$$Q_{ED}$$

Model Specification Tests Against Non-Nested Alternatives

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

12-1983

Model Specification Tests
Against Non-Nested Alternatives

James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

Abstract

Non-nested hypothesis tests provide a way to test the specification of an econometric model against the evidence provided by one or more non-nested alternatives. This paper surveys the recent literature on non-nested hypothesis testing in the context of regression and related models. Much of the purely statistical literature which has evolved from the fundamental work of Cox (1961, 1962) is discussed briefly or not at all. Instead, emphasis is placed on those techniques which are easy to employ in practice and are likely to be useful to applied workers.

1. Introduction

Every careful applied econometrician has had the experience of estimating a regression model which seemed, at first glance, to be highly satisfactory, but which subsequently turned out, on closer investigation, to be false and misleading. The nature of economic data makes this inevitable. Right-hand side variables are very often collinear, and time-series data usually exhibit trends and respond to the same business cycles, so that even nonsensical models will often fit well and yield apparently significant parameter estimates. In this circumstance, it is usual to appeal to economic theory for guidance in specifying econometric models. But theory rarely suggests only one model to explain a particular phenomenon. Often there are rival theories, and even when there are not, a single theory may be compatible with several functional forms and/or stochastic specifications.

Faced with several rival regression models, what is the applied econometrician to do? If he or she is convinced that the set of available models include the truth, then it is simply necessary to choose the model which is "best". For this, the econometrician would use some sort of **model selection criterion**. More realistically, however, and especially when the investigation is in its early stages, the econometrician may not know whether any of the available models could conceivably be true. The first order of business, then, is to test the specification of each of the available models. Tests for heteroskedasticity, serial correlation, parameter stability, and so on obviously have a large role to play in this context. But such tests do not make use of the information that the model being tested is only one of several models to explain the same data. After all, if model H_0 is true, then it follows that any non-nested alternative model, say H_1, must be false, and this proposition can serve as the basis for a test of H_0. Such tests are commonly referred to as **non-nested hypothesis tests**.

At this point, it is perhaps appropriate to define the term "nonnested hypothesis". All econometricians are familiar with nested hypotheses. An econometric model, H_0, is said to be nested within an alternative model, H_1, if H_1 can be reduced to H_0 by imposing one or more restrictions on its parameters. As a classic example, the Cobb-Douglas production function is nested within the C.E.S. production function, because when the elasticity of substitution is unity, the latter reduces to the former. Thus two models, H_0 and H_1, may be said to be **non-nested** if H_0 is not nested within H_1 and H_1 is not nested within H_0. Such models are also sometimes referred to as **separate**. If one is to use a non-nested hypothesis test to test the specification of a model, there must exist at least one non-nested alternative model. Such an alternative should normally have been suggested by economic theory, but it need not be a model which the investigator would take seriously.

It is customary in the econometric literature to treat model selection criteria and non-nested hypothesis tests as being very closely related, perhaps even as rival procedures. That is a very misleading point of view. Model selection criteria are appropriate when one wishes to choose one out of a group of rival models. The objective is to choose the "best" model, explicitly trading off goodness of fit and parsimony in parametrization.

Either one assumes in advance that one of the models must be "true", or one simply does not care whether any of the models is satisfactory.

Non-nested hypothesis tests, on the other hand, are tests of model specification, just like tests for serial correlation or omitted variables. The only real difference between them and more classical procedures is that they rely on the existence of non-nested alternative models. The results of applying such tests to, for simplicity, two rival models may be that one model can be rejected while the other cannot; but it may just as well be that both models, or neither model, can be rejected. Thus testing each model against the evidence provided by the other cannot, in general, allow the investigator to choose one of the two. What it can do is to provide evidence that one of the models, or perhaps both models, are misspecified.

As the preceding discussion implies, the term "non-nested hypothesis test" is perhaps an unfortunate one. Indeed, many econometricians would regard it as a contradiction in terms, since, in the classical theory, testing a hypothesis is synonymous with testing restrictions on a general model. It would be more accurate to refer to these procedures as "model specification tests using the evidence provided by non-nested alternative hypotheses". As the remainder of this paper will, I hope, make clear, these tests are in fact perfectly classical ones. I will, however, continue to use the term "non-nested hypothesis test" because it is well established in the literature and reasonably convenient.

In this paper, I survey the recent literature on non-nested hypothesis testing in the context of regression models. The emphasis throughout will be on techniques that are likely to prove useful in practice to applied econometricians. The literature on model selection criteria will not be discussed at all. Readers who are interested in this rapidly growing field might wish to look at Gaver and Geisel (1974), Sawa (1978), Leamer (1979), Amemiya (1980), Sawyer (1980), or Chapter 11 of Judge *et al.* (1980).

2. The Cox Test

All of the theoretical literature on non-nested hypothesis testing derives, to a greater or lesser extent, from two fundamental papers by Cox (1961, 1962). One of the principal achievements of these papers is the development of a very general, but not always easily implemented, procedure which has since come to be known as the Cox test. The basic idea of this procedure is that one may test the validity of a null hypothesis, H_0, about how a set of data was generated by comparing the observed ratio of the values of the likelihood functions for H_0 and for some non-nested alternative hypothesis, H_1, with an estimate of the expected value of this likelihood ratio if H_0 were true. If H_1 fits either better or worse than it should if H_0 were true, then H_0 must be false.

The numerator of the Cox test statistic may be written as

$$T_0 = \mathcal{L}(\hat{\boldsymbol{\theta}}_0) - \mathcal{L}(\hat{\boldsymbol{\theta}}_1) - T\Big(\plim_{n\to\infty}{}_0 \frac{1}{T}\big(\mathcal{L}(\hat{\boldsymbol{\theta}}_0) - \mathcal{L}(\hat{\boldsymbol{\theta}}_1)\big)\Big)_{\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0}. \tag{1}$$

Here $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are the vectors of parameters under H_0 and H_1, respectively, T is the number of observations, and $\mathcal{L}(\hat{\boldsymbol{\theta}}_k)$, for $k = 0, 1$, denotes the value of the loglikelihood

function for H_k evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\theta}}_k$. The notation "$\text{plim}_0$" denotes the probability limit under H_0. Since this limit will in general depend on $\boldsymbol{\theta}_0$, the notation "$\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0$" indicates that $\boldsymbol{\theta}_0$ is to be evaluated at its maximum likelihood estimate $\hat{\boldsymbol{\theta}}_0$. Thus the first two terms in (1) are simply the log of the likelihood ratio for the two models, while the third term is an estimate of the expectation of the log of that ratio, assuming that H_0 is true. Cox demonstrates that T_0 is asymptotically normally distributed, with mean zero and variance that can be consistently estimated according to a formula he provides. Thus an estimate of T_0, divided by the square root of this consistent estimate of its variance, will be asymptotically $N(0,1)$. A rigorous discussion of this proposition for the i.i.d. case may be found in White (1982b).

The major difficulty with applying the Cox test to problems in econometrics is to find a way to estimate the third term in (1). Walker (1967) applied the test to time-series models, and Amemiya (1973) used it to test alternative specifications of the distribution of the error term in regression models, but it seems fair to say that the apparent difficulties of employing the Cox test discouraged most econometricians from paying much attention to it. Indeed, Dhrymes *et al.* (1972) actually implied that the Cox test was impractical in most econometric applications. The test first attracted substantial interest among econometricians when Pesaran (1974) showed how it could be applied to the case of non-nested univariate linear regression models. Subsequently, Pesaran and Deaton (1978) extended Pesaran's derivation to the case of nonlinear and possibly multivariate regression models. For simplicity, I will restrict attention for the moment to the nonlinear univariate case of Pesaran and Deaton.

