Journal of Econometrics 41 (1989) 205–235. North-Holland

TESTING INEQUALITY CONSTRAINTS IN LINEAR ECONOMETRIC MODELS

Frank A. WOLAK*

Stanford University, Stanford, CA 94305, USA

Received February 1986, final version received July 1988

This paper develops three asymptotically equivalent tests for examining the validity of imposing linear inequality restrictions on the parameters of linear econometric models. First we consider the model $y = X\beta + \varepsilon$, where ε is $N(0, \Sigma)$, and the hypothesis test H: $R\beta \geq r$ versus K: $\beta \in R^K$. Later we generalize this testing framework to the linear simultaneous equations model. We show that the joint asymptotic distribution of these test statistics and the test statistics from the hypothesis test H: $R\beta = r$ versus K: $R\beta \geq r$ is a weighted sum of two sets of independent χ^2-distributions. We also derive a useful duality relation between the multivariate inequality constraints test developed here and the multivariate one-sided hypothesis test. In small samples, these three test statistics satisfy inequalities similar to those derived by Berndt and Savin (1977) for the case of equality constraints. The paper also contains an illustrative application of this testing technique.

1. Introduction

Estimation under inequality restrictions has a long history in regression analysis and its application has become more widespread with the increase in sophistication of computer software. Judge and Takayama (1966) introduced least squares regression under inequality restrictions and suggested its formulation as a quadratic programming problem. Liew (1976) discussed the large-sample properties of the estimator as well as presented results of a simulation study of the small-sample properties of the inequality constrained least squares (ICLS) estimator. The increased use of this estimation technique suggests the need for a hypothesis testing procedure to examine its validity.

An inequality constraints testing framework should be useful to applied researchers because it provides a statistical test of the validity of *a priori* beliefs about the signs of regression parameters. Very often in empirical practice researchers obtain estimated parameter vectors where the signs of

*I would like to thank the associate editor and referees for instructive comments which lead to substantial improvements in the paper's content and presentation. I am especially grateful to Professor Arnold Zellner for his helpful editorial advice on previous drafts of this paper. I would also like to acknowledge the helpful comments of C.L. Cavanagh, A.K. Han, D.W. Jorgenson, and M.W. Watson on an earlier draft. This research was supported by National Science Foundation Grant SES-84-20262 to the Department of Economics, Harvard University.

0304-4076/89/$3.50

several of the elements are incorrect as determined by *a priori* knowledge or economic theory. The usual *ad hoc* procedure is to delete these variables from the estimating equation and re-estimate the model or report the regression as is and try to explain these incorrect signs. A multivariate inequality constraints test provides a solution to this problem by allowing the researcher to statistically test whether or not the signs of the true values of these elements are consistent with the researcher's beliefs. Or in practical terms, whether or not it is statistically valid to zero out incorrectly signed estimated coefficients from a regression equation.

Recently, methods for testing the validity of imposing inequality constraints have been proposed by econometricians. Yancy, Judge, and Bock (1981) discussed tests of the null hypothesis that a subset of the parameter vector lies in the positive orthant for the special case that the design matrix in the linear regression model $y = X\beta + \varepsilon$ is orthogonal ($X'X = I$) and the covariance matrix of the disturbance vector is scalar ($\mathrm{E}(\varepsilon\varepsilon') = \sigma^2 I$). Kodde and Palm (1986) devised a generalized distance test for examining the validity of multivariate nonlinear inequality constraints on the parameters of general econometric models which utilizes a consistent, asymptotically normal estimate of the parameter vector. Our work, developed independently of theirs, links up their generalized distance approach to testing multivariate inequality constraints to a likelihood-ratio-based approach.

For linear econometric models (the linear simultaneous equations model being the most general), we devise three asymptotically equivalent linear inequality constraints test statistics and derive their exact null asymptotic distribution. These statistics resemble the likelihood ratio, Wald, and Lagrange multiplier statistics for testing linear equality constraints. The specialization to linear inequality constraints and linear econometric models is an important distinction because several complications arise and the testing framework presented here must be modified when nonlinearities in the constraints or model are introduced.[1] This paper also extends the well-known duality relation in equality constraints testing between H: $R\beta = r$ versus K: $R\beta \neq r$ and H: $\lambda = 0$ versus K: $\lambda \neq 0$, where λ is the true value of the Lagrange multiplier vector associated with constraints $R\beta = r$. We show that for tests of the same size, H: $R\beta \geq r$ versus K: $\beta \in R^K$ is the same test as H: $\lambda = 0$ versus K: $\lambda \geq 0$. We also contrast the multivariate inequality constraints test considered here with the hypothesis test, H: $R\beta = r$ versus K: $R\beta \geq r$, considered by Gourieroux, Holly, and Monfort (1982), hereafter referred to as GHM. We derive the joint distribution of the likelihood ratio test statistics from these two testing problems. This joint distribution derivation illustrates the precise relationship between these two hypothesis tests and suggests a methodology for applying these test statistics to problems involving inequality constraints.

[1] Wolak (1987, 1989) discusses these issues.

The inequality constraints testing problem has many of its roots in the multivariate one-sided hypothesis testing literature of mathematical statistics. This literature deals with the hypothesis testing problem, H: $\mu = 0$ versus K: $\mu \geq 0$, where μ is the mean of a multivariate normal random vector. This literature begins with Bartholomew (1959a, b, 1961), who considers a related testing problem, H: $\mu_1 = \mu_2 = \cdots = \mu_K$ versus K: $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$, where the μ_j, $j = 1, \ldots, K$, are means of independent normal random variables. Kudo (1963) extends Bartholomew's results to the specific case considered above. Neusch (1966) also treats this same problem. Perlman (1969) generalizes these results to testing H: $\mu \in P_1$ versus K: $\mu \in P_2$, where P_1 and P_2 are positively homogeneous sets with $P_1 \subset P_2$. A special case of his framework is the hypothesis test H: $\mu \in A^*$ versus K: $\mu \in R^K$, where A^* is a closed, convex cone in R^K. Under certain conditions, linear inequality constraints define closed, convex cones in R^K. Consequently, we use this case of Perlman's framework as the general result off of which we specialize to develop our inequality constraints tests.

Robertson and Wegman (1978) test order restrictions as a null hypothesis within the context of the exponential family of distributions. They consider hypothesis tests of the form H: $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$ versus an unrestricted alternative. Dykstra and Robertson (1983) extend this testing framework to cases where a collection of independent normal means is, in their words, decreasing on the average. This allows reversals in the above inequalities over short ranges of the μ_j, $j = 1, \ldots, K$. The general methodology these researchers use to calculate the null distribution of their likelihood ratio statistic for testing order restrictions can be extended to apply to our problem of testing multivariate inequality constraints.

An outline of the rest of the paper follows. In section 2 we introduce the unconstrained, inequality-constrained, and equality-constrained generalized least squares estimators of the coefficients of the linear regression model for the case that the covariance matrix of the errors is known. Section 3 contains the derivation of the Kuhn–Tucker, Wald, and likelihood ratio statistics for testing inequality constraints and shows their equivalence when the covariance matrix of the disturbance vector of the model is known. Here we show the equivalence of the Kodde and Palm (1986) generalized distance statistic, specialized to linear models and constraints, to the three likelihood-ratio-based statistics. In section 4 we show the joint distribution of the likelihood ratio statistics from the GHM testing problem and the multivariate inequality constraints testing problem is a weighted sum of two sets of independent chi-squared distributions. This section also contains the extension of the duality relation in multivariate equality constraints testing to the multivariate inequality constraints testing framework.

In section 5 we show that when the covariance matrix of the disturbances depends on a finite number of parameters the test statistics of section 3 are asymptotically equivalent. We also show that the Berndt and Savin (1977) or Breusch (1979) inequalities for the Lagrange multiplier, Wald, and likelihood ratio tests continue to hold for our three analogous inequality constraints test statistics. Following this section is a simple, yet potentially very common, application of our test statistics in demand analysis. We test the hypothesis that electricity demand is decreasing in price, increasing in income, and increasing in the price of natural gas. Here we also suggest a procedure for applying the multivariate inequality constraints test and the GHM test to problems involving inequality constraints. The final section of the paper extends our testing technique to the linear simultaneous equations model. Hendry (1976) presents a discussion of all linear models which arise as a special case of this model.

2. The three estimates of β

For expositional ease, we first derive our results for the linear model

$$y = X\beta + \varepsilon, \tag{1}$$

where y is a $(T \times 1)$ vector, X a $(T \times K)$ matrix of rank K, and β is a $(K \times 1)$ vector. We assume that ε is a $(T \times 1)$ random vector which is $\mathrm{N}(0, \Sigma)$. We assume that Σ is a positive definite matrix of known constants. In section 5 we relax this assumption to the existence of a consistent estimate of Σ. Because we assume a general covariance matrix for ε, the results presented here apply to the univariate linear regression model with a very general form of the covariance matrix of the errors (for example, autocorrelation, heteroscedasticity, and error component models). The results also cover the linear seemingly unrelated regressions system and the linear multivariate regression model, because both models can be written in this general form [see Theil (1971, pp. 307–311)]. Linear inequality constraints in linear simultaneous equations models is the most general testing problem which can be dealt with by the framework discussed here. Section 7 presents the extensions to this framework. Wolak (1989) shows that a sufficient condition for the validity of the results of this paper is linearity of the model and constraint functions.

The matrix of constraints, R, is a $(P \times K)$ matrix of rank P, where $P \leq K$.[2] The inequality constraints are expressed as $R\beta \geq r$, where r is a known $(P \times 1)$ vector.

[2] The consideration of the case where $P > K$ adds considerably to the complexity of the problem. Kudo and Choi (1975) consider this situation for the multivariate one-sided hypothesis test.

We now define the three estimates of β. The ICLS estimator, $\tilde{b}$, is the solution to the quadratic program (QP):

$$\min_{b} (y - XB)'\Sigma^{-1}(y - Xb), \quad \text{subject to} \quad Rb \geq r. \tag{2}$$

Gill, Murray, and Wright (1981) provide an excellent survey of the various methods to solve this problem. The $(P \times 1)$ vector of Kuhn–Tucker multipliers associated with the constraints $Rb \geq r$ is represented by $\tilde{\lambda}$. The unconstrained estimator is obtained by generalized least squares as: $\hat{b} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$. For completeness, we associate with $\hat{b}$ a Kuhn–Tucker multiplier $\hat{\lambda}$ which is identically zero. The equality constrained estimator, $\bar{b}$, is the solution to

$$\min_{b} (y - Xb)'\Sigma^{-1}(y - Xb), \quad \text{subject to} \quad Rb = r. \tag{3}$$

The Lagrange multiplier vector associated with the equality constraints is $\bar{\lambda}$.

