



ELSEVIER

Journal of Econometrics 93 (1999) 369–401

JOURNAL OF
Econometrics

www.elsevier.nl/locate/econbase

The distributions of the J and Cox non-nested tests in regression models with weakly correlated regressors

Leo Michelis*

Department of Economics, Ryerson Polytechnic University, Toronto, Ont., Canada M5B 2K3

Received 1 March 1995; received in revised form 1 April 1999; accepted 19 April 1999

Abstract

This paper examines the asymptotic null distributions of the J and Cox non-nested tests in the framework of two linear regression models with nearly orthogonal non-nested regressors. The analysis is based on the concept of near population orthogonality (NPO), according to which the non-nested regressors in the two models are nearly uncorrelated in the population distribution from which they are drawn. New distributional results emerge under NPO. The J and Cox tests tend to two different random variables asymptotically, each of which is expressible as a function of a nuisance parameter, c , a $N(0, 1)$ variate and a $\chi^2(q)$ variate, where q is the number of non-nested regressors in the alternative model. The Monte Carlo method is used to show the relevance of the new results in finite samples and to compute alternative critical values for the two tests under NPO by plugging consistent estimates of c into the relevant asymptotic expressions. An empirical example illustrates the ‘plug in’ procedure. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C12; C15

Keywords: Non-nested tests; Null distribution; Near population orthogonality; Nuisance parameters; Monte Carlo simulations

* Tel.: + 1-416-979-5092.

E-mail address: michelis@acs.ryerson.ca (L. Michelis)

1. Introduction

Most of the theoretical and empirical research on non-nested hypothesis testing in econometrics has been carried out in the context of two possibly nonlinear non-nested regression models and under the assumption of model non-orthogonality; that is when the sample or population covariance matrix among the non-nested regressors in the two models is a non-null matrix. This restriction of model non-orthogonality introduces two related considerations regarding non-nested tests. First, how valid are existing non-nested tests when the non-nested regressors are suspected of being nearly orthogonal or uncorrelated? At the theoretical level this is as legitimate a question as it would be when one suspects, for instance, heteroskedasticity or autocorrelation in the data. Second, in applied work cases may arise when model non-orthogonality is not a tenable assumption. For instance, in testing Keynesian versus New Classical theories of output or unemployment determination one has to use the OLS residuals from government policy equations as generated regressors in the New Classical empirical equations. Since Keynesian specifications have common regressors with the policy equations this procedure gives rise to rival empirical models with orthogonal regressors; see Pesaran (1982). Given these considerations, it is important to investigate the asymptotic distribution of non-nested tests under the assumption of model orthogonality.

The purpose of this paper is to derive, under a specific condition of model orthogonality, the asymptotic null distributions of two well-known non-nested hypothesis tests in econometrics: the Cox test (henceforth the *C* test) of Cox (1961, 1962) and the *J* test of Davidson and MacKinnon (1981). The framework of analysis involves two linear regression models with nearly orthogonal non-overlapping or non-nested regressors. More specifically, the analysis is based on the concept of near population orthogonality (NPO), according to which the non-nested regressors in the two models are nearly uncorrelated in the population distribution from which they are drawn.

New results emerge under NPO. The *J* and *C* tests tend to two different random variables asymptotically each of which is expressible as a function of a nuisance parameter, denoted c , a $N(0, 1)$ variate and a $\chi^2(q)$ variate, where q is the number of non-nested regressors in the alternative model. Monte Carlo simulations indicate that the theoretical results under NPO predict better than standard results the finite sample behavior of the *J* and *C* tests. Furthermore, a new procedure is proposed for testing nearly orthogonal regression models. This consists of plugging consistent estimates of c into the asymptotic expressions to compute critical values for the two tests under NPO. Here, the Monte Carlo evidence shows that the *J* test performs better than the *C* test.

The new theoretical results are useful in providing a correct understanding of the behavior of non-nested tests under model orthogonality. First, reading the early literature on this issue, one is left with the impression that non-nested tests

are undefined when the non-nested models are orthogonal (e.g., see Pesaran, 1979; Klein, 1983). However, it should be clear from the results in this paper that this is a false impression in general. Second, the J and C tests are no longer asymptotically equivalent under NPO as they are in the standard case of non-orthogonal models. Third, the new results provide additional theoretical insights to the Monte Carlo evidence which shows that the J and C test tend to over-reject more severely when the correlation among the non-nested regressors is weakened (Godfrey and Pesaran, 1983).

The rest of the paper is organized as follows. Section 2 sets out the underlying framework of analysis, which includes the models and the regularity conditions. Section 3 reviews briefly the standard case of non-orthogonal models. Section 4 discusses the asymptotic null distributions of the J and C tests under NPO. Section 5 presents the results of a Monte Carlo experiment designed to investigate the relevance of the new theoretical results in finite samples and to compute critical values for the two tests using the ‘plug in’ procedure. Section 6 illustrates the plug in procedure with an empirical example and the last section concludes the paper. The appendix contains the proofs to the main results of the paper.

2. The framework of analysis

Consider the following two non-nested linear regression models

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

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

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.

Whatever the joint distribution of columns of X and Z may be, we shall assume that the following probability limits exist, are finite and the matrices Σ_{xx} and Σ_{zz} are non-singular:

$$(A1) \quad \text{plim}_{n \rightarrow \infty} (n^{-1} X'X) = \Sigma_{xx},$$

$$(A2) \quad \text{plim}_{n \rightarrow \infty} (n^{-1} Z'Z) = \Sigma_{zz}.$$

In addition to (A1) and (A2) the following assumption has been made in the literature for the standard case of non-orthogonal models.

$$(A3) \quad \text{plim}_{n \rightarrow \infty} (n^{-1} X'Z) = \Sigma_{xz} \neq 0.$$

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.¹ 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.

For the analysis of the non-nested tests under model orthogonality, we shall replace (A3) by the following formalization of the notion of NPO:

$$(A4) \quad \text{plim}_{n \rightarrow \infty} (n^{-1/2} X'Z) = \Delta,$$

where Δ is a $p \times q$ matrix of constants. Notice that condition (A4) implies that

$$\text{plim}_{n \rightarrow \infty} (n^{-1} X'Z) \equiv \Sigma_{xz} = 0, \quad (2.3)$$

where Σ_{xz} is the covariance matrix of the random variables in X and Z and 0 is a $p \times q$ matrix of zeros. It is obvious from (2.3) 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. For this reason, we call (A4) the near population orthogonality condition.

It is illuminating to compare (A4) with the condition used by Staiger and Stock (1997) in their recent work on instrumental variables with weakly correlated instruments. For endogenous variables X and instruments Z , they let $X = Z\pi + w$ where w is a random error and $\pi = \Theta n^{-1/2}$, so that the coefficients on the instruments Z in the first stage regression tend to zero at a rate proportional to $n^{-1/2}$. With standard conditions on (X, Z, w) , this implies the condition

$$(A4') \quad n^{-1/2} X'Z \rightarrow \Sigma_{zz}\Theta + \xi,$$

where ξ is multivariate normal with mean zero and covariance matrix $\sigma_w^2 \Sigma_{zz}$. Under (A4'), the key parameter c in the present paper can no longer be estimated consistently and the problem of constructing alternative critical values for the J and C tests will be magnified by the presence of additional nuisance

¹ In fact, this is not a sufficient condition for the validity of standard asymptotic results. What is required is $\Sigma_{xz}\beta \neq 0$ (see Godfrey and Pesaran, 1983, Appendix A).

parameters. On the other hand, (A4) allows consistent estimation of c and, as will be seen below, does a fairly good job in delivering reasonable critical values especially for the J test.

3. The standard case

In this section we briefly consider the J and C tests and comment on their asymptotic properties in the context of the two non-nested models in (2.1) and (2.2) above and under the assumptions (A1)–(A3). We call this the standard case in order to point out the fact that existing theory assumes that $\Sigma_{xz} \neq 0$.

The J tests is easily computed from an artificial regression as a t statistic for a nesting parameter. Given the two linear models in (2.1) and (2.2), the J test artificial regression may be written as

$$H_J: y = X\beta + \alpha P_z y + \varepsilon, \quad (3.1)$$

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 . Let $\hat{\alpha}$ be the ordinary least squares (OLS) estimate of the nesting parameter α in (3.1). Then the J test is simply the t -statistic for the hypothesis $\alpha = 0$, i.e.

$$J = \frac{y'P_z M_x y}{\hat{\sigma}_J (y'P_z M_x P_z y)^{1/2}}, \quad (3.2)$$

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

Turning to the C test, we adopt a criterion based on Cox's simplified test procedure discussed in Fisher (1983). As long as H_0 and H_1 have distributions that belong to the Koopman–Darmois family this test is asymptotically equivalent to the conventional Cox test and can also be interpreted either as a test based on 'parameters of interest' (Dastoor, 1983) or as a 'variance encompassing test' (Mizon and Richard, 1986). For a test of H_0 against H_1 this statistic may be written as

$$C = \frac{n(\hat{\omega}^2 - \hat{\omega}_*^2)}{2\hat{\sigma}(y'P_x M_z M_x M_z P_x y)^{1/2}}, \quad (3.3)$$

where $\hat{\omega}^2$ is a consistent estimate of the error variance in the alternative model, $\hat{\omega}_*^2$ is a consistent estimate of ω_*^2 , the pseudo-true value of ω^2 under H_0 , and $\hat{\sigma}$ is a consistent estimate of the standard deviation of the error term in the null model. In this paper, $\hat{\omega}^2$, $\hat{\omega}_*^2$ and $\hat{\sigma}$ will be taken to be the maximum likelihood estimators of the corresponding population parameters. (e.g., see McAleer and Pesaran, 1986).

In the standard case, the asymptotic null distribution of each of the two tests is standard normal. Furthermore, the J and minus the C tests tend to the same random variable asymptotically and are thus asymptotically equivalent. Substituting for $y = X\beta + u$ in (3.2) and (3.3) and using the fact that, under H_0 , $\hat{\sigma}_J \xrightarrow{p} \sigma$, and $\hat{\sigma} \xrightarrow{p} \sigma$, it can be easily seen that this random variable is

$$\frac{\beta' X' P_z M_x u}{\sigma(\beta' X' P_z M_x P_z X \beta)^{1/2}} \tag{3.4}$$

4. Near population orthogonality

In this section we consider the asymptotic null distributions of the two tests under near population orthogonality in the sense of the condition (A4) discussed in Section 2. It should be clear from that discussion that NPO is much less restrictive than ESO (exact sample orthogonality) according to which X and Z are orthogonal in a given sample; i.e., $X'Z = 0$. Further, our definition of model orthogonality by the NPO condition (A4) is similar but not directly comparable to the notion of asymptotic (local) orthogonality of Szroeter (1992, p. 560); see also Kent, 1986. In fact, in the present context, it could be argued that Szroeter’s notion of orthogonality corresponds to $\Sigma_{xz} = 0$ as opposed to NPO which means $\Sigma_{xz} = 0$ but $\Delta \neq 0$ exists.