The model to be tested is

$$H_0: \quad \boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}) + \boldsymbol{u}_0, \quad \boldsymbol{u}_0 \sim \text{NID}(0, \sigma_0^2 \mathbf{I}), \tag{2}$$

where $\boldsymbol{y}$ is a vector of observations on a dependent variable, $\boldsymbol{f}(\boldsymbol{\beta})$ is a vector of functions $f_t(\boldsymbol{\beta})$ which depend in a suitably continuous manner on an unknown parameter vector $\boldsymbol{\beta}$ and on exogenous and/or lagged dependent variables, and $\boldsymbol{u}_0$ is a white noise error vector if H_0 is true. The alternative model is

$$H_1: \quad \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{\gamma}) + \boldsymbol{u}_1, \quad \boldsymbol{u}_1 \sim \text{NID}(0, \sigma_1^2 \mathbf{I}), \tag{3}$$

where the notation is analogous to that for H_0 and once again the error vector is only assumed to follow the normal law if H_1 happens to be true. The models H_0 and H_1 are assumed to be nonnested.

The obvious estimator of

$$T \plim_{\substack{0 \\ n \to \infty}} \frac{1}{T} \mathcal{L}(\hat{\boldsymbol{\theta}}_0),$$

evaluated at $\hat{\boldsymbol{\theta}}_0$, is simply $\mathcal{L}(\hat{\boldsymbol{\theta}}_0)$. Thus (1) reduces to

$$T_0 = -\mathcal{L}(\hat{\boldsymbol{\theta}}_1) + T \left(\plim_{\substack{0 \\ n \to \infty}} \frac{1}{T} \mathcal{L}(\hat{\boldsymbol{\theta}}_1) \right)_{\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_0}. \tag{4}$$

Except for an additive constant, the (concentrated) loglikelihood function for H_1 is

$$\mathcal{L}(\hat{\boldsymbol{\theta}}_1) = -\frac{T}{2} \log \hat{\sigma}_1^2, \tag{5}$$

where

$$\hat{\sigma}_1^2 \equiv \frac{1}{T}(\boldsymbol{y} - \hat{\boldsymbol{g}})^\top (\boldsymbol{y} - \hat{\boldsymbol{g}}). \tag{6}$$

Here $\hat{\boldsymbol{g}}$ denotes $\boldsymbol{g}(\hat{\boldsymbol{\gamma}})$, that is, the fitted values under H_1 evaluated at the maximum likelihood estimates $\hat{\boldsymbol{\gamma}}$.

Since $\mathcal{L}(\hat{\boldsymbol{\theta}}_1)$ depends only on $\hat{\sigma}_1^2$, it is merely necessary to find the plim of $\hat{\sigma}_1^2$ under H_0 in order to calculate the second term in (4). Let us denote by $\tilde{\boldsymbol{\gamma}}$ an estimate of $\boldsymbol{\gamma}$ under H_0 which tends in the limit to the plim of $\hat{\boldsymbol{\gamma}}$ under H_0, and let $\tilde{\boldsymbol{g}} \equiv \boldsymbol{g}(\tilde{\boldsymbol{\gamma}})$. There will in general be many choices for $\tilde{\boldsymbol{\gamma}}$; see below. Then, by definition,

$$\tilde{\sigma}_1^2 \equiv \frac{1}{T}(\boldsymbol{y} - \tilde{\boldsymbol{g}})^\top (\boldsymbol{y} - \tilde{\boldsymbol{g}}) \tag{7}$$

$$= \frac{1}{T}(\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{f} - \tilde{\boldsymbol{g}})^\top (\boldsymbol{y} - \boldsymbol{f} + \boldsymbol{f} - \tilde{\boldsymbol{g}}) \tag{8}$$

$$= \frac{1}{T}\left((\boldsymbol{y} - \boldsymbol{f})^\top (\boldsymbol{y} - \boldsymbol{f}) + (\boldsymbol{f} - \tilde{\boldsymbol{g}})^\top (\boldsymbol{f} - \tilde{\boldsymbol{g}}) + 2(\boldsymbol{y} - \boldsymbol{f})^\top (\boldsymbol{f} - \tilde{\boldsymbol{g}})\right), \tag{9}$$

here $\boldsymbol{f}$ denotes $\boldsymbol{f}(\boldsymbol{\beta})$, that is, the fitted values under H_0 evaluated at the true parameter vector $\boldsymbol{\beta}$. Under H_0, the vector $\boldsymbol{y} - \boldsymbol{f}$ is simply a vector of normal random variates with mean zero and variance σ_0^2, which must certainly be asymptotically independent of the vector $\boldsymbol{f} - \tilde{\boldsymbol{g}}$, because $\boldsymbol{f} - \tilde{\boldsymbol{g}}$ is asymptotically nonstochastic. Thus an estimate of the plim of $\hat{\sigma}_1^2$ is

$$\tilde{\sigma}_{10}^2 = \hat{\sigma}_0^2 + \frac{1}{T}(\hat{\boldsymbol{f}} - \tilde{\boldsymbol{g}})^\top (\hat{\boldsymbol{f}} - \tilde{\boldsymbol{g}}). \tag{10}$$

This estimate has two parts: an estimate of the variance of the true model, $\hat{\sigma}_0^2$, and an estimate of the additional variance due to the difference between H_0 and H_1.

Combining (4), (5), and (10), we find the numerator of the Cox test statistic is simply

$$T_0 = \frac{T}{2} \log \frac{\hat{\sigma}_1^2}{\tilde{\sigma}_{10}^2}. \tag{11}$$

If the alternative model H_1 fits better than it should, this statistic will be negative, and if H_1 fits worse than it should, this statistic will be positive.

The estimate $\tilde{\boldsymbol{\gamma}}$ is obtained by Pesaran and Deaton (1978) from the nonlinear regression

$$\hat{\boldsymbol{f}} = \boldsymbol{g}(\boldsymbol{\gamma}) + \boldsymbol{e}, \tag{12}$$

that is, the model H_1 re-estimated using $\boldsymbol{f}$ as the dependent variable. An estimate of the variance of T_0 is obtained from the formula

$$\hat{V}_0(T_0) = \frac{\hat{\sigma}_0^2}{\tilde{\sigma}_{10}^4}(\hat{\boldsymbol{f}} - \tilde{\boldsymbol{g}})^\top (\mathbf{I} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^\top \hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}^\top)(\hat{\boldsymbol{f}} - \tilde{\boldsymbol{g}}), \tag{13}$$

where $\hat{\boldsymbol{F}}$ is the matrix of derivatives of $\boldsymbol{f}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, evaluated at $\hat{\boldsymbol{\beta}}$. This formula is derived by applying a general result of Cox (1962). Thus the Cox test statistic, as implemented for the nonlinear regression problem by Pesaran and Deaton (1978), is just

$$N_0 = \frac{T_0}{\hat{V}_0^{1/2}(T_0)}.\tag{14}$$

It should be stressed at this point that N_0 provides a test of H_0 and tells us nothing whatsoever about the validity of H_1. If N_0 is significantly less than zero, we may conclude that H_0 is rejected in the direction of H_1, and if it is significantly greater than zero, we may conclude that H_0 is rejected in a direction away from H_1. But we can never conclude from N_0 alone that H_0 is rejected in favor of H_1, as some early authors have misleadingly asserted (see, e.g., Pesaran (1974, p. 158)). If we wish to test the validity of H_1, we must reverse the roles of the two models and compute the test statistic N_1. On the other hand, it is true that the Cox test will reject H_0 with probability one asymptotically if H_1 is true; thus the test may be said to be consistent. For a general proof, see Pereira (1977a).

Unfortunately, the implementation of the Cox test by Pesaran (1974) and Pesaran and Deaton (1978) is unnecessarily complicated. The first thing to observe is that, under H_0, the plim of $\tilde{\boldsymbol{\gamma}}$ is equal to the plim of $\hat{\boldsymbol{\gamma}}$. Thus the artificial nonlinear regression (12) is unnecessary, and it is possible to replace $\tilde{\boldsymbol{g}}$ by $\hat{\boldsymbol{g}}$ in (7), (8), (9), (10), and (13). The resulting test statistic will also be a valid Cox test, asymptotically the same as N_0 under H_0 but generally different under H_1 and in small samples.

