Sankhyā : The Indian Journal of Statistics
2002, Volume 64, Series A, Pt.1, pp 109-138

THE BEHAVIOUR OF LINEAR MODEL SELECTION TESTS UNDER GLOBALLY NON-NESTED HYPOTHESES

By MITCHELL R. WATNIK and WESLEY O. JOHNSON
University of California, Davis

SUMMARY. We consider the asymptotic behaviour of tests for non-nested linear model selection under the alternative hypothesis. We consider three testing procedures: the *J-test* (cf. Davidson and MacKinnon, 1981); the *JA-test* (cf. Fisher and McAleer, 1981); and the *modified Cox test* (cf. Godfrey and Pesaran, 1983). All are shown to have positive means under the true alternative. Thus, for discriminating between models, they should be treated as one-sided tests. We derive asymptotic distributions for slightly modified versions of these statistics under the alternative in order to obtain relative efficiencies. The modified versions are selected since they have greater power than the original test statistics. Both algebraic and theoretical relationships among the tests are indicated.

1. **Introduction**

Choosing between non-nested linear regression models is an important practical problem in statistics. For example, suppose there are two schools of thought regarding a certain biological, economic or sociological phenomenon. Each school proceeds to analyze particular data accordingly, resulting in models that overlap to a degree but are still "separate" e.g. non-nested. Alternatively, the models may have resulted from different model selection strategies e.g. C_p (cf. Mallows, 1973) or the corrected Akaike Information Criteria (cf. Hurvich and Tsai, 1989). There is a clear interest in choosing between the two models.

The problem of choosing between non-nested models traces back to Cox (1961) who formulated a testing procedure based on a likelihood ratio approach. Under appropriate conditions, differing variants of his procedure for non-nested linear models have well-behaved asymptotic properties under the null and local alternative hypotheses (cf. Dastoor and McAleer, 1989; Pesaran, 1982, 1987; Watnik, Johnson and Bedrick, 2001).

Paper received October 2000; revised July 2001.
AMS (1991) subject classifications. Primary 62J05, 62F05; secondary 62J15, 62F03.
Key words and phrases. Asymptotic relative efficiency; j-test; JA-test; Multiple correlation; Power; Regression; Separate hypotheses.

In the significance testing arena (cf. Cox, 1962), one does not assume that the two models make up the entire space of possibilities; the alternative is simply left unspecified. Here, unlike Dastoor and McAleer (1989), for example, we are interested in the discrimination problem (Cox, 1962); i.e., determining which of the two hypotheses is better (Efron, 1984). McAleer (1995) notes that, although this is not the way non-nested tests were derived, this is typically the way they are employed.

Unlike the usual nested situation, it is possible to 'reject' both hypotheses when they are non-nested. To be specific, if the two hypotheses are referred to as H_X and H_Z, one must calculate the test statistic when H_X is the null and H_Z is the alternative as well as when H_Z is the null and H_X is the alternative. With each calculation, the practitioner must consider whether or not to 'reject' the current null hypothesis. In the non-nested situation, then, four conclusions are possible (Green, 1971): 1) reject both hypotheses; 2) reject H_X but not H_Z; 3) reject H_Z but not H_X; and 4) do not reject either hypothesis.

Pesaran (1974) defines "size" and "power" for non-nested tests. Size, α, is the probability, calculated under H_X, of rejecting the null hypothesis when the H_X is the null. Power, π_n, is defined to be the probability of correctly identifying the true model; namely the probability under H_X of simultaneously i) not rejecting H_X *and* ii) rejecting H_Z. By definition, then, power in the non-nested situation is less than the power for a comparable test of nested hypotheses, π_n^*, which in our case would simply be the probability under H_X of rejecting the null hypothesis H_Z. The last point is important, because standard definitions of efficiency, such as the Bahadur efficiency (Bahadur, 1960) and the Hájek and Šidák (1967, p. 267) relative efficiency, depend on the definition of power in the nested situation. We note, however, that π_n is approximately $\pi_n^* - \alpha$ provided the chance of rejecting the false null, conditional on having rejected the true null, is nearly one. Thus, with nested power calculations, we need only make a simple adjustment to get the non-nested power under this very plausible assumption, which is supported by empirical work.

Cox (1961, 1962) considers the significance testing setting and uses it to motivate the idea of artificial nesting, wherein one creates an artificial "supermodel" and tests hypotheses about subsets of it. This is the tactic employed in both the Davidson and MacKinnon (1981) development of the popular *J-test*, and the Fisher and McAleer (1981) *JA-test*. Our results (cf. section 3) indicate that, in the discrimination setting, the above procedures should result in one-sided tests. Non-nested tests have often been considered as significance tests and thus treated as two-sided (cf. Godfrey, 1998).

The primary purpose of this paper is to make asymptotic power comparisons for the *modified Cox test* of Godfrey and Pesaran (1983), the *J-test*, the *JA-test*, and some variations of these three procedures. Towards that goal, we also give a way to estimate the power of non-nested model selection tests when the true parameter vector and variance can be consistently estimated. In addition, we give a method for theoretically judging which procedure is the most powerful by using measures of asymptotic relative efficiency. Although the *J-test* usually stands out as preferable in large samples, it has been established that this test may not hold size in moderate samples, especially when the dimensions of the two models are different (Godfrey, 1998; McAleer, 1995; Watnik, Johnson and Bedrick, 2001).

Pesaran (1984) compares the Bahadur (1960) asymptotic relative efficiencies of some Cox-type testing procedures against each other and against the *J-test* using the two-sided versions; i.e., under the significance testing setting. Pesaran (1982) considers the power of the *J-test* against a Cox-type test for locally non-nested hypotheses, both using asymptotic and Monte Carlo methods. Watnik, *et al.* (2001) give Monte Carlo results for the three aforementioned test procedures.

Godfrey (1998) suggests a promising bootstrap-*J-test* to correct for the *J-test* size problem, but asymptotic theory for this procedure is beyond the scope of this article. Other approaches to the non-nested testing problem in various settings have been discussed by Graybill and Iyer (1994, pp. 309-315), Efron (1984), Efron and Tibshirani (1993, pp. 190-198) and Victoria-Feser (1997).

In section 2, we develop notation, introduce the models, and give preliminary results. Test statistics are introduced and asymptotic properties given in section 3 with technical details given in the appendix. Section 4 discusses the asymptotic relative efficiency comparisons for the testing procedures. Concluding remarks are given in section 5.

2. **Background**

Throughout this paper, the model generating the $n \times 1$ vector Y will be assumed to be one of two candidate models. First consider

$$Y = V\delta + X\beta + \epsilon_x, \tag{2.1}$$

where V is an $n \times o$ matrix; X is $n \times p$; and $\epsilon_{x1}, ..., \epsilon_{xn}$ are independently and identically distributed with mean 0 and finite variance, $\sigma_x^2 > 0$. The rank of X is defined to be p throughout. If model (2.1) generates Y, then we assume that $\delta_1, \ldots, \delta_o, \beta_1, \ldots, \beta_p$ are all *non-zero.*

As a competitor, the model

$$Y = V\delta + Z\gamma + \epsilon_z \tag{2.2}$$

is suggested; Z is an n x q matrix; V is common to both models; and ϵ_{zi} are independently and identically distributed with mean 0 and variance σ_z^2 for $i = 1, ..., n$. The rank of Z is q. If model (2.2) generates Y, then we assume that $\delta_1, \ldots, \delta_o, \gamma_1, \ldots, \gamma_q$ are all *non-zero.*

We refer to matrices V, X, Z etc. and their respective column spaces (cf. Christensen, 1987, p. 325) interchangeably in order to ease the notational burden. We assume that the $n \times 1$ vector of ones, e, is in V. The column vectors in X are linearly independent of those in Z and vice versa. Furthermore, neither X nor Z may be the null set, as this would result in nested models.

For ease of tractability, and without loss of generality, we assume

$$V^T X = 0, \qquad V^T Z = 0. \tag{2.3}$$

Thus the vector δ is the same in models (2.1) and (2.2). Otherwise, one can reparameterize by projecting both X and Z onto $V^\perp$, the orthogonal complement of V, in order to satisfy (2.3) without altering the models. Equations (2.3) allow us to consider the contributions due to X and Z individually; i.e., outside of whatever V already adds to the model.

We develop asymptotic properties for two statistics that are key to our later results. Our inferences are conditional upon X and Z and we require

$$C_{XX} \equiv \lim \frac{X^T X}{n}, \qquad C_{ZZ} \equiv \lim \frac{Z^T Z}{n}, \qquad C_{XZ} = C_{ZX}^T \equiv \lim \frac{X^T Z}{n}, \tag{2.4}$$

where C_{XX} and C_{ZZ} are finite, positive definite, symmetric matrices. In fact, we assume that $X^T X/n = C_{XX}$, etc. for large enough n. We also require the following standard assumptions (cf. McAleer, 1995) to hold for large n:

ASSUMPTIONS 2.1.

a. *The design matrices X and Z are completely separate (cf. Cox, 1961); i.e., no linear combination of columns in X lies in the column space of Z and vice-versa, or $X \cap Z = \emptyset$.*

b. *$X\beta$ does not lie in $Z^\perp$.*

c. *X and Z are non-singular.*

Kent (1986) discusses the implications of violating these assumptions. We address the consequences of violating Assumption 2.1b in section 3.3.

Pesaran (1987) defines "local alternatives" to mean that (2.1) and (2.2) can be brought arbitrarily close to one another in our setting. His definition of "global alternatives" is that the two models cannot be brought arbitrarily close; i.e., they are distinct. With the parameterization that Pesaran (1987) uses, he asserts that linear models cannot be globally non-nested. However, in our parameterization and because of the assumptions that β_i $(i = 1, \ldots, p)$ and γ_j $(j = 1, \ldots, q)$ are all non-zero, we ensure that the models are globally non-nested and are not pair-wise local alternatives. Ours separates the common part of the models (V), and the distinct parts (X and Z), while his does not. We believe that the global setting is a more natural way to consider linear alternatives.

Now, define the set of linear statistics

$$W_X \equiv n^{-1/2}(X^TX)^{-1/2}X^TY, \qquad W_Z \equiv n^{-1/2}(Z^TZ)^{-1/2}Z^TY.$$

Under (2.1), and with large n,

$$E(W_X) = C_{XX}^{1/2}\beta \equiv \mu_x, \qquad E(W_Z) = C_{ZZ}^{-1/2}C_{ZX}\beta \equiv \mu_z. \tag{2.5}$$

Further, define

$$A \equiv (Z^TZ)^{-1/2}Z^TX(X^TX)^{-1/2}, \qquad R \equiv C_{ZZ}^{-1/2}C_{ZX}C_{XX}^{-1/2}, \qquad M \equiv R^TR. \tag{2.6}$$

Note that $A = corr(n^{1/2}W_Z, n^{1/2}W_X)$ and that X and Z are centered since $e \in V$ and both are projected onto $V^{\perp}$, due to (2.3). We use A to denote the computational formula for the correlation matrix while R represents the limiting correlation in the theoretical results. Notably, A^TA is also the matrix from which the sample canonical correlation coefficients are obtained for the two samples of column variables associated with X and Z, while M is the "population counterpart." An important property of M is that its characteristic roots are between 0 and 1, inclusive, and hence is non-negative definite. We note in the Appendix to Section 2 that W_X and W_Z are asymptotically jointly normal under mild conditions.

Recalling the definitions of M from (2.6) and μ_x from (2.5), we define

$$\begin{aligned} \xi_i &\equiv \mu_x^TM^i\mu_x, \quad i = 0, 1, ..., 5; \\ \lambda_i &\equiv \xi_{i-1} - \xi_i = \mu_x^TM^{i-1}(I - M)\mu_x, \quad i = 1, ..., 5; \end{aligned} \tag{2.7}$$

where $M^0 \equiv I$ by convention. We note that $\lambda_i \geq 0$ and that $\lambda_i - \lambda_{i+1} = \mu_x^T(I-M)M^{i-1}(I-M)\mu_x \geq 0$, where the inequality holds because $M = R^TR$ is symmetric with roots between 0 and 1. It follows that $\xi_0 \geq \xi_1 \geq \cdots \geq \xi_5 \geq 0$, and

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \lambda_5 \geq 0. \tag{2.8}$$

We note that, under Assumptions 2.1, statement (2.8) becomes $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > \lambda_5 > 0$.