4.1. The J test

Consider the J test for testing H_0 against H_1 given by (3.2). Under H_0 we have

$$y' P_z M_x y = \beta' X' P_z M_x u + u' P_z M_x u, \tag{4.1}$$

$$y' P_z M_x P_z y = \beta' X' P_z M_x P_z X \beta + 2\beta' X' P_z M_x P_z u + u' P_z M_x P_z u \tag{4.2}$$

and

$$\hat{\sigma}_J \xrightarrow{p} \sigma. \tag{4.3}$$

Using the identity $M_x = I_n - P_x$ to expand the right-hand side of each of (4.1) and (4.2) and using (A1), (A2) and (A4) to determine the terms of leading order of magnitude, we obtain the asymptotic relations

$$y' P_z M_x y \stackrel{a}{=} \beta' X' P_z u + u' P_z u + O_p(n^{-1/2}) \tag{4.4}$$

and

$$y' P_z M_x P_z y \stackrel{a}{=} \beta' X' P_z X \beta + 2\beta' X' P_z u + u' P_z u + O_p(n^{-1/2}), \tag{4.5}$$

where $\beta'X'P_zu$, $\beta'X'P_zX\beta$, and $u'P_zu$ are all $O_p(1)$. In view of (4.3), (4.4) and (4.5) it is clear that under H_0 the J statistic (3.2) converges asymptotically to the random variable

$$\frac{\beta'X'P_zu + u'P_zu}{\sigma(\beta'X'P_zX\beta + 2\beta'X'P_zu + u'P_zu)^{1/2}} \tag{4.6}$$

In view of this result, we might expect the asymptotic null distribution of the J test to be different from its own distribution in the standard case and under ESO. However, since it is possible that different random variables may have the same asymptotic distribution, it is necessary to obtain the asymptotic distribution of (4.6) and hence that of the J test. The following theorem provides this and shows the exact form that this distribution takes.

Theorem 4.1. Under NPO the asymptotic null distribution of the J statistic (3.2) is given by the random variable

$$J_{npo} = \frac{cV_0 + V_0^2 + V_{q-1}}{((V_0 + c)^2 + V_{q-1})^{1/2}} \tag{4.7}$$

where $c \equiv D^{1/2}/\sigma$ in which $D = \beta' \Delta \Sigma_{zz}^{-1} \Delta' \beta$, 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 .

Proof. See the Appendix.

Theorem 4.1 makes it clear that the asymptotic null distribution of the J test is a complicated function of a $N(0, 1)$ variate and a $\chi^2(q)$ variate.² Nonetheless, the fact that these two variates appear in (4.7) is appealing because the former is relevant in the standard case whereas the latter results under ESO (Michelis, 1996). Furthermore, the asymptotic expression (4.7) includes ESO and standard results as special cases when $c = 0$ and $c \rightarrow \infty$ respectively. Evidently, NPO irons out the discontinuity in the distribution of the J test which arises under ESO relative to the standard case.

The different results for ESO and NPO may seem counterintuitive since one might expect the same result under NPO as one gets under ESO. Nonetheless, this finding is significant because it highlights the importance of the conceptual distinction between ESO and NPO. In the context of this distinction the two different outcomes make perfect sense. The ESO result (i.e., $J^2 \sim \chi^2(q)$) is obtained under the restriction of zero sample correlations among the non-nested regressors in H_0 and H_1 . On the other hand, NPO allows for non-zero sample

² To see this explicitly, just substitute V_q for $V_0^2 + V_{q-1}$ in (4.7) after expanding the squared term in the denominator.

correlations despite the fact that the columns of X and Z may arise from an underlying parent joint distribution in which the corresponding population correlations approach zero at a rate proportional to $n^{-1/2}$.

Next, consider the special case when $q = 1$. Evidently, in this case the J statistic has the same asymptotic distribution under ESO as it does in the standard case, namely $N(0, 1)$. However, this is no longer true when q is greater than one. Intuitively, in some sense, the problem with the distribution of the J test is to determine its correct degrees of freedom (df) when testing for the significance of the fitted values from the model H_1 . With standard asymptotics $df = 1$ and with ESO $df = q$. Clearly, the problem disappears when $q = 1$.

4.2. The C test

Consider, first, the numerator of (3.3). Using the definitions of $\hat{\omega}^2$ and $\hat{\omega}_*^2$, and upon further manipulation the numerator can be written as

$$n(\hat{\omega}^2 - \hat{\omega}_*^2) = y'P_xP_zP_{xy} - y'P_{zy}. \tag{4.8}$$

Under H_0 , the right-hand side of (4.8) reduces to

$$2\beta'X'P_zP_xu + u'P_xP_zP_xu - (2\beta'X'P_zu + u'P_zu). \tag{4.9}$$

Therefore, using (A1), (A2) and (A4) to determine orders of magnitude and retaining only the $O_p(1)$ terms, we have the following result asymptotically:

$$y'P_xP_zP_{xy} - y'P_{zy} \stackrel{a}{=} - (2\beta'X'P_zu + u'P_zu) + O_p(n^{-1/2}). \tag{4.10}$$

Next consider the denominator of (3.3). Under H_0 the bracketed expression becomes

$$y'P_xM_zM_xM_zP_{xy} = \beta'X'M_zM_xM_zX\beta + 2\beta'X'M_zM_xM_zP_xu + u'P_xM_zM_xM_zP_xu. \tag{4.11}$$

Simplifying each term on the right-hand side of (4.11) by using the facts that $X'M_x = 0$ and $Z'M_z = 0$, we can write

$$y'P_xM_zM_xM_zP_{xy} \stackrel{a}{=} \beta'X'P_zX\beta + O_p(n^{-1/2}) \tag{4.12}$$

in which $\beta'X'P_zX\beta$ is $O_p(1)$ as before.

Putting together (4.10) and (4.12) and using the fact that under $H_0 \hat{\sigma} \xrightarrow{p} \sigma$, it is easily seen that minus the C statistic (3.3) converges to the following random variable asymptotically,

$$\frac{2\beta'X'P_zu + u'P_zu}{2\sigma(\beta'X'P_zX\beta)^{1/2}}. \tag{4.13}$$

Given (4.13) we state and prove the following theorem regarding the asymptotic null distribution of the C test.

Theorem 4.2. Under NPO with Δ non-null and $\Delta'\beta \neq 0$, the asymptotic null distribution of the C statistic (3.3) is given by the random variable

$$C_{\text{npo}} \equiv -\frac{2cV_0 + V_q}{2c}, \quad (4.14)$$

where c , V_0 and V_q are the same quantities as those defined in Theorem 4.1.

Proof. See the Appendix.

In Theorem 4.2 the condition Δ being non-null and $\Delta'\beta \neq 0$ is needed to rule out the possibility of dividing the numerator of C_{npo} by zero. It is clear from (4.14) that the complete (standardized) Cox test is well defined under our notion of model orthogonality in the sense of NPO. This result which is true under NPO cannot be obtained under the existing notions of model orthogonality (i.e., ESO or $\Sigma_{xz} = 0$).

Yet, another important result emerges from the present analysis. In the standard case the J and C tests are asymptotically equivalent not only in the sense of having the same null distribution, but also because they tend to the same random variable asymptotically. By contrast, under NPO this equivalence is no longer true. To see this, notice first that a comparison of (4.6) and (4.13) shows that the two tests are different random variables asymptotically. Furthermore, Theorems 4.1 and 4.2 establish that the two tests have different asymptotic null distributions as well. Consequently, studying these tests independently is a useful and informative exercise in its own right.

Also, the results of the present paper may be viewed as providing theoretical explanation of Monte Carlo evidence that indicates that the J and C tests tend to over-reject the null more severely as the correlation among the non-nested regressors decreases. Theorems 4.1 and 4.2 show that the J and C tests involve not only a $N(0, 1)$ variate but also a $\chi^2(q)$ variate. With c positive, the $\chi^2(q)$ variate will shift the distribution of the J test to the right of $N(0, 1)$ on average, and the distribution the C test to the left of $N(0, 1)$ because of the minus sign in (4.14);³ see Figs. 1 and 2 below. Since non-nested hypothesis testing entails two-sided tests and the simulation evidence is based on nominal values from $N(0, 1)$, each test will tend to over-reject the null more often under NPO. The right shift of the J test will cause it to over-reject the null more often in the right tail, and the left shift of the C test will cause it to over-reject the null more often in the left tail.

³ Note that the mean of C_{npo} is $-q/2c$ which is negative, and the mean of J_{npo} is positive with the mean of its numerator being equal to q .

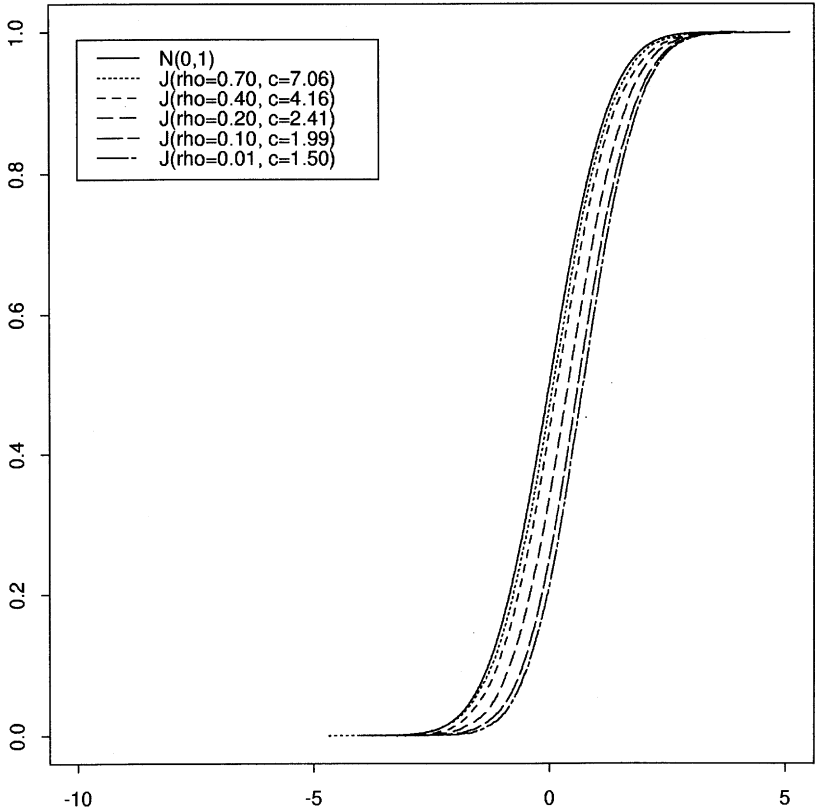


Fig. 1. Five edfs of the J test plotted against the cdf of $N(0, 1)$ for $q = 2$. The edfs are based on 20,000 replications with $n = 100$.