It is straightforward to show that $\hat{\boldsymbol{g}}$ and $\tilde{\boldsymbol{g}}$ tend to the same probability limit under H_0. The normal equations which define $\hat{\boldsymbol{g}}$ are

$$\frac{1}{T}\big(\boldsymbol{y} - \boldsymbol{g}(\hat{\boldsymbol{\gamma}})\big)^{\top}\boldsymbol{G}(\hat{\boldsymbol{\gamma}}) = \boldsymbol{0},\tag{15}$$

where $\boldsymbol{G}(\hat{\boldsymbol{\gamma}})$ denotes the matrix with typical element

$$G_{ij} = \frac{\partial g_i(\boldsymbol{\gamma})}{\partial \gamma_j}$$

evaluated at $\hat{\boldsymbol{\gamma}}$. Under H_0, $\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{u}_0$, so that (15) may be rewritten as

$$\frac{1}{T}\big(\boldsymbol{f} - \boldsymbol{g}(\hat{\boldsymbol{\gamma}})\big)^{\top}\boldsymbol{G}(\hat{\boldsymbol{\gamma}}) + \frac{1}{T}\boldsymbol{u}_0^{\top}\boldsymbol{G}(\hat{\boldsymbol{\gamma}}) = \boldsymbol{0}.\tag{16}$$

The second term of (16) clearly tends to zero as T tends to infinity. Thus, in the limit, $\hat{\boldsymbol{\gamma}}$ is defined by

$$\frac{1}{T}\big(\boldsymbol{f} - \boldsymbol{g}(\hat{\boldsymbol{\gamma}})\big)^{\top}\boldsymbol{G}(\hat{\boldsymbol{\gamma}}) = \boldsymbol{0}.\tag{17}$$

Similarly, the normal equations which define $\tilde{\boldsymbol{\gamma}}$ are

$$\frac{1}{T}\big(\hat{\boldsymbol{f}} - \boldsymbol{g}(\tilde{\boldsymbol{\gamma}})\big)^{\top}\boldsymbol{G}(\tilde{\boldsymbol{\gamma}}) = \boldsymbol{0}.\tag{18}$$

The only difference between (17) and (18) is that $\boldsymbol{f}$ appears in the former and $\hat{\boldsymbol{f}}$ in the latter. But the consistency of $\hat{\boldsymbol{\beta}}$ ensures that $\hat{\boldsymbol{f}}$ tends to $\boldsymbol{f}$ asymptotically. Thus $\hat{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}}$ must tend to the same probability limit.

Now observe that

$$\hat{\sigma}_1^2 = \frac{1}{T}\big((\boldsymbol{y} - \hat{\boldsymbol{f}})^\top(\boldsymbol{y} - \hat{\boldsymbol{f}}) + (\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}})^\top(\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}}) + 2(\boldsymbol{y} - \hat{\boldsymbol{f}})^\top(\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}})\big) \tag{19}$$

while a valid alternative to $\tilde{\sigma}_{10}^2$ is

$$\hat{\sigma}_{10}^2 = \frac{1}{T}\big((\boldsymbol{y} - \hat{\boldsymbol{f}})^\top(\boldsymbol{y} - \hat{\boldsymbol{f}}) + (\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}})^\top(\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}}). \tag{20}$$

If we formulate a Cox test using $\hat{\boldsymbol{\gamma}}$ instead of $\tilde{\boldsymbol{\gamma}}$, it simply tests the proposition that the ratio of (19) to (20) is unity. But the only difference between (19) and (20) is that in the latter the term $(2/T)(\boldsymbol{y} - \hat{\boldsymbol{f}})^\top(\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}}))$ has been replaced by its probability limit, which is zero. Thus what the Cox test is really testing, in this case, is whether $\boldsymbol{y} - \hat{\boldsymbol{f}}$ is asymptotically uncorrelated with $\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}}$. That is, of course, a very natural thing to test. If H_0 is a true model, its residuals should be uncorrelated with the difference between the fitted values from H_0 and H_1. But instead of doing so via the Cox test, why not simply test this proposition directly? That is precisely what the tests to be discussed in the next section do.

3. Tests Based on Artificial Regressions

The simplest way to test the proposition that $\boldsymbol{y} - \hat{\boldsymbol{f}}$ is orthogonal to $\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}}$ would seem to be to run the artificial linear regression

$$\boldsymbol{y} - \hat{\boldsymbol{f}} = \alpha(\hat{\boldsymbol{f}} - \hat{\boldsymbol{g}}) + \boldsymbol{u} \tag{21}$$

or, equivalently,

$$\boldsymbol{y} = (1 - \alpha)\hat{\boldsymbol{f}} + \alpha\hat{\boldsymbol{g}} + \boldsymbol{u}. \tag{22}$$

Davidson and MacKinnon (1981) call this the C test, because (21) is *conditional* on $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$. They find, however, that the t statistic on α is not asymptotically distributed as $N(0,1)$; rather, it is asymptotically normal with variance less than unity. It is not difficult to obtain a valid estimate of the variance of α, and so make use of a corrected C test, but that turns out to be an unnecessarily complicated way to proceed.

Instead, it is most fruitful to consider a comprehensive model analogous to (22):

$$H_C: \quad \boldsymbol{y} = (1 - \alpha)\boldsymbol{f}(\boldsymbol{\beta}) + \alpha\boldsymbol{g}(\boldsymbol{\gamma}) + \boldsymbol{u}. \tag{23}$$

This is an artificial model, in which H_0 and H_1 are combined with weights $1 - \alpha$ and α, respectively. As it stands, this model is perfectly useless, since one normally cannot hope to estimate α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ jointly. To make (23) useful for testing H_0, one has to

recognize the crucial fact that, under H_0, $\boldsymbol{\gamma}$ may validly be replaced by $\hat{\boldsymbol{\gamma}}$. Thus the compound model (23) becomes

$$\boldsymbol{y} = (1 - \alpha)\boldsymbol{f}(\boldsymbol{\beta}) + \alpha\hat{\boldsymbol{g}} + \boldsymbol{u}. \tag{24}$$

Davidson and MacKinnon (1981) prove that, under H_0, the vector $\hat{\boldsymbol{g}}$ converges to a nonstochastic plim, so that $\hat{\boldsymbol{g}}$ may validly be used as a right-hand side variable in (24), and the t statistic on $\hat{\alpha}$ is asymptotically $N(0,1)$. They call this procedure the J test, because α and $\boldsymbol{\beta}$ are estimated *jointly*.

If H_0 is a nonlinear model, then (24) is also a nonlinear regression, and one which may be computationally intractable if H_0 and H_1 are very similar. There is, however, a simple way to overcome this problem. We simply linearize equation (24) around the point $\alpha = 0$, $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, so as to obtain the linear regression

$$\boldsymbol{y} - \hat{\boldsymbol{f}} = \hat{\boldsymbol{F}}\boldsymbol{b} + \alpha(\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}}) + \boldsymbol{u}, \tag{25}$$

where $\hat{\boldsymbol{F}}$ is the matrix of derivatives of $\boldsymbol{f}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, evaluated at $\hat{\boldsymbol{\beta}}$. It is a consequence of standard results for maximum likelihood and/or least squares estimation, together with the previous result that $\hat{\boldsymbol{g}}$ is asymptotically non-stochastic, that the t statistic on $\hat{\alpha}$ from regression (25) is asymptotically $N(0,1)$ when H_0 is true. See in particular Durbin (1970). Results in that paper can also be used to show that the C test t statistic based on (22) is asymptotically normal with mean zero and variance less than unity. Thus testing H_0 against the evidence provided by H_1 merely requires one artificial linear regression, a procedure that was called the P test by Davidson and MacKinnon (1981).

