QUEENS

ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH

QUEEN'S UNIVERSITY

Kingston, Ontario, Canada K7L 3N6

SOME NON-NESTED HYPOTHESIS TESTS AND THE RELATIONS AMONG THEM

Russell Davidson
and
James G. MacKinnon

Discussion Paper #409

Queen's University

November, 1980

ABSTRACT

In this paper we discuss several statistical techniques which may be used to test the validity of a possibly nonlinear and multivariate regression model, using the information provided by estimating one or more alternative models on the same set of data. The first such techniques in econometrics were proposed by Pesaran (1974) and Pesaran and Deaton (1978), based on the work of Cox (1961, 1962). In Davidson and MacKinnon (1980), we recently proposed, for the univariate case, some new techniques which are conceptually and computationally simpler. The first major result of this paper is that the techniques we have proposed can be regarded as alternative implementations of Cox's basic idea for non-nested hypothesis testing; under the null hypothesis all of the test statistics are asymptotically the same random variable. A second major result is that, for the univariate linear regression case, our tests and Pesaran's test have asymptotic relative efficiency of unity for local alternatives. We then propose several generalizations of our procedures to the case of multivariate regression models, and show that one of these generalizations is asymptotically equivalent under the null hypothesis to the test proposed by Pesaran and Deaton. Finally, we present the results of a sampling experiment for univariate linear models which shows that the small-sample performance of our J-test and Pesaran's test can be quite different.

1. Introduction

Economic theory typically suggests not one but a multiplicity of models that might explain any given phenomenon. Only a small fraction of these can reasonably be dealt with in any particular piece of empirical work. It is therefore important that the applied econometrician have available not only techniques which allow him to choose which of the available models is the best, but also ones which can allow him to decide whether any of the available models is satisfactory. Conventional nested hypothesis testing is not always adequate here. Since there must always be some maintained hypothesis, it allows of no formal test by which even a maintained hypothesis that is plainly wrong can be rejected. All one can do is reject a model against a more general one, and perhaps decide that the latter is unsatisfactory because it makes no economic sense. In this paper we are concerned with less conventional procedures for testing non-nested hypotheses, which have the property that any or all of the models in a given set may be rejected.

The first procedure of this type was introduced by Cox (1961, 1962) as a generalization of the likelihood ratio test. Cox's idea was that one may test the validity of a hypothesis, H_0, about how a set of data was generated, by comparing the value, from the data, of the likelihood function for some alternative hypothesis with an estimate of the expected value of this likelihood function if H_0 were true. This idea was not implemented in econometrics until Pesaran (1974) showed how it could be applied to linear regression models. Subsequently, Pesaran and Deaton (1978) (hereafter PD) extended Pesaran's technique to deal with nonlinear and multivariate regression models. The procedure they describe will be referred to as the Cox-Pesaran-Deaton or CPD procedure. It allows one to test whether the truth of one model can be maintained, given the performance of an alternative model. The roles of the two models can of course be reversed, and it is entirely possible that both (or neither) may be rejected.

Two other tests for the univariate regression case, with the same purpose and substantially the same properties, were recently proposed by us in Davidson and MacKinnon (1980), hereafter DM. These tests, which we called the J-test and the P-test, will be described in Section 2 of the paper. They are conceptually simpler than the CPD test, and easier to implement with existing computer software. Moreover, both the new tests can easily be extended so that a model may be tested against several alternatives simultaneously.

In Section 2 of the paper we describe the tests proposed by DM, and show that they may be regarded as developments of the artificial nesting procedure of Atkinson (1970), which solve the identification problem normally associated with such procedures. We then demonstrate that our P-test could alternatively have been developed as a way to implement Cox's idea for non-nested hypothesis testing, and show that under the null hypothesis all the test statistics are asymptotically the same random variable.

In Section 3 of the paper we consider what happens when the alternative hypothesis is true. For the case of univariate linear models we are able to show that, among the CPD test and our tests, any test compared with any other has asymptotic relative efficiency of unity for local alternatives.

In Section 4 we propose several generalizations of our procedures to the case of multivariate regression models. One of these generalizations turns out to be asymptotically equivalent, under the null hypothesis, to the CPD test. Since the computational advantages of our procedure relative to the CPD procedure are much greater in the multivariate case than in the univariate one, the former should be very useful in applied work.

Finally, in Section 5, we present the results of a sampling experiment in which we compare the small-sample performance of our J-test and Pesaran's test for univariate linear models. It turns out that under the null hypothesis neither test statistic is very close to N(0,1) when the sample size is very small or the variance of the error term is large. However, inferences from the J-test are much more reliable in such cases than inferences from Pesaran's test, because the density of the latter seems to have much thicker tails. Except in cases of extremely small sample size and large variance, both tests seem to have good power.

2. The P-test and the CPD Test

Throughout this section, we shall consider the alternative non-nested and in general nonlinear univariate regression models,

$$H_0: \quad y_t = f(X_t, \beta) + \varepsilon_{0t} \tag{2.1}$$

$$H_1: \quad y_t = g(Z_t, \gamma) + \varepsilon_{1t} \,. \tag{2.2}$$

The y_t (t = 1 to n) are observations on a dependent variable, and the X_t and Z_t are nonstochastic vectors of observations on independent variables, assumed fixed in repeated samples. The two hypotheses are compound, with respectively a k-vector β and an ℓ-vector γ of parameters to be estimated.

If H_0 is true, the ε_{0t} are NID$(0,\sigma_0^2)$; if H_1 is true, the ε_{1t} are NID$(0,\sigma_1^2)$. The functions f and g are assumed to be twice continuously differentiable with respect to β and γ respectively, with first partial derivatives denoted by $F(\beta)$ and $G(\gamma)$, which are respectively $n \times k$ and $n \times \ell$ matrices, with transposes $F^T(\beta)$ and $G^T(\gamma)$. It is further assumed that, as $n \to \infty$, $(1/n)F^T(\beta)F(\beta)$, $(1/n)G^T(\gamma)G(\gamma)$ and $(1/n)F^T(\beta)G(\gamma)$ all converge to well-defined finite limits for all bounded β and γ, the first two being positive definite and the third non-zero.

The tests proposed by DM can be constructed in three separate steps. The first of these is to nest H_0 and H_1 in an artificial compound model. At least three such models might seem reasonable. The simplest of them is

$$y = (1 - \alpha)f(\beta) + \alpha g(\gamma) + \varepsilon \tag{2.3}$$

where y, f and g now denote vectors and X and Z have been suppressed for notational convenience. A second compound model is

$$y = \left(\frac{(1-\lambda)\sigma_1^{\,2}}{(1-\lambda)\sigma_1^{\,2} + \lambda\sigma_0^{\,2}}\right)f(\beta) + \left(\frac{\lambda\sigma_0^{\,2}}{(1-\lambda)\sigma_1^{\,2}+\lambda\sigma_0^{\,2}}\right)g(\gamma) + \varepsilon\,. \tag{2.4}$$

This model, which has been investigated by Atkinson (1970), among others, has a likelihood function which is the same as an exponential combination of the likelihood functions of H_0 and H_1, with weights $(1-\lambda)$ and λ respectively. A third compound model is

$$y = \left(\frac{(1-\mu)\sigma_1}{(1-\mu)\sigma_1+\mu\sigma_0}\right)f(\beta) + \left(\frac{\mu\sigma_0}{(1-\mu)\sigma_1 + \mu\sigma_0}\right)g(\gamma) + \varepsilon\,. \tag{2.5}$$

If H_0 is expressed in the form $(y - f(\beta))/\sigma_0 = u_0$, then if H_0 were true u_0 would be distributed as NID(0,1). Model (2.5) results from combining H_0 with H_1 in this form with weights $(1-\mu)$ and μ respectively, and then solving for y.

It is clear that by suitable reparametrizations (2.4) and (2.5) can both be put into the form of (2.3), so that the latter is the only compound model we need to consider. As we shall see in Section 4, however, that will not be true for the multivariate analogues of (2.4) and (2.5). It should also be clear that, in general, the parameters α, β and γ of (2.3) will not all be identified. In order to test the truth of H_0, we wish to test the hypothesis that $\alpha = 0$. That will not be possible if α is not identified.

The second step in DM's construction of a test procedure is designed to get around this problem.[1] To ensure that α is identified, γ is replaced by its maximum likelihood estimate $\hat{\gamma}$. Thus one way to test the validity of H_0 is to estimate the possibly nonlinear regression

$$y = (1-\alpha)f(\beta) + \alpha\hat{g} + \varepsilon, \tag{2.6}$$

where $\hat{g}$ denotes $g(\hat{\gamma})$. This procedure was called the J-test by DM. The test statistic is simply the ordinary t-statistic for a test of $\alpha = 0$, which DM prove is asymptotically $N(0,1)$ under H_0.