Finally there is an useful interpretation of the parameter c in (4.7) and (4.14) which distinguishes ESO, NPO and standard asymptotics. In particular, under the null, c^2 is the population analog of the chi-squared statistic for testing the significance of Z in the ‘first stage’ regression used to obtain the fitted values from H_1 . It is worth noting that c^2 is not the actual chi-squared statistic testing the significance of Z in the model H_1 , but rather something which is equivalent to it under the null hypothesis. Under NPO one should expect c^2 to be small but non-zero. Thus, the sample version of this statistic, $\hat{c}^2 = (\hat{\beta}X'P_zX\hat{\beta})^2/\hat{\sigma}^2$, should be quite useful in practice in detecting near orthogonality. In practice, of course, any common regressors in empirical non-nested models must be removed suitably before computing the \hat{c}^2 statistic. This can be accomplished by projecting the regressand and the non-nested regressors in each model off the intersection subspace in the two models.

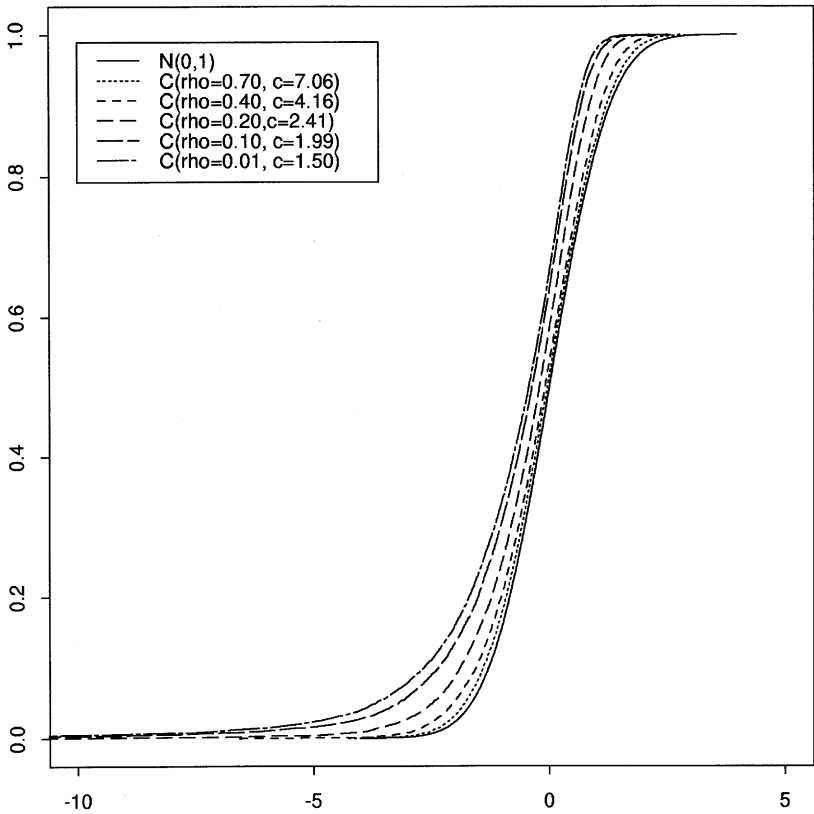


Fig. 2. Five edfs of the C test plotted against the cdf of $N(0, 1)$ for $q = 2$. The edfs are based on 20,000 replications with $n = 100$.

In Staiger and Stock (1997) the Wald statistic plays a similar role in detecting weak instruments. However, their specific asymptotics do not allow consistent estimation of the Wald statistic. Here, in contrast, \hat{c} is a consistent estimator of c and, in addition, the values of \hat{c} can be used in (4.7) and (4.14) to compute alternative critical values for testing non-nested models with weakly correlated non-nested regressors. We exploited this fact to obtain the Monte Carlo results reported in Section 5.2.3 below.

5. Monte Carlo simulations

The purpose of the Monte Carlo simulations is to study the finite sample behavior of the J and C tests primarily under NPO. Three related questions of

practical significance are analyzed. First, how do the J and C tests perform in finite samples and how important is NPO asymptotically? Second, do the asymptotic formulae (4.7) and (4.14) yield good approximations to the finite sample distributions of the two tests when the non-nested regressors are weakly correlated? Third, what is the size of each test under NPO, if one computes critical values from (4.7) and (4.14) after substituting \hat{c} for c in the asymptotic expressions? Three sets of Monte Carlo experiments were performed to investigate these questions. The simulation results will be analyzed following the description of the experimental design.

5.1. Design of the experiments

The design of the experiments is similar to that given in Godfrey and Pesaran (1983). The ‘true’ model that generated the n observations on the dependent variable y was given by

$$H_0: y_t = \beta_0 + \sum_{i=1}^p \beta_i x_{ti} + u_t, \quad t = 1, \dots, n, \quad (5.1)$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ is the vector of the true regression coefficients. For each replication the values of the explanatory variables x_{ti} were generated according to IIN(0, 1) using the pseudo-random number generating subroutine DRNNOF from the IMSL library.⁴ 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 values of the error term, u_t , were also generated by the DRNNOF subroutine according to IIN(0, σ^2) where σ^2 is the error variance in the model given in H_0 . Furthermore, the values of u_t were chosen independently of the values of x_{ti} so that $E(x_{ti}u_t) = 0$ for all t and i .

The alternative ‘false’ model was also linear and given by

$$H_1: y_t = \gamma_0 + \sum_{i=1}^q \gamma_i z_{ti} + w_t, \quad t = 1, \dots, n, \quad (5.2)$$

where $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_p)$ is the vector of regression coefficients and the error term w_t was generated, similarly to u_t , according to $N(0, \omega^2)$ where ω^2 is the error variance in the alternative model. All the values of γ and ω^2 were fixed at unity for all the experiments that were carried out.

⁴ DRNNOF is a double precision subroutine that generates pseudo-random numbers from $N(0, 1)$ using the transformation method (e.g., see Davidson and MacKinnon (1993)).

The DGP for each explanatory variable in H_1 was given by

$$z_{ti} = \psi_i x_{ti} + v_{ti}, \quad i = 1, \dots, \min(p, q) \quad (5.3)$$

and, if $q > p$,

$$z_{ti} = v_{ti}, \quad i = p + 1, \dots, q, \quad (5.4)$$

with $v_{ti} \sim \text{IIN}(0, 1)$ for $t = 1, \dots, n$. The values of the ψ_i were set according to the relation

$$\psi_i = \rho_i / (1 - \rho_i^2)^{1/2}, \quad i = 1, \dots, \min(p, q), \quad (5.5)$$

where ρ_i is the simple population correlation coefficient between x_{ti} and z_{ti} .

We carried out several experiments to cover various possibilities of interest in the context of the issues raised earlier. The different experiments that were carried out involved the parameters $(p, q, \beta, n, \sigma, \rho_i, c)$ which were used to generate the data.

The parameters p and q represent the number of non-overlapping regressors in H_0 and H_1 , respectively. The theoretical results in Section 4 depend explicitly on q through the V_q term in Theorems 4.1 and 4.2. Therefore, it was important to consider different values for q . Accordingly, we set $p = 2$ and we investigated the cases: $q = 2$, $q = 4$ and $q = 6$. As mentioned earlier, it is important to set q greater than one; otherwise the J^2 statistic is numerically equal to the $F(1, n - p - q)$ statistic of which the distribution is known exactly and is independent of the parameters ρ_i . Also, the asymptotic null distributions of the J and C tests depend explicitly on β , the regression coefficient vector in H_0 , through the $c \equiv D^{1/2}/\sigma$ term in the Theorems 4.1 and 4.2. Since what matters is the value of c , we set β and σ equal to unity throughout all the experiments and changed the value of the key parameter c by changing the values of q and ρ_i .

Of the remaining parameters, the sample size, n , was set at two values: some experiments were performed with $n = 50$ and some other with $n = 100$. The value $n = 50$ is moderately small and is intended to capture the small sample behavior of the J and C statistics in the various cases we investigated. If asymptotic theory is ever going to be a poor guide in finite samples it should be in this case rather than when $n = 100$. Alternatively, when $n = 100$ we should expect the experimental results to improve and be in closer agreement with those provided by asymptotic theory.

The crucial parameters ρ_i were set, via the ψ_i in (5.5), at a range of high and low values. Assuming that $\rho_i = \rho$, typical values of ρ were $\rho = (0.70, 0.40, 0.20, 0.10, 0.01)$. We also carried out an extra set of experiments where we allowed for different ρ_i . As long as all the ρ_i are high (i.e., 0.40 or above) or low (i.e., 0.30 or below), the experimental results were virtually the same as if $\rho_i = \rho$. For this reason we do not report the results in this case.

The computer programs used in the Monte Carlo experiments were written in Fortran 77 and were executed on an IBM R/S 6000 workstation. Each

experiment was started by calling the DRNNOF subroutine from the IMSL library and then replicating either 5000 times for the numerical results or 20,000 times for the graphical results (see below).

5.2. Simulation results

The simulation results are displayed either in tabular form or in graphical form. The former includes numerical tables with rejection frequencies, means and variances of the two test statistics. The latter consists of plots of the empirical distribution functions (EDF's) of the J and C statistics against either the cumulative distribution function (CDF) of $N(0, 1)$ or the EDFs of (4.7) and (4.14) once with c known and another time with c estimated.

The plots were obtained as follows. To construct the EDFs we took the experimental data for the relevant test statistics and sorted them in increasing order of magnitude. From the sorted data we then computed cumulative relative frequencies by choosing certain points over the range of the data. A total of 400 points were chosen. To construct the CDF of the standard normal distribution, we generated 400 numbers from the interval -10.0 to $+5.0$ and then used the IMSL subroutine DNORDF to compute the CDF of $N(0, 1)$. Once the 400 data points of sample cumulative relative frequencies and values of the standard normal CDF were available, the EDFs and the CDF were plotted using the *motif* graphing device of the software package S-PLUS (Statistical Sciences, S-PLUS, 1993).

