GROSS NON-NORMALITY AND THE QUALITY OF A SIMPLE APPROXIMATION TO THE P-VALUE OF A ROUTINE TEST OF NON-NESTED REGRESSIONS

Jerzy Szroeter

UCL Economics Discussion Paper 1998-5

Key Words and Phrases: non-nested regressions, p-value, size, non-normality, point approximation, eigenvalues.

JEL Classification: C1, C2, C4.

ABSTRACT

The distribution of certain test statistics for non-nested regressions can be so grossly non-normal that p-values computed on the assumption of approximate normality cannot be safely used for routine inference. This paper presents results on the quality of a new more accurate yet still user-friendly p-value approximation which embodies an inverse measure of the strength of relationship between regressors of competing models. This easily-computed measure is equivalent to the sum of eigenvalues which have recently been shown to characterize the exact finite-sample distribution of the test statistic.

1. INTRODUCTION

Suppose an observable random n-vector y is ascribed an $N(\mu, \sigma^2 I_n)$ distribution subject to the competing non-nested hypotheses

$$H_0 : \mu = X\beta \qquad versus \qquad H_1 : \mu = Z\gamma$$

where β , γ denote parameter vectors of dimensions k , g and X , Z denote observable nonrandom matrices such that $rank[X] = k$, $rank[Z] = g$, $rank[X, Z] = r$ with $n > r > max\{k, g\}$.

It has been shown by Szroeter (1996) that the critical region for the most powerful test of H_0 versus H_1 applicable when β, γ, σ^2 are of specified value turns into the following region when β, γ, σ^2 are unknown and replaced by least squares estimates :

$$\hat{Q} > C \tag{1.1}$$

where C is an appropriate constant and

$$\hat{Q} \equiv [\hat{\sigma}^2(\hat{\gamma}'Z'M_X Z\hat{\gamma})]^{-1/2}[\hat{\gamma}'Z'M_X y] \quad , \tag{1.2}$$

$$\hat{\gamma} \equiv (Z'Z)^{-1}Z'y \quad , \tag{1.3}$$

$$M_X \equiv I_n - S[X] \quad , \tag{1.4}$$

$$\hat{\sigma}^2 \equiv y'\{I_n - S[X,Z]\}y/(n - r) \quad , \tag{1.5}$$

where $S[A]$ denotes the orthogonal projection operator onto the space spanned by the columns of matrix or matrices A . The value of $\hat{\sigma}\hat{Q}$ is equal to the routinely-computed least squares estimate of the scalar ν in the artificial regression model

$$y = X\beta + \nu(Z\hat{\gamma}) + error.$$

The statistic $\hat{Q}$ itself only differs from the Davidson-Mackinnon (1981) J statistic in small detail : In the formula (1.5) for $\hat{\sigma}^2$, the J statistic uses $[X, Z\hat{\gamma}]$ and $(k + 1)$ instead of $[X,Z]$ and r . For connections with the Cox (1961, 1962) statistic, see McAleer (1987).

A characterization of the unknown exact finite-sample distribution of $\hat{Q}$ has been obtained by Szroeter (1996). The shape and structure of that distribution depends critically on certain eigenvalues to such an extent that location-scale adjustments (see Godfrey and Pesaran (1983)) need not in general reduce $\hat{Q}$ to reliably approximate normal or Student T form. More fundamental adjustments to the components of the statistic may lead to undesirable side-effects. For example, the Fisher-McAleer (1981) and Godfrey (1983) adjustment based on Atkinson (1970) gives an exact test whose power function cuts below size (see Szroeter (1995)). We therefore focus our research here on the unadjusted form of $\hat{Q}$. Of particular concern are tail-probability approximations of the following type based on Szroeter's (1996) characterization :

$$Pr\{ \hat{Q} > C \} \cong$$

$$(1 - \bar{\lambda})Pr\{ T_{n-r} > C \} + \bar{\lambda}Pr\{ F^{r-k}_{n-r} > C^2/(r - k) \} \qquad (1.6)$$