Several relations, which are useful in subsequent sections, exist between the three estimates of β. A derivation by GHM and Liew (1976) implies that each of the three estimators satisfies

$$b_n = \hat{b} + (X'\Sigma^{-1}X)^{-1}R'\lambda_n/2, \tag{4}$$

where n indexes the unconstrained, inequality-constrained, or equality-constrained estimator. This equation for the inequality constrained estimator is

$$\tilde{b} - \hat{b} = (X'\Sigma^{-1}X)^{-1}R'\tilde{\lambda}/2. \tag{5}$$

Eq. (4) for the equality constrained estimator is

$$2\left(R(X'\Sigma^{-1}X)^{-1}R'\right)^{-1}(r - R\hat{b}) = \bar{\lambda}. \tag{6}$$

Taking expectations of both sides of (6) gives

$$2\left(R(X'\Sigma^{-1}X)^{-1}R'\right)^{-1}(r - R\beta) = \lambda, \tag{7}$$

which defines λ in terms of β.

All of the estimators and their associated Kuhn–Tucker or Lagrange multipliers will be denoted by the symbols we have defined here throughout the entire paper.

3. Four equivalent test statistics

In this section we derive four equivalent tests for the null hypothesis $R\beta \geq r$, for the case in which Σ is known. The likelihood ratio (LR) statistic is defined in the usual fashion as

$$LR = -2\log(\tilde{L}/\hat{L}) = 2(\log\hat{L} - \log\tilde{L}),$$

where $\tilde{L}$ and $\hat{L}$ are the maximum values of the likelihood function under the null hypothesis ($R\beta \geq r$) and maintained hypothesis ($\beta \in R^K$). Because Σ is assumed to be known, $\hat{b}$ and $\tilde{b}$ are also the unconstrained and inequality-constrained maximum likelihood (ML) estimates of β. This implies

$$LR = (y - X\tilde{b})'\Sigma^{-1}(y - X\tilde{b}) - (y - X\hat{b})'\Sigma^{-1}(y - X\hat{b}).$$

The LR statistic is also the optimal value of the objective function from the following quadratic program (QP):

$$\begin{aligned} &\min_b\,(y - Xb)'\Sigma^{-1}(y - Xb) - (y - X\hat{b})'\Sigma^{-1}(y - X\hat{b}), \\ &\text{subject to} \quad Rb \geq r. \end{aligned} \tag{8}$$

Using the definition of $\hat{b}$, QP (8) becomes

$$\begin{aligned} &\min_b\, y'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y - 2y'\Sigma^{-1}Xb + b'X'\Sigma^{-1}Xb, \\ &\text{subject to} \quad Rb \geq r. \end{aligned} \tag{9}$$

This form will prove useful later, but for now it puts QP (8) into the form of the standard QP:

$$\begin{aligned} &\min_x\; a + c'x + \tfrac{1}{2}x'Qx, \\ &\text{subject to} \quad Fx \geq d. \end{aligned} \tag{10}$$

The dual of QP (10) is

$$\begin{aligned} &\max_\lambda\, \lambda'(d + FQ^{-1}c) - \tfrac{1}{2}\lambda'FQ^{-1}F'\lambda - \tfrac{1}{2}c'Q^{-1}c + a, \\ &\text{subject to} \quad \lambda \geq 0. \end{aligned} \tag{11}$$

Luenberger (1969, ch. 8) and Avriel (1976, ch. 7) discuss the duality theory of quadratic programming necessary for our purposes. If we define QP (8) as the primal problem, then its dual is

$$\max_{\lambda} \lambda'(r - R\hat{b}) - \lambda' A \lambda / 4, \quad \text{subject to} \quad \lambda \geq 0, \tag{12}$$

where

$$A = \left[R(X'\Sigma^{-1}X)^{-1} R' \right].$$

The Kuhn–Tucker test statistic, KT, is the optimal objective function value of QP (12), so that

$$KT = \tilde{\lambda}' A \tilde{\lambda} / 4,$$

where $\tilde{\lambda}$ is the Kuhn–Tucker multiplier defined earlier. By our assumptions the matrix A is positive definite and $(X'\Sigma^{-1}X)^{-1}$ is nonsingular, hence [see Gill, Murray, and Wright (1981) or Avriel (1976) for details] the objective functions of the primal and dual problems evaluated at the optimum $(\tilde{b}, \tilde{\lambda})$ are equal. This gives the first of our equalities: $LR = KT$.

The Wald statistic measures the difference between the unrestricted and restricted estimates of $R\beta$ and is given by

$$W = (R\tilde{b} - R\hat{b})' A^{-1} (R\tilde{b} - R\hat{b}).$$

By eq. (5) we have $R\tilde{b} - R\hat{b} = R(\tilde{b} - \hat{b}) = A\tilde{\lambda}/2$. This implies $W = KT$. Finally, QP (8) can be rewritten in the following form:

$$\min_{b} (b - \hat{b})'(X'\Sigma^{-1}X)(b - \hat{b}), \quad \text{subject to} \quad Rb \geq r. \tag{13}$$

This QP is a special case of the generalized distance statistic derived by Kodde and Palm (1986), following the logic of Perlman (1969). This statistic measures the distance between the unrestricted estimate of β and the set defining the null hypothesis. To see the equivalence of QP (13) to QP (8), expand QP (13) and utilize the definition of $\hat{b}$ to show QP (13) also equals QP (9) and hence QP (8). This process has also shown that $\tilde{b}$ is the solution to QP (13) as well as the ICLS estimate. A final equivalent test statistic is

$$\overline{W} = (\tilde{b} - \hat{b})'(X'\Sigma^{-1}X)(\tilde{b} - \hat{b}).$$

Hence, for the case that Σ is known we conclude that $LR = KT = W = \overline{W}$.

At this point it is convenient to state the three equivalent test statistics for the GHM testing problem H: $R\beta = r$ versus K: $R\beta \geq r$. In our notation they are

$$LR = (y - X\bar{b})'\Sigma^{-1}(y - X\bar{b}) - (y - X\tilde{b})'\Sigma^{-1}(y - \tilde{b}),$$

$$KT = (\bar{\lambda} - \tilde{\lambda})'A(\bar{\lambda} - \tilde{\lambda}),$$

$$W = (R\tilde{b} - r)'A^{-1}(R\tilde{b} - r).$$

These will be used in section 4 when we derive the joint distribution of these statistics and the multivariate inequality constraints test statistics.

4. Distribution of test statistics under null hypothesis

Before proceeding with the derivation of the null distribution of our test statistics we consider the following special case of our testing problem:

$$\begin{aligned} &\text{H}: \mu \geq 0 \quad \text{versus} \quad \text{K}: \mu \in R^P, \\ &\hat{\mu} = \mu + \nu, \quad \nu \text{ is a } (P \times 1) \text{ vector that is } \mathrm{N}(0, \Omega), \end{aligned} \tag{14}$$

and Ω is of full rank P and known.

Perlman (1969) considered a more general version of problem (14) where the set defining the null hypothesis is a closed, convex cone (as opposed to the positive orthant) and the covariance matrix, Ω, is completely unknown. Following his logic, we construct the likelihood ratio statistic for this problem as the optimal value of the objective function from

$$\begin{aligned} &\min_{\mu} (\hat{\mu} - \mu)'\Omega^{-1}(\hat{\mu} - \mu), \\ &\text{subject to} \quad \mu \geq 0. \end{aligned} \tag{15}$$

Let $\tilde{\mu}$ represent the solution to this QP. Define

$$IU = (\hat{\mu} - \tilde{\mu})'\Omega^{-1}(\hat{\mu} - \tilde{\mu}), \tag{16}$$

where IU is shorthand for the null hypothesis of inequality constraints versus an unrestricted alternative.

We must now find the least favorable value of μ under the null hypothesis to use in constructing an exact size test of this null hypothesis. The usual approach proceeds as follows. For this problem, the 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 where μ lies under the null hypothesis. Call this Θ_H. Following Lehmann (1986), 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. Following this logic we construct a rejection region for any size test of (14). A special case of Lemma 8.2 in Perlman (1969) is given below.

Lemma 1. *For any $\mu \geq 0$ and $c \in R_+$, the following is true:*

$$\Pr_{\mu,\Omega}[IU \geq c] \leq \Pr_{0,\Omega}[IU \geq c].$$

The notation $\Pr_{\mu,\Omega}[IU \geq c]$ denotes the probability of the event $[IU \geq c]$ assuming that $\hat{\mu}$ is distributed as $\mathrm{N}(\mu, \Omega)$. This notation will be used throughout the remainder of the paper. As an immediate corollary we have

$$\sup_{\mu \in \Theta_H} \Pr_{\mu,\Omega}[IU \geq c] = \Pr_{0,\Omega}[IU \geq c],$$

so that $\mu = 0$ is the unique least favorable value for μ to specify under the null hypothesis to obtain critical values for any size test.

To relate the inequality constraints test to the much studied multivariate one-sided hypothesis test we now discuss the test

$$\mathrm{H}\colon \mu = 0 \quad \text{versus} \quad \mathrm{K}\colon \mu \geq 0, \tag{17}$$

under the same assumptions on $\hat{\mu}$. We follow Perlman's (1969) presentation of this problem because it is the most flexible for our purposes.

Perlman formulated the likelihood ratio test for (17) as the maximum value of the objective function from the following QP:

$$\begin{aligned} &\max_{\mu} \ \hat{\mu}'\Omega^{-1}\hat{\mu} - (\mu - \hat{\mu})'\Omega^{-1}(\mu - \hat{\mu}), \\ &\text{subject to} \quad \mu \geq 0. \end{aligned} \tag{18}$$

Clearly, the $\tilde{\mu}$ that satisfies (18) is the same $\tilde{\mu}$ which satisfies (15). Define

$$EI = \hat{\mu}'\Omega^{-1}\hat{\mu} - (\tilde{\mu} - \hat{\mu})'\Omega^{-1}(\tilde{\mu} - \hat{\mu}) = \tilde{\mu}'\Omega^{-1}\tilde{\mu},$$

where EI denotes the equality constraints null hypothesis versus the inequality constraints alternative hypothesis. We now state the following theorem proven in the appendix.