5.2.1. Comparison to standard results

The experiments in the first set were intended to assess the performance of the J and C tests and the NPO results in relation to standard asymptotics. The objectives in these comparisons were to examine how the J and C tests perform with near orthogonality and to see how important is near orthogonality asymptotically. To accomplish these objectives we computed, for each of the 5000 replications, the J , C , J_{npo} and C_{npo} statistics given by (3.2), (3.3), (4.7) and (4.14) respectively, and then we obtained estimates of Type I error (i.e., estimated size) by calculating the proportion of times that each test rejected the null hypothesis at the 10%, 5% and 1% nominal levels of significance (i.e., nominal sizes). This meant calculating the proportion of times that the test statistics were greater than 1.645, 1.96 and 2.576 in absolute value (i.e., the 10%, 5% and 1% critical values for the standard normal distribution). Besides calculating rejection frequencies we also computed the sample means and variances for each test statistic using the data from the total number of replications.

The numerical results for the J and C tests are contained in Tables 1 and 2 respectively. Several features emerge from these experiments. Consider first the results in Table 1. It is clear that, in general, the estimated size and mean of the J test increase, and its variance decreases as ρ and c take on smaller values.

Table 1
Size, mean and variance of the J test

Z-reg. in H_1 q	c-par. c	rho ρ	Estimated size			Sample mean \bar{J}	Sample variance $V(J)$
			10%	5%	1%		
2	7.06	0.70	0.1066	0.0550	0.0108	0.102	1.020
	4.16	0.40	0.1151	0.0586	0.0168	0.301	1.001
	2.41	0.20	0.1462	0.0800	0.0216	0.550	0.980
	1.99	0.10	0.1536	0.0898	0.0260	0.680	0.904
	1.50	0.01	0.1696	0.0996	0.0270	0.748	0.875
4	8.21	0.70	0.1570	0.0894	0.0270	0.495	1.106
	5.66	0.40	0.1618	0.0934	0.0262	0.618	1.008
	3.70	0.20	0.2282	0.1412	0.0444	0.916	0.960
	2.97	0.10	0.2598	0.1738	0.0588	1.054	0.951
	2.66	0.01	0.2774	0.1752	0.0606	1.111	0.902
6	8.28	0.70	0.2412	0.1548	0.0508	0.890	1.080
	5.88	0.40	0.2442	0.1602	0.0556	0.942	1.069
	4.17	0.20	0.3480	0.2380	0.0936	1.250	1.021
	3.48	0.10	0.3838	0.2648	0.1096	1.371	0.978
	3.33	0.01	0.4068	0.2828	0.1158	1.432	0.933

Note: Each experiment is based on 5000 replications and $n = 50$.

The size distortions become excessively large as q increases successively from the value of 2 to 4 to 6.

Turning to Table 2, it is seen that as ρ and c become smaller the size, mean (in absolute value) and the variance of the C test all increase. When $q = 2$, the increase in size is much greater than is the case for the J test, and this is especially true for values of ρ at 0.20 or below. Notice also the large increase in the variance of the C test for low values of ρ and c . This stands in contrast to the behavior of the J test, the variance of which falls under similar conditions. Moreover, as q increases the size of the C test increases uniformly. However, when $q = 4$ and $q = 6$ the size distortions are smaller for the C test than for the J test.⁵

It is instructive to compare the behavior of the two tests for high and low values of ρ and c . Notice that when $\rho = 0.70$ and $c = 7.06$, the two tests behave

⁵ This result is not global and depends on the choice of the parameters in the experiments. For instance, when the c parameter is made smaller by setting $\sigma = 3$, the size of the C test becomes substantially larger than the size of the J test at every value of q . These results are available from the author upon request.

Table 2
Size, mean and variance of the C test

Z-reg. in H_1 q	c-par. c	rho ρ	Estimated size			Sample mean C	Sample variance $V(C)$
			10%	5%	1%		
2	7.06	0.70	0.1084	0.0536	0.0124	-0.091	1.030
	4.16	0.40	0.1264	0.0770	0.0340	-0.281	1.262
	2.41	0.20	0.1668	0.1202	0.0652	-0.560	1.990
	1.99	0.10	0.1998	0.1558	0.0998	-0.780	3.304
	1.50	0.01	0.2108	0.1678	0.1074	-0.881	3.803
4	8.21	0.70	0.1342	0.0780	0.0218	-0.340	1.116
	5.66	0.40	0.1632	0.1092	0.0446	-0.513	1.243
	3.70	0.20	0.1974	0.1460	0.0766	-0.745	1.470
	2.97	0.10	0.2226	0.1674	0.0928	-0.844	1.631
	2.66	0.01	0.2444	0.1830	0.1066	-0.920	1.832
6	8.28	0.70	0.1738	0.1048	0.0342	-0.540	1.150
	5.88	0.40	0.1946	0.1304	0.0602	-0.722	1.219
	4.17	0.20	0.2534	0.1844	0.0964	-0.920	1.481
	3.48	0.10	0.2790	0.2062	0.1132	-1.031	1.518
	3.33	0.01	0.2788	0.2062	0.1050	-1.032	1.533

Note: Each experiment is based on 5000 replications and $n = 50$.

similarly and according to the standard theory. Both estimated sizes, mean and variances are close to the nominal levels of the $N(0, 1)$ distribution. But notice that when ρ and c fall to 0.20 and 2.41 respectively the size of both tests is excessively large.

A related issue is whether or not the observed differences in the simulation results are due to experimental error. To investigate this, consider the case $(c, \rho) = (2.41, 0.20)$ where the estimated sizes of the J and C tests are 0.1462 and 0.1668 respectively. Let $\hat{\pi}$, π and N denote estimated size, nominal size and number of replications respectively, and define the z -score by $z = (\hat{\pi} - \pi) / (\pi(1 - \pi)/N)^{1/2}$. Then under the null hypothesis that $\pi = 0.10$, the z -scores are 10.89 and 15.74 respectively. Clearly, the null hypothesis is decisively rejected at practically all significance levels. As another example, when $(c, \rho) = (7.06, 0.70)$ and the nominal size is 5% the estimated sizes of the J and C tests are 0.0550 and 0.0536 respectively. Under the null that the size of each test is 5% the two z -scores are 1.62 and 1.17 respectively. In this case the null hypothesis for either test cannot be rejected at the 5% level of significance.

In order to gain more insight on the dependence of the distributions of the J and C tests on ρ (or equivalently c), we have plotted the CDF of $N(0, 1)$ against

the EDFs of the two tests for different values of ρ . Figs. 1 and 2 contain the relevant results for the J and C tests respectively. In both figures the results were obtained for the case with $q = 2$ and $n = 100$. Reading from the left (right), the first curve is the CDF of $N(0, 1)$ followed by the EDFs of the J (C) test for ρ set at values 0.70, 0.40, 0.20, 0.10 and 0.01 respectively. The corresponding values of the key parameter c are also reported on the graphs.

It is clear from Fig. 1 that the J test is not invariant with respect to ρ . As ρ decreases gradually from 0.70 to 0.01 the EDFs of the J test drift further to the right from the CDF of the standard normal distribution. Notice in particular that even when ρ drops from 0.70 to the relatively high value of 0.40 the distribution of the J test shifts; compare the two EDFs in Fig. 1 starting from the left. Moreover, the shift to the right is greater when ρ drops from 0.40 to 0.10 than when it drops from 0.70 to 0.40. This finding is explainable in terms of the theory of Section 4 since at low values of ρ the J test does not follow the $N(0, 1)$ distribution.

A different picture emerges from Fig. 2. As ρ falls toward zero, the EDFs of the C test drift to the left rather than to the right as is the case for the J test under the same conditions. This is understandable, since the C test has a negative mean and the J test a positive mean. Notice also the shape differences in the distribution functions of the J and C tests. The C test has a longer lower tail than the J test and the tail becomes thicker as ρ decreases. Thus, given that rejection frequencies are based on two-sided nominal critical values from $N(0, 1)$, the C test tends to over-reject more in the lower tail than in the upper tail compared to the J test.

Tables 3 and 4 report the simulation results for the J_{npo} and C_{npo} statistics. In what follows, when referring to nearly orthogonal samples it will be taken to be samples arising from values of $\rho = 0.20$ or less. Of course, in practice, near orthogonality can be detected by means of a χ^2 test using the observed value of $\hat{\zeta}^2$.

Table 3 shows that the J_{npo} statistic mimics closely the behavior of the J test with nearly orthogonal samples. For every value of c , ρ and nominal size, the size of J_{npo} is less than the size of the J test in Table 1. This result is expected, since the results in Table 1 measure the total size distortion of the J test whereas the results in Table 3 can be viewed as measuring the proportion of the test's size distortion accounted for by the near orthogonality. Viewed this way, it is evident from comparing the Tables 1 and 3 that, for nearly orthogonal samples, NPO explains over 95% of the actual size distortion of the J test. This is rather compelling evidence for the relevance of the NPO results in explaining the actual behavior of the J test in nearly orthogonal samples.

Table 4 documents the simulation results for the C_{npo} statistic. Unlike J_{npo} of which the size is uniformly lower than that of the J test, the size of the C_{npo} statistic shows more variation in nearly orthogonal samples. For all values of q , its size is uniformly above or below the size of the C test as the nominal size

Table 3
Size, mean and variance of the J_{npo} statistic

Z-reg. in H_1	c-par. c	rho ρ	Estimated size			Sample mean \bar{J}_{npo}	Sample variance $V(J_{\text{npo}})$
			10%	5%	1%		
2	7.06	0.70	0.0988	0.0520	0.0092	0.162	0.970
	4.16	0.40	0.1058	0.0550	0.0098	0.301	0.951
	2.41	0.20	0.1332	0.0750	0.0192	0.530	0.920
	1.99	0.10	0.1484	0.0846	0.0192	0.670	0.844
	1.50	0.01	0.1610	0.0874	0.0210	0.728	0.825
4	8.21	0.70	0.1104	0.0524	0.0110	0.355	0.950
	5.66	0.40	0.1360	0.0714	0.0160	0.568	0.888
	3.70	0.20	0.2052	0.1226	0.0348	0.851	0.890
	2.97	0.10	0.2472	0.1448	0.0420	1.034	0.832
	2.66	0.01	0.2798	0.1734	0.0522	1.091	0.831
6	8.28	0.70	0.1562	0.0854	0.0246	0.620	0.930
	5.88	0.40	0.1874	0.1074	0.0292	0.812	0.879
	4.17	0.20	0.2944	0.1898	0.0616	1.140	0.861
	3.48	0.10	0.3476	0.2428	0.0886	1.301	0.868
	3.33	0.01	0.3702	0.2470	0.0870	1.352	0.803

Note: Each experiment is based on 5000 replications and $n = 50$.

is 10% or 1% respectively. At the intermediate 5% level, the size of C_{npo} is either above or below the size of the C test reported in Table 2. This type of behavior of the C_{npo} statistic can be explained by its increasing variability, at every value of q , as the samples become successively more orthogonal (see the last column of Table 4).