where T_{n-r} , F^{r-k}_{n-r} denote Student T , F variates and

$$\bar{\lambda} = trace\{(Z'Z)^{-1}Z'M_X Z\}/(r - k) . \qquad (1.7)$$

The purpose of the present paper is to derive integral bounds on the true finite-sample size of test region (1.1) and to assess the quality of the approximation (1.6) in the light of these. Of special interest is the case where the upper and lower bounds coincide, giving a precise value for true size. Sections 2 and 3 of the paper set out the theory. Section 4 reports the results of numerical computations.

2. BOUNDS ON TRUE FINITE-SAMPLE SIZE

Let $H_b(\cdot)$, $h_b(\cdot)$ denote the cdf , pdf respectively of a central chi-squared variate with b degrees of freedom. Let $\Phi(\cdot)$, $\phi(\cdot)$ be the cdf , pdf of a unit normal variate. Let $F^a_b(\cdot)$ denote the cdf of a central F variate having a , b numerator , denominator degrees of freedom. Let the scalar m be defined as

$$m \equiv n - r . \tag{2.1}$$

Define the function $S(C,\varepsilon,\psi)$ on $0 < C,\ 0 < \varepsilon < 1,\ 0 < \psi < 1$ as

$$S(C,\varepsilon,\psi) \equiv 1 - \int_0^\infty \int_0^\infty h_m(w) h_{r-k}(\tau) \Phi[(1 - \psi)^{-1/2} m^{-1/2} W^{1/2} C - (1 - \varepsilon)^{-1/2} \varepsilon^{1/2} \tau^{1/2}] d\tau dw \quad . \tag{2.2}$$

Define the function $R(C,\varepsilon,\psi)$ on $0 < C,\ 0 < \varepsilon \le 1,\ 0 < \psi \le 1$ as

$$R(C,\varepsilon,\psi) \equiv$$

$$1 - \int_0^\infty h_m(w) \left\{ \int_0^{L(\psi)} \phi(u) H_{r-k}(\varepsilon^{-1}[m^{-1/2} w^{1/2} C - (1 - \psi)^{1/2} u]^2) du + \int_{-\infty}^0 \phi(u) H_{r-k}(\varepsilon^{-1}[m^{-1/2} w^{1/2} C - (1 - \varepsilon)^{1/2} u]^2) du \right\} dw \tag{2.3}$$

where

$$L(\psi) = \{[(1 - \psi)m]^{-1} w\}^{1/2} C \quad \text{if} \quad \psi < 1 ,$$
$$L(\psi) = \infty \quad \text{otherwise} . \tag{2.4}$$

Define

$$d \equiv (k + g + 1 - r) . \tag{2.5}$$

Now let $\lambda_d \le \lambda_{d+1} \le \ldots \le \lambda_g$ denote the (possibly repeated) nonzero eigenvalues of the matrix product $(Z'Z)^{-1}Z'M_XZ$.

Following Lehmann (1986, p.69), the size $\alpha(C)$ of critical region (1.1) is defined as

$$\alpha(C) \equiv \sup Pr\{ \hat{Q} > C \} \tag{2.6}$$

where the supremum is taken over the set $\{ (\beta, \sigma^2) : \beta \in \mathbb{R}^k, \sigma^2 > 0 \}$. Test p-value is the size function $\alpha(C), C \in \mathbb{R},$ evaluated at the point $C = q$ where q is the realized value of the variate $\hat{Q}$. The results which follow give integral upper and lower bounds for $\alpha(C)$.

THEOREM 1 : *For* $\lambda_g < 1$, $S(C,\lambda_d,\lambda_g) \le \alpha(C) \le S(C,\lambda_g,\lambda_d)$.