The following proposition is useful for establishing later results.

PROPOSITION 2.1.

$$(a)\ \lambda_2^2 \leq \lambda_1\lambda_3; \qquad (b)\ \lambda_3^2 \leq \lambda_2\lambda_4; \qquad (c)\ \lambda_3^2 \leq \lambda_1\lambda_5; \qquad (d)\ \lambda_2\lambda_3 \leq \lambda_1\lambda_4. \tag{2.9}$$

We prove claims (a) and (d) in the Appendix to Section 2 and the others follow similarly.

Finally, we define a consistent estimator of the variance. Following Pesaran (1974), under (2.1),

$$s_x^2 \equiv \frac{||P_{(V,X)^\perp}Y||^2}{n-p-o} \to \sigma_x^2, \qquad s_z^2 \equiv \frac{||P_{(V,Z)^\perp}Y||^2}{n-q-o} \to \sigma_x^2 + \lambda_1,$$

in probability as $n \to \infty$, where $P_{X^\perp} \equiv I - P_X$, P is the projection matrix, $||.||$ denotes the Euclidean norm of a vector. Similarly, $s_z^2 \to \sigma_z^2$ and $s_x^2 \to \sigma_z^2 + ||C_{ZZ}^{1/2}\gamma||^2 - ||C_{XX}^{-1/2}C_{XZ}\,\gamma||^2$, in probability under (2.2). Since the true model is unknown *à priori*, we use Royston and Thompson's (1995) estimate of variance:

$$\hat{\sigma}^2 \equiv min(s_x^2, s_z^2).$$

Then, $\hat{\sigma}^2$ consistently estimates σ_x^2 under (2.1) and σ_z^2 under (2.2). Equivalently, $\hat{\sigma}^2$ consistently estimates σ^2, the variance of the true model, regardless of whether σ^2 is σ_x^2 or σ_z^2. We use $\hat{\sigma}$ instead of either s_x or s_z to estimate σ, because it increases the power of non-nested tests while not changing the asymptotic distributions under the null.

3. The Test Statistics

We seek to evaluate the performance of various non-nested testing procedures. We first present the test statistics associated with the *J-test* (Davidson and MacKinnon, 1981), the *JA-test* (Fisher and McAleer, 1981), and the *modified Cox test* (Godfrey and Pesaran, 1983). All of these test statistics considered herein have asymptotic standard normal distributions under the null hypothesis. We will derive the asymptotic distributions under the alternative.

3.1. *The null hypothesis (2.2).* The version of the test statistic for the *J-test* developed by Royston and Thompson (1995, equation 4.5), which they argue has closer to nominal size under (2.2) than does the original Davidson and MacKinnon (1981) proposal, is

$$J_Z \equiv \frac{||P_X Y||^2 - Y^T P_X P_Z Y}{n^{1/2}\hat{\sigma}\,(||P_X Y||^2 - ||P_Z P_X Y||^2)^{1/2}}. \tag{3.1}$$

In the Appendix to Section 3, we present this and the other statistics in this section in a form which facilitates theoretical arguments. We consider the form of (3.1) to be more practical for calculations.

The *JA-test* has test statistic (cf. Table 1 of Fisher, 1983)

$$JA_Z \equiv \frac{Y^T P_X P_Z Y - ||P_X P_Z Y||^2}{n^{1/2}\hat{\sigma}\,(||P_X P_Z Y||^2 - ||P_Z P_X P_Z Y||^2)^{1/2}}. \tag{3.2}$$

Fisher (1983) shows that, when the appropriate term is substituted for $\hat{\sigma}$ above, JA_Z has an exact t distribution under (2.2) when the errors are normally distributed.

Our version of the *modified Cox test* statistic of Godfrey and Pesaran (1983) is the negative of the original

$$\tilde{G}_Z \equiv \frac{||P_X Y||^2 - ||P_X P_Z Y||^2 - s_z^2\,\{p - tr(P_Z P_X)\}}{n^{1/2}\,\{4s_z^2\,(||P_X P_Z Y||^2 - ||P_Z P_X P_Z Y||^2) + 2s_z^4 tr(B^2)\}^{1/2}},$$

where

$$tr(B^2) \equiv p - tr(P_X P_Z)^2 - \frac{\{p - tr(P_X P_Z)\}^2}{n-q}.$$

The numerator of this statistic is identical to expression (89) of Cox (1961) and it has expectation zero under H_Z. We consider the simpler test statistic

$$G_Z \equiv \frac{||P_X Y||^2 - ||P_X P_Z Y||^2}{\{4n\hat{\sigma}^2\,(||P_X P_Z Y||^2 - ||P_Z P_X P_Z Y||^2)\}^{1/2}}. \tag{3.3}$$

We note that $G_Z - \tilde{G}_Z = o_p(1)$ under (2.2) (cf. Godfrey and Pesaran, 1983, p. 138).

REMARKS.

1. The statistics J_X, JA_X, and G_X are defined analogously; namely, by exchanging Z for X and vice-versa in equations (3.1), (3.2), and (3.3).

2. The statistics J_Z, JA_Z, and $\tilde{G}_Z$ are, with the exception of using $\hat{\sigma}$ in place of their respective estimates of the standard deviation, exactly the same as those given in Davidson and MacKinnon (1981) (cf. Table 1 of Fisher, 1983), Fisher and McAleer (1981), and Godfrey and Pesaran (1983), respectively. Our use of $\hat{\sigma}$ to estimate the standard deviation was done solely to increase the power of all of the tests (cf. section 2).

3. Had we used s_z in place of $\hat{\sigma}$ in (3.3), it would have been the case that $G_Z - \tilde{G}_Z = o_p(1)$ under (2.1) (cf. Godfrey and Pesaran, 1983) and that G_Z would be equivalent to the test statistic given in equation (25) of Fisher and McAleer (1981) (cf. Table 1 of Fisher, 1983).

We define a notation for the denominators of the J and JA tests, respectively:

$$\begin{aligned} \hat{\tau}_{JZ} &\equiv \left\{\hat{\sigma}^2\left(||W_X||^2 - ||AW_X||^2\right)\right\}^{1/2}; \\ \hat{\tau}_{JAZ} &\equiv \left\{\hat{\sigma}^2\left(||A^T W_Z||^2 - ||AA^T W_Z||^2\right)\right\}^{1/2} \end{aligned} \tag{3.4}$$

Note that the denominator of G_Z is proportional to the denominator of JA_Z.

Using (2.7) and (A.1), we see that, under (2.1),

$$\hat{\tau}_{JX} \xrightarrow{p} \sigma_x \lambda_2^{1/2}, \hat{\tau}_{JAX} \xrightarrow{p} \sigma_x \lambda_2^{1/2}, \hat{\tau}_{JZ} \xrightarrow{p} \sigma_x \lambda_1^{1/2},\ \hat{\tau}_{JAZ} \xrightarrow{p} \sigma_x \lambda_3^{1/2}. \tag{3.5}$$

This verifies the result of Fisher (1983), who notes that the denominators of J_X and JA_X are asymptotically equivalent under H_X.

So, it may be interesting to investigate the properties of these statistics using the other's denominator, as Pesaran (1984, p. 251) notes, the efficiency of a Cox-type test depends "on the choice of consistent estimate of the variance". We therefore define

$$FG_Z \equiv \frac{||P_X Y||^2 - ||P_X P_Z Y||^2}{2\hat{\tau}_{JZ}}; \tag{3.6}$$

$$FJ_Z \equiv \frac{||P_X Y||^2 - Y^T P_X P_Z Y}{\hat{\tau}_{JAZ}}; \tag{3.7}$$

and

$$FJA_Z \equiv \frac{Y^T P_X P_Z Y - ||P_X P_Z Y||^2}{\hat{\tau}_{JZ}}. \tag{3.8}$$

We note the correspondence

$$FG_Z = \frac{J_Z + FJA_Z}{2}, \qquad G_Z = \frac{FJ_Z + JA_Z}{2}. \tag{3.9}$$

As a consequence of this, we thus expect the performance of FG_Z and G_Z to strike a balance between that for the corresponding two tests. We finally note that the statistics FG_X, FJ_X, and FJA_X are defined by interchanging the roles of X and Z in equations (3.6), (3.7), and (3.8).

3.2 The alternative hypothesis: (2.1). We now derive the distributions of J_Z, FJ_Z, JA_Z, FJA_Z, G_Z, and FG_Z under (2.1). As the subscript indicates, (2.2) is considered to be the null.

Using (2.5) and (2.7) it follows that, in probability and under (2.1),

$$\begin{array}{llllll} J_Z \to \frac{\lambda_1}{\sigma\lambda_1^{1/2}}, & JA_Z \to \frac{\lambda_2}{\sigma\lambda_3^{1/2}}, & G_Z \to \frac{\lambda_1+\lambda_2}{2\sigma\lambda_3^{1/2}}, \\ FJ_Z \to \frac{\lambda_1}{\sigma\lambda_3^{1/2}}, & FJA_Z \to \frac{\lambda_2}{\sigma\lambda_1^{1/2}}, & FG_Z \to \frac{\lambda_1+\lambda_2}{2\sigma\lambda_1^{1/2}}. \end{array} \tag{3.10}$$

Under Assumptions 2.1, then, the asymptotic means of these test statistics are all strictly positive. Pesaran (1987) notes this about J, JA and G, but, since he considers local alternatives, he does not suggest using the statistics as one-sided tests. In the model discrimination setting, however, it follows directly that corresponding tests of hypotheses should be one-sided and the null hypothesis should only be rejected if the observed test statistic is too large.

We obtain the result

PROPOSITION 3.2. *Under (2.1), the assumptions for (A.1) and Assumptions 2.1, the following convergence results hold as $n \to \infty$:*

$$n^{1/2}\left(J_Z - \frac{\lambda_1^{1/2}}{\sigma}\right) \sim N(0,1),$$

$$n^{1/2}\left(JA_Z - \frac{\lambda_2}{\sigma\lambda_3^{1/2}}\right) \sim N\left[0, \frac{\lambda_2}{\lambda_3}\left\{-1 + 2\left(\frac{\lambda_4}{\lambda_3}\right) + \left(\frac{\lambda_2}{\lambda_3^2}\right)(\lambda_4 - \lambda_5)\right\}\right],$$

$$n^{1/2}\left(G_Z - \frac{\lambda_1+\lambda_2}{2\sigma\lambda_3^{1/2}}\right) \sim N\left[0, \left(\frac{\lambda_1+\lambda_2}{\lambda_3^2}\right)\left\{\lambda_4 + \frac{(\lambda_1+\lambda_2)(\lambda_4-\lambda_5)}{4\lambda_3}\right\} - 1\right],$$

$$
\begin{aligned}
&n^{1/2}\left(FJ_Z - \frac{\lambda_1}{\sigma\lambda_3^{1/2}}\right) \\
&\qquad \sim N\left[0, \frac{1}{\lambda_3}\left\{2\lambda_1 - 3\lambda_2 + 2\left(\frac{\lambda_1}{\lambda_3}\right)\lambda_4 + \left(\frac{\lambda_1}{\lambda_3}\right)^2(\lambda_4 - \lambda_5)\right\}\right], \\
&n^{1/2}\left(FJA_Z - \frac{\lambda_2}{\sigma\lambda_1^{1/2}}\right) \\
&\qquad \sim N\left[0, \frac{1}{\lambda_1}\left\{\lambda_2 - 4\left(\frac{\lambda_2}{\lambda_1}\right)(\lambda_2 - \lambda_3) + \left(\frac{\lambda_2}{\lambda_1}\right)^2(\lambda_1 - \lambda_2)\right\}\right], \\
&n^{1/2}\left(FG_Z - \frac{\lambda_1 + \lambda_2}{2\sigma\lambda_1^{1/2}}\right) \\
&\qquad \sim N\left[0, \left(\frac{1}{4\lambda_1}\right)\left\{\left(\frac{\lambda_1 + \lambda_2}{\lambda_1}\right)^2(\lambda_1 - \lambda_2) + 4\left(\frac{\lambda_2}{\lambda_1}\right)\lambda_3\right\}\right].
\end{aligned}
\tag{3.11}
$$

The proof of Proposition 3.2 is in the Appendix to Section 3. Note that the asymptotic variances do not depend on $||\beta||$, while the means do. Also note that the asymptotic distributions under the alternative only depend on the design matrices X and Z through the "canonical correlation" matrix M.

