ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH

QUEEN'S UNIVERSITY

Kingston, Ontario, Canada K7L 3N6

Several Tests for Model Specification in the Presence of Alternative Hypotheses[1]

DISCUSSION PAPER NO. 378

by

Russell Davidson
and
James G. MacKinnon

March, 1980

This is a substantially revised version of Discussion Paper No. 336.

ABSTRACT

Several procedures are proposed for testing the specification of an econometric model in the presence of one or more other models which purport to explain the same phenomenon. These procedures are shown to be closely related, but not identical, to the non-nested hypothesis tests recently proposed by Pesaran and Deaton [7], and to have similar asymptotic properties. They are remarkably simple both conceptually and computationally, and, unlike earlier techniques, they may be used to test against several alternative models simultaneously. Some empirical results are presented which suggest that the ability of the tests to reject false hypotheses is likely to be rather good in practice.

Introduction

One of the major functions of econometrics is to test the validity of models put forward by economic theory. Most techniques for hypothesis testing in econometrics, however, simply allow one to test restrictions on a model more general than the one being tested, conditional on the more general model being valid. A striking exception to this generalization is a technique recently suggested by Pesaran and Deaton [7], based on the earlier work of Cox [2,3] and Pesaran [6]. The procedure they propose, henceforth referred to as the Cox-Pesaran-Deaton or CPD test, allows one to test the truth of a possibly nonlinear and multivariate regression model, when there exists a non-nested alternative hypothesis. The latter need not be true, and need not even be a hypothesis which the investigator would seriously maintain.

In this paper, we propose several related procedures for doing essentially the same thing as the CPD test. For simplicity, we consider only univariate models. Our tests are conceptually much simpler than the CPD test, can readily be implemented using existing computer software, and can handle several alternative hypotheses simultaneously. In Section 1 we describe our test procedures; in Section 2 we present some theoretical results on the relationships among them and between them and the CPD test; and in Section 3 we present some empirical results on the application of our tests to the data and models investigated by Pesaran and Deaton.

1. A Simple Test for Specification Error

We consider initially the case of a single-equation, possibly nonlinear regression model, the truth of which we wish to test,

$$H_0: \; y_i = f_i(X_i, \beta) + \varepsilon_{0i} \qquad (1)$$

where y_i is the i^{th} observation on the dependent variable, X_i is a vector of observations on exogenous variables, β is a k-vector of parameters to be estimated, and the error term ε_{0i} is assumed to be NID $(0, \sigma_0^2)$.

Suppose that economic theory suggests an alternative hypothesis, though not one in which we need have any faith,

$$H_1: \; y_i = g_i(Z_i, \gamma) + \varepsilon_{1i} \qquad (2)$$

where Z_i is a vector of observations on exogenous variables, γ is an ℓ-vector of parameters to be estimated and ε_{1i} is NID$(0, \sigma_1{}^2)$ if H_1 is true. We assume that H_1 is not nested within H_0 and that H_0 is not nested within H_1. Thus the truth of H_0 implies the falsity of H_1, and vice versa.

Consider the possibly nonlinear regression

$$y_i = (1-\alpha) f_i(X_i, \beta) + \alpha \, \hat{g}_i + \varepsilon_i, \qquad (3)$$

where $\hat{g}_i = g_i(Z_i, \hat{\gamma})$ and $\hat{\gamma}$ is the ML estimate of γ. If H_0 is true, then the true value of α is zero. Now $\hat{g}_i$ is simply a function of the exogenous variables Z_i and the parameter estimates $\hat{\gamma}$. The former are independent of ε_i by assumption. Asymptotically, the latter are also independent of ε_i, because the influence of any particular error term on the estimates tends to zero as the sample size tends to infinity. Thus, asymptotically, $\hat{g}_i$ will be independent of ε_i, so that one may validly test whether $\alpha = 0$ in (3) by using a conventional asymptotic t-test or, equivalently, a likelihood ratio test.