It is clear that (2.6) would still yield a valid test statistic if $\hat{g}$ were replaced by <u>any</u> vector which is asymptotically non-stochastic, such as $g(\gamma')$ where γ' is some other estimate of γ. In particular, following Atkinson (1970), one might want to use an estimate of the expectation of $\hat{\gamma}$ under H_0. Pesaran (1980) considers this and several other choices for γ as ways of modifying the J-test for linear models. His results do not suggest that there would be any gain from using estimates other than $\hat{\gamma}$, and in Section 3 we shall verify that, by the criterion of asymptotic relative efficiency, there is indeed no gain.

If f is linear in β, performing a J-test simply requires that one estimate H_1 and then compute a single linear regression. If f is nonlinear, however, (2.6) will be a nonlinear regression. The third step in DM's construction of a test procedure is to linearize the J-test regression (2.6) around the point $(\alpha=0,\beta=\hat{\beta})$. This yields the linear regression

$$y - \hat{f} = \hat{F}b + \alpha(\hat{g} - \hat{f}) + \varepsilon, \tag{2.7}$$

where $\hat{f} = f(\hat{\beta})$, $\hat{F} = F(\hat{\beta})$ and $b = (1 - \alpha)\beta$. As Durbin (1970) has shown, a Wald or likelihood ratio test of the hypothesis $\alpha = 0$ based on the linearized regression (2.7) will be asymptotically equivalent to one based on the nonlinear regression (2.6), under the null hypothesis and for local alternatives. DM call the t-test of $\alpha = 0$ in (2.7) the P-test. It is obvious that if f is in fact linear, the P-test and J-test will yield identical results.

The OLS estimate of α from (2.7) is

$$\hat{\alpha}_p = [(\hat{g}-\hat{f})^T\hat{M}_0(y-\hat{f})]/[(\hat{g}-\hat{f})^T\hat{M}_0(\hat{g}-\hat{f})] \tag{2.8}$$

where

$$\hat{M}_0 = I - \hat{F}(\hat{F}^T\hat{F})^{-1}\hat{F}^T. \tag{2.9}$$

It is easily shown (see DM) that the t-statistic for $\hat{\alpha}_p$ is N(0,1) asymptotically if H_0 is true. It is also easy to see that since $(y-\hat{f})$ is orthogonal to $\hat{F}$, $\hat{M}_0(y-\hat{f}) = (y-\hat{f})$, and hence the numerator of $\hat{\alpha}_p$ is simply

$$(\hat{g}-\hat{f})^T(y-\hat{f}). \tag{2.10}$$

Thus the P-test (and also the J-test) really just amounts to testing whether the residuals from H_0 are significantly different from being orthogonal to the difference between the fitted values from H_1 and H_0.

We now turn our attention to the CPD statistic, the numerator of which is given by

$$T_0 = (n/2) \log (\hat{\sigma}_1^2/\hat{\sigma}_{10}^2) \tag{2.11}$$

where

$$\hat{\sigma}_{10}^2 = \hat{\sigma}_0^2 + \hat{\sigma}_a^2. \tag{2.12}$$

Here $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$ are the ML estimates of the regression variances for (2.1) and (2.2) respectively, and $\hat{\sigma}_a^2$ is the ML estimate of the regression variance for an auxiliary nonlinear regression

$$\hat{f} = g(\gamma) + \varepsilon_a. \tag{2.13}$$

If H_0 is true, $(1/\sqrt{n})T_0$ will be of order unity and will be asymptotically normally distributed with mean zero. Following Cox, PD obtain the following estimate of its variance:

$$V_0(T_0) = (\hat{\sigma}_0^2/\hat{\sigma}_{10}^4)(\hat{f} - \tilde{g})^T\hat{M}_0(\hat{f} - \tilde{g}), \tag{2.14}$$

where $\tilde{g}$ denotes the fitted values from regression (2.13). It is then straightforward to compute the test statistic $N_0 = T_0/\sqrt{V_0}$.

While this is certainly a valid way to implement Cox's idea for non-nested hypothesis testing in the context of nonlinear univariate regression models, it is not the only way to do so. Cox provides a series of formulae which involve true parameters or probability limits, and these must be replaced by consistent estimates to obtain a useful test statistic. Using different consistent estimators will produce different test statistics, asymptotically equivalent under H_0 but possibly different under H_1. For example, observe that, as $n \to \infty$, the ML estimates $\hat{\gamma}$ will in general tend under H_0 to a well-defined limit $\bar{\gamma}$ defined implicitly by the equation

$$\lim_{n\to\infty} (1/n)\, G^T(\bar{\gamma})(f(\beta_0) - g(\bar{\gamma})) = 0, \tag{2.15}$$

where β_0 is the true value of β in (2.1).[2] It is easily seen that the estimates $\tilde{\gamma}$ from the auxiliary regression (2.13) will also tend to $\bar{\gamma}$ under H_0 but not necessarily, of course, under other hypotheses. Thus the auxiliary regression (2.13) is unnecessary. We may replace $\hat{\sigma}_a^2$ in (2.12) by $(\hat{g} - \hat{f})^T(\hat{g} - \hat{f})/n$ to obtain a new statistic T_0', and $\tilde{g}$ by $\hat{g}$ in (2.14) when computing its variance. The resulting test statistic, N_0', is just as much an implementation of Cox's idea as is N_0, and since both are functions of consistent estimates they must be asymptotically identical under H_0.

Eliminating the auxiliary regression can simplify the procedure substantially in some cases. Suppose, for example, that H_1 is

$$y_t = Z_t\gamma + \rho\, y_{t-1} - \rho Z_{t-1}\gamma + \varepsilon_t, \tag{2.16}$$

that is, a linear regression with AR(1) errors. In order to compute the auxiliary regression, one must replace y_t in (2.16) by $\hat{f}_t$, but one must <u>not</u> replace y_{t-1} by $\hat{f}_{t-1}$. Since most computer programs for estimating regression models with serial correlation would automatically use $\hat{f}$ lagged on the right hand side if $\hat{f}$ were the dependent variable, computation of the auxiliary regression might well require a good deal of additional effort on the part of the investigator.

We now consider a statistic S' defined by

$$S' = (\sqrt{n}/2)(1 - \exp(-2T_0'/n)) \tag{2.17}$$

which is simply a first-order Taylor series approximation of $T_0'/\sqrt{n}$ around zero. Clearly S' and $T_0'/\sqrt{n}$ are asymptotically equal if H_0 is true. Moreover, S' and T_0' are numerically related in a one-to-one fashion: S' is always less than or equal to $T_0'/\sqrt{n}$, but will be greater in absolute value if $T_0'/\sqrt{n}$ is negative.

It is straightforward to show that

$$S' = -\sqrt{n}\,[(\hat{g}-\hat{f})^T\,(y-\hat{f})]\;/\;[(y-\hat{g})^T\,(y-\hat{g})]. \tag{2.18}$$

The numerator of S', except for the minus sign, is identical to the numerator of $\sqrt{n}\hat{\alpha}_p$ in (2.10). Because both $(y-\hat{g})^T(y-\hat{g})/n$ and $(\hat{g}-\hat{f})^T\hat{M}_0(\hat{g}-\hat{f})/n$ have non-stochastic non-zero probability limits under either H_0 or H_1, it follows that S' and $\sqrt{n}\hat{\alpha}_p$ yield exactly the same tests, asymptotically. Thus the P-test may be regarded as yet another test based on Cox's basic idea.

3. Asymptotic Relative Efficiency of Alternative Tests

In this section we investigate the power of the P-test and the CPD test. We restrict our attention to the case where both models are linear and "close" to each other, the sample size is very large, and the alternative H_1 is in fact true. Despite these restrictive assumptions, the analysis is by no means easy. We make use of the concept of asymptotic relative efficiency (ARE) as defined by Kendall and Stuart (1967), which seems to be the most natural way to compare power in the asymptotic regime.

ARE requires the existence of a sequence of alternative hypotheses, H_θ, which approach the null hypothesis H_0 as $\theta \to 0$. Usually, θ is a parameter in a comprehensive model in which both H_θ and H_0 are nested, so that the sequence of local alternatives is easily constructed. With a non-nested hypothesis test, that is not the case, and the construction is consequently somewhat harder.

