher estimates a statistical model and wants to test the empirical validity of *a priori* knowledge about the signs of two or more functions of the parameters of the model.

An outline of the remainder of this paper follows. Section 2 introduces the unconstrained, inequality constrained, and equality constrained estimators for the ML model framework. For continuity with previous work, our estimation framework follows that given in GHM [21]. Section 3 contains the derivation of the Kuhn-Tucker, Wald, and likelihood ratio statistics and gives conditions under which they are locally asymptotically equivalent. Section 4 shows that the asymptotic distribution of the test statistics for the purposes of testing the null hypothesis is a weighted sum of chi-squared distributions. In the appendix we discuss the proof of a global monotonicity property of the asymptotic power function from the inequality constraints test. In this section we analyze the impact of this monotonicity property on inequality constraints testing in general. As shown in Wolak [42], straightforward application of the technique used to show this monotonicity property in linear models with linear inequality constraints is not possible. The major implication of this discussion is general conditions for global instead of local inequality constraints tests. Section 4 also illustrates the asymptotic duality relation between the multivariate inequality constraints test and the multivariate one-sided test in terms of the vector of dual variables associated with the vector of nonlinear constraints. Section 5 extends our results to testing nonlinear equality and inequality restrictions jointly. If there are no inequality restrictions, this framework reduces to the standard ML-based framework for testing equality constraints. Section 6 extends these test procedures to the BGS model framework. In Section 7 we discuss the computation of critical values and probability values for the various hypothesis tests. This section also states upper and lower bounds on the null asymptotic distribution of the test statistics for hypotheses involving inequality constraints. In Section 8 we contrast our testing framework with the GHM [21] hypothesis testing framework. There we point out the local nature of their nonlinear multivariate one-sided test and extend their results to consider a local combination multivariate one-sided and two-sided hypothesis test.

2. NOTATION AND PRESENTATION OF THREE ESTIMATORS

For the sake of brevity and clarity, we first present our results for the well-known ML model framework. The notational burden necessary for the general BGS class of models is considerable while no special complications arise that are not present in the ML model. The initial use of the ML model framework allows us to simplify the exposition and focus on the primary purpose of the paper while preserving the essential complexities of testing nonlinear inequality constraints in nonlinear models. In addition, the ML model framework is used by GHM [21] and some of the results presented in this section and in Section 3 were derived by them. So that further justification for the use of this ML framework is to take maximum advantage of their work.

Before proceeding with the definition of our three estimators, we lay out the necessary notation. Denote m-dimensional Euclidian space by R^m. Let $X_i = (X_{i1}, X_{12}, \ldots, X_{im})'$ be an observation from a random vector in R^m with a probability density function $f(x_i, \beta)$, where $\beta = (\beta_i, \beta_2, \ldots, \beta_K)'$, a point in R^K, represents the unknown parameter vector and the function $f(x_i, \beta)$ is continuous in β for all x_i. The parameter space which contains β is Θ, a compact subset of R^K. The Appendix contains a full listing of additional regularity conditions necessary for the validity of our results.

The nonlinear constraints are represented by a set of continuous, differentiable functions h: $R^K \rightarrow R^P$ $(P \leq K)$, defined by $h(\beta) = (h_1(\beta), h_2(\beta), \ldots, h_P(\beta))'$, $\beta \in \Theta$. The partial derivatives, $\partial h_i(\beta)/\partial\beta_j$ $(i = 1, \ldots, P)$ $(j = 1, \ldots, K)$, exist and are continuous for all $\beta \in \Theta$. Denote by $H(\beta)$ the $(P \times K)$ matrix of partial derivatives whose (i,j)th element is $\partial h_i(\beta)/\partial\beta_j$. There exists a value of $\beta, \beta^0 = (\beta_1^0, \beta_2^0, \ldots, \beta_K^0)'$, in the interior of the Θ such that $h(\beta^0) = 0$. To avoid degeneracies in the null asymptotic distribution, we assume $H(\beta^0)$ has full row rank P. Under our local null hypothesis, β_n^0, the true value of β, satisfies:

$$\beta_n^0 \in K_n \equiv \{x \mid h(x) \geq 0,\ x \in N_{\delta_n}(\beta^0)\} \text{ for all } n.$$

Note that $(\beta_n^0 - \beta^0) = o(1)$ and $n^{1/2}(\beta_n^0 - \beta^0) = O(1)$.

A point in R^{nm} denoted by $x = (x_1, x_2, \ldots, x_n)'$ represents a set of n independent and identically distributed observations of X_i from the density function given above. The log-likelihood function L on $R^{nm} \times R^K$ is:

$$L(\beta) = L(x, \beta) = \sum_{i=1}^{n} \ln(f(x_i, \beta)). \tag{2.1}$$

For notational ease we suppress x from $L(x, \beta)$, although the dependence of $L(\beta)$ on x and n is clear.

Each estimate of β_n^0 chooses β to maximize Eq. (2.1) subject to β remaining in some compact set. Because this paper is primarily concerned with testing inequality constraints we will not discuss the computation of the various estimates of β_n^0 discussed below, only their existence. The inequality constrained ML estimate of β_n^0, which we denote by $\tilde{\beta}$, is the solution to:

$$\min_{\beta} \; -n^{-1}L(\beta) \qquad \text{subject to } h(\beta) \geq 0, \qquad \beta \in \Theta. \tag{2.2}$$

Associated with the nonlinear constraint vector is a set of Kuhn-Tucker multipliers, $\tilde{\lambda}$. The Kuhn-Tucker theorem asserts that none of the components of $\tilde{\lambda}$ are negative. The selection of this form for the optimization problem defining the ML estimates (minimizing the negative of the log-likelihood function) considerably simplifies the derivation of the null asymptotic distribution of our test statistics. The first-order conditions for this optimization problem are

$$-n^{-1}\frac{\partial L}{\partial \beta}(\tilde{\beta}) = H(\tilde{\beta})'\tilde{\lambda}, \qquad h_j(\tilde{\beta})\tilde{\lambda}_j = 0\ (j = 1,\ldots,P), \qquad \tilde{\lambda} \geq 0,$$

$$h(\tilde{\beta}) \geq 0. \tag{2.3}$$

The constraint qualification condition stated in the Appendix ensures that if $\tilde{\beta}$ solves (2.2) there exists a $\tilde{\lambda}$ satisfying (2.3). See Bazaraa and Shetty [9, chap. 4] for more on this topic.

The equality constrained maximum likelihood estimate, $\bar{\beta}$, is the solution to:

$$\min_{\beta} \; -n^{-1}L(\beta) \qquad \text{subject to } h(\beta) = 0, \qquad \beta \in \Theta. \tag{2.4}$$

Let $\bar{\lambda}$ denote the vector of Lagrange multipliers associated with the nonlinear equality constraints. The elements of this vector are unrestricted in sign. The first-order conditions for the equality constrained estimator are:

$$-n^{-1}\frac{\partial L}{\partial \beta}(\bar{\beta}) = H(\bar{\beta})'\bar{\lambda}, \qquad h(\bar{\beta}) = 0. \tag{2.5}$$

Finally, the unconstrained maximum likelihood estimator $\hat{\beta}$ is the solution to:

$$\min_{\beta} \; -n^{-1}L(\beta) \qquad \text{subject to } \beta \in \Theta. \tag{2.6}$$

For completeness, we associate a vector of Lagrange multipliers, $\hat{\lambda}$, with this estimate of β. This multiplier vector is equal to zero by definition of the unconstrained ML estimator. Assuming an interior solution, the first-order conditions for this problem are:

$$-n^{-1}\frac{\partial L}{\partial \beta}(\hat{\beta}) = 0. \tag{2.7}$$

Gill, Murray, and Wright [20] present a complete discussion of algorithms which can be used to solve optimization problems in (2.2), (2.4) and (2.6). They also discuss the relative merits of each technique for the various problems.

Several relationships between the various estimators of β_n^0 are useful for proving the local asymptotic equivalence of our test statistics and deriving their asymptotic distribution. Following the logic given in GHM [21], each of the three estimates of β_n^0 satisfies the following equation in β:

$$-n^{-1/2}\frac{\partial L}{\partial \beta}(\beta^0) + \boldsymbol{I}(\beta^0)\,[n^{1/2}(\beta - \beta^0)] \cong n^{1/2}H(\beta)'\lambda, \tag{2.8}$$

where $I(\beta^0)$ is Fisher's information matrix $(\lim_{n\to\infty} n^{-1}E_{\beta^0}[-\partial^2 L(\beta)/\partial\beta\partial\beta'])$ evaluated at $\beta = \beta^0$ and λ is the multiplier vector associated with that estimate of β_n^0. The symbol $\cong$ means that the difference between both sides of it converges in probability to zero as $n \to \infty$. For $\hat{\beta}$ and $\tilde{\beta}$ this equation implies:

$$I(\beta^0)[n^{1/2}(\tilde{\beta} - \hat{\beta})] \cong n^{1/2}H(\tilde{\beta})'\tilde{\lambda}. \quad \textbf{(2.9)}$$

This relationship is useful for relating $\tilde{\beta}$ to $\tilde{\lambda}$.

The framework derived in Aitchison and Silvey [1] and Silvey [36] implies:

$$h(\beta) \cong H(\beta^0)(\beta - \beta^0) \quad \text{and} \quad H(\beta) - H(\beta^0) \cong 0, \quad \textbf{(2.10)}$$

where β can be any one of the three estimates of β_n^0. The relations in (2.9) and (2.10) for $\hat{\beta}$ and $\bar{\beta}$ imply:

$$-n^{1/2}h(\hat{\beta}) \cong n^{1/2}H(\beta^0)I(\beta^0)^{-1}H(\beta^0)'\bar{\lambda}. \quad \textbf{(2.11)}$$

This equation is useful for relating $\bar{\lambda}$, the unrestricted estimate of λ, to $\hat{\beta}$, the unrestricted estimate of β_n^0.

3. THE THREE ASYMPTOTICALLY EQUIVALENT TEST STATISTICS

In this section we derive three locally asymptotically equivalent (for all $\beta_n^0 \in N_{\delta_n}(\beta^0)$) likelihood ratio-based statistics to test multivariate nonlinear inequality constraints. We prove that the likelihood ratio (LR) form of the test statistic is locally asymptotically equivalent to a generalized distance test statistic similar to that derived in Kodde and Palm [26]. This equivalence is useful for deriving the asymptotic distribution of the test statistics presented in this section. Proof of the asymptotic equivalence of the LR form of the inequality constraints statistic to the Wald and Kuhn-Tucker forms is not presented here because it parallels the proof given in GHM [21] of the asymptotic equivalence of their three analogously defined nonlinear multivariate one-sided test statistics. Their work is applicable to proving these results because of the asymptotic equivalence, shown in Section 4, between the local multivariate inequality constraints test and a multivariate one-sided test in terms of the vector of dual variables associated with the constraint vector.

Under the regularity conditions in the Appendix, $\hat{\beta}$ and $\tilde{\beta}$ are strongly consistent estimates of β_n^0. This implies the following relationship for these two estimates of β_n^0:

$$L(\beta) \cong L(\beta^0) + \left[n^{-1/2}\frac{\partial L}{\partial \beta}(\beta^0)\right]'[n^{1/2}(\beta - \beta^0)] - \frac{n}{2}(\beta - \beta^0)'I(\beta^0)(\beta - \beta^0). \quad \textbf{(3.1)}$$

This equation also holds for all $\beta \in N_{\delta_n}(\beta^0)$.