An even simpler way to test the truth of H_0 would be to estimate

$$y_i = (1-\alpha)\hat{f}_i + \alpha\hat{g}_i + \varepsilon_i, \text{ or } y_i - \hat{f}_i = \alpha(\hat{g}_i - \hat{f}_i) + \varepsilon_i, \qquad (4)$$

where $\hat{f}_i = f_i(X_i,\hat{\beta})$. However, the t-statistic for $\hat{\alpha}$ from (4) provides a test the asymptotic size of which is smaller than its nominal size, as we shall demonstrate in the next section. In order to rectify this, it is possible to compute an asymptotically valid standard error for $\hat{\alpha}$ from (4) by doing an auxiliary regression and some other simple calculations, but that is not the simplest approach. Instead, one merely needs to estimate a regression which is the linearization of (3) about $\beta = \hat{\beta}$:

$$y_i - \hat{f}_i = \alpha(\hat{g}_i - \hat{f}_i) + \hat{F}_i b + \varepsilon_i, \qquad (5)$$

where $\hat{F}_i$ is a row-vector containing the derivatives of f with respect to the parameters β for the i^{th} observation, evaluated at $\hat{\beta}$. It is clear that (3) and (5) will yield identical estimates of α and its standard error if H_0 is a linear regression model, since in that case $\hat{F}_i = X_i$ and $\hat{f}_i = X\hat{\beta}$ is simply a linear combination of the regressors. In the nonlinear case (3) and (5) will yield different results in small samples, but we shall show that they yield identical results asymptotically when H_0 is true.

We have thus suggested three procedures for testing the validity of H_0. The first procedure, based on (3), will be referred to as the J-test, since it involves estimating α and β jointly. It is extremely easy to use when H_0 is linear. The second procedure, based on (4), will be referred to as the C-test, since it involves estimating α conditional on $\hat{\beta}$. Since the t-statistic from (4) asymptotically has variance less than unity under H_0, (4) may be all that one has to estimate to reject H_0. The third procedure, based on (5), will be called the P-test, for reasons that will become clear later on. It is likely to be much easier to perform than the J-test when H_0 is nonlinear, because the latter involves a nonlinear regression which may not be well-behaved. Thus we recommend the J-test when H_0 is linear, the P-test when H_0 is nonlinear, and the C-test as a simple preliminary test when H_0 is nonlinear and $\hat{F}_i$ is not easy to calculate.

It is obvious that, if H_1 is true, the estimates of α from (3), (4) or (5) will converge asymptotically to one. This suggests that one could test the truth of H_1 without doing any more regressions. That is not quite true. The t-statistics from (3) and (5) are conditional on the truth of H_0, not on the truth of H_1. Thus, as we indicate in the next section, a t-statistic which is valid for testing the truth of H_0 will not be valid for testing the truth of H_1. If one wants to test H_1 the simplest procedure is simply to reverse the roles of H_0 and H_1 and carry out the test again. When this is done, it is conceivable that both hypotheses may be rejected, or that neither may be rejected. It is also conceivable that one may be rejected and the other may not be, in which case one would presumably want to choose the latter over the former. However, like the CPD test, our procedures are really designed for testing model specification, not for choosing among a number of competing models. If one simply wants to choose one out of a set of competing models, one should use some sort of information criterion (see, e.g., Sawa [8]), rather than our procedures or the CPD test.

Unlike the CPD test, our J- and P-tests can be used to test the truth of a hypothesis against several alternatives at once. To test H_0 against m alternative models $g_j(Z_{ji}, \gamma_j)$ by a J-test, one would simply estimate

$$y_i = (1 - \Sigma_{j=1}^m \alpha_j) f_i(X_i, \beta) + \Sigma_{j=1}^m \alpha_j g_{ji} + \varepsilon_i, \qquad (6)$$

and perform a likelihood ratio test of the restriction that all the α_j's are zero. For the P-test, one would estimate

$$y_i - \hat{f}_i = \Sigma_{j=1}^m \alpha_j (\hat{g}_{ji} - \hat{f}_i) + \hat{F}_i b + \varepsilon_i, \qquad (7)$$

and perform the same likelihood ratio test. If there are several quite different alternative hypotheses, this seems a more natural procedure than testing H_0 against each of them singly.

A different approach to testing non-nested regression models is to form a compound model from two or more alternative hypotheses and test the restrictions implied by only one of them being true. There is a close relationship between this approach and ours, which can easily be demonstrated for the case where both H_0 and H_1 are linear models,

$$H_0: \; y_i = X_i\beta_1 + W_i\beta_2 + \varepsilon_{0i} \qquad (8)$$

$$H_1: \; y_i = Z_i\gamma_1 + W_i\gamma_2 + \varepsilon_{1i}. \qquad (9)$$

Here X_i and Z_i denote vectors of regressors which are unique to H_0 and H_1 respectively, and W_i denotes the regressors which are common to both hypotheses. One may form the compound model

$$y_i = (1-\alpha)X_i\beta_1 + W_i [(1-\alpha)\beta_2 + \alpha\gamma_2] + \alpha Z_i\gamma_1 + \varepsilon_i \qquad (10)$$