We assume that the hypotheses H_0 and H_1 are both linear, so that (2.1) and (2.2) may be rewritten in vector notation as [3]

$$H_0: \quad y = X\beta + \varepsilon_0 \tag{3.1}$$

$$H_1: \quad y = Z\gamma + \varepsilon_1. \tag{3.2}$$

Since these hypotheses are compound (i.e., β and γ are not specified), each is completely specified by the linear span of the columns of X or Z. The linear span of the columns of an n x k matrix X of rank k is in turn completely characterized by the orthogonal projection onto it. For X and Z these projections are

$$M_0^\perp \equiv I - M_0 \equiv X(X^TX)^{-1}X^T$$

and

$$M_1^\perp \equiv I - M_1 \equiv Z(Z^TZ)^{-1}Z^T.$$

Thus any geometric measure of the distance between the projections $M_0^\perp$ and $M_1^\perp$ can serve as a measure of the distance between H_0 and H_1.

In fact, the most suitable such measure will be the norm of the matrix $M_1^\perp M_0 M_1^\perp$. This matrix is evidently symmetric and non-negative definite, so its norm is just its largest eigenvalue, which cannot exceed unity. Two cases are of particular interest. If H_1 becomes nested in H_0 (or becomes equivalent to H_0), then $M_0^\perp M_1^\perp = M_1^\perp$, or, equivalently, $M_1^\perp M_0 M_1^\perp = 0$. On the other hand, if ℓ, the number of columns in Z, exceeds k, the number of columns in X, there must exist a linear subspace, of dimension at least $\ell - k$, of $n \times 1$ vectors which lie in the range of both M_0 and $M_1^{\perp}$. Thus, whether or not H_0 is nested in H_1, so long as $\ell > k$, there exists a vector v such that $M_0 v = M_1{}^{\perp} v = M_1^{\perp} M_0 M_1{}^{\perp} v = v$, so that $||M_1^{\perp} M_0 M_1^{\perp}|| = 1$.

We are now ready to construct our sequence of local alternatives. For each sample size, n, let H_0 and H_1 respectively define linear subspaces in $\mathbb{R}^n$ by the projection matrices $M_0{}^{(n)}$ and $M_1{}^{(n)}$. If $\ell \leq k$, $M_1(n)$ is to be chosen so as to satisfy the following conditions:

(i) $||M_1^{\perp}\ M_0\ M_1^{\perp}|| > 0$ and, as $n \to \infty$, $||M_1^{\perp}\ M_0\ M_1^{\perp}|| \to 0$;

(ii) the rank of $M_1^{\perp}\ M_0\ M_1^{\perp}$, r, remains constant as $n \to \infty$, with $0 < r \leq k$.

In writing conditions (i) and (ii) we have suppressed the explicit dependence of $M_1^{\perp}$ and M_0 on n to simplify notation. Here, and subsequently, M_0 denotes $M_0{}^{(n)}$, and so on. Condition (i) means that H_1 approaches H_0 but is never nested in it for finite n, while condition (ii) means that the eigenvalues of $M_1^{\perp}\ M_0\ M_1^{\perp}$ are either nonzero or identically zero for all finite n.

If $\ell > k$, $M_1{}^{(n)}$ is chosen as follows. Let the range of $M_1^{\perp}$, which is an ℓ-dimensional subspace of $\mathbb{R}^n$, be expressed as the direct sum of the space of vectors v for which $M_1^{\perp} M_0 M_1{}^{\perp} v = v$ and the orthogonal complement of that space. We denote the latter by $C^{(n)}$, and the restriction of $M_1^{\perp} M_0 M_1^{\perp}$ to C by $N_1{}^{(n)}$. Then as $n \to \infty$ we require that $||N_1|| \to 0$, with the (nonzero) rank of N_1 remaining constant.

In order that the local alternatives actually approach the null hypothesis, we must impose the conditions

$$\begin{aligned} &||X^{(n)}\beta - Z^{(n)}\gamma^{(n)}|| > 0 \\ &||M_0^{(n)}\ Z^{(n)}\ \gamma^{(n)}|| > 0 \qquad (3.3) \\ \text{and}\quad &||X^{(n)}\beta - Z^{(n)}\gamma^{(n)}|| \to 0 \text{ as } n \to \infty. \end{aligned}$$

Here we have not suppressed the dependence on n because β, uniquely, does not depend on n. For $\ell = k$, condition (3.3) presents no difficulty. For $\ell < k$, clearly $M_1^{\perp}$ must be chosen in such a way that the distance between the range of $M_1^{\perp}$ and the vector $X\beta$ tends to zero for large n. For $\ell > k$, we may without significant loss of generality require that $Z\gamma \in C$, since non-zero vectors v with $M_1^{\perp}M_0M_1^{\perp}v = v$ must always be at a positive distance from any vector like $X\beta$ in the range of $M_0^{\perp}$.

In addition, we assume for simplicity that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Allowing instead the weaker condition that $\sigma_1^2 \to \sigma_0^2$ as $n \to \infty$ would not change our results in any way. Finally, we impose the condition that

$$\sigma^{-1}||Z^{(n)}\gamma^{(n)}|| = \mu\, n^{1/2} \qquad (3.4)$$

for all n and some constant μ. In view of our assumptions that $(1/n)X^TX$ and $(1/n)Z^TZ$ tend to finite, nonzero limits as $n \to \infty$, this involves no loss of generality.

The next step in the determination of the ARE of the P-test and CPD test is to obtain the expectations and variances of the two test statistics under the sequence of local alternatives. Let us denote the two test statistics by N_p and N_{CPD} respectively. Under H_0, these are both N(0,1) asymptotically. It is shown in DM that under the alternative H_1

$$E_1(n^{-1/2}\ N_p) = \sigma^{-1}U^{1/2} + O(n^{-1/2}), \text{ and}$$

$$- E_1(n^{-1/2} N_{CPD}) = (1/2)(U + V + \sigma^2)[W(U + \sigma^2)]^{-1/2} \log[1 + (U + V)/\sigma^2] + 0\,(n^{-1/2}) \tag{3.5}$$

where

$$U = (1/n)||M_0 Z\gamma||^2$$

$$V = (1/n)||M_1 M_0 Z\gamma||^2$$

$$W = (1/n)||M_0 M_1 M_0 Z\gamma||^2.$$

It is easy to see that

$$U \geq V \geq W. \tag{3.6}$$

Further, since $M_1^{\perp} Z = Z$, and using (3.4),

$$0 < ||M_0 Z\gamma|| = (\gamma^T Z^T M_1^{\perp} M_0 M_1^{\perp} Z\gamma)^{1/2} \leq ||M_1^{\perp} M_0 M_1^{\perp}||^{1/2} n^{1/2} \mu\sigma. \tag{3.7}$$

It is now convenient to introduce the parameter θ by which we shall index our sequence of local alternatives. We choose to define $\theta = \sigma^{-1} U^{1/2}$, so that by (3.6)

$$0 < \theta \leq \mu ||M_1^{\perp} M_0 M_1^{\perp}||^{1/2}. \tag{3.8}$$

Thus as $n \to \infty$ and $||M_1^{\perp} M_0 M_1^{\perp}|| \to 0$, $\theta \to 0$ through strictly positive values.

Next we show that, as $\theta \to 0$,

$$V = U(1 + o(1)) \text{ and } W = U(1 + o(1)), \tag{3.9}$$

where $o(1)$ denotes a quantity which tends to zero as $\theta \to 0$. Let us denote by $e_1 \ldots e_m$ those eigenvalues of $M_1^{\perp} M_0 M_1^{\perp}$ (or of N, if $\ell > k$) which correspond to eigenvectors in the range of $M_1^{\perp}$ (or of N). Here $m = \min(\ell, k)$ and clearly all other eigenvalues are zero. Let $e_1 = ||M_1^{\perp} M_0 M_1^{\perp}||$ be the largest eigenvalue, and let e_1 through e_r be the nonzero eigenvalues.

The n x 1 eigenvectors corresponding to e_1 to e_m may be denoted by z_1 to z_m. Then we may express the vector $n^{-1/2} Z\gamma$ as $\Sigma_{i=1}^{m} g_i z_i$. We observe from (3.7) that the squared weights g_i^2 sum to $\mu^2\sigma^2$, independent of n.

