TESTING NON-NESTED MODELS

Howard Doran*

* Department of Econometrics, University of New England,

Armidale 2351, Australia

1. Introduction

Since the mid-seventies, a major area of econometric research has been the development of procedures to test the specification of empirical models. This article deals with testing non-nested (or separate) models. Two models are said to be 'nested' when one is a special case of the other, obtained by parameter restrictions. Thus, the Cobb-Douglas production function is nested within the constant elasticity of substitution (CES) production function. If, on the other hand, one model cannot be expressed as a special case of the other by parameter restrictions, the models are 'non-nested'. Applied economics gives rise to many instances in which non-nested models are being examined as adequate descriptions of data. The following four examples should suffice to show this:

(i) Models may differ in functional form. Often the competing functional forms will be non-nested, for example the variable elasticity of substitution (VES) and CES production functions.

(ii) Sometimes models may differ in the definitions of explanatory variables. For example, supply response models may use different definitions of expected price [see, for example, Shideed and White (1989)]. Again, in money demand studies, different definitions of interest rate may be used.

(iii) Sometimes different theories (eg. about the consumption function) are being tested. At the end of this paper, an empirical example is given of testing the Absolute and Relative Income hypotheses.

(iv) Linear and log-linear formulations are sometimes compared. Testing between Cobb-Douglas functions with additive and multiplicative errors is an example.

This article is in no sense a review of specification testing techniques. Several excellent and comprehensive such reviews already exist. For general specification testing reference is made to Pagan and Hall (1983) and MacKinnon (1991). The area of non-nested testing has been reviewed by MacKinnon (1983) and McAleer (1987), with the latter article being the source of much of the material covered here. Rather, an attempt is being made to bring some of the basic results to a particular audience - applied workers who have neither the training nor inclination to keep abreast of theoretical econometric developments. It is however assumed that readers have a fairly thorough grounding in regression analysis.

It is hoped that readers who are unfamiliar with the techniques of testing non-nested models will realize how easy they are to apply through artificial regressions. At the same time, the aim is to present sufficient theory to avoid the trap of merely presenting 'cookbook' recipes.

The plan of the paper is as follows. Section 2 gives some general preliminaries; Section 3 shows how the Cox principle can be utilized by constructing certain artificial regressions; Section 4 considers another approach - that of artificial nesting; Section 5 deals with a test of linear and log-linear models; Section 6 gives an empirical example: and Section 7 is a conclusion.

2. Preliminaries

(a) *Testing versus Discrimination*

Testing nested hypotheses (for example, constant returns to scale in a Cobb-Douglas production function) forms a fundamental part of any elementary econometrics course and has been standard practice among applied workers for decades. However, at least until recently, the approach to non-nested models has been different, with the emphasis on discrimination rather than testing. Thus, the best model is chosen on some criterion of performance (or sample explanation), such as R^2, $\bar{R}^2$, AIC, etc., together with an *ad hoc* perusal of signs.

There are two fundamental differences between testing and the latter approach. First, when the best model is sought, one model will always be chosen. In the testing framework, all models may be (and often are) rejected as being inadequate. Second, as MacKinnon (1983) points out, non-nested tests are tests in the usual sense, and probabilities of incorrect rejection of a null are assigned. No such probabilities exist in the discrimination approach.

(b) *Predicting the Performance of a Model*

Let us suppose that the true data generating process (DGP) of a random variable y is known. It stands to reason that in such (unrealistic) cases, we should be able to predict the performance of any model at all which is postulated to describe the behaviour of y. In order to fix ideas, let us assume the DGP is known to be

$$y = X_0\beta_0 + u_0, \qquad E(u_0^2) = \sigma_0^2,$$

with the parameters β_0 and σ_0^2 known. A model of the form

$$y = X_1\beta_1 + u_1, \qquad E(u_1^2) = \sigma_1^2$$

is postulated, where the parameters β_1 and σ_1^2 are not known and have to be estimated. We will take $\hat{\sigma}_1^2$ to be a measure of the performance of the model. It is now quite easy to show, on the basis of the DGP, that

$$E(\hat{\sigma}_1^2) = \sigma_0^2 + \beta_0'X_0'(I-P_1)X_0\beta_0 \qquad (2.1)$$

where

$$P_1 = X_1(X_1'X_1)^{-1}X_1' . \qquad (2.2)$$

We note in passing, that $E(\hat{\sigma}_1^2)$ is a function of the parameters β_0, σ_0^2 of the DGP.