and test whether $\alpha\gamma_1 = 0$ using an F test, a procedure suggested by Atkinson [1] and discussed by Pesaran [6] as an alternative to the CPD test. Our J-test involves estimating the compound model

$$y_i = (1-\alpha) (X_i\beta_1 + W_i\beta_2) + \alpha(W_i\hat{\gamma}_2 + Z_i\hat{\gamma}_1) + \varepsilon_i \qquad (11)$$

which can be rewritten as

$$y_i = X_i\beta_1^* + W_i(\beta_2^* + \alpha\hat{\gamma}_2) + \alpha Z_i\hat{\gamma}_1 + \varepsilon_i \qquad (12)$$

where $\beta_j^* = (1-\alpha)\beta_j$. The only real difference between (10) and (12), in terms of parameters which are identifiable, is that in the latter, γ_1 is restricted to be proportional to $\hat{\gamma}_1$.[2] Thus in cases where there is only one regressor in H_0 that is not in H_1, the J-test will yield exactly the same results as the compound model approach. In other cases, the two procedures will not yield the same results.

The procedures we have proposed are conceptually much simpler than the CPD test, and computationally somewhat simpler, especially for nonlinear models. Before advocating their use, however, we must investigate their properties and compare them to those of the CPD test. That is done in the next section.

2. Asymptotic Properties of the Tests

In this section, we shall derive the asymptotic distributions of various test statistics, on the assumption either that H_0 (equation (1)) is true, or that H_1 (equation (2)) is true. Completeness requires a discussion of the case in which neither is true; unfortunately nothing specific can be said without knowing what is true, and the possibilities here are too numerous for any general conclusions to follow. Here we shall simply remark that all the tests we consider are capable of rejecting both H_0 and H_1.

We shall make the following assumptions:

(A1) Either H_0 or H_1, as in (1) or (2), is true, with true parameters (β_0, σ_0^2) or (γ_1, σ_1^2).

(A2) The vectors X_i and Z_i are non-stochastic for all i = 1, ... n, and are fixed in repeated samples.

(A3) Let the matrices of partial derivatives with respect to β or γ of the functions f_i and g_i be denoted by $F(\beta)$ and $G(\gamma)$, where these matrices are $n \times k$ and $n \times \ell$; their transposes are $F^T(\beta)$ and $G^T(\gamma)$. Then, as $n \to \infty$, the matrices

$$\frac{1}{n} F^T(\beta)F(\beta),\ \frac{1}{n} G^T(\gamma)G(\gamma),\ \frac{1}{n} F^T(\beta)G(\gamma)$$

converge to well-defined finite limits for all bounded β and γ, the first two being positive definite and the third non-zero.[3]

Now let us consider regression (3). The log-likelihood function for it is

$$L(\alpha,\beta,\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\,||\,y - (1-\alpha)f(\beta) - \alpha\hat{g}||^2.$$

Here y, $f(\beta)$ and $\hat{g}$ denote $n \times 1$ vectors of y_i, $f_i(X_i,\beta)$, and $g_i(Z_i,\hat{\gamma})$, and $||\ldots||$ denotes the Euclidean norm of a vector. The likelihood equations, which are the first-order conditions for a maximum of L, are obtained by setting to zero the following partial derivatives:

$$L_\alpha = (1/\sigma^2)(\hat{g} - f(\beta))^T(y - (1-\alpha)f(\beta) - \alpha\hat{g}),$$

$$L_\beta = (1/\sigma^2)(1-\alpha)\ F^T(\beta)\ (y - (1-\alpha)f(\beta) - \alpha\hat{g}),$$

$$L_{\sigma^2} = -\ (n/2\sigma^2) + (1/2\sigma^4)\ ||\ y - (1-\alpha)f(\beta) - \alpha\hat{g}||^2. \qquad (13)$$

The ML estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$ satisfy the likelihood equations, and consequently their probability limits under H_0 satisfy the equations

$$\operatorname{plim}_0 \frac{1}{n}\frac{\partial L(\hat{\alpha},\hat{\beta},\hat{\sigma}^2)}{\partial(\alpha,\beta,\sigma^2)} = 0,$$

in obvious abbreviated notation. It is immediate from (13) that these limits are 0, β_0 and σ_0^2 respectively. Then as usual we have:

$$\sqrt{n}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta} - \beta_0\\ \hat{\sigma}^2 - \sigma_0^2\end{bmatrix} \sim I^{-1}\frac{1}{\sqrt{n}}\begin{bmatrix}L_\alpha(0,\ \beta_0,\ \sigma_0^2)\\ L_\beta(0,\ \beta_0,\ \sigma_0^2)\\ L_{\sigma^2}(0,\beta_0,\ \sigma_0^2)\end{bmatrix} \qquad (14)$$