It is now easy to see that

$$U = \Sigma_i \; g_i^2 e_i$$

$$V = \Sigma_i \; g_i^2 (e_i - e_i^2) \tag{3.10}$$

$$W = \Sigma_i \; g_i^2 (e_i - 2e_i^2 + e_i^3)$$

Thus,

$$(U-V)/V = (\Sigma_i \; g_i^2 \; e_i^2)/(\Sigma_i \; g_i^2 \; e_i)$$

$$= e_1 [g_1^2 + \Sigma_{i=2}^{r} \; g_i^2 (e_i/e_1)^2]/[g_1^2 + \Sigma_{i=2}^{r} \; g_i^2 (e_i/e_1)] \leq e_1. \tag{3.11}$$

The final inequality in (3.11) can legitimately be inferred because our assumptions exclude the possibility of either numerator or denominator being zero. In view of (3.6) we shall have proved (3.9) when we establish that $\theta \to 0$ implies $e_1 \to 0$. If not, then there must exist a $\delta > 0$ and a subsequence $\{n_k\}$, k = 1,2, ... of the integers such that $\theta^{(n_k)} \to 0$ as $k \to \infty$ while $e_1^{(n_k)} \geq \delta$. However, as $k \to \infty$ and hence $n \to \infty$, $e_1^{(n)} \to 0$, so that only a finite number of the $e_1^{(n_k)}$ can be equal to or greater than δ. But from (3.3) we see that $\theta^{(n)}$ is always strictly positive, so that no finite subsequence of θ's can tend to zero. This contradiction finishes the proof of (3.9).

We may now use this result to rewrite the equations (3.5) in the form

$$E_\theta \; (n^{-1/2} \; N_p) = \theta + 0(n^{-1/2})$$

$$- E_\theta\,(n^{-1/2}\,N_{CPD}) = \theta[1 + o(1)] + 0(n^{-1/2}) \tag{3.12}$$

where $E_\theta(..)$ denotes an expectation calculated under the hypothesis in the sequence of alternatives which is indexed by θ.

For $\theta = 0$ we know that $Var(N_P) = Var(N_{CPD}) = 1 + 0(n^{-1/2})$. Because θ is a positively linearly homogeneous function of the components of the vector $Z\Upsilon$, and both N_P and N_{CPD} depend differentiably on $Z\Upsilon$, we may conclude that in a neighbourhood of $\theta = 0$, $Var_\theta\,(N_P)$ and $Var_\theta(N_{CPD})$ can be expressed as

$$Var_\theta(N_i) = 1 + 0(\theta) + 0(n^{-1/2}), \tag{3.13}$$

where N_i denotes either N_P or N_{CPD}. This differentiable dependence also ensures that the terms $0(n^{-1/2})$, here and in equations (3.11), are uniform in θ.

These remarks are sufficient to establish the regularity conditions of Kendall and Stuart, namely

$$\lim_{\theta\to 0}\ [\tfrac{\partial}{\partial\theta} E_\theta(N_i)] \ / \ [\tfrac{\partial}{\partial\theta} E_\theta(N_i)]_{\theta=0} = 1$$

and

$$\lim_{\theta\to 0}\ \frac{Var_\theta(N_i)}{Var_0(N_i)} = 1.$$

Following Kendall and Stuart, we find that the ARE of the CPD test compared with the P-test is

$$\left[\frac{\frac{\partial}{\partial\theta}\,E_\theta(N_{CPD})/Var_\theta(N_{CPD})|_{\theta=0}}{\frac{\partial}{\partial\theta}\,E_\theta\,(N_P)/Var_\theta\,(N_P)|_{\theta\,=\,0}}\right]^{1/\delta}$$

where the exponent δ is defined by the asymptotic relation

$$\frac{\partial}{\partial\theta}\,E_\theta(N_i)/Var_\theta\,(N_i)|_{\theta=0} \sim K\,n^\delta\ ,$$

K being a constant. From equations (3.12) and (3.13), we see that K = 1 and $\delta = 1/2$ for both the statistics, so that the ARE is just unity. There is thus no reason to believe, from this analysis, that the P-test is more or less powerful than the CPD test.

Pesaran (1980) has pointed out that the J-test could be modified by using some other estimator for γ rather than $\hat{\gamma}$, and has suggested several such modified J-tests for the case of linear models. It is straightforward to compare these modified J-tests to the original J-test (P-test) according to the criterion of ARE, using the same techniques used above. It turns out that for two of the tests the ARE is unity, and for one of them it is less than unity. Thus this analysis suggests that there is nothing to gain by using a more complicated estimator than $\hat{\gamma}$.

We conclude that asymptotic analysis of power, at least according to the ARE criterion, provides no basis for believing that the CPD test is more or less useful than the P-test, or indeed any of Pesaran's modified J-tests. If we are to choose between the tests on grounds of their statistical properties, we shall have to learn more about their performance in small samples. This problem is tackled in section 5, below, where we report the results of some sampling experiments.

4. Testing Multivariate Models

In this section we propose three different generalizations of the P-test for the case of multivariate models; these emerge from three different artificial compound models. Each of these tests merely requires one GLS regression, so that they are far more straightforward than the multivariate

CPD test. We then show that one of these P-tests is asymptotically equivalent to the CPD test under the null hypothesis.

We shall be concerned with two non-nested multivariate models, which may be either a set of (in general nonlinear) seemingly unrelated equations, or the restricted reduced form of a simultaneous equations model. The two models may be written as:

$$H_0: \quad y_{it} = f_{it}(X_t, \beta) + \varepsilon_{it} \tag{4.1}$$

$$H_1: \quad y_{it} = g_{it}(Z_t, \gamma) + \varepsilon_{it} \tag{4.2}$$

where i (=1 to m) is the index of the equation and t (=1 to n) is the index of the observation. For given t, the ε_{it} (whose 0 or 1 subscripts have been dropped for convenience) are assumed to be multivariate normal with covariance matrix Ω_0 or Ω_1, and serially independent. The notation for the independent variables and parameters is unchanged from the univariate case.

It is convenient at this point to introduce further notation which we shall use throughout this section. In order to simplify as much as possible the algebra that is inevitable in multivariate analysis, we make use of the Einstein summation convention for indices. With this convention, any index repeated in a term is to be summed over, provided that one occurrence is a subscript and another is a superscript. In this way it is possible to represent operations involving vectors and matrices easily, without, for example, needing to resort to Kronecker product notation. Indices from a through ℓ will refer to equations and will be summed from 1 to m; indices from p through v will refer to observations and will be summed from 1 to n; and Greek indices will be used to index the parameters of models (that is, the elements of β or γ, as indicated by context). If ω_{ab} denotes an element of some covariance matrix Ω and ω^{ab} denotes the same element of Ω^{-1}, we can write

$$\omega^{ab}\,\omega_{bc} = \delta^{a}_{c},$$

where δ is the Kronecker delta, equal to unity if its indices are the same, and zero otherwise. In order to effect summations over observations, we shall make no distinction between subscripts and superscripts from p through v.

The first step in the construction of a P-test is to nest H_0 and H_1 in an artificial compound model. As in the univariate case, at least three such models seem reasonable, but in the multivariate case they are not equivalent. The simplest compound model, analogous to (2.3), is:

$$y_{it} = (1-\alpha)\, f_{it}(\beta) + \alpha g_{it}(\gamma) + u_{it}. \tag{4.3}$$

A more complicated model, analogous to (2.4), arises if we combine the likelihood functions for H_0 and H_1 exponentially with weights $(1-\lambda)$ and λ respectively. Pesaran (1980) shows that this yields the following compound model:

$$y_{it} = (1-\lambda)\omega_{ij}(\lambda)\omega_0^{jk}\, f_{kt}\,(\beta) + \lambda\omega_{ij}(\lambda)\omega_1^{jk}\, g_{kt}(\gamma) + u_{it}. \tag{4.4}$$

Here $\omega_{ij}(\lambda)$ is an element of the covariance matrix of the u_{it}'s and is defined by

$$\omega^{ij}(\lambda) = (1-\lambda)\omega_0^{ij} + \lambda\omega_1^{ij},$$

where ω_0^{ij} and ω_1^{ij} are elements of Ω_0^{-1} and Ω_1^{-1} respectively. Thus the covariance matrix of the u_{it}'s is the inverse of a convex combination of the inverse covariance matrices under H_0 and H_1.