The t statistic on $\hat{\alpha}$ from equation (25) is

$$\frac{(\boldsymbol{y} - \hat{\boldsymbol{f}})^\top \boldsymbol{M}_{\hat{\boldsymbol{F}}}(\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}})}{\hat{\sigma}\big((\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}})^\top \boldsymbol{M}_{\hat{\boldsymbol{F}}}(\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}})\big)^{1/2}}, \tag{26}$$

where

$$\boldsymbol{M}_{\hat{\boldsymbol{F}}} \equiv \boldsymbol{I} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^\top\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}^\top, \tag{27}$$

and $\hat{\sigma}$ is the estimate of the standard error from (25). The first-order conditions which determine $\hat{\boldsymbol{\beta}}$ imply that $(\boldsymbol{y} - \hat{\boldsymbol{f}})^\top\hat{\boldsymbol{F}} = \boldsymbol{0}$. Thus the numerator of the test statistic (26) is simply $(\boldsymbol{y} - \hat{\boldsymbol{f}})^\top(\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}})$. The denominator, of course, is an asymptotically valid estimate of the standard error of the numerator. Hence what the P test does is to provide a very simple way of testing what we set out to test in the first place, namely, the orthogonality of $\boldsymbol{y} - \hat{\boldsymbol{f}}$ and $\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}}$. In view of this, it is not surprising that the P test statistic is asymptotically perfectly correlated with the Cox test statistic under H_0, with correlation coefficient minus one. The P test statistic is also asymptotically equal, under H_0, to both the J test statistic and the C test statistic when the latter uses a valid variance estimate.

It is worth discussing briefly the case of linear regression models. Suppose that the two models of interest are

$$\mathrm{H}_0 \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{\delta}_0 + \boldsymbol{u}_0 \tag{28}$$

and

$$\mathrm{H}_1 \quad \boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{W}\boldsymbol{\delta}_1 + \boldsymbol{u}_1, \tag{29}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are matrices of regressors unique to H_0 and H_1, respectively, and $\boldsymbol{W}$ is a matrix of regressors common to the two models. The J test amounts to estimating the linear regression

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{W}\boldsymbol{\delta}_0^* + \alpha(\boldsymbol{Z}\hat{\boldsymbol{\gamma}} + \boldsymbol{W}\hat{\boldsymbol{\delta}}_1) + \boldsymbol{u}, \tag{30}$$

and the t statistic on $\hat{\alpha}$ is numerically equal to the t statistic from the P test regression in this case.

A more obvious way to test H_0 would be to form the compound model

$$\mathrm{H}_C: \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{W}\boldsymbol{\delta} + \boldsymbol{u} \tag{31}$$

and then test the hypothesis that $\boldsymbol{\gamma} = \boldsymbol{0}$ using an F test. If $\boldsymbol{Z}$ has only one column, this procedure will be numerically identical to the J test. However, if $\boldsymbol{Z}$ has more than one column, the two procedures will not be the same. The F test will involve as many degrees of freedom as there are columns in $\boldsymbol{Z}$, while the J test will involve only one degree of freedom. This suggests that the latter may have more power than the former, and that is indeed the case. Pesaran (1982) has recently shown that, for local alternatives, the F statistic (in chi-squared form) and the square of the J test statistic are both asymptotically distributed as non-central chi-squared, with the same non-centrality parameter but different numbers of degrees of freedom. Since the critical value for the F test will be larger, because there are more degrees of freedom, the J test must have higher power against H_1. On the other hand, if both H_0 and H_1 are false, the non-centrality parameters would no longer be the same, and the F test might well have more power than the J test, although that is perhaps unlikely to be the case if $\boldsymbol{X}$ and $\boldsymbol{Z}$ are highly collinear.

At this point, the possibility of a somewhat pathological special case should be mentioned. Suppose that $\boldsymbol{X}$ and $\boldsymbol{Z}$ are orthogonal to each other after they have been projected off $\boldsymbol{W}$. This implies that (29) and (31) will yield the same estimates of $\boldsymbol{\gamma}$. It follows that the sums of squared residuals from (30) and (31) must be identical, so that a one degree of freedom test will not be valid, even asymptotically, unless $\boldsymbol{Z}$ has only one column. In this situation, incidentally, the second factor in (13) will be identically zero, so that the Cox test as implemented by Pesaran and Deaton (1978) will be undefined. Thus non-nested hypothesis tests are not valid in the case of orthogonal models, and all authors have to make regularity assumptions which rule out this case; see, for example, Pesaran (1974, p. 156). The F test, of course, remains both valid and powerful in this case.

The P test can easily be extended to handle several alternative hypotheses. Let the null hypothesis still be H_0, given by (2), and the alternative hypotheses be

$$H_j: \quad \boldsymbol{y} = \boldsymbol{g}_j(\boldsymbol{\gamma}_j) + \boldsymbol{u}_j, \quad j = 1, \ldots, J. \tag{32}$$

Then the obvious compound model is

$$H_C: \quad \Big(1 - \sum_{j=1}^{J} \alpha_j\Big) \boldsymbol{f}(\boldsymbol{\beta}) + \sum_{j=1}^{J} \alpha_j \boldsymbol{g}(\boldsymbol{\gamma}_j) + \boldsymbol{u}, \tag{33}$$

and the corresponding P test regression is

$$\boldsymbol{y} - \hat{\boldsymbol{f}} = \hat{\boldsymbol{F}}\boldsymbol{b} + \sum_{j=1}^{J} \alpha_j(\hat{\boldsymbol{g}}_j - \hat{\boldsymbol{f}}) + \boldsymbol{u}. \tag{34}$$

The appropriate test statistic is then an asymptotic F test of the hypothesis that all the α_j are zero, or some asymptotically equivalent test.

Whether one would want to test a model against several alternative hypotheses simultaneously, in practice, is not entirely clear. If one of the alternatives is true, then highest power will be surely be achieved by testing against that hypothesis alone. On the other hand, if none of the alternatives is true, testing against several of them jointly may have higher power than testing against each of them individually.

The tests discussed so far are valid only asymptotically; in small samples, the test statistics will in general not be normally distributed and will have neither mean zero nor variance unity. It should be observed, however, that many estimates of $\boldsymbol{\gamma}$ could be used in place of $\hat{\boldsymbol{\gamma}}$. In particular, Fisher and McAleer (1981), citing Atkinson's (1970) suggestion, in connection with the Cox test, that all quantities should be evaluated under H_0, proposed that the J and P tests be modified by using $\tilde{\boldsymbol{\gamma}}$ instead of $\hat{\boldsymbol{\gamma}}$. Subsequently, Godfrey (1981) and Davidson and MacKinnon (1982) observed that, for the linear regression case with non-stochastic regressors, the resulting J_A test statistic will actually be distributed as t with the appropriate number of degrees of freedom. This follows from the fact that $\tilde{\boldsymbol{\gamma}}$ depends on $\boldsymbol{y}$ only through the fitted values from H_0 (that is, the vector $\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$), which are independent of the error terms; see Milliken and Graybill (1970).

Thus, for the case in which H_0 is a linear regression model with nonstochastic regressors and normally distributed errors, the J_A test provides an exact non-nested hypothesis test. Unfortunately, this does not mean that the other tests discussed above, which are valid only asymptotically, are obsolete. Monte Carlo work reported in Davidson and MacKinnon (1982) suggests that the J_A test can be very much less powerful than the ordinary J test when neither H_0 nor H_1 is true, a situation that is all too likely to occur in practice. Simulations reported in Godfrey and Pesaran (1983) suggest that the J_A test can also be seriously lacking in power when H_1 has more regressors than H_0. These results imply that it will often be desirable to use both tests in practice.

4. Multivariate Models

Both the Cox test and several variants of the P test may be applied to the case of multivariate nonlinear regression models. We may write the null hypothesis as

$$\text{H}_0: \quad y_{ti} = f_{ti}(\boldsymbol{\beta}) + u_{ti}^0, \quad \boldsymbol{u}_t^0 \sim \text{N}(\boldsymbol{0}, \boldsymbol{\Omega}_0), \tag{35}$$

and the alternative as

$$\text{H}_1: \quad y_{ti} = g_{ti}(\boldsymbol{\gamma}) + u_{ti}^1, \quad \boldsymbol{u}_t^1 \sim \text{N}(\boldsymbol{0}, \boldsymbol{\Omega}_1). \tag{36}$$

Here i indexes the m equations and t indexes the T observations, and $\boldsymbol{\Omega}_j$ is the $m \times m$ contemporaneous covariance matrix for the error terms corresponding to hypothesis H_j. Of course, these error terms have the specified distribution only if the respective hypothesis happens to be true. The hypotheses given in (35) and (36) may be multivariate, nonlinear non-simultaneous models, such as demand systems or systems of cost share equations, or they may be the restricted reduced forms of linear simultaneous equations models. They may not be the reduced forms of nonlinear simultaneous equations models, however, because, if the errors adhering to the structural equations were Gaussian, the errors adhering to the reduced form equations could not be.