This ability to predict the performance of a model on the basis of the DGP of the dependent variable is the key concept in testing non-nested models, and was first utilized by Cox (1962). In essence it is used as follows: A null hypothesis, H_0, is postulated to be the DGP. Another hypothesis, H_1, is advanced and its performance is predicted on the basis of the null hypothesis. If the actual performance $\hat{\sigma}_1^2$ is close to its predicted performance, then this serves as confirmation of H_0. On the other hand, if the actual performance is nowhere near what had been predicted, then the null hypothesis H_0 will be rejected. It is crucial to realize that rejecting H_0 makes absolutely no implications about the adequacy of H_1.

This approach of comparing actual results with those expected under a null is familiar to anyone who has done an elementary statistics course in the form of χ^2 tests for contingency tables. There is however a crucial difference. In the contingency table example, the null and alternative are jointly exhaustive - one must be true. Thus, rejection of the null implies acceptance of the alternative. In the case of non-nested models however there are, in principle, an infinite number of alternative hypotheses.

(c) *Testing Principles*

Here an excellent discussion by MacKinnon (1991) is closely followed. Let us consider a very simple model for y, viz.

$$H_0: \quad y_t = \beta_0 + \beta_1 x_t + u_t,$$

which is being tested by the addition of a variable z_t. Many common tests are of this form; for example, the RESET test, tests for autocorrelation and tests for omitted variables.

Suppose that the DGP is

$$y_t = \beta_0^* + \beta_1^* x_t + \delta w_t + u_t .$$

If a regression

$$y_t = \beta_0 + \beta_1 x_t + \gamma z_t + u_t$$

is carried out, then for large samples

(i) the standard error of $\hat{\gamma} \to 0$;

(ii) $\hat{\gamma} \approx \delta \hat{\beta}_{wz|x}$

where $\hat{\beta}_{wz|x}$ is the slope coefficient obtained when the residuals from w on x are regressed on the residuals from z on x. Thus, $\hat{\gamma} \neq 0$ whenever w and z (purged of the effect of x) are correlated. It follows that in such cases, $t(\hat{\gamma})$, the t-statistic of $\hat{\gamma}$, will be significant when the sample is large, even if $\gamma = 0$! Thus, a significant $t(\hat{\gamma})$ tells us something is wrong with H_0, but does not tell us the augmented model is true. Any choice of z, provided it is correlated with the true variable w once the effect of x is removed, will give rise to a significant $t(\hat{\gamma})$.

As we will see, the most convenient way of testing non-nested models is to use the device of adding a variable and examining its associated t statistic. From the above analysis, rejection of a model A says nothing about another model B, even when the added variable is derived from model B. This leads to a vital feature of testing non-nested models. Suppose that initially H_0 is model A. Following the test, model B becomes H_0 and the testing process is repeated. It is clear that four outcomes are possible:

(i) Both models are rejected as inadequate;

(ii) Model A is 'accepted', while model B is rejected;

(iii) Model B is 'accepted', while model A is rejected;

(iv) both models are 'accepted'.

It is only with the last outcome that discrimination criteria should be used to choose the best model.

A crucial question in any specification testing context is the power of the procedure - that is, the ability of the test to detect inadequacies in the model which is being tested. Following Davidson and MacKinnon (1985), let us suppose that the DGP is[1]

$$y = X\beta_0 + n^{-1/2}W\delta_0 + u_0 , \quad E(u_0^2) = \sigma_0^2$$

and H_0 is given by

$$H_0: \quad y = X\beta_0 + u_0 .$$

Clearly, H_0 is inadequate, as the variables included in W have been incorrectly omitted.

Davidson and MacKinnon (1985) have shown that if H_0 is tested by adding r variables to give

$$H_1: \quad y = X\beta_0 + Z\gamma + u_0 ,$$

then the power of the test of $\gamma = 0$ to detect inadequacies in H_0 is directly related to a scalar h of the form

$$h = \sigma_0^{-2}\left[\delta_0'W'(I-P_0)W\delta_0/n\right]R^2 \qquad (2.3)$$

with R^2 arising from the regression of $W\delta_0$ on Z, once the effect of X has been removed. Davidson and MacKinnon draw three inferences from (2.3), as follows:

---

1 The factor $n^{-1/2}$ multiplying $W\delta_0$ is for technical reasons which need not concern us here.

(i) If σ_0^2 is large, the power of tests will be low, regardless of the choice of Z. This reflects the obvious fact that if y has a very substantial component of noise, then it will be difficult to marshall enough evidence to reject an incorrect H_0;

(ii) The second factor $\delta_0' W'(I-P_0)W\delta_0/n$ reflects the difference between H_0 and the 'truth', that is, the DGP. If the omitted term, $n^{-1/2} W\delta_0$, is important relative to $X\beta_0$, then this factor will be large and the inadequacy in H_0 should be picked up.

