1995

Non-Nested Pretest Tests

Leo Michelis

ISSN:0318-725X
ISBN:0-7714-1837-X

RESEARCH REPORT 9520

Non-Nested Pretest Tests

by

Leo Michelis

September 1995

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario, Canada
N6A 5C2
econref@sscl.uwo.ca

Non–Nested Pretest Tests

Leo Michelis
Department of Economics
Social Science Centre
The University of Western Ontario
London, Canada N6A 5C2

e-mail: michelis@sscl.uwo.ca

September 1995

Abstract

The J and Cox non–nested tests do not follow the usual $N(0, 1)$ distribution when the non–nested regression models are nearly uncorrelated or orthogonal. This paper attempts to assess the question of whether or not one can do better (in terms of size and power) than existing tests by using two alternative testing procedures based on the notion of pretesting. The Monte Carlo evidence on this question indicates that the testing procedure of first testing the null of zero correlations among the non–nested regressors in two non–nested models and then using either the J test if the null is rejected or the encompassing F test if the null is not rejected, may outperform the encompassing F test in terms of power especially when the number of regressors in the alternative model is large. Hence, such a procedure may be useful in practical applications.

*I am grateful to Russell Davidson and James G. MacKinnon for helpful suggestions and comments. Helpful comments from A.L. Nagar and T. Stengos are also gratefully acknowledged. All remaining errors are mine.

1 Introduction

Within the class of one degree of freedom non–nested tests the Cox (C) test (Cox (1961, 1962), Pesaran (1974)) and the J test of Davidson and MacKinnon (1981) have been studied under the standard assumption of non–orthogonal regression models. In this case both tests are asymptotically equivalent and distributed as $N(0,\ 1)$ under the null hypothesis.

Neither of these two conclusions, however, is true under model orthogonality. The standard results are no longer valid under the condition of near population orthogonality (NPO) according to which the non–nested regressors in two linear regression models become asymptotically uncorrelated or orthogonal in the population distribution from which they are drawn (Michelis (1994)). Specifically, under NPO the asymptotic null distributions of the J and C tests are expressible as functions of a $N(0,\ 1)$ variate and a χ^2 variate. Furthermore the Monte Carlo simulations indicated that the NPO results predict better than standard results the finite–sample behavior of the J and C tests. In fact the simulation evidence indicated that this is true for correlation values among the non–nested regressors as high as .30.

Given this evidence and given the fact that under NPO the J and C tests do not follow the usual $N(0,\ 1)$ distribution, what can applied econometricians do? That is, when working with non–nested models that have non–nested regressors with low correlations, can an applied econometrician do better in terms of alternative new testing procedures than using the standard non–nested tests or the encompassing F test? In this paper we deal with these questions in the framework of pretest testing theory (see Giles and Giles (1993) for an excellent review of the literature). In general, this theory amounts to, first, using a preliminary test to test a relevant null hypothesis on a certain aspect of a testing problem, and then based on the outcome of the test, to employ alternative testing procedures in order to test the main hypothesis of interest. In the present context we adopt the following testing strategy which is based on a two stage decision problem. In the first stage, the null hypothesis of zero correlations among the non–nested regressors is tested using a preliminary test of significance. In the second stage, if the null hypothesis is rejected, the J test is chosen as a non–nested test; otherwise, either the J_A test of Fisher and McAleer (1981) or the encompassing F test of Mizon and Richard (1986) is employed. The so called pretest test statistics thus generated are random combinations of the two component tests and, in general, have different sampling properties from them. More specifically, we propose and investigate the finite sample properties of the following non–nested pretesting procedures:

(i) Test 1: J combined with J_A

(ii) Test 2: J combined with F.

An investigation of the pretest testing procedures (i) and (ii) may be quite useful and informative. Depending on the outcome of the preliminary test, each procedure uses the potentially more powerful J test or a test with better size in finite samples (i.e., the J_A test or the F test).[1] It is therefore interesting to examine whether (i) and (ii) have better properties in finite samples than do the standard J, J_A and F tests. The Monte Carlo experiments in this paper were designed with this purpose in mind. The results indicate that the procedure (ii) may be potentially useful in practical applications.

The plan of the rest of the paper is as follows. In Section 2 we set out the framework of analysis and explore four testing procedures and choose one, the likelihood ratio test, to be the preliminary test in the context of the testing procedures (i) and (ii). These procedures are then formalized and stated explicitly. Section 3 deals with the design of the Monte Carlo experiments and the presentation of the results of these experiments. Here the main emphasis is on the properties of the non–nested pretest testing procedures (i) and (ii). However, these are also compared to the properties of the "naive" procedures (strategies) of either always ignoring the null of zero correlations and using the J test all the time, or always allowing for it and using either the J_A test or the F test all the time. The various testing procedures will be compared on the basis of three criteria: size, power and power after correcting for size for the non–exact tests (i.e., the J test and the pretest tests). Section 4 contains some concluding remarks.

[1] It is well known that the J and C tests tend to over–reject the null hypothesis in finite samples but they are, in general, more powerful than the J_A and F tests. Under NPO, the over–rejection problem of the J and C tests becomes worse. In fact, the C test becomes explosive in this case (Michelis (1994)). For this reason, we do not incorporate the C test in the present analysis.

2 Preliminary Tests for NPO and Non–nested Pretest Tests

In this section we formalize and present the two non–nested pretesting procedures Test 1 and Test 2 outlined above. First, though, we set out the framework of analysis and explore several testing procedures, each of which may be employed as a preliminary test (call it the NPO test) to test the null hypothesis of zero correlations among the non–nested regressors of two non–nested models. In order to keep the total number of experiments down to a small number, we recommend using only one of them, the likelihood ratio test, based on some plausible arguments.

2.1 The Framework of Analysis

Consider the following two non–nested linear regression models

$$H_0: \quad y = X\beta + u, \quad u \sim \text{iid}(0, \sigma^2 I_n), \quad 0 < \sigma^2 < \infty \tag{2.1}$$

$$H_1: \quad y = Z\gamma + v, \quad v \sim \text{iid}(0, \omega^2 I_n), \quad 0 < \omega^2 < \infty \tag{2.2}$$

where y is an $n \times 1$ vector of observations on the dependent variable, X and Z are the $n \times p$ and $n \times q$ observation matrices of the explanatory variables of models H_0 and H_1 respectively, β and γ are $p \times 1$ and $q \times 1$ vectors of unknown regression coefficients and u and v are $n \times 1$ vectors representing the random errors in the two models. We assume, for convenience, that the non–nested models intersect only at the origin. If this assumption is not true to begin with, then any common regressors can be removed in an obvious way, by projecting the dependent variable and the non–overlapping regressors in each model into the space orthogonal to the intersection subspace, see Michelis (1995).

Given the two linear models in (2.1) and (2.2) the encompassing F test of testing H_0 against H_1 is simply a test for $\gamma = 0$ in the compound model $y = X\beta + Z\gamma + \epsilon$. Next, the J and J_A tests are easily computed from single artificial regressions as t statistics for a nesting parameter. The J test artificial regression may be written as

$$H_J: \quad y = X\beta + \alpha P_z y + \varepsilon \tag{2.3}$$

where $P_z = Z(Z'Z)^{-1}Z'$ and $P_z y = Z\hat{\gamma}$ is the orthogonal projection on the span of Z, representing the $n \times 1$ vector of fitted values from H_1. Then the J

test is simply the t statistic for the hypothesis $\alpha = 0$, i.e.

$$J = \frac{y'P_zM_xy}{\hat{\sigma}_J(y'P_zM_xP_zy)^{1/2}} \tag{2.4}$$

where $\hat{\sigma}_J$ is the OLS estimate of the standard deviation of the error in the J test regression.