THEOREM 2 : *For* $\lambda_g \le 1$, $R(C,\lambda_d,\lambda_g) \le \alpha(C) \le R(C,\lambda_g,\lambda_d)$.

COROLLARY : *For* $\lambda_g = 1, \lambda_d = 1$, $\alpha(C) = 1 - F_m^{r-k}[C^2/(r-k)]$.

3. PROOFS

The proofs of Theorems 1 and 2 depend on the following Lemma which is a special case of Theorem 1 of Szroeter (1996, p.11) :

LEMMA 1 : *Let* U_s $\{s = 1, 2, \ldots, (g+1)\}$ *be independent unit normal variates which are also jointly independent of the central chi-square variate* W *having* m *degrees of freedom. Let* B , D *be the variates*

$$B \;\equiv\; \Big[\sum_{s=d}^{g} \lambda_s (U_s + \theta_s)^2 \Big]^{-1/2} \Big[\sum_{s=d}^{g} \lambda_s (U_s + \theta_s) U_s \Big] \;, \tag{3.1}$$

$$D \;\equiv\; \Big[\sum_{s=d}^{g} \lambda_s (U_s + \theta_s)^2 \Big]^{-1} \Big[\sum_{s=d}^{g} \lambda_s^2 (U_s + \theta_s)^2 \Big] \;, \tag{3.2}$$

where $\theta_s \equiv \sigma^{-1} p_s' Z' X \beta$ *for a right-hand eigenvector* p_s *associated with the eigenvalue* λ_s *of the matrix* $(Z'Z)^{-1} Z' M_X Z$ *, given normalization conditions* $p_s' Z' Z p_s = 1,\; p_s' Z' Z p_t = 0$ *for* $s \neq t$. *Then, under* H_0 *, the exact finite-sample distribution of the statistic* $\hat{Q}$ *defined by equation (1.2) is the same as the distribution of*

$$Q^* \;\equiv\; m^{1/2} W^{-1/2} [B + (1 - D)^{1/2} U_{g+1}] \;. \tag{3.3}$$

We are now in a position to prove Theorems 1 and 2.

PROOF OF THEOREM 1 :

From (3.1) we obtain the inequality

$$|B| \;\leq\; \Big[\sum_{s=d}^{g} \lambda_s U_s^2 \Big]^{1/2} \;\leq\; \lambda_g^{1/2} \Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} \;. \tag{3.4}$$

Using (3.3) and (3.4) we find that

$$Pr\{\, Q^* > C \,\} \;\leq$$

$$Pr\{\, [\lambda_g^{1/2} \Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} + (1 - D)^{1/2} U_{g+1}] > m^{-1/2} W^{1/2} C \,\} \;=$$

$$1 - Pr\{\, [\lambda_g^{1/2} \Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} + (1 - D)^{1/2} U_{g+1}] \leq m^{-1/2} W^{1/2} C \,\}. \tag{3.5}$$

Now observe from (3.2) that

$$0 < \lambda_d \leq D \leq \lambda_g < 1 \quad . \tag{3.6}$$

Given (3.6), we see that from (3.5) that

$$Pr\{ Q^* > C \} \leq$$

$$1 - Pr\{U_{g+1} \leq (1 - D)^{-1/2}(m^{-1/2}W^{1/2}C - \lambda_g^{1/2}\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2})\} \leq$$

$$1 - Pr\{U_{g+1} \leq (1 - \lambda_d)^{-1/2}m^{-1/2}W^{1/2}C - (1 - \lambda_g)^{-1/2}\lambda_g^{1/2}\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2})\}$$

$$= S(C,\lambda_g,\lambda_d) \tag{3.7}$$

where the function $S(\cdot)$ is defined by (2.2). Expression (3.7) is an upper bound on the probability $Pr\{Q^* > C\}$ for each value of (β, σ), hence is an upper bound on $\alpha(C)$ as defined by (2.6).