Theorem 1. *Given* $\mu = 0$ [*the least favorable value of* μ *for hypothesis test* (14) *and the null value of* μ *for hypothesis test* (17)], *we have the following result*:

$$\Pr_{0,\Omega}[IU \geq c_{IU}, EI \geq c_{EI}]$$

$$= \sum_{k=0}^{P} \Pr[\chi^2_{P-k} \geq c_{IU}] \Pr[\chi^2_k \geq c_{EI}] w(P, k, \Omega).$$

The joint distribution of IU and EI is a weighted sum of two sets of independent chi-squared distributions ranging from zero to P degrees of freedom. A χ^2_k for $k = 0$ is simply a point mass at the origin. By redefining the index variable k, we have the following corollary.

Corollary 1. *For the hypothesis test* H: $\mu \geq 0$ *versus* K: $\mu \in R^P$, *the distribution of the likelihood ratio statistic satisfies*

$$\sup_{\mu \geq 0} \Pr_{\mu,\Omega}[IU \geq c] = \Pr_{0,\Omega}[IU \geq c]$$

$$= \sum_{k=0}^{P} \Pr[\chi^2_k \geq c] w(P, P-k, \Omega).$$

We also have the well-known distribution of the multivariate one-sided test statistic.

Corollary 2. *For the hypothesis test* H: $\mu = 0$ *versus* K: $\mu \geq 0$, *the null distribution of the likelihood ratio statistic is*

$$\Pr_{0,\Omega}[EI \geq c] = \sum_{k=0}^{P} \Pr[\chi^2_k \geq c] w(P, k, \Omega).$$

The weight, $w(P, k, \Omega)$, is the probability that $\tilde{\mu}$ has exactly k positive elements. The sum of the weights from 0 to P is one. These weights depend explicitly on the covariance matrix of $\hat{\mu}$. Closed form solution for the weights are available for the cases when $P \leq 4$ [see Kudo (1963) for these formulae]. Shapiro (1985) provides alternative closed form expressions for these weights for the case that $P = 4$. Both of these authors provide detailed derivations of these formulae. There are various numerical methods available for the cases that $P \geq 5$. Bohrer and Chow (1978) give an algorithm which is designed to calculate these weights up to the case that $P = 10$. However, for $P \geq 8$, the complexity of the calculations could make these numerical methods prohibitively expensive or simply intractable.

A final methodology for computing these weights when $P \geq 8$ is to use Monte Carlo techniques. Here the researcher takes, say 1000, draws from a multivariate normal distribution with mean zero and covariance matrix Ω. For each draw he computes $\tilde{\mu}$ and counts the number of elements of the vector greater than zero. In this case $w(P, k, \Omega)$ is computed as the proportion of the 1000 draws that $\tilde{\mu}$ has exactly k elements greater than zero. This technique has the following advantages. No expensive numerical integration techniques are required. There are no limits on the values of P for which it is applicable. However, the resulting weights are not exact because this is a Monte Carlo technique. Preliminary comparisons of this technique with exact techniques are very encouraging in terms of the degree of agreement with the exact procedure. Nevertheless, what is clear from the discussion of these weights is that their exact calculation for $P \geq 8$ is a major stumbling block to the widespread application, to higher-dimensional problems, of this testing framework.

Fortunately, in many instances the researcher may not need to compute these weights because there are upper and lower bounds on the distributions given in Corollaries 1 and 2.[3] These bounds are utilized in the same manner as the bounds for the Durbin–Watson statistic. Kodde and Palm (1986) derive upper and lower bounds on the null asymptotic distribution of their nonlinear inequality constraints distance test statistic. For the general version of their testing framework, Wolak (1987) shows these bounds are not tight. Nevertheless, these bounds do eliminate the need to compute the weights if the value of the statistic does not fall between the upper and lower critical values. Following the logic of Kodde and Palm (1986) and the results of Perlman (1969, theorem 6.2), who derived bounds on the null distribution of his test statistics, we obtain

$$\alpha = \inf_{\Omega} \mathrm{Pr}_{0,\Omega}[Z \geq c_l] = \tfrac{1}{2}\mathrm{Pr}[\chi_1^2 \geq c_l], \tag{19}$$

and

$$\alpha = \sup_{\Omega} \mathrm{Pr}_{0,\Omega}[Z \geq c_u] = \tfrac{1}{2}\mathrm{Pr}[\chi_{P-1}^2 \geq c_u] + \tfrac{1}{2}\mathrm{Pr}[\chi_P^2 \geq c_u], \tag{20}$$

where Z can be either the IU or EI statistic and α is the size of the hypothesis test. For tests (14) and (17), these bounds are tight because, as stated earlier and different from the case of nonlinear constraints, the parameters of the covariance matrix of $\hat{\mu}$ are assumed to be functionally independent of μ. If the value of the test statistic is less than c_l, the null hypothesis cannot be rejected. If the value of the statistic is greater than c_u, the null hypothesis is rejected. Kodde and Palm (1986) provide a table of c_l and c_u for tests ranging in size

[3] I am grateful to a referee for suggesting these bounds.

from 0.25 to 0.001 and degrees of freedom from 1 to 40. Therefore, the weights must be calculated and the exact critical value obtained only if these bounds yield inconclusive test results.

Given the results of Theorem 1, we now derive the joint distribution of the likelihood ratio statistics from the GHM hypothesis test and multivariate inequality constraints test. Once again our inequality constraints test is complicated by a null hypothesis that does not specify a unique value for β or that it must satisfy a system of linear equalities. The null hypothesis requires that β satisfy a system of linear inequalities. For this reason, we must find a least favorable value of $\beta \in C \equiv \{\beta | R\beta \geq r,\ \beta \in R^K\}$, the set defining the null hypothesis, to compute the exact null distribution of our test statistics for any size test.

First we establish the least favorable value of $\beta \in C$ and then proceed to derive the joint distribution of the GHM and inequality constraints test statistics. Let R^* be a $((K-P)\times K)$ matrix such that when appended to R, the matrix $T=[R'\ R^{*\prime}]$ is of full rank K. There exists an R^* because R is of rank P by assumption. Let $t=[r'\ r^{*\prime}]'$, where r^* is a $((K-P)\times 1)$ vector of known constants. Using the definition of T and t, the $\overline{W}$ statistic becomes

$$\overline{W} = \min_b \left([Tb-t]-[T\hat{b}-t]\right)'T'^{-1}(X'\Sigma^{-1}X)T^{-1}$$

$$\times\left([Tb-t]-[T\hat{b}-t]\right),$$

$$\text{subject to} \quad Rb \geq r.$$

Set $\phi = T\beta - t$, $\hat{f} = T\hat{b} - t$, and $f = Tb - t$. The QP determining the $\overline{W}$ statistic is now

$$W = \min_f (f-\hat{f})'V(\hat{f})^{-1}(f-\hat{f}),$$

$$\text{subject to} \quad f_1 \geq 0,$$

where $V(\hat{f}) = T(X'\Sigma^{-1}X)^{-1}T'$ and f_1 is the first P elements of f. Define ϕ_1 analogously to f_1 so that $\phi = (\phi_1', \phi_2')'$. Because $g(x) \equiv Tx - t$ is a nonsingular affine transformation the testing problem in terms of ϕ is the same as the one in terms of β. The results of Lemma 1 applied to the test H: $\phi_1 \geq 0$ versus K: $\phi_1 \in R^P$ implies that $\phi_1 = 0$ is the least favorable value of ϕ_1 in the set defining the null hypothesis. The function $g(x)$ implies that the β corresponding to $\phi_1 = 0$ is any β^* such that $R\beta^* = r$. Consequently the least favorable value of $\beta \in C$ is any β^* that satisfies the inequality constraints as equalities. We now state the following theorem whose proof follows directly from that of Theorem 1.

Theorem 2. *Given any β^* such that $R\beta^* = r$ (the least favorable value of β for the inequality constraints hypothesis test and the null value of β for the GHM hypothesis test) we have the following result:*

$$\Pr_{\beta^*,(X'\Sigma^{-1}X)^{-1}}[IU \geq c_{IU}, EI \geq c_{EI}]$$

$$= \sum_{k=0}^{P} \Pr[\chi^2_{P-k} \geq c_{IU}] \Pr[\chi^2_k \geq c_{EI}] w(P,k,A),$$

where IU is the LR, KT, W, or $\overline{W}$ statistic and EI is any of the three GHM test statistics.

Redefining the index variable k gives the following corollary:

Corollary 3. *For the hypothesis testing problem H: $R\beta \geq r$ versus K: $\beta \in R^K$, the distribution of IU satisfies the following property:*

$$\sup_{\beta \in C} \Pr_{\beta,(X'\Sigma^{-1}X)^{-1}}[IU \geq c] = \Pr_{\beta^*,(X'\Sigma^{-1}X)^{-1}}[IU \geq c]$$

$$= \sum_{k=0}^{P} \Pr[\chi^2_k \geq c] w(P, P-k, A),$$

where β^* is such that $R\beta^* = r$.

The distribution derived by Kodde and Palm (1986), specialized to the case of linear models and linear inequality constraints, is precisely the above distribution. Consequently, we can apply their bounds on the null distribution discussed earlier directly to this testing problem.

Depending on how we parametrize Σ, the bounds given by (19) and (20) will be slack or tight.[4] If we restrict Σ only to be a $(T \times T)$ positive definite matrix, the bounds on the null distribution are tight. The full column rank of X and the full row rank of R guarantees that, for any $(P \times P)$ positive definite matrix Q and fixed matrices X and R, there exists a positive definite matrix Σ which satisfies $Q = R(X'\Sigma^{-1}X)^{-1}R'$. See Graybill (1969, theorem 6.3.3) for the justification for this claim. Computing the infimum and supremum of the probability of rejection with respect to Σ for the distribution in Corollary 3 is exactly analogous to computing these quantities with respect to Ω in (19) and (20) so that the bounds given by those two equations are tight. However, when Σ is functionally dependent on a fixed, finite number of parameters, as is the case for the models discussed at the start of section 2, these bounds will, in

[4] I would like to thank a referee for bringing up this point.

most cases, be slack for that specific parametric class of disturbance covariance matrices. For example, if the errors follow a stationary AR(1) process, for fixed X and R any positive definite matrix Q cannot be expressed as $R(X'\Sigma^{-1}X)$ for a Σ within the class of covariance matrices derived from this family of disturbance processes. The appendix contains a simple example of the slackness of these bounds when Σ arises from an AR(1) process.