We have the following corollary to Proposition 3.2, which becomes important in the next section.

Corollary to Proposition 3.2: *Under (2.1), the assumptions for (A.1) and Assumptions 2.1, the asymptotic variance of the JA test is greater than or equal to 1.*

The proof of this corollary is in the Appendix to Section 3.

3.3. The case of $X \perp Z$. Although the situation with $X \perp Z$ is excluded by Assumptions 2.1, it is a situation in which it should be easy to discriminate between the two models. So, when X is nearly perpendicular to Z, and the assumptions hold, one might expect high power from these test procedures. However, when $X \perp Z$, the *JA-test* has numerator 0. The denominators of both the *JA-test*, and the *modified Cox test*, G_Z and G_X, are 0 in this situation as well (the statistics $\tilde{G}_Z$ and $\tilde{G}_X$ will have relatively small, but not 0, denominators). Fisher (1983, p. 130) notes the failure of the *JA-test* in this situation. In fact, when X is perpendicular to Z, FG is the only one of these six statistics that does not necessarily have either the numerator or denominator tending to 0.

To elaborate on this, we note that, when $X \perp Z$, $\xi_i = 0$ for $i = 1, 2, \ldots, 5$. So, $\lambda_i = 0$ for $i = 2, \ldots, 5$. The proofs of asymptotic normality for JA_Z and

G_Z will fail in the case where $X \perp Z$ as noted in the proofs in the Appendix to Section 3.

However, when $X \perp Z$, the *J-test* also fails. This observation about the *J-test* is made in a footnote of Davidson and MacKinnon (1981). From equation (3.1), we see that, in our way of calculating the *J-test*, it becomes $||P_X Y||/(n^{1/2}\hat{\sigma})$ when $X \perp Z$. It is easy to see that this cannot be asymptotically normal, with mean 0, in this case because the statistic is necessarily non-negative. One obvious possibility is to instead consider the model $E(Y) = V\delta + X\beta + Z\gamma$ and to use conventional F-tests for $\beta = 0$ and $\gamma = 0$.

4. Relative Efficiency

We proceed to calculate the asymptotic relative efficiencies for J, FJ, JA, FJA, G and FG. We define three measures of relative efficiency and then proceed to make comparisons. We conclude this section by giving an illustration of how to consistently estimate power in the general case.

4.1. Measures of Relative Efficiency. We first define the standard measure of relative efficiency for comparing one-sided tests. Suppose a statistic, say $n^{1/2}S$, is asymptotically standard normal under the null, and $n^{1/2}(S - \mu_a)/\sigma_a$ is asymptotically standard normal under the alternative. Then the nested asymptotic power of a size α test is $1 - \Phi\{(z_\alpha - n^{1/2}\mu_a)/\sigma_a\}$ where $\Phi(z_\alpha) = 1 - \alpha$. Thus power is large provided μ_a is large and/or σ_a is small. Solving for the sample size that achieves power π_n^*, we obtain $n = (z_\alpha - \sigma_a z_{\pi_n^*})^2/\mu_a^2$, provided $z_\alpha \geq \sigma_a z_{\pi_n^*}$. Then define the asymptotic relative efficiency of, say, J to FG to be n_j/n_{fg}, the relative sample sizes for the two size α tests to both achieve power π_n^*. Then

$$e_S(J, FG) \equiv \frac{n_j}{n_{fg}} = \left\{ \left(\frac{z_{1-\alpha} - \sigma_j z_{\pi_n^*}}{\mu_j} \right) / \left(\frac{z_{1-\alpha} - \sigma_{fg} z_{\pi_n^*}}{\mu_{fg}} \right) \right\}^2, \tag{4.1}$$

where μ_{fg} is the mean for FG under the alternative, etc. This measure of efficiency implies that J is superior to FG if $e_S(J, FG) < 1$. By (3.11) and (4.1) we see that relative efficiencies for any pair of tests will not depend on the length of the vector β. We may thus assume without loss of generality that $||\beta|| = 1$ when making these calculations.

In theory, this measure can be somewhat complicated since it may depend on many inputs. Simplifications are obtained if we let $\alpha = .5$ or if $\pi_n^* \to 1$. In these instances

$$e_S(J, FG) \equiv e_{HS}(J, FG) = \left(\frac{\mu_{fg}\sigma_j}{\mu_j \sigma_{fg}} \right)^2, \tag{4.2}$$

which is the Hájek and Šidák (1967, p. 267) measure of relative efficiency, which we will refer to as the HS criterion. On the other hand, if we let $\alpha \to 0$, or $\pi_n^* = .5$, we obtain

$$e_S(J, FG) \equiv e_B(J, FG) = \left(\frac{\mu_{fg}}{\mu_j}\right)^2, \tag{4.3}$$

which is the Bahadur (1961) relative efficiency for comparing J to FG.

4.2 General conclusions about efficiencies. Using Proposition 3.2, we can order some of the statistics relative to these criteria. We begin with a comparison between JA and J under the condition $\pi_n^* \geq 0.5$. We note that

$$e_S(J, JA) = \left\{\frac{(z_{1-\alpha} - z_{\pi_n^*})\,\mu_{ja}}{(z_{1-\alpha} - \sigma_{ja} z_{\pi_n^*})\,\mu_a}\right\}^2.$$

Now, under the condition $\pi_n^* \geq 0.50$, $z_{\pi_n^*} \leq 0$, which implies

$$e_S(J, JA) \leq \left(\frac{\mu_{ja}}{\mu_j}\right)^2 = e_B(J, JA)$$

since, by Corollary to Proposition 3.2, $\sigma_{ja} \geq 1$. Therefore, $e_S(J, JA) \leq \lambda_2^2/\lambda_1\lambda_3 \leq 1$, by Proposition 2.1. Hence J is superior to JA under the general criterion, as well as both the HS and Bahadur criteria.

In the Appendix to Section 4, we establish that, according to the Bahadur criterion, $FJ \succ G \succ J \succ FG \succ FJA$. So, according to the Bahadur criterion, FJ is the theoretically optimal procedure to use, while FJA is theoretically worst. Considering only the three test statistics from the literature, we have established that G is preferable to J, which in turn is preferable to JA, using the Bahadur criterion. We shall see later that JA neither dominates nor is dominated by FG in this context.

Using HS efficiency, we have the following comparison between FG and J:

$$e_{HS}(FG, J) = \left[\frac{\lambda_1^{1/2}\sqrt{\frac{1}{4\lambda_1}\left\{\left(\frac{\lambda_1+\lambda_2}{\lambda_1}\right)^2(\lambda_1 - \lambda_2) + 4\left(\frac{\lambda_2\lambda_3}{\lambda_1}\right)\right\}}}{\frac{\lambda_1+\lambda_2}{2\lambda_1^{1/2}}}\right]^2 \tag{4.4}$$

In the Appendix to Section 4, we show that $e_{HS}(FG, J) \leq 1$ and thus $FG \succeq J$, under HS. Interestingly, the reverse was true according to Bahadur.

If we add the following assumptions

$$\lambda_3^2 = \lambda_2\lambda_4; \qquad \lambda_4^2 = \lambda_3\lambda_5, \tag{4.5}$$

more conclusions under the HS criterion are possible. Equations (4.5) imply that $\lambda_2\lambda_5 = \lambda_3\lambda_4$. Note that equations (4.5) hold if the roots of M are equal or if $rank(M) = 1$, or, more generally, if the non-zero roots of M are equal. So, these assumptions hold in the situations when either $p = 1$ or $q = 1$, as we see in the next two subsections.

In the Appendix to Section 4, we show that $e_{HS}(G, J) \geq 1$ and that $e_{HS}(G, FJ) \leq 1$. So, under the HS criterion and assumptions (4.5), we have the ordering $FG \succeq J \succeq G \succeq FJ$. We do not compare FJA here, because, as we see in the succeeding subsections, no generalizations for FJA are possible. We do not compare JA to the others under HS because it is dominated by J.

4.3 Relative efficiencies; $p = 1$ Case. With $p = 1$, results are particularly simple. Keep in mind that we are testing H_Z versus H_X where the rank of X is one and the rank of Z is arbitrary. We note that M is now the squared multiple correlation between X and Z. Then, from (2.7), we have $\lambda_i = C_{XX}\beta^2 M^{i-1}(1 - M)$ for $i = 1, \dots, 5$. It follows that assumptions (4.5) hold. Proposition 3.2 yields

$$\begin{aligned}
\mu_j = \mu_{ja} &= \frac{|C_{XX}^{1/2}\beta|(1-M)^{1/2}}{\sigma}, & \sigma_j^2 = \sigma_{ja}^2 &= 1,\\
\mu_g &= \frac{|C_{XX}^{1/2}\beta|}{2M\sigma}(1+M)(1-M)^{1/2}, & \sigma_g^2 &= \frac{1+M+3M^2-M^3}{4M^3},\\
\mu_{fj} &= \frac{|C_{XX}^{1/2}\beta|(1-M)^{1/2}}{M\sigma}, & \sigma_{fj}^2 &= \frac{1+M-M^2}{M^3},\\
\mu_{fja} &= \frac{|C_{XX}^{1/2}\beta|M(1-M)^{1/2}}{\sigma}, & \sigma_{fja}^2 &= M(1-3M+3M^2),\\
\mu_{fg} &= \frac{|C_{XX}^{1/2}\beta|}{2\sigma}(1+M)(1-M)^{1/2}, & \sigma_{fg}^2 &= \frac{1+M-M^2+3M^3}{4}.
\end{aligned} \tag{4.6}$$

This implies that, when $p = 1$, J and JA will be asymptotically equivalent, as they are under local alternatives (McAleer, 1995). For the remainder of this subsection, then, we will only state results for J.

We observe from (4.6) that, with small M, J and FG will be more powerful than FJA. In addition, (4.1) and (4.6) imply that changes in size will have the most dramatic effect on G and FJ because of the presence of M in the denominators of their means. Note that the power of all of the test statistics will be monotone and increasing in $|C_{XX}^{1/2}\beta|$ for fixed M; more variability in the X values gives greater power. Also note that all relative efficiencies will depend on the model only through the value of M. We summarize our conclusions for the situations where $\alpha = 0.05$ in Table 1. It is worth remembering that, as size decreases, $e_S \to e_B$. In section 4.2, we saw that the Bahadur criterion had $FJ \succ G \succ J \succ FG \succ FJA$. In this special case, then, the only new conclusion is that $JA \succ FG$.

Table 1. Orderings for tests with $p = 1$.

M	$\pi_n^* = 0.50$	$\pi_n^* = 0.75$
small (≈0.1)	$FJ \succ G \succ J \sim JA \succ FG \succ FJA$	$J \sim JA \succ G \succ FJ \succ FG \succ FJA$
moderate (≈0.5)	$FJ \succ G \succ J \sim JA \succ FG \succ FJA$	$FJ \succ G \succ J \sim JA \succ FG \succ FJA$
large (≈0.9)	$FJ \succ G \succ J \sim JA \succ FG \succ FJA$	$FJ \succ G \succ J \sim JA \succ FG \succ FJA$
	$\pi_n^* = 0.95$	$\pi_n^* = 1$
small	$J \sim JA \succ FG \succ G \succ FJ \succ FJA$	$FG \succ J \sim JA \succ G \succ FJ \succ FJA$
moderate	$J \sim JA \succeq G \succeq FJ \succ FG \succ FJA$	$FG \succ J \sim JA \succ G \succ FJ$
large	$J \sim JA \succeq G \succeq FJ \succeq FG \succeq FJA$	$FJA \succ FG \succ J \sim JA \succ G \succ FJ$

When $\pi_n^* = 1$, we obtain the conclusions summarized in the lower right-hand box of Table 1. In the previous subsection, we showed that $FG \succ J \succ G \succ FJ$ according to this measure, i.e., the HS criterion. We now see that FJA is most efficient for large values of M, while it is least efficient for small values. This is why we were unable to make general conclusions about FJA for the HS criterion. Its place for moderate values of M depends on M. We view these results as theoretically interesting, but of limited value in practice since π_n^* must be arbitrarily close to one for them to hold.