(iii) R^2 is the only factor involving the alternative hypothesis. If, having purged the effect of X, $W\delta_0$ and Z are strongly correlated, then h will be large and a test of $\gamma=0$ will be powerful in detecting inadequacies in H_0. This result implies a trade-off in designing a powerful test of H_0. It is, of course, always possible to increase R^2 by simply adding more variables to Z. An F-test of $\gamma = 0$ has degrees of freedom r and s, where r is the number of variables in Z and s depends on sample size. It is well-known that the power of an F-test is maximized when r is small and s is large. Thus, adding variables to Z increases power by increasing R^2, but decreases power by increasing r.

It follows that in designing a powerful test, the column dimension of Z should be as small as possible, commensurate with the ideal of being as close to the truth as possible. One of the major reasons for testing H_0 against a non-nested alternative as opposed to using, say, the RESET test, is that if the alternative also is based on economic theory, there is a good chance that the augmenting variable(s), Z, will be highly correlated with any omitted variables in H_0.

3. The Cox Principle and Artificial Regressions[2]

In the previous section we defined the Cox principle as a method of validating a model (H_0) by comparing the actual performance of another model (H_1) with the prediction, based on H_0, of this performance. Cox (1962) applied this principle to the logarithm of the likelihood ratio obtained from H_0 and H_1. Thus if $L_0(\hat{\alpha}_0)$ is the maximum value of the likelihood of a sample of y values when H_0 is postulated, and $L_1(\hat{\alpha}_1)$ is analogously defined for H_1, then $\hat{\ell}_{10}$, the logarithm of the likelihood ratio is given by

$$\hat{\ell}_{10} = \log L_0(\hat{\alpha}_0) - \log L_1(\hat{\alpha}_1) \; . \tag{3.1}$$

Cox proposed that H_0 be evaluated by finding the significance of a statistic T_0 given by

$$T_0 = \hat{\ell}_{10} - E_0(\hat{\ell}_{10})_{\alpha=\hat{\alpha}_0} \; . \tag{3.2}$$

In the above, E_0 means the expectation assuming H_0 to be the DGP. We have already noted in (2.1) that such an expectation will be a function of the parameters of the DGP - here α_0. As these parameters would in practice never be known, they are replaced by consistent estimates $\hat{\alpha}_0$.

If T_0 is 'close to zero', then the likelihood ratio is close to what it would be expected to be if H_0 were true. On the other hand, a 'large' positive or negative value of T_0 would indicate discrepancy between actual and expected likelihood ratios, and hence the validity of H_0 should be

---

2 The mathematical treatment in this section is intended for readers who are familiar with regression theory and are interested in the progression from the principle to the final test, given at the end of the section.

doubted. Cox also derived an expression for $V(T_0)$, the variance of T_0, and showed that for large samples, and under H_0,

$$N_0 \equiv T_0 / \left[V(T_0)\right]^{1/2} \sim N(0,1). \quad (3.3)$$

Pesaran (1974) was the first to apply Cox's results to econometrics and showed that for linear non-nested hypotheses

$$H_0: \quad y = X\beta_0 + u_0 , \qquad u_0 \sim N(0, \sigma_0^2 I)$$

$$H_1: \quad y = X\beta_1 + u_1 , \qquad u_1 \sim N(0, \sigma_1^2 I)$$

T_0 and V_0 are given by

$$T_0 = \frac{n}{2} \log\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_{10}^2}\right) \quad (3.4)$$

and

$$\hat{V}_0(T_0) = \frac{\hat{\sigma}_0^2}{\hat{\sigma}_{10}^4} \left[y' P_0 P_1 (I-P_0) P_1 P_0 y\right] \quad (3.5)$$

where

$$P_i = X_i (X_i' X_i)^{-1} X_i' \quad (i = 0,1) \quad (3.6)$$

and

$$\hat{\sigma}_{10}^2 = \hat{\sigma}_0^2 + n^{-1} (P_0 X)' (I-P_1)(P_0 X) . \quad (3.7)$$

It is instructive to note that each of the quantities in T_0 and V_0 can be obtained by appropriate regressions, as follows:

(i) $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$ are routinely obtained by regressing y on X_0 and y on X_1, respectively;

(ii) $\hat{\sigma}_{10}^2$ requires evaluation of $(P_0X)'(I-P_1)(P_0X)$. This is the error sum of squares when the prediction of y from H_0 ($P_0X \equiv \hat{y}_0$) is regressed on X_1; and

(iii) The expression $y'P_0P_1(I-P_0)P_1P_0y$ required to evaluate $\hat{V}_0(T_0)$ is the error sum of squares obtained when the predictions from the regression (ii) are regressed on X_0.