where I, the information matrix, is defined by

$$I = \text{plim}_0 \, [-(1/n) \, D^2 L(0, \beta_0, \sigma_0^2)],$$

and '~' relates quantities whose difference has a probability limit of zero. We readily obtain that

$$I = \lim_{n\to\infty} \frac{1}{n\sigma^2} \begin{bmatrix} ||\hat{g} - f||^2 & (\hat{g}-f)^T F & 0 \\ F^T(\hat{g}-f) & F^T F & 0 \\ 0 & 0 & n/2\sigma^2 \end{bmatrix} \quad (15)$$

where F and f without subscript denote $F(\beta_0)$ and $f(\beta_0)$ respectively.

We obtain from (1), (14) and the inverse of I that

$$\sqrt{n}\,\hat{\alpha} \sim \sqrt{n}(\hat{g}-f)^T M_0 \varepsilon_0 / ||M_0(\hat{g}-f)||^2 \quad (16)$$

where

$$M_0 = I - F(F^T F)^{-1} F^T. \quad (17)$$

It is asymptotically correct to replace $\hat{g}$ in expression (16) by $g \equiv g(\gamma_0)$ where $\gamma_0 = \text{plim}_0 \hat{\gamma}$. This is so since γ_0 is defined by the equation

$$\lim \frac{1}{n} G^T(\gamma_0) \, (g(\gamma_0) - f(\beta_0)) = 0$$

(from the likelihood equations for the regression $y = g(\gamma) + \varepsilon$) and since a calculation similar to that above yields

$$\hat{g} \sim g - G(G^T G + \sum_i (g_i - f_i) D^2 g_i)^{-1} \, G^T(g - f - \varepsilon_0).$$

(D^2 again denotes the Hessian and $G \equiv G(\gamma_0)$). Our assumptions are sufficient to ensure that the second term here has a probability limit of zero. Hence $\sqrt{n}\,\hat{\alpha}$ is asymptotically normal with mean zero.

The estimate of the variance of $\sqrt{n}\,\hat{\alpha}$ from the regression (3) is

$$n\hat{\sigma}^2 / ||M_0(\hat{\beta}) \, (\hat{g} - f(\hat{\beta}))||^2$$

and it is clear that this tends in probability to the variance of the right-hand side of (16). We have thus proved

Lemma 1: The t-statistic for $\hat{\alpha}$ generated by regression (3) is asymptotically distributed as N(0,1) if H_0 is true.

This lemma deals with the J-test procedure. We now briefly consider the C-test procedure, which is based on regression (4). The estimate of α from (4) is

$$\hat{\alpha}_C = [(\hat{g}-\hat{f})^T (y-\hat{f})] / ||\hat{g}-\hat{f}||^2, \tag{18}$$

where $\hat{f}$ denotes $f(\hat{\beta})$ and $\hat{\beta}$ is now the ML estimate from (1). If H_0 is true, it is easy to see that $\hat{f} \sim f + (I - M_0)\varepsilon_0$, so that

$$y - \hat{f} \sim M_0\varepsilon_0. \tag{19}$$

The estimate of the variance of $\sqrt{n}\hat{\alpha}_C$ from (4) is

$$n\hat{\sigma}^2||\hat{g}-\hat{f}||^{-2} \tag{20}$$

where $\hat{\sigma}^2$ is the estimate of σ^2 from (4), which is obviously consistent under H_0. The variance estimate (20) is asymptotically biased, however. This situation has been analysed by Durbin [4], who shows how to obtain a correct estimate by the use of consistent estimators of the various components of the information matrix. Using (15) and Durbin's prescription, or alternatively by direct calculation from (18) and (19), one can easily show that a consistent estimator of the variance of $\sqrt{n}\hat{\alpha}_C$ is any estimator which converges in probability to

$$n\sigma^2||M_0(g-f)||^2 / ||g-f||^4. \tag{21}$$

Now observe that expression (21) is smaller than the probability limit of (20) under H_0, since M_0 is an orthogonal projection matrix, so that

$$||M_0(g-f)||^2 \leq ||g-f||^2.$$

Thus a crude test based on (4) will be valid in the sense that the true (asymptotic) probability of Type I error will be no greater than the size of the test.