In Table 1 we also summarize the ordering for the situations when $\pi_n^* = 0.75$ and $\pi_n^* = 0.95$, with $\alpha = .05$. When seeking larger power, J tends to be best, while FJ may be preferable when seeking moderate power. The orderings are almost identical to the Bahadur case with $\pi_n^* = 0.75$. Conclusions are practically identical with $\alpha = 0.10$.

The ordering tends towards that for the HS criterion with $\pi_n^* = 0.95$. The tests are practically equivalent with $M > 0.6$. The situation changes slightly with $\alpha = 0.10$; FJA is technically most efficient though by a practically insignificant margin, especially for $M > 0.85$.

For values of π_n^* that are as large as 0.999, the orderings tend to be more like those for $\pi_n^* = 0.50, 0.75$, or 0.95 than for $\pi_n^* = 1$, so for realistic power and based on Table 1, it would appear that FJ, G, J and JA are preferable to FG and FJA. The test FJ is preferable for moderate power and $J \sim JA$ is preferable for higher power. The test G ranks either second or third.

We recall from (3.9) that FG is the average of J and FJA and G is the average of FJ and JA. It is thus not surprising that, when G is more efficient than FJ, it tends to be less efficient than JA and vice-versa, and analogously for FG, J, and FJA, as can be seen by consideration of Table 1.

4.4 Relative efficiencies; $q = 1$ Case. Here, the rank of X is arbitrary and the rank of Z is 1. For this situation, we define $r = RR^T = C_{ZX}C_{XX}^{-1}C_{XZ}/C_{ZZ}$, the squared multiple correlation between Z and X. Also, define $c = ||\beta||^2$ and $\eta = \rho^T\beta/c^{1/2}$, where $\rho^T \equiv C_{ZX}C_{XX}^{-1/2}/C_{ZZ}^{1/2} \equiv (\rho_1, \dots, \rho_p)$, the vector of correlations between Z and X. Note that $r = \sum_{i=1}^{p} \rho_i^2$. As a consequence of these definitions and Assumptions 2.1, $0 < \eta^2 \leq r < 1$. Also note that, if $p = 1$ then $\eta^2 = r$.

We use the assumptions $C_{XX} = I$ and $C_{ZZ} = 1$, without loss of generality. For example, in setting up the models (2.1) and (2.2), we could simply have reparameterized X and Z appropriately utilizing Gram-Schmidt without altering the basic structure of the models. It then follows from (2.7) that $\lambda_1 = c(1-\eta^2)$, $\lambda_2 = c\eta^2(1-r)$, $\lambda_3 = c\eta^2 r(1-r)$, $\lambda_4 = c\eta^2 r^2(1-r)$, $\lambda_5 = c\eta^2 r^3(1-r)$. It follows directly that assumptions (4.5) hold. Additionally, Proposition 3.2 implies

$$
\begin{aligned}
\mu_j &= \frac{\{c(1-\eta^2)\}^{1/2}}{\sigma}, & \mu_{fj} &= \frac{c^{1/2}(1-\eta^2)}{\sigma\{\eta^2 r(1-r)\}^{1/2}}, \\
\mu_{ja} &= \frac{\{c\eta^2(1-r)\}^{1/2}}{\sigma r^{1/2}}, & \mu_{fja} &= \frac{c^{1/2}\eta^2(1-r)}{\sigma\{(1-\eta^2)\}^{1/2}}, \\
\mu_g &= \frac{c^{1/2}(1-r\eta^2)}{2\sigma\eta\{r(1-r)\}^{1/2}}, & \mu_{fg} &= \frac{c^{1/2}(1-r\eta^2)}{2\sigma\{(1-\eta^2)\}^{1/2}},
\end{aligned}
$$

$$
\begin{aligned}
\sigma_j^2 &= 1, \qquad \sigma_{ja}^2 = 1, \\
\sigma_g^2 &= \frac{1}{4}\left\{\frac{1}{r\eta^4(1-r)}\right\}\left(1 + 2r\eta^2 - 4r\eta^4 + r^2\eta^4\right), \\
\sigma_{fj}^2 &= \left\{\frac{1}{\eta^2 r(1-r)}\right\}\left(2r + \eta^2 r - 4\eta^2 + \frac{1}{\eta^2}\right), \\
\sigma_{fja}^2 &= \left(\frac{\eta^2}{1-\eta^2}\right)\left\{1 - r - 4\eta^2\frac{(1-r)^3}{1-\eta^2} + \eta^2\left(\frac{1-r}{1-\eta^2}\right)^2(1 - 2\eta^2 + r\eta^2)\right\}, \\
\sigma_{fg}^2 &= \frac{1}{4}\left(\frac{1}{1-\eta^2}\right)\left\{\left(\frac{1-r\eta^2}{1-\eta^2}\right)^2(1 - 2\eta^2 + r\eta^2) + 4r\eta^4\frac{(1-r)^2}{1-\eta^2}\right\}.
\end{aligned}
\tag{4.7}
$$

Recall that the efficiencies of the tests will not depend on c, but we have included it for potential use in power calculations. We must keep in mind that β here is the reparameterized version. Note that when $p = 1$, these formulae reduce to those in (4.6), as they should.

We have already established that J is asymptotically at least as efficient as JA according to the general criterion. When $\eta^2 < r$, $J \succ JA$ and they are equally efficient only where $\eta^2 = r$. This illustrates how restrictive is the assumption of local alternatives (cf. Pesaran, 1987) because, under local alternatives, J is asymptotically equivalent to JA (cf. McAleer, 1995). Clearly, if $\eta^2 \doteq 0$, JA will have $\pi_n^* \doteq \alpha$. The test FJA will also suffer if $\eta^2 \doteq 0$.

We saw earlier that the Bahadur criterion's ordering of the procedures was $FJ \succeq G \succeq J \succeq FG \succeq FJA$ and $J \succeq JA \succeq FJA$ (cf. section 4.2). Now, we can see that neither FG nor JA dominates the other according to the Bahadur criterion since, with $r = 1$, $FG \succeq JA$, while with $\eta^2 = r$, $JA \succeq FG$.

As was the case when $p = 1$, the HS measure of relative efficiency does not imply that FJA is superior or inferior to either J or FG everywhere. So, either FG or FJA will be superior when the power is 1.

We present a summary of the conclusions we draw for the situation when $\alpha = 0.05$ and $\pi_n^* = 0.75$ and 0.95 in Table 2. We note that the conclusions when $\alpha = 0.10$ are virtually identical to those in Table 2.

Table 2. ORDERINGS FOR TESTS WITH $q = 1$.

r	η^2	$\pi_n^* = 0.75$	$\pi_n^* = 0.95$
small	small	$J \succ G \succ FJ \succ FG \succ FJA$	$J \succ FG \succ G \sim FJ \succ FJA$
moderate	small	$J \succ FJ \succ G \sim FG \succ FJA$	$J \succ FG \succ FJ \sim G \succ FJA$
moderate	moderate	$FJ \succ G \succ J \succ FG \succ FJA$	$J \succeq G \succeq FJ \succ FG \succ FJA$
large	small	$J \succ FJ \succ G \sim FG \succ FJA$	$J \succ FG \succ FJ \succeq G \succeq FJA$
large	moderate	$FJ \succ G \succ J \succ FG \succ FJA$	$J \succ FJ \succ G \succ FG \succ FJA$
large	large	$FJ \succ G \sim J \succ FG \succ FJA$	$J \succeq FJ \succeq G \succeq FG \succeq FJA$

The conclusions in Table 2 are similar to the analogous ones in Table 1 in that they imply that, in general, either FJ or J is superior while FJA is inferior.

4.5. An example: flow rate. We illustrate how our formulae can be used to make theoretical power calculations. Consider modelling flow rate from watersheds following storm episodes as in Rawlings (1988, pp. 150-1). Though the sample size is 30, which is not particularly large, we use it nonetheless for illustrative purposes. The response variable is the log of the peak flow rate, Y. Other variables are the average slope of the watershed in percent, $\bar{s}$; longest stream flow in watershed in thousands of feet, f; infiltration rate of water into soil in inches per hour, i; watershed area in square miles, a; estimated soil storage capacity in inches of water, c; rainfall in inches, r; time period during which rainfall exceeded 0.25 inches per hour, t; and surface absorbancy index, where 0 represents complete absorbancy and 100 represents no absorbancy, b.

The model obtained through a forward stepwise selection procedure was

$$H_X : Y = \delta_0 + \delta_1 \bar{s} + \beta_1 f + \beta_2 i + \epsilon_X.$$

Rawlings (1988, p. 519) mentions that Mallows' C_p selects the model

$$H_Z : Y = \delta_0 + \delta_1 \bar{s} + \gamma_1 a + \gamma_2 c + \gamma_3 r + \gamma_4 t + \gamma_5 b + \epsilon_Z.$$

Comparing these models is of interest. Here, then, $p = 2$, $q = 5$ and $o = 2$. We projected the matrices X and Z into $V^\perp$ and, assumed that $C_{XX} = X^T X/n$, etc. Since $\hat{\beta}$ and $\hat{\gamma}$ are consistent estimates of β and γ under their respective hypotheses, and $\hat{\sigma} = min(s_x, s_z)$ is a consistent estimate of σ, we were then able to calculate the following statistics:

$$\hat{\lambda}_1 = 0.1344,\ \hat{\lambda}_2 = 0.0470,\ \hat{\lambda}_3 = 0.0358,\ \hat{\lambda}_4 = 0.0337,\ \hat{\lambda}_5 = 0.0327,$$

$$\hat{\lambda}_{1Z} = 0.1408,\ \hat{\lambda}_{2Z} = 0.0346,\ \hat{\lambda}_{3Z} = 0.0334,\ \hat{\lambda}_{4Z} = 0.0326,\ \hat{\lambda}_{5Z} = 0.0318.$$

Note that assumptions (4.5) do not hold. We have $s_X = 0.6337$ and $s_Z = 0.5487$. So $\hat{\sigma} = 0.5487$. Using Proposition 3.2, we obtained the following approximations to power, say π_n^*, for .05 level tests assuming model X is true, namely

$$\hat{\pi}_n^*(J) = 0.9781, \qquad \hat{\pi}_n^*(JA) = 0.7767, \qquad \hat{\pi}_n^*(G) = 0.9431,$$

$$\hat{\pi}_n^*(FJ) = 0.9496, \qquad \hat{\pi}_n^*(FJA) = 0.2571, \qquad \hat{\pi}_n^*(FG) = 0.9071. \qquad (4.8)$$

Similarly, approximations to π_n^* under model Z are:

$$\hat{\pi}_n^*(J) = 0.9822, \qquad \hat{\pi}_n^*(JA) = 0.5949, \qquad \hat{\pi}_n^*(G) = 0.9356,$$

$$\hat{\pi}_n^*(FJ) = 0.9470, \qquad \hat{\pi}_n^*(FJA) = 0.0864, \qquad \hat{\pi}_n^*(FG) = 0.8771. \qquad (4.9)$$

Thus J is most powerful, while JA and FJA are evidently considerably less powerful. These power estimates are for the nested power and that the actual power for this scenario will be about 5% less than the $\hat{\pi}_n^*$s in (4.8) and (4.9); cf. section 1. For these data, we obtained the following test statistics:

$$J_X = 2.91, \quad J_Z = 4.26, \quad JA_X = -1.01, \quad JA_Z = 1.68,$$

$$G_X = 0.20, \quad G_Z = 2.29, \quad FJ_X = 5.04, \quad FJ_Z = 8.40,$$

$$FJA_X = -0.59, \quad FJA_Z = 0.85, \quad FG_X = 1.16, \quad FG_Z = 2.56.$$

The multiple correlation between the models X and Z was 0.98, R_X^2 was approximately 86% and R_Z^2 was approximately 83%. The tests JA, FG and G lead us to the conclusion that model X was preferable to model Z. The tests J and FJ reject both models at the 1% level of significance, while FJA does not reject either model. These results are not surprising, as the power indicated that the tests G and FG would probably be decisive while FJA would tend to be indecisive. The *J-test* may reject too often in practice (cf. McAleer, 1995), and we conjecture that it is also the case that FJ will have larger than nominal size. Therefore, as some preliminary work has been done to find finite-sample corrections for the *J-test* (cf. Royston and Thompson, 1995) and others have used the bootstrap as a means of correcting the size (Godfrey, 1998), we believe it would be worthwhile to thoroughly investigate such corrections for both J and FJ. In the meantime, with relatively small sample sizes, practitioners may prefer to use the more robust method presented in Victoria-Feser (1997) or the less powerful alternatives G, FG, and JA.