Pesaran and Deaton (1978) derive their version of the Cox test for the multivariate case as well as the univariate one. The numerator of the test statistic is simply

$$T_0 = \frac{T}{2} \log \frac{|\hat{\boldsymbol{\Omega}}_1|}{|\tilde{\boldsymbol{\Omega}}_{10}|}, \tag{37}$$

where $\hat{\boldsymbol{\Omega}}_1$ is the ML estimate of $\boldsymbol{\Omega}_1$, and $\tilde{\boldsymbol{\Omega}}_{10}$ is defined analogously to $\tilde{\sigma}_{10}$; see equation (10) above. The calculation of the variance of T_0 is quite complicated, and interested readers should consult the original paper. It should perhaps be noted that the calculation of $\tilde{\boldsymbol{\gamma}}$, and hence $\tilde{\boldsymbol{\Omega}}_{10}$, requires the solution of two sets of nonlinear equations. Pesaran and Deaton (1978) give a procedure for doing so, but it is incorrect; see Davidson and MacKinnon (1983). In general, finding $\tilde{\boldsymbol{\gamma}}$ is a non-trivial calculation. As in the univariate case, this calculation could be avoided by using $\hat{\boldsymbol{\Omega}}_{10}$ instead of $\tilde{\boldsymbol{\Omega}}_{10}$, which is valid because $\hat{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}}$ have the same probability limits under H_0.

There are several multivariate versions of the P test, which are discussed in Davidson and MacKinnon (1982, 1983). The simplest artificial compound model, analogous to (24) in the univariate case, is

$$\text{H}_C: \quad y_{ti} = (1 - \alpha)f_{ti}(\boldsymbol{\beta}) + \alpha \hat{g}_{ti} + u_{ti}, \tag{38}$$

where, under H_0, the vector $\boldsymbol{u}_t$ should have covariance matrix $\boldsymbol{\Omega}_0$. Linearizing (38) around the point $\alpha = 0$, $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ yields the multivariate linear regression

$$y_{ti} - \hat{f}_{ti} = \hat{\boldsymbol{F}}_{ti}^\top \boldsymbol{b} + \alpha(\hat{g}_{ti} - \hat{f}_{ti}) + u_{ti}, \tag{39}$$

where f_{ti} and g_{ti} denote $f_{ti}(\hat{\boldsymbol{\beta}})$ and $g_{ti}(\hat{\boldsymbol{\gamma}})$, respectively, and $\hat{\boldsymbol{F}}_{ti}$ denotes the vector of derivatives of $f_{ti}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, evaluated at $\hat{\boldsymbol{\beta}}$. This regression is to be estimated by generalized least squares, using $\hat{\boldsymbol{\Omega}}$ as the assumed covariance matrix, and is referred to as the P_0 test. The test statistic is the t statistic on $\hat{\alpha}$

The P_0 test is particularly simple, but it is not asymptotically equivalent to a Cox test. One may derive alternative tests by considering alternative compound models. In particular, it is possible to write down a combined regression model of which the likelihood function is the same as the likelihood function which results if one combines the likelihoods of H_0 and H_1 exponentially, an idea which goes back to Cox (1962). This model is rather complicated, so it will be omitted here. Linearizing it around $\alpha = 0$, $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ yields

$$y_{ti} - \hat{f}_{ti} = \hat{\boldsymbol{F}}_{ti}^{\top}\boldsymbol{b} + \alpha\hat{h}_{ti} + u_{ti}, \tag{40}$$

where $\hat{h}_{ti}$ is an element of the $T \times m$ matrix

$$\hat{\boldsymbol{h}} \equiv (\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}})\hat{\boldsymbol{\Omega}}_1^{-1}\hat{\boldsymbol{\Omega}}_0. \tag{41}$$

Here $\hat{\boldsymbol{g}}$ and $\hat{\boldsymbol{f}}$ are $T \times m$ matrices with typical elements $\hat{g}_{ti}$ and $\hat{f}_{ti}$, respectively. Once again, the regression is to be estimated by GLS using $\hat{\boldsymbol{\Omega}}_0$ as the assumed covariance matrix. The test statistic is again the t statistic on $\hat{\alpha}$, and Davidson and MacKinnon (1983) refers to this procedure as the P_1 test. In that paper, it is proved that the P_1 test statistic is asymptotically distributed as $N(0, 1)$ under H_0 and that it is asymptotically equivalent to a Cox test.

Little is known about the relative performance of the three tests just described. The P tests are far easier to implement than the Cox test proposed by Pesaran and Deaton (1978). The fact that Cox's procedure is based on likelihood theory suggests that it should have good power. Thus the P_1 test is probably the procedure of choice, since it is asymptotically a Cox test and requires only one artificial linear regression once the models have been estimated.

Notice that, in the case of simultaneous linear regression models, it is not necessary explicitly to solve for the restricted reduced form. Once the models have been estimated to yield $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, it is possible to calculate $\hat{f}_{ti}$, $\hat{g}_{ti}$, $\hat{\boldsymbol{\Omega}}_0$, and $\hat{\boldsymbol{\Omega}}_1$ numerically by solving the model over the sample period, and $\hat{\boldsymbol{F}}_{ti}$ can be calculated numerically. We can also calculate $\hat{\boldsymbol{F}}_{ti}$ numerically for non-simultaneous models as well, of course. The reliability of these calculations can in turn be checked numerically, because if α is restricted to be zero in (40), then the estimate of $\boldsymbol{b}$ should be identically zero by the first-order conditions for $\hat{\boldsymbol{\beta}}$.

5. Models with Different Error Structures

The non-nested hypothesis tests that have been discussed so far may be used only when all of the alternative models have the same error structure. In practice, however, it will often be the case that different models have different error structures. For example, one model may specify that errors adhere additively to the level of the dependent variable,

another that they are multiplicative, so that they adhere additively to the log, and a third that they adhere to the dependent variable multiplied by some weighting factor, as in share equations. The techniques discussed so far cannot handle models which differ in these ways.

The case of linear versus logarithmic models is particularly common in applied work. If the two alternative models are the same except that in one case the dependent variable and some of the regressors appear as logarithms while in the other they appear as levels, the situation can easily be handled without recourse to explicit non-nested hypothesis tests. Suppose the two models are

$$\mathrm{H}_0: \quad \log y_t = \sum_{i=1}^{I} \beta_i \log x_{ti} + \sum_{j=1}^{J} \gamma_j z_{tj} + u_{t0}, \tag{42}$$

and

$$\mathrm{H}_1: \quad y_t = \sum_{i=1}^{I} \beta_i x_{ti} + \sum_{j=1}^{J} \gamma_j z_{tj} + u_{t1}, \tag{43}$$

where u_{tk} is assumed to be i.i.d. normal if H_k is true, for $k = 0, 1$. Strictly speaking, H_0 and H_1 cannot both be potentially valid models, since H_1 implies that y_t can be negative with positive probability while H_0 implies that it cannot be. If y_t in fact cannot be negative, then H_1 must be false, and if y_t in fact can be negative, then H_0 must be false. In many applied cases, however, where y_t is typically ten or more standard errors above zero, the probability that H_1 will generate negative values can surely be neglected, and H_0 and H_1 may thus often be regarded as plausible rival models.

One natural way to handle (42) and (43) is to nest them within a comprehensive hypothesis based on the Box-Cox transformation:

$$\mathrm{H}_C: \quad \phi(y_t, \lambda) = \sum_{i=1}^{I} \beta_i \phi(x_{ti}, \lambda) + \sum_{j=1}^{J} \gamma_j z_{tj} + u_t, \tag{44}$$

where

$$\begin{aligned} \phi(x, \lambda) &= (x^\lambda - 1)/\lambda \quad \text{if } \lambda \neq 0 \\ &= \log x \quad \text{if } \lambda = 0. \end{aligned} \tag{45}$$

Here $\phi(x, \lambda)$ is the famous transformation introduced by Box and Cox (1964); it is obvious that (44) reduces to (42) if $\lambda = 0$ and to (43) if $\lambda = 1$, provided in the latter case that the z_{tj} include a constant term.