We now show the duality relation claimed to hold between the inequality constraints test in terms of β and the multivariate one-sided test in terms of λ by deriving the null distribution of the inequality constraints test statistics in terms of λ and $\bar{\lambda}$. Recall that the least favorable value of $\beta \in C$ for constructing exact critical values is any β^* such that $R\beta^* = r$. This implies that $\bar{\lambda}$, as defined in eq. (6), is $\mathrm{N}(0, 4A^{-1})$. For other $\beta \in R^K$, the expectation of $\bar{\lambda}$ is λ as defined by eq. (7).

We can transform the QP defining the KT statistic into a QP solely in terms of $\bar{\lambda}$ and λ. First, let $\hat{\nu} = r - R\hat{b}$. Complete the square of the objective function of QP (12) by adding and subtracting $\hat{\nu}'A^{-1}\hat{\nu}$. In the new notation this QP becomes

$$KT = \max_{\lambda} - \left[(\lambda - 2A^{-1}\hat{\nu})'A(\lambda - 2A^{-1}\hat{\nu})\right]/4 + \hat{\nu}'A^{-1}\hat{\nu}, \qquad \text{subject to} \quad \lambda \geq 0. \tag{21}$$

From eq. (6) we note that

$$\bar{\lambda} = 2A^{-1}\hat{\nu} \quad \text{and} \quad \bar{\lambda}'A\bar{\lambda}/4 = \hat{\nu}'A^{-1}\hat{\nu}. \tag{22}$$

Utilizing (6) and (22), we can rewrite (21) as

$$KT = \max_{\lambda} \bar{\lambda}'A\bar{\lambda}/4 - (\lambda - \bar{\lambda})'A(\lambda - \bar{\lambda})/4, \qquad \text{subject to} \quad \lambda \geq 0. \tag{23}$$

In QP (18), if we replace μ by λ and Ω by $4A^{-1}$, it becomes QP (23). Consequently, QP (23) is a likelihood ratio statistic for a multivariate one-sided test in terms of λ. In addition, the solution to this QP is $\tilde{\lambda}$, the KT multiplier vector defined in section 2. Because $R\beta = r$ [which implies $E(\bar{\lambda}) = \lambda = 0$] for an exact size test of our null hypothesis, the inequality constraints test in terms of our dual variables is the null hypothesis that the Lagrange multiplier is zero versus the restricted alternative that it is greater that or equal to zero. For λ as defined in eq. (7), we have the following theorem by an application of Corollary 2.

Theorem 3. For the hypothesis testing problem H: $\lambda = 0$ *versus K*: $\lambda \geq 0$, *the null distribution of IU is*

$$\mathrm{P}_{0,4A^{-1}}[IU \geq c] = \sum_{k=0}^{P} \mathrm{P}[\chi_k^2 \geq c] w(P, k, 4A^{-1}),$$

where IU represents the LR, KT, W, or $\overline{W}$ *statistic.*

Note that the two weighted sum of χ^2-distributions given in Corollary 3 and Theorem 3 are different. The weights in Corollary 3 (the primal approach to the null distribution) depend on A and $P-k$, whereas those in Theorem 3 (the dual approach to the problem) depend on $4A^{-1}$ and k. In the appendix we prove $w(P, k, \Gamma) = w(P, P-k, \sigma\Gamma^{-1})$, $\sigma > 0$ for $k = 0, \ldots, P$. This gives the following result.

Corollary 4. For hypothesis tests of the same size, the inequality constraints test H: $R\beta \geq r$ *versus K*: $\beta \in R^K$ *is equivalent to the multivariate one-sided test H*: $\lambda = 0$ *versus K*: $\lambda \geq 0$.

We can think of this result as an extension of the well-known duality result in testing multivariate equality constraints which states that H: $R\beta = r$ versus K: $R\beta \neq r$ and H: $\lambda = 0$ versus K: $\lambda \neq 0$ are equivalent tests. Wolak (1988) derives the general form of this duality result for multivariate hypothesis tests.

It is natural at this point to discuss the power of these tests. This, as has been discussed by various researchers, is extremely difficult because the weight function, $w(P, k, \Omega)$, is no longer valid for computing the distribution of the test statistic under the alternative hypothesis. For testing problem (14), under the alternative hypothesis, say $\mu < 0$, the probability that $\tilde{\mu}$ has exactly $k \leq P$ positive elements is no longer equal to sums of products of $\mathrm{N}(0, \Omega)$ orthant probabilities [see (A.8) in the appendix], because $\mathrm{E}(\hat{\mu}) = \mu \neq 0$. For this alternative, suppose a $\tilde{\mu}$ is computed that has exactly k positive elements. Partition $\tilde{\mu}$ and μ as is done in the appendix so that $\tilde{\mu} = (\tilde{\mu}_1', \tilde{\mu}_2')'$. Briefly, $\tilde{\mu}_1 \in R^{P-k}$ contains all of the zero elements of $\tilde{\mu}$ and $\tilde{\mu}_2 \in R^k$ contains the remaining elements of $\tilde{\mu}$ which are all positive. As shown in the appendix, the value of the likelihood ratio statistic for testing problem (14) conditional on $\tilde{\mu}$ having exactly these k positive elements is

$$IU = \hat{\mu}_1 \Omega_{11}^{-1} \hat{\mu}_1,$$

where $\hat{\mu}_1$ is $\mathrm{N}(\mu_1, \Omega_{11})$ for this alternative hypothesis. Given this distribution for $\hat{\mu}_1$, from Theorem 1.4.1. of Muirhead (1982) the distribution of IU conditional on $\tilde{\mu}_1 = 0$ is noncentral chi-squared with $P-k$ degrees of freedom and noncentrality parameter $\delta = \mu_1' \Omega_{11}^{-1} \mu_1$. In addition, this noncentrality parameter depends on precisely which k of the elements of $\tilde{\mu}$ are positive; in other words, exactly which elements of μ comprise μ_1. Consequently, much of the symmetry which simplifies the derivation of the null distribution when $\mu = 0$ is lost. This complication becomes even more problematic for higher dimensions because there are numerous ways $\tilde{\mu}$ can have exactly k positive elements, and a different noncentral chi-squared distribution may arise if, for instance, the first k elements of $\tilde{\mu}$ are positive than if any other k elements of $\tilde{\mu}$ are positive. These complications make power calculations theoretically possible but computationally intractable for nearly all types of testing problems and alternatives with $P > 2$.

Power calculations for certain alternatives for the case that the matrix $(X'\Sigma^{-1}X) = \sigma^2 I$ (I is the identity matrix), the null hypothesis is $\beta \geq 0$, and $\beta \in R^2$ are reported in Yancey, Bohrer, and Judge (1982). Although the condition $(X'\Sigma^{-1}X) = \sigma^2 I$ rarely if ever occurs in practice, these calculations help to provide intuition about the power properties of these type tests. The power of this test, for testing our inequality constraints null hypothesis, is greater, for all alternatives in the direction $R\beta \leq r$, than the two-sided test, H: $R\beta = r$ versus K: $R\beta \neq r$ because it takes into account the fact, in our case, that λ as defined by (7) is always greater than zero under the alternative. Bartholomew (1961), Barlow, Bartholomew, Bremner, and Brunk (1972), and Hillier (1986) discuss the power properties of the multivariate one-sided test. In particular, Hillier (1986) 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 for the linear regression model. Based on his results, the likelihood ratio test is preferred. Power calculations for the multivariate one-sided test are relevant for discussing the power properties of our test procedure by Corollary 4 which states that for any multivariate linear inequality constraints test there is an equivalent multivariate one-sided test. See Wolak (1988) for more on the exact relation between these two test procedures and its implications for power calculations.

We now list some of the properties possessed by our test. As stated in Perlman (1969) for all testing problems of his general class (see section 1 for the definition of this class), the power of the test approaches one uniformly in Σ and β as the distance, in the norm of the covariance matrix of $\hat{b}$, between β and where it lies under the null hypothesis tends to infinity. The test is biased. The least favorable distribution is obtained at a β such that $R\beta = r$, so the power is smaller for values of β elsewhere in C. By continuity of the power function of the test statistic, there are values of β not in C where the power is smaller than when β is such that $R\beta = r$. However, our test is consistent both when Σ is known and unknown but consistently estimated. This can be seen by the following logic. Make the standard assumption that $\lim_{T\to\infty} T^{-1}(X'\Sigma^{-1}X) \to Q$, a positive definite matrix [Theil (1971, p. 398)]. Then under both the null and alternative hypotheses, $T^{1/2}(\hat{b} - \beta)$ converges in distribution to $\mathrm{N}(0, Q^{-1})$. Consider the case that $R\beta \not\leq r$. We have the following: the probability limit of (plim) $\hat{b} = \beta$ and plim $\tilde{b} \neq \beta$. Therefore we have plim$(\hat{b} - \tilde{b}) \neq 0$. Let c denote the critical value for a size α test of our null hypothesis. Utilizing the $\overline{W}$ form of our statistics we obtain

$$\lim_{T \to \infty} \mathrm{P}_{\beta,(X'\Sigma^{-1}X)^{-1}}\left[(\hat{b} - \tilde{b})'\left[T^{-1}(X'\Sigma^{-1}X)\right](\hat{b} - \tilde{b}) \geq c/T\right] = 1. \tag{24}$$

Because $[T^{-1}(X'\Sigma^{-1}X)]$ converges to a positive definite matrix Q, we know that $(\hat{b} - \tilde{b})'[T^{-1}(X'\Sigma^{-1}X)](\hat{b} - \tilde{b})$ converges in probability to some quantity strictly greater than zero. Consistency of the test follows because $c/T \to 0$ as $T \to \infty$. However, if $R\beta \geq r$, then we have plim $\tilde{b} = \beta$, plim$(\hat{b} - \tilde{b}) = 0$, and the above construction is not possible. To show consistency of the test for the case that Σ is unknown but estimated consistently we replace Σ by its consistent estimate and note that as $T \to \infty$ this estimate converges in probability to Σ. In this case the distribution in (24) is the asymptotic distribution of our test statistics. Besides the usual regularity condition necessary for the asymptotic normality of the GLS estimate of β calculated using a consistent estimate of Σ [Theil (1971, p. 399)], the above construction does not change.