Judging from the numerical and graphical results so far, it is clear that the asymptotic equivalence of the two tests breaks down for nearly orthogonal regression models. In this case, the distribution of the C test behaves differently, and is more volatile than that of the J test. Overall, it seems that the J test tends to behave better than the C test.

5.2.2. Comparison of EDF quantiles

This section reports the simulation results of the second set of experiments designed to examine how well the asymptotic distributions under NPO approximate the finite sample distributions. This was done by comparing selected quantile estimates from the EDFs of the J and J_{npo} statistics and the EDFs of the C and C_{npo} statistics. Tables 5 and 6 report quantiles at the 50%, 75%, 90%, 95%, 97.5% and 99% levels for given values of q , c and ρ . To give the asymptotic

Table 4
Size, mean and variance of the C_{npo} statistic

Z-reg. in H_1 q	c-par. c	ρ ρ	Estimated size			Sample mean C_{npo}^-	Sample variance $V(C_{\text{npo}})$
			10%	5%	1%		
2	7.06	0.70	0.1056	0.0510	0.0088	– 0.131	0.990
	4.16	0.40	0.1238	0.0708	0.0212	– 0.261	1.112
	2.41	0.20	0.1814	0.1164	0.0518	– 0.550	1.910
	1.99	0.10	0.2290	0.1650	0.0878	– 0.810	3.694
	1.50	0.01	0.2400	0.1724	0.1002	– 0.901	5.243
4	8.21	0.70	0.1182	0.0644	0.0146	– 0.260	1.056
	5.66	0.40	0.1560	0.0984	0.0354	– 0.493	1.163
	3.70	0.20	0.2116	0.1398	0.0596	– 0.695	1.460
	2.97	0.10	0.2504	0.1724	0.0784	– 0.834	1.651
	2.66	0.01	0.2562	0.1836	0.0896	– 0.900	1.842
6	8.28	0.70	0.1354	0.0728	0.0218	– 0.410	1.060
	5.88	0.40	0.1978	0.1250	0.0412	– 0.661	1.193
	4.17	0.20	0.2480	0.1706	0.0718	– 0.850	1.361
	3.48	0.10	0.2802	0.2014	0.0892	– 0.986	1.498
	3.33	0.01	0.2978	0.2088	0.0960	– 1.022	1.503

Note: Each experiment is based on 5000 replications and $n = 50$.

theory a better chance of proving itself, we set $n = 100$ and allowed for 20,000 replications in order to reduce the experimental error.

Table 5 shows the estimated quantiles from the EDFs of the J and J_{npo} statistics. As shown in the table, the J_{npo} statistic delivers quantile estimates which are near to those of the J test when the samples are nearly orthogonal. This is especially true when $q = 2$, with the difference between corresponding quantiles becoming slightly greater as q increases. For instance, the 90% pairs of quantiles of the J and J_{npo} statistics when $\rho = 0.20$ and $q = [(2), (4), (6)]$ are $[(1.666, 1.652), (1.925, 1.881), (2.260, 2.181)]$ respectively. On the other hand, when the samples are not orthogonal the quantile differences are greater. Consider for example the 90% quantiles of J and J_{npo} for the three values of q when $\rho = 0.70$. They are $[(1.368, 1.264), (1.642, 1.540), (1.923, 1.691)]$ respectively. Clearly the differences in the latter pairs of quantiles are greater than the differences in the former pairs. These results are as expected, since the NPO expression (4.7) should have more predictive power with nearly orthogonal samples.

Table 6 gives the estimated quantiles for the C and C_{npo} statistics. The results here are, in some respects, different from those in Table 5. For every value of q ,

Table 5
Quantiles from the edfs of the J and J_{np0} statistics

		Quantiles from the edf of												
		J					J_{np0}							
q	c	ρ	50%	75%	90%	95%	97.5%	99%	50%	75%	90%	95%	97.5%	99%
2	7.06	0.70	0.178	0.752	1.368	1.757	2.062	2.443	0.103	0.655	1.264	1.715	2.032	2.302
	4.16	0.40	0.178	0.833	1.429	1.790	2.088	2.459	0.190	0.944	1.406	1.792	2.077	2.462
	2.41	0.20	0.408	1.063	1.666	2.028	2.346	2.713	0.405	1.059	1.652	2.011	2.319	2.662
	1.99	0.10	0.607	1.238	1.826	2.169	2.482	2.856	0.608	1.230	1.794	2.131	2.435	2.797
4	1.50	0.01	0.710	1.327	1.900	2.244	2.535	2.859	0.709	1.323	1.876	2.221	2.502	2.822
	8.21	0.70	0.351	1.030	1.642	2.005	2.330	2.707	0.252	0.924	1.540	1.925	2.199	2.593
	5.66	0.40	0.433	1.090	1.695	2.055	2.361	2.737	0.390	1.054	1.663	2.022	2.289	2.606
	3.70	0.20	0.679	1.341	1.925	2.275	2.575	2.997	0.673	1.305	1.881	2.237	2.488	2.861
6	2.97	0.10	0.940	1.580	2.161	2.513	2.845	3.195	0.934	1.550	1.121	2.483	2.727	3.051
	2.66	0.01	1.069	1.689	2.268	2.628	2.913	3.323	1.078	1.689	2.240	2.608	2.859	3.189
	8.28	0.70	0.650	1.315	1.923	2.309	2.629	3.045	0.425	1.900	1.691	2.052	2.341	2.707
	5.88	0.40	0.716	1.390	1.998	2.361	2.667	3.023	0.638	1.281	1.863	2.227	2.533	2.885
6	4.17	0.20	1.015	1.678	2.260	2.610	2.908	3.287	0.987	1.613	2.181	2.526	2.815	3.141
	3.48	0.10	1.279	1.910	2.492	2.853	3.145	3.474	1.224	1.862	2.427	2.771	3.032	3.388
	3.33	0.01	1.380	2.001	2.593	2.948	3.243	3.625	1.351	1.961	2.519	2.856	3.151	3.511

Note: Each experiment is based on 20,000 replications with $n = 100$.

Table 6
Quantiles from the edfs of the C and C_{npo} statistics

		Quantiles from the edf of												
		C					C _{npo}							
q	c	ρ	50%	75%	90%	95%	97.5%	99%	50%	75%	90%	95%	97.5%	99%
2	7.06	0.70	-0.035	0.635	1.210	1.543	1.821	2.141	-0.096	0.575	1.191	1.562	1.883	2.258
	4.16	0.40	-0.072	0.578	1.090	1.830	1.330	1.583	-0.178	0.492	1.095	1.471	1.784	2.201
	2.41	0.20	-0.207	0.419	0.865	1.099	1.294	1.516	-0.363	0.354	0.981	1.372	1.707	2.081
	1.99	0.10	-0.349	0.265	0.681	0.905	1.072	1.263	-0.525	0.224	0.873	1.270	1.600	1.947
	1.50	0.01	-0.443	0.191	0.588	0.805	0.969	1.171	-0.635	0.148	0.822	1.238	1.568	1.968
4	8.21	0.70	-0.177	0.520	1.090	1.423	1.706	2.044	-0.203	0.483	1.081	1.453	1.774	2.142
	5.66	0.40	-0.273	0.401	0.940	1.221	1.462	1.738	-0.329	0.365	0.984	1.350	1.659	2.021
	3.70	0.20	-0.481	0.202	0.704	0.977	1.184	1.425	-0.556	0.167	0.814	1.198	1.539	1.907
	2.97	0.10	-0.623	0.059	0.556	0.814	1.023	1.244	-0.717	0.048	0.691	1.404	1.075	1.787
	2.66	0.01	-0.682	0.040	0.520	0.772	0.974	1.206	-0.807	0.034	0.637	1.036	1.378	1.768
6	8.28	0.70	-0.331	0.341	0.934	1.266	1.525	1.866	-0.305	0.377	0.993	1.358	1.677	2.053
	5.88	0.40	-0.465	0.231	0.785	1.105	1.351	1.599	-0.494	0.218	0.821	1.188	1.493	1.861
	4.17	0.20	-0.653	0.043	0.583	0.873	1.105	1.377	-0.729	0.001	0.644	1.021	1.343	1.731
	3.48	0.10	-0.783	0.082	0.465	0.761	0.982	1.210	-0.871	-0.128	0.546	0.948	1.273	1.644
	3.33	0.01	-0.852	0.135	0.406	0.702	0.923	1.147	-0.948	-0.165	0.503	0.879	1.231	1.616

Note: Each experiment is based on 20,000 replications and n = 100.

the 50% quantiles are now negative rather than positive. Also, with nearly orthogonal samples the C and C_{np0} quantiles are smaller than those of the J and J_{np0} statistics. Again, this reflects the leftward shift in the distribution of the C test that was documented above in Table 2 and Fig. 2. More importantly, the predictive performance of C_{np0} is not as good as that of the J_{np0} statistic. With nearly orthogonal samples the gap between the C and C_{np0} quantiles is greater than the gap between the J and J_{np0} quantiles. For instance, with $q = 2$ and $\rho = 0.01$ the absolute difference of the 90% quantiles in Table 6 is 0.234, about ten times larger than the value of 0.024 observed under the same conditions in Table 5. At $q = 4$ or $q = 6$, the quantile differences are reduced at every nominal level but remain relatively larger compared to those in Table 5.