5. Concluding Remarks

We have considered non-nested hypothesis testing procedures in the linear regression setting. We demonstrated that the asymptotic means of the most commonly-used methods (cf. McAleer, 1995) are positive and thus, the procedures should be treated as one-sided tests when one is comparing two specified models. Furthermore, we presented a way to estimate the power of non-nested model selection tests when the true parameter vector and variance can be consistently estimated. In addition, we have given a method for theoretically judging which procedure is the most powerful by using measures of asymptotic relative efficiency. Interestingly, the asymptotic relative efficiencies did not depend upon the norm of the parameter vector. For the cases of $p = 1$ or $q = 1$, we saw that the efficiency depends only upon how "close" the design matrices are and how "close" the mean vector is to the alternative design's column space. This is consistent with the results obtained in the local alternatives case (Pesaran, 1987).

In the general case, the relative efficiencies will involve the canonical correlation matrix M and the vector β. We saw that J is theoretically preferable to JA in general. For the Bahadur (1961) measure of asymptotic relative efficiency (cf. (4.3)), we were able to conclude that $FJ \succ G \succ J \succ FG \succ FJA$. For the Hájek and Šidák (1967, p. 267) measure, we saw that $FG \succ J \succ G \succ FJ$, under assumptions (4.5).

We remind the reader that these asymptotic results do not take into account the possibility of incorrect size. Therefore, if one does not have a large sample the most powerful test, as determined by use of this methodology, may not be the best test to use (McAleer, 1995).

Acknowledgements. We are indebted to an anonymous referee for helpful suggestions. We also thank Ron Christensen for his numerous useful comments on earlier versions of this paper.

A. **Appendix to Section 2**

Under the conditions necessary for the Lindeberg-Feller Central Limit Theorem (cf. Loève, 1977, pp. 292-294), we have the following distributional result

$$n^{1/2}\left\{\begin{pmatrix} W_X \\ W_Z \end{pmatrix} - \begin{pmatrix} \mu_x \\ \mu_z \end{pmatrix}\right\} \to N\left\{0, \sigma_x^2\begin{pmatrix} I & R^T \\ R & I \end{pmatrix}\right\}, \qquad (n \to \infty), \quad (A.1)$$

under (2.1). Practically speaking, the conditions sufficient for (A.1) to hold are that the matrices C_{XX} and C_{ZZ} are invertible and finite, i.e., the matrices X^TX and Z^TZ are all of order n and of full rank, and the error term ϵ is either bounded or derives from a distribution with at least 2 moments and sufficiently small tails.

To prove claim (a) of Proposition 2.1, we recall that the characteristic roots of M, which we call m_i for $i = 1, ..., p$ here, are between 0 and 1. Let $\tilde{M}_i$ be the eigenvectors of M and $d_i = \tilde{M}_i^T \mu_x$. Then $\lambda_j = \sum_{i=1}^{p} m_i^{j-1}(1 - m_i)d_i^2$. Letting $w_i = (1 - m_i)d_i^2/\lambda_1$, we have that the w_is define a probability measure. So

$$\lambda_2^2 \le \lambda_1\lambda_3 \iff \left(\sum_{i=1}^{p} w_i m_i\right)^2 \le \sum_{i=1}^{p} w_i m_i^2,$$

which is obviously true. Claims (b) and (c) follow similarly. Claim (d) follows from claims (a) and (b) and that $\lambda_i \ge 0$.

B. **Appendix to Section 3**

The calculational and theoretical forms for J_Z (cf. (3.1)) are

$$J_Z \equiv \frac{||P_XY||^2 - Y^TP_XP_ZY}{n^{1/2}\hat{\sigma}\left(||P_XY||^2 - ||P_ZP_XY||^2\right)^{1/2}} = \frac{||W_X||^2 - W_X^TA^TW_Z}{\hat{\sigma}\left(||W_X||^2 - ||AW_X||^2\right)^{1/2}}.$$

The theoretical form of JA_Z (cf. (3.2)) is

$$JA_Z \equiv \frac{W_X^T A^T W_Z - ||A^T W_Z||^2}{\hat{\sigma}\,(||A^T W_Z||^2 - ||AA^T W_Z||^2)^{1/2}}.$$

Similarly, the theoretical form of G_Z (cf. (3.3)) is

$$G_Z \equiv \frac{||W_X||^2 - ||A^T W_Z||^2}{\{4\hat{\sigma}^2\,(||A^T W_Z||^2 - ||AA^T W_Z||^2)\}^{1/2}}.$$

For FG_Z, FJ_Z, and FJA_Z (cf. (3.6), (3.7), (3.8)), the theoretical forms are

$$FG_Z \equiv \frac{||W_X||^2 - ||A^T W_Z||^2}{2\hat{\sigma}\,(||W_X||^2 - ||AW_X||^2)^{1/2}};$$

$$FJ_Z \equiv \frac{||W_X||^2 - W_X^T A^T W_Z}{\hat{\sigma}\,(||A^T W_Z||^2 - ||AA^T W_Z||^2)^{1/2}};$$

and

$$FJA_Z \equiv \frac{W_X^T A^T W_Z - ||A^T W_Z||^2}{\hat{\sigma}\,(||W_X||^2 - ||AW_X||^2)^{1/2}}.$$

We now prove Proposition 3.2. First, note that the numerators for each of J_Z, JA_Z, G_Z, FJ_Z, FJA_Z, and FG_Z may be written as a linear combination of the three statistics $||W_X||^2$, $W_Z^T AW_X$, and $||A^T W_Z||^2$. Let b_1, b_2, b_3 represent the coefficients associated with the three statistics, respectively. That is, for J_X, $b_1 = 1$, $b_2 = -1$, and $b_3 = 0$. Let b_4 be an indicator variable for using $\hat{\tau}_{JAZ}$ in the denominator of the test statistic.

This leads us to the realization that, when calculated under H_Z, the numerators of all of our test statistics may be written in the form

$$u_n \equiv b_1||W_X||^2 + b_2 W_Z^T AW_X + b_3||A^T W_Z||^2,$$

where $b_1, b_2, b_3 \in \{-1, 0, 1\}$. Under H_X, $u_n \xrightarrow{p} b_1\xi_0 + b_2\xi_1 + b_3\xi_2$. Therefore, all of our test statistics may be written as

$$S_n = \frac{u_n}{c\,\{(1 - b_4)\hat{\tau}_{JZ} + b_4\hat{\tau}_{JAZ}\}},$$

where $c = 2$ if we are using G or FG and it is 1 otherwise and $b_4 \in \{0, 1\}$.

We can now obtain the partial derivatives of S_n with respect to W_X and W_Z.

$$\frac{\partial S_n}{\partial W_X} = \frac{(\partial u_n/\partial W_X)}{c\{(1-b_4)\hat{\tau}_{JZ} + b_4\hat{\tau}_{JAZ}\}} + \frac{u_n}{c}\left\{(1-b_4)\frac{\partial \hat{\tau}_{JZ}^{-1}}{\partial W_X} + b_4\frac{\partial \hat{\tau}_{JAZ}^{-1}}{\partial W_X}\right\}$$
$$\frac{\partial S_n}{\partial W_Z} = \frac{(\partial u_n/\partial W_Z)}{c\{(1-b_4)\hat{\tau}_{JZ} + b_4\hat{\tau}_{JAZ}\}} + \frac{u_n}{c}\left\{(1-b_4)\frac{\partial \hat{\tau}_{JZ}^{-1}}{\partial W_Z} + b_4\frac{\partial \hat{\tau}_{JAZ}^{-1}}{\partial W_Z}\right\} \qquad (B.1)$$

We note that, as alluded to in section 3.3, when $X \perp Z$, these derivatives are infinite when $b_4 = 1$, because $\hat{\tau}_{JAZ} = 0$. So, these results do not hold for JA_Z, G_Z and FJ_Z, if $X \perp Z$.

Recalling the notational definitions in equations (2.4), (2.5), (2.6) and (2.7), we note that $\mu_z = R\mu_x$. Using all of these, we obtain:

$$\frac{\partial u_n}{\partial W_X} = 2b_1 W_X + b_2 A^T W_Z \xrightarrow{p} 2b_1\mu_x + b_2 R^T\mu_z = 2b_1\mu_x + b_2 R^T R\mu_x;$$
$$\frac{\partial u_n}{\partial W_Z} = b_2 A W_X + 2b_3 A A^T W_Z$$
$$\xrightarrow{p} b_2\mu_z + 2b_3 R R^T \mu_z = b_2 R\mu_x + 2b_3 R R^T R\mu_x; \qquad (B.2)$$

where the convergences are under H_X. We now obtain the asymptotic norms:

$$\left\|\frac{\partial u_n}{\partial W_X}\right\|^2 \to 4b_1^2\xi_0 + b_2^2\xi_2 + 4b_1b_2\xi_1 \equiv U_X^*.$$
$$\left\|\frac{\partial u_n}{\partial W_Z}\right\|^2 \to b_2^2\xi_1 + 4b_3^2\xi_3 + 4b_2b_3\xi_2 \equiv U_Z^*.$$
$$\left(\frac{\partial u_n}{\partial W_X}\right)^T A^T \left(\frac{\partial u_n}{\partial W_Z}\right) \to 2b_1b_2\xi_1 + 4b_1b_3\xi_2 + b_2^2\xi_2 + 2b_2b_3\xi_3 \equiv U_{XZ}^*. \qquad (B.3)$$

From (3.4), we derive the derivatives of $\hat{\tau}_{JZ}^{-1}$ and $\hat{\tau}_{JAZ}^{-1}$:

$$\frac{\partial \hat{\tau}_{JZ}^{-1}}{\partial W_X} = \frac{-\hat{\sigma}\left(W_X - A^T A W_X\right)}{\hat{\tau}_{JZ}^3} \quad \frac{\partial \hat{\tau}_{JAZ}^{-1}}{\partial W_X} = 0$$
$$\frac{\partial \hat{\tau}_{JZ}^{-1}}{\partial W_Z} = 0 \quad \frac{\partial \hat{\tau}_{JAZ}^{-1}}{\partial W_Z} = \frac{-\hat{\sigma}\left(AA^T W_Z - AA^T AA^T W_Z\right)}{\hat{\tau}_{JAZ}^3}. \qquad (B.4)$$

Now, noting that b_4 is an indicator, we calculate the partial derivatives of S_n using (B.1)-(B.4):

$$\begin{aligned}\frac{\partial S_n}{\partial W_X} &= \frac{(\partial u_n/\partial W_X)}{c\{(1-b_4)\hat{\tau}_{JZ}+b_4\hat{\tau}_{JAZ}\}} - \frac{(1-b_4)u_n\hat{\sigma}}{c\hat{\tau}_{JZ}^3}\left\{(I-A^TA)W_X\right\}\\ \frac{\partial S_n}{\partial W_Z} &= \frac{(\partial u_n/\partial W_Z)}{c\{(1-b_4)\hat{\tau}_{JZ}+b_4\hat{\tau}_{JAZ}\}} - \frac{b_4u_n\hat{\sigma}}{c\hat{\tau}_{JAZ}^3}(AA^T-AA^TAA^T)W_Z.\end{aligned} \qquad (B.5)$$