Similarly, the J_A test artificial regression may be written as

$$H_{J_A}: \quad y = X\beta + \alpha P_zP_xy + \varepsilon \tag{2.5}$$

where $P_z = Z(Z'Z)^{-1}Z'$, $P_x = X(X'X)^{-1}X'$ and $P_zP_xy = Z\hat{\gamma}_*$ is the $n \times 1$ vector of fitted values obtained by regressing P_xy on the columns of Z. The J_A test is the t statistic for testing $\alpha = 0$ in the artificial regression model given by (2.5). That is,

$$J_A = \frac{y'P_xP_zM_xy}{\hat{\sigma}_{J_A}(y'P_xP_zM_xP_zP_xy)^{1/2}} \tag{2.6}$$

where $\hat{\sigma}_{J_A}$ is the OLS estimate of the standard deviation of the error term in (2.5).

Whether or not non–nested models are orthogonal depends on the form of the asymptotic covariance matrix of the random variables in X and Z. In the standard case of non–orthogonal models it is assumed that

$$\underset{n\to\infty}{\text{plim}}(n^{-1}X'Z) = \Sigma_{xz} \neq 0 \tag{2.7}$$

The condition $\Sigma_{xz} \neq 0$ is to be understood as stating that not every element of the $p \times q$ matrix Σ_{xz} is zero. If that is not the case and the elements of Σ_{xz} are near but not necessarily equal to zero, then the distributional results that we obtain are quite different from those in the existing literature. This case can be stated formally by the NPO condition:

$$\underset{n\to\infty}{\text{plim}}(n^{-1/2}X'Z) = \Delta \tag{2.8}$$

where Δ is a $p \times q$ non–null matrix of constants such that $\Delta'\beta \neq 0$. Notice that (2.8) implies Σ_{xz} tends to the null matrix asymptotically, so that the columns of X and Z become asymptotically uncorrelated or orthogonal as the sample size, n, tends to infinity. Since Δ is $O(1)$ by assumption, asymptotic orthogonality is attained at a rate proportional to $n^{-1/2}$. Alternatively, as the sample size becomes large, the sample matrices X and Z are drawn from a ($p+$

q)–dimensional distribution of which the p components are nearly uncorrelated with or orthogonal to the remaining q components.

The F test has the same distribution under (2.8) as it does under (2.7). The same is true for the J_A test. On the other hand, under (2.8), the J test converges to the random variable

$$\frac{cV_0 + V_0^2 + V_{q-1}}{((V_0 + c)^2 + V_{q-1})^{1/2}} \tag{2.9}$$

where c is a constant that is proportional to β and inversely proportional to σ, V_0 is a $N(0, 1)$ variate and V_{q-1} is a $\chi^2(q-1)$ variate which is independent of V_0 (Michelis (1994), Theorem 4.2). It is precisely because the J test has a discontinuous distribution under (2.7) and (2.8) that the present analysis of non–nested pretest tests is interesting.

2.2 Tests for Preliminary Testing

Here we consider four testing procedures that can be used to test for the possible lack of correlation among the columns of X and Z. For analytic tractability we assume that the rows of X and Z come from a joint $(p+q)$ variate normal distribution with mean $\mu = (\mu_1, \mu_2)$ and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{2.10}$$

where μ_1 and μ_2 are vectors of dimension p and q respectively, Σ_{11} is $p \times p$, Σ_{22} is $q \times q$ and $\Sigma_{12} = \Sigma_{21}'$ is $p \times q$.

We wish to test the null hypothesis that X and Z are uncorrelated against the alternative that they are not, i.e.,

$$H_0: \ \Sigma_{12} = 0 \quad \text{versus} \quad H_A: \ \Sigma_{12} \neq 0 \tag{2.11}$$

It is well known that the population canonical correlations, corresponding to X and Z, are null or non–null depending on whether $\Sigma_{12} = 0$ or $\Sigma_{12} \neq 0$. Therefore, H_0 and H_A in (2.11) can be restated equivalently in terms of the population canonical correlation coefficients ρ_i, $i = 1, \ldots, k = \min(p, q)$, as

$$H_0': \ \rho_1 = \ldots = \rho_k = 0 \quad \text{versus} \quad H_A': \ \text{at least one } \rho_i \neq 0. \tag{2.12}$$

Expressing the null and the alternative in terms of the canonical correlation coefficients is illuminating. The test criteria we shall consider below can all be expressed in terms of the sample canonical correlations between X and Z, and the decision on whether or not to reject the null will depend on the statistical significance of a function of those coefficients.

Let X and Z, as given above, denote the $n \times p$ and $n \times q$ matrices of sample observations and assume $n > p + q$. Define the sample covariance matrix between X and Z by

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{2.13}$$

where the partition of S is similar to that of Σ in (2.10), and let

$$H = S_{12}S_{22}^{-1}S_{21} \tag{2.14}$$

and

$$E \equiv S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}. \tag{2.15}$$

Then the null hypothesis can be tested using Wilks' U–statistic

$$U = \frac{|E|}{|E+H|}, \tag{2.16}$$

where $|E|$ and $|E+H|$ denote the determinants of E and $E+H$ respectively.

It can be easily shown that the likelihood ratio test for the same null hypothesis is a function of the U–statistic. The former test is (see Eaton (1982), Chapter 10)

$$\Lambda = \left(\frac{|S|}{|S_{11}|\,|S_{22}|}\right)^{n/2} \tag{2.17}$$

Using the identity $|S| = |S_{22}|\,|S_{11.2}|$ we have

$$\frac{|S|}{|S_{11}|\,|S_{22}|} = \frac{|S_{11.2}|}{|S_{11}|} = \frac{|E|}{|E+H|}, \tag{2.18}$$

which is exactly (2.16). Consequently, $U = \Lambda^{2/n}$.

Although the exact distribution of U is complicated, in large samples $-n \log U$ is approximately distributed as a $\chi^2(pq)$ variate. A better approximation to the χ^2 distribution is given by the adjusted statistic (Box (1949)),

$$\lambda = -b \log U \tag{2.19}$$

where $b = -n/d$ in which

$$d = 12npq/(2\ell + 3m) \tag{2.20}$$

with $\ell = 2pq$ and $m = (p+q)^3 - (p^3 + q^3)$. For a test of size α we would reject H_0 if $\lambda > \chi^2_\alpha$.

The U–statistic has intuitive appeal when it is expressed in terms of the sample canonical correlations between X and Z. Using (2.18) we obtain

$$U = \frac{|S_{11} - S_{12}S_{22}^{-1}S_{21}|}{|S_{11}|} = |I_p - S_{11}^{-1/2} S_{12}S_{22}^{-1} S_{21} S_{11}^{-1/2}| = \prod_{i=1}^{k}(1 - r_i^2), \quad (2.21)$$

where $r_1^2 \geq r_2^2 \geq \ldots \geq r_k^2$ are the nonzero eigenvalues of $S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$, i.e., the nonzero squared canonical correlations among the columns of X and Z. This form of the test makes it clear that $H_0' : \rho_1 = \ldots = \rho_k = 0$ is unlikely to be rejected when the r_i^2 are small. This is likely to be the case when X and Z are drawn from nearly orthogonal populations.