One can of course compute a correct C-test statistic by using an estimate of (21), but it is easier to make use of the P-test procedure, which is based on regression (5). In vector notation, (5) is:

$$y - \hat{f} = \alpha(\hat{g}-\hat{f}) + \hat{F}b + \varepsilon. \tag{22}$$

To show the validity of this procedure, we make use of a theorem of Lovell [5], which can be stated as follows. The estimates of the parameters c and of their variances will be identical whether one estimates $Y = Xc + Zd + u$ or $M_Z Y = M_Z Xc + u$, where Y is a vector of dependent variables, X and Z are matrices of independent variables, u is a vector of errors, and $M_Z = I - Z(Z^TZ)^{-1}Z^T$. Identifying $\hat{g}-\hat{f}$ with X and $\hat{F}$ with Z, and noting that $\hat{M}_0(y-\hat{f}) = y-\hat{f}$ identically, we conclude that, so far as the estimate of α is concerned, (22) may be replaced by the equivalent regression:

$$y - \hat{f} = \alpha\hat{M}_0(\hat{g} - \hat{f}) + \varepsilon. \tag{23}$$

The name "P-test" derives from this regression, in which the projection matrix M_0 explicitly appears.

From (23) it is obvious that the estimate of α from (22) will be

$$\hat{\alpha}_P = (\hat{g} - \hat{f})^T\hat{M}_0(y-\hat{f}) \;/\; ||\hat{M}_0(\hat{g}-\hat{f})||^2, \tag{24}$$

and that the OLS estimate of the variance of $\sqrt{n}\,\hat{\alpha}_P$ will be

$$n\hat{\sigma}^2||\hat{M}_0(\hat{g}-\hat{f})||^{-2}, \tag{25}$$

where $\hat{\sigma}^2 = (1/(n-k-1))||y-\hat{f}-\hat{\alpha}_P\hat{M}_0(\hat{g}-\hat{f})||^2$. From (19) we may conclude that, under H_0,

$$\sqrt{n}\hat{\alpha}_P \sim \sqrt{n}\,(g-f)^TM_0\varepsilon_0/||M_0(g-f)||^2, \tag{26}$$

the variance of which is indeed the probability limit, under H_0, of the estimate (25). Since $\hat{\alpha}_P$ is obviously normally distributed with mean zero, we have proved

<u>Lemma 2</u>: The t-statistic for $\hat{\alpha}$ generated by regression (5) is asymptotically distributed as N(0,1) if H_0 is true.

Note that (16) and (26) are identical, so that, asymptotically under H_0, $\hat{\alpha}_J$ and $\hat{\alpha}_P$ will be equal, and that both will also be perfectly correlated with $\hat{\alpha}_C$ (see (18) and (19)).

If in any of the regressions (3), (4) or (5), the estimate $\hat{\alpha}$ is significantly different from unity, then one may conclude that H_1 is not sustained by the data. This follows from an argument similar to the one used above to prove that the estimate of the variance of $\hat{\alpha}$ from (4) is biased upwards. Again, the method of Durbin [4] can be used to obtain a valid variance estimate, but it will generally be simpler just to invert the roles of H_0 and H_1 in one of the regular procedures when one wishes to test the latter.

We now turn our attention to the statistic used by Pesaran and Deaton [7] for what we have called the CPD test. The numerator of their statistic is

$$T_0 = \frac{n}{2} \log (\hat{\sigma}_1^2/\hat{\sigma}_{10}^2), \quad (\text{with } \hat{\sigma}_{10}^2 \equiv \hat{\sigma}_0^2 + \hat{\sigma}_a^2). \tag{27}$$

Here $\hat{\sigma}_1^2$ is the ML estimate from the regression $y = g(\gamma) + \varepsilon$, $\hat{\sigma}_0^2$ is the ML estimate from the regression $y = f(\beta) + \varepsilon$, and $\hat{\sigma}_a^2$ is the ML estimate from an auxiliary regression $f(\hat{\beta}) = g(\gamma) + \varepsilon_a$. The assumptions made by Pesaran and Deaton are the same as ours, and their notation is only slightly different.

As the logarithm in (27) is difficult to work with, we perform a Taylor expansion of it around unity and retain only the term of leading order, so as to obtain a statistic S which is asymptotically equivalent to $T_0/\sqrt{n}$.[4] If we make the definition

$$T_0/\sqrt{n} = -(\sqrt{n}/2)\log(1-2S/\sqrt{n}) \tag{28}$$

then we see that

$$T_0/\sqrt{n} = S + 2S^2/\sqrt{n} + \ldots.$$

so that

$$S = -\frac{\sqrt{n}}{2}[(\hat{\sigma}^2_{10} - \hat{\sigma}^2_1)/\hat{\sigma}_1{}^2]$$