A third model, analogous to (2.5), arises if we first transform H_0 and H_1 so that their error terms are distributed as N(0,1), then take a convex combinations with weights $(1-\mu)$ and μ respectively, and finally solve for y_{it}. If we follow this procedure we obtain

$$y_{it} = (1-\mu)Q_{ij}(\mu)P_0^{jk}\ f_{kt}(\beta) + \mu Q_{ij}(\mu)P_1^{jk}\ g_{kt}(\gamma) + u_{it} \qquad (4.5)$$

where P_0^{ij} and P_1^{ij} are the ij-th elements of triangular matrices P_0 and P_1 such that $P_0{}^T P_0 = \Omega_0^{-1}$ and $P_1{}^T P_1 = \Omega_1^{-1}$,

$$P(\mu) = (1-\mu)\ P_0 + \mu P_1,$$

and $Q(\mu) = P(\mu)^{-1}$. The covariance matrix of the u_{it}'s in (4.5) is then

$$\Omega(\mu) \equiv Q(\mu)\ Q(\mu)^T.$$

In order to construct P-tests based on (4.3), (4.4) and (4.5), we must first replace γ and Ω_1 by $\hat{\gamma}$ and $\hat{\Omega}_1$ to yield identified J-test regressions. Linearizing these about $\beta = \hat{\beta}$, $\Omega_0 = \hat{\Omega}_0$ and $\alpha = 0$, $\lambda = 0$ or $\mu = 0$, as the case may be, yields the P-test regressions. For the simplest case, (4.3), this is seen to be

$$y_{it} - \hat{f}_{it} = \hat{f}_{it,\nu}\ b^{\nu} + \alpha(\hat{g}_{it} - \hat{f}_{it}) + u_{it}. \qquad (4.6)$$

Here $\hat{f}_{it}$ and $\hat{g}_{it}$ denote the fitted values $f_{it}(X_t,\hat{\beta})$ and $g_{it}(Z_t,\hat{\gamma})$ based on FIML estimates of H_0 and H_1 respectively, and $f_{it,\nu}$ denotes the partial derivative of f_{it} with respect to the ν-th parameter of β, so that $\hat{f}_{it,\nu}$ is this derivative evaluated at $\hat{\beta}$. Note that, under H_0, the u_{it} are distributed as $N(0,\Omega_0)$, so that (4.6) must be estimated by GLS using an assumed covariance matrix proportional to $\hat{\Omega}_0$. We call the t-test of $\alpha = 0$ from this GLS regression the P_0-test.

Applying this same procedure to (4.4) and (4.5), we obtain

$$y_{it} - \hat{f}_{it} = \hat{f}_{it,\nu}\ b^{\nu} + \lambda\hat{\omega}_{ij}^{0}\ \hat{\omega}_1^{jk}\ (\hat{g}_{kt} - \hat{f}_{kt}) + u_{it} \qquad (4.7)$$

and

$$y_{it} - \hat{f}_{it} = \hat{f}_{it,\nu}\, b^{\nu} + \mu\, \hat{Q}^{0}_{ij}\, \hat{P}^{jk}_{1}\, (\hat{g}_{kt} - \hat{f}_{kt}) + u_{it}, \tag{4.8}$$

where $\hat{Q}^{0}_{ij}$ is an element of $\hat{Q}^{0} = \hat{P}_{0}^{-1}$. Both of these regressions are to be estimated by GLS using $\hat{\Omega}_0$ as the assumed covariance matrix. We shall refer to the tests based on (4.7) and (4.8) as the P_1-test and the P_2-test respectively.

We should note at this point that in one important respect the validity of these tests has not yet been established. If $\hat{g}$ and $\hat{\Omega}_1$ were known quantities rather than estimates, the fact that the t-statistics for α, λ and μ in (4.6), (4.7) and (4.8) are asymptotically N(0,1) under H_0 would follow immediately from standard results; see Durbin (1970). We have to verify that,as DM proved for the univariate case, the use of ML estimates rather than known quantities does not affect this asymptotic distribution. This task is relegated to the Appendix.

Let us now briefly set down the steps needed to perform any P-test in the multivariate case:

1. Estimate the models H_0 and H_1 to obtain $\hat{f}_{it}$, $\hat{g}_{it}$, $\hat{\Omega}_0$ and $\hat{\Omega}_1$. Differentiate f_{it} with respect to the parameters of H_0 so as to obtain the quantities $\hat{f}_{it,\nu}$; this may be done numerically.

2. Compute m x m triangular matrices $\hat{P}_0$ and $\hat{P}_1$ such that $\hat{P}_0{}^T\hat{P}_0 = \hat{\Omega}_0{}^{-1}$ and $\hat{P}_1{}^T\hat{P}_1 = \hat{\Omega}_1{}^{-1}$.

3. For the P_0-test, form h_{it} as $(\hat{g}_{it} - \hat{f}_{it})$. For the P_1-test, premultiply the m x n matrix whose typical element is $(\hat{g}_{jt} - \hat{f}_{jt})$ by the m x m matrix $\hat{\Omega}_0\hat{\Omega}_1{}^{-1}$ to form h_{it}. For the P_2-test, premultiply $(\hat{g}_{jt} - \hat{f}_{jt})$ by $\hat{P}_0{}^{-1}\hat{P}_1$ to form h_{it}.

4. Perform a GLS regression of $(y_{it} - \hat{f}_{it})$ on $\hat{f}_{it,\nu}$ and h_{it}. That is, premultiply the regressand and all of the regressors, considered as m x n matrices, by $\hat{P}_0$ and then run an OLS regression. The t-statistic on h_{it} is the P-test statistic. Incidentally, if this last regressor is omitted, all of the coefficients should be identically zero, which is an easy way to check most of the computations.

We now turn our attention to the multivariate CPD test, as exposited by PD. The numerator of the test statistic is

$$T_0 = (n/2) \log (|\hat{\Omega}_1|/|\hat{\Omega}_{10}|), \tag{4.9}$$

where $\hat{\Omega}_{10}$ is an estimate of the probability limit under H_0 of the estimate of Ω_1. As in the univariate case, $\hat{\Omega}_{10}$ is computed as $\hat{\Omega}_0 + \hat{\Omega}_a$ where $\hat{\Omega}_a$ is the estimated covariance matrix from an auxiliary multivariate regression analogous to (2.13), that is, H_1 re-estimated using $\hat{f}_{it}$ as the dependent variable. The symbol $|..|$ denotes a determinant. Following the instructions given by PD, one may compute an estimate of the variance of T_0, and thus compute the test statistic.

At first glance, there is no apparent resemblance between (4.9) and any sort of P-test. The former involves a logarithm, two determinants and an auxiliary regression, none of which would play any part in the latter. Nevertheless, it turns out that the P_1-test is asymptotically equivalent under the null hypothesis to the CPD test, a proposition which we now set out to prove. Many of the technical details have been relegated to an appendix, but a certain amount of algebra is unavoidable.

The first difference between the CPD test and the P_1-test is the former's use of an auxiliary nonlinear regression to compute $\hat{\Omega}_{10}$. As in the univariate case, this is unnecessary. Let us denote elements of $\hat{\Omega}_{10}$, $\hat{\Omega}_0$ and $\hat{\Omega}_1$ respectively by $\hat{\omega}^{10}_{ij}$, $\hat{\omega}^{0}_{ij}$ and $\hat{\omega}^{1}_{ij}$. Then PD compute $\hat{\Omega}_{10}$ using

$$\hat{\omega}^{10}_{ij} = \hat{\omega}^{0}_{ij} + (1/n)(\hat{f}_{it} - \tilde{g}_{it})(\hat{f}_j{}^t - \tilde{g}_j{}^t), \tag{4.10}$$

where, as before, the $\tilde{g}_{it}$ are fitted values based on estimates $\tilde{\gamma}$ from the auxiliary regression. It is straightforward to show that under H_0 plim $\tilde{\gamma}$ = plim $\hat{\gamma} = \overline{\gamma}$, so that $\hat{g}$ may validly be used in (4.10) rather than $\tilde{g}$. For details, see the appendix.

If we replace $\tilde{g}$ by $\hat{g}$ and then use the same Taylor series approximation (2.17) as in the univariate case, we obtain

$$S' = (\sqrt{n}/2)[1 - |\hat{\omega}^{0}_{ij} + (\hat{f}_{it} - \hat{g}_{it})(\hat{f}_j{}^t - \hat{g}_j{}^t)/n| / |\hat{\omega}^{1}_{ij}|]. \tag{4.11}$$

The numerator determinant in this expression can be rewritten as

$$\begin{aligned} &|\hat{\omega}^{1}_{ij} + \hat{\omega}^{0}_{ij} - \hat{\omega}^{1}_{ij} + (\hat{g}_{it} - \hat{f}_{it})\ (\hat{g}_j{}^t - \hat{f}_j{}^t)/n| \\ &\quad = |\hat{\omega}^{1}_{ij} + (2/n)\ (\hat{g}_{it} - \hat{f}_{it})\ (y_j{}^t - \hat{f}_j{}^t)|. \end{aligned} \tag{4.12}$$

Consider now a determinant $|X_{ij} + \theta_{ij}|$ where the quantities θ are an order of magnitude smaller than the quantities X. The Taylor expansion of the determinant to first order is

$$|X_{ij} + \theta_{ij}| \doteq |X_{ij}|\ (1 + X^{ji}\ \theta_{ij}), \tag{4.13}$$

where the superscripts again denote the inverse matrix. This formula may validly be applied to (4.12), because the first term, $\hat{\omega}^{1}_{ij}$, is of order unity as $n \to \infty$, while the second is of order $n^{-\frac{1}{2}}$.[4] The result is

$$|\hat{\omega}^{1}_{ij}| \; [1 + (2/n)\hat{\omega}_1^{ij} \; (\hat{g}_{it} - \hat{f}_{it}) \; ({y_j}^t - {f_j}^t)]. \tag{4.14}$$