In addition to the likelihood ratio test, there are three other criteria that can be used to test H_0. These are,

$$T_1 = \mathrm{tr}[H(E+H)^{-1}] = \sum_{j=1}^{k}(\ell_i/(1+\ell_i)) = \sum_{j=1}^{k} r_i^2 \quad (2.22)$$

$$T_2 = \mathrm{tr}(HE^{-1}) = \sum_{j=1}^{k} \ell_i = \sum_{j=1}^{k} \frac{r_i^2}{1 - r_i^2} \quad (2.23)$$

$$T_3 = \mathrm{Max}\{\ell_1, \ldots, \ell_k\} = \ell_1 = \frac{r_1^2}{1 - r_1^2} \quad (2.24)$$

where $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_k$ are the nonzero roots of

$$|H - \ell E| = 0. \quad (2.25)$$

The third set of equalities in (2.22), (2.23) and (2.24) follows from the fact that

$$\ell_i = \frac{r_i^2}{1 - r_i^2}. \quad (2.26)$$

To establish (2.26) just substitute $E = S_{11} - H$ in (2.25) and rearrange.

In large samples nT_1 and nT_2 are approximately distributed as $\chi^2(pq)$ variates while T_3 does not follow any tabulated distribution and requires its own table of percentage points. In the literature on multivariate analysis T_1, T_2, and T_3 are known respectively as the Bartlett–Nanda–Pillai trace test, the Lawley–Hotelling trace test and Roy's maximum root test (e.g., see Anderson (1984), Chapter 8). Interestingly, from the work of Berndt and Savin (1977), we can identify T_1 as a Lagrange multiplier test and T_2 as a Wald test. The same authors have noted that

$$T_1 \leq T_0 \leq T_2 \tag{2.27}$$

where $T_0 = \log U^{-1}$ is the likelihood ratio test. Consequently, in finite samples there may be a conflict among T_1, T_2 and T_3 with respect to deciding on H_0. If the χ^2 significance point is used, T_2 may lead to a rejection while T_0 or T_1 may not. Clearly, the inequalities in (2.27) pose a difficulty in choosing one of the three criteria. On the other hand, Roy's test criterion is exact but requires its own significance points.

As a guide to choosing one of these four tests we would also like to compare their power functions. Unfortunately, none of these tests is uniformly most powerful against all alternatives. Anderson (1984, p. 331) reports a condition, due to Rothenberg (1977), that can be used to rank T_1, T_2 and T_3 in terms of power. This condition is

$$\frac{\sigma_w}{\bar{w}} > \sqrt{\frac{(p-1)(p+q)}{pq+2}} \tag{2.28}$$

where $\bar{w} = (1/p)\sum_{i=1}^{p} w_i$, $\sigma_w^2 = (1/p)\sum_{i=1}^{p}(w_i - \bar{w})^2$ and $w_1, w_2, \ldots, w_p$ are the roots of the noncentrality matrix associated with the three statistics. If (2.28) is satisfied then

$$\text{Power}(T_2) > \text{Power}(T_0) > \text{Power}(T_1), \tag{2.29}$$

and otherwise the inequalities are reversed. As noted above, Roy's largest root test does not have a χ^2 distribution under the null and a non–central χ^2 distribution under the alternative hypothesis. Hence, it cannot be compared to T_0, T_1 and T_2 using Rothenberg's condition. However, Monte Carlo evidence has shown that the maximum root test has greatest power if the alternative is one dimensional (i.e., $\rho_1 \neq 0$, $\rho_2 = \ldots = \rho_k = 0$). Otherwise, it is inferior to other tests.

With no clear–cut answers regarding the power and size properties of these tests, we have chosen the likelihood ratio test as the preliminary test to test for NPO in X and Z. We feel this is a judicious choice for several reasons. First, the likelihood ratio test is intermediate between T_1 and T_2 in terms of size and power and therefore its own behavior reflects the average behavior of the three tests. Furthermore, in general, it is more powerful than Roy's greatest root test. Second, the likelihood ratio test may throw more light on the relationship between X and Z through its various factorizations. This property can be exploited to construct likelihood ratio criteria for the association between any subsets of the variables in X and Z. The resulting criteria are all χ^2 distributed with appropriate degrees of freedom adjustments. Third, from (2.27) we may conjecture that on average the actual size of T_0 must be closer to the true size of the χ^2 approximation. Consequently, the null distribution of T_0 may be closer to that of χ^2 than is the distribution of either T_1 or T_2. Furthermore, the adjusted likelihood ratio statistic (2.19) is an improved approximation to the χ^2 distribution. In fact, it is the λ statistic in (2.19) that we have used in our simulation experiments.

2.3 Non–nested Pretest Tests

Suppose again we wish to test for the empirical adequacy of the two non–nested models (2.1) and (2.2) respectively, while suspecting that the non–overlapping regressors in the two models may be uncorrelated. In view of pretesting theory, this suggests the following two stage testing strategy: first, we use Box's likelihood ratio statistic (2.19) to test the null $H_0 : \Sigma_{12} = 0$, and then if the null is rejected we use the J test or otherwise we use the J_A test in one case and the F in the other. Therefore, we can define two new non–nested pretest testing procedures as follows

$$NP_1 = I_{[0,c]}(\lambda)J_A + I_{(c,\infty)}(\lambda)J \qquad (2.30)$$

and

$$NP_2(\lambda) = I_{[0,c]}(\lambda)F + I_{(c,\infty)}(\lambda)J \qquad (2.31)$$

where $I(\cdot)(\lambda)$ is an indicator function which takes the value of unity if λ falls in the subscripted range and zero otherwise, and c is the critical value of the $\chi^2(pq)$ distribution for a test of size α.

Notice that in contrast to NP_1, the NP_2 procedure combines two test statistics, namely the J and F tests, which have different null distributions. Consequently, when the F test is chosen, the critical values of the relevant F distribution must be employed in order to calculate the empirical size and power of the NP_2 test. Similarly, when the J test is chosen, the critical values of $N(0, 1)$ must be used. Since the actual choice of either component test depends on the value of λ, we write NP_2 as an explicit function of λ.

Representing NP_1 and NP_2 as we have done in (2.30) and (2.31) is illuminating because these expressions highlight the difficulty of deriving exact results. Each statistic is the sum of two parts, both of which are products of two non-independent random variables. For this reason it is difficult to obtain exact analytic expressions for the sampling properties of NP_1 and NP_2. Nonetheless, some observations can be made. Suppose for instance that $H_0 : \Sigma_{12} = 0$ holds. In this case we would like to use the J_A or F test, but $100\alpha\%$ of the time the λ test will incorrectly reject H_0 and the J test will be chosen. Given that the J test over-rejects rather severely in this case NP_1 and NP_2 will also tend to over-reject relative to the J_A and F tests. On the other hand, when H_0 does not hold we would like to use the more powerful J test. However, the proportion of times that the J test is chosen will depend on the power of the λ test (likelihood ratio). The higher the power of the likelihood ratio test the more often the J test will be chosen and the more powerful NP_1 and NP_2 will be. Otherwise, the latter tests may be expected to be less powerful. Clearly, the size α of the preliminary test is important in determining the properties of NP_1 and NP_2.