The basic lower bound on $\alpha(C)$ is (3.7) with λ_d and λ_g interchanged. The justification, however, differs from that for the upper bound. By (2.6) and Lemma 1,

$$\alpha(C) \geq Pr\{ m^{1/2}W^{-1/2}[B^* + (1 - D^*)^{1/2}U_{g+1}] > C \} \tag{3.8}$$

where

$$B^* \equiv \Big[\sum_{s=d}^{g} \lambda_s U_s^2 \Big]^{1/2} \geq \lambda_d^{1/2}\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} , \tag{3.9}$$

$$D^* \equiv \Big[\sum_{s=d}^{g} \lambda_s U_s^2 \Big]^{-1}\Big[\sum_{s=d}^{g} \lambda_s^2 U_s^2 \Big] , \tag{3.10}$$

Since equation (3.6) also holds with D replaced by D^*, we see from (3.8), (3.9) and (3.10) that

$$\alpha(C) \geq 1 - Pr\{ U_{g+1} \leq (1 - D^*)^{-1/2}(m^{-1/2}W^{1/2}C - \lambda_d^{1/2}\left[\sum_{s=d}^{g} U_s^2 \right]^{1/2})\}$$

$$\geq 1 - Pr\{ U_{g+1} \leq (1 - \lambda_g)^{-1/2}m^{-1/2}W^{1/2}C -$$

$$(1 - \lambda_d)^{-1/2}\lambda_d^{1/2}\left[\sum_{s=d}^{g} U_s^2 \right]^{1/2})\}$$

$$= S(C,\lambda_d,\lambda_g) .$$

■

PROOF OF THEOREM 2 :

Given (3.2), observe that

$$Pr\{ [\lambda_g^{1/2}\left[\sum_{s=d}^{g} U_s^2 \right]^{1/2} + (1 - D)^{1/2}U_{g+1}] \leq m^{-1/2}W^{1/2}C \} \geq$$

$$Pr\{ [\left[\sum_{s=d}^{g} U_s^2 \right]^{1/2} + \lambda_g^{-1/2}(1 - \lambda_d)^{1/2}U_{g+1}] \leq \lambda_g^{-1/2}m^{-1/2}W^{1/2}C ,$$

$$U_{g+1} > 0 \} +$$

$$Pr\{ [\left[\sum_{s=d}^{g} U_s^2 \right]^{1/2} + \lambda_g^{-1/2}(1 - \lambda_g)^{1/2}U_{g+1}] \leq \lambda_g^{-1/2}m^{-1/2}W^{1/2}C ,$$

$$U_{g+1} \leq 0 \}$$

$$=$$

$$\int_0^\infty h_m(w)\left\{ \int_0^{L(\lambda_d)} \phi(u)H_{r-k}(\lambda_g^{-1}[m^{-1/2}w^{1/2}C - (1 - \lambda_d)^{1/2}u]^2)du + \right.$$

$$\left. \int_{-\infty}^0 \phi(u)H_{r-k}(\lambda_g^{-1}[m^{-1/2}w^{1/2}C - (1 - \lambda_g)^{1/2}u]^2)du \right\}dw \qquad (3.11)$$

where $L(\cdot)$ is the function defined by (2.4). Now, relation (3.5) of the proof of Theorem 1 continues to hold even when the assumption that $\lambda_g < 1$ is relaxed to $\lambda_g \leq 1$. From (2.6), (3.5), (3.11) and Lemma 1, we then obtain the upper bound of Theorem 2 .