From (2.7), (3.5), and (B.1)-(B.5), we see that

$$\begin{aligned}\left\|\frac{\partial S_n}{\partial W_X}\right\|^2 \to\ & \frac{U_X^*}{c^2\sigma_x^2\{(1-b_4)\lambda_1+b_4\lambda_3\}}\\ &+\left\{\frac{(1-b_4)(b_1\xi_0+b_2\xi_1+b_3\xi_2)}{c^2\sigma_x^2\lambda_1^2}\right\}\left\{\frac{(b_1\xi_0+b_2\xi_1+b_3\xi_2)(\lambda_1-\lambda_2)}{\lambda_1}\right\}\\ &-\left\{\frac{2(1-b_4)(b_1\xi_0+b_2\xi_1+b_3\xi_2)}{c^2\sigma_x^2\lambda_1}\right\}\left(\frac{2b_1\lambda_1+b_2\lambda_2}{\lambda_1}\right).\\ \left\|\frac{\partial S_n}{\partial W_Z}\right\|^2 \to\ & \frac{U_Z^*}{c^2\sigma_x^2\{(1-b_4)\lambda_1+b_4\lambda_3\}}\\ &+\left\{\frac{b_4(b_1\xi_0+b_2\xi_1+b_3\xi_2)}{c^2\sigma_x^2\lambda_3^2}\right\}\left[\left\{\frac{b_1\xi_0+b_2\xi_1+b_3\xi_2}{\lambda_3}\right\}(\lambda_4-\lambda_5)\right]\\ &-\left\{\frac{2b_4(b_1\xi_0+b_2\xi_1+b_3\xi_2)}{c^2\sigma_x^2\lambda_3}\right\}\left(\frac{b_2\lambda_3+2b_3\lambda_4}{\lambda_3}\right).\end{aligned}$$

$$\begin{aligned}\left(\frac{\partial S_n}{\partial W_X}\right)^T A^T\left(\frac{\partial S_n}{\partial W_Z}\right) \to\ & \frac{U_{XZ}^*}{c^2\sigma_x^2\{(1-b_4)\lambda_1+b_4\lambda_3\}}\\ &-\left\{\frac{(1-b_4)(b_1\xi_0+b_2\xi_1+b_3\xi_2)}{c^2\sigma_x^2\lambda_1^2}\right\}(b_2\lambda_2+2b_3\lambda_3)\\ &-\left\{\frac{b_4(b_1\xi_0+b_2\xi_1+b_3\xi_2)}{c^2\sigma_x^2\lambda_3^2}\right\}(2b_1\lambda_3+b_2\lambda_4).\end{aligned} \qquad (B.6)$$

Using the delta method, the asymptotic variance of $n^{1/2}S_n$ under (2.1) is

$$\begin{aligned}&\left(\left[\frac{\partial S}{\partial W_X}|_\mu\right]^T,\ \left[\frac{\partial S}{\partial W_Z}|_\mu\right]^T\right)\sigma_x^2\begin{pmatrix} I & R^T\\ R & I\end{pmatrix}\begin{pmatrix}\frac{\partial S}{\partial W_X}|_\mu\\ \frac{\partial S}{\partial W_Z}|_\mu\end{pmatrix}\\ &\quad=\sigma_x^2\left\{\left\|\frac{\partial S}{\partial W_X}\right\|^2+\left\|\frac{\partial S}{\partial W_Z}\right\|^2+2\left(\frac{\partial S}{\partial W_X}\right)^T R^T\left(\frac{\partial S}{\partial W_Z}\right)\right\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{U_X^* + U_Z^* + 2U_{XZ}^*}{c^2\left\{(1-b_4)\lambda_1 + b_4\lambda_3\right\}} \\
&\quad + \left\{\frac{(1-b_4)\left(b_1\xi_0 + b_2\xi_1 + b_3\xi_2\right)}{c^2\lambda_1^2}\right\}\left\{\frac{\left(b_1\xi_0 + b_2\xi_1 + b_3\xi_2\right)(\lambda_1 - \lambda_2)}{\lambda_1}\right\} \\
&\quad - \left\{\frac{(1-b_4)\left(b_1\xi_0 + b_2\xi_1 + b_3\xi_2\right)}{c^2\lambda_1}\right\}\left(\frac{4b_1\lambda_1 + 4b_2\lambda_2 + 4b_3\lambda_3}{\lambda_1}\right) \\
&\quad + \left\{\frac{b_4\left(b_1\xi_0 + b_2\xi_1 + b_3\xi_2\right)}{c^2\lambda_3^2}\right\}\left[\left\{\frac{b_1\xi_0 + b_2\xi_1 + b_3\xi_2}{\lambda_3}\right\}(\lambda_4 - \lambda_5)\right] \\
&\quad - \left\{\frac{b_4\left(b_1\xi_0 + b_2\xi_1 + b_3\xi_2\right)}{c^2\lambda_3}\right\}\left\{\frac{(4b_1 + 2b_2)\lambda_3 + (2b_2 + 4b_3)\lambda_4}{\lambda_3}\right\}.
\end{aligned} \tag{B.7}
$$

For J_Z, we have $b_1 = 1, b_2 = -1, b_3 = 0, b_4 = 0, c = 1$. Therefore, according to (B.7), the asymptotic variance of J_Z under H_X is:

$$
\begin{aligned}
&\frac{(4\xi_0 + \xi_2 - 4\xi_1) + (\xi_1) + 2(-2\xi_1 + \xi_2)}{\lambda_1} + (\xi_0 - \xi_1)\left\{\frac{(\xi_0 - \xi_1)(\lambda_1 - \lambda_2)}{\lambda_1^3}\right\} \\
&\quad - \left\{\frac{(\xi_0 - \xi_1)}{\lambda_1}\right\}\left(\frac{4\lambda_1 - 4\lambda_2}{\lambda_1}\right) \\
&\qquad = \frac{4\lambda_1 - 3\lambda_2}{\lambda_1} + \frac{(\lambda_1 - \lambda_2)}{\lambda_1} - \left(\frac{4\lambda_1 - 4\lambda_2}{\lambda_1}\right) = 1.
\end{aligned} \tag{B.8}
$$

For JA_Z, we have $b_1 = 0, b_2 = 1, b_3 = -1, b_4 = 1, c = 1$. Therefore, according to (B.7), the asymptotic variance of JA_Z under H_X is:

$$
\begin{aligned}
&\frac{(\xi_2) + (\xi_1 + 4\xi_3 - 4\xi_2) + 2(\xi_2 - 2\xi_3)}{\lambda_3} + \left(\frac{\xi_1 - \xi_2}{\lambda_3^2}\right)\left(\frac{\xi_1 - \xi_2}{\lambda_3}\right)(\lambda_4 - \lambda_5) \\
&\quad - \left(\frac{\xi_1 - \xi_2}{\lambda_3}\right)\left(\frac{2\lambda_3 - 2\lambda_4}{\lambda_3}\right) \\
&= \frac{\xi_1 - \xi_2}{\lambda_3} + \left(\frac{\lambda_2^2}{\lambda_3^3}\right)(\lambda_4 - \lambda_5) - \left(\frac{2\lambda_2}{\lambda_3}\right)\left(\frac{\lambda_3 - \lambda_4}{\lambda_3}\right) \\
&= \frac{\lambda_2}{\lambda_3}\left\{\left(\frac{\lambda_2}{\lambda_3^2}\right)(\lambda_4 - \lambda_5) + 2\left(\frac{\lambda_4}{\lambda_3}\right) - 1\right\}.
\end{aligned} \tag{B.9}
$$

For G_Z, we have $b_1 = 1, b_2 = 0, b_3 = -1, b_4 = 1, c = 2$. Therefore, according to (B.7), the asymptotic variance of G_Z under H_X is:

$$\begin{aligned}
&\frac{(4\xi_0)+(4\xi_3)+2(-4\xi_2)}{4\lambda_3}+\left(\frac{\xi_0-\xi_2}{4\lambda_3^2}\right)\left(\frac{\xi_0-\xi_2}{\lambda_3}\right)(\lambda_4-\lambda_5) \\
&\quad-\left(\frac{\xi_0-\xi_2}{4\lambda_3}\right)\left(\frac{4\lambda_3-4\lambda_4}{\lambda_3}\right) \\
&\quad=\frac{\lambda_1+\lambda_2-\lambda_3}{\lambda_3}+\left(\frac{\lambda_1+\lambda_2}{4\lambda_3^2}\right)\left(\frac{\lambda_1+\lambda_2}{\lambda_3}\right)(\lambda_4-\lambda_5) \\
&\qquad\quad-\left(\frac{\lambda_1+\lambda_2}{4\lambda_3}\right)\left(\frac{4\lambda_3-4\lambda_4}{\lambda_3}\right) \\
&\quad=-1+\left(\frac{\lambda_1+\lambda_2}{\lambda_3}\right)\left\{1+\left(\frac{\lambda_1+\lambda_2}{4\lambda_3^2}\right)(\lambda_4-\lambda_5)-\left(\frac{\lambda_3-\lambda_4}{\lambda_3}\right)\right\} \\
&\quad=-1+\left(\frac{\lambda_1+\lambda_2}{\lambda_3^2}\right)\left\{\left(\frac{\lambda_1+\lambda_2}{4\lambda_3}\right)(\lambda_4-\lambda_5)+\lambda_4\right\}.
\end{aligned} \tag{B.10}$$

For FJ_Z, we have $b_1=1, b_2=-1, b_3=0, b_4=1, c=1$. Therefore, according to (B.7), the asymptotic variance of FJ_Z under H_X is:

$$\begin{aligned}
&\frac{(4\xi_0+\xi_2-4\xi_1)+(\xi_1)+2(-2\xi_1+\xi_2)}{\lambda_3}+\left(\frac{\xi_0-\xi_1}{\lambda_3^2}\right)\left(\frac{\xi_0-\xi_1}{\lambda_3}\right)(\lambda_4-\lambda_5) \\
&-\left(\frac{\xi_0-\xi_1}{\lambda_3}\right)\left(\frac{2\lambda_3-2\lambda_4}{\lambda_3}\right) \\
&\quad=\frac{4\lambda_1-3\lambda_2}{\lambda_3}+\left(\frac{\lambda_1^2}{\lambda_3^3}\right)(\lambda_4-\lambda_5)-\left(\frac{2\lambda_1}{\lambda_3}\right)\left(\frac{\lambda_3-\lambda_4}{\lambda_3}\right) \\
&\quad=\frac{2\lambda_1-3\lambda_2}{\lambda_3}+\left(\frac{\lambda_1^2}{\lambda_3^3}\right)(\lambda_4-\lambda_5)+\frac{2\lambda_1\lambda_4}{\lambda_3}.
\end{aligned} \tag{B.11}$$

For FJA_Z, we have $b_1=0, b_2=1, b_3=-1, b_4=0, c=1$. Therefore, according to (B.7), the asymptotic variance of FJA_Z under H_X is:

$$\begin{aligned}
&\frac{(\xi_2)+(\xi_1+4\xi_3-4\xi_2)+2(\xi_2-2\xi_3)}{\lambda_1}+\left(\frac{\xi_1-\xi_2}{\lambda_1^2}\right)\left\{\frac{(\xi_1-\xi_2)(\lambda_1-\lambda_2)}{\lambda_1}\right\} \\
&-\left(\frac{\xi_1-\xi_2}{\lambda_1}\right)\left(\frac{4\lambda_2-4\lambda_3}{\lambda_1}\right) \\
&\quad=\frac{\lambda_2}{\lambda_1}+\left(\frac{\lambda_2}{\lambda_1^2}\right)\left\{\frac{\lambda_2(\lambda_1-\lambda_2)}{\lambda_1}\right\}-\left(\frac{\lambda_2}{\lambda_1}\right)\left(\frac{4\lambda_2-4\lambda_3}{\lambda_1}\right).
\end{aligned} \tag{B.12}$$

We note that, when $X \perp Z$, as is considered in section 3.3, this asymptotic variance is 0.