Since the nature of the problem precludes analytic results, we shall follow the practice of other researchers and investigate the sampling properties of NP_1 and NP_2 through Monte Carlo simulations. These properties will be compared to those of the J, J_A and F tests. We examine the size and power properties of the NP_1, NP_2, J, J_A and F tests in finite samples, so that the Monte Carlo evidence may be potentially useful in practical applications.

3 Monte Carlo Simulations

3.1 Experimental Design

The design of the experiments is similar to that given in Godfrey and Pesaran (1983). The model that generated the data for the size calculations was

$$H_0: \quad y_t = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ti} + u_t, \quad u_t \sim NID(0, \sigma^2), \quad t = 1, \ldots, n \quad (3.1)$$

where $\beta = (\beta_0, \beta_1, ..., \beta_p)$ is a vector of regression coefficients. For each replication the values of the explanatory variables x_{ti} were generated according to $N(0, 1)$ using the pseudo–random number generating subroutine DRNNOF from the IMSL library. The covariance matrix of the explanatory variables was chosen to be the identity matrix, I_p, so that $E(x_{ti}x_{\ell j})$ is unity if $t = \ell$ and $i = j$ and zero otherwise.

The model that generated the data for the power calculations was

$$H_1: \quad y_t = \gamma_0 + \sum_{i=1}^{q} \gamma_i z_{ti} + w_t, \quad w_t \sim NID(0, \omega^2), \quad t = 1, \ldots, n \qquad (3.2)$$

where $\gamma = (\gamma_0, \gamma_1, ..., \gamma_q)$ is a vector of regression coefficients. The DGP (data generation process) for the z_{ti}'s was

$$z_{ti} = \psi_i x_{ti} + v_{ti}, \quad v_{ti} \sim NID(0, 1), \quad i = 1, \ldots, \min(p, q) \qquad (3.3)$$

and, if $q > p$,

$$z_{ti} = v_{ti}, \quad i = p + 1, \ldots, q \qquad (3.4)$$

Notice that the design is such that the simple population correlation coefficients among the non–nested regressors are equal to the population canonical correlation coefficients, i.e.,

$$\phi_i = \rho_i, \quad i = 1, \ldots, \min(p, q) \qquad (3.5)$$

and in all the experiments we set $\rho_i = \rho$, so that the non–nested regressors were equally correlated.

We carried out several experiments to cover various possibilities of interest by changing the principal parameters that should affect the results. These parameters are $(\beta, \gamma, n, \sigma, \omega, p, q, \rho, \alpha)$ where $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ and $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_q)$ are vector valued and the remaining parameters are scalar. Throughout all the experiments the components of β and γ were set equal to unity. The parameters n, σ and ω were set at values 25, 2.0 and 2.0 respectively. The effects of changes in these parameters on size/power estimates were as expected, namely an increase in n or a decrease in σ and ω improves size/power estimates and conversely. For this reason, we do not report the effects of changing n, σ or ω.

The parameters p and q represent the non–overlapping regressors in the two models. The relation between p and q as well as their magnitudes are important in determining the properties of the tests. For instance, when $q > p$ the J_A test is not as powerful as it would be in the opposite case (Godfrey and Pesaran (1983)). Furthermore, under NPO (see (2.9) above), the asymptotic null distribution of the J test depends explicitly on q. The larger the value of q is the greater the discrepancy should be between the s-tandard results and the new results. For these reasons we have considered the following cases for p and q: (a) $p = q = 2$, (b) $p = 4$, $q = 2$, and (c) $p = 2$, $q = 4$.

The most important parameter from the point of view of this paper is ρ, the correlation coefficient among the non–nested regressors. Here we were interested in investigating the size–power properties of the tests for low values of ρ. Thus, ρ was set at the following values: $\rho = (.30, .20, .10, .05, 0)$.

Finally, α is the size of the likelihood ratio test for testing the preliminary hypothesis of zero correlations among the non–nested regressors. As discussed above, the size of a preliminary test plays an important role in determining the properties of pretest test statistics. Accordingly, to see the effects of α on the size and power characteristics of the NP_1 and NP_2 tests, we set α at two values: $\alpha = .05$ and $\alpha = .20$.

The computer programs used for the Monte Carlo experiments were written in Fortran 77. Each experiment was started by calling the DRNNOF subroutine from the IMSL library and then replicating for a specified number of times. We report the results of these experiments next.

3.2 Results of the Monte Carlo Experiments

The simulation results will be presented either in tabular form or by plots. The former includes numerical tables with estimates for the size and power of the J, J_A, NP_1, NP_2 and F tests. The latter consists of plots of the empirical size–power tradeoff curves for the five tests.

For the numerical results and for each of the 5000 replications in each experiment we computed the test statistics J, J_A, NP_1 and NP_2 given by the equations (2.4), (2.6), (2.30) and (2.31) respectively, and the encompassing F statistic for the test for $\gamma = 0$ in the combined regression $y = X\beta + Z\gamma + \epsilon$. We then computed estimates of the size and power of these tests by calculating the proportion of times that each test statistic exceeded the appropriate 5%

and 1% nominal critical values when H_0 was the DGP and when H_1 was the DGP respectively. For the J, J_A and NP_1 tests the critical values 1.96 and 2.576 of the $N(0, 1)$ distribution were used whereas for the F test the 5% and 1% critical values of the $F(q, n-p-q-1)$ distribution were employed given our specific choice of n, p and q.

As mentioned above, the NP_2 test incorporates the J and F tests, and thus requires two different sets of critical values; one set when the J test is used and another when the F test is used. To deal with this problem, we first converted the actual values of the J and F statistics to p-values (i.e., prob-values) and then combined the latter to obtain p-values for the NP_2 test. Each p-value, P_J, for the J test was obtained from the relation $P_J = 2(1 - NCDF(|J|))$ where $|J|$ is the absolute value of J and $NCDF$ is the cdf of $N(0, 1)$. Similarly, each p-value, P_F, for the F test was calculated from the relation $P_F = 1 - FCDF(F)$, where $FCDF$ is the cdf of the F distribution with appropriate degrees of freedom. Letting P denote the p-value for the NP_2 test, we set $P = P_J$ if the λ test rejected the null in (2.11) and $P = P_F$ otherwise. Using this formulation, size and power estimates for the NP_2 test were obtained by calculating the proportion of times, under H_0 and H_1 respectively, that $1 - P$ exceeded the values .95 and .99.

In addition to the information provided by the numerical tables, we have also constructed size–power tradeoff curves. These curves are more informative than the results presented in numerical tables because they allow for the comparison of the five tests' powers on a size corrected basis. The size–power curves are so constructed that for the same estimated size of each test one can compare and rank their estimated powers just by inspecting the graph with the overlaid plots of those curves. By contrast in the numerical tables we, inevitably, compare the powers of the tests of which the sizes are different.