Thus, evaluation of N_0 requires four regressions: (a) y on X_0; (b) y on X_1; (c) the predictions of (a) on X_1; and (d) the predictions from (c) on X_0. Appropriate outputs from these regressions must then be combined according to (3.4), (3.7) and (3.5). This is a very cumbersome procedure, particularly if many models are being tested.

Fisher and McAleer (1981) replaced T_0 by an expression which is equivalent when H_0 is true and the sample large.

If H_0 is true, then at least for large samples, $\hat{\sigma}_1^2 \approx \hat{\sigma}_{10}^2$. Thus, we can write

$$T_0 = \frac{n}{2}\log\left(1 + \frac{\hat{\sigma}_1^2 - \hat{\sigma}_{10}^2}{\hat{\sigma}_{10}^2}\right)$$

and $(\hat{\sigma}_1^2 - \hat{\sigma}_{10}^2)/\hat{\sigma}_{10}^2$ will be small relative to 1. Applying a Taylor's expansion,

$$T_0 \approx TL_0 \equiv \frac{n}{2}\left(\frac{\hat{\sigma}_1^2 - \hat{\sigma}_{10}^2}{\hat{\sigma}_{10}^2}\right) , \qquad (3.8)$$

where TL_0 stands for a 'linearized version' of T_0. The variance of TL_0 is the same, in large samples, as that of T_0. Thus,

$$NL_0 \equiv TL_0/\left[V(T_0)\right]^{1/2} \sim N(0,1) . \qquad (3.9)$$

From (3.5),

$$NL_0 = \frac{(n/2)\left[\hat{\sigma}_1^2 - \hat{\sigma}_{10}^2\right]}{\hat{\sigma}_0\left[y'P_0P_1(I-P_0)P_1P_0y\right]^{1/2}} . \qquad (3.10)$$

Two points emerge from (3.10). The first is that it is clear from the numerator that NL_0 is assessing the significance of $\hat{\sigma}_1^2$ relative to $\hat{\sigma}_{10}^2$ -

another form of the Cox principle.

The second feature of (3.10) is more technical. Let us suppose a regression is performed of y on a <u>single</u> variable v,

$$y = v\gamma + u .$$

The t statistic of $\hat{\gamma}$ is given by

$$t(\hat{\gamma}) = \frac{\Sigma y_t v_t}{\hat{\sigma}[\Sigma v_t^2]^{1/2}} = \frac{y'v}{\hat{\sigma}[v'v]^{1/2}} .$$

If now we consider a regression

$$y = X_0\beta_0 + v\gamma + u ,$$

the above result is modified to become

$$t(\hat{\gamma}) = \frac{y'(I-P_0)v}{\hat{\sigma}[v'(I-P_0)v]^{1/2}} , \tag{3.11}$$

where P_0 is given by (3.6). Comparing the denominators of (3.10) and (3.11) we see that if we define

$$v = P_1P_0y , \tag{3.12}$$

then NL_0 has the same denominator as a t-statistic. Also, as NL_0 has a standard normal distribution in large samples, it seems very likely that NL_0, or an approximation to it, may in fact be a t-statistic from an artificial regression.

In fact, expressing $\hat{\sigma}_1^2$ and $\hat{\sigma}_{10}^2$ as

$$\hat{\sigma}_1^2 = n^{-1}\Big[y'(I-P_1)y\Big]$$
$$\hat{\sigma}_{10}^2 = n^{-1}\Big[y'(I-P_0)y + y'P_0(I-P_1)P_0y\Big]$$

some algebraic manipulation on the nume:·tor of (3.10) together with a further app oximation which has no effe.t when the sample is large, yields a statistic known as NA_0 which approximately equals NL_0 and is given by

$$NA_0 = \frac{-y'(I-P_0)P_1P_0y}{\hat{\sigma}_0(y'P_0P_1(I-P_0)P_1P_0y)^{1/2}} . \tag{3.13}$$