To extract more information about the behavior of the EDFs with nearly orthogonal samples, we plotted them in Figs. 3-8. In some figures, the

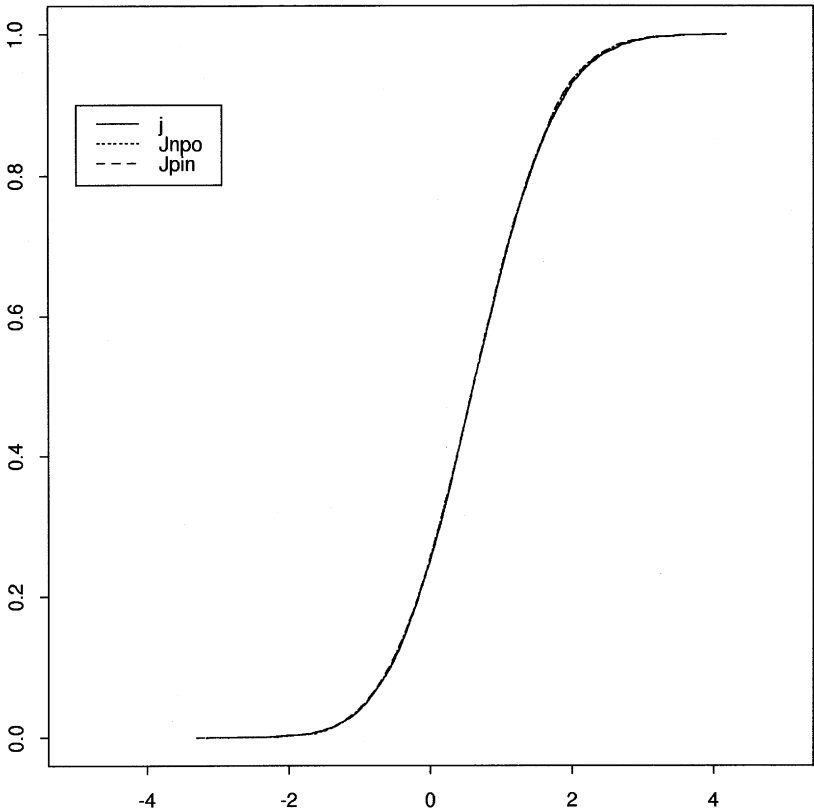


Fig. 3. The edfs of the J , J_{np0} and J_{pin} tests for $\rho = 0.10$, $c = 1.99$ and $q = 2$. The edfs are based on 20,000 replications with $n = 100$.

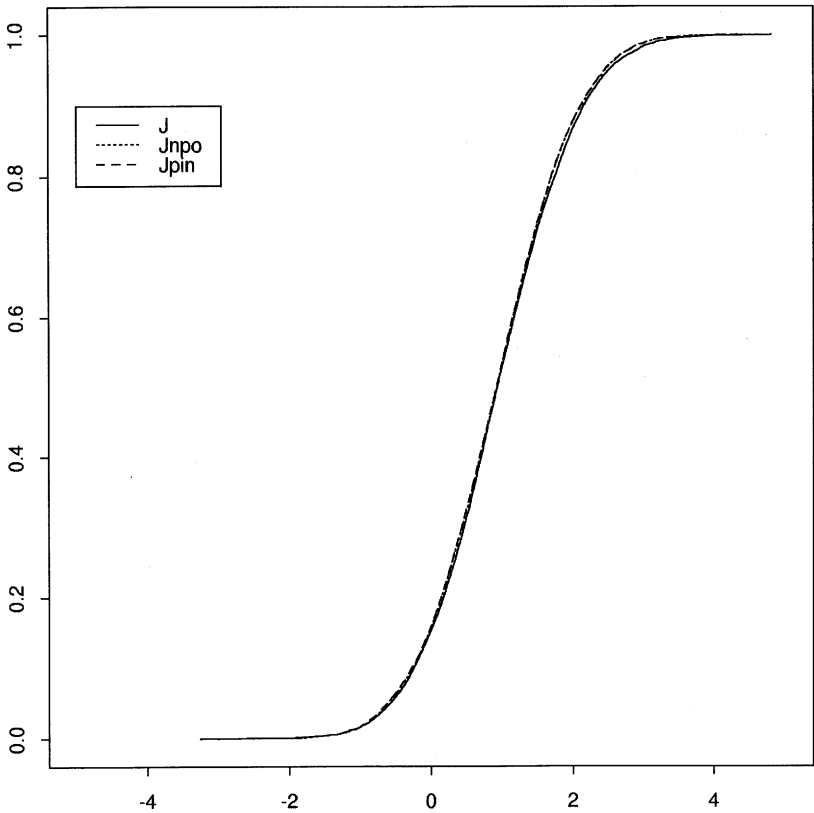


Fig. 4. The edfs of the J , J_{npo} and J_{pin} tests for $\rho = 0.10$, $c = 2.97$ and $q = 4$. The edfs are based on 20,000 replications with $n = 100$.

horizontal axis was set at different ranges in order to highlight the distinction among the tests. In the figures, in addition to the EDFs of J , J_{npo} , C and C_{npo} , the EDFs of J_{pin} and C_{pin} were also plotted. The latter two EDFs are useful in practice and correspond to the asymptotic expressions (4.7) and (4.14) with c estimated by \hat{c} from the simulated data rather than assumed known. Because of the consistency of \hat{c} , the EDFs of J_{pin} and C_{pin} are almost exactly overlaid onto the EDFs of J_{npo} and C_{npo} respectively, so that only two EDFs appear visible in each graph. In all cases shown, the EDFs were constructed from 20,000 replications with $n = 100$, $\rho = 0.10$ and $q = 2, 4, 6$ along with corresponding values of c . As shown in the Figs. 3–5, the EDFs of J and J_{npo} are much more tied together than the EDFs of C and C_{npo} in the Figs. 6–8. The EDFs of J_{npo} track very well the EDFs of the J statistic throughout the whole range of their variation, but that is not the case with the EDFs of C and C_{npo} which seem

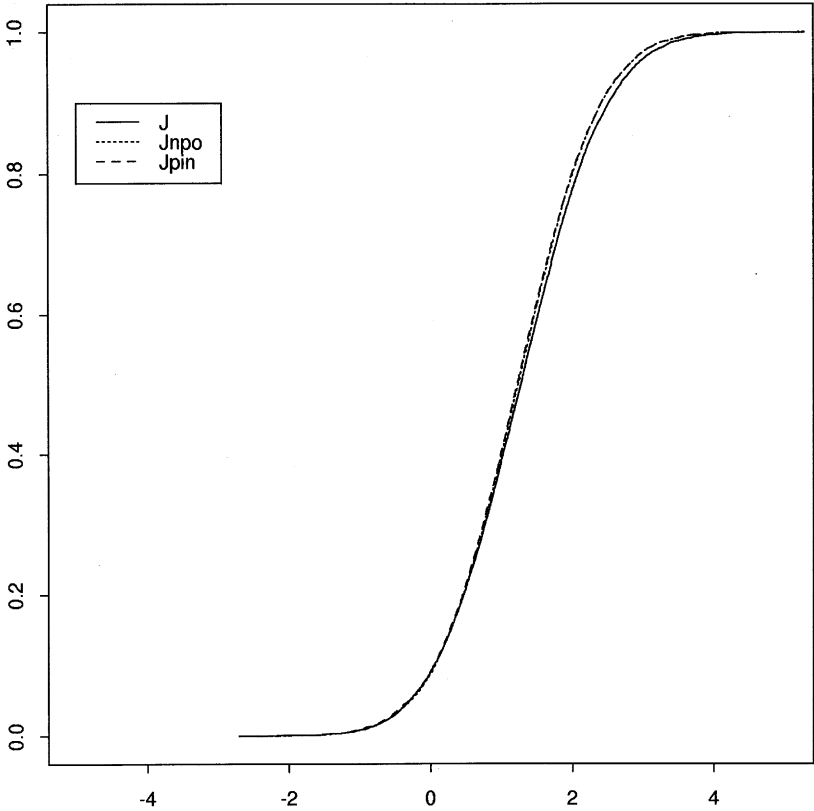


Fig. 5. The edfs of the J , J_{npo} and J_{pin} tests for $\rho = 0.10$, $c = 3.48$ and $q = 6$. The edfs are based on 20,000 replications with $n = 100$.

to deviate more from each other in the tails of their distributions. Further, the C and C_{npo} statistics have a much longer left tail than do the J and J_{npo} statistics. Overall, it seems that J_{npo} predicts better the behavior of the J test than does C_{npo} for the C test.

It is of considerable importance that the observable J_{pin} and C_{pin} statistics track almost exactly the J_{npo} and C_{npo} statistics which are not observable in practice. This means that the above arguments could be justifiably re-casted in terms of the J_{pin} and C_{pin} statistics instead of the J_{npo} and C_{npo} statistics. Further, as an anonymous referee suggested, one may use the observed quantiles of the J_{pin} and C_{pin} statistics, rather than those of $N(0, 1)$, for the purpose of testing nearly orthogonal regression models. We outline this testing procedure in the next section.

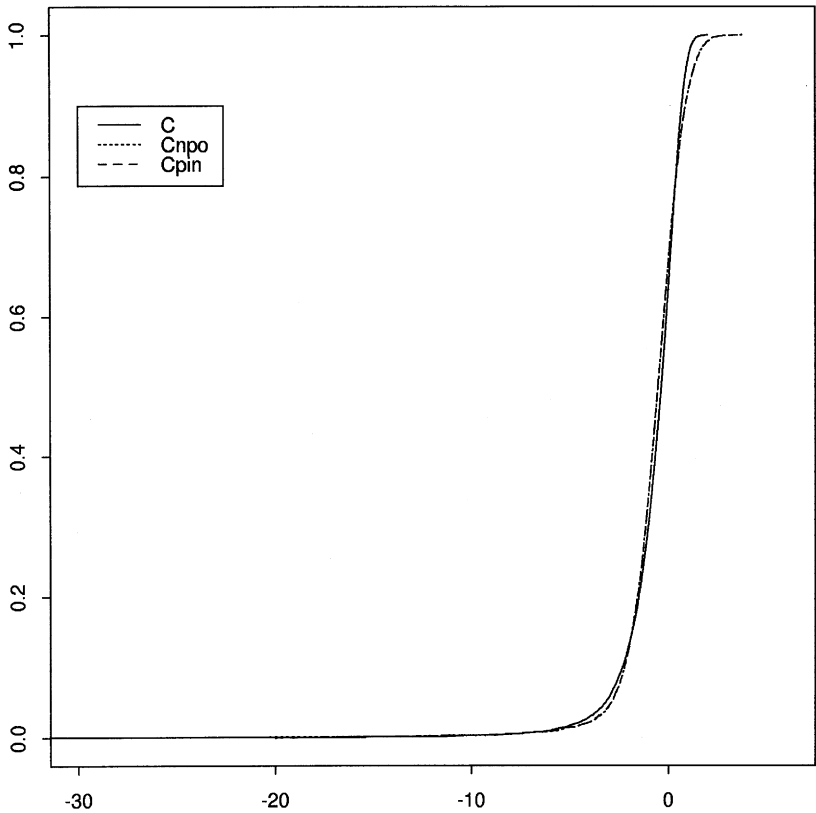


Fig. 6. The edfs of the C , C_{npo} and C_{pin} tests for $\rho = 0.10$, $c = 1.99$ and $q = 2$. The edfs are based on 20,000 replications with $n = 100$.