Substitution of this into (4.11) yields

$$S' \sim (-1/\sqrt{n}) \; \hat{\omega}_1^{ij} \; (\hat{g}_{it} - \hat{f}_{it}) \; ({y_j}^t - {\hat{f}_j}{}^t). \tag{4.15}$$

The expression on the right-hand side of (4.15) is of order unity, as it should be, and is analogous to (2.18). We must now show that it is equal to the numerator of the P_1-test statistic. Consider first the simple GLS linear regression:

$$\hat{N}_i{}^j{}_t{}^s \; (y_{js} - \hat{f}_{js}) = \lambda \hat{N}_i{}^j{}_t{}^s \; \hat{\omega}^0_{jk} \; \hat{\omega}_1^{k\ell} \; (\hat{g}_{\ell s} - \hat{f}_{\ell s}) + u_{it}, \tag{4.16}$$

where $\hat{N}$ is the oblique projection defined by a GLS regression on the $\hat{f}_{jt,\nu}$ with covariance matrix $\hat{\Omega}_0$; for further details, see the appendix. It is easily proved that the estimate of λ and of its t-statistic from (4.16) will be identical to those from (4.7), except for any degrees of freedom correction in computing the variance. This is a consequence of a well-known result which may be expressed as follows: the estimates of the parameters c and of their variances will be identical whether one estimates by GLS the regression $Y = Xc + Zd + u$ with u assumed $N(0,\Omega)$, or the regression $N_Z Y = N_Z Xc + u$, where Y is a vector of dependent variables, X and Z are matrices of independent variables, and N_Z is the oblique projection defined by

$$N_Z = I - Z(Z^T \, \Omega^{-1} \, Z)^{-1} \, Z^T \, \Omega^{-1}.$$

Since $\hat{N}$ in (4.16) is the oblique projection corresponding to a regression on the $\hat{f}_{it,\nu}$ with covariance matrix $\hat{\Omega}_0$, the required result follows immediately.

Now observe that

$$\hat{N}_i{}^j{}_t{}^s\,(y_{js} - \hat{f}_{js}) = y_{it} - \hat{f}_{it}. \tag{4.17}$$

This result corresponds to the result that $\hat{M}(y - \hat{f}) = y - \hat{f}$ in the univariate case, and is proved similarly from the likelihood equations for H_0 which define $\hat{\beta}$ and $\hat{\Omega}_0$ (see the Appendix). Thus to perform the GLS regression (4.16), we may run the OLS regression

$$\hat{P}_0^{ai}(y_{it} - \hat{f}_{it}) = \lambda \hat{P}_0^{ai}\,\hat{N}_i{}^j{}_t{}^s\,\hat{\omega}^0_{jk}\,\hat{\omega}_1^{k\ell}\,(g_{\ell s} - f_{\ell s}) + \varepsilon^a{}_t. \tag{4.18}$$

The numerator of the estimate of λ from this regression is

$$(y_{it} - \hat{f}_{it})\,\hat{\omega}_0^{ib}\,\hat{N}_b{}^{jt}{}_s\,\hat{\omega}^0_{jk}\,\hat{\omega}_1^{k\ell}\,(g_\ell{}^s - f_\ell{}^s). \tag{4.19}$$

It follows directly from a result in the Appendix that

$$\hat{\omega}_0^{ib}\,\hat{N}_b{}^{jt}{}_s\,\hat{\omega}^0_{jk} = \hat{N}_k{}^i{}_s{}^t,$$

and this along with (4.17) allows us to simplify (4.19) to

$$(y_{it} - \hat{f}_{it})\,\hat{\omega}_1^{kj}\,(\hat{g}_j{}^s - \hat{f}_j{}^s). \tag{4.20}$$

This is clearly identical to the right-hand side of (4.15), except for the factor $(-1/\sqrt{n})$. Expression (4-20) is the numerator of the t-statistic from the P_1-test regression. Since the denominator will have a non-stochastic, non-zero probability limit, it follows that S' and the P_1-test yield exactly the same tests, asymptotically.

We have thus proved that, in the multivariate case as well as in the univariate one, there exists a P-test which implements Cox's basic idea. There also exist (at least) two other P-tests which are not asymptotically Cox tests, and it would be interesting to compare the power of the three tests. We have not yet obtained any analytical results on this matter. Limited experience with empirical applications of the tests suggests that the P_1-test and P_2-test yield very similar inferences, and are more prone to reject (presumably false) null hypotheses than the P_0-test.

It should perhaps be noted that P-tests can be applied to simultaneous equations models without explicitly deriving their reduced forms. One merely requires, in order to obtain the fitted values, covariance matrix and derivatives of a model, that one be able to estimate the model and solve it for the values of the dependent variables conditional on the predetermined variables; the derivatives can, of course, be computed numerically. There would, however, appear to be a problem if the model being tested were non-linear, so that the covariance matrix of the errors adhering to the reduced form would not be constant over time. At the moment it is not clear how any of these procedures could validly be adapted to deal with such a case.

It should also be noted that P-tests can straightforwardly be used to test a model against several other models simultaneously. In that case the test regression will include several regressors like h_{it}, and the appropriate test statistic will be a Wald or pseudo-likelihood-ratio statistic.

5. A Sampling Experiment

In Section 2 above we showed that the univariate CPD test statistic and the P-test statistic are asymptotically the same random variable under H_0, and in Section 3 we showed that for linear models the asymptotic relative efficiency of the two tests is unity for local alternatives. Thus according to the large-sample theory, there is no reason to prefer one test over the other. The next step, obviously, is to investigate the performance of the tests in small samples. However, a full analysis of this matter would be far beyond the scope of this paper. Instead, we report the results of a sampling experiment in which the performance of the two tests is compared for univariate linear models. This is computationally the easiest case to deal with, and surely the most common in practice. Note that in this case the CPD test reduces to Pesaran's (1974) test, and that the J-test and the P-test are identical.

A number of regressors were generated according to simple ARIMA models, with specifications similar to those characterizing actual quarterly economic time series. These specifications were adapted from some of those reported by Nelson (1973, Chapter 8). To ensure that related series (for example, two different price or interest rate series) were indeed related, the error terms in the ARIMA models were chosen to be correlated with each other. The following regressors were generated in this way: Y, designed to resemble the log of current dollar GNP; U, designed to resemble the unemployment rate; UR = log(U/(100 - U)); PC, designed to resemble the rate of change of the Consumer Price Index; PY, designed to resemble the rate of change of the GNP deflator; RS, designed to resemble the log of a short-term interest rate; and RL, designed to resemble the log of a long-term interest rate. For each of these regressors we generated 25 observations. For sample sizes longer than 25 these same observations were then repeated, so as to ensure that the X matrices did not change systematically as the sample size was increased.

The dependent variable, which is not intended to have any particular economic interpretation, will be referred to as D. Three linear models to explain D were postulated. They are:

$$H_1: \quad D = a_0 + a_1Y + a_2U + a_3PY + a_4RL + \varepsilon$$
$$H_2: \quad D = b_0 + b_1Y + b_2UR + b_3PC + b_4RS + \varepsilon$$
$$H_3: \quad D = c_0 + c_1Y + c_2U + c_3PC + c_4RL + c_5 RS + \varepsilon$$

The data were actually generated by H_1, with the following parameter values: $a_0 = .5$, $a_1 = .8$, $a_2 = .02$, $a_3 = -.02$, $a_4 = .2$. Thus H_2 and H_3 were always false models. Note that H_2 has only one regressor, Y, in common with H_1, and has the same number of parameters. On the other hand, H_3 has three regressors, Y, U and RL, in common with H_1, and has one more parameter. Thus it seems likely that H_3 will fit better than H_2, so that a test of H_1 against H_3 is more likely to result in rejection of the true model than a test of H_1 against H_2.

Six different experiments were performed. In each case the number of replications was 500, which is sufficiently large for trustworthy statistical inferences. The sample size was either 25 or 100, which are roughly the extremes for time series work with quarterly data. The variance was .0001, .0004 or .0016. Since the data are in logarithms, these correspond to standard errors of one, two and four percent. Standard errors of one and two percent seem quite realistic, but a standard error of four percent seems rather large for time series regressions. Thus the worst case considered (sample size 25, variance .0016) is surely as unfavorable a situation for estimation and inference as one is likely to encounter.