$$= \frac{-1}{2\sqrt{n}}[(y-\hat{f})^T(y-\hat{f}) + (\hat{f}-\tilde{g})^T(\hat{f}-\tilde{g})-(y-\hat{g})^T(y-\hat{g})]/[\frac{1}{n}(y-\hat{g})^T(y-\hat{g})] \tag{29}$$

where $\tilde{g} = g(\tilde{\gamma})$, and $\tilde{\gamma}$ is the ML estimate of γ from the auxiliary regression. It is easy to see that $\tilde{g}$, like $\hat{g}$, has probability limit $g \equiv g(\gamma_0)$, where as before $\gamma_0 \equiv \text{plim}_0\hat{\gamma}$. This implies that the auxiliary regression in the CPD procedure is quite unnecessary, since $\hat{\sigma}^2_a$ can validly be replaced by $(1/n)||\hat{g}-\hat{f}||^2$, under H_0. Of course if H_0 is not true, this replacement will yield different results. In either case, we obtain that, under H_0:

$$S \sim [(1/\sqrt{n})(f-g)^T M_0\varepsilon_0] / [\sigma^2 + (1/n)||f-g||^2]. \tag{30}$$

Comparison of (30) with (16) or (26) shows that S is asymptotically perfectly correlated with all the $\hat{\alpha}$'s, with correlation coefficient minus one.

Pesaran and Deaton give for an estimate of the variance of T_0 the expression

$$\hat{V}_0(T_0) = (\hat{\sigma}^2_0/\hat{\sigma}^4_{10})(\hat{f}-\tilde{g})^T M_0(\hat{\beta})(\hat{f}-\tilde{g})$$

and it is clear from (30) that, asymptotically, $\hat{V}_0(T_0)/n$ is equal to the variance of S. This result is noteworthy because the variance of S follows immediately from (30) from first principles, whereas Pesaran and Deaton's derivation of $\hat{V}_0(T_0)$ uses a lengthy calculation based on a general and by no means elementary result of Cox [2].

In summary, then, we have

<u>Lemma 3</u>: When H_0 is true, the N_0-statistic of Pesaran and Deaton, which is

defined as $T_0/(\hat{V}_0(T_0))^{1/2}$, is asymptotically equal to minus the J- and P-test statistics.

We now examine the power of our tests and of the CPD test. The power of a test is defined as one minus the probability of Type II error (see, for example, Silvey [9], Chapter 6). For our purposes, a Type II error is committed whenever a test fails to reject H_0 when it is false. We restrict our attention to asymptotic results for the case where H_1 is true, and shall, for the sake of brevity, consider only the P-test and the CPD test.

First, it is easy to see that the plim_1 of $\hat{\alpha}_P$ (i.e., the plim under H_1) is unity, and that the plim_1 of the estimate of the variance of $\sqrt{n}\hat{\alpha}_P$ is

$$n\sigma_1^2 \,/\, ||\, M_0(g-f)||^2,$$

where $g = g(\gamma_1)$, $\beta_1 = \text{plim}_1\hat{\beta}$, $f = f(\beta_1)$ and $M_0 = M_0(\beta_1)$. Then if we make the definition

$$U \equiv (1/n)||M_0(g-f)||^2$$

and denote the P-test statistic by N_P, we conclude that

$$\text{plim}_1\, N_P/\sqrt{n} = U^{1/2}/\sigma_1. \qquad (31)$$

A somewhat lengthier calculation gives the corresponding result for the CPD statistic, N_0:

$$-\,\text{plim}_1\, N_0/\sqrt{n} = \tfrac{1}{2}(U + V + \sigma_1^2)\,\log[1 + (U+V)/\sigma_1^2]/[W(U + \sigma_1^2)]^{1/2}, \qquad (32)$$

where we have made the definitions:

$$V \equiv (1/n)||M_1(g-f)||^2,$$

$$W \equiv (1/n)||M_0M_1(g-f)||^2,$$

$$M_1 \equiv I - G(\gamma_2)[G^T(\gamma_2)G(\gamma_2)]^{-1}\, G^T(\gamma_2),$$

with $\gamma_2 = \text{plim}_1\tilde{\gamma}$ <u>not</u> necessarily equal to γ_1.

It can readily be seen that the variances of both $N_p/\sqrt{n}$ and $N_0/\sqrt{n}$ are of order $1/n$ as $n \to \infty$. Since the expressions (31) and (32) are plainly of order unity, we can conclude that as the sample size tends to infinity, both the tests reject H_0 against H_1 with probability unity when H_1 is true. It does not appear to be possible to conclude that one test will be more powerful than the other.