5. Test statistics when error covariance matrix is unknown

We now consider the case that Σ is unknown but depends on a finite number of parameters. The parameters of Σ cannot be functionally dependent on β or the monotoncity property of the power function crucial to deriving an asymptotically exact null distribution of the test statistics will fail. A condition of Perlman's (1969) Lemma 8.2 (a more general version of our Lemma 1), which requires that $(X'\Sigma^{-1}X)^{-1}$ not vary as β varies over the parameter space, is violated.[5] This assumption is also used to construct the conditional likelihood ratio statistics necessary to prove the small sample inequalities for our statistics when Σ is unknown [see Savin (1976, pp. 1314–1315)]. Consequently, these inequalities hold for all models that can be written in the form of a linear regression model with a general covariance matrix of the disturbances. See the first part of section 2 for a list of these models.

We now illustrate these small-sample inequalities which hold between the three inequality constraints test statistics. We assume the existence of unique unrestricted ML and restricted ML estimates of β and Σ, which we denote

[5]This restriction is not troublesome for extending our results to an asymptotic framework for the case of linear econometric models and linear constraints. The extension to the case of nonlinear inequality constraints in the context of nonlinear econometric models violates this assumption [Wolak (1987)]. Wolak (1989) derives a large sample methodology for testing nonlinear inequality constraints in this nonlinear model context which circumvents this assumption.

$\hat{b}_u$, $\hat{\Sigma}$ and $\tilde{b}_r$, $\tilde{\Sigma}$, respectively.[6] Throughout the remainder of this section a hat denotes an unrestricted estimate. An estimate calculated under our restriction $Rb \geq r$ is denoted by a tilde. The subscripts on the estimate denote which estimate of Σ was used to calculate the estimate of β. A subscript u indicates the estimate was based on $\hat{\Sigma}$ and a subscript r implies the estimate is based on $\tilde{\Sigma}$. Hence $\tilde{b}_u$ is the constrained ML estimate of β calculated using $\hat{\Sigma}$ and $\hat{b}_r$ is the unconstrained ML estimate computed using $\tilde{\Sigma}$. This follows the conventions of Breusch (1979).

We can define residuals for the four estimators of β as follows:

$$\hat{\varepsilon}_j = y - X\hat{b}_j \quad \text{and} \quad \tilde{\varepsilon}_j = y - X\tilde{b}_j, \qquad j = u, r.$$

Let

$$\tilde{A} = \left[R(X'\tilde{\Sigma}^{-1}X)^{-1}R'\right] \quad \text{and} \quad \hat{A} = \left[R(X'\hat{\Sigma}^{-1}X)^{-1}R'\right].$$

The likelihood function takes the standard form:

$$L(b, \Sigma) = (2\pi)^{-T/2}\det(\Sigma)^{-1/2}\exp\left[-\tfrac{1}{2}(y - Xb)'\Sigma^{-1}(y - Xb)\right].$$

The Wald statistic traditionally tests the amount the unrestricted estimate differs from the hypothesized restrictions and so it is based on the unrestricted estimate of Σ. This statistic is defined as

$$W = \left(R\tilde{b}_u - R\hat{b}_u\right)'\hat{A}^{-1}\left(R\tilde{b}_u - R\hat{b}_u\right).$$

In light of our additional notation eq. (5) becomes

$$\tilde{b}_j - \hat{b}_j = (X'\Sigma^{-1}X)^{-1}R'\tilde{\lambda}_j/2, \qquad j = u, r, \tag{25}$$

where $\Sigma = \hat{\Sigma}$ if $j = u$ and $\Sigma = \tilde{\Sigma}$ if $j = r$. Multiplying both sides by R and solving for $\tilde{\lambda}_j$ yields

$$\tilde{\lambda}_j = 2A^{-1}\left(R\tilde{b}_j - R\hat{b}_j\right), \tag{26}$$

where $A = \hat{A}$ if $j = u$ and $A = \tilde{A}$ if $j = r$. Eqs. (25) and (26) give the following:

$$\tilde{b}_u = \hat{b}_u + (X'\hat{\Sigma}^{-1}X)^{-1}R'\hat{A}^{-1}\left(R\tilde{b}_u - R\hat{b}_u\right).$$

In terms of residuals this implies

$$\tilde{\varepsilon}_u = \hat{\varepsilon}_u - X(X'\hat{\Sigma}^{-1}X)^{-1}R'\hat{A}^{-1}\left(R\tilde{b}_u - R\hat{b}_u\right).$$

[6] We should note that in some cases there may be some difficulty in finding $\tilde{b}_r$ and $\tilde{\Sigma}$ due to multiple solutions to the likelihood equation or failure of the usual iterative methods used for their calculation.

Recall the following:

$$X'\hat{\Sigma}^{-1}\hat{\varepsilon}_u = X'\hat{\Sigma}^{-1}\left(I - X\left(X'\hat{\Sigma}^{-1}X\right)^{-1}X'\hat{\Sigma}^{-1}\right)y = 0.$$

This allows us to write:

$$\tilde{\varepsilon}_u'\hat{\Sigma}^{-1}\tilde{\varepsilon}_u = \hat{\varepsilon}_u'\hat{\Sigma}^{-1}\hat{\varepsilon}_u + \left(R\tilde{b}_u - R\hat{b}_u\right)'\hat{A}^{-1}\left(R\tilde{b}_u - R\hat{b}_u\right).$$

Rearranging, we note this implies

$$\begin{aligned} W &= \tilde{\varepsilon}_u'\hat{\Sigma}^{-1}\tilde{\varepsilon}_u - \hat{\varepsilon}_u'\hat{\Sigma}^{-1}\hat{\varepsilon}_u \\ &= -2\log\left[L\left(\tilde{b}_u, \hat{\Sigma}\right)/L\left(\hat{b}_u, \hat{\Sigma}\right)\right] \\ &= -2\log\left[\sup_{Rb\geq r} L\left(b|\hat{\Sigma}\right)/\sup_{b} L\left(b|\hat{\Sigma}\right)\right]. \end{aligned} \tag{27}$$

The KT statistic tests the significance from zero of the Kuhn–Tucker multiplier and is based on the restricted estimate of Σ. We define it as follows:

$$KT = \tilde{\lambda}_r'\tilde{A}\tilde{\lambda}_r/4.$$

Eq. (25) implies that

$$\tilde{b}_r = \hat{b}_r + \left(X'\tilde{\Sigma}^{-1}X\right)^{-1}R'\tilde{\lambda}_r/2.$$

In terms of residuals we have

$$\tilde{\varepsilon}_r = \hat{\varepsilon}_r - X\left(X'\tilde{\Sigma}^{-1}X\right)^{-1}R'\tilde{\lambda}_r/2.$$

Once again $X'\tilde{\Sigma}^{-1}\hat{\varepsilon}_r = 0$, so that

$$\tilde{\varepsilon}_r'\tilde{\Sigma}^{-1}\tilde{\varepsilon}_r = \hat{\varepsilon}_r'\tilde{\Sigma}^{-1}\hat{\varepsilon}_r + \tilde{\lambda}_r'\tilde{A}\tilde{\lambda}_r/4.$$

This gives us the following set of equations:

$$\begin{aligned} KT &= \tilde{\varepsilon}_r'\tilde{\Sigma}^{-1}\tilde{\varepsilon}_r - \hat{\varepsilon}_r'\tilde{\Sigma}^{-1}\hat{\varepsilon}_r \\ &= -2\log\left[L\left(\tilde{b}_r, \tilde{\Sigma}\right)/L\left(\hat{b}_r, \tilde{\Sigma}\right)\right] \\ &= -2\log\left[\sup_{Rb\geq r} L\left(b|\tilde{\Sigma}\right)\Big/\sup_{b} L\left(b|\tilde{\Sigma}\right)\right]. \end{aligned} \tag{28}$$

Finally, the LR statistic is given by

$$LR = -2\log\left[\sup_{Rb\geq r,\Sigma} L(b,\Sigma)/\sup_{b,\Sigma} L(b,\Sigma)\right]. \tag{29}$$

We now recall some inequalities from Breusch (1979), modified for our problem:

$$\sup_{b} L(b|\tilde{\Sigma}) \leq \sup_{b,\Sigma} L(b,\Sigma) = \sup_{b} L(b|\hat{\Sigma}),$$

and

$$\sup_{Rb\geq r} L(b|\hat{\Sigma}) \leq \sup_{Rb\geq r,\Sigma} L(b,\Sigma) = \sup_{Rb\geq r} L(b|\tilde{\Sigma}).$$

Viewing eqs. (27), (28), and (29) in light of these two sets of inequalities we obtain: $W \geq LR \geq KT$. This ordering agrees with that derived by Breusch (1979) and Berndt and Savin (1977) for their equality constraints statistics. Our fourth statistic, $\overline{W}$, is equal to the W statistic if the unconstrained estimate of Σ is used in its computation. It is equal to the KT statistic if the constrained estimate of Σ is used.

We now consider the asymptotic distribution of our three test statistics under the null hypothesis. We assume that $\tilde{\Sigma}$ and $\hat{\Sigma}$ are consistent estimates of Σ and that the usual regularity conditions necessary for the asymptotic normality of the GLS estimate of β using an estimated covariance matrix of ε hold [Theil (1971, p. 399)]. By an application of Slutsky's theorem these statistics all have the same asymptotic distribution as the test statistics evaluated with the true covariance matrix Σ. As discussed in GHM (1982), the same logic holds for the GHM statistics computed with a consistent estimate of Σ. Thus all of the results of section 4 carry through asymptotically.

6. Application to demand analysis

In this section we apply our statistics to a hypothesis testing problem which arises frequently in single-equation demand analysis. Consider the following simple double-log demand function:[7]

$$\begin{aligned}\ln Q_t = \alpha &+ \beta_1 \ln PE_t + \beta_2 \ln PG_t + \beta_3 \ln Y_t \\ &+ \beta_4 D1_t + \beta_5 D2_t + \beta_6 D3_t + \varepsilon_t,\end{aligned} \tag{30}$$

where $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$ and η_t is $\mathrm{N}(0,\sigma^2)$. If we write (30) in matrix notation as

[7]This model was used to illustrate an application of our test procedure. There may be other problems with its specification which may lead us to reject it as the proper model for aggregate residential electricity demand.