The likelihood ratio statistic takes the usual form:

$$\mathrm{LR} = -2[L(\tilde{\beta}) - L(\hat{\beta})] = 2[L(\hat{\beta}) - L(\tilde{\beta})]. \tag{3.2}$$

It also arises from the mathematical programming problem

$$\mathrm{LR} = \min_{\beta} 2[L(\hat{\beta}) - L(\beta)] \qquad \text{subject to } h(\beta) \geq 0. \tag{3.3}$$

By Equations (3.1), (3.3), and (2.10), the LR statistic is equivalent to, for large n, the optimal value of the objective function from the following quadratic program (QP):

$$\begin{aligned} \mathrm{LR} \cong \min_{\beta} 2\Big[& n^{-1/2} \frac{\partial L}{\partial \beta} (\beta^0)'[n^{1/2}(\hat{\beta} - \beta^0)] - \frac{n}{2} [(\hat{\beta} - \beta^0)' \boldsymbol{I}(\beta^0)(\hat{\beta} - \beta^0)] \\ & - n^{-1/2} \frac{\partial L}{\partial \beta} (\beta^0)'[n^{1/2}(\beta - \beta^0)] + \frac{n}{2} (\beta - \beta^0)' \boldsymbol{I}(\beta^0)(\beta - \beta^0) \Big] \\ & \text{subject to } \beta \in N_{\delta_n}(\beta^0) \text{ and } H(\beta^0)(\beta - \beta^0) \geq 0. \end{aligned} \tag{3.4}$$

This QP can be simplified to one similar to the Kodde and Palm [26] distance test statistic as follows. Taking the transpose of Equation (2.8) and post multiplying both sides of the equality by $n^{1/2}(\hat{\beta} - \beta^0)$, we find that $\hat{\beta}$ satisfies:

$$\left[n^{-1/2} \frac{\partial L}{\partial \beta} (\beta^0) \right]' n^{1/2}(\hat{\beta} - \beta^0) \cong n(\hat{\beta} - \beta^0)' \boldsymbol{I}(\beta^0)(\hat{\beta} - \beta^0). \tag{3.5}$$

Using (3.5), rewrite the objective function of (3.4) in an asymptotically equivalent form as:

$$\begin{aligned} \min_{\beta} n(\hat{\beta} - \beta^0)' \boldsymbol{I}(\beta^0)(\hat{\beta} - \beta^0) - 2n^{-1/2} \frac{\partial L}{\partial \beta} (\beta^0)'[n^{1/2}(\beta - \beta^0)] \\ + n(\beta - \beta^0)' \boldsymbol{I}(\beta^0)(\beta - \beta^0). \end{aligned} \tag{3.6}$$

This objective function simplifies to:

$$\min_{\beta} n(\hat{\beta} - \beta)' \boldsymbol{I}(\beta^0)(\hat{\beta} - \beta). \tag{3.7}$$

To see this, expand (3.7); add and subtract $n\beta^{0\prime}\boldsymbol{I}(\beta^0)\beta^0$ and $2n\hat{\beta}'\boldsymbol{I}(\beta^0)\beta^0$ from it to obtain:

$$\begin{aligned} \min_{\beta} n(\hat{\beta} - \beta^0)' \boldsymbol{I}(\beta^0)(\hat{\beta} - \beta^0) - 2n\hat{\beta}' \boldsymbol{I}(\beta^0)(\beta - \beta^0) + n\beta' \boldsymbol{I}(\beta^0)\beta \\ - n\beta^{0\prime} \boldsymbol{I}(\beta^0)\beta^0. \end{aligned} \tag{3.8}$$

Applying equation (2.8) for $\hat{\beta}$, simplifying and collecting terms, gives the objective function (3.6). Thus we have:

$$\mathrm{LR} \cong D = \min_{\beta} n(\hat{\beta} - \beta)'\boldsymbol{I}(\beta^0)(\hat{\beta} - \beta)$$

$$\text{subject to } \beta \in N_{\delta_n}(\beta^0) \quad \text{and} \quad H(\beta^0)(\beta - \beta^0) \geq 0. \qquad \textbf{(3.9)}$$

Hence, the LR statistic is asymptotically equivalent to the generalized distance statistic, D.

There are three other asymptotically equivalent forms for the inequality constraints test statistic. First is the Wald statistic which measures the magnitude of the difference between the restricted and unrestricted estimates of β_n^0 in the norm of the asymptotic covariance matrix of $n^{1/2}(\hat{\beta} - \beta_n^0)$:

$$\mathrm{W} = n(\tilde{\beta} - \hat{\beta})'\boldsymbol{I}(\beta^0)(\tilde{\beta} - \hat{\beta}). \qquad \textbf{(3.10)}$$

The Kuhn-Tucker statistic measures the magnitude of the Kuhn-Tucker multiplier vector arising from the inequality constrained estimation procedure:

$$\mathrm{KT} = n\tilde{\lambda}'H(\tilde{\beta})\boldsymbol{I}(\beta^0)^{-1}H(\tilde{\beta})'\tilde{\lambda}. \qquad \textbf{(3.11)}$$

By (2.9) the KT statistic is asymptotically equivalent to the W statistic. An asymptotically equivalent way to that given in (3.10) for writing the Wald statistic is:

$$\mathrm{W}^* = n(h(\tilde{\beta}) - h(\hat{\beta}))'[H(\hat{\beta})\boldsymbol{I}(\beta^0)^{-1}H(\hat{\beta})']^{-1}(h(\tilde{\beta}) - h(\hat{\beta})). \qquad \textbf{(3.12)}$$

Because the difference between all of these statistics converges in probability to zero as $n \to \infty$, they all possess the same asymptotic distribution.

4. ASYMPTOTIC NULL DISTRIBUTION OF STATISTICS

Our hypothesis testing problem does not fit into the standard hypothesis testing framework because our composite null hypothesis does not specify a unique value for $h(\beta_n^0)$. Our problem only requires $h(\beta_n^0)$ to lie in the positive orthant of P-dimensional space. In contrast, for an equality constraints test, under the null hypothesis β_n^0 must satisfy $h(\beta_n^0) = 0$. As a consequence, a least favorable value of $\beta_n^0 \in K_n$ must be found to construct an asymptotically exact size test of the inequality constraints.

As mentioned in the introduction, we would prefer an asymptotically exact test of the null hypothesis $\beta \in C \equiv \{x | h(x) \geq 0, x \in R^K\}$ with our test statistics. Wolak [42] showed this is impossible because of the general indeterminacy of the least favorable value of $\beta \in C$. The global monotonicity property of the power function of the test is only able to limit the least favorable value of β to the set $C^E \equiv \{x | h(x) = 0, x \in R^K\}$. Unless this set contains one element, there will be as many null asymptotic distributions as there are elements of C^E. This occurs because the matrix $H(\beta)$, which the asymptotic distribution of the test statistics is functionally dependent on, varies with β if the constraint vector $h(\beta)$ is at all nonlinear. There is no way to select among these values of β to find the least favorable value besides complete enumeration, which is impossible if the set C^E is uncountable. For a given $\beta \in C^E$, the asymptotic null distribution that obtains depends on the geometry of the *cone of tangents to the set C at* β through the matrix $H(\beta)$. Depending on what $\beta \in C^E$ we assume for the true value of β, a different asymptotic distribution will obtain because $H(\beta)$ should vary with β for nonlinear constraints. Consequently, to obtain a determinant exact null asymptotic distribution, we must settle for a local nonlinear inequality constraints test. Our hypothesis test is still $h(\beta) \geq 0$ versus $\beta \in R^K$, but it is relative to the point β^0. Wolak [42] points out the local nature of hypothesis tests involving nonlinear inequality constraints and describes why these problems do not arise in the nonlinear equality constraints testing framework.

We now derive the asymptotic distribution of our three statistics for any size test of our null hypothesis. To illustrate the duality relation that exists for our testing framework we derive the null distribution in terms of both the primal and dual variables. First we deal with the primal approach which is in terms of the parameter vector β. As a starting point, consider the hypothesis testing problem:

$$H\colon \mu \geq 0 \quad \text{versus} \quad K\colon \mu \in R^P \quad \text{where } \hat{\mu} = \mu + \nu, \tag{4.1}$$

and ν is $N(0,\Xi)$, and Ξ is known and positive definite. Wolak [44], following Perlman [32], shows the likelihood ratio statistic for (4.1) is the optimal value of the objective function from the following QP:

$$Z = \min_{\mu} \; (\mu - \hat{\mu})'\Xi^{-1}(\mu - \hat{\mu}) \qquad \text{subject to } \mu \geq 0. \tag{4.2}$$

Let $\tilde{\mu}$ denote the solution to QP (4.2).

We must now choose a least favorable value of μ under the null hypothesis to construct an exact size test of this null hypothesis. The prescribed approach to this problem proceeds as follows. For test (4.1), our sample space in the Neyman-Pearson likelihood ratio hypothesis testing framework is $\Theta = R^P$. The positive orthant in P-dimensional space is the subset of Θ in which μ lies under the null hypothesis. We denote this by Θ_H. Following Lehmann [30], let s be the test statistic for our hypothesis test and S the rejection region. If

$$\sup_{\mu \in \Theta_H} pr_\mu(s \in S) = \alpha,$$

then S is the rejection region for a size α test of our null hypothesis.

By this logic we will construct a rejection region for a level α test of (4.1). A special case of Lemma 8.2 in [32] is given below.

LEMMA 4.1. *For any $\mu \geq 0$ and positive scalar c, the following is true:* $pr_{\mu,\Omega}[Z \geq c] \leq pr_{0,\Omega}[Z \geq c]$.

An immediate corollary is:

$$\sup_{\mu \in \Theta_H} pr_{\mu,\Omega}[Z \geq c] = pr_{0,\Omega}[Z \geq c].$$

This lemma provides a unique value for μ to specify under the null hypothesis for any size α test. The following theorem proved in [44] gives the null distribution for any size test.

THEOREM 4.1. *Under the hypothesis $\mu \geq 0$, the likelihood ratio statistic (4.2), which we denote by Z, has the following distribution:*

$$\sup_{\mu \geq 0} pr_{\mu,\Xi}(Z \geq c) = pr_{0,\Xi}(Z \geq c) = \sum_{k=0}^{P} pr(\chi_k^2 \geq c) w(P, P-k, \Xi).$$

This distribution is a weighted sum of chi-squared distributions. The weight, $w(P, P-k, \Xi)$, is the probability that $\tilde{\mu}$ has exactly $P-k$ positive elements. Wolak [44] provides a detailed discussion of the computation of these weights.

If we consider $h(\beta)$ as μ and $n^{-1}H(\beta^0)I(\beta^0)^{-1}H(\beta^0)'$ as Ξ in hypothesis testing problem (4.1), then our local inequality constraints hypothesis test

$$H\colon h(\beta) \geq 0, \qquad \beta \in N_{\delta_n}(\beta^0) \quad \text{versus} \quad K\colon \beta \in R^K \tag{4.3}$$

is asymptotically equivalent to testing problem (4.1). The logic for this claim proceeds as follows. We know $n^{1/2}h(\hat{\beta})$ converges in distribution to a $N(H(\beta^0)b,\ H(\beta^0)I(\beta^0)^{-1}H(\beta^0)')$ random variable for all $\beta_n^0 \in N_{\delta_n}(\beta^0)$, where $b = \lim_{n \to \infty} n^{1/2}(\beta_n^0 - \beta^0)$. Therefore, as $n \to \infty$, the following relationship holds by a large sample version of Lemma 4.1:

$$\sup_{b \in B} pr_{b,I(\beta^0)^{-1}}(D \geq c) = pr_{0,I(\beta^0)^{-1}}(D \geq c), \tag{4.4}$$

where $B = \{b \mid H(\beta^0)b \geq 0,\ b \in R^K\}$ and D is the asymptotic value of the three-test statistics. This relation implies that $\beta_n^0 = \beta^0$ (which occurs if $b = 0$) is the least favorable value of β_n^0 to select for an asymptotically exact size test of (4.3). The distributional results for hypothesis testing problem (4.1) derived in [44] and the above intuition yields the following result proved in the Appendix.

THEOREM 4.2. *For the local hypothesis testing problem H: $h(\beta) \geq 0$, $\beta \in N_{\delta_n}(\beta^0)$ versus K: $\beta \in R^K$, the asymptotic distribution of the KT, LR, and W statistics satisfies the following property:*

$$\sup_{b \in B} pr_{b,\boldsymbol{I}(\beta^0)^{-1}}(D \geq c) = pr_{\beta^0}(D \geq c) = \sum_{k=0}^{P} pr(\chi_k^2 \geq c)\, w(P, P-k, \Pi)$$

where D is the asymptotic value of the three statistics and $\Pi = [H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']$.