The comprehensive model (44) can be used to test the validity of (42) and (43) in several different ways. Perhaps the simplest one conceptually is just to estimate (44) and then test the hypotheses that $\lambda = 0$ and $\lambda = 1$ by likelihood ratio tests, as Box and Cox (1964) suggested. However, this requires one to estimate (44). Asymptotically equivalent procedures based on Lagrange multiplier tests are easily developed, of course; for one such, see Godfrey and Wickens (1981).

A somewhat more ingenious procedure has been suggested by Andrews (1971). By taking a Taylor series approximation to (44) around $\lambda = 0$ or $\lambda = 1$, and then replacing y_t by its fitted value whenever it appears on the right-hand side, Andrews develops a test which simply requires that one rerun each of regressions (42) and (43) with one additional regressor. The additional regressor depends on which model is to be tested, and it is a fairly complicated function of the x_{ti}, the z_{tj}, and the estimates of $\boldsymbol{\beta}$ under the null. The resulting test statistic is simply the t statistic on the additional regressor, and it actually has the t distribution in finite samples. For details, see Andrews (1971) or Godfrey and Wickens (1981).

Procedures based on the Box-Cox transformation are useful only when H_0 and H_1 take special forms like (42) and (43) and have essentially the same regressors. More generally, there might be many alternative models, some with the dependent variable in levels and some with it in logs, perhaps with very different explanatory variables. In particular, suppose that there are two alternative models, which may be linear or nonlinear:

$$H_0: \quad \log y_t = f_t(\boldsymbol{\beta}) + u_{t0}, \quad \boldsymbol{u}_0 \sim \mathrm{N}(\boldsymbol{0}, \sigma_0^2 \mathbf{I}), \tag{46}$$

and

$$H_1: \quad y_t = g_t(\boldsymbol{\beta}) + u_{t1}, \quad \boldsymbol{u}_1 \sim \mathrm{N}(\boldsymbol{0}, \sigma_1^2 \mathbf{I}). \tag{47}$$

Aneuryn-Evans and Deaton (1980) have proposed variants of the Cox test which may be used to test H_0 against the evidence provided by H_1, and vice versa. The former procedure is complicated but not difficult; the latter, however, requires a good deal of numerical integration, and a special-purpose computer program would have to be written.

Simpler and more general procedures based on artificial nesting have recently been suggested by Davidson and MacKinnon (1984). Suppose the model to be tested is

$$H_0: \quad y_t = f_t(\boldsymbol{\beta}) + u_{t0}, \quad \boldsymbol{u}_0 \sim \mathrm{N}(\boldsymbol{0}, \sigma_0^2 \mathbf{I}), \tag{48}$$

and the alternative is

$$H_0: \quad h_t(y_t) = g_t(\boldsymbol{\gamma}) + u_{t1}, \quad \boldsymbol{u}_1 \sim \mathrm{N}(\boldsymbol{0}, \sigma_1^2 \mathbf{I}). \tag{49}$$

Here $h_t(y_t)$ may be any twice continuously differentiable transformation which does not depend on any unknown parameters. Examples include $\log y_t$, $\exp y_t$, y_t^2, y_t/z_t (for z_t exogenous), and so on. Thus the models (48) and (49) include (46) and (47) as special cases.

The simplest artificial compound model which nests H_0 and H_1 is

$$H_C: \quad (1 - \alpha)\big(y_t - f_t(\boldsymbol{\beta})\big) + \alpha\big(h_t(y_t) - g_t(\boldsymbol{\gamma})\big) + u_t. \tag{50}$$

Replacing $\boldsymbol{\gamma}$ by $\hat{\boldsymbol{\gamma}}$ would seem to yield a conceptually straightforward J test procedure. However, this procedure would not be computationally easy. H_C is not a nonlinear

regression model, because the dependent variable, y_t, does not have a coefficient of unity. The loglikelihood function will include the Jacobian term

$$\sum_{t=1}^{T} \log \left| 1 - \alpha + \alpha h_t'(y_t) \right|, \tag{51}$$

where $h_t'(y_t)$ is the derivative of $h_t(y_t)$ with respect to y_t. Thus standard packaged programs for nonlinear estimation would probably not be applicable.

One way around this problem is to test the hypothesis that $\alpha = 0$ using a procedure which requires estimates only under the null hypothesis. MacKinnon, White, and Davidson (1983) show that a modified version of the P test may be used for this purpose. The artificial regression is

$$y_t - \hat{f}_t = \hat{\boldsymbol{F}}_t \boldsymbol{b} + \alpha \left(\hat{g}_t - h_t(\hat{f}_t) \right) + u_t, \tag{52}$$

and the t statistic on $\hat{\alpha}$ is once again asymptotically $\mathrm{N}(0, 1)$ under H_0. A more general but somewhat more cumbersome procedure has been suggested by Davidson and MacKinnon (1984). This procedure requires that one run a double-length artificial linear regression, that is, one with $2T$ observations instead of T. The resulting test is equivalent to a Lagrange Multiplier test, and it will generally have more power than the modified P test of (52). For details, see the original paper.

At this point, it seems natural to speculate that methods of non-nested hypothesis testing can be developed for almost all classes of econometric models by a technique similar to the one just discussed. First, find a way to nest both models within an artificial composite model. Next, replace the parameters of the alternative by asymptotically non-stochastic estimates. Finally, perform a Lagrange Multiplier or similar test; see Breusch and Pagan (1980). Just how these tests would be related to Cox tests, if the latter were developed for the same cases, is a matter for conjecture. It would seem that the two procedures would not, in general, be asymptotically equivalent.

Two interesting applications of the Cox test should be mentioned here. Amemiya (1973) derived Cox tests for lognormal versus Gamma errors in linear regression models, assuming that the non-stochastic parts of the models were the same. Morimune (1979) developed Cox tests, and also modified tests based on the same fundamental idea, for bivariate dichotomous dependent variable models based on the logistic and normal distributions. These papers illustrate the wide range of applicability of Cox's original idea.

6. Interpreting Non-Nested Hypothesis Tests

When two models are tested against each other, there are four (or perhaps nine) possible outcomes. The number is larger if one distinguishes between rejection in the direction of the alternative model and rejection in the opposite direction. Interpreting the results of such a pair of tests may therefore seem complicated; this matter is discussed at length by Fisher and McAleer (1979). For most practical purposes, however,

interpretation is not particularly difficult. If neither model is rejected, then the data do not allow us to say that either is false. If one model is rejected and the other is not, then the set of models that are worth further examination has been reduced in size (a very useful outcome). If both models are rejected, then that set has been reduced in size even more; indeed, if there are no other models around, we should presumably try to invent some! In this case, the signs of the test statistics may be useful, since they will tell us whether we should look for a model that combines features of both the models we have rejected, or for a model that moves further away from one of the models.

There are, however, some situations where the results of non-nested hypothesis tests can be predicted in advance, so that the tests do not tell us very much. Suppose that H_0 and H_1 are univariate nonlinear regression models, as in (2) and (3), so that the P test regression is

$$\boldsymbol{y} - \hat{\boldsymbol{f}} = \hat{\boldsymbol{F}}\boldsymbol{b} + \alpha(\hat{\boldsymbol{g}} - \hat{\boldsymbol{f}}) + \boldsymbol{u}. \tag{53}$$

It is clear that, if $\mathcal{L}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_0^2)$ is less than $\mathcal{L}(\hat{\boldsymbol{\gamma}}, \hat{\sigma}_1^2)$ by more than 1.92, then H_0 must be rejected at the .05 level, because 3.84 is the .05 critical value for the chi-squared distribution with one degree of freedom. After all, if α were unity and $\boldsymbol{b}$ were zero, regression (53) would reduce to

$$\boldsymbol{y} = \hat{\boldsymbol{g}} + \boldsymbol{u}, \tag{54}$$

which would have the same loglikelihood value as H_1, whereas under the null hypothesis that $\alpha = 0$, the combined model would have the same likelihood value as H_0. Thus the P test regression (53) must fit at least as well as whichever of H_0 and H_1 fits better. Therefore, if H_1 fits significantly better than H_0 on a one degree of freedom likelihood ratio test, H_0 must be rejected. Essentially the same thing is true for the Cox test as well. On the other hand, H_0 may be rejected in a test against H_1 even when H_1 fits much worse than H_0, and this outcome cannot be predicted in advance. In such a case, a non-nested test provides genuinely new information.