There remain a great many interesting questions related to the small-sample behavior of the various tests, which we intend to examine in a future paper. For now, we merely remark that the performance of all the tests appears to be quite similar in small samples, and that the ability of the tests to reject false hypotheses, even when testing against other false hypotheses, appears to be rather good. These remarks are illustrated by the empirical results of the next section.

3. Empirical Results

In this section we apply the test procedures we have proposed to the data and models investigated by Pesaran and Deaton. They considered five simple models of the relationship between real consumption and real personal disposable income, denoted by H_1 to H_5, using U.S. quarterly seasonally adjusted data for 1954-2 to 1974-3. According to H_1, consumption depends linearly on current income and a measure of wealth; according to H_2, it depends linearly on current income and consumption lagged one period; according to H_3, it depends multiplicatively on current income and lagged consumption, but with an additive error term; according to H_4, it depends on current income and on all past income with geometrically declining weights; and according to H_5 it depends on current income and on incomes for the past 21 quarters with weights lying on a second degree polynomial. More detailed discussions of these five hypotheses, together with estimates of all of them, are provided in [7]. Since none of the hypotheses is of much economic interest, because they were deliberately kept very simple for purposes of illustration , we do not reproduce that material here.[5]

Table 1 presents the results of pairwise tests of each model, H_1 through H_5, against each of the other models. Each group of four rows relates to a particular hypothesis being tested. The first element in each off-diagonal entry is the value of the CPD N_0-statistic, from Table II of [7].[6] The second element is a test statistic for the J-test. Where the hypothesis being tested is linear (H_1, H_2 and H_5), this is simply the t-statistic associated with the estimate of α from (4). Where the hypothesis being tested is nonlinear (H_3 and H_4), this is the square root of twice the difference between the log-likelihood function for equation (4) evaluated at the maximum and evaluated at $(0,\hat{\beta})$, a quantity which is asymptotically distributed as N(0,1) if the hypothesis under test is true. We present this test statistic rather than an estimated asymptotic t-statistic because we sometimes had difficulty numerically evaluating the latter. The third element in each off-diagonal entry is the value of the t-statistic from (4), as computed by the regression package, and the fourth element is the P-test statistic. This fourth element is omitted when the hypothesis being tested is linear, since the P-test is identical to the J-test in that case.

Several features of Table 1 are worthy of note. First of all, as the fact that their asymptotic correlation is unity suggests, the J- and P-test statistics tend to be very similar. Secondly, inferences from the J- and P-tests are basically the same as inferences from the CPD test. Since we are applying asymptotic tests to estimates based on only 82 observations, let us, conservatively, take 2.5 as a critical value. Then the CPD test rejects H_0 in 12 cases out of 20, the P- and J-tests reject H_0 in 13 cases, and all three tests reject H_0 in 11 cases. In two of the three cases where our tests and the CPD test yield different inferences, the actual values of the test statistics are not that far apart, so that there is serious conflict in only one case out of 20. Another interesting feature of Table 1 is that the ordinary t-statistic from (4), although not as likely to reject hypotheses as the J- or P-tests, is nevertheless quite useful. It rejects H_0 in nine of the thirteen cases where both the other tests do so.

As one would expect from Lemma 3, a positive value of the CPD N_0-statistic is usually, but not invariably, associated with negative values of the J- and P-test statistics, and vice versa. Remember that the lemma holds only asymptotically and only if H_0 is in fact true. Thus these results emphasize the fact that, despite their perfect negative asymptotic correlation when H_0 is true, the CPD test and the J- and P-tests are different procedures, which can yield different inferences.

Finally, it is interesting to observe that large values of the J- and P-test statistics tend to be associated with extremely large values of the CPD statistic. This phenomenon has been observed in several other sets of data as well. It is presumably related to the possibility that expression (32) may be very much larger than (31) when U is large and W is small relative to U. This may be a disadvantage of the CPD procedure, because it may condition investigators to expect enormous values whenever a hypothesis is false, and to be skeptical of values between say, 2.5 and 3.5.

As noted earlier, our procedures allow one to test a hypothesis against several alternative hypotheses simultaneously, by estimating equation (6) or (7) and testing whether all of the α_j's are zero. We employ a standard likelihood-ratio test. When each of H_1 through H_5 is tested against the other four hypotheses jointly, the test statistic would be asymptotically distributed as chi-squared with four degrees of freedom if the hypothesis under test were true. These test statistics are: for H_1, 69.236; for H_2, 30.840; for H_3, 18.798 (J) and 15.884 (P); for H_4, 44.795 (J) and 59.240 (P); and for H_5, 83.080. Since the .005 critical value for $\chi^2(4)$ is 14.86, it is clear that all five hypotheses must be rejected, most at an extremely high level of significance.