When we compare (3.13) with (3.11) it is apparent that $NA_0 = -t(\hat{\gamma})$ when the 'artificial regression'

$$y = X_0\beta_0 + v\gamma + u \tag{3.14}$$

is performed, and v is given by (3.12). Testing H_0 against H_1 is thus equivalent to testing $\gamma = 0$ in (3.14) by a standard t-test. The actual structure of the variable v should be made explicit. $P_0y = X_0(X_0'X_0)^{-1}X_0'y = X_0\hat{\beta}_0 = \hat{y}_0$, which is the OLS prediction of y obtained from H_0. Thus $v = P_1\hat{y}_0$. Applying the same argument again we can write

$$v = \hat{y}_{01} \tag{3.15}$$

where $\hat{y}_{01}$ means the prediction of $\hat{y}_0$ by the regressors of H_1. We now have a much more convenient variant of the Cox principle. Two regressions are required to construct $v = \hat{y}_{01}$. Then the artificial regression (3.14) yields a simple test of $\gamma = 0$. An advantage of this procedure over direct calculation of N_0 as given by (3.4), (3.7) and (3.5) is that only three regressions are needed instead of four. However, much more significant from a computational point of view is that the regression program automatically calculates the relevant t-statistic enabling the cumbersome computations involved in finding N_0 to be avoided.

We are now in a position to set out a simple, practical method for testing between two non-nested hypotheses. It is known as the JA-test.

(i) Regress y on X_0, obtaining $\hat{y}_0$;

(ii) Regress $\hat{y}_0$ on X_1 obtaining $\hat{y}_{01}$;

(iii) Regress y on X_0 and $\hat{y}_{01}$, performing a t-test on the coefficient of $\hat{y}_{01}$.

If this coefficient is significant, H_0 is rejected. H_0 and H_1 are now reversed and the procedure repeated. As mentioned earlier, four results on H_0 and H_1 are possible.

A very significant difference between the JA-test and the RESET test of model misspecification can now be appreciated. In the JA-test, the augmenting variable $\hat{y}_{01}$ incorporates information from the alternative hypothesis H_1 - it has been obtained by regressing $\hat{y}_0$ on the regressors of H_1. The RESET test, on the other hand, only uses some function $\hat{y}_0$ to augment H_0.

The possibility of obtaining a powerful test is now apparent in terms of the earlier discussion. First, as the model is augmented by the <u>single</u> variable $\hat{y}_{01}$, r = 1 - as small as it can be. Second, if H_1 reflects alternative economic theory, there is the possibility (perhaps probability ?) that the R^2 of (2.3) will be large.

Finally, it should be noted that if the hypotheses are non-linear, exactly the same procedure is valid, except that non-linear regression must be used.

4. <u>The Artificial Nesting Approach</u>

Consider again the two linear hypotheses

$$H_0: \quad y = X_0\beta_0 + u_0 \; ; \qquad E(u_0^2) = \sigma_0^2 \; ;$$
$$H_1: \quad y = X_1\beta_1 + u_1 \; ; \qquad E(u_1^2) = \sigma_1^2$$

where for ease of exposition, we assume X_0 and X_1 are linearly independent. An obvious way to test these hypotheses is to artificially nest them within a composite model H_c given by

$$H_c:\quad y = X_0\beta_0 + X_1\beta_1 + u\ . \tag{4.1}$$

Then $\beta_1 = 0$ implies H_0, and $\beta_0 = 0$ implies H_1. These tests can be carried out by standard F-tests. A problem, however, relates to the power of such tests. If either X_0 or X_1 (or both) have a large number of columns, then from our earlier discussion, the degrees of freedom in the numerator of the F-statistic are large, and power is consequently diminished. Furthermore, if the columns of X_0 and X_1 are highly collinear, as often happens, multicollinearity would further reduce power. It is therefore quite likely that tests based on H_c would reject neither H_0 nor H_1.

The model H_c above illustrates just one of an infinite number of ways that two hypotheses could be artificially nested in a composite model. Davidson and MacKinnon (1981) considered two non-linear hypotheses

$$H_0:\quad y = f(X_0,\ \beta_0) + u_0$$

$$H_1:\quad y = g(X_1,\ \beta_1) + u_1$$

and suggested nesting them in the model

$$H_c:\quad y = (1-\alpha)f(X_0,\beta_0) + \alpha g(X_1,\beta_1) + u\ . \tag{4.2}$$

If $\alpha = 0$, then H_0 is confirmed, while $\alpha = 1$ implies H_1. In principle then H_0 could be tested by testing $\alpha = 0$. Unfortunately the parameter α in (4.2) is unidentified. Davidson and MacKinnon suggested that a simple solution would be to replace $g(X_1,\beta_1)$ by its predicted value under H_1. Thus, the composite model becomes

$$H_c':\quad y = (1-\alpha)f(X_0,\beta_0) + \alpha g(X_1,\hat{\beta}_1) + u\ . \tag{4.3}$$