3.2.1 Cases with $p = q = 2$ and $p = 4$, $q = 2$

Table 1 and Table 2 contain respectively the size and power estimates of the J, J_A, NP_1, NP_2 and F tests in the case with $p = q = 2$ for different values of α and ρ. Several interesting features emerge from these experiments. Consider first Table 1. For given α, a fall in ρ from .30 toward .00 leads to a gradual increase in the size of the J test. The J_A and F tests are exact and their size estimates differ from 5% and 1% only because of sampling variations. The NP_1 and NP_2 tests have sizes much closer to nominal size with NP_2 being even better than NP_1 in this respect. Next, when α increases from .05 to .20 the size of NP_1 remains the same but the size of the NP_2 test increases marginally when $\rho = .30$ or $\rho = .20$. The latter finding is easily explainable. For these values of ρ and α, the J test is chosen more frequently as a component of the NP_2 test, and since it over–rejects the null, it causes the NP_2 test to over–reject as well.

Table 2 contains power estimates for the five tests at nominal levels 5% and 1%, and for different combinations of the parameters α and ρ. A simple inspection of the table shows that all tests are more powerful at the 5% level than at the 1% level. The most powerful test of all is the J test, followed by the NP_2, F, NP_1 and J_A tests regardless of the values of the parameters. As ρ falls toward zero, the powers of the J and F tests tend to increase and the powers of the J_A, NP_2 and NP_1 tests tend to decrease.

When α increases from .05 to .20 the power of the tests remains relatively unaffected except for the power of the NP_2 which increases. This reflects the fact that the NP_2 test, in this case, incorporates the relatively powerful J test more frequently.

Tables 3 and 4 contain size and power estimates of the five tests for the case with $p = 4$ and $q = 2$. The pattern of results in Tables 3 and 4 are qualitatively similar to those in Tables 1 and 2 respectively. Differences though do exist in the two pairs of tables. In comparing the results in Tables 1 and 3 a noticeable difference is that the size of the J test is smaller in Table 3 than it is in Table 1. This result should not be surprising. It is well known from the work of Godfrey and Pesaran (1983) that the mean of the J test is directly related to the difference of $q - p$. For the results in Table 1 this difference is zero, whereas for the results in Table 3 the same difference is -2.0, thereby reducing the mean of the J test and hence its estimated size.

Similarly, a comparison of corresponding power estimates in Tables 2 and 4 shows that the power of the tests is, in general, lower when $p = 4$, $q = 2$ than when $p = q = 2$. Regarding the powers of the J, NP_1 and NP_2 tests this is partly due to the fact just mentioned, namely the reduction in the size of the J test when $p = 4$, $q = 2$. Another reason is that the λ test now has 8 degrees of freedom not just 4. Consequently, in view of the Das Gupta and Perlman (1974) result, the likelihood ratio test will not reject as often the null of zero correlations when it is false. This means that the relatively powerful J test will not be chosen as often as it should be. In turn, this translates into a lower power for the NP_1 and NP_2 tests.

An important issue is whether the observed differences in the two sets of tables are statistically significant or just due to the inherent variability of the simulation results. To investigate this issue, consider the size of the J test when $p = q = 2$ and $p = 4$, $q = 2$. If the true size of the J test is the same in both cases then the observed difference in the estimated size of the J test should be statistically insignificant. For example, when $(\alpha, \rho) = (.05, .00)$ the estimated size of the J test at the nominal 5% level is .151 in Table 1 and .123 in Table 3. Under the null hypothesis that the difference is zero the z-score is 4.075 [i.e., $z = (.151 - .123)/((.123 \times .877 + .151 \times .849)/5000)^{1/2}$]. Therefore, the null hypothesis is decisively rejected at all plausible levels of significance. Similarly, for the same choice of parameters the observed power difference, in Tables 2 and 4, is .041 (=.921 - .880), and the corresponding z-score is 6.83. Therefore, the null hypothesis of no difference in the powers of the J test is overwhelmingly rejected at any level of significance.

3.2.2 Cases with $p = 2$, $q = 4$ and $p = q = 4$

Tables 5 and 6 report the simulation results of the size and power estimates of the five tests for the case with $p = 2$, $q = 4$. A cursory inspection of the information in Table 5 indicates immediately that the size of the J test increases dramatically when $q = 4$. This is easily explained in terms of (2.9). Under NPO, the asymptotic null distribution of the J test involves a term which is $\chi^2(q)$ distributed. As a result the larger the value of q the larger will be the magnitude of the test for low values of ρ. This is exactly what the simulation results indicate. As ρ falls toward zero the size of the J test becomes progressively larger, in general, for given values of α. The exact J_A and F tests have the correct size for all parameter values, and the NP_1 and NP_2 tests have size closer to the nominal value in all cases. However, both tests tend to over-reject the null more when α increases from .05 to .20. The over-rejection is due to the fact that, when $\alpha = .20$, the λ test rejects the null hypothesis of zero correlations more often, and thus the NP_1 and NP_2 tests incorporate the J test more often than they do the J_A and F tests respectively.

Turning to Table 6, it is seen that for all parameter values the J test is most powerful again, followed closely by the NP_2 and F tests. On the other hand the NP_1 and J_A tests are markedly inferior to the other three tests in terms of power. Notice that when $q = 4$ rather than $q = 2$, the power of the J_A test is reduced almost by a factor of 2.0; compare the powers of the J_A test in Tables 4 and 6. This result is not surprising. It is well known that the J_A test has poor power when the true model has more non–overlapping regressors than the false model.

As expected, when α is increased from .05 to .20, the power of the NP_1 and NP_2 tests increases. Evidently, this occurs because, in this case, the NP_1 and NP_2 tests incorporates the relatively powerful J test more frequently.

The finding that the NP_2 test is more powerful than the exact F test is important because it points to the possibility that it may be potentially useful in practice in cases where the number of non–nested regressors in empirical models is large.

To obtain a better idea of the properties of the five tests it is useful to consider their power after correcting for size. The size–power tradeoff curves, discussed above, perform automatically size correction, thus allowing a valid comparison of the tests. We turn to the discussion of these curves next.

3.2.3 Power after Size Correction

In this section we discuss the size–power tradeoff curves of the J, J_A, NP_1, NP_2 and F tests. There are basically two ways to correct for size distortions. One is to use the Monte Carlo results under the null to estimate critical values for the non–exact tests (i.e., J, NP_1 and NP_2), and then compute power using those critical values. Unless the critical values are very accurate this approach may not be as reliable as it should be. The second approach to the problem is to use empirical size–power trade–off curves as in Davidson and MacKinnon (1993, Chapter 12). These curves are easy to draw if data from two experiments, one under the null and one under some alternatives, are available. To construct a size–power trade–off curve we just sort each data set, independently, and then plot their ranks (on the $0-1$ interval) for various values of a test statistic. For example, if in the experiment under the null a certain test was greater than 10 2% of the time, and in the experiment under the alternative it was greater than 10 50% of the time, the point (.02, .50) would lie on the size–power tradeoff curve.

We have used the latter method in the present experiments. It is more informative than the first, and it is always valid because it automatically performs size correction. By overlaying the size–power tradeoff curves of the five tests, we can see clearly their power rankings and make immediate power comparisons.