$y = X\beta + \varepsilon$, then ε is $\mathrm{N}(0, \sigma^2 V(\rho))$, where $V(\rho)$ is as defined in Theil (1971, p. 252). For our data set, Q_t is aggregate electricity demand per capita for the residential sector, PE_t is the average price of electricity to the residential sector, PG_t is the price of natural gas to the residential sector, and Y_t is income per capita.[8] Seasonal dummy variables $D1_t$, $D2_t$, and $D3_t$ were included because the data from which our demand function was estimated are available on a quarterly basis for a total of 88 observations from the beginning of 1961 to the end of 1983.

We are interested in testing the null hypothesis that the subvector $(-\beta_1, \beta_2, \beta_3)'$ is greater than or equal to zero. For model (30) in matrix notation, our matrix R is a (3×7) matrix with zeros everywhere except for -1 at the $(2,2)$ position and 1 at both the $(3,3)$ and $(4,4)$ positions. Our vector r is equal to zero.

Our technique for calculating the inequality restricted estimator differs from the standard QP algorithm because of the very simple structure of our constraints. We run all possible regressions which occur by zeroing out each one of our three variables being tested until we find the regression yielding the lowest weighted sum of squared residuals as well as satisfying our inequality constraints. If that procedure fails to yield estimates satisfying our restrictions, we perform regressions zeroing out pairs of the variables under consideration until we find the minimal weighted sum of squares residuals regression satisfying the constraints. For hypotheses of this low a dimensionality and simplicity this technique is a very easy way of solving the QP necessary to obtain our ICLS estimates using most any available econometric software package.

Table 1 contains the unconstrained and inequality constrained estimates for eq. (30). Standard errors for the estimates are in parentheses below the coefficient estimates. Both of the model estimations yield consistent estimates of σ^2 and ρ which are used to construct the consistent estimates of $\sigma^2 V(\rho)$ necessary to compute our test statistics. Table 2 contains the W, LR, and KT test statistics, which obey the inequalities calculated in section 5.

As discussed in Kudo (1963) and in the appendix, the weights used to calculate the null distribution of our test statistics depend on the correlation coefficients and partial correlation coefficients of $[R(X'\Sigma^{-1}X)^{-1}R']$. Utilizing the three correlation coefficients (ρ_{12}, ρ_{13}, and ρ_{23}) and the three partial correlation coefficients ($\rho_{12.3}$, $\rho_{13.2}$, and $\rho_{23.1}$) from the unconstrained covariance matrix, $[R(X'\hat{\Sigma}^{-1}X)^{-1}R']$, we calculate the weights that enter the null distribution of our test statistics. We utilize the closed form expressions for the weights given in Kudo (1963). Under both the null and alternative hypotheses these estimates of the correlation coefficients and partial correlation coefficients are consistent.

[8]A less up-to-date version of this data set was used by Sutherland (1983). This paper contains a detailed discussion of the sources for this data set.

Table 1

Unconstrained and constrained model.

α	β_1	β_2	β_3	β_4	β_5	β_6	ρ
			Unconstrained model				
−0.918	−0.115	−0.179	1.811	0.113	−0.048	0.143	0.667
(0.685)	(0.127)	(0.079)	(0.090)	(0.012)	(0.012)	(0.018)	(0.082)
			Constrained model ($\beta_2 = 0$)				
0.205	−0.289	0.000	1.666	0.121	−0.049	0.109	0.653
(0.464)	(0.102)	(0.000)	(0.062)	(0.012)	(0.012)	(0.011)	(0.082)

Table 2

Test statistics.

Wald	Likelihood ratio	Kuhn–Tucker
6.321	5.581	4.448

The level α critical value for our test statistic is the solution of the following equation in c:

$$\sum_{k=0}^{3} w(3,3-k,\Omega)\Phi_k(c) = 1-\alpha, \tag{31}$$

where $\Phi_k(c)$ is the distribution function for a chi-squared variable with k degrees of freedom and $\Omega = [R(X'\hat{\Sigma}^{-1}X)^{-1}R']$. Eq. (31) can be solved via any numerical method for finding the zeros of a univariate function.

There is another methodology for performing this hypothesis test. We simply calculate the probability that a random variable with the null distribution of our test statistics takes on a value greater than the test statistic. In other words, if $G(x)$ is the distribution function of our statistics under the null hypothesis, we calculate

$$1 - G(X) = \sum_{k=1}^{3} w(3,3-k,\Omega)\Pr(\chi_k^2 \geq X),$$

where X is either the W, LR, or KT statistic. The above summation begins with $k=1$ because $\Pr(\chi_0^2 \geq \delta) = 0$ for all $\delta > 0$. To calculate $\Pr(\chi_k^2 \geq X)$ we can either utilize a numerical integration procedure for the chi-squared distribution function or simply interpolate the relevant probabilities utilizing the tables for the chi-squared distribution. Table 3 contains the values for the

Table 3

Weights and critical values.

	$k=0$	$k=1$	$k=2$	$k=3$
Weight	0.068	0.347	0.432	0.153
$\Pr(\chi_k^2 \geq W)$	0.000	0.012	0.042	0.099
$\Pr(\chi_k^2 \geq LR)$	0.000	0.018	0.061	0.134
$\Pr(\chi_k^2 \geq KT)$	0.000	0.035	0.108	0.217
$1-G(W)=0.037$				$1-G(LR)=0.053$
$1-G(KT)=0.092$				Critical value = 5.714
$c_l=2.706$				$c_u=7.045$

weights and $\Pr(\chi_k^2 \geq X)$ and $1-G(X)$ for each test statistic. The critical value for our hypothesis test at a 0.05 level of significance and the same size values for c_l and c_u obtained from Kodde and Palm (1986) are also in the table. Note that all of our statistics are greater than c_l and less than c_u so that we must compute the asymptotically exact critical value.

Comparing our test statistics to the critical value, we find that the null hypothesis cannot be rejected at a 0.05 level for the *LR* and *KT* test statistics. As expected, the probability values, $1-G(X)$, confirm the hypothesis test results. The results of the *LR* and *KT* tests conflict with a Bonferroni-type approach to testing our null hypothesis based on a combination of individual asymptotic t-tests applied to the unconstrained model. Suppose we performed the three one-sided t-tests H: $\beta_i \geq 0$ versus K: $\beta_i < 0$, $i=2,3$, and H: $\beta_1 \leq 0$ versus K: $\beta_1 > 0$ for the unconstrained model such that $\sum_{i=1}^{3}\alpha_i = 0.05$, where α_i is the size of the ith individual t-test. We would reject the null in favor of the alternative at a 0.05 level of significance for any allocation of the overall level of significance to the three individual tests such that $\alpha_2 > 0.02$. This anomaly is due to the high degree of correlation between the three estimated parameters under examination. This causes the standard Bonferroni bound for computing the overall level of significance to be quite slack. This high degree of correlation is the major shortcoming associated with applying this Bonferroni-type approach to testing multiple inequality constraints.

Given that we accepted the inequality constraints null hypothesis implied by economic theory for the *KT* and *LR* statistics and that we rejected the standard two-sided null hypothesis that these parameters are jointly zero versus the unrestricted alternative (this test was performed but not reported), following the logic of Hogg (1961), we can test the null hypothesis that $\beta_i = 0$, $i=1,2,3$, versus the restricted alternative that $\beta_1 \leq 0$ and $\beta_i \geq 0$, $i=2,3$ (one of these weak inequalities strict). If we reject this null hypothesis in favor of the restricted alternative, this outcome, in combination with the other test results, provides further evidence that the hypothesized inequality constraints hold strictly for at least one element of the parameter vector under examination. We should emphasize that rejecting the point null hypothesis, $\beta_i = 0$, $i = 1, 2, 3$, in favor of the restricted alternative alone provides little evidence for the validity of imposing inequality constraints on the parameters. This test result in isolation says that imposing inequality constraints on the parameters is significantly less objectionable, statistically, than imposing these same constraints as equalities. If we want to test the validity of inequality constraints we should test them as a null hypothesis. If this null hypothesis is accepted, proceeding to apply the GHM test to determine whether the constraints hold as equalities versus inequalities allows us to further distinguish where the true parameter vector lies. Going this next step eliminates the set of β such that $R\beta = r$ from the set of possible values of β which caused us to accept the inequality constraints null hypothesis. And as discussed by Hogg (1961), these two hypothesis tests are independent so that exact overall significance levels can be easily computed. Following this procedure, we computed the three GHM statistics given at the end of section 2. For each of the statistics the null hypothesis that these coefficients are jointly zero was rejected in favor of the hypothesized inequality constraints alternative at a 0.05 level of significance. This discussion and our testing results illustrate why the GHM test and multivariate inequality constraints test are useful for different purposes. For testing inequality constraints, our recommendation is if the inequality constraints null hypothesis is accepted, the independent GHM test procedure can then be used to determine if it was because $R\beta = r$ or because at least one of the weak inequalities in $R\beta \geq r$ holds as a strict inequality.

7. Conclusions and extensions

In this paper we devised a methodology for testing general linear inequality restrictions within the context of linear econometric models. As noted earlier our results can be extended to the linear simultaneous equations model. The distance test approach taken by Kodde and Palm (1986) is available for simultaneous equations estimators by computing the $\overline{W}$ statistic replacing $(X'\Sigma^{-1}X)$ in (13) by the appropriate estimate of the inverse of the asymptotic covariance matrix of parameter estimate. In the notation of Theil (1971, p. 509) the $\overline{W}$ statistic for the 3SLS model is

$$\min_{\delta}\,(\delta - \hat{\delta})'\left[Z'\left(\Sigma^{-1} \otimes X(X'X)^{-1}X'\right)Z\right](\delta - \hat{\delta}),$$

$$\text{subject to} \quad R\delta \geq r,$$

where Σ is replaced by a consistent estimate. The unconstrained 3SLS estimate is denoted by $\hat{\delta}$ and can be calculated utilizing any unrestricted consistent estimate of Σ which we denote by $\hat{\Sigma}$.