For FG_Z, we have $b_1 = 1, b_2 = 0, b_3 = -1, b_4 = 0, c = 2$. Therefore, according to (B.7), the asymptotic variance of FG_Z under H_X is:

$$\begin{aligned}
&\frac{(4\xi_0)+(4\xi_3)+2(-4\xi_2)}{4\lambda_1} + \left(\frac{\xi_0-\xi_2}{4\lambda_1^2}\right)\left\{\frac{(\xi_0-\xi_2)(\lambda_1-\lambda_2)}{\lambda_1}\right\} \\
&- \left(\frac{\xi_0-\xi_2}{4\lambda_1}\right)\left(\frac{4\lambda_1-4\lambda_3}{\lambda_1}\right) \\
&= \frac{4\lambda_1+4\lambda_2-4\lambda_3}{4\lambda_1} + \left(\frac{\lambda_1+\lambda_2}{4\lambda_1^2}\right)\left\{\frac{(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)}{\lambda_1}\right\} \qquad (B.13) \\
&\quad - \left(\frac{\lambda_1+\lambda_2}{4\lambda_1}\right)\left(\frac{4\lambda_1-4\lambda_3}{\lambda_1}\right) \\
&= \left(\frac{\lambda_1+\lambda_2}{4\lambda_1^2}\right)\left\{\frac{(\lambda_1+\lambda_2)(\lambda_1-\lambda_2)}{\lambda_1}\right\} + \left(\frac{\lambda_2\lambda_3}{\lambda_1^2}\right).
\end{aligned}$$

The proof of the Corollary to Proposition 3.2 follows.

$$Var(JA_Z) = \frac{\lambda_2}{\lambda_3}\left\{-1 + \frac{2\lambda_4}{\lambda_3} + \frac{\lambda_2(\lambda_4-\lambda_5)}{\lambda_3^2}\right\}.$$

This is greater than or equal to 1 if and only if

$$\begin{aligned}
-1 + \frac{2\lambda_4}{\lambda_3} + \frac{\lambda_2(\lambda_4-\lambda_5)}{\lambda_3^2} &\geq \frac{\lambda_3}{\lambda_2} \\
\Leftarrow \quad -1 + \frac{2\lambda_4}{\lambda_3} &\geq \frac{\lambda_3}{\lambda_2} \\
\Longleftrightarrow \quad \frac{\lambda_2\lambda_4}{\lambda_3^2} &\geq 1.
\end{aligned}$$

The last statement is true by Proposition 2.1.

C. **Appendix to Section 4**

To show the Bahadur ordering, we have

$$e_B(FJ,G) = e_B(J,FG) = \left(\frac{\frac{\lambda_1+\lambda_2}{2\sigma\lambda_3^{1/2}}}{\frac{\lambda_1}{\sigma\lambda_3^{1/2}}}\right)^2 = \left(\frac{\lambda_1+\lambda_2}{2\lambda_1}\right)^2 < 1.$$

$$e_B(FG,FJA) = \left(\frac{\frac{\lambda_2}{\sigma\lambda_3^{1/2}}}{\frac{\lambda_1+\lambda_2}{2\sigma\lambda_3^{1/2}}}\right)^2 = \left(\frac{2\lambda_2}{\lambda_1+\lambda_2}\right)^2 < 1.$$

$$e_B(FJ,J) = e_B(G,FG) = \frac{\lambda_3}{\lambda_1} < 1.$$

Now, we compare J and G:

$$e_B(G,J) = 4\left\{\frac{\lambda_1\lambda_3}{(\lambda_1+\lambda_2)^2}\right\}. \qquad (C.1)$$

To show that $e_B(G,J) < 1$, we note that $4\left\{\frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}\right\} \le 1$ implies $4\left\{\frac{\lambda_1\lambda_3}{(\lambda_1+\lambda_2)^2}\right\} < 1$ because $\lambda_3 < \lambda_2$. But,

$$\begin{aligned} 4\left\{\frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}\right\} \le 1 \quad &\iff \quad 4(\lambda_1\lambda_2) \le (\lambda_1+\lambda_2)^2 \\ &\iff \quad 0 \le \lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2 = (\lambda_1-\lambda_2)^2, \end{aligned}$$

Now, we show $e_{HS}(J,FG) \ge 1$. Beginning with (4.4), we have

$$\begin{aligned} e_{HS}(J,FG) &= \left[\frac{\frac{\lambda_1+\lambda_2}{2\lambda_1^{1/2}}}{\lambda_1^{1/2}\sqrt{\frac{1}{4\lambda_1}\left\{\left(\frac{\lambda_1+\lambda_2}{\lambda_1}\right)^2(\lambda_1-\lambda_2)+4\left(\frac{\lambda_2\lambda_3}{\lambda_1}\right)\right\}}}\right]^2 \\ &= \left[\frac{\lambda_1+\lambda_2}{\sqrt{\left\{\frac{(\lambda_1+\lambda_2)^2}{\lambda_1}\right\}(\lambda_1-\lambda_2)+4(\lambda_2\lambda_3)}}\right]^2 \\ &= \frac{(\lambda_1+\lambda_2)^2}{\left\{\frac{(\lambda_1+\lambda_2)^2}{\lambda_1}\right\}(\lambda_1-\lambda_2)+4(\lambda_2\lambda_3)} \\ &= \frac{1}{\frac{1}{\lambda_1}(\lambda_1-\lambda_2)+4\left\{\frac{\lambda_2\lambda_3}{(\lambda_1+\lambda_2)^2}\right\}} \end{aligned}$$

This is greater than, or equal to, 1 if and only if

$$\begin{aligned} 1 &\ge \frac{(\lambda_1-\lambda_2)}{\lambda_1} + 4\left\{\frac{\lambda_2\lambda_3}{(\lambda_1+\lambda_2)^2}\right\} \\ \iff \frac{\lambda_2}{\lambda_1} &\ge 4\left\{\frac{\lambda_2\lambda_3}{(\lambda_1+\lambda_2)^2}\right\} \\ \iff 1 &\ge 4\left\{\frac{\lambda_1\lambda_3}{(\lambda_1+\lambda_2)^2}\right\}. \end{aligned}$$

But the right-hand side is equivalent to $e_B(G,J)$ (see equation 4.A.1). So, this completes the proof.

We now show $e_{HS}(G,J) \geq 1$. For HS efficiency, we consider the relative efficiencies between G and J, and between G and FJ.

$$\begin{aligned} e_{HS}(G,J) &= \frac{\lambda_1 \left[\left(\frac{\lambda_1+\lambda_2}{\lambda_3^2} \right) \left\{ \lambda_4 + \frac{(\lambda_1+\lambda_2)(\lambda_4-\lambda_5)}{4\lambda_3} \right\} - 1 \right]}{(\lambda_1+\lambda_2)^2/4\lambda_3} \\ &= \frac{4\lambda_1 \left[(\lambda_1+\lambda_2) \left\{ \lambda_4 + \frac{(\lambda_1+\lambda_2)(\lambda_4-\lambda_5)}{4\lambda_3} \right\} - \lambda_3^2 \right]}{(\lambda_1+\lambda_2)^2 \lambda_3}. \end{aligned}$$

This is greater than, or equal to, 1, if and only if

$$\begin{aligned} 4\lambda_1 \left[(\lambda_1+\lambda_2) \left\{ 4\lambda_3\lambda_4 + (\lambda_1+\lambda_2)(\lambda_4-\lambda_5) \right\} - 4\lambda_3^3 \right] &\geq 4(\lambda_1+\lambda_2)^2 \lambda_3^2 \\ \lambda_1 \left[(\lambda_1+\lambda_2)(4\lambda_3\lambda_4 + \lambda_1\lambda_4 + \lambda_2\lambda_4 - \lambda_1\lambda_5 - \lambda_2\lambda_5) - 4\lambda_3^3 \right] & \\ \geq (\lambda_1^2 + 2\lambda_1\lambda_2 + \lambda_2^2)\lambda_3^2. & \end{aligned} \tag{C.2}$$

The left-hand side is equivalent to

$$\begin{aligned} &\lambda_1 \left(4\lambda_1\lambda_3\lambda_4 + 4\lambda_2\lambda_3\lambda_4 + \lambda_1^2\lambda_4 + 2\lambda_1\lambda_2\lambda_4 + \lambda_2^2\lambda_4 - \lambda_1^2\lambda_5 - 2\underline{\lambda_1\lambda_2\lambda_5} - \lambda_2^2\lambda_5 - 4\underline{\lambda_3^3} \right) \\ &\quad = \lambda_1 \left(2\lambda_1\lambda_3\lambda_4 + \lambda_1^2\lambda_4 + 2\lambda_1\lambda_2\lambda_4 + \lambda_2^2\lambda_4 - \lambda_1^2\lambda_5 - \lambda_2^2\lambda_5 \right), \end{aligned}$$

where the underscores indicate use of assumptions (4.5). We will use underscores in the following to indicate use of (2.9) and assumptions (4.5). Now, we continue with (C.2):

$$\begin{aligned} & 2\lambda_1^2\lambda_3\lambda_4 + \lambda_1^3\lambda_4 + 2\lambda_1^2\lambda_2\lambda_4 + \lambda_1\lambda_2^2\lambda_4 - \lambda_1^3\lambda_5 - \lambda_1\lambda_2^2\lambda_5 \\ & \geq (\lambda_1^2 + 2\lambda_1\lambda_2 + \lambda_2^2)\lambda_3^2 \\ \iff\ & 2\lambda_1^2\lambda_3\lambda_4 + \lambda_1^3\lambda_4 + 2\lambda_1^2\lambda_2\lambda_4 + \underline{\lambda_1\lambda_2^2\lambda_4} \\ & \geq \lambda_1^3\lambda_5 + \lambda_1\lambda_2^2\lambda_5 + \lambda_1^2\lambda_3^2 + 2\lambda_1\lambda_2\lambda_3^2 + \lambda_2^2\lambda_3^2 \\ \iff\ & 2\lambda_1^2\lambda_3\lambda_4 + \lambda_1^3\lambda_4 + 2\lambda_1^2\lambda_2\lambda_4 \geq \lambda_1^3\lambda_5 + \lambda_1\lambda_2^2\lambda_5 + \underline{\lambda_1^2\lambda_3^2} + \lambda_1\lambda_2\lambda_3^2 + \lambda_2^2\lambda_3^2 \\ \Leftarrow\ & 2\lambda_1^2\lambda_3\lambda_4 + 2\lambda_1^2\lambda_2\lambda_4 \geq \underline{\lambda_1\lambda_2^2\lambda_5} + \lambda_1^2\lambda_2\lambda_4 + \lambda_1\lambda_2\lambda_3^2 + \lambda_2^2\lambda_3^2 \\ \iff\ & 2\lambda_1^2\lambda_3\lambda_4 + \lambda_1^2\lambda_2\lambda_4 \geq \lambda_1\lambda_2\lambda_3\lambda_4 + \underline{\lambda_1\lambda_2\lambda_3^2} + \lambda_2^2\lambda_3^2 \\ \Leftarrow\ & \lambda_1^2\lambda_3\lambda_4 + \lambda_1^2\lambda_2\lambda_4 \geq \lambda_1\lambda_2^2\lambda_4 + \underline{\lambda_2^2\lambda_3^2} \\ \Leftarrow\ & \lambda_1^2\lambda_3\lambda_4 \geq \lambda_1\lambda_2\lambda_3\lambda_4, \end{aligned}$$

which is a true statement.