Given the relevant size–power data, the plots were produced on a Zeta plotter using the GPLOT procedure from the SAS/GRAPH software package. In order to produce plots that are relatively smooth we generated data for each experiment by setting the number of replications equal to 15000. Since the NP_2 test involves two different distributions, its size–power tradeoff curve was constructed from the p–values, discussed above, as follows. We first computed the quantities $QS_i = 1 - PS_i$ and $QP_i = 1 - PP_i$, $i = 1, \ldots, 15000$, in which PS_i and PP_i are the p–values for the null and alternative cases respectively. Then we used the QS_i and QP_i in place of the actual test statistics in order to construct the size–power tradeoff curve for the NP_2 test. For example, suppose that $QS_i > .95$ 8% of the time and $QP_i > .95$ 45% of the time. This means that, at the 5% nominal level, (.08, .45) is a point on the size–power curve of the NP_2 test.

Figures 1 and 2 give the size–power tradeoff curves of the five tests when $p = q = 2$ and $p = 2$, $q = 4$ respectively. In each figure the horizontal axis measures size and the vertical axis measures power. Each figure is constructed for $\rho = .10$ and $\alpha = .20$. Varying the values for these parameters produced little additional information. Looking at Figure 1 it is seen clearly that the most powerful test is the J test and the least powerful test is the J_A test. The NP_1 test is more powerful than the J_A test. On the other hand, the NP_2 and F tests are very similar in power with the NP_2 test being marginally more powerful when test size is greater than 15% approximately. Finally, the curves corresponding to the NP_2 and F tests are much closer to that of the J test than to that of either the J_A test or the NP_1 test.

Considering Figure 2 for the case with $p = 2$, $q = 4$, it is clear that the power rankings of the five tests are the same as those in Figure 2 but there is one noticeable difference. The J, NP_2 and F tests become more powerful and the J_A and NP_1 tests less powerful. Thus, the discrepancy in powers between the two sets of tests becomes large, with the curves of the J_A and NP_1 tests moving closer to the 45° line emanating from the origin of the graph, and the curves of the J, NP_2 and F tests moving closer to the (0.0, 1.0) point of the graph. This is as expected, and reflects the fact that when $q - p$ is positive the J test is more powerful and the J_A test less powerful than otherwise.

More importantly, Figure 2 demonstrates clearly the fact that the NP_2 test is never less powerful and, in general, is more powerful than the F test. This is primarily due to the fact that with α set at .20 the J test is chosen frequently as a component of the NP_2 test. Given that the J test is the most powerful test, we obtain a transmission of its power to the NP_2 test and this makes the latter more powerful than the F test.

4 Conclusion

In summary, we can state the following results. The J test is by far the most powerful of the five tests, but its size may be excessively large especially when q is large. The J_A and F tests have the correct size but the J_A test is much less powerful when $q - p$ is positive and large. The NP_1 test behaves similarly to the J_A test, and the NP_2 test has size close to the nominal level and behaves similarly to the F test. Furthermore, when $p = 2$, $q = 4$ and $\alpha = .20$, the NP_2 test is more powerful than the F test. Given that the J test severely over–rejects for low values of ρ and large values for q, and given the fact that the NP_2 test has correct size and is more powerful than the F test under the conditions just stated, the NP_2 test may be potentially useful in practical applications where these conditions are met.

REFERENCES

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed., New York, Wiley.

Berndt, E.R., and N.E. Savin (1977). "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica*, 45, 1263–78.

Box, G.E.P. (1949). "A General Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, 36, 317–46.

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, Series B*, 24, 406–24.

Das Gupta, S. and M.D. Perlman (1974). "Power of the Noncentral F Test: Effect of Additional Variates on Hotelling's T^2–Test," *Journal of American Statistical Association*, 69, 174–80.

Davidson, R., and J.G. MacKinnon (1981). "Several Tests for Model Specification in the Presence of Alternative Hypotheses, *Econometrica*, 49, 781–93.

Davidson, R., and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*, New York, Oxford University Press.

Eaton, M.L. (1983). *Multivariate Statistics: A Vector Space Approach*, New York, John Wiley & Sons.

Fisher, G.R., and M. McAleer (1981). "Alternative Procedures and Associated Tests of Significance for Non–nested Hypotheses," *Journal of Econometrics*, 16, 103–19.

Giles, J.A., and E.A. Giles (1993). "Pre–Test Estimation and Testing in Econometrics: Recent Developments," *Journal of Economic Surveys*, 7, 145–97.

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, 133–54.

IMSL, Inc. (1987). *User's Manual: Stat/Library, Version 1.0*, Houston, USA, IMSL Inc.

Michelis, L. (1994). "The Null Distribution of Non–Nested Tests with Nearly Orthogonal Regression Models," Paper presented at the Canadian Econometric Study Group, Windsor, Ontario, (September 1994).

Michelis, L. (1995). "Orthogonal Regression Models and the Distribution of Non–Nested Tests," *Journal of Quantitative Economics*, forthcoming.

Mizon, G.E., and J.–F. Richard (1986). "The Encompassing Principle and its Application to Testing Non–Nested Hypotheses," *Econometrica*, 54, 657–78.

Pesaran, M.H. (1974). "On the General Problem of Model Selection," *Review of Economic Studies*, 41, 153–71.

Pesaran, M.H. (1982). "Comparison of Local Power of Alternative Tests of Non–Nested Regression Models," *Econometrica*, 50, 1287–1305.

Rothenberg, T. (1977). *Edgeworth Expansions for Multivariate Test Statistics*, IP–255, Centre for Research in Management Science, University of California, Berkeley.

SAS Institute Inc. (1990). *SAS/GRAPH Software*, Volume 1 and Volume 2, Cary, NC, USA, SAS Institute Inc.

Table 1

Size estimates for the J, J_A, NP_1, NP_2 and F tests for the case with $p = q = 2$.

		Estimated Size									
		J		J_A		NP_1		NP_2		F	
α	ρ	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
	.30	0.118	0.037	0.063	0.016	0.074	0.022	0.065	0.016	0.049	0.008
	.20	0.126	0.044	0.066	0.019	0.069	0.021	0.057	0.012	0.050	0.011
.05	.10	0.139	0.048	0.066	0.018	0.067	0.019	0.052	0.010	0.051	0.011
	.05	0.133	0.043	0.059	0.017	0.061	0.018	0.053	0.012	0.044	0.008
	.00	0.151	0.052	0.065	0.019	0.068	0.020	0.054	0.011	0.056	0.013
	.30	0.124	0.041	0.062	0.018	0.089	0.029	0.088	0.026	0.050	0.010
	.20	0.129	0.044	0.064	0.020	0.081	0.025	0.074	0.020	0.053	0.013
.20	.10	0.142	0.048	0.064	0.018	0.079	0.024	0.066	0.019	0.051	0.011
	.05	0.143	0.047	0.062	0.020	0.073	0.025	0.068	0.016	0.047	0.009
	.00	0.143	0.047	0.062	0.017	0.072	0.020	0.065	0.013	0.051	0.009

Table 2

Power estimates of the J, J_A, NP_1, NP_2 and F tests for the case with $p = q = 2$.