As for the lower bound on $\alpha(C)$, equations (3.8) to (3.10) of the proof of Theorem 1 also continue to hold when the assumption $\lambda_g < 1$ is weakened to $\lambda_g \leq 1$. Therefore,

$$\alpha(C) \geq Pr\{ [\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} + \lambda_d^{-1/2}(1 - D^*)^{1/2} U_{g+1}] > \lambda_d^{-1/2} m^{-1/2} W^{1/2} C ,$$

$$U_{g+1} > 0 \} \quad +$$

$$Pr\{ [\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} + \lambda_d^{-1/2}(1 - D^*)^{1/2} U_{g+1}] > \lambda_d^{-1/2} m^{-1/2} W^{1/2} C ,$$

$$U_{g+1} \leq 0 \}$$

$$\geq Pr\{ [\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} + \lambda_d^{-1/2}(1 - \lambda_g)^{1/2} U_{g+1}] > \lambda_d^{-1/2} m^{-1/2} W^{1/2} C ,$$

$$U_{g+1} > 0 \} \quad +$$

$$Pr\{ [\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} + \lambda_d^{-1/2}(1 - \lambda_d)^{1/2} U_{g+1}] > \lambda_d^{-1/2} m^{-1/2} W^{1/2} C ,$$

$$U_{g+1} \leq 0 \} \quad . \tag{3.12}$$

We now note that

$$Pr\{[\Big[\sum_{s=d}^{g} U_s^2 \Big]^{1/2} + \lambda_d^{-1/2}(1 - \lambda_g)^{1/2} U_{g+1}] \leq \lambda_d^{-1/2} m^{-1/2} W^{1/2} C, \quad U_{g+1} > 0\}$$

$$= \int_0^\infty h_m(w)\left\{\int_0^{L(\lambda_g)} \phi(u) H_{r-k}(\lambda_d^{-1}[m^{-1/2}w^{1/2}C \;-\; (1-\lambda_g)^{1/2}u]^2)du\right\}dw \qquad (3.13)$$

where $L(\cdot)$ *is the function defined by (2.4). Similarly,*

$$Pr\{[\left[\sum_{s=d}^{g} U_s^2\right]^{1/2} + \lambda_d^{-1/2}(1 - \lambda_d)^{1/2}U_{g+1}] \le \lambda_d^{-1/2}m^{-1/2}W^{1/2}C, \quad U_{g+1} \le 0\}$$

$$= \int_{-\infty}^{0} \phi(u)H_{r-k}(\lambda_d^{-1}[m^{-1/2}w^{1/2}C \;-\; (1 \;-\; \lambda_d)^{1/2}u]^2)du\Big\}dw \;. \qquad (3.14)$$

Given the definition (2.3), relations (3.12), (3.13), (3.14) imply that

$$\alpha(C) \;\ge\; R(C,\lambda_d,\lambda_g) \;,$$

which is the lower bound of Theorem 2. ∎

4. NUMERICAL COMPUTATIONS AND COMPARISONS

By Theorems 1 and 2,

$$\alpha(C) = S(C,\lambda,\lambda) = R(C,\lambda,\lambda) \qquad for \quad \lambda_d = \lambda_g = \lambda < 1 \;, \qquad (4.1)$$

hence exact size $\alpha(C)$ for the equal eigenvalue case is known. This enables a precise baseline assessment of the quality of the " *T*, root-*F* mixture" approximation (1.6). In evaluating $S(C,\lambda,\lambda) = R(C,\lambda,\lambda)$ when

$(r - k) > 2$, double integration cannot be avoided but, when $(r - k) = 2$, $R(C,\lambda,\lambda)$ can be rewritten in the single integral form

$$R(C,\lambda,\lambda) = 1 - \int_0^\infty h_m(w)\left\{ \Phi[(1 - \lambda)^{-1/2} m^{-1/2} W^{1/2} C] - \lambda^{1/2} exp[-wC^2/(2m)] \Phi[\lambda^{1/2}(1 - \lambda)^{-1/2} m^{-1/2} W^{1/2} C] \right\} dw . \quad (4.2)$$