A test of $\alpha = 0$ is known as the J-test, and is a routine t-test. Using the notation of the last section

$$H_c':\quad y = f^*(X_0,\beta_0) + \hat{y}_1\alpha + u\ ,$$

showing the similarity and the difference between the J- and JA-tests. Both tests use an artificial regression, obtained by augmenting the null hypothesis by a single variable. Both tests are 'one degree of freedom' tests and hence likely to be more powerful than the direct composite approach (4.1). The difference lies in the augmenting variable. For the J-test, it is $\hat{y}_1$. Thus, the J-test only requires two regressions. An advantage of the JA-test is that when the hypotheses are linear and the disturbances normal the test is exact. That is, even for small samples it will produce tests of correct size. The J-test on the other hand is not exact, and in small samples tends to reject a true H_0 too often. This disadvantage of the J-test is counterbalanced by Monte Carlo evidence which suggests that it is usually a more powerful procedure than the JA-test.

We note that another test, known as the P-test, is also based on the composite model (4.3). Details can be found in Davidson and MacKinnon (1981). Godfrey and Pesaran (1983) have derived procedures for correcting the size of the J-test. However, at the moment there are no simple procedures which combine the size characteristics of the JA-test and the power of the J-test. Reasonable advice is to always calculate both J- and JA-tests.

An interesting synthesis of the Cox and artificial nesting approaches was given by Dastoor (1983). In this article Dastoor applied a variant of the Cox principle to non-nested linear models. Whereas Cox concentrated on the likelihood ratio, Dastoor derived a test based on the actual 'parameters of interest'. Considering (in our notation) the two linear hypotheses

$$H_0:\quad y = X_0\beta_0 + u_0 \qquad E(u_0^2) = \sigma_0^2 I$$

$$H_1:\quad y = X_1\beta_1 + u_1 \qquad E(u_1^2) = \sigma_1^2 I\ .$$

Dastoor proposed a test of the random vector

$$\hat{\eta} = \hat{\beta}_1 - \hat{\beta}_{10} \qquad (4.4)$$

where $\hat{\beta}_{10}$ is the expectation of $\hat{\beta}_1$ when H_0 is true. This is just Cox's principle applied to the β parameters. He showed that the significance of $\hat{\eta}$ can be assessed by testing $\delta = 0$ in the composite model

$$H_c: \quad y = X_0\beta_0 + \bar{X}_1\delta + u \ ,$$

where $\bar{X}_1$ consists of those columns of X_1 which are linearly independent of X_0. Thus the F-test of the composite model discussed earlier can also be interpreted as a variant of the Cox principle. Furthermore, Dastoor showed that the J- and JA-test statistics can be written in the form $a'\hat{\eta}_0$ and $b'\hat{\eta}_0$ where a' and b' are particular row vectors. This demonstrated that these tests also can be viewed as tests about the parameters of interest. As $a'\hat{\eta}_0$ and $b'\hat{\eta}_0$ are scalars, tests on their significance will clearly be 'one degree of freedom' tests in contrast to the composite model approach which has as many degrees of freedom as $\bar{X}$ has columns.

5. Testing Linear and Log-Linear Regression Models

Before we discuss a general method, due to Bera and McAleer (1989), of carrying out such tests, we will return to the artificial nesting procedure. We recall that to test

$$H_0: \quad y = f(X_0,\beta_0) + u_0$$
$$H_1: \quad y = g(X_1,\beta_1) + u_1$$

Davidson and MacKinnon (1981) proposed the composite model

$$H_c: \quad y = (1-\alpha)f(X_0,\beta_0) + \alpha g(X_1,\beta_1) + u \ . \tag{5.1}$$

and a test of H_0 is equivalent to testing $\alpha = 0$ in H_c.

We now rewrite (5.1) as

$$H_c: \quad (1-\alpha)\Big(y-f(X_0,\beta_0)\Big) + \alpha\Big(y - g(X_2,\beta_1)\Big) = u \ . \tag{5.2}$$

As y-f and y-g are u_0 and u_1, respectively, this formulation is equivalent to nesting the disturbances. If we now divide (5.2) by 1-α, we obtain

$$y = f(X_0,\beta_0) + \theta u_1 + v \quad (5.3)$$

where

$$\theta = -\alpha/(1-\alpha) .$$

Thus, testing $\alpha = 0$ is equivalent to testing $\theta = 0$ in (5.3). As u_1 is unobservable, we replace it by an observable proxy. The J-test is obtained by replacing u_1 by

$$\hat{u}_1 = y - g(X_1,\hat{\beta}_1) \quad (5.4)$$

On the other hand, the JA-test is obtained by replacing u_1 by

$$\hat{u}_{01} = \hat{y}_0 - g(X_1,\hat{\beta}_1^*) , \quad (5.5)$$

where $\hat{\beta}_1^*$ is obtained by regressing $\hat{y}_0$ on $g(X_1,\beta_1)$.