		Estimated Power									
		J		J_A		NP_1		NP_2		F	
α	ρ	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
	.30	0.909	0.775	0.563	0.396	0.663	0.510	0.817	0.612	0.749	0.410
	.20	0.910	0.783	0.533	0.369	0.589	0.428	0.794	0.568	0.754	0.506
.05	.10	0.916	0.789	0.529	0.367	0.555	0.395	0.769	0.536	0.760	0.520
	.05	0.917	0.801	0.521	0.363	0.540	0.388	0.768	0.530	0.772	0.516
	.00	0.921	0.797	0.529	0.370	0.547	0.387	0.762	0.524	0.764	0.516
	.30	0.909	0.786	0.580	0.413	0.776	0.634	0.852	0.694	0.759	0.511
	.20	0.915	0.793	0.541	0.383	0.683	0.539	0.824	0.624	0.765	0.522
.20	.10	0.916	0.788	0.532	0.368	0.622	0.465	0.797	0.588	0.755	0.514
	.05	0.912	0.785	0.527	0.365	0.601	0.443	0.790	0.580	0.756	0.505
	.00	0.914	0.786	0.527	0.360	0.608	0.445	0.791	0.572	0.757	0.514

Table 3

Size estimates for the J, J_A, NP_1, NP_2 and F tests for the case with $p = 4$ and $q = 2$.

		Estimated Size									
		J		J_A		NP_1		NP_2		F	
α	ρ	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
	.30	0.103	0.035	0.062	0.018	0.065	0.019	0.063	0.016	0.049	0.011
	.20	0.115	0.036	0.062	0.016	0.066	0.018	0.055	0.012	0.048	0.008
.05	.10	0.112	0.037	059	0.017	0.063	0.018	0.053	0.014	0.050	0.008
	.05	0.120	0.036	0.062	0.018	0.064	0.018	0.051	0.012	0.047	0.012
	.00	0.123	0.048	0.070	0.023	0.071	0.023	0.051	0.011	0.057	0.013
	.30	0.109	0.037	0.066	0.021	0.084	0.028	0.074	0.021	0.049	0.009
	.20	0.114	0.037	0.066	0.016	0.073	0.020	0.065	0.017	0.048	0.008
.20	.10	0.120	0.042	0.067	0.020	0.075	0.025	0.063	0.014	0.051	0.010
	.05	0.116	0.039	0.063	0.019	0.071	0.023	0.061	0.014	0.048	0.009
	.00	0.125	0.042	0.067	0.018	0.074	0.022	0.063	0.016	0.052	0.010

Table 4

Power estimates for the J, J_A, NP_1, NP_2 and F tests for the case with $p = 4$ and $q = 2$.

		Estimated Power									
		J		J_A		NP_1		NP_2		F	
α	ρ	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
	.30	0.883	0.742	0.600	0.433	0.649	0.482	0.754	0.522	0.712	0.446
	.20	0.887	0.744	0.588	0.418	0.614	0.442	0.743	0.509	0.713	0.457
.05	.10	0.889	0.742	0.562	0.399	0.579	0.417	0.722	0.469	0.709	0.459
	.05	0.883	0.739	0.548	0.375	0.564	0.391	0.720	0.476	0.706	0.447
	.00	0.880	0.740	0.558	0.383	0.572	0.396	0.725	0.480	0.709	0.466
	.30	0.886	0.751	0.609	0.436	0.729	0.577	0.811	0.604	0.719	0.454
	.20	0.884	0.747	0.586	0.417	0.681	0.517	0.772	0.547	0.715	0.461
.20	.10	0.886	0.744	0.561	0.383	0.623	0.455	0.767	0.526	0.708	0.450
	.05	0.882	0.740	0.553	0.384	0.604	0.447	0.753	0.515	0.707	0.444
	.00	0.889	0.750	0.567	0.387	0.632	0.455	0.750	0.507	0.716	0.464

Table 5

Size estimates for the J, J_A, NP_1, NP_2 and F tests for the case with $p = 2$ and $q = 4$.

		Estimated Size									
		J		J_A		NP_1		NP_2		F	
α	ρ	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
	.30	0.315	0.141	0.061	0.014	0.104	0.036	0.099	0.033	0.049	0.010
	.20	0.346	0.154	0.071	0.019	0.097	0.028	0.072	0.025	0.053	0.010
.05	.10	0.336	0.154	0.062	0.015	0.075	0.020	0.063	0.015	0.044	0.009
	.05	0.352	0.165	0.062	0.017	0.074	0.023	0.060	0.015	0.053	0.009
	.00	0.344	0.157	0.032	0.018	0.076	0.023	0.060	0.013	0.052	0.012
	.30	0.309	0.134	0.067	0.018	0.142	0.069	0.174	0.068	0.052	5010
	.20	0.326	0.150	0.060	0.016	0.133	0.051	0.134	0.049	0.049	0.007
.20	.10	0.343	0.154	0.062	0.013	0.120	0.043	0.105	0.037	0.047	0.008
	.05	0.355	0.159	0.062	0.018	0.112	0.042	0.100	0.033	0.050	0.010
	.00	0.343	0.162	0.064	0.019	0.110	0.045	0.099	0.034	0.052	0.012