An intuitive justification for this result follows by noting that hypothesis test (4.3), as $n \to \infty$, is equivalent to the test

$$H\colon \beta \in T(\beta^0) \qquad \text{versus} \qquad K\colon \beta \in R^K \tag{4.5}$$

based on $\hat{b} = \beta + \nu$, where $\nu \sim N(0, H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)')$ and

$$T(\beta^0) \equiv \{x \mid H(\beta^0)(x - \beta^0) \geq 0,\ x \in R^K\},$$

is the *cone of tangents of C at* β^0. The vertex of this cone is β^0. So the results of Lemma 4.1 imply the least favorable value of β to choose for an exact size test of (4.5) is β^0, the value of β which satisfies the linear inequalities defining $T(\beta^0)$ as equalities. Figure 1 provides a graphic example of our nonlinear inequality constraints framework for $\beta = (\beta_1, \beta_2)'$ and $h(\beta) = (h_1(\beta), h_2(\beta))'$. The intersection of the two curves $h_1(\beta) = 0$ and $h_2(\beta) = 0$ is the point β^0. The two rays tangent to the curves at the point β^0, R_1 and R_2, describe the *cone of tangents to the set C at* β^0, called $T(\beta^0)$ in the figure. Mathematically, the equation for the rays R_1 and R_2 are the first and second rows of the matrix equation $H(\beta^0)(\beta - \beta^0) \geq 0$, respectively.

The null distribution for any size test can also be obtained by the dual approach to the testing problem. In terms of λ, the true value of the Lagrange multiplier vector arising from the equality constrained estimation problem, our hypothesis testing problem is $H\colon \lambda = 0$ versus $K\colon \lambda \geq 0$, with the inequality strict for at least one element of the λ under the alternative hypothesis. For every n, the true value of λ is defined as the value of the Lagrange multiplier vector arising from the solution to

$$\min_{\beta}\ -L_n(\beta, \beta_n^0) \qquad \text{subject to } h(\beta) = 0 \tag{4.6}$$

where $L_n(\beta, \beta_n^0) = E_{\beta_n^0}[\ln(f(X, \beta))]$. As stated in the Appendix, the unique unconstrained maximum of $L_n(\beta, \beta_n^0)$ occurs at $\beta = \beta_n^0$, the true value of β. If β_n^0 is not an element of C^E then λ will be nonzero. Given $\lambda = 0$, so that $\beta_n^0 \in C^E$, Aitchison and Silvey [1] and Silvey [36] show $\lim_{n\to\infty} n^{1/2}\bar{\lambda}$ converges in distribution to a $N(0, [H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']^{-1})$ random vector under the regularity conditions in the Appendix. These results provide an intuitive justification for choosing the least favorable value of $\beta_n^0 \in K_n$ such that $\beta_n^0 = \beta^0$ in order to construct an asymptotically exact size test of our local inequality constraints test. We will find that, for the same size test, the local nonlinear inequality constraints test (4.3) and the hypothesis test $H\colon \lambda = 0$ versus $K\colon \lambda \geq 0$, for λ as defined in (4.6), are asymptotically equivalent. This

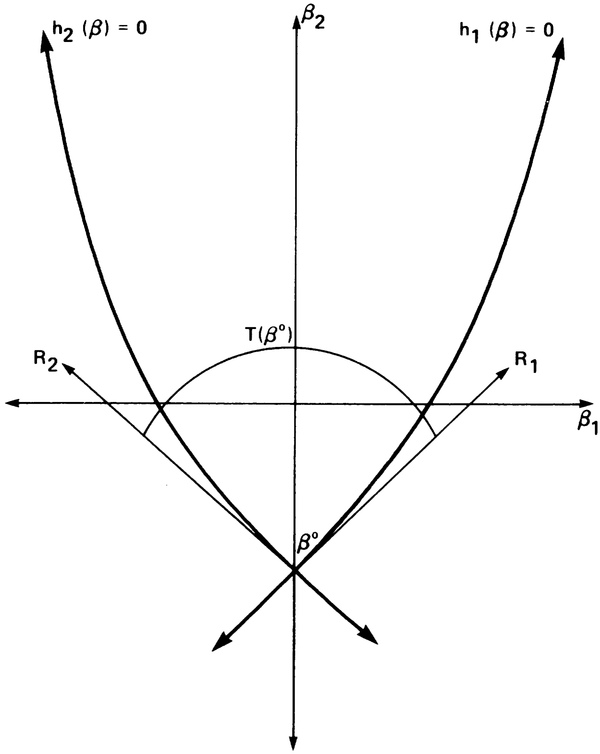

FIGURE 1. Graphical representation of the inequality constraints set $C = \{\beta \mid h(\beta) \geq 0, \beta \in R^2\}$ and its *cone of tangents* $T(\beta^0)$ *at* β^0.

result extends the large-sample duality relation between the hypothesis tests H: $h(\beta) = 0$ versus K: $h(\beta) \neq 0$ and H: $\lambda = 0$ versus K: $\lambda \neq 0$, to the case of multivariate inequality constraints.

To discuss the dual approach to our problem, we first consider the much studied multivariate one-sided hypothesis testing problem

$$H: \xi = 0 \qquad \text{versus} \qquad K: \xi \geq 0, \ \xi \in R^P,$$
$$\text{where } \bar{\xi} = \xi + \eta, \ \eta \text{ is } N(0,\Sigma), \tag{4.7}$$

and Σ is known and of full rank. Wolak [44] provides a detailed summary of this literature. The LR statistic for (4.7) is the optimal value of the objective function from:

$$\max_{\xi} \bar{\xi}'\Sigma^{-1}\bar{\xi} - (\bar{\xi} - \xi)'\Sigma^{-1}(\bar{\xi} - \xi) \qquad \text{subject to } \xi \geq 0. \tag{4.8}$$

Let $\tilde{\xi}$ denote the solution to QP (4.8), so that:

$$\text{LR} = \bar{\xi}'\Sigma^{-1}\bar{\xi} - (\bar{\xi} - \tilde{\xi})'\Sigma^{-1}(\bar{\xi} - \tilde{\xi}) = \tilde{\xi}'\Sigma^{-1}\tilde{\xi}. \tag{4.9}$$

The second equality follows from the complementary slackness conditions for QP (4.8). The null distribution of the LR statistic is given in the following theorem proved in Kudo [27], Nuesch [31], and Perlman [32].

THEOREM 4.3. *For the hypothesis testing problem (4.7), the LR statistic has the following distribution under the null hypothesis:*

$$pr_{0,\Sigma}(LR \geq c) = \sum_{k=0}^{P} pr(\chi_k^2 \geq c)\, w(P,k,\Sigma),$$

where $w(P,k,\Sigma)$ is the probability that $\tilde{\xi}$ has exactly k positive elements.

Note that the weights, $w(P,k,\Sigma)(k = 0,\ldots,P)$, have the same functional form as those in Theorem 4.1. Therefore the considerable literature on the computation of these weights in the multivariate one-sided hypothesis testing literature is available to apply to our inequality constraints testing problem.

We now apply these results to our testing problem. The duality theory of quadratic programming applied to QP (3.9) implies the LR statistic, and therefore the KT statistic, is asymptotically equivalent to the optimal objective function value of the following QP:

$$\text{KT} \cong \max_{\lambda} n[\lambda'(H(\beta^0)\beta^0 - H(\beta^0)\hat{\beta}) - \frac{1}{4}\lambda' H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'\lambda]$$

$$\text{subject to } \lambda \geq 0. \tag{4.10}$$

Equations (2.10) for $\hat{\beta}$ and Equation (2.11) imply optimization problem (4.10) is asymptotically equivalent to:

$$\text{KT} \cong \max_{\lambda} n[\lambda' H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'\bar{\lambda} - \frac{1}{4}\lambda'(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'\lambda]$$

$$\text{subject to } \lambda \geq 0. \tag{4.11}$$

Finally, we can show that the optimal value of the objective function from the following QP is equal to that same value from QP (4.11); so that

$$\text{KT} \cong \max_{\lambda^*} n[\bar{\lambda}' H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'\bar{\lambda}$$
$$- (\lambda^* - \bar{\lambda})' H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'(\lambda^* - \bar{\lambda})]$$
$$\text{subject to } \lambda^* \geq 0. \tag{4.12}$$

The value of λ^* which solves (4.12) is asymptotically equivalent to the Kuhn-Tucker multiplier from estimation procedure (2.2). The optimal value of the objective function from (4.12) is:

$$\text{KT}^* = n\tilde{\lambda}^{*\prime} H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'\tilde{\lambda}^*. \tag{4.13}$$

Because $\tilde{\lambda}^* \cong \tilde{\lambda}$, we have $\text{KT}^* \cong \text{KT} \cong \text{W} \cong \text{LR}$.

If we replace $\tilde{\xi}$ *by* $\bar{\lambda}$ and Σ by $[nH(\beta^0)I(\beta^0)^{-1}H(\beta^0)']^{-1}$ in QP (4.8) and recall that for β_n^0 such that $h(\beta_n^0) = 0, \bar{\lambda}$ is, for large samples, approximately $N(0, [nH(\beta^0)I(\beta^0)^{-1}H(\beta^0)']^{-1})$. Therefore, the asymptotic distribution of our three test statistics is equivalent to the null distribution from a multivariate one-sided hypothesis test performed on $\bar{\lambda}$. We summarize our results in the following theorem.

THEOREM 4.4. *For the dual form of hypothesis testing problem (4.3), which takes the form* H: $\lambda = 0$ *versus* K: $\lambda \geq 0$, *the null asymptotic distribution of the KT, W, and LR statistics is*:

$$pr_{\beta^0}(Z \geq c) = \sum_{k=0}^{P} pr(\chi_k^2 \geq c) w(P, k, \Lambda),$$

where Z is the asymptotic value of the three statistics, and $\Lambda = [H(\beta^0) I(\beta^0)^{-1}H(\beta^0)']^{-1}$.

In [44] the following relationship is shown to hold for the weight functions:

$$w(P, k, \Omega) = w(P, P - k, a\Omega^{-1}) \text{ for } k = 0, \ldots, P \text{ and } a > 0.$$

Given this result, comparing the null distribution derived in Theorem 4.2 with that derived in Theorem 4.4 yields the following.

COROLLARY 4.1. *For hypothesis tests of the same asymptotic size, the local inequality constraints test,* H: $h(\beta) \geq 0$, $\beta \in N_{\delta_n}(\beta^0)$ *versus* K: $\beta \in R^K$, *is equivalent to the multivariate one-sided test,* H: $\lambda = 0$ *versus* K: $\lambda \geq 0$ *for* λ *as defined by (4.6).*

Wolak [43] discusses this duality relationship from both the geometric and mathematical programming viewpoint and shows that it is a generalization of the classical duality result in multivariate equality constraints testing.

To concentrate on the local versus global inequality constraints distinction, we now turn to the hypothesis test H: $R\beta \geq r$ versus K: $\beta \in R^K$, where R is a $(K \times K)$ matrix of full rank, and r is $(K \times 1)$ vector. Here global hypothesis tests are possible for the reasons cited in Section 1 and because $\beta^0 = R^{-1}r$ is the unique least favorable value of β^0 that specifies all of the elements of β under the null hypothesis. Unfortunately, without further restrictions, all of these conditions are necessary for a global inequality constraints test in nonlinear models. The null distribution of an asymptotically exact size test is:

$$pr_{\beta^0}(D \geq c) = \sum_{k=0}^{K} pr(\chi_k^2 \geq c) w(K, K - k, RI(\beta^0)^{-1}R'),$$

where D is the asymptotic value of the inequality constraints test statistics. Consider the case that R has fewer than K rows but is still of full row rank. Because there is an uncountable number of β's that satisfy the equation

$R\beta = r$, the fact that $I(\beta)$ may vary with these values of β implies an indeterminacy in the null asymptotic distribution.