The standard methodology for calculating the LR, W, and KT statistics discussed in section 3 can be modified to apply to linear simultaneous equations estimators. In the same notation, the weighted sum of squared residuals is defined as

$$S(\delta) = (y - Z\delta)'\left(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X'\right)(y - Z\delta).$$

The inequality constrained 3SLS estimator, $\tilde{\delta}$, is the solution to

$$\min_{\delta} S(\delta) \quad \text{subject to} \quad R\delta \geq r.$$

Let $\tilde{\lambda}$ be the Kuhn–Tucker multiplier vector associated with the constraints $R\delta \geq r$. The likelihood ratio statistic for the 3SLS estimate is the difference between the weighted sum of squared residuals from the two model fits:

$$LR = S(\tilde{\delta}) - S(\hat{\delta}).$$

The same estimate of Σ should be used to compute $S(\tilde{\delta})$ and $S(\hat{\delta})$ in constructing this statistic. The Wald statistic is

$$W = (\hat{\delta} - \tilde{\delta})'\left[Z'\left(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X'\right)Z\right](\hat{\delta} - \tilde{\delta}).$$

The KT statistic is

$$KT = \tilde{\lambda}'\left(R\left[Z'\left(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X'\right)Z\right]^{-1}R'\right)\tilde{\lambda}/4.$$

These three statistics are asymptotically equivalent. Their null distribution for any asymptotically exact size test of H: $R\delta \geq r$ versus K: $\delta \in R^K$ is the weighted sum of chi-squared distributions given in Corollary 3 with the matrix A consistently estimated by $(R[Z'(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X')Z]^{-1}R')$. The W, KT, and LR statistics for the 2SLS model can be constructed and the null asymptotic distribution obtained as a special case of this procedure.

Appendix

A.1. Proof of Theorem 1

First we state and prove the following lemma:

Lemma A.1. *For any $\hat{\mu}, \tilde{\mu} \in R^P$ the following statements are true:*

(i) $\tilde{\mu}'\Omega^{-1}(\hat{\mu} - \tilde{\mu}) = 0$,

(ii) $\hat{\mu}'\Omega^{-1}\hat{\mu} - \tilde{\mu}'\Omega^{-1}\tilde{\mu} = (\hat{\mu} - \tilde{\mu})'\Omega^{-1}(\hat{\mu} - \tilde{\mu})$,

(iii) *either* $\tilde{\mu}_i = 0$ *and* $[-\Omega^{-1}(\hat{\mu} - \tilde{\mu})]_i \geq 0$ *or* $\tilde{\mu}_i > 0$ *and* $[-\Omega^{-1}(\hat{\mu} - \tilde{\mu})]_i = 0$, *for all* $i = 1, \ldots, P$.

Proof. Let $f(\mu)$ denote the objective function of QP (15). We can establish (i) by recalling the Kuhn–Tucker conditions for this QP:

$$\frac{\partial f(\tilde{\mu})}{\partial \mu} = -2\Omega^{-1}(\hat{\mu} - \tilde{\mu}) \geq 0,$$

and

$$\tilde{\mu}' \frac{\partial f(\tilde{\mu})}{\partial \mu} = -2\tilde{\mu}'\Omega^{-1}(\hat{\mu} - \tilde{\mu}) = 0. \tag{A.1}$$

Rearranging the second equation in (A.1) yields: $\tilde{\mu}'\Omega^{-1}\hat{\mu} = \tilde{\mu}'\Omega^{-1}\tilde{\mu}$. Substituting this equality back into the objective function of QP (15) gives (ii). We obtain (iii) from both equations in (A.1) and the restriction that $\tilde{\mu} \geq 0$. □

By (ii) of Lemma A.1 we have the following relationship between the two test statistics:

$$EI = \tilde{\mu}'\Omega^{-1}\tilde{\mu} = \hat{\mu}'\Omega^{-1}\hat{\mu} - IU, \tag{A.2}$$

because

$$IU = (\tilde{\mu} - \hat{\mu})'\Omega^{-1}(\tilde{\mu} - \hat{\mu}).$$

Suppose that $\tilde{\mu}$ has k positive elements. Partition μ, $\hat{\mu}$, and $\tilde{\mu}$ such that $\tilde{\mu}' = (\tilde{\mu}_1', \tilde{\mu}_2')$, where $\tilde{\mu}_1 = 0$ and $\tilde{\mu}_2 > 0$. For this $\tilde{\mu}$ we have $\mu_1 \in R^{P-k}$ and $\mu_2 \in R^k$. Let $\Omega^{-1} \equiv C$. Partition C and Ω conformably to μ so that

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

where $C_{12}' = C_{21}$. Part (iii) of Lemma A.1 allows us to write the following set of equations:

$$-C_{11}(\hat{\mu}_1 - \tilde{\mu}_1) - C_{12}(\hat{\mu}_2 - \tilde{\mu}_2) \geq 0, \tag{A.3a}$$

$$-C_{21}(\hat{\mu}_1 - \tilde{\mu}_1) - C_{22}(\hat{\mu}_2 - \tilde{\mu}_2) = 0. \tag{A.3b}$$

Using $\tilde{\mu}_1 = 0$, we can solve (A.3b) for $\tilde{\mu}_2$ as follows:

$$\tilde{\mu}_2 = \hat{\mu}_2 + C_{22}^{-1}C_{21}\hat{\mu}_1 = \hat{\mu}_2 - \Omega_{21}\Omega_{11}^{-1}\hat{\mu}_1. \tag{A.4}$$

Utilizing (ii) of Lemma A.1 and multiplying out, we obtain

$$\begin{aligned} IU &= (\hat{\mu} - \tilde{\mu})'C(\hat{\mu} - \tilde{\mu}) \\ &= \hat{\mu}_1'C_{11}\hat{\mu}_1 + 2\hat{\mu}_1'C_{12}\hat{\mu}_2 + \hat{\mu}_2'C_{22}\hat{\mu}_2 - \tilde{\mu}_2'C_{22}\tilde{\mu}_2. \end{aligned} \tag{A.5}$$

By (A.4), we have

$$\tilde{\mu}_2'C_{22}\tilde{\mu}_2 = \hat{\mu}_2'C_{22}\hat{\mu}_2 + 2\hat{\mu}_1'C_{12}\hat{\mu}_2 + \hat{\mu}_1'C_{12}C_{22}^{-1}C_{21}\hat{\mu}_1.$$

Substituting this expression into (A.5) gives

$$IU = \hat{\mu}_1'\left[C_{11} - C_{12}C_{22}^{-1}C_{21}\right]\hat{\mu}_1 = \hat{\mu}_1'\Omega_{11}^{-1}\hat{\mu}_1, \tag{A.6}$$

where the second equality follows from the partitioned matrix inversion lemma. Because $\hat{\mu}_1$ is $\mathrm{N}(0, \Omega_{11})$, conditional on $\tilde{\mu}$ having k positive elements IU is distributed as a χ^2_{P-k}. By Theorem 1.4.3 of Muirhead (1982) and eqs. (A.6) and (A.2), the distribution of EI conditional on $\tilde{\mu}$ having k positive elements is χ^2_k. In addition, by straightforward but tedious application of the results of Problem 1.23 in Muirhead (1982) these two chi-squared distributions are independent.

Because $\mathrm{E}(\hat{\mu}) = 0$, this joint conditional distribution depends only on the number of positive elements of $\tilde{\mu}$ and not which specific elements of it are positive. Consequently, we can repeat this same procedure for all partitionings of $\tilde{\mu}$ such that the number of nonzero elements of $\tilde{\mu}$ is k. This yields the $(k+1)$th term of the joint distribution:

$$\Pr\left(\chi^2_{P-k} \geq c_{IU}\right)\Pr\left(\chi^2_k \geq c_{EI}\right)w(P, k, \Omega).$$

This term is the joint distribution of IU and EI conditional on the event $\{\tilde{\mu}|$ k elements of $\tilde{\mu}$ are positive$\}$, multiplied by the probability of that event, $w(P,k,\Omega)$.

Repeating this procedure for all k from zero to P yields the following joint distribution:

$$\sum_{k=0}^{P} \Pr\left(\chi^2_{P-k} \geq c_{IU} -\right)\Pr\left(\chi^2_k \geq c_{EI}\right)w(P, k, \Omega). \quad \square$$

A.2. Proof of relationship between weights

We claim that $w(p, k, \Omega) = w(p, p-k, \sigma\Omega^{-1})$, $\sigma > 0$ for $k = 0, \ldots, p$, where p is the dimension of the $\mathrm{N}(0, \Omega)$ vector, $\hat{\mu}$. First some results from multivariate normal distribution theory. If the vector $\hat{\mu}' = (\hat{\mu}_1', \hat{\mu}_2')$, is $\mathrm{N}(0, \Omega)$, then the conditional distribution of $\hat{\mu}_1$ given $\tilde{\mu}_2 = 0$ is $\mathrm{N}(0, [\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}])$, where Ω is partitioned conformably to $\hat{\mu}$.

For the sake of brevity, in our proof we use eq. (3.41) from Kudo (1963) and the notation relevant to his problem. He defines the null distribution of the statistic used to test H: $\mu = 0$ versus K: $\mu \geq 0$ in the following fashion:

$$\mathrm{P}\left(\bar{\chi}^2 \geq \bar{\chi}_0^2\right) = \sum_{\varnothing \subseteq M \subseteq P} \mathrm{P}\left(\chi^2_{n(M)} \geq \bar{\chi}_0^2\right)\mathrm{P}\left([\Omega_{M'}]^{-1}\right)\mathrm{P}\left(\Omega_{M:M'}\right), \tag{A.7}$$

where the summation runs over all of the 2^p subsets of $P = \{1, \dots, p\}$ (P = the set of integers from one to p), which are indexed by M. Kudo defines $n(M)$ as the number of elements in M and M' as the complement of M relative to P. The matrix Ω_M is the covariance matrix of the vector composed of elements of $\hat{\mu}$ given by $\hat{\mu}_i$ such that $i \in M$. The matrix $\Omega_{M;M'}$ is the covariance matrix of the vector composed of the $\hat{\mu}_i$, $i \in M$, conditional on the remainder of the elements of $\hat{\mu}$, $\hat{\mu}_j$, j not an element of M ($j \in M'$), equal to zero. $\mathrm{P}(\Sigma)$ is the probability that a multivariate normal random vector with mean zero and covariance matrix Σ has all positive elements. In this notation, $\chi^2_{n(M)}$ has a χ^2-distribution with $n(M)$ degrees of freedom, where $\chi^2_{n(\varnothing)}$ is a point mass at zero ($\varnothing$ denotes the null set). $\mathrm{P}(\Omega_{\varnothing,P}) = 1$ and $\mathrm{P}([\Omega_{P'}]^{-1}) = \mathrm{P}([\Omega_{\varnothing}]^{-1}) = 1$. Kudo's likelihood ratio statistic, our statistic EI, is denoted by $\bar{\chi}^2$ and $\bar{\chi}^2_0$ is a constant greater than or equal to zero. Each one of the subsets M of P corresponds to one of the 2^p disjoint subsets of the positive orthant in p-dimensional space where $\tilde{\mu}$ can lie. For a given set M, the $\tilde{\mu}$ that lies in it is given as follows:

$$\tilde{\mu}_i = 0 \text{ if } i \notin M \quad \text{and} \quad \tilde{\mu}_i > 0 \text{ if } i \in M.$$