Next, we compare G to FJ according to the Hájek and Šidák criterion:

$$e_{HS}(G,FJ)=\frac{\frac{\lambda_1^2}{\lambda_3}\left[\left(\frac{\lambda_1+\lambda_2}{\lambda_3^2}\right)\left\{\lambda_4+\frac{(\lambda_1+\lambda_2)(\lambda_4-\lambda_5)}{4\lambda_3}\right\}-1\right]}{\{(\lambda_1+\lambda_2)^2/4\lambda_3^2\}\left\{2\lambda_1-3\lambda_2+2\lambda_4\left(\frac{\lambda_1}{\lambda_3}\right)+\left(\frac{\lambda_1}{\lambda_3}\right)^2(\lambda_4-\lambda_5)\right\}}$$
$$=\frac{\lambda_1^2\left[(\lambda_1+\lambda_2)\{4\lambda_3\lambda_4+(\lambda_1+\lambda_2)(\lambda_4-\lambda_5)\}-4\lambda_3^3\right]}{\lambda_3^2(\lambda_1+\lambda_2)^2\left\{2\lambda_1-3\lambda_2+2\lambda_4\left(\frac{\lambda_1}{\lambda_3}\right)+\left(\frac{\lambda_1}{\lambda_3}\right)^2(\lambda_4-\lambda_5)\right\}}.$$

This is less than, or equal to, 1, if and only if

$$\lambda_1^2\left[(\lambda_1+\lambda_2)\{4\lambda_3\lambda_4+(\lambda_1+\lambda_2)(\lambda_4-\lambda_5)\}-4\lambda_3^3\right]$$
$$\leq(\lambda_1+\lambda_2)^2\left\{2\lambda_1\lambda_3^2-3\lambda_2\lambda_3^2+2\lambda_1\lambda_3\lambda_4+\lambda_1^2(\lambda_4-\lambda_5)\right\}. \qquad (C.3)$$

The left-hand side is equivalent to:

$$\lambda_1^2(\lambda_1+\lambda_2)\{4\lambda_3\lambda_4+(\lambda_1+\lambda_2)(\lambda_4-\lambda_5)\}-4\lambda_1^2\lambda_3^3$$
$$=4\lambda_1^3\lambda_3\lambda_4+4\lambda_1^2\lambda_2\lambda_3\lambda_4+(\lambda_1^4\lambda_4+2\lambda_1^3\lambda_2\lambda_4+\lambda_1^2\lambda_2^2\lambda_4)$$
$$-(\lambda_1^4\lambda_5+2\lambda_1^3\lambda_2\lambda_5+\lambda_1^2\lambda_2^2\lambda_5)-4\underline{\lambda_1^2\lambda_3^3}$$
$$=4\lambda_1^3\lambda_3\lambda_4+\lambda_1^4\lambda_4+2\lambda_1^3\lambda_2\lambda_4+\lambda_1^2\lambda_2^2\lambda_4-\lambda_1^4\lambda_5-2\underline{\lambda_1^3\lambda_2\lambda_5}-\lambda_1^2\lambda_2^2\lambda_5$$
$$=2\lambda_1^3\lambda_3\lambda_4+\lambda_1^4\lambda_4+2\lambda_1^3\lambda_2\lambda_4+\lambda_1^2\lambda_2^2\lambda_4-\lambda_1^4\lambda_5-\lambda_1^2\lambda_2^2\lambda_5,$$

where the underscores above and below indicate the use of (4.5).

The right-hand side of (C.3) is equivalent to:

$$(\lambda_1^2+2\lambda_1\lambda_2+\lambda_2^2)\left(2\lambda_1\lambda_3^2-3\lambda_2\lambda_3^2+2\lambda_1\lambda_3\lambda_4+\lambda_1^2\lambda_4-\lambda_1^2\lambda_5\right)$$
$$=(\lambda_1^4\lambda_4-\lambda_1^4\lambda_5+2\lambda_1^3\lambda_3^2+2\lambda_1^3\lambda_3\lambda_4+2\lambda_1^3\lambda_2\lambda_4-2\underline{\lambda_1^3\lambda_2\lambda_5})$$
$$+(\lambda_1^2\lambda_2\lambda_3^2+4\lambda_1^2\lambda_2\lambda_3\lambda_4)$$
$$+\lambda_1^2\lambda_2^2\lambda_4-\underline{\lambda_1^2\lambda_2^2\lambda_5}-4\lambda_1\lambda_2^2\lambda_3^2+2\lambda_1\lambda_2^2\lambda_3\lambda_4-3\lambda_2^3\lambda_3^2)$$
$$=(\lambda_1^4\lambda_4-\lambda_1^4\lambda_5+2\lambda_1^3\lambda_3^2+2\underline{\lambda_1^3\lambda_2\lambda_4})$$
$$+(\underline{\lambda_1^2\lambda_2\lambda_3^2}+3\lambda_1^2\lambda_2\lambda_3\lambda_4+\lambda_1^2\lambda_2^2\lambda_4-4\lambda_1\lambda_2^2\lambda_3^2+2\lambda_1\lambda_2^2\lambda_3\lambda_4-3\lambda_2^3\lambda_3^2).$$

So, $e_{HS}(G,FJ)\leq 1$ if and only if (transformations using (2.9) or (4.5) are indicated beforehand using underscores)

$$2\lambda_1^3\lambda_3\lambda_4+\lambda_1^4\lambda_4+2\underline{\lambda_1^3\lambda_2\lambda_4}+\lambda_1^2\lambda_2^2\lambda_4-\lambda_1^4\lambda_5-\lambda_1^2\lambda_2^2\lambda_5$$

$$
\begin{aligned}
& \leq \lambda_1^4\lambda_4 - \lambda_1^4\lambda_5 + 4\lambda_1^3\lambda_3^2 + 2\lambda_1^2\lambda_2^2\lambda_4 \\
& +3\lambda_1^2\lambda_2\lambda_3\lambda_4 - 4\lambda_1\lambda_2^2\lambda_3^2 + 2\lambda_1\lambda_2^2\lambda_3\lambda_4 - 3\lambda_2^3\lambda_3^2 \\
\iff \quad & 2\lambda_1^3\lambda_3\lambda_4 \leq \lambda_1^2\lambda_2^2\lambda_5 + 2\lambda_1^3\lambda_3^2 + \lambda_1^2\lambda_2^2\lambda_4 \\
& +3\lambda_1^2\lambda_2\lambda_3\lambda_4 - 4\lambda_1\lambda_2^2\lambda_3^2 + 2\lambda_1\lambda_2^2\lambda_3\lambda_4 - 3\lambda_2^3\lambda_3^2 \\
\Leftarrow \quad & 4\lambda_1\lambda_2^2\lambda_3^2 + 3\lambda_2^3\lambda_3^2 \leq \lambda_1^2\lambda_2^2\lambda_5 + \lambda_1^2\lambda_2^2\lambda_4 + 3\underline{\lambda_1^2\lambda_2\lambda_3\lambda_4} + 2\lambda_1\lambda_2^2\lambda_3\lambda_4 \\
\Leftarrow \quad & \lambda_1\lambda_2^2\lambda_3^2 + 3\lambda_2^3\lambda_3^2 \leq \lambda_1^2\lambda_2^2\lambda_5 + \lambda_1^2\lambda_2^2\lambda_4 + 2\underline{\lambda_1\lambda_2^2\lambda_3\lambda_4} \\
\Leftarrow \quad & \lambda_1\lambda_2^2\lambda_3^2 + \lambda_2^3\lambda_3^2 \leq \underline{\lambda_1^2\lambda_2^2\lambda_5} + \lambda_1^2\lambda_2^2\lambda_4 \\
\Leftarrow \quad & \underline{\lambda_2^3\lambda_3^2} \leq \lambda_1^2\lambda_2^2\lambda_4 \quad \iff \quad \lambda_2^4\lambda_4 \leq \lambda_1^2\lambda_2^2\lambda_4,
\end{aligned}
$$

which is a true statement.

References

BAHADUR, R.R. (1960). Stochastic Comparison of Tests. *Ann. Math. Statist.* **31**, 276-295.

CHRISTENSEN, R. (1987). *Plane Answers to Complex Questions: The Theory of Linear Models*. New York: Springer-Verlag.

COX, D.R. (1961). Tests of separate families of hypotheses. *Proceedings of the 4th Berkeley Symposium on Mathematics, Probability and Statistics*, Vol. **1**, 105-123. Berkeley: University of California Press.

COX, D.R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Stat. Soc.*, Series B, **24**, 406-424.

COX, D.R. and HINKLEY, D.V. (1974). *Theoretical Statistics*. New York: Wiley.

DASTOOR, N.K. and MCALEER, M. (1989). Some power comparisons of joint and paired tests for nonnested models under local hypotheses. *Econometric Theory* **5**, 83-94.

DAVIDSON, R. and MACKINNON, J.G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica* **49**, 781-793.

EFRON, B. (1984). Comparing non-nested linear models. *J. Amer. Statist. Assoc.* **79**, 791-803.

EFRON, B. and TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*, Pacific Grove, CA: Chapman and Hall.

FISHER, G.R. and MCALEER, M. (1981). Alternative procedures and associated tests of significance for non-nested hypotheses. *J. Econometrics* **16**, 103-119.

FISHER, G.R. (1983). Tests for two separate regressions. *J. Econometrics* **21**, 117-132.

GODFREY, L.G. (1998). Tests of non-nested regression models: some results on small sample behaviour and the bootstrap. *J. Econometrics* **84**, 59-74.

GODFREY, L.G. and PESARAN, M.H. (1983). Tests of non-nested regression models. *J. Econometrics* **21**, 133-154.

GRAYBILL, F.A. and IYER, H.K. (1994). *Regression Analysis: Concepts and Applications*. Belmont, CA: Duxbury.

GREEN, J.R. (1971). Testing departure from a regression without using replication. *Technometrics* **13**, 609-615.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* San Francisco: Academic Press.

Hurvich, C.M. and Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Kent, J.T. (1986). The underlying structure of nonnested hypothesis tests. *Biometrika* **73**, 333-343.

Loève, M. (1977). *Probability Theory Volume I.* New York: Springer-Verlag.

Mallows, C.L. (1973). Some comments on C_p. *Technometrics* **15**, 661-675.

McAleer, M. (1995). The significance of testing empirical non-nested models. *J. Econometrics* **67**, 149-171.

McQuarrie, A. and Tsai, C.L. (1998). *Regression And Time Series Model Selection.* River Edge, NJ: World Scientific.

Miller, A.J. (1990). *Subset Selection in Regression.* New York: Chapman and Hall.

Neter, J. Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996). *Applied Linear Statistical Models* (Fourth Edition). Chicago: Irwin.

Pesaran, M.H. (1974). On the general problem of model selection. *Rev. Economic Studies* **41**, 153-171.

Pesaran, M.H. (1982). Comparison of local power of alternative tests of non-nested regression models. *Econometrika* **50**, 1287-1305.

Pesaran, M.H. (1984). Asymptotic power comparisons of tests of separate parametric families by Bahadur's approach. *Biometrika* **71**, 245-252.

Pesaran, M.H. (1987). Global and partial non-nested hypotheses and asymptotic local power. *Econometric Theory* **3**, 69-97.

Rawlings, J.O. (1988). *Applied Regression Analysis.* Pacific Grove, CA: Wadsworth & Brooks / Cole.

Royston, P. and Thompson, S.G. (1995). Comparing Non-Nested Regression Models. *Biometrics* **51**, 114-127.

Szroeter, J. (1995). The exact power function of an exact test of a regression model against multiple separate alternatives. *Comm. Stat. Theor. Meth.* **24**, 2329-2339.

Victoria-Feser, M.P. (1997). A robust test for non-nested hypotheses. *J. Roy. Stat. Soc.* **59**, 715-727.

Watnik, M.R., Johnson, W.O. and Bedrick, E.J. (2001). Non-Nested Linear Model Selection Revisited. *Comm. Stat. Theor. Meth.* **30**, 1-20.

Mitchell Watnik
Statistical Laboratory
University of California
Davis, CA 95616, U.S.A.
E-mail: mrwatnik@ucdavis.edu

Wesley Johnson
Department of Statistics
University of California
Davis, California 95616, U.S.A.
E-mail: wojohnson@ucdavis.edu