Given these results, we now state conditions under which global versus local inequality constraints tests are possible. A sufficient condition is that $H(\beta)I(\beta^0)^{-1}H(\beta)'$ remains constant as β varies over the set of β such that $h(\beta) = 0$. Conditions which guarantee this are $H(\beta) = R$, a matrix of constants, and $I(\beta^0) = \Sigma$, a matrix of constants. Wolak [44] deals with the case when these two conditions are satisfied. There a framework is presented for testing global linear inequality constraints in linear econometric models. The most general linear model considered is the linear simultaneous equations model. In the notation of Theil [39, pp. 439–441] this model takes the form $Y\Gamma + XB = E$. The framework in Wolak [44] allows global linear inequality constraints tests on the elements of Γ and B. See Hendry [23] for a discussion of the commonly used linear econometric models which are special cases of this general framework. Linearity in the parameters and the constraints is an easily verifiable sufficient condition for a global inequality constraints test.

The local nature of inequality constraints tests arises even when testing whether or not a subvector of β lies in the positive orthant. Let $\beta' = (\beta_1', \beta_2')$, where the hypothesis test is H: $\beta_1 \geq 0$ versus K: $\beta_1 \in R^P$. The derivation of the null distribution is complicated by the fact that unless the submatrix of $I(\beta^0)^{-1}$ corresponding to β_1 does not depend on β_2, the null distribution obtained depends on the values assumed for these nuisance parameters under the null hypothesis. Consequently if every value of β_2 implies a different submatrix of $I(\beta^0)^{-1}$, this β_2 will also imply a different null asymptotic distribution of the test statistics. There is no straightforward way to find the value of these parameters which yields the least favorable null distribution. In this case the hypothesis test must be performed local to the point $\beta_1 = 0$ and $\beta_2 = \beta_2^0$, where β_2^0 is the value assumed for β_2.

5. JOINT EQUALITY AND INEQUALITY CONSTRAINTS

In this section we consider local hypothesis tests of the form H: $h^1(\beta) \geq 0$ and $h^2(\beta) = 0$ versus K: $\beta \in R^K$, where $h^1(\beta)$ is a vector of the first L ($L \leq P$) elements of $h(\beta)$ and $h^2(\beta)$ is a vector of the remaining $P - L$ elements of $h(\beta)$. In this way we jointly test the validity of nonlinear multivariate inequality and equality constraints on β. We will derive three asymptotically equivalent forms of the test statistic used to examine these kinds of hypotheses.

Three estimates of β_n^0 are necessary to discuss this hypothesis testing problem. First is the mixed inequality-equality constrained estimate. It is the solution to

$$\min_{\beta} \; -n^{-1}L(\beta) \qquad \text{subject to } h^1(\beta) \geq 0, \quad h^2(\beta) = 0, \quad \beta \in \Theta. \tag{5.1}$$

The solution to (5.1) is $\dot{\beta}$. A vector of multipliers, $\dot{\lambda}' = (\dot{\lambda}_1', \dot{\lambda}_2')$, is associated with the solution. The first subvector, $\dot{\lambda}_1$, is associated with the inequality constraints and is therefore restricted to be greater than or equal to zero. The subvector $\dot{\lambda}_2$ is associated with the equality constraints so it is unrestricted in sign. The first-order conditions for this optimization problem are:

$$-n^{-1} \frac{\partial L}{\partial \beta}(\dot{\beta}) = H(\dot{\beta})'\dot{\lambda}, \quad \dot{\lambda}_1 \geq 0, \quad h^1(\dot{\beta}) \geq 0, \qquad h_j(\dot{\beta})\dot{\lambda}_j = 0$$
$$(j = 1, \ldots, L) \qquad h_k(\dot{\beta}) = 0 \quad (k = L+1, \ldots, P). \quad \mathbf{(5.2)}$$

The equality constrained version of this problem is, by construction, the equality constrained maximum likelihood problem (2.4). The unconstrained version of this problem is the unconstrained maximum likelihood problem (2.6). The regularity conditions in the appendix ensure $\dot{\beta}$ is a strongly consistent estimate of β_n^0.

The LR statistic is defined analogously to (3.2) as:

$$\mathrm{LR} = -2[L(\dot{\beta}) - L(\hat{\beta})] = 2[L(\hat{\beta}) - L(\dot{\beta})]. \quad \mathbf{(5.3)}$$

The Wald statistic is:

$$\mathrm{W} = n(\dot{\beta} - \hat{\beta})'\boldsymbol{I}(\beta^0)(\dot{\beta} - \hat{\beta}). \quad \mathbf{(5.4)}$$

The KT form of the mixed inequality-equality constraints test statistic is:

$$\mathrm{KT} = n\dot{\lambda}'H(\dot{\beta})\boldsymbol{I}(\beta^0)^{-1}H(\dot{\beta})'\dot{\lambda}. \quad \mathbf{(5.5)}$$

These three statistics are locally asymptotically equivalent (for all $\beta_n^0 \in N_{\delta_n}(\beta^0)$). By logic similar to that used to derive the Wald statistic in (3.12), another form of that statistic for this problem is:

$$\mathrm{W}^* = n(h(\dot{\beta}) - h(\hat{\beta}))'[H(\hat{\beta})\boldsymbol{I}(\beta^0)^{-1}H(\hat{\beta})']^{-1}(h(\dot{\beta}) - h(\hat{\beta})). \quad \mathbf{(5.6)}$$

By the logic of Equations (3.2) through (3.9), the LR statistic for the mixed inequality-equality constraints test statistic is asymptotically equivalent to the optimal objective function value from the following QP:

$$\mathrm{LR} \cong \min_{\beta} n(\hat{\beta} - \beta)'\boldsymbol{I}(\beta^0)\hat{\beta} - \beta)$$
$$\text{subject to } \beta \in N_{\delta_n}(\beta^0) \text{ and } H^1(\beta^0)(\beta - \beta^0) \geq 0,\ H^2(\beta^0)(\beta - \beta^0) = 0, \quad \mathbf{(5.7)}$$

where $H^1(\beta^0)$ and $H^2(\beta^0)$ are the matrices of partial derivatives of $h^1(\beta^0)$ and $h^2(\beta^0)$, respectively. These are defined in the same fashion as $H(\beta^0)$ is for $h(\beta^0)$. This statistic is analogous to the generalized distance test statistic for combinations of inequality and equality constraints discussed in Kodde and Palm [26].

To extend the asymptotic duality relation to combinations of multivariate inequality and equality constraints, the null asymptotic distribution for these statistics will be derived from both the dual and primal viewpoint. First we treat our testing problem in terms of the dual vector of constraint multipliers because the presentation for this approach is more straightforward given the results of Section 4. We will show that the local hypothesis testing problem H: $h^1(\beta) \geq 0$ and $h^2(\beta) = 0$, $\beta \in N_{\delta_n}(\beta^0)$ versus K: $\beta \in R^K$ is equivalent to H: $\lambda = 0$ versus K: $\lambda_1 \geq 0$ and $\lambda_1 \neq 0$, where λ_1 is associated with $h^1(\beta)$ and λ_2 is associated with $h^2(\beta)$.

Kudo [27] considered the hypothesis testing problem:

$$H: \psi = 0 \quad \text{versus} \quad K: \psi_i \geq 0 \quad (i = 1, \ldots, L \leq P), \quad \psi_i \neq 0 \quad (i = L + 1, \ldots, P), \quad \psi \in R^P, \tag{5.8}$$

where $\bar{\psi} = \psi + \omega$, ω is $N(0, \Psi)$, and Ψ is known and positive definite. The likelihood ratio statistic for this problem is:

$$M = \max_{\psi} \bar{\psi}'\Psi^{-1}\bar{\psi} - (\psi - \bar{\psi})'\Psi^{-1}(\psi - \bar{\psi}) \quad \text{subject to } \psi_i \geq 0 \ (i = 1, \ldots, L). \tag{5.9}$$

The null distribution for the M statistic stated below is given in Kudo [27].

LEMMA 5.1. *For the hypothesis testing problem (5.8) the LR test statistic (M) has the following null distribution:*

$$pr_{0,\Psi}(M \geq c) = \sum_{k=0}^{L} pr(\chi_{P-L+k} \geq c) w(L, k, \Gamma)$$

where Γ is the submatrix of Ψ corresponding to ψ_j $(j = 1, \ldots, L)$.

Given these results we derive the null distribution for our statistics. Recall that the optimal objective function value of QP (5.7) is asymptotically equivalent to the mixed constraint LR statistic. The dual of this quadratic programming problem is:

$$\text{KT} \cong \max_{\lambda} n[\lambda'(H(\beta^0)\beta^0 - H(\beta^0)\hat{\beta}) - \frac{1}{4} \lambda' H(\beta^0) I(\beta^0)^{-1} H(\beta^0)'\lambda] \quad \text{subject to } \lambda_i \geq 0 \ (i = 1, \ldots, L). \tag{5.10}$$

The $P - L$ elements of λ, λ_j $(j = 1 + L, \ldots, P)$ are unrestricted. The duality theory of quadratic programming implies that the optimal value of the objective function of (5.10) is equal that same value from (5.7).

The optimal value of the objective function of QP (5.10) is asymptotically equivalent to the same magnitude from the QP:

$$\max_{\lambda^*} n[\bar{\lambda}'H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'\bar{\lambda} - (\lambda^* - \bar{\lambda})'H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'(\lambda^* - \bar{\lambda})]$$
$$\text{subject to } \lambda_i^* \geq 0 \ (i = 1, \ldots, L). \qquad \textbf{(5.11)}$$

Similar to the case of only inequality constraints, the value of λ^* which satisfies (5.10) is asymptotically equivalent to the Kuhn-Tucker multiplier from the mixed inequality-equality constrained maximum likelihood estimation procedure, $\dot{\lambda}$. The optimal value of the objective function of (5.11) is:

$$\text{KT}^* = n\dot{\lambda}^{*\prime}H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'\dot{\lambda}^*. \qquad \textbf{(5.12)}$$

Because $\dot{\lambda}^* \cong \dot{\lambda}$, we know that $\text{KT}^* \cong \text{KT} \cong \text{W} \cong \text{LR}$ for the mixed constraints testing procedure.

Replacing $\bar{\psi}$ by $\bar{\lambda}$ and Ψ by $[nH(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']^{-1}$ in testing problem (5.8), and noting that $\lim_{n\to\infty} n^{1/2}\bar{\lambda}$ converges in distribution to a $N(0, [H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']^{-1})$ random vector given $\beta_n^0 = \beta^0$, the asymptotic distribution of our three test statistics is equivalent to the null distribution of the LR statistic for hypothesis test (5.8) performed on λ as defined in (4.6). This logic gives the following theorem.

THEOREM 5.1. *For the dual form of the local inequality-equality constraints hypothesis testing problem, H: $\lambda = 0$ versus K: $\lambda_i \geq 0$ $(i = 1, \ldots, L)$ and $\lambda_i \neq 0$ $(i = L + 1, \ldots, P)$, the null asymptotic distribution of the KT, W, and LR statistics is:*

$$pr_{\beta^0}(X \geq c) = \sum_{k=0}^{L} pr(\chi^2_{P-L+k} \geq c)\, w(L,k,\Gamma)$$

where Γ is the submatrix of $[H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']^{-1}$ corresponding to λ_j $(j = 1, \ldots, L)$ and X is the asymptotic value of any of the three statistics.