The above arguments suggest that the results of non-nested hypothesis tests should be interpreted with caution when the sample size is small. If H_1 has more parameters than H_0, it will tend to fit better. By simply adding enough irrelevant variables to an alternative hypothesis which initially failed to reject H_0, one could always eventually come up with an alternative hypothesis which would reject H_0. Thus, in small samples, it would seem wise to use these tests only when both models have similar numbers of parameters. Of course, these remarks do not apply to the J_A test for linear models.

7. Future Directions

It is now possible to perform non-nested hypothesis tests of model specification for a wide variety of econometric models. Nevertheless, a great deal of work remains to be done in this area. Little is known about the behavior of the tests when neither the null nor the alternative hypothesis is true, which is of course the situation most likely to be encountered in practice. Moreover, there are still many empirical problems to which the tests cannot readily be applied. For example, it is often the case that the variable,

or set of variables, which is treated as endogenous by one theory differs from the set of variables which is treated as endogenous by an alternative theory. At present, it is not at all clear how non-nested hypothesis tests could be used in such a case.

Empirical work using non-nested hypothesis tests is still all but non-existent. Pesaran and Deaton (1978) estimated five models of consumption behavior and tested them against each other, but the models were deliberately kept very simple for purposes of illustration, and cannot be taken seriously; they all suffer from numerous econometric problems, so that all could easily be rejected without recourse to non-nested hypothesis tests. The only serious published paper which utilizes these tests is Deaton (1978), in which two competing demand systems are tested against each other; both systems are rejected. Some unpublished work exists, but by and large it is fair to say that the literature on non-nested hypothesis testing has had no impact whatsoever on empirical work in economics. Economists continue to estimate models, and to draw policy conclusions on the basis of them, without ever testing them against alternative models, or (in most cases) even making any serious effort to test their specifications at all.

Should non-nested hypothesis tests play a major role in applied work? No doubt some econometricians would argue that they should not. In many practical cases, a specification error that could be detected by application of such a test could also be detected by application of one of many currently available tests for different types of model misspecification. After all, it must surely be rare that misspecification fails to cause serial correlation, heteroskedasticity, failure of linear or nonlinear restrictions to hold, unstable parameter estimates, correlation between error terms and powers of the fitted values (Ramsey, 1969; Ramsey and Schmidt, 1976), divergence between alternative estimators of the covariance matrix (White, 1982), or some other observable misbehavior of the model. Thus, in principle, econometricians may rarely need non-nested hypothesis tests to detect misspecified models.

In practice, however, many of those who do applied econometric work are very sloppy about testing their models. Anything that forces them to test models a bit more rigorously is therefore highly desirable. Thus I would argue that, even if non-nested hypothesis tests never had greater power than alternative tests for model misspecification, the former would still be a very useful part of the applied econometrician's kit of tools. Moreover, these tests often will have good power, especially when the alternative hypothesis, or another model which resembles it, does happen to be true.

When he or she decides to make use of non-nested hypothesis tests, an investigator is forced to recognize that there are usually many models to explain any given phenomenon, and that most of these models must be false. The investigator is then obliged to estimate several non-nested alternative models, and to give each of them several chances to show that it is not true by confronting it with the evidence provided by the data and by the other models. Doing this must increase greatly the probability that the model finally selected will not be thoroughly false. Even if that is the only contribution of nonnested hypothesis tests, it will be a very valuable one.

8. Postscript

This paper was published along with comments by Roger W. Klein, Michael McAleer and Anil Bera, Grayham Mizon and Jean-François Richard, Kimio Murimune, and M. Hashem Pesaran. There was also a reply by the author. One of the points made in the reply is of some interest and should really have been made in the paper. It is therefore included here. I have also cited a few more recent papers, which are listed in a separate "Additional References" section at the end of the paper.

In the case of linear regression models, there is much less diversity of test statistics than may at first appear to be the case. The numerator of the J statistic is

$$\boldsymbol{y}^{\top}(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\boldsymbol{y}, \tag{55}$$

and the numerator of the J_A statistic is

$$\boldsymbol{y}^{\top}(\mathbf{I} - \boldsymbol{M}_0)(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\boldsymbol{y}. \tag{56}$$

Here $\boldsymbol{M}_0$ denotes the matrix that projects orthogonally off the subspace spanned by all the regressors in the H_0 model, and $\boldsymbol{M}_1$ denotes the matrix that projects orthogonally off the subspace spanned by all the regressors in the H_1 model. Thus, if the regressor matrix for H_0 were $\boldsymbol{X}$, then $\boldsymbol{M}_0 = \mathbf{I} - \boldsymbol{X}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}$.

If we linearize T_0, the numerator of the Cox test statistic, which is given in equation (11), around zero, we obtain

$$\frac{\hat{\sigma}_1^2 - \hat{\sigma}_0^2 - \tilde{\sigma}_a^2}{\hat{\sigma}_1^2}, \tag{57}$$

where $\tilde{\sigma}_a^2$ is the estimated variance from the auxiliary regression (12). In the linear case, T times the numerator of (57) is

$$\boldsymbol{y}^{\top}\boldsymbol{M}_1\boldsymbol{y} - \boldsymbol{y}^{\top}\boldsymbol{M}_0\boldsymbol{y} - \boldsymbol{y}^{\top}(\mathbf{I} - \boldsymbol{M}_0)(\mathbf{I} - \boldsymbol{M}_1)(\mathbf{I} - \boldsymbol{M}_0)\boldsymbol{y}. \tag{58}$$

Notice that (58) is just the sum of squared residuals, or SSR, from H_1, minus the SSR from H_0, minus an estimate of the amount by which the SSR from H_1 should exceed the one from H_0. After a few manipulations, (58) can be rewritten as

$$-\boldsymbol{y}^{\top}(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\boldsymbol{y} - \boldsymbol{y}^{\top}(\mathbf{I} - \boldsymbol{M}_0)(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\boldsymbol{y}, \tag{59}$$

which is simply minus the sum of the numerators of the J and J_A statistics. Thus there are really only two quantities that matter, namely, the numerators of the J and J_A statistics. Other test statistics, such as the Cox test (14), simply depend on them.

In Godfrey and Pesaran (1983), an estimate of the mean of (58) under H_0 is subtracted from (58), and that adjusted quantity, divided by the square root of an appropriate variance estimate, is used as a test statistic. The objective is to obtain good small-sample properties without the poor power properties of the J_A test. This is a very good idea, but it is odd to begin with (58) rather than (55), because one term of the

former already has mean zero. Moreover, since the only merit of the J_A numerator is that it has mean zero, it is pointless to include it in a numerator of which the mean is to be adjusted anyway.

The Godfrey-Pesaran idea can be applied directly to (55), the mean of which is

$$\sigma_0^2 \operatorname{Tr}\big((\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\big). \tag{60}$$

Thus a statistic which will have mean zero is

$$\boldsymbol{y}^\top(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\boldsymbol{y} - \frac{\boldsymbol{y}^\top\boldsymbol{M}_0\boldsymbol{y}}{n - k_0}\operatorname{Tr}\big((\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\big). \tag{61}$$

The variance of this statistic under H_0 is

$$\begin{aligned}
\sigma_0^2\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{X}\boldsymbol{\beta} + \sigma_0^4\Big(&\operatorname{Tr}\big((\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\big) \\
+ \operatorname{Tr}\big((\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0(\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\big) &- \frac{2}{n - k_0}\Big(\operatorname{Tr}\big((\mathbf{I} - \boldsymbol{M}_1)\boldsymbol{M}_0\big)\Big)^2\Big),
\end{aligned} \tag{62}$$

which can be estimated straightforwardly. The adjusted J statistic suggested here will be much harder to compute than the ordinary J statistic, but it should be a little easier to compute than the similar statistic suggested by Godfrey and Pesaran. However, the success of the bootstrap in correcting the tendency of the J test to overreject (see below) has made this sort of adjustment much less attractive.