Once this alternative approach to the construction of J- and JA-tests has been appreciated, it is very easy to construct these tests for linear and log-linear models. Consider

$$H_0: \quad \log y = f(X,\beta_0) + u_0$$
$$H_1: \quad y = g(X,\beta_1) + u_1 .$$

By following the above nesting procedure, we obtain from (5.3) that the J-test of H_0 is a test of θ=0 in the artificial regression

$$\log y = f(X_0,\beta_0) + \theta\hat{u}_1 + v . \quad (5.6)$$

Similarly, the JA-test is obtained from the artificial regression

$$\log y = f(X_0,\beta_0) + \theta\hat{u}_{01} + v . \quad (5.7)$$

Construction of the variable $\hat{u}_{01}$ needs some care. First, log y is regressed on $f(X,\beta_0)$ to give predictions $\widehat{\log y}$. Then $\exp(\widehat{\log y}) \equiv \hat{y}_0$ is regressed on $g(X,\beta_1)$. The variable $\hat{u}_{01}$ is the residual from this regression.

If we now reverse H_0 and H_1 so that

$$H_0: \quad y = g(X_1,\beta_1) + u_1$$

$$H_1: \quad \log y = f(X_0,\beta_0) + u_0 ,$$

then the relevant artificial regression is

$$y = g(X_1,\beta_1) + \theta\tilde{u}_0^* + v ,$$

where $\tilde{u}_0^*$ is an appropriate proxy for u_0. For the J-test

$$\tilde{u}_0^* = \hat{u}_0 = \log y - f(X_0,\hat{\beta}_0) .$$

To construct $\tilde{u}_0^*$ for the JA-test, as usual two steps are required. First, y is regressed on $g(X_1,\beta_1)$, the regressand of H_0, yielding predictions $\hat{y}$. Second $\log(\hat{y})$ is regressed on $f(X_0,\beta_0)$, and $\tilde{u}^*$ is the residual vector from this second regression. Bera and McAleer cite Monte Carlo evidence which suggests that this testing procedure has better performance than other commonly used tests.

The approach outlined above can be easily adapted to any situation in which the dependent variable of one hypothesis is a monotonic function of the dependent variable of the other hypothesis.

6. Empirical Example

In a study of the consumption function, Guise (1989) used Australian data for the years 1947-48 to 1982-83 to quantify some well-known consumption theories. We use his results for two of these theories, namely, the Keynesian Absolute Income Hypothesis (AIH) and Duesenberry's Relative Income Hypothesis (RIH), to illustrate the techniques which have been discussed. This example is essentially reproduced from Doran (1989).

Guise's regression results (with standard errors in parentheses) were as follows:

AIH

$$\hat{C}_t = .005 + .88Y_t - 1.07I_t \;;$$
$$\quad (.09) \quad (.02) \quad (.34)$$

$$\bar{R}^2 = 0.983; \quad D.W. = 2.18$$

where C_t = private consumption expenditure,

Y_t = household disposable income,

$I_t = (P_t - P_{t-1})/P_{t-1}$ is a measure of inflation and P_t is a price deflator.

RIH

$$\hat{C}_t = -.030 + .90Y_{2t} + .22Z_{1t} - .93I_t \;;$$
$$\quad (.05) \quad (.01) \quad (.09) \quad (.20)$$

$$\bar{R}^2 = 0.994; \quad D.W. = 1.98$$

where C_t and I_t are as above.

In specifying this relationsh:,, Guise defined the following additional variables

Y_t^* = highest income prior to year t;

$Z_{1t} = (Y_t - Y_t^*)$ when $Y_t > Y_t^*$, zero otherwise;

$Z_{2t} = (Y_t - Y_t^*)$ when $Y_t < Y_t^*$, zero otherwise;

$Y_{2t} = Y_t^* + Z_{2t}$.

The important thing to notice about both models is that they perform very well on the usual checks; $\bar{R}^2$'s are high, Durbin-Watson statistics are close to 2, slope coefficients are highly significant with 'correct' signs and the marginal propensities to consume are estimated at 0.90 and 0.88.

The two models are non-nested and we now apply the J, JA and F tests as follows:

(i) Null Hypothesis is AIH

J-test: regress C on Y_2, Z_1 and I to obtain predictions $\hat{C}_R$; regress C on Y, I and $\hat{C}_R$; the relevant statistic is the t ratio of the coefficient of $\hat{C}_R$.