The null asymptotic distribution of the mixed constraints test statistics can be obtained in terms of the primal problem by extending Theorem 4.1 in a similar fashion to Kudo's extension of Theorem 4.3. By the same logic as for inequality constraints alone, as $n \to \infty$, the local inequality-equality constraints hypothesis test is equivalent to the test H: $\beta \in T^*(\beta^0)$ versus K: $\beta \in R^K$, based on $\hat{b} = \beta + \nu$ and $\nu \sim N(0, H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)')$, where

$$T^*(\beta^0) = \{x \mid H_1(\beta^0)(x - \beta^0) \geq 0,\ H_2(\beta^0)(x - \beta^0) = 0,\ x \in R^K\}$$

is the *cone of tangents to C^* at β^0* and $C^* = \{x \mid h^1(x) \geq 0,\ h^2(x) = 0,\ x \in R^K\}$. Because $T^*(\beta^0)$ is a convex cone and β^0 its vertex, the monotonicity property of the power function for this exact test implies the least favorable null value of β is β^0. In addition, similar to the local power function results of Section 4, we have:

$$\sup_{b \in B^*} pr_{b, I(\beta^0)^{-1}}(X \geq c) = pr_{0, I(\beta^0)^{-1}}(X \geq c),$$

where $B^* \equiv \{H^1(\beta^0)b \geq 0,\ H^2(\beta^0)b = 0,\ b \in R^K\}$. So $\beta_n^0 = \beta^0$ for the purposes of our local mixed constraints hypothesis test. Therefore, we have the following theorem.

THEOREM 5.2. *For the local hypothesis testing problem H: $h^1(\beta) \geq 0$ and $h^2(\beta) = 0$, $\beta \in N_{\delta_n}(\beta^0)$ versus K: $\beta \in R^K$, the null asymptotic distribution of the KT, W, and LR statistics satisfies the following property:*

$$\sup_{b \in B^*} pr_{b, I(\beta^0)^{-1}}(X \geq c) = pr_{\beta^0}(X \geq c)$$

$$= \sum_{k=0}^{L} pr(\chi^2_{P-L+k} \geq c) w(L, L-k, \Pi),$$

where Π is the asymptotic covariance matrix of $n^{1/2}h^1(\ddot{\beta})$.

We define $\ddot{\beta}$ as the equality constrained estimate of β calculated by assuming $h^2(\beta) = 0$. The asymptotic covariance matrix of $n^{1/2}h^1(\ddot{\beta})$ is used to compute the weights because under our null hypothesis the unrestricted estimate of β assumes that $h^2(\beta) = 0$. From Silvey [36], the asymptotic covariance matrix (avar) of $n^{1/2}h^1(\ddot{\beta})$ is:

$$\begin{aligned}\operatorname{avar}(n^{1/2}h^1(\ddot{\beta})) = {} & H^1(\beta^0)I(\beta^0)^{-1}H^1(\beta^0)' \\ & - H^1(\beta^0)I(\beta^0)^{-1}H^2(\beta^0)'(H^2(\beta^0)I(\beta^0)^{-1}H^2(\beta^0)')^{-1} \\ & H^2(\beta^0)I(\beta^0)^{-1}H^1(\beta^0)'.\end{aligned}$$

In partitioned matrix form, the asymptotic covariance matrix of $n^{1/2}\bar{\lambda}$ is:

$$\begin{bmatrix} H^1(\beta^0)I(\beta^0)^{-1}H^1(\beta^0)' & H^1(\beta^0)I(\beta^0)^{-1}H^2(\beta^0)' \\ H^2(\beta^0)I(\beta^0)^{-1}H^1(\beta^0)' & H^2(\beta^0)I(\beta^0)^{-1}H^2(\beta^0)' \end{bmatrix}^{-1}.$$

The element of $(H(\beta^0)I(\beta^0)^{-1}H(\beta^0)')^{-1}$ corresponding to $H^1(\beta^0)I(\beta^0)^{-1}H^1(\beta^0)'$ is:

$$\begin{aligned}Q = {} & [H^1(\beta^0)I(\beta^0)^{-1}H^1(\beta^0)' \\ & - H^1(\beta^0)I(\beta^0)^{-1}H^2(\beta^0)'(H^2(\beta^0)I(\beta^0)^{-1}H^2(\beta^0)')^{-1}H^2(\beta^0) \\ & I(\beta^0)^{-1}H^1(\beta^0)']^{-1}.\end{aligned}$$

We just have shown $k[\operatorname{avar}(n^{1/2}h^1(\ddot{\beta}))]^{-1} = Q$, with $k > 0$, where the weights of Theorem 5.2 depend on $\operatorname{avar}(n^{1/2}h^1(\ddot{\beta}))$, and those in Theorem 5.1 depend on Q. The properties of the weights functions given in Wolak [44] yield the following result.

COROLLARY 5.1. *For hypothesis tests of the same asymptotic size, the local combination inequality and equality constraints test, H*: $h^1(\beta) \geq 0$ *and* $h^2(\beta) = 0$, $\beta \in N_{\delta_n}(\beta^0)$ *versus K*: $\beta \in R^K$ *is equivalent to the combination multivariate one-sided and two-sided test, H*: $\lambda = 0$ *versus K*: $\lambda_1 \geq 0$ *and* $\lambda_2 \neq 0$, *where* λ *is as defined in (4.6).*

If $L = 0$ (the case of testing only equality constraints), our framework reduces to that derived by Aitchison and Silvey [1] for testing multivariate nonlinear equality constraints. The null asymptotic distribution of our test statistics is chi-squared with degrees of freedom equal to the dimension of the nonlinear equality constraints vector.

At this point we discuss the power of these tests. Power calculations are extremely difficult, even for the exact distribution multivariate one-sided test (4.7), because the weights entering into the computation of the distribution of the test statistics depend on the alternative hypothesis considered. This well-known problem also arises in the present testing framework. However, the local power of this test for our inequality constraints null hypothesis is greater against any alternative of the form $h(\beta) \leq 0$ than the two-sided test H: $h(\beta) = 0$ versus K: $h(\beta) \neq 0$ because our test takes into account the fact for this hypothesis test λ is always greater than zero under the alternative. For similar reasons, the mixed inequality-equality test statistic should have superior local power properties against these types of alternatives ($h^1(\beta) \leq 0$, $h^2(\beta) \neq 0$) for testing mixed null hypotheses. For the linear regression model and multivariate one-sided test, Hillier [24] compares the power properties of the classical two-sided F-test, the likelihood ratio test, and a one-sided t-test in a particular direction. Based on his calculations, the likelihood ratio test is the preferred test. For a further discussion of power for multivariate one-sided hypothesis tests see Bartholomew [8] and Barlow, Bartholomew, Bremner, and Brunk [3]. These power discussions and the results in Hillier [24] are relevant to our inequality constraints hypothesis testing framework because of the precise linkage derived earlier between any size test for local multivariate inequality constraints and the multivariate one-sided test in terms of the true value of the Lagrange multiplier vector. See Wolak [43] for more on this point.

6. EXTENSION TO BGS ESTIMATION FRAMEWORK

In this section we outline the extension of the results for the ML model to the BGS class of estimators. The notation of this section follows that in Gallant [19, chap. 3] unless otherwise stated. The value of λ that minimizes $s_n(\lambda)$ subject to the appropriate set of constraints defines the associated estimates of λ_n^0. In this framework, λ_n^0 is assumed to drift toward λ^* at the rate of $n^{-1/2}$. To concentrate on the problem of testing nonlinear inequality constraints, model misspecification in the way given in [19] is assumed not to exist. Essentially, this implies that the model generating the data is contained within the parametric class of models being estimated. This implies that $\mathcal{I} = \mathcal{J}$ and $\mathcal{U} = 0$ in the notation of [19, p. 239]. Under these conditions, the statistics for this framework analogous to the LR, W, and KT statistics defined earlier are locally asymptotically equivalent. The regularity conditions necessary for the validity of the results in this section are those in [19, chap. 3] and the constraint qualification conditions in the Appendix.

Two estimators of λ_n^0 are needed to define the test statistics for this framework. The first, the unconstrained estimator $(\hat{\lambda}_n)$, is the solution to:

$$\min_{\lambda} s_n(\lambda) \qquad \text{subject to } \lambda \in \Lambda, \tag{6.1}$$

where Λ is the compact parameter space containing λ_n^0 and λ^*. The second. is the inequality constrained estimate, $\tilde{\lambda}_n$. This estimator arises as the solution to:

$$\min_{\lambda} s_n(\lambda) \qquad \text{subject to } h(\lambda) \geq 0, \quad \lambda \in \Lambda, \tag{6.2}$$

where $h(\lambda)$ satisfies the same conditions in λ and λ^* as $h(\beta)$ in our notation satisfies in β and β^0. Associated with the vector of inequality constraints is a vector of Kuhn-Tucker multipliers, $\tilde{\theta}_n$.

We now define our test statistics for the local hypothesis test

$$H\colon h(\lambda) \geq 0, \lambda \in N_{\delta_n}(\lambda^*) \quad \text{versus} \quad K\colon \lambda \in R^p \text{ for all } n. \tag{6.3}$$

The analog of the likelihood ratio statistic is:

$$L = 2n(s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)). \tag{6.4}$$

The Wald statistic is:

$$\mathrm{W} = n(\tilde{\lambda}_n - \hat{\lambda}_n)'\mathcal{I}(\tilde{\lambda}_n - \hat{\lambda}_n) \tag{6.5}$$

or

$$\mathrm{W}^* = n(h(\tilde{\lambda}_n) - h(\hat{\lambda}_n))'[H(\lambda^*)\mathcal{I}^{-1}H(\lambda^*)']^{-1}(h(\tilde{\lambda}_n) - h(\hat{\lambda}_n)), \tag{6.6}$$

where $H(\lambda)$ is the matrix of partial derivatives of $h(\lambda)$ with respect to λ and $H(\lambda^*)$ is of full row rank q. The Kuhn-Tucker statistic is:

$$\mathrm{KT} = n\tilde{\theta}_n'[H(\lambda^*)\mathcal{I}^{-1}H(\lambda^*)']\tilde{\theta}_n. \tag{6.7}$$

All of these statistics are locally asymptotically equivalent and their null distribution for an asymptotically exact size test is given by Theorem 4.2 where Π in our notation is $[H(\lambda^*)\mathcal{I}^{-1}H(\lambda^*)']$.

For the local combination nonlinear inequality and equality constraints test

$$H\colon h^1(\lambda) \geq 0, h^2(\lambda) = 0, \lambda \in N_{\delta_n}(\lambda^*) \quad \text{versus} \quad K\colon \lambda \in R^p \text{ for all } n, \tag{6.8}$$

we must first define the mixed constraints estimator, $\bar{\lambda}_n$, as the solution to:

$$\min_{\lambda} s_n(\lambda) \qquad \text{subject to } h^1(\lambda) \geq 0 \text{ and } h^2(\lambda) = 0,\ \lambda \in \Lambda, \tag{6.9}$$

for the $h(\lambda)$ partitioned in the same manner as $h(\beta)$ in Section 5. Also arising from this estimation procedure is the mixed constraints multiplier vector $\bar{\theta}_n$. For this hypothesis test replace $\tilde{\lambda}_n$ with $\bar{\lambda}_n$ in the definition of the L, W, and W* statistics and $\tilde{\theta}_n$ with $\bar{\theta}_n$ in the definition of the KT statistic. The null asymptotic distribution for any asymptotically exact size test is given by the results of Theorem 5.1 or Theorem 5.2. In terms of Theorem 5.1, the weighted sum of chi-squared distributions depends on Γ, which in the current notation is the submatrix of $[H(\lambda^*)\mathcal{I}^{-1}H(\lambda^*)']^{-1}$ corresponding to the elements of $\bar{\theta}_n$ associated with the vector of inequality constraints $h^1(\lambda)$. Finally, it is straightforward to show that the duality relations derived for the ML model framework carry over to the BGS class of estimators.

7. APPLYING THE STATISTICS

The calculation of the critical value for a given hypothesis testing problem is no longer as simple as looking up the relevant number in a table because the weighted sum of chi-squared distributions determine the critical value. The widespread availability of FORTRAN subroutine libraries make the task substantially easier.