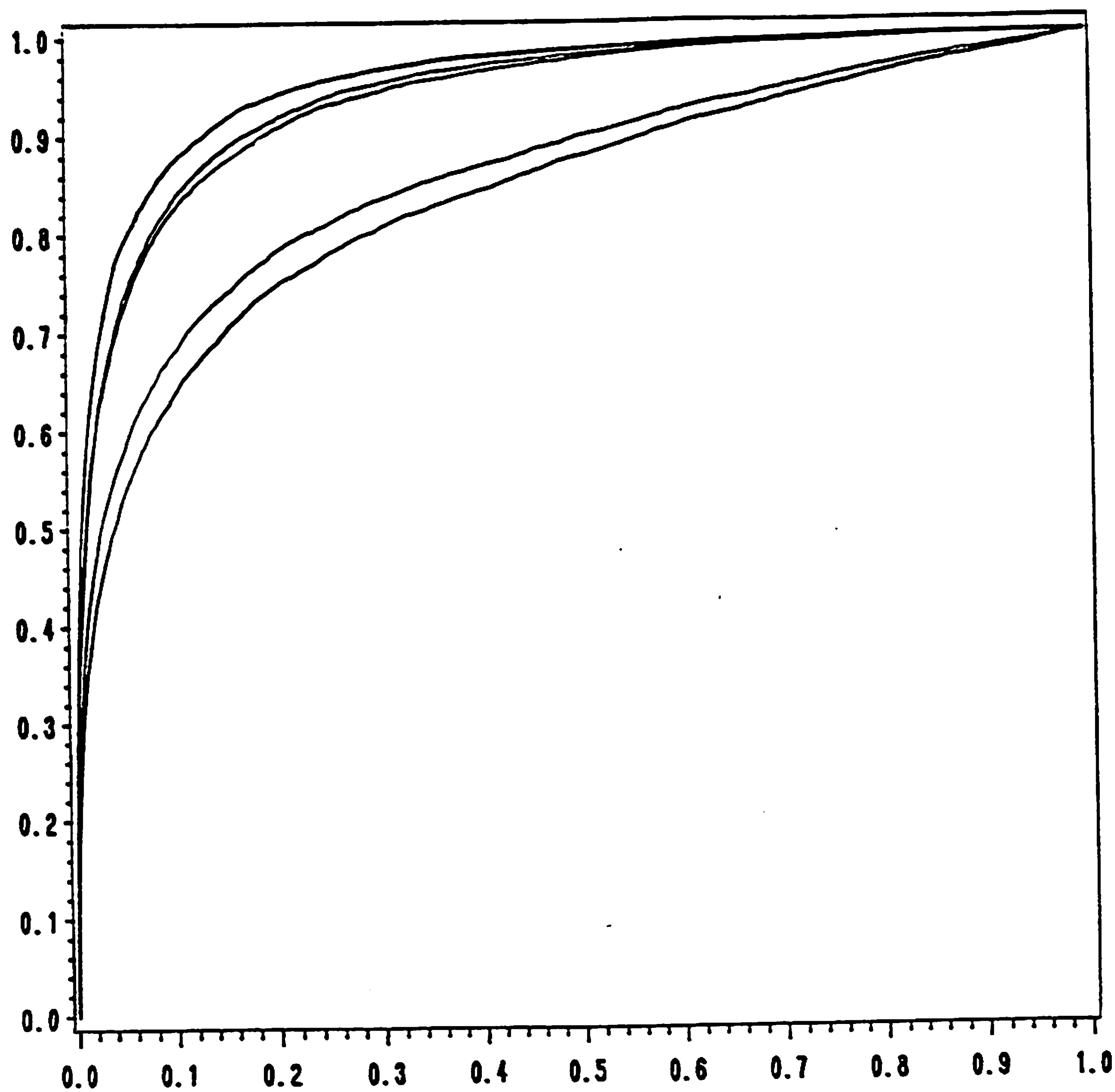
Table 6

Power estimates for the J, J_A, NP_1, NP_2 and F tests for the case with $p = 2$ and $q = 4$.

		Estimated Power									
		J		J_A		NP_1		NP_2		F	
α	ρ	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%
	.30	0.996	0.982	0.344	0.304	0.496	0.385	0.919	0.782	0.888	0.701
	.20	0.996	0.984	0.357	0.213	0.433	0.302	0.904	0.734	0.895	0.710
.05	.10	0.996	0.984	0.345	0.213	0.391	0.265	0.900	0.722	0.893	0.705
	.05	0.996	0.980	0.343	0.212	0.378	0.253	0.899	0.720	0.896	0.711
	.00	0.996	0.984	0.346	0.214	0.380	0.252	0.895	0.721	0.889	0.697
	.30	0.997	0.982	0.339	0.201	0.686	0.611	0.948	0.849	0.889	0.700
	.20	0.997	0.986	0.357	0.221	0.584	0.489	0.929	0.797	0.893	0.712
.20	.10	0.995	0.983	0.354	0.226	0.513	0.408	0.920	0.774	0.897	0.710
	.05	0.996	0.985	0.359	0.215	0.488	0.378	0.913	0.759	0.892	0.706
	.00	0.996	0.982	0.354	0.215	0.489	0.370	0.914	0.761	0.893	0.704

Figure 1

Size–power tradeoff curves of the J, J_A, NP_1, NP_2 and F tests for the case with $\rho = .10$, $p = 4$, $q = 2$ and $\alpha = .20$

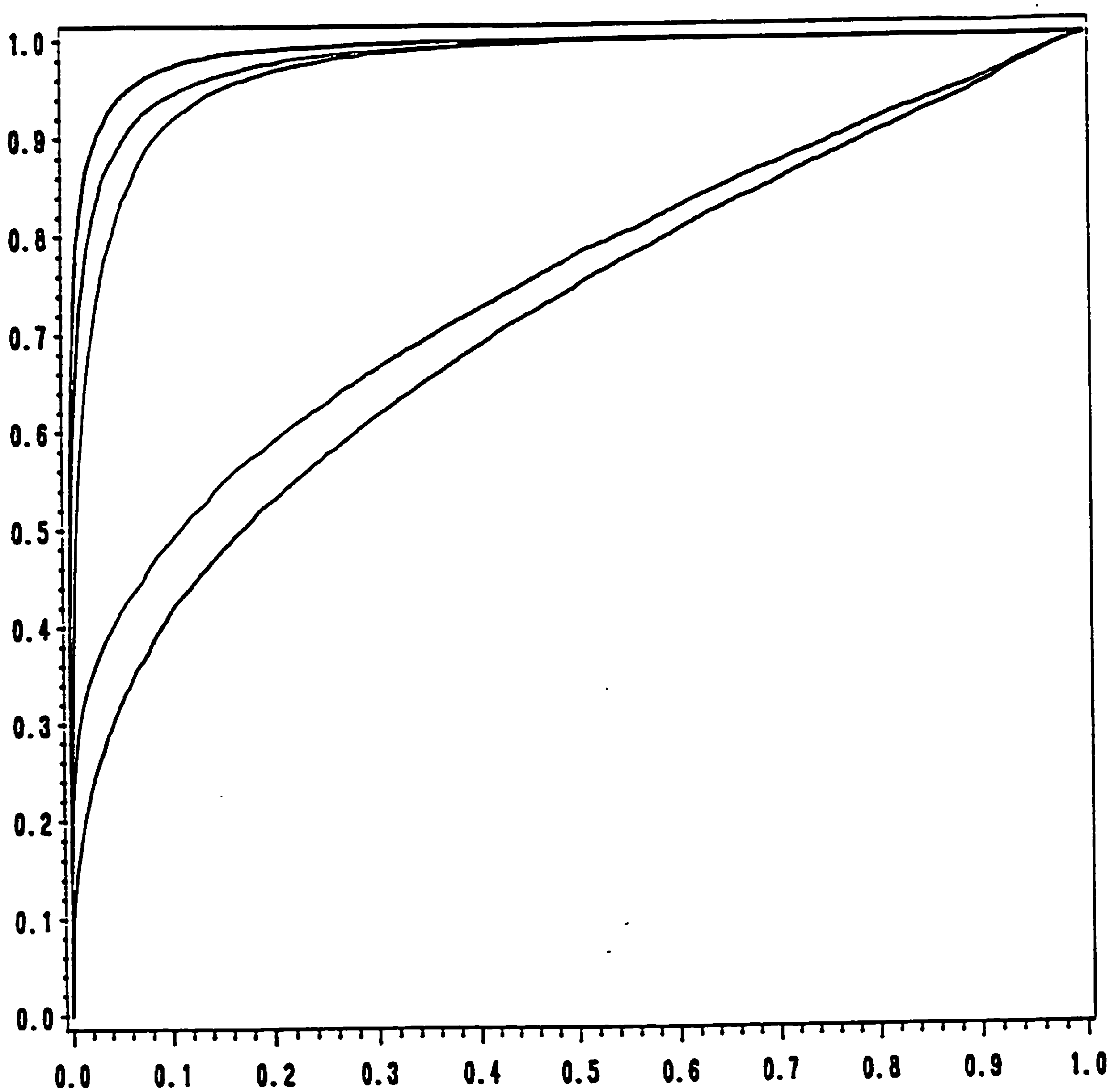


Note: (1) Each experiment is based on 15000 replications, $n = 25$ and $\sigma = \omega = 2.0$.

(2) From left to right: the J, NP_2, F, NP_1 and J_A tests.

Figure 2

Size–power tradeoff curves of the J, J_A, NP_1, NP_2 and F tests for the case with $\rho = .10$, $p = 2$, $q = 4$ and $\alpha = .20$



Note: (1) Each experiment is based on 15000 replications, $n = 25$ and $\sigma = \omega = 2.0$.

(2) From left to right: the J, NP_2, F, NP_1 and J_A tests.