JA-test: regress C on Y and I to obtain predictions $\hat{C}_A$; regress $\hat{C}_A$ on Y_2, Z_1 and I to obtain predictions $\hat{C}_{AR}$; regress C on Y, I and $\hat{C}_{AR}$; the t ratio of the coefficient of $\hat{C}_{AR}$ is the relevant statistic.

F-test: regress C on Y, I, Y_2, Z_1 and examine the joint significance of the coefficients of Y_2 and Z_1.

(ii) Null Hypothesis is RIH

J-test: regress C on Y and I to obtain predictions $\hat{C}_A$; regress C on Y_2, Z_1, I and $\hat{C}_A$; test the significance of the coefficient of $\hat{C}_A$.

JA-test: regress C on Y_2, Z_1 and I to obtain $\hat{C}_R$; regress $\hat{C}_R$ on Y and I to obtain $\hat{C}_{RA}$; regress C on Y_2, Z_1, I and $\hat{C}_{RA}$ and examine the significance of the coefficient of $\hat{C}_{RA}$.

F-test: regress C on Y_2, Z_1, I and Y; examine significance of the coefficient of Y.

The results of these tests are shown in Table 1.

TABLE 1

H_0	Test		
	J	J_A	F
AIH	6.22**	0.59	19.54**
RIH	-0.87	-0.87	0.76

** significant at the 1 per cent level.

On the basis of the J- and F-tests, the Absolute Income Hypothesis model is inadequate, and should be rejected. On the other hand, the tests do not indicate inadequacy in the Relative Income Hypothesis. One final note. If the alternative hypothesis contains only one variable which is not also in the null hypothesis then the J-, JA- and F-tests are all equivalent. This is the case when RIH is the null. Hence J = JA = -0.87 and F = 0.76 = $(-0.87)^2$

7. Concluding Remarks

An attempt has been made to present some of the main results in testing non-nested models in a fairly non-technical way. The essential message is twofold. First, when non-nested models are being compared, the traditional discrimination techniques (eg. comparing R^2's) should be accompanied by appropriate hypothesis tests. Second, and very significant from the point of view of an applied researcher, such tests are easy to apply and can be constructed with all standard regression packages.

The purpose of this article is not to break new ground, but to make a small contribution in translating well-established econometric theory into applied economic practice.

References

Bera, A.K. and M. McAleer (1989), 'Nested and Non-Nested Procedures for Testing Linear and Log-Linear Regression Models', Sankhya B, 51, part 2, 212-224.

Cox, D.R. (1962), 'Further Results on Tests of Separate Families of Hypotheses', Journal of the Royal Statistical Society B, 24, 406-424.

Dastoor, N.K. (1983), 'Some Aspects of Testing Non-Nested Hypotheses', Journal of Econometrics 21(2), 213-228.

Davidson, R. and J.G. MacKinnon (1981), 'Several Tests for Model Specification in the Presence of Alternative Hypotheses', Econometrica 49, 781-793.

——————— (1985), 'The Interpretation of Test Statistics', Canadian Journal of Economics 18(1), 38-57.

Doran, H.E. (1989), Applied Regression Analysis in Econometrics, New York: Marcel Dekker.

Fisher, G.R. and M. McAleer (1981), 'Alternative Procedures and Associated Tests of Significance for Non-Nested Hypotheses', Journal of Econometrics 16, 103-119.

Godfrey, L.G. and M.H. Pesaran (1983), 'Tests of Non-Nested Regression Models: Small Sample Adjustments and Monte Carlo Evidence', Journal of Econometrics 21, 133-154.

Guise, J.W.B. (1989), 'Modeling and Estimation of Economic Relationships: An Example' in H.E. Doran, Applied Regression Analysis in Econometrics, New York: Marcel Dekker.

MacKinnon, J.G. (1983), 'Model Specification Tests Against Non-Nested Alternatives', Econometric Reviews 2, 85-157.

——————— (1991), 'Model Specification Tests and Artificial Regressions', Journal of Economic Literature (forthcoming).

McAleer, M.J. (1987), 'Specification Tests for Separate Models: A Survey' in M.L. King and D.E.A. Giles (eds), Specification Analysis in the Linear Model, London: Routledge and Kegan Paul.

Pagan, A.R. and A.D. Hall (1983), 'Diagnostic Tests as Residual Analysis', Econometric Reviews 2(2), 159-218.

Pesaran, M.H. (1974), 'On the General Problem of Model Selection', Review of Economic Studies 41, 153-171.

Shideed, K.H. and F.C. White (1989), 'Alternative Forms of Price Expectations in Supply Analysis for U.S. Corn and Soybean Acreages', Western Journal of Agricultural Economics 14(2), 281-292.