For the asymptotic level α test in the pure inequality constraints case, the critical value is the solution in x of the following equation:

$$\alpha = \sum_{k=1}^{P} pr[\chi_k^2 \geq x]\, w(P,k,\Xi), \tag{7.1}$$

where $\Xi = (H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)')^{-1}$ in our notation. Problem (7.1) can be solved by any method for finding the zero of a univariate function. In addition, $\boldsymbol{I}(\beta^0)$ can be replaced by

$$-n^{-1}\frac{\partial^2 L(\beta^0)}{\partial\beta\partial\beta'} \quad \text{or} \quad n^{-1}\sum_{i=1}^{n}\frac{\partial ln(f(x_i,\beta^0))}{\partial\beta}\,\frac{\partial ln(f(x_i,\beta^0))'}{\partial\beta}.$$

Both of these expressions converge almost surely to $\boldsymbol{I}(\beta^0)$ under our assumptions. To be consistent with our local inequality constraints hypothesis test, these expressions should be evaluated at β^0, the value of β the test is local to. However, replacing β^0 by any of its consistent estimates ($\hat{\beta}$, $\tilde{\beta}$, or $\bar{\beta}$) in these two estimates of $\boldsymbol{I}(\beta^0)$ and in the evaluation of the estimate of $H(\beta^0)$ is asymptotically valid under our regularity conditions. Similarly, for the general BGS class of models, the estimate of $\mathcal{I}$ constructed as discussed in [19, chap. 3] which is used to compute the weights, should be evaluated at $\lambda = \lambda^*$, the fixed value of λ the hypothesis test is local to, although using any of the three consistent estimates of λ is asymptotically valid under the regularity conditions in [19].

For the asymptotic level α test in the mixed equality-inequality constraints case the critical value is the solution in x of:

$$\alpha = \sum_{k=0}^{L} pr[\chi^2_{P-L+k} \geq x]\, w(L,k,\Gamma), \tag{7.2}$$

where Γ is as defined in Theorem 5.1. This problem also involves finding the zero of a univariate function. The procedures described above for computing estimates of Ξ apply to computing estimates of Γ.

There is another methodology when iterative procedures are not practical. In this case we calculate the probability of getting a value greater than or equal to any of the small sample values of our three statistics from a random variable with the null asymptotic distribution of our test statistics. If $G(x)$ is the asymptotic distribution of our test statistics under the null hypothesis we calculate $1 - G(Z)$, where Z is the value of any of our three statistics. For the inequality constrained case:

$$1 - G(Z) = \sum_{k=1}^{P} pr[\chi^2_k \geq Z]\, w(P,k,\Xi).$$

For the mixed equality-inequality constrained case $1 - G(Z)$ is:

$$1 - G(Z) = \sum_{k=0}^{L} pr[\chi^2_{P-L+k} \geq Z]\, w(L,k,\Gamma).$$

The chi-squared probabilities can be calculated numerically or interpolated from available tables of the chi-squared distribution. There are also various series expansions methodologies for calculating these probabilities (see Lackritz [28]). In this instance an investigator rejects the null hypothesis if $1 - G(Z) < \alpha$, where α is the size of the hypothesis test.

A final methodology uses bounds on the asymptotic distribution to compute upper and lower bounds on the critical value similar to the upper and lower bounds on the critical value used in the Durbin-Watson test. Following Perlman [32], Kodde and Palm [26] compute these bounds on the asymptotic distribution of their generalized distance test statistic. Their bounds are applicable to the testing framework presented in this paper because of the asymptotic equivalence shown in Sections 4 and 5 between the two sets of KT, LR, and W statistics and the generalized distance test statistics presented in [26]. The upper and lower bounds on the size α, critical value from the asymptotic distribution of a mixed inequality-equality constraints test statistic, are the solution in c_u and c_l, respectively, to the following two equations:

$$\alpha = \frac{1}{2} pr[\chi_P^2 \geq c_u] + \frac{1}{2} pr[\chi_{P-1}^2 \geq c_u]$$

$$\text{and} \quad \alpha = \frac{1}{2} pr[\chi_{P-L+1}^2 \geq c_l] + \frac{1}{2} pr[\chi_{P-L}^2 \geq c_l]. \tag{7.3}$$

When there are no equality constraints, these bounds are still valid with L set equal to P. Kodde and Palm [26] compute a table of these critical values for several values of P, L, and α. Consequently, the asymptotically exact critical value must be calculated only in cases when these bounds yields an inconclusive test results (i.e., $c_u > Z > c_l$).

8. EXTENSIONS OF GHM FRAMEWORK

The general case of the hypothesis test considered in Gourieroux, Holly, and Monfort [21] is H: $h(\beta) = 0$ versus K: $h(\beta) \geq 0$, with the inequality strict for at least one element of $h(\beta)$. Their paper considered H: $\beta_i = 0$ ($i = 1, \ldots, L \leq K$) versus K: $\beta_i \geq 0$ ($i = 1, \ldots, L$). Before proceeding, we should note that the GHM problem in nonlinear models falls prey to the same indeterminacy of the least favorable null distribution as occurs in testing global nonlinear inequality constraints. Local nonlinear multivariate one-sided tests must be resorted to if asymptotically exact size tests are desired. Wolak [42] elaborates on this point. The general case of the nonlinear multivariate one-sided tests is asymptotically equivalent to the exact test of the null hypothesis that β is such that $H(\beta^0)(\beta - \beta^0) = 0$ versus the one-sided alternative that β lies in the *cone of tangents to the set C at* β^0, based on $\hat{b} \sim N(\beta, H(\beta^0)I(\beta^0)^{-1}H(\beta^0)')$. In our notation, the local GHM test is:

$$H\colon h(\beta_n^0) = 0 \quad \text{versus} \quad K\colon h(\beta_n^0) \geq 0 \text{ for all } n,$$

with $\beta_n^0 \in N_{\delta_n}(\beta^0)$ under both the null and alternative hypotheses. The null asymptotic distribution is:

$$pr_{\beta^0}(Z \geq c) = \sum_{k=0}^{P} pr(\chi_k^2 \geq c) w(P, k, \Pi),$$

where Π is as defined in Theorem 4.2 and Z is the asymptotic value of the three GHM test statistics.

In this section we expand their framework to consider local hypothesis tests of the form

$$H\colon h(\beta_n^0) = 0 \quad \text{versus} \quad K\colon h^1(\beta_n^0) \geq 0 \text{ and } h_2(\beta_n^0) \neq 0 \text{ for all } n, \tag{8.1}$$

with $\beta_n^0 \in N_{\delta_n}(\beta^0)$ under both hypotheses, within the context of our maximum likelihood model. Assume the nonlinear constraint vector $h(\beta)$ satisfies the same regularity conditions and is partitioned in the same fashion as for the mixed constraints testing problem.

Proceeding in a similar fashion to our mixed inequality-equality constraints testing problem, we define the LR, W, and KT statistics for the testing problem (8.1). The LR statistic is defined as:

$$\mathrm{LR} = -2[L(\bar{\beta}) - L(\hat{\hat{\beta}})] = 2[L(\hat{\hat{\beta}}) - L(\bar{\beta})]. \tag{8.2}$$

We define $\hat{\hat{\beta}}$ as the solution to the following inequality constrained maximum likelihood problem:

$$\max_{\beta} n^{-1}L(\beta) \qquad \text{subject to } h^1(\beta) \geq 0,\ \beta \in \Theta. \tag{8.3}$$

Associated with this optimization problem is an L-dimensional vector of Kuhn-Tucker multipliers for the inequality constraints, $\hat{\hat{\lambda}}$. The first order conditions for this optimization problem are:

$$n^{-1}\frac{\partial L}{\partial \beta}(\hat{\hat{\beta}}) = H(\hat{\hat{\beta}})'\hat{\hat{\lambda}}, \quad h_j(\hat{\hat{\beta}})\hat{\hat{\lambda}}_j = 0\ (j = 1, \ldots, L), \quad \hat{\hat{\lambda}} \leq 0, \quad h^1(\hat{\hat{\beta}}) \geq 0.$$

Two forms of the Wald statistic for this hypothesis testing problem are:

$$\overline{W} = n(\hat{\hat{\beta}} - \bar{\beta})'\boldsymbol{I}(\beta^0)(\hat{\hat{\beta}} - \bar{\beta}) \tag{8.4}$$

and

$$W = nh(\hat{\hat{\beta}})'[H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']^{-1}h(\hat{\hat{\beta}}). \tag{8.5}$$

The KT statistic for this hypothesis test is:

$$\mathrm{KT} = n(\hat{\hat{\lambda}}^* - \bar{\lambda})'[H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)'](\hat{\hat{\lambda}}^* - \bar{\lambda}) \tag{8.6}$$

where $\hat{\hat{\lambda}}^*$ is a P-dimensional vector whose first L elements are $\hat{\hat{\lambda}}$ and last $P - L$ elements are zeros. By a straightforward application of the logic used in [21], these four statistics are asymptotically equivalent.

Under our null hypothesis, $n^{1/2}h(\hat{\beta})$ converges in distribution to a $N(0, H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)')$ random vector. Hence, asymptotically, the testing problem (8.1) fits into the framework of hypothesis testing problem (5.8). We state our result in the form of a lemma.

LEMMA 8.1. *For the local hypothesis testing problem (8.1), the asymptotic null distribution of the LR, W, and KT statistics is:*

$$pr_{\beta^0}(X \geq c) = \sum_{k=0}^{L} pr(\chi^2_{P-L+k} \geq c)\, w(L, k, Y)$$

where X is the asymptotic value of the three statistics and Y is the submatrix of $[H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']$ corresponding to $h^1(\beta)$.

The computation of critical values and probability values follows the same logic given in Section 7.

9. SUMMARY AND EXTENSIONS

In this paper we devised an asymptotic testing framework for examining nonlinear inequality constraints within a large class of nonlinear econometric models. We extended these results to consider combinations of nonlinear equality and inequality restrictions. We also extended the asymptotic one-sided hypothesis testing framework of GHM to consider combinations of one-sided and two-sided hypothesis tests. We also illustrated the local nature of any testing problem involving nonlinear inequality constraints or a nonlinear model when the dimension of the test is less than the dimension of the parameter vector. Our approach also provided a well-known framework (the ML model) to show why a global inequality constraints test is not in general possible for nonlinear constraints or models.

Inequality constraints can be tested locally for an even larger class of models. As long as an estimation procedure yields local asymptotic normal estimators, we can test local nonlinear inequality constraints and combinations of local nonlinear equality and inequality constraints following the approach in Kodde and Palm [26]. Conditions for local asymptotic normality are that $n^{1/2}(\hat{\beta} - \beta_n^0)$ converges in distribution to $N(0,\Sigma)$, where Σ is fixed and $\beta_n^0 \in N_{\delta_n}(\beta^0)$, and $N_{\delta_n}(\beta^0)$ and β^0 are as defined earlier. These conditions are analogous to those presented by Stoud [37] to guarantee the existence of the local power function for multivariate equality constraints tests.

Given a local asymptotic normal estimator, for the hypothesis test

$$H\colon h(\beta) \geq 0,\ \beta \in N_{\delta_n}(\beta^0) \quad \text{versus} \quad K\colon \beta \in R^K, \tag{9.1}$$

we compute the Kodde and Palm [26] test statistic

$$U = \min_x n(h(\hat{\beta}) - x)'[H(\beta^0)\Sigma H(\beta^0)']^{-1}(h(\beta^0) - x) \tag{9.2}$$

subject to $x \geq 0$,

where $H(\beta)$ is as defined earlier and $[H(\beta^0)\Sigma H(\beta^0)']$ is the asymptotic covariance matrix of $n^{1/2}h(\hat{\beta})$. The asymptotic distribution of U for the purposes of testing the null hypothesis of (9.1) is given in Theorem 4.2 with $\Pi = [H(\beta^0)\Sigma H(\beta^0)']$.