We therefore compute $\alpha(C)$, to a numerical accuracy of $\pm$ 0.001, using $\alpha(C) = R(C,\lambda,\lambda)$ with $R(\cdot)$ as given by (4.2) for $(r - k) = 2$, but using $\alpha(C) = S(C,\lambda,\lambda)$ with $S(\cdot)$ as given by (2.2) for $(r - k) = 4, 8$. This This is done for each $\lambda = 0.25, 0.5, 0.75$ across the range $C = 1.3, 1.7, 2.1, 2.5, 2.9, 3.3, 3.7$. We use the typical figure of 30 for m. Alongside each value of $\alpha(C)$ we place for comparison the values of

$$T(C) \equiv Pr\{ T_m > C \} , \quad (4.3)$$

$$F(C) \equiv Pr\{ F_m^{r-k} > C^2/(r - k) \} , \quad (4.4)$$

$$M(C) \equiv (1 - \lambda)T(C) + \lambda F(C) , \quad (4.5)$$

where T_m , F_m^{r-k} denote central Student T, F variates and $M(C)$ is the " T, root-F mixture " approximation (1.6) described in Section 1.

The results of the computations are presented in Tables I, II, III. The general gross understatement by $T(C)$ and overstatement by $F(C)$ of the tail probability $\alpha(C)$ is striking. For example, if we take $C = 1.7$ for a nominal size of 0.05 on the T distribution, then the actual size for the case $\lambda = 0.5$ ranges over 0.18 (when $r = k + 2$) , 0.34 (when

TABLE I

Comparison of Student T, Root-F and Mixture Approximations for the Exact P-Value of a Test of Non-Nested Regressions

The Case $\lambda = 0.25$

r = k + 2

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.187	0.101	0.049	0.022	0.009	0.003	0.001
α(C)	0.240	0.133	0.066	0.030	0.013	0.005	0.002
F(C)	0.440	0.252	0.128	0.059	0.025	0.010	0.004

r = k + 4

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.274	0.183	0.110	0.059	0.029	0.013	0.005
α(C)	0.355	0.218	0.120	0.060	0.027	0.012	0.005
F(C)	0.791	0.583	0.374	0.210	0.105	0.048	0.020

r = k + 8

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.323	0.271	0.219	0.162	0.108	0.064	0.034
α(C)	0.534	0.371	0.232	0.131	0.067	0.032	0.014
F(C)	0.986	0.933	0.808	0.622	0.422	0.253	0.137

TABLE II

Comparison of Student T, Root-F and Mixture Approximations for the Exact P-Value of a Test of Non-Nested Regressions

The Case $\lambda = 0.5$

$r = k + 2$

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.271	0.151	0.075	0.034	0.014	0.006	0.002
α(C)	0.315	0.179	0.091	0.042	0.018	0.007	0.003
F(C)	0.440	0.252	0.128	0.059	0.025	0.010	0.004

$r = k + 4$

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.447	0.317	0.198	0.110	0.054	0.025	0.010
α(C)	0.514	0.340	0.200	0.105	0.050	0.022	0.009
F(C)	0.791	0.583	0.374	0.210	0.105	0.048	0.020

$r = k + 8$

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.544	0.492	0.415	0.316	0.213	0.127	0.069
α(C)	0.769	0.609	0.434	0.277	0.159	0.083	0.040
F(C)	0.986	0.933	0.808	0.622	0.422	0.253	0.137

TABLE III

Comparison of Student T, Root-F and Mixture Approximations for the Exact P-Value of a Test of Non-Nested Regressions

The Case $\lambda = 0.75$

r = k + 2

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.356	0.202	0.102	0.047	0.020	0.008	0.003
α(C)	0.381	0.218	0.111	0.051	0.021	0.008	0.003
F(C)	0.440	0.252	0.128	0.059	0.025	0.010	0.004

r = k + 4

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.619	0.450	0.286	0.160	0.080	0.036	0.015
α(C)	0.657	0.461	0.284	0.155	0.076	0.034	0.015
F(C)	0.791	0.583	0.374	0.210	0.105	0.048	0.020

r = k + 8

C	1.3	1.7	2.1	2.5	2.9	3.3	3.7
T(C)	0.102	0.050	0.022	0.009	0.003	0.001	0.000
M(C)	0.765	0.712	0.612	0.469	0.317	0.190	0.103
α(C)	0.915	0.800	0.632	0.445	0.279	0.157	0.080
F(C)	0.986	0.933	0.808	0.622	0.422	0.253	0.137