We first consider what happens when the model being tested is the true model, H_1. Asymptotically, both the test statistics should be N(0,1) in this case. Whether the observed distributions are consistent with this may be tested by means of a Kolmogorov-Smirnov test. In Table 1, the numbers reported under KSP are the probabilities, on a two-tail test, of observing a KS statistic as large as or larger than the one actually observed, given that the true distribution is N(0,1).

In Table 1 we also report the means and standard deviations of the P-test and CPD test statistics under "Mean" and "S.D." respectively, together with test statistics for the hypotheses that the true mean is zero and the true variance is unity. The latter, which should both be N(0,1) under the null, are reported under "Test μ" and "Test σ^2". The first of these is simply $\hat{\mu}/(\hat{\sigma}\sqrt{500})$ and the second, which is based on a large-sample approximation, is $(\hat{\sigma}^2 - 1)/(\hat{\sigma}^4/250)^{\frac{1}{2}}$.

What we are really interested in, of course, is how many times the two non-nested hypothesis tests will lead us falsely to reject H_1. We therefore report the proportion of times that the two test statistics are greater than 1.96 and 2.50 in absolute value (under "R1.96" and "R2.50"). The former number is of course the .05 critical value for the normal distribution. The latter corresponds to a .0124 critical level, but, more important, it is a convenient number to remember, which one might reasonably use as a conservative critical value in applied work if one suspected

TABLE 1

Tests of H_1 Against Two Alternative Models

Sample Size	Variance	Alternative Model	KSP	Mean	Test μ	S.D.	Test σ^2	R1.96	R2.50
25	.0001	H_2	.0042	.1249	2.93	.9540	-1.56	.038	.014
			.0074	-.2144	-4.25	1.1268	3.36	.090 *	.036 *
		H_3	.0000	.3233	7.24	.9992	-0.03	.076 *	.022
			.0000	-.3965	-7.09	1.2500	5.69	.118 *	.064 *
25	.0004	H_2	.0000	.2515	5.67	.9915	-0.27	.056	.016
			.0000	-.3606	-6.20	1.3004	6.46	.120 *	.072 *
		H_3	.0000	.7805	18.64	.9364	-2.22	.104 *	.036 *
			.0000	-1.0380	-8.22	2.8250	13.83	.208 *	.134 *
25	.0016	H_2	.0000	.5480	12.47	.9822	-0.58	.070	.024 *
			.0000	-.6511	-9.58	1.5192	8.96	.150 *	.090 *
		H_3	.0000	1.0947	28.99	.8444	-6.36	.156 *	.052 *
			.0000	-4.1969	-4.77	19.6844	15.77	.384 *	.312 *
100	.0001	H_2	.4258	.0558	1.28	.9776	-0.73	.036	.008
			.3302	-.0981	-2.14	1.0250	0.76	.040	.018
		H_3	.0082	.1800	3.97	1.0133	0.41	.050	.014
			.0138	-.1963	-4.09	1.0738	2.10	.074 *	.022
100	.0004	H_2	.4357	.0538	1.20	1.0049	0.15	.048	.016
			.2308	-.1037	-2.15	1.0768	2.17	.064	.024 *
		H_3	.0000	.3254	7.40	.9829	-0.56	.060	.018
			.0058	-.3000	-5.98	1.1225	3.26	.084 *	.046 *
100	.0016	H_2	.0000	.2217	5.20	.9530	-1.60	.050	.012
			.0104	-.2204	-4.65	1.0600	1.74	.082 *	.038 *
		H_3	.0000	.7204	18.60	.8662	-5.26	.076 *	.024 *
			.0000	-.7059	-12.58	1.2547	5.77	.148 *	.094 *

Note: The first number of each pair refers to the P-test (and J-test) statistic, and the second refers to the CPD (Pesaran) statistic.

that the asymptotic regime did not strictly apply. If the proportion of rejections is significantly greater than .05 or .0124, according to a normal approximation to the binomial distribution, this is indicated by an asterisk.

It is clear from Table 1 that the small sample distributions of the test statistics depend on the sample size, the variance, and the characteristics of the alternative model. When the sample size is 25, the KS test always rejects the hypothesis that the true distribution is N(0,1). When the sample size is 100 and the variance is .0001 or .0004, that hypothesis cannot be rejected when the alternative model is H_2, but can be rejected when it is H_3. This is true for both tests.

Both test statistics tend to have non-zero means, of roughly the same magnitude. The mean for the P-test is always positive, and the mean for the CPD test is always negative. This is perhaps an unfortunate characteristic, since those are the signs one would expect the test statistics to have if the alternative model were true.

The major difference between the two test statistics is that the standard deviation of the CPD statistic is always greater than unity, usually significantly so, while that of the P-test statistic is usually less than unity and never significantly greater. In the most extreme case (sample size 25, variance .0016, alternative H_3), the standard deviation of the CPD test is almost twenty, reflecting the influence of some extreme outliers.

The large variance of the CPD test means that it always rejects the null hypothesis more often than the P-test, and usually rejects it more often than it should. The P-test also rejects the null too often in some cases, but this is much less marked. Moreover, the distribution of the P-test statistic apparently has much thinner tails than that of the CPD statistic. Using 2.5 as the critical value never yields a rejection rate of more than 5.2% for the P-test, but yields one as high as 31.2% for the CPD test. This suggests that even in cases where the small sample distribution is far from N(0,1), one may be able to guard against Type I error for the P-test by using a somewhat conservative critical value, but that this will not be possible for the CPD test.

These results also suggest that it will be relatively easy to modify the P-test to make it approximately valid in small samples, because one would simply have to subtract an estimate of the mean of the test statistic. In contrast, any attempt to make the CPD test more useful in small samples would have to deal with the variance as well as the mean. These matters are the subject of ongoing research.

In Table 2, we present results for the case where the model under test is false. We present the proportion of the time that the model under test is rejected using critical values of 1.96 and 2.50, and the mean, median and standard deviation of the test statistics. In the left-hand side of the table H_2 and H_3 are tested against the true model, H_1, and in the right-hand side they are tested against each other.

It has often been observed in practice that when the P-test statistic is large (say, greater than four), the CPD test statistic is even larger. This observation is confirmed in Table 2. Most of the time the mean of the latter is indeed larger in absolute value than the

TABLE 2

Tests of False Models

Model Tested	R1.96	R2.50	Mean	Median	S.D.	False Alt.	R1.96	R2.50	Mean	Median	S.D.
Case 1: Sample Size = 25, Variance = .0001											
H_2	1.000	1.000	7.14	6.92	1.73	H_3	.642	.208	2.14	2.12	0.50
	1.000	1.000	-14.75	-14.39	3.72		.930	.846	-3.91	-3.67	1.50
H_3	1.000	.998	6.19	6.01	1.63	H_2	.056	.008	-0.95	-0.95	0.62
	1.000	1.000	-20.97	-19.61	9.37		.054	.002	1.01	1.05	0.62
Case 2: Sample Size = 25, Variance = .0004											
H_2	.924	.820	3.60	3.54	1.21	H_3	.328	.114	1.68	1.69	0.71
	.976	.960	-7.84	-7.44	3.71		.710	.604	-3.62	-3.08	2.71
H_3	.832	.678	3.06	2.97	1.20	H_2	.068	.018	-0.75	0.75	0.82
	.970	.950	-16.15	-10.02	22.71		.070	.004	0.78	0.85	0.84
Case 3: Sample Size = 25, Variance = .0016											
H_2	.430	.232	1.83	1.78	1.06	H_3	.124	.036	1.18	1.17	0.74
	.658	.580	-3.94	-3.02	3.54		.572	.490	-4.03	-2.41	6.79
H_3	.308	.160	1.55	1.48	1.02	H_2	.074	.028	-0.35	-0.32	1.01
	.710	.640	-28.59	-4.18	339.75		.074	.022	0.33	0.35	1.06
Case 4: Sample Size = 100, Variance = .0001											
H_2	1.000	1.000	13.76	13.71	1.41	H_3	1.000	1.000	4.40	4.38	0.51
	1.000	1.000	-26.73	-26.64	3.23		1.000	1.000	-7.16	-7.10	1.44
H_3	1.000	1.000	11.98	12.02	1.32	H_2	.572	.268	-2.10	-2.09	0.67
	1.000	1.000	-34.91	-34.41	5.92		.558	.230	2.03	2.03	0.60
Case 5: Sample Size = 100, Variance = .0004											
H_2	1.000	1.000	6.92	6.89	1.13	H_3	.932	.798	3.07	3.10	0.70
	1.000	1.000	-13.65	-13.43	3.29		.962	.920	-5.19	-4.89	2.17
H_3	1.000	.996	6.01	5.99	1.10	H_2	.254	.106	-1.39	-1.46	0.91
	1.000	1.000	-25.21	-18.52	122.69		.246	.094	1.37	1.45	0.87
Case 6: Sample Size = 100, Variance = .0016											
H_2	.948	.842	3.48	3.43	0.99	H_3	.448	.190	1.85	1.87	0.78
	.970	.952	-6.83	-6.35	3.17		.716	.604	-3.58	-2.98	2.67
H_3	.852	.704	2.99	2.96	0.99	H_2	.112	.044	-0.80	-0.82	0.99
	.970	.952	-36.91	-9.29	425.64		.100	.032	0.79	0.83	0.98

Note: The first number of each pair refers to the P-test statistic, the second to the CPD test statistic.

mean of the former. In such cases the standard deviation of the latter is also always much larger. By itself, the larger mean would give the CPD test greater power, while the larger variance would give it less power.