Under the same conditions on the estimator, for the hypothesis test

$$H\colon h^1(\beta) \geq 0, \quad h^2(\beta) = 0, \quad \beta \in N_{\delta_n}(\beta^0) \quad \text{versus} \quad K\colon \beta \in R^K, \tag{9.3}$$

we compute the Kodde and Palm [26] test statistic

$$Z = \min_x n(h(\hat{\beta}) - x)[H(\beta^0)\Sigma H(\beta^0)']^{-1}(h(\hat{\beta}) - x) \tag{9.4}$$

subject to $x_1 \geq 0$ and $x_2 = 0$,

where x_1 corresponds to the inequality constraints $h^1(\beta)$ and x_2 corresponds to the equality constraints $h^2(\beta)$. The asymptotic distribution of Z for the purposes of testing the null hypothesis is given in Theorem 5.1, with $\boldsymbol{I}(\beta^0)^{-1}$ replaced by Σ.

REFERENCES

1. Aitchison, J. & S.D. Silvey. Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics* 29 (1958): 813-828.
2. Avriel, M. *Nonlinear programming: analysis and methods.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1976.
3. Barlow, R.E., D.J. Bartholomew, J.N. Bremner, & H.D. Brunk. *Statistical inference under order restrictions: the theory of isotonic regression.* New York: John Wiley & Sons, 1972.
4. Barnett, W.A. New indices of money supply and the flexible Laurent demand system. *Journal of Business & Economic Statistics* 1 (1983): 7-24.
5. Barnett, W.A. & Y.W. Lee. The global properties of the minflex Laurent, generalized Leontief, and translog flexible functional forms. *Econometrica* 53 (1985): 1421-1438.
6. Bartholomew, D.J. A test of homogeneity for ordered alternatives, I. *Biometrika* 46 (1959): 36-48.
7. Bartholomew, D.J. A test of homogeneity for ordered alternatives, II. *Biometrika* 46 (1959): 328-335.
8. Bartholomew, D.J. A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society Series B* 23 (1961): 239-281.
9. Bazaraa, M.S. & C.M. Shetty. *Nonlinear programming: theory and algorithms.* New York: John Wiley & Sons, 1979.
10. Burguete, J.F., A.R. Gallant, & G. Souza. On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews* 1 (1982): 151-190.
11. Chernoff, H. On the distribution of the likelihood ratio. *Annals of Mathematical Statistics* 25 (1954): 573-578.
12. Diewert, W.E. & Wales, T.J. Flexible functional forms and global curvature conditions. *Econometrica* 55 (1987): 43-68.
13. Dufour, J.M. Nonlinear hypotheses, inequality restrictions and non-nested hypotheses: exact simultaneous tests in linear regressions. Department of Economics, Université de Montreal, Cahier 8647, October 1986.
14. Farebrother, R.W. Testing linear inequality constraints in the standard linear model. *Communications in Statistics—Theory and Methods* 15 (1986): 7-31.
15. Feder, P.I. On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. *Annals of Mathematical Statistics* 39 (1968): 2044-2055.
16. Gallant, A.R. On the bias in flexible functional forms and an essentially unbiased form: the Fourier flexible form. *Journal of Econometrics* 15 (1981): 211-245.
17. Gallant, A.R. Unbiased determination of production technologies. *Journal of Econometrics* 20 (1982): 285-323.
18. Gallant, A.R. & G.H. Golub. Imposing curvature restrictions on flexible functional forms. *Journal of Econometrics* 26 (1984): 295-322.
19. Gallant, A.R. *Nonlinear Statistical Models.* New York: John Wiley & Sons, 1987.
20. Gill, P.E., W. Murray, & M.H. Wright. *Practical Optimization.* New York: Academic Press, 1981.
21. Gourieroux, C., A. Holly, & A. Monfort. 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, June 1980.
22. Gourieroux, C., A. Holly, & A. Monfort. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50 (1982): 63-80.
23. Hendry, D.F. The structure of simultaneous equations estimators. *Journal of Econometrics* 4 (1976): 51-88.

24. Hillier, G. Joint tests for zero restrictions on nonnegative regression coefficients. *Biometrika* 73 (1986): 657-670.
25. Jorgenson, D.W., L.J. Lau, & T.M. Stoker. The transcendental logarithmic model of aggregate consumer behavior. In R.L. Basmann & G. Rhodes (eds.), *Advances in Econometrics*, Vol. 1, pp. 97-236. Greenwich: JAI Press, 1982.
26. Kodde, D.A. & F.C. Palm. Wald criterion for jointly testing equality and inequality restrictions. *Econometrica* 54 (1986): 1243-1248.
27. Kudo, A. A multivariate analogue of the one-sided test. *Biometrika* 50 (1963): 403-418.
28. Lackritz, J.R. Exact p values for F and t tests. *The American Statistician* 38 (1984): 312-314.
29. Lau, L.J. Testing and imposing monotonicity, convexity and quasi-convexity constraints. In M. Fuss and D. MacFadden (eds.), *Production economics: a dual approach to theory and applications*, Vol. 1, pp. 409-453. Amsterdam: North-Holland Publishing Company, 1978.
30. Lehmann, E.L. *Testing statistical hypotheses* (2nd ed). New York: John Wiley & Sons, Inc., 1986.
31. Nuesch, P.E. On the problem of testing location in multivariate problems for restricted alternatives. *Annals of Mathematical Statistics* 37 (1966): 113-119.
32. Perlman, M.D. One-sided problems in multivariate analysis. *Annals of Mathematical Statistics* 40 (1969): 549-567.
33. Rogers, A.J. Modified Lagrange multiplier tests for problems with one-sided alternatives. *Journal of Econometrics* 31 (1986): 341-361.
34. Rothenberg, T.J. *Efficient estimation with a priori information.* New Haven: Yale University Press, 1973.
35. Silvey, S.D. The Lagrangian multiplier test. *Annals of Mathematical Statistics* 30 (1959): 389-407.
36. Silvey, S.D. *Statistical inference.* London: Chapman and Hall, 1970.
37. Stroud, T.W.F. On obtaining large-sample tests from asymptotically normal estimators. *Annals of Mathematical Statistics* 42 (1971): 1412-1424.
38. Stroud, T.W.F. Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistic. *Annals of Mathematical Statistics* 43 (1972): 447-454.
39. Theil, H. *Principles of econometrics.* New York: John Wiley & Sons, Inc., 1971.
40. Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54 (1943): 426-482.
41. White, H. Nonlinear regression on cross-section data. *Econometrica* 48 (1980): 721-746.
42. Wolak, F.A. The local nature of hypothesis tests involving nonlinear inequality constraints. Mimeo, December 1987.
43. Wolak, F.A. Duality in testing multivariate inequality constraints. *Biometrika* 75 (1988): 611-615.
44. Wolak, F.A. Testing inequality constraints in linear econometric models. *Journal of Econometrics*, forthcoming, 1989.

APPENDIX

1. REGULARITY CONDITIONS

This appendix gives a set of regularity conditions, in addition to those at the beginning of Section 2, necessary for the validity of the results in the paper. With exception of the constraint qualification conditions, these conditions are exactly those required for investigating the local power properties of nonlinear equality constraints tests. These conditions are stated in the form of assumptions. Because the primary concern of this paper is hypothesis testing, these regularity conditions are those most useful for that purpose. More general and easily verifiable sets of conditions which imply these are cited. The proof of Theorem 4.2 and a discussion of the global monotonicity of the power function of the test follows the regularity conditions.

Assumption 1. For all n, $n^{-1}L(\beta)$ is a continuous function of β.

Assumption 2. $n^{-1}L(\beta)$ converges almost surely to a function $L_\infty(\beta,\beta^0) \equiv E_{\beta^0}[\ln(f(X,\beta))]$ for all $\beta \in \Theta$. The function $L_n(\beta,\beta_n^0) \equiv E_{\beta_n^0}[\ln(f(X,\beta))]$ has a unique maximum at β_n^0. In addition, as $n \to \infty$, the function $L_n(\beta,\beta_n^0)$ converges to $L_\infty(\beta,\beta^0)$, which has a unique local maximum at $\beta = \beta^0$.

Assumptions 1 and the compactness of Θ guarantee the existence with probability one, of the unrestricted and equality, inequality, and mixed inequality-equality restricted estimates of β_n^0. These results follow from the compactness of the set of β over which the likelihood function is maximized for each of the estimation problems. Assumption 2 ensures that the four estimates of β_n^0 converge almost surely to β^0. By applying Lemma 2.2 of White [41], modified for our problem, these results follow in a straightforward fashion. Assumptions 1-5 and 6A of Silvey [35] are more readily verifiable conditions which guarantee the strong consistency of each of the estimators.

We now consider assumptions necessary for deriving the asymptotic distributions of the various estimators of β_n^0.

Assumption 3. The partial derivatives $\dfrac{\partial L(\beta)}{\partial \beta_j}$ $(j = 1,..,K)$ exist and are continuous on Θ with probability one.

Assumption 4. The second partial derivatives $\dfrac{\partial^2 L(\beta)}{\partial \beta_i \partial \beta_j}$ $(j=1,\ldots,K)(i=1,\ldots,K)$ exist and are continuous on Θ with probability one.

Assumption 5. The $(K \times K)$ matrix $-n^{-1}\dfrac{\partial^2 L(\beta)}{\partial \beta \partial \beta'}$ converges almost surely and uniformly for all $\beta \in \Theta$ to the matrix $\boldsymbol{I}(\beta,\beta^0) \equiv \lim_{n\to\infty} n^{-1}E_{\beta^0}[-\partial^2 L(\beta)/\partial\beta\partial\beta']$.

Assumption 6. The matrix $\boldsymbol{I}(\beta^0,\beta^0)$ is positive definite.

By definition, $\boldsymbol{I}(\beta^0)$ in the paper is $\boldsymbol{I}(\beta^0,\beta^0)$. Assumptions 3, 4, and 5 imply that for any β_n that converges almost surely to β^0, $n^{-1}\dfrac{\partial^2 L(\beta_n)}{\partial\beta\partial\beta'}$ converges almost surely to $\boldsymbol{I}(\beta^0)$ as $n \to \infty$. This result is used in the Taylor's expansion of the first derivatives of $n^{-1}L(\beta)$ used to derive the asymptotic distribution of the estimators and to establish the relationships presented at the end of Section 2 necessary to show the asymptotic equivalence of test statistics in Sections 3, 5, and 8. In addition, it is essential to allowing us to replace $\boldsymbol{I}(\beta^0)$ by either of its two consistent estimates and replacing β^0 by any of its consistent estimates in the evaluation of $\boldsymbol{I}(\beta^0)$ as discussed in Section 7. Note that implicit in these assumptions is the usual result for a correctly specified ML model that $I(\beta^0) = \lim_{n\to\infty} n^{-1}E_{\beta^0}\left[\frac{\partial L(\beta^0)}{\partial \beta}\frac{\partial L(\beta^0)}{\partial \beta'}\right]$.

Assumption 7. The vector $n^{-1/2}\frac{\partial L}{\partial \beta}(\beta_n^0), \beta_n^0 \in N_{\delta_n}(\beta^0)$, where $\delta_n = O(n^{-1/2})$, is asymptotically normal with mean the zero vector and covariance matrix $I(\beta^0)$ under β_n^0.

Silvey [35] in assumptions 1–12, with θ^* in his notation replaced by β_n^0 in our notation, provides verifiable conditions which guarantee the satisfaction of our Assumptions 1–7. Conditions R1 to R6 of Feder [15] also imply the satisfaction of our assumptions. In short, there are many other available sets of regularity conditions which ensure the satisfaction of these assumptions.

We now consider assumptions which are specifically requried for the Kuhn-Tucker conditions to be necessary conditions for the existence of a local optimum to an inequality constrained mathematical programming problem. These conditions are termed constraint qualification conditions. First we require definitions from Bazaraa and Shetty [9, chap. 5] modified to our notation.

Definition. Let S be a nonempty set in R^K, and let $\bar{x} \in \mathrm{cl}(S)$. The *cone of tangents of S at* $\bar{x}$, which we denote by T, is the set of all directions d such that $d = \lim_{k\to\infty} \lambda_k(x_k - \bar{x})$, where $\lambda_k > 0$, $x_k \in S$ for each k and $x_k \to \bar{x}$.