From (A.7), the expression for the weight, $w(p, k, \Omega)$, is

$$w(p,k,\Omega) = \sum_{\varnothing \subseteq M \subseteq P} \mathrm{P}\big([\Omega_{M'}]^{-1}\big)\,\mathrm{P}\big(\Omega_{M;M'}\big) \quad \text{such that} \quad n(M) = k. \tag{A.8}$$

Because we are evaluating sums of products of orthant probabilities from multivariate normal random vectors, $\mathrm{P}(\Sigma) = \mathrm{P}(R)$, where R is the correlation matrix associated with Σ. Thus the weights, $w(P, k, \Omega)$, depend only on the correlation matrix of Ω. Because the correlation matrix from $\sigma\Sigma$ is the same as that from Σ, we can assume, without loss of generality, $\sigma = 1$. Recall the following two identities concerning the binomial coefficients:

$$\sum_{k=0}^{p} \binom{p}{k} = 2^p, \qquad \binom{p}{k} = \binom{p}{p-k}. \tag{A.9}$$

The first identity implies that the total number of disjoint subsets of the positive orthant in p-dimensional space is 2^p. The second implies that the number of subsets with k elements greater than zero is exactly equal to the number of subsets with $p-k$ elements greater than zero, for all k from zero to p. If we show that each individual term of the summation comprising $w(p, k, \Omega)$ equals the corresponding term of the summation comprising $w(p, p-k, \Omega^{-1})$, then because both summations have the same number of elements we can conclude that $w(p, k, \Omega) = w(p, p-k, \Omega^{-1})$.

The term of (A.8) for $w(p, n(M), \Omega)$ associated with the subset M of P is

$$\mathrm{P}([\Omega_{M'}]^{-1})\,\mathrm{P}([\Omega_{M;\,M'}]). \tag{A.10}$$

The term for $w(p, p-n(M), \Omega^{-1})$ corresponds to the set $M' = P - M$. It is

$$\mathrm{P}([\Omega_M^{-1}]^{-1})\,\mathrm{P}(\Omega_{M';\,M}^{-1}). \tag{A.11}$$

Recall the notation used in the proof of Theorem 1 where we set $\Omega^{-1} \equiv C$. Let C be partitioned conformably to Ω for the purposes of our proof. Hence we can rewrite (A.11) as

$$\mathrm{P}([C_M]^{-1})\,\mathrm{P}(C_{M';\,M}). \tag{A.12}$$

The partitioned matrix inversion lemma for both C and Ω and the expression for the variance of the conditional distribution of one subvector from a multivariate normal vector given the remaining subvector from this same vector yields the following two matrix equalities:

$$[\Omega_{M'}]^{-1} = C_{M';\,M} \quad \text{and} \quad \Omega_{M;\,M'} = [C_M]^{-1}.$$

Therefore (A.10) and (A.11) are equal. We can continue in this fashion for all $M \subseteq P$ such that $n(M) = m$ and all corresponding $M' \subseteq P$ such that $n(M') = p - m$ to show that $w(p, m, \Omega) = w(p, p-m, \Omega^{-1})$. Since p and m are arbitrary we can show this equality holds for any p and all m between 0 and p. □

A.3. Example of slackness of bounds on null distribution for parameterized Σ

Consider model (1) with AR(1) errors as described in section 6. Let $X = [e_1 \;\; e_2]$ where e_i is a $(T \times 1)$ vector with one in the ith position and zeros elsewhere. Suppose we are testing $\beta \geq 0$ for $\beta \in R^2$ so that $R = I$ and $r = 0$. In this case $\Sigma = \sigma^2 V(\rho)$ for $V(\rho)$ given in Theil (1971, p. 252). For this purposely simplistic choice of X, R, Σ and $Q \equiv R(X'\Sigma^{-1}X)^{-1}R'$ we have

$$Q = \sigma^2 \begin{bmatrix} 1+\rho^2 & \rho \\ \rho & 1 \end{bmatrix}. \tag{A.13}$$

Stationarity of ε_t requires that $|\rho| < 1$. This implies that the modulus of the correlation between β_1 and β_2 is at most $1/\sqrt{2}$. By Corollary 3, for any $c > 0$,

$$\begin{aligned}\Pr_{0,Q}(IU \geq c) = {} & w(2,1,Q)\Pr(\chi_1^2 \geq c) \\ & + w(2,0,Q)\Pr(\chi_2^2 \geq c). \end{aligned} \tag{A.14}$$

Using the weights functions given in Kudo (1963) and taking the supremum and infimum of the left-hand side probability in (A.14) with respect to all ρ less than one in modulus implies the following upper and lower bounds:

$$\alpha = \inf_{|\rho|<1} \Pr_{0,Q}(IU \geq c_L) = \tfrac{1}{2}\,\mathrm{P}(\chi_1^2 \geq c_L) + \tfrac{1}{8}\Pr(\chi_2^2 \geq c_L), \tag{A.15}$$

$$\alpha = \sup_{|\rho|<1} \Pr_{0,Q}(IU \geq c_U) = \tfrac{1}{2}\,\mathrm{P}(\chi_1^2 \geq c_U) + \tfrac{3}{8}\Pr(\chi_2^2 \geq c_U). \tag{A.16}$$

It is straightforward to show that $c_l < c_L$ and $c_u > c_U$, so that for Σ parameterized by AR(1) errors the bounds given in (19) and (20) are slack.

References

Avriel, M., 1976, Nonlinear programming: Analysis and methods (Prentice Hall, Englewood Cliffs, NJ).

Barlow, R.E., D.J. Bartholomew, J.N. Bremner, and H.D. Brunk, 1972, Statistical inference under order restrictions (Wiley, New York, NY).

Bartholomew, D.J., 1959a, A test of homogeneity for ordered alternatives I, Biometrika 46, 36–48.

Bartholomew, D.J., 1959b, A test of homogeneity for ordered alternatives II, Biometrika 46, 328–335.

Bartholomew, D.J., 1961, A test of homogeneity of means under restricted alternatives, Journal of the Royal Statistical Society B 23, 239–281.

Berndt, E.R. and N.E. Savin, 1977, Conflict among criterion for testing hypotheses in the multivariate linear regression model, Econometrica 45, 1263–1278.

Bohrer, R. and W. Chow, 1978, Weights for one-sided multivariate inference, Applied Statistics 27, 100–104.

Breusch, T.S., 1979, Conflict among criterion for testing hypotheses: Extensions and comments, Econometrica 47, 203–206.

Dykstra, R.L. and T. Robertson, 1983, On testing monotone tendencies, Journal of the American Statistical Association 78, 342–350.

Gill, P.E., W. Murray, and M.H. Wright, 1981, Practical optimization (Academic Press, New York, NY).

Gourieroux, C., A. Holly, and A. Monfort, 1982, Likelihood ratio, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters, Econometrica 50, 63–80.

Graybill, F.A., 1969, Introduction to matrices with applications in statistics (Wadsworth, Belmont, CA).

Hendry, D.F., 1976, The structure of simultaneous equations estimators, Journal of Econometrics 4, 51–88.

Hillier, G., 1986, Joint tests for zero restrictions on nonnegative regression coefficients, Biometrika 73, 657–670.

Hogg, R.V., 1961, On the resolution of statistical hypothesis, Journal of the American Statistical Association 58, 978–989.

Judge, G.G. and T. Takayama, 1966, Inequality restrictions in regression analysis, Journal of the American Statistical Association 61, 166–181.

Kodde, D.A. and F.C. Palm, 1986, Wald criteria for jointly testing equality and inequality restrictions, Econometrica 54, 1243–1248.

Kudo, A., 1963, A multivariate analogue of the one-sided test, Biometrika 50, 403–418.

Kudo, A. and J.-R. Choi, 1975, A generalized multivariate analogue of the one sided test, Memoirs of the Faculty of Science, Kyushu University A 29, 303–328.

Lehmann, E.L., 1986, Testing statistical hypothesis, 2nd ed. (Wiley, New York, NY).

Liew, C.K., 1976, Inequality constrained least-squares estimation, Journal of the American Statistical Association 71, 746–751.
Luenberger, D.G., 1969, Optimization by vector space methods (Wiley, New York, NY).
Nuesch, P.E., 1966, On the problem of testing location in multivariate problems for restricted alternatives, Annals of Mathematical Statistics 37, 113–119.
Muirhead, R.J., 1982, Aspects of multivariate statistical theory (Wiley, New York, NY).
Perlman, M.D., 1969, One-sided problems in multivariate analysis, Annals of Mathematical Statistics 40, 549–567.
Robertson, T. and E.J. Wegman, 1978, Likelihood ratio tests for order restrictions in exponential families, Annals of Statistics 6, 485–505.
Savin, N.E., 1976, Conflict among testing procedures in a linear regression model with autoregressive disturbances, Econometrica 44, 1303–1315.
Shapiro, A., 1985, Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints, Biometrika 72, 133–144.
Sutherland, R.J., 1983, Distributed lags and the demand for electricity, The Energy Journal 4 (special electricity issue), 141–151.
Theil, H., 1971, Principles of econometrics (Wiley, New York, NY).
Wolak, F.A., 1987, The local nature of hypothesis tests involving nonlinear inequality constraints, Mimeo. (Stanford University, Department of Economics, Stanford, CA).
Wolak, F.A., 1988, Duality in testing multivariate hypotheses, Biometrika 75, forthcoming.
Wolak, F.A., 1989, Local and global testing of linear and nonlinear inequality constraints in nonlinear econometric models, Econometric Theory 5, forthcoming.
Yancey, T.A., G.G. Judge, and M.E. Bock, 1981, Testing multiple equality and inequality hypothesis in economics, Economics Letters 7, 249–255.
Yancey, T.A., R. Bohrer, and G.G. Judge, 1982, Power function comparisons in inequality hypothesis testing, Economics Letters 9, 161–167.