5.2.3. The plug in test

In this section we exploit the fact that the \hat{c} is a consistent estimator of c and use it to construct reasonably sized tests for testing nearly orthogonal non-nested regression models. To construct a test, realized values of \hat{c} are plugged into either (4.7) or (4.14) instead of known values of c . For this reason, we call this the plug in test.⁶ Two issues should be clear in relation to this test. First, significance is the key concern with the plug in test; that is, the question is whether the population value of c is small enough to invalidate using the standard J and C tests. Second, this is a somewhat ad hoc procedure since the

⁶ The term plug in was suggested by an anonymous associate editor.

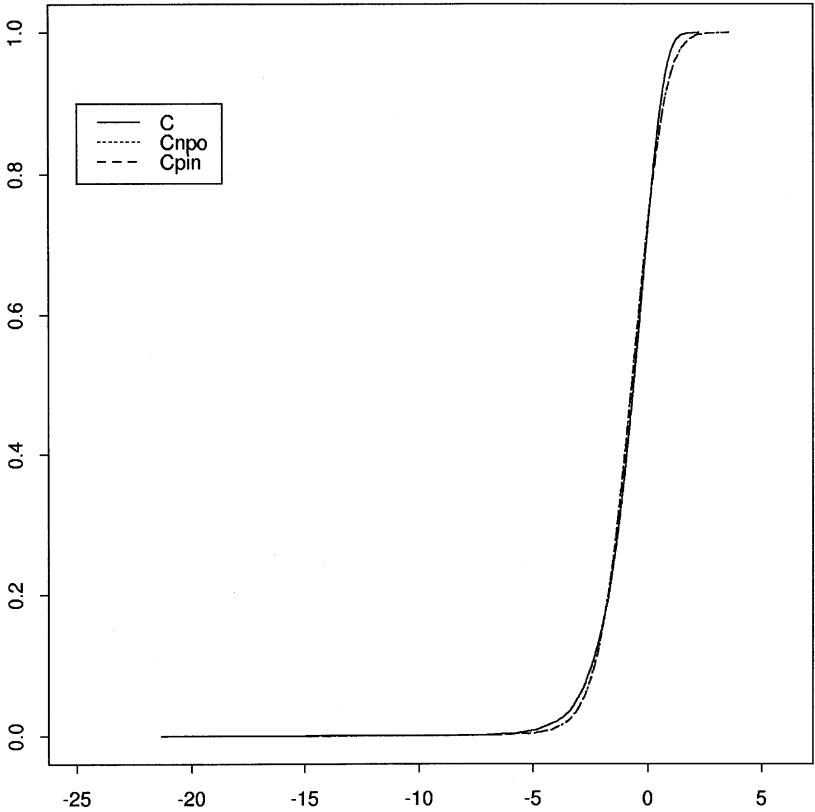


Fig. 7. The edfs of the C , C_{npo} and C_{pin} tests for $\rho = 0.10$, $c = 2.97$ and $q = 4$. The edfs are based on 20,000 replications with $n = 100$.

actual distribution of each test conditional on the outcome of the pre-test with respect to c is unknown.

To carry out this test, the following steps are required:

1. Use the data to estimate c by $\hat{c} = \hat{\beta}'X'P_zX\hat{\beta}/\hat{\sigma}$ where $\hat{\beta}$ and $\hat{\sigma}$ are the OLS or maximum likelihood estimators of β and σ respectively. If \hat{c}^2 is statistically significant, relative to an appropriate χ^2 distribution, conclude that the models are not nearly orthogonal and use the conventional J or C test. Otherwise proceed as in (b) and (c).
2. Plug in \hat{c} in (4.7) or (4.14) and compute the desired quantile of the J_{pin} or C_{pin} statistic.
3. Compute the J or C test and compare its realized value to the critical value (quantile) obtained from (b). Reject the model under the null if the test statistic exceeds the critical value.

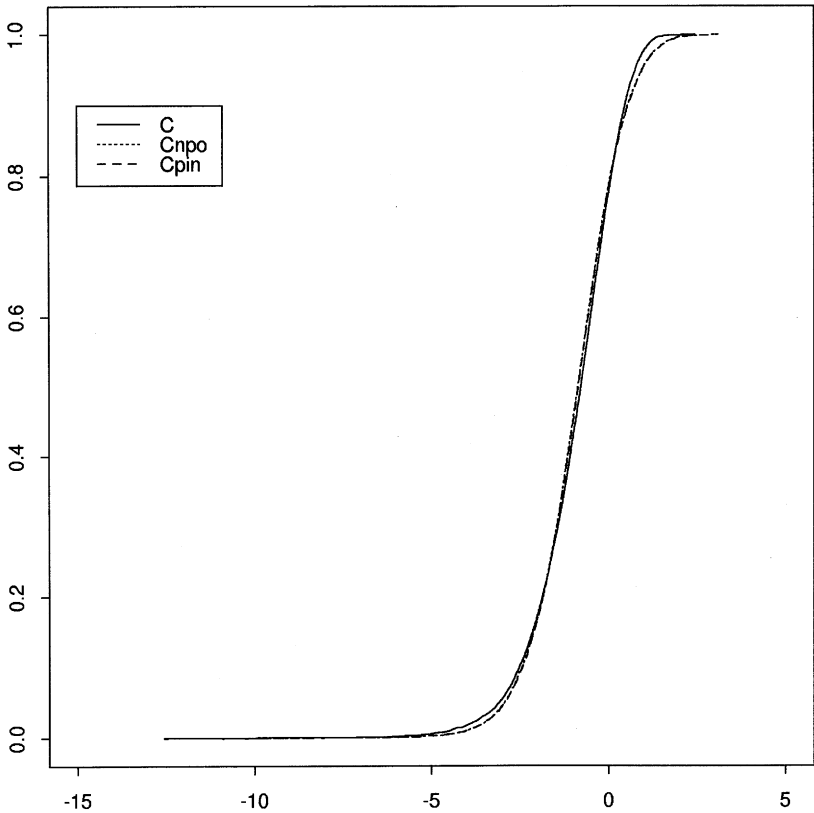


Fig. 8. The edfs of the C , C_{npo} and C_{pin} tests for $\rho = 0.10$, $c = 3.48$ and $q = 6$. The edfs are based on 20,000 replications with $n = 100$.

Is the plug in test properly sized? We investigated this question by the third set of the Monte Carlo experiments. The simulations were carried out for nearly orthogonal samples only and for $n = 100$. Tables 7 and 8 contain the simulation results for the estimated size of the plug in tests, denoted J_p and C_p respectively. The critical values used for the two tests were obtained from the EDFs of the J_{pin} and C_{pin} statistics respectively based on 20,000 replications with $n = 100$. For the sake of comparison, the sizes of the conventional J and C tests are also given in the tables.

Table 7 shows the estimated size of the J_p and J tests. It is clear from the table that the size of J_p is remarkably near its nominal level at every value of q . It over-rejects the null only slightly at the 10% nominal level when $q = 4$ or $q = 6$ and $\rho = 0.01$. On the other hand, under the same conditions the J test has very large size distortion which increases dramatically with q even when $n = 100$.

Table 7
Estimated Size of the J_p and J tests

q	c	ρ	Estimated size of					
			J_p			J		
			10%	5%	1%	10%	5%	1%
2	2.41	0.20	0.1018	0.0495	0.0140	0.1189	0.0650	0.0186
	1.99	0.10	0.1094	0.0558	0.0114	0.1508	0.0846	0.0208
	1.50	0.01	0.1050	0.0578	0.0136	0.1658	0.0944	0.0224
4	3.70	0.20	0.1084	0.0566	0.0116	0.1996	0.1214	0.0384
	2.97	0.10	0.1098	0.0572	0.0146	0.2678	0.1674	0.0512
	2.66	0.01	0.1114	0.0574	0.0130	0.2778	0.1774	0.0574
6	4.17	0.20	0.1058	0.0518	0.0138	0.3052	0.2012	0.0682
	2.48	0.10	0.1090	0.0522	0.0118	0.3614	0.2508	0.0914
	3.33	0.01	0.1108	0.0516	0.0148	0.3936	0.2684	0.1038

Note: Each experiment is based on 20,000 replications and $n = 100$.

Table 8
Estimated Size of the C_p and C tests

q	c	ρ	Estimated size of					
			C_p			C		
			10%	5%	1%	10%	5%	1%
2	2.41	0.20	0.0950	0.0480	0.0094	0.1412	0.0952	0.0458
	1.99	0.10	0.0881	0.0413	0.0101	0.1770	0.1398	0.0818
	1.50	0.01	0.0881	0.0398	0.0081	0.2080	0.1652	0.1034
4	3.70	0.20	0.0856	0.0163	0.0103	0.1742	0.1224	0.0554
	2.97	0.10	0.0877	0.0430	0.0100	0.2000	0.1474	0.0858
	2.66	0.01	0.0824	0.0404	0.0080	0.2436	0.1830	0.1054
6	4.17	0.20	0.0788	0.0385	0.0080	0.2146	0.1532	0.0714
	3.48	0.10	0.0751	0.0342	0.0067	0.2566	0.1866	0.0918
	3.33	0.01	0.0669	0.0288	0.0053	0.2854	0.2078	0.1044

Note: Each experiment is based on 20,000 replications and $n = 100$.

Table 8 reports the results for the C_p and C tests. As the results indicate, the C_p test tends to be undersized at every value of q . It seems that the under-rejection problem becomes more severe when q increases and when the samples

become more orthogonal. This under-rejection property of C_p can be attributed to the fact that the quantiles of C_{pin} are larger than the quantiles of C with nearly orthogonal samples; see Table 6 and Figs. 6–8.

Comparing the simulation results in Tables 7 and 8 it is plainly clear that the J_p test outperforms the C_p test in terms of size. The size of J_p is much closer to the nominal level at every value of q and is more stable across nearly orthogonal samples. For those reasons, it may be preferable to use the J_p test in practical applications. An empirical example is given in the next section.

6. Empirical example

This section gives a brief empirical example to demonstrate the practical significance of weakly correlated regressors and, at the same time, demonstrate the plug in testing procedure outlined in the previous section. Beyond these objectives, no serious attempt was made to evaluate the models with the battery of model specification tests that exist in the literature. In this sense, the empirical results should be viewed as illustrative.