By $\mathrm{cl}(S)$ we mean the closure of the set S. In words, d belongs to the cone of tangents if there is a sequence $\{x_k\} \in S$ converging to $\bar{x}$ such that the directions of the cords $(x_k - \bar{x})$ converge to d. Figure 1 provides a graphical example of the *cone of tangents to C at* β^0 for an arbitrary inequality constraints set.

We consider the constraint qualification conditions for the mixed inequality-equality constraints case because the constraint qualification condition for inequality constraints alone is a special case of this more general condition. Let $h(\beta)$ be partitioned in the same manner described in the paper, where L is the dimension of $h^1(\beta)$ and $P - L$ is the dimension of $h^2(\beta)$. For a given L let

$$\Theta^M = \Theta \cap \{\beta \mid h_j(\beta) \geq 0,\ j = 1,\ldots,L \text{ and } h_k(\beta) = 0,\ k = L+1,\ldots,P\}.$$

Let T be the *cone of tangents of* Θ^M *at* $\dot{\beta}$, where $\dot{\beta}$ is our mixed constraints estimate of β_n^0.

Given a $\dot{\beta}$, define the set $J = \{i \mid h_i(\dot{\beta}) = 0,\ i = 1,\ldots,L\}$. In addition, define the following two sets: $G = \left\{d \,\middle|\, \frac{\partial h_i}{\partial \beta}(\dot{\beta})'d \geq 0 \text{ for } i \in J\right\}$ and $H = \left\{d \,\middle|\, \frac{\partial h_i}{\partial \beta}(\dot{\beta})'d = 0 \text{ for } i = L+1,\ldots,P\right\}$. The Abadie constraint qualification condition requires the following relationship to hold between these sets T, G, and H for any $\dot{\beta}$.

Assumption 8. Abadie constraint qualification condition: $T = G \cap H$.

From [9, chap. 5], if $\dot{\beta}$ is a local optimum, this constraint qualification condition guarantees the existence of a multiplier vector λ that satisfies the Kuhn-Tucker conditions given in (5.2).

For the case of inequality constraints alone, $L = P$. Consequently, the constraint qualification condition for $\tilde{\beta}$ becomes: $T = G$. In this case, T is the *cone of tangents to* $\Theta^C \equiv \Theta \cap \{\beta | h_j(\beta) \geq 0, j = 1, \ldots, P\}$ at $\tilde{\beta}$, and G and J are as defined above with L replaced by P. This condition guarantees that if $\tilde{\beta}$ is a local optimum, then the Kuhn-Tucker conditions (2.3) are satisfied. These constraint qualification conditions guarantee that any movement from $\tilde{\beta}$ along a feasible direction cannot increase the value of the likelihood function. Bazaraa and Shetty [9] and Avriel [2] give detailed discussions of these constraint qualification conditions in their various forms. As a final note, these conditions are automatically satisfied for linear constraints regardless of the form of the objective function.

2. PROOF OF THEOREM 4.2

The full rank of $H(\beta^0)$ assumption implies that the sets C and C^* can be approximated by positively homogeneous sets (in our case cones) at β^0 in the sense given in Chernoff [11, Definition 2]. The cone approximating C at β^0 is $T(\beta^0)$ and the one approximating C^* at β^0 is $T^*(\beta^0)$. Both of these approximating cones are the respective cones of tangents at β^0 defined in Sections 4 and 5.

Throughout what follows we utilize the results of (4.4) which imply that the locally least favorable value of $\beta_n^0 \in K_n$ (or equivalently, b in (4.4)) for an asymptotically exact size test of our null hypothesis is β^0 (or equivalently, $b = 0$). Note that there may be other values of $b \neq 0$ which yield the same least favorable asymptotic distribution as $b = 0$ because the favorable value of b must only satisfy $H(\beta^0)\ b = 0$. Conditional of this locally least favorable value of β_n^0, our proof of the asymptotic null distribution follows that of Theorem 1 from [11]. Recall that

$$n^{1/2}(\hat{\beta} - \beta^0) \cong n^{1/2}\boldsymbol{I}(\beta^0)^{-1}S(\beta^0), \tag{A.1}$$

where $S(\beta^0) = n^{-1}\partial L(\beta^0)/\partial\beta$. From Equation (3.1) for $\hat{\beta}$ and Equation (3.5), we have the following:

$$L(\hat{\beta}) \cong L(\beta^0) + \frac{n}{2}(\hat{\beta} - \beta^0)'\boldsymbol{I}(\beta^0)(\hat{\beta} - \beta^0). \tag{A.2}$$

Utilizing (A.1) we obtain:

$$L(\hat{\beta}) \cong L(\beta^0) + \frac{n}{2} S(\beta^0)'\boldsymbol{I}(\beta^0)^{-1}S(\beta^0). \tag{A.3}$$

Let $\beta = \boldsymbol{I}(\beta^0)^{-1}S(\beta^0) + \beta^0 + \eta_n$, with $\eta_n = O(n^{-1/2})$. Then we have:

$$L(\beta) \cong L(\beta^0) + \frac{n}{2} S(\beta^0)'\boldsymbol{I}(\beta^0)^{-1}S(\beta^0) - \frac{n}{2}\eta_n'\boldsymbol{I}(\beta^0)\eta_n. \tag{A.4}$$

Subtracting Equation (A.4) from (A.3) and by an application of Lemma 1 of [11], with the origin and ϕ in his statement of the lemma equal to β^0 and C in our notation, we have:

$$LR \cong 2[L(\hat{\beta}) - L(\tilde{\beta})] \cong n[\min_{\beta \in K_n} \eta_n'\boldsymbol{I}(\beta^0)\eta_n]. \tag{A.5}$$

Recall that for sufficiently large n, K_n can be approximated by $T(\beta^0)$ at β^0. Given the definition of η_n,

$$LR \cong n[\min_{\beta \in T(\beta^0)} \eta_n' \boldsymbol{I}(\beta^0)\eta_n]$$

$$= n[\min_{\beta \in T(\beta^0)} ((\beta - \beta^0) - \boldsymbol{I}(\beta^0)^{-1}S(\beta^0))\boldsymbol{I}(\beta^0)((\beta - \beta^0) - \boldsymbol{I}(\beta^0)^{-1}S(\beta^0))]. \tag{A.6}$$

Define $A \equiv (\beta - \beta^0)$. The cone $T(\beta^0)$ becomes $T(\beta^0) = \{A | H(\beta^0)A \geq 0,\ A \in R^K\}$. Because $T(\beta^0)$ is positively homogenous we have:

$$LR \cong n[\min_{A \in T(\beta^0)} (A - \boldsymbol{I}(\beta^0)^{-1}S(\beta^0))'\boldsymbol{I}(\beta^0)(A - \boldsymbol{I}(\beta^0)^{-1}S(\beta^0))]$$

$$= [\min_{A \in T(\beta^0)} (A - n^{1/2}\boldsymbol{I}(\beta^0)^{-1}S(\beta^0))'\boldsymbol{I}(\beta^0)(A - n^{1/2}\boldsymbol{I}(\beta^0)^{-1}S(\beta^0))] \tag{A.7}$$

By Assumption 7, $n^{1/2}\boldsymbol{I}(\beta^0)^{-1}S(\beta^0)$ is asymptotically $N(0,\boldsymbol{I}(\beta^0)^{-1})$ under $\beta_n^0 = \beta^0$. This implies that the distribution of our test statistics for the purposes of testing our local null hypothesis is the same as the distribution of

$$LR^* = \min_{A \in T(\beta^0)} (A - \hat{A})'\boldsymbol{I}(\beta^0)(A - \hat{A}) \tag{A.8}$$

under the assumption that $\hat{A} \sim N(0,\boldsymbol{I}(\beta^0)^{-1})$. This statistic is a likelihood ratio statistic for the following testing problem in terms of $\hat{A}$ and A:

$$H\colon H(\beta^0)A \geq 0 \quad \text{versus} \quad K\colon A \in R^K. \tag{A.9}$$

Applying the results of our Theorem 4.1 and noting that $H(\beta^0)\hat{A} \sim N(0,H(\beta^0)\cdot \boldsymbol{I}(\beta^0)^{-1}H(\beta^0)')$, we find that the statistic (A.9) has the following distribution:

$$pr_{0,\Pi}(LR^* \geq c) = \sum_{k=0}^{P} pr(\chi_k^2 \geq c)w(P,P-k,\Pi)$$

where $\Pi = [H(\beta^0)\boldsymbol{I}(\beta^0)^{-1}H(\beta^0)']$. This is the asymptotic distribution claimed in Theorem 4.2 for our three test statistics. The results of Theorem 4.2 obtain because $LR^* \cong LR$. ∎

If we were not interested in the null distribution for the locally least favorable value and instead only required $\beta_n^0 \in N_{\delta_n}(\beta^0)$ for all n, we would be in the domain of the results of Feder [15]. In this case, our proof of the asymptotic distribution of the test statistics would follow the proof of [15, Theorem 1, Case 1]. Following Feder's logic, the asymptotic distribution of our test statistics is the same as the distribution of LR^* in (A.8) (the likelihood ratio statistic for (A.9)) under the assumption that $\hat{A} \sim N(\bar{A},\boldsymbol{I}(\beta^0)^{-1})$, where $\bar{A} = \lim_{n\to\infty} n^{1/2}(\beta_n^0 - \beta^0)$. In [44] this distribution was shown to be a complicated weighted sum of noncentral chi-squared distributions with noncentrality parameters depending on $\bar{A}$ and exactly which elements of $\tilde{A}$, the optimal value of (A.8), are positive. In addition, the weights functions will now depend on $\bar{A}$ as well as $H(\beta^0)I^{-1}H(\beta^0)'$. In brief, with the exception of low dimension problems, this distribution is extremely difficult to calculate.

3. GLOBAL MONOTONICITY OF ASYMPTOTIC POWER FUNCTION

We now turn to the global monotonicity of the power function for a global inequality constraints test. As discussed in the paper and in [42], for this general testing problem, there is, in general, no way to establish the least favorable value of β. We can only show that it must be in the set $C^E = \{x | h(x) = 0, x \in R^K\}$. Wolak [42] also showed that the method used to prove Lemma 8.2 in [32], contrary to the claim in [26], cannot be used to establish this global monotonicity property of the power function. In Perlman's notation, the condition that Σ not vary with μ does not hold in this case. The matrix $H(\beta^0)I(\beta^0)^{-1}H(\beta^0)'$ is functionally dependent on the value assumed for β so that an alternative route must be found to establish this monotonicity result.

Although a rigorous proof of this result is possible, it is very tedious. Moreover, the intuition behind the result is quite simple. For the sake of brevity, we concentrate on conveying this intuition. Rothenberg [34] dealt with estimation under inequality restrictions in the ML model and the impact of imposing these restrictions on the lower bound for the asymptotic covariance matrix of the parameter vector. His results [34, p. 50] state that as long as the restricted parameter space has full dimension, as is the case for inequality constraints, the asymptotic covariance matrix of the restricted estimator is the same as the unrestricted estimator. This argument is valid for all parameter values in the interior of the restricted parameter set. For our purposes, this implies that all of our test statistics will converge in probability to zero for values of β in the interior of C (all of the inequalities are satisfied as strict inequalities), because the probability of the event $\{\hat{\beta} | h(\hat{\beta}) > 0\}$ converges to one as $n \to \infty$. We can also rule out values of β such that $h_i(\beta) = 0$ for some i and $h_i(\beta) > 0$ for that rest of the elements of $h(\beta)$ by similar logic. Let H denote the set of i such that $h_i(\beta) > 0$. In this case, the probability of the event $\{\hat{\beta} | h_i(\hat{\beta}) > 0, i \in H\}$ converges to one as $n \to \infty$, so that the part of our statistics corresponding to these elements of $h(\beta)$ will converge in probability to zero. Therefore, the value of β which maximizes the large sample probability that our statistics are greater that any positive constant must be from the set C^E, where all of the inequalities in $h(\beta) \geq 0$ are satisfied as equalities. However, as discussed in the paper and in [42], what value of β from C^E is least favorable under the null hypothesis cannot in general be determined. Wolak [42] discusses these issues.