In the paper, it was remarked that the J test is invalid when certain orthogonality conditions hold. This is also true for the Cox and J_A tests. Near orthogonality also causes serious problems for all these tests. See Michelis (1999) for an illuminating analysis of this case.

Since this paper was published, it has become apparent that the test of Andrews (1971) for linear versus loglinear models can be seriously deficient in power and that the test of Godfrey and Wickens (1981) often overrejects severely in finite samples. Better tests of linear versus loglinear models are available. See Davidson and MacKinnon (1985, 1988) and Godfrey, McAleer, and McKenzie (1988).

During the past ten years, much has been learned about the finite-sample properties of non-nested hypothesis tests. The most important result is that these can be dramatically improved by using a parametric or semiparametric bootstrap. Early evidence on this point may be found in Godfrey (1998). Davidson and MacKinnon (2002a) provide a detailed discussion of how to bootstrap the J test and a theoretical explanation of why doing so works so well. The bootstrap, when implemented properly, almost entirely solves the problem of finite-sample overrejection by the J test, a problem that is especially acute when H_1 has more parameters than H_0. In particularly ill-behaved cases, it may be desirable to use the fast double bootstrap instead of more straightforward bootstrap procedures; see Davidson and MacKinnon (2002b).

References

Amemiya, T. (1973). "Regression analysis when the variance of the dependent variable is proportional to the square of its expectation," *Journal of the American Statistical Association*, 68, 928–934.

Amemiya, T. (1980). "Selection of regressors," *International Economic Review*, 21, 331–354.

Andrews, D. F. (1971). "A note on the selection of data transformations," *Biometrika*, 58, 249–254.

Aneuryn-Evans, G., and A. S. Deaton (1980). "Testing linear versus logarithmic regression models," *Review of Economic Studies*, 47, 275–291.

Atkinson, A. C. (1969). "A test for discriminating between models," *Biometrika*, 56, 337–347.

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

Box, G. E. P., and D. R. Cox (1964). "An analysis of transformations," *Journal of the Royal Statistical Society*, Series B, 26, 211–252.

Breusch, T. S., and A. R. Pagan (1980). "The Lagrange Multiplier test and its applications to model specification in econometrics," *Review of Economic Studies*, 47, 239–253.

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

Davidson, R., and J. G. MacKinnon (1980). "On a simple procedure for testing non-nested regression models," *Economics Letters*, 5, 45–48.

Davidson, R., and J. G. MacKinnon (1981). "Several tests for model specification in the presence of alternative hypotheses," *Econometrica*, 49, 781–793.

Davidson, R., and J. G. MacKinnon (1982). "Some non-nested hypothesis tests and the relations among them," *Review of Economic Studies*, 49, 551–565.

Davidson, R., and J. G. MacKinnon (1983). "Testing the specification of multivariate models in the presence of alternative hypotheses," *Journal of Econometrics*, 23, 301–313.

Davidson, R., and J. G. MacKinnon (1984). "Model specification tests based on artificial linear regressions," *International Economic Review*, 25, 485–502.

Deaton, A. S. (1978). "Specification and testing in applied demand analysis," *Economic Journal*, 88, 524–536.

Dhrymes, P. J., E. P. Howrey, S. H. Hymans, J. Kmenta, E. E. Leamer, R. E. Quandt, J. B. Ramsey, H. T. Shapiro, and V. Zarnowitz (1972). "Criteria for evaluation of econometric models," *Annals of Economic and Social Measurement*, 1, 291–324.

Durbin, J. (1970). "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables," *Econometrica*, 38, 410–421.

Fisher, G. R., and M. McAleer (1979). "On the interpretation of the Cox test in econometrics," *Economics Letters*, 4, 145–150.

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

Gaver, K. M., and M. S. Geisel (1974). "Discriminating among alternative models: Bayesian and non-Bayesian methods," in *Frontiers in Econometrics*, ed. P. Zarembka, New York: Academic Press, 49–77.

Godfrey, L. G. (1983). "Testing non-nested models after estimation by instrumental variables or least squares," *Econometrica*, 51, 355–366.

Godfrey, L. G., and Wickens, M. R. (1981). "Testing linear and log-linear regressions for functional form," *Review of Economic Studies*, 48, 487–496.

Hoel, P. G. (1947). "On the choice of forecasting formulas," *Journal of the American Statistical Association*, 42, 605–611.

Judge, G. G., W. E. Griffiths, R. C. Hill, and T.-C. Lee (1980). *The Theory and Practice of Econometrics*, New York: Wiley.

Leamer, E. E. (1979). "Information criteria for choice of regression models: A comment," *Econometrica*, 47, 507–510.

MacKinnon, J. G., H. White, and R. Davidson (1983). "Tests for model specification in the presence of alternative hypotheses: Some further results," *Journal of Econometrics*, 21, 53–70.

Milliken, G. A., and F. A. Graybill (1970). "Extensions of the general linear hypothesis model," *Journal of the American Statistical Association*, 65, 797–807.

Morimune, K. (1979). "Comparisons of normal and logistic models in the bivariate dichotomous analysis," *Econometrica*, 47, 957–975.

Pereira, B. de B. (1977a). A note on the consistency and on the finite sample comparisons of some tests of separate families of hypotheses," *Biometrika*, 64, 109–113.

Pereira, B. de B.(1977b). "Discriminating among separate models: A bibliography," *International Statistics Review*, 45, 163–172.

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

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

Pesaran, M. H., and A. S. Deaton (1978). "Testing non-nested nonlinear regression models," *Econometrica*, 46, 677–694.

Quandt, R. E. (1974). "A comparison of methods for testing non-nested hypotheses," *Review of Economics and Statistics*, 56, 92–99.

Ramsey, J. B. (1969). "Tests for specification errors in classical linear least-squares regression analysis," *Journal of the Royal Statistical Society*, Series B, 31, 350–371.

Ramsey, J. B., and P. Schmidt (1976). "Some further results on the use of OLS and BLUS residuals in specification error tests," *Journal of the American Statistical Association*, 71, 389–390.

Sawa, T. (1978). "Information criteria for discriminating among alternative regression models," *Econometrica*, 46, 1273–1291.

Sawyer, K. (1980). *The Theory of Econometric Model Selection*, Department of Statistics, Australian National University, Ph.D. Thesis

Walker, A. M. (1967). "Some tests of separate families of hypotheses in time series analysis," *Biometrika*, 54, 39–68.

White, H. (1982a). "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

White, H. (1982b). "Regularity conditions for Cox's test of non-nested hypotheses," *Journal of Econometrics*, 19, 301–318.

Additional References

Davidson, R., and J. G. MacKinnon (1985). "Testing linear and loglinear regressions against Box-Cox alternatives," *Canadian Journal of Economics*, 25, 499–517.

Davidson, R., and J. G. MacKinnon (1988). "Double-length artificial regressions," *Oxford Bulletin of Economics and Statistics*, 50, 203–217.

Davidson, R., and J. G. MacKinnon (2002a). "Bootstrap J tests of nonnested linear regression models," *Journal of Econometrics*, 109, 167–193.

Davidson, R., and J. G. MacKinnon (2002b). "Fast double bootstrap tests of nonnested linear regression models," *Econometric Reviews*, 21, 417–427.

Godfrey, L. G. (1998). "Tests of non-nested regression models: Some results on small sample behaviour and the bootstrap," *Journal of Econometrics*, 84, 59–74.

Godfrey, L. G., M. McAleer, and C. R. McKenzie (1988). "Variable addition and Lagrange Multiplier tests for linear and logarithmic regression models," *Review of Economics and Statistics*, 70, 492–503.

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.

Michelis, L. (1999). "The distributions of the J and Cox non-nested tests in regression models with weakly correlated regressors," *Journal of Econometrics*, 93, 369–401.