$r = k + 4$), 0.61 (when $r = k + 8$). $F(C)$ overestimates these figures as $0.25, 0.58$ *and* 0.93 respectively. The mixture approximation $M(C)$ gives figures of $0.15, 0.32,$ *and* $0.50.$ These are undoubtedly a considerable improvement on $T(C)$ and $F(C)$.

In all cases, the inequality $T(C) < min\{ M(C), \alpha(C) \} \leq max\{ M(C), \alpha(C) \} < F(C)$ holds. When $r = (k + 2)$, the inequality $M(C) \leq \alpha(C)$ holds throughout the computed range of C but, when $r = (k + 8)$, it holds only for the lower C values of $1.3, 1.7, 2.1.$ It switches to $M(C) > \alpha(C)$ at C values of $2.5, 2.9, 3.3, 3.7.$ Switching also occurs for $r = (k + 4)$, but in that case the switch point C depends on the value of λ .

The approximation $M(C)$ seems to be most effective at size levels from 0.2 down. For example, in the case $\lambda = 0.25$, when $M(C)$ indicates levels of $0.10, 0.11, 0.16$ for $r = (k + 2), (k + 4), (k + 8)$, then the true level $\alpha(C)$ equals $0.13, 0.12, 0.13$ respectively. In the case $\lambda = 0.50$, for the same values of r, when $M(C)$ indicates levels of $0.08, 0.11, 0.13$, then $\alpha(C)$ equals $0.09, 0.11, 0.08.$ In the final case $\lambda = 0.75$, for the same values of r, when $M(C)$ equals $0.10, 0.16, 0.10$, then $\alpha(C)$ equals $0.11, 0.16, 0.08.$ The approximation $M(C)$ performs less well at higher p-values than 0.2, but it remains considerably more informative than the very misleading undervaluation $T(C)$.

BIBLIOGRAPHY

Atkinson, A.C. (1970). "A method for discriminating between models," *Journal of the Royal Statistical Society B*, 32, 323-353.

Cox, D.R. (1961). "Tests of separate families of hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 105-23.

Cox, D.R. (1962). "Further results on tests of separate families of hypotheses," *Journal of the Royal Statistical Society B,* 24, 406-424.

Davidson, R. and J.G. MacKinnon (1981). "Several tests for model specification in the presence of alternative hypotheses," *Econometrica*, 47, 781-793.

Fisher, G.R. and M. McAleer (1981). "Alternative procedures and associated tests of significance for non-nested hypotheses," *Journal of Econometrics*, 16, 103-119.

Godfrey, L.G. (1983). "Testing non-nested models after estimation by instrumental variables or least squares," *Econometrica*, 51, 355-365.

Godfrey, L.G. and M.H. Pesaran (1983). "Tests of non-nested regression models: small-sample adjustments and Monte Carlo evidence," *Journal of Econometrics*, 21, 133-154.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* New York: John Wiley & Sons.

McAleer, M (1987). "Specification tests for separate models: a survey," *Specification Analysis in the Linear Model*, (M. King and D. Giles, Eds.). London: Routledge & Kegan Paul, 146-195.

Szroeter, J. (1995). "The exact power function of an exact test of a regression model against multiple separate alternatives," *Communications in Statistics - Theory and Methods*, 24, 2329-2339.

Szroeter, J. (1996). "Characterization of the Exact Finite-Sample Distribution of a Routine Test Statistic for Non-Nested Regressions," *Statistics and Probability Letters*, 29, 9-14.