In the example we consider two non-nested aggregate unemployment models: one justified by standard Keynesian arguments (the K -model) and the other deriving from the Natural Rate/Rational Expectations hypothesis (the R -model); see for example Neftci and Sargent (1978) and Pesaran (1982). From a Keynesian perspective, the systematic components of fiscal and monetary policies can influence the real side of an economy, so that including lagged values of such policy variables in an unemployment equation makes perfect sense. On the other hand, according to the Natural Rate/Rational Expectations hypothesis only the random components of government policies matter. The real effects of the systematic components of policies are nullified by the offsetting behavior of rational and forward looking economic agents. The K and R models were specified as follows:

$$K: U_t = \beta_0 + \beta_1 U_{t-1} + \beta_2 t + \beta_3 M_{t-1} + \beta_4 G_{t-1} + \varepsilon_{Kt}, \quad (6.1)$$

$$R: U_t = \gamma_0 + \gamma_1 U_{t-1} + \gamma_2 RESG_t + \gamma_3 RESM_t + \varepsilon_{Rt}. \quad (6.2)$$

The variables are: $U_t = \log[UR_t/(1 - UR_t)]$, where UR_t is the rate of unemployment, $M_t = \log(M1_t)$, where $M1_t$ is the narrow definition of the money supply, $G_t = \log(GE_t)$ where GE_t is real government expenditures measured in 1990 prices and $RESG_t$ and $RESM_t$ are proxies for the unobserved random components of the fiscal and monetary policies respectively. These proxies were obtained as the OLS residuals from regressing G_t and M_t on a constant, U_{t-1} , M_{t-1} and G_{t-1} respectively. As in Pesaran (1982), the Keynesian specification includes a time trend variable, t , to capture trend changes in labour participation and other slowly changing variables.

The data used in the empirical example are annual time series US data for the period 1948–1996 and, except for the rate of unemployment, were obtained from the IFS data base of the IMF. The unemployment rate was not available in the IFS and was obtained from the Cansim tape of Statistics Canada. Given these data, the two models were estimated by the method of iterated maximum likelihood which adjusts for first-order serial correlation. The fitted versions of Eqs. (6.1) and (6.2) are

$$K: \hat{U}_t = -1.29 + 0.225U_{t-1} + 0.026t - 0.003G_{t-1} - 0.237M_{t-1},$$

(1.54) (0.151) (0.034) (0.001) (0.151)

$$R: \hat{U}_t = -2.22 + 0.213U_{t-1} - 0.002RESG_t + 1.68RESM_t,$$

(0.387) (0.0003) (0.133) (0.802)

The numbers in the parentheses are standard errors of the estimated coefficients. Also, we computed $R^2 = 0.60$, and $DW = 1.86$ for the K -model and $R^2 = 0.66$ and $DW = 1.97$ for the R -model.

In testing K versus R , the \hat{c} estimate turned out to be 1.989. The square of \hat{c} is 3.956 which is less than 5.99, the 5% critical value of the χ^2 distribution with $q = 2$ degrees of freedom. Consequently, we conclude that the two models are nearly orthogonal and proceed with steps (b) and (c) of the plug in test. The J_p test will be employed since it has better size properties than the C_p test. Suppose the non-nested test is carried out at the 5% level. Using the corresponding critical values from the $N(0, 1)$ and $\chi^2(1)$ distributions along with $\hat{c} = 1.989$ in the expression (4.7) yields the critical value $J_{\text{pin}} = 2.63$. The realized value of the J statistic was computed from the data to be 3.88. Since this exceeds the number 2.63, the K model is rejected by the evidence provided by the R model.

Similarly, in testing the R -model against the K -model the value of \hat{c} was 0.447. This together with the 5% critical values of the $N(0, 1)$ and $\chi^2(2)$ yielded the value $J_{\text{pin}} = 3.119$. In this case the J statistic was computed to be 2.75. This falls in the non-rejection region of the test. Therefore, the R -model can not be rejected by the evidence provided by the K -model. This turns out to be a rather appealing result since the R -model is more parsimonious than the K model ($q = 2$ versus $q = 3$). Also, notice that if the conventional procedure had been applied both models would have been rejected since in both cases the J test is greater than 1.96, the critical value of the $N(0, 1)$.

7. Conclusion

This paper investigated the asymptotic null distributions of the J and Cox non-nested tests under a specific assumption of model orthogonality. Having

introduced the NPO condition, new distributional results have been obtained. The two tests converge to two different random variables asymptotically, each of which is expressible as a function of a nuisance parameter, c , a $N(0, 1)$ variate and a χ^2 variate. As a result, they are no longer asymptotically equivalent as they are in the standard case of non-orthogonal models.

The simulation results show that, with nearly orthogonal models, the J and C tests have excessive size distortion that increases with the number of non-nested regressors in the alternative model. The size distortion is accounted for, in large measure, by the simulated NPO expressions derived in the paper. Our formalization of NPO by the condition (A4) allows for consistent estimation of the nuisance parameter, c , from the data. Based on this insight, a new plug in testing procedure is proposed for testing nearly orthogonal non-nested regression models. The Monte Carlo evidence shows that the plug in test incorporating the J test has better size properties than the plug in test associated with the C test. An empirical example involving two aggregate non-nested unemployment models was used to demonstrate the practical usefulness of the plug in test.

This paper has focused on the null distributions of the J and Cox non-nested tests under NPO. It would be informative to investigate and compare the size/power properties of the plug in test with other testing procedures that may have desirable properties under near orthogonality. One possibility is to consider J -type or C -type bootstrap tests. Another possibility is to consider non-nested pretest tests that combine the J or C test through pretesting for model orthogonality with other tests, such as the J_A test of Fisher and McAleer (1981) or the encompassing F test. This is a topic under current research by the author.

Acknowledgements

I am grateful to R. Davidson and J.G. MacKinnon for helpful suggestions and comments, an associate editor for constructive advice and an anonymous referee for very important suggestions including the plug in procedure. Also I thank T. Breusch, R. Carter, J. Knight, A. Mansoorian, A.L. Nagar and T. Stengos for helpful comments. All remaining errors are my own.

Appendix

Proof of Theorem 4.1. Start from (4.6) and consider first the term $\beta'X'P_zu$. It is easily seen that $V_0 \equiv \beta'X'P_zu/\sigma(\beta'X'P_zX\beta)^{1/2}$ is a $N(0, 1)$ variate asymptotically. Also, under (A1)–(A3) $\beta'X'P_zX\beta \xrightarrow{P} D \equiv \beta'\Delta\Sigma_{zz}^{-1}\Delta'\beta$. Therefore, we obtain

$\beta'X'P_zu \stackrel{a}{=} \sigma D^{1/2}V_0$. Next, consider the term $V_q \equiv u'P_zu/\sigma^2$. Obviously this is distributed as $\chi^2(q)$ asymptotically. Further, notice that $u'P_zu$ can be written as the sum of two independent components: $V_q = V_0^2 + V_{q-1}$. In fact,

$$\sigma^2V_q = u'P_zu = u'P_1u + u'(P_z - P_1)u, \quad (\text{A.1})$$

where P_1 is the orthogonal projection matrix on to $S[P_zX\beta]$. The decomposition in (A.1) can be understood by noting that $P_zX\beta$ is one vector in the q -dimensional span of P_z , and so P_z can be split into the sum of two mutually orthogonal projections, one on to the direction of $P_zX\beta$ and one on to its orthogonal complement. Dividing both sides of (A.1) by σ^2 we obtain

$$V_q = u'P_1u/\sigma^2 + u'(P_z - P_1)u/\sigma^2. \quad (\text{A.2})$$

The first term in (A.2) is just V_0^2 and the second term, denoted V_{q-1} , is independent of it and distributed as $\chi^2(q-1)$. To complete the proof, substitute $\sigma D^{1/2}V_0$ for $\beta'X'P_zu$ and σ^2V_q for $u'P_zu$ in (4.6). Then using the decomposition $V_q = V_0^2 + V_{q-1}$, letting $c = D^{1/2}/\sigma$ and manipulating the resulting expression algebraically gives the desired result. \square

Proof of Theorem 4.2. It suffices to consider the distribution of each term in (4.13). From the proof of Theorem 4.1 we know that $\beta'X'P_zu \stackrel{a}{=} \sigma D^{1/2}V_0$ where $V_0 \sim N(0, 1)$, $u'P_zu \stackrel{a}{=} \sigma^2V_q$ where $V_q \sim \chi^2(q)$, and $\beta'X'P_xX\beta \xrightarrow{p} D$. Consequently, defining $c = D^{1/2}/\sigma$ as before and substituting each of these relations in (4.13) gives the desired result (4.14). \square

References

- Cox, D.R., 1961. Tests of separate families of hypotheses. Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 105–123.
- Cox, D.R., 1962. Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society, Series B 24, 406–424.
- Dastoor, N.K., 1983. Some aspects of testing non-nested hypotheses. Journal of Econometrics 17, 213–228.
- Davidson, R., MacKinnon, J.G., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 49, 781–793.
- Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press, New York.
- Fisher, G.R., 1983. Tests for two separate regressions. Journal of Econometrics 21, 117–132.
- Fisher, G.R., McAleer, M., 1981. Alternative procedures and associated tests of significance for non-nested hypotheses. Journal of Econometrics 16, 103–119.
- Godfrey, L.G., Pesaran, M.H., 1983. Tests of non-nested regression models: small sample adjustments and Monte Carlo evidence. Journal of Econometrics 21, 133–154.
- Kent, J., 1986. The underlying structure of nonnested hypothesis tests. Biometrika 73, 333–343.
- Klein, R.W., 1983. Model specification tests against non-nested alternatives: comment. Econometric Reviews 2, 115–119.

- Michelis, L., 1996. Orthogonal regression models and the distribution of non-nested tests. *Journal of Quantitative Economics* 12, 1–16.
- McAleer, M., Pesaran, M.H., 1986. Statistical inference in non-nested econometric models. *Applied Mathematics and Computation* 272–311.
- Mizon, G.E., Richard, J.-F., 1986. The encompassing principle and its application to testing non-nested hypotheses. *Econometrica* 54, 657–678.
- Neftci, S., Sargent, T.J., 1978. A little bit of evidence on the natural rate hypothesis from the U.S. *Journal of Monetary Economics* 4, 315–319.
- Pesaran, M.H., 1974. On the general problem of model selection. *Review of Economic Studies* 41, 153–171.
- Pesaran, M.H., 1982. A critique of the proposed tests of the natural rate-rational expectations hypothesis. *Economic Journal* 92, 529–554.
- Staiger, D., Stock, J.H., 1997. Instrumental variable regression with weak instruments. *Econometrica* 65, 557–586.
- Statistical Sciences, S-PLUS, 1993. User's Manual, Version 3.2, Seattle, MathSoft Inc.
- Szroeter, J., 1992. The asymptotic local structure of the cox modified likelihood-ratio statistic for testing non-nested hypotheses. *Econometric Theory* 8, 553–569.