Conclusion

In this paper we have proposed several new procedures for testing the validity of regression models, provided there exist non-nested alternative hypotheses. These tests behave very much like the existing CPD test, except that they less often produce enormous test statistics. They are remarkably simple to compute. When H_0 is linear, one merely has to run one extra linear regression to test it. When H_0 is nonlinear, one either has to run one extra nonlinear regression (for the J-test), or calculate the derivatives of the model evaluated at $\hat{\beta}$ and run one extra linear regression (for the P-test). Since the tests are trivially easy to implement, and since finding alternative models is rarely difficult, there would appear to be no barrier to their widespread use in applied econometric work.

TABLE 1

Pairwise Tests for H_1 through H_5

Alternative Hypothesis:		H_1	H_2	H_3	H_4	H_5
Tested hypothesis:	H_1	-233.45	-47.08	-29.30	-28.30	0.71
			6.84	7.10	4.78	1.34
			6.96	7.22	4.33	0.72
	H_2	0.37	-214.25	-3.38	-2.58	0.96
		-0.40		3.29	1.97	3.75
		-0.10		2.54	1.39	0.59
	H_3	1.08	2.68	-213.15	-1.86	1.43
		-1.19	-2.91		1.54	1.86
		-0.37	-2.03		1.05	0.32
		-0.95	-2.88		1.39	1.03
	H_4	2.19	-11.20	-12.09	-225.13	1.31
		-2.65	5.08	5.22		1.80
		-0.67	5.21	5.38		0.43
		-3.06	5.34	5.51		1.78
	H_5	-14.89	-73.04	-39.66	-121.3	-233.97
		4.87	8.72	8.96	5.63	
		1.25	7.11	7.33	4.44	

Entries on the diagonal are log L. The first element in each off-diagonal entry is the value of the CPD statistic, the second is the J-test statistic, the third is the ordinary t-statistic from (4) and the fourth is the P-test statistic, for H_3 and H_4 only.

Footnotes

1. We would like to thank Angus Deaton, Gordon Fisher, Bentley Macleod, Michael McAleer and Christopher Sims for helpful comments on earlier drafts. Versions of this paper have been presented in seminars at Queen's University, the University of British Columbia, Stanford University and the University of California at Berkeley. If any errors survive, we are responsible for them. This research was supported, in part, by a grant from the Social Sciences and Humanities Research Council of Canada.

2. It was pointed out to us by a referee that Atkinson's compound model procedure suffers from the difficulty that α and γ_1 cannot be separately identified. Our procedures circumvent this difficulty by estimating α conditional on $\hat{\gamma}_1$.

3. These technical requirements are imposed to avoid difficulties associated with unidentified models or with statistics that have infinite variances. Strictly speaking, we must also exclude the following possibility: the span of the columns of F and G have an interesection, V, of positive dimension, and the orthogonal complements of V in the respective spans of the columns of F and G are themselves orthogonal. We are indebted to W. Bentley Macleod for this point.

4. Under H_0, it is $T_0/\sqrt{n}$ which is of order unity in probability.

5. We are grateful to Professor Angus Deaton for supplying us with the data used by Pesaran and Deaton, which differ from the data published with that article because they have one more significant digit. Using the former, we were able essentially to reproduce the estimates reported by Pesaran and Deaton.

6. Actually, the N_0-statistics reported in the last column of Table 1 differ from those reported by Pesaran and Deaton because the latter were apparently computed incorrectly.

References

[1] Atkinson, A.C.: "A Method for Discriminating Between Models," *Journal of the Royal Statistical Society*, Series B, 32 (1970), 323-344.

[2] Cox, D.R.: "Tests of Separate Families of Hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley: University of California Press, 1961.

[3] __________: "Further Results on Tests of Separate Families of Hypotheses," *Journal of the Royal Statistical Society*, Series B, 24 (1962), 406-424.

[4] Durbin, J.: "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38 (1970), 410-421.

[5] Lovell, M.C.: "Seasonal Adjustment of Economic Time Series," *Journal of the American Statistical Association*, 58 (1963), 993-1010.

[6] Pesaran, M.H.: "On the General Problem of Model Selection," *Review of Economic Studies*, 41 (1974), 153-171.

[7] Pesaran, M.H., and A.S. Deaton: "Testing Non-nested Nonlinear Regression Models," *Econometrica*, 46 (1978), 677-694.

[8] Sawa, T.: "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, 46 (1978), 1273-1291.

[9] Silvey, S.D.: *Statistical Inference*, London: Chapman and Hall, 1975.