In fact, it is evident that both tests have ample power against the truth in cases 1, 4 and 5, and reasonably high power in cases 2 and 6. Only when the sample size is 25 and the variance is .0016 do the tests seriously lack power. It also appears that, except when testing H_3 against H_2, the CPD test has higher power than the P-test. This result, however, is probably spurious, since we know from Table 1 that the sizes of the two tests are not equal. How we can meaningfully compare power when these sizes are unknown is far from clear. All we can say with any confidence is that the CPD test is more likely to lead to rejection than the P-test, irrespective of whether the model being tested is true.

The foregoing experimental results suggest the following conclusions:

1. When the sample size is reasonably large and the variance is reasonably small, both the tests perform in a satisfactory manner. This will depend on the characteristics of the models being tested, of course.

2. As the sample size decreases and/or the variance increases, the performance of both tests deteriorates. The CPD test becomes much more likely than the P-test to reject the model being tested, whether or not it is true. The P-test may still be used safely in unfavourable conditions by adopting a somewhat conservative critical value, such as 2.5 or 3.0, but the CPD test acquires a very large variance under the null in such conditions and becomes completely unreliable.

3. Further theoretical work to develop tests which can validly be used in small samples is clearly called for.

Appendix

In this technical appendix we demonstrate that both $\hat{\gamma}$ and $\tilde{\gamma}$ tend to the same probability limit $\overline{\gamma}$, under H_0, so that an auxiliary regression like (2.13) is indeed unnecessary. Using this result, we then show that the t-statistics from all of the P-test regressions (4.6), (4-7) and (4.8) are asymptotically N(0,1) under H_0.

The likelihood equations for model H_1 which define the estimates $\hat{\gamma}$ and $\hat{\Omega}_1$ are as follows:

$$\hat{\omega}^1_{ab} = (1/n)\ (y_{at} - g_{at}(\hat{\gamma}))\ (y_b{}^t - g_b{}^t(\hat{\gamma})) \tag{A.1}$$

$$g_{it,\mu}\ (\hat{\gamma})\ \hat{\omega}_1^{ij}\ (y_j{}^t - g_j{}^t(\hat{\gamma})) = 0. \tag{A.2}$$

In order to discuss the convergence in probability under H_0 of $\hat{\gamma}$ and $\hat{\Omega}_1$, we define the functions W_{ab} and G:

$$W_{ab}(\gamma,\Omega,r_{ab}) = \omega^0_{ab} - \omega_{ab} + r_{ab}$$
$$+ \lim_{n\to\infty} (1/n)\ (f_{at}(\beta_0) - g_{at}(\gamma))\ (f_b{}^t(\beta_0) - g_b{}^t(\gamma))$$

$$G_\mu(\gamma,\Omega,s_\mu) = s_\mu + \lim_{n\to\infty} (1/n) g_{it,\mu}(\gamma)\ \omega^{ij}\ (f_j{}^t(\beta_0) - g_j{}^t(\gamma)).$$

Here ω^0_{ab} and β_0 denote the true values of these parameters, and r_{ab} and s_μ denote as yet unspecified arguments. So that the functions W and G are well-defined, we must make a few assumptions, some of which were made already in sections 2 and 4:

(i) As $n \to \infty$, the limits of $(1/n)f_{at}(\beta_0)f_b{}^t(\beta_0)$, $(1/n)g_{at}(\gamma)f_b{}^t(\beta_0)$ and $(1/n)g_{at}(\gamma)g_b{}^t(\gamma)$ exist and are finite. Convergence is uniform with respect to γ in any compact subset of $\mathbb{R}^{\ell}$, so that the limits are continuous functions of γ. In fact, uniform convergence of enough derivatives is assumed so that the limits here are twice continuously differentiable.

(ii) As $n \to \infty$, the limits of $(1/n)f_{it,\mu}(\beta_0)f_{j}{}^{t}{}_{,\nu}(\beta_0)$, $(1/n)g_{it,\mu}(\gamma)g_j{}^t(\gamma)$ and $(1/n)g_{it,\mu}(\gamma)f_j{}^t(\beta_0)$ exist and are finite. Again, convergence is uniform in compact sets.

(iii) There exists a finite solution $(\overline{\gamma},\overline{\Omega})$ of the equations $W_{ab}(\gamma,\Omega,0) = 0$ and $G_\mu(\gamma,\Omega,0) = 0$ which corresponds to a global maximum of

$$2\,L_1(\gamma,\Omega) = -\log|\Omega| - \lim_{n\to\infty} (f_{at}(\beta_0) - g_{at}(\gamma))\omega^{ab}(f_b{}^t(\beta_0) - g_b{}^t(\gamma)).$$

(iv) The Hessian of L_1 at $(\overline{\gamma},\overline{\Omega})$ is positive definite. Of course, Ω as an argument of L_1 is restricted to the set of symmetric matrices. Assumptions (iii) and (iv) ensure that H_1 is asymptotically identified under H_0.

We now wish to show that as $n \to \infty$, $\text{plim}\ \hat{\gamma} = \text{plim}\ \tilde{\gamma} = \overline{\gamma}$. The equations given by PD for $\tilde{\gamma}$ and the matrix they call $\hat{\Omega}_{10}$ are as follows:

$$\hat{\omega}^{10}_{ab} = \omega^{0}_{ab} + (1/n)\ (f_{at}(\hat{\beta}) - g_{at}(\tilde{\gamma}))(f_b{}^t(\hat{\beta}) - g_b{}^t(\tilde{\gamma})) \tag{A.3}$$

$$g_{it,\mu}(\tilde{\gamma})\ \hat{\omega}^{ij}_{10}\ (f_j{}^t\ (\hat{\beta}) - g_j{}^t\ (\tilde{\gamma})) = 0 \tag{A.4}$$

Assumptions (i) and (ii) ensure that if H_0 is true the usual maximum likelihood results hold:

$$\hat{\beta} = \beta_0 + O_p\ (n^{-\frac{1}{2}}) \tag{A.5}$$

$$\hat{\Omega}_0 = \Omega_0 + O_p\ (n^{-\frac{1}{2}}). \tag{A.6}$$

Consequently, both equations (A.1) and (A.2) for $\hat{\gamma}$ and $\hat{\Omega}_1$ and equations (A.3) and (A.4) for $\tilde{\gamma}$ and $\hat{\Omega}_{10}$ can be expressed in the form

$$W_{ab}\ (\gamma,\Omega,r_{ab}\ (\gamma,\Omega,\ n)) = 0 \tag{A.7}$$

$$G_\mu(\gamma,\Omega,s_\mu(\gamma,\Omega,n)) \qquad = 0, \tag{A.8}$$

if $r_{ab}(\cdot)$ and $s_\mu(\cdot)$ are defined appropriately. For equations (A.1) and (A.2), for example,

$$s_\mu(\gamma,\Omega,n) = (1/n)g_{it,\mu}(\gamma)\ \omega^{ij}\ \varepsilon_j{}^t$$

$$+\ (1/n)g_{it,\mu}(\gamma)\ \omega^{ij}\ (f_j{}^t(\beta_0) - g_j{}^t(\gamma))$$

$$-\ \lim_{n\to\infty}\ (1/n)\ g_{it,\mu}(\gamma)\ \omega^{ij}\ (f_j{}^t(\beta_0) - g_j{}^t(\gamma)),$$

since $\quad y_{it} = f_{it}\ (\beta_0) + \varepsilon_{it}$ (A.9)

Clearly the random functions $r_{ab}(\cdot)$ and $s\ (\cdot)$ are all $o_p(1)$ as $n \to \infty$ uniformly in compact sets.

Assumptions (iii) and (iv) are precisely what is needed to apply the implicit function theorem to the equations

$$W_{ab}\ (\gamma,\Omega,r_{ab}) = 0 \tag{A.10}$$

$$G\ (\gamma,\Omega,s_\mu) \qquad = 0\ . \tag{A.11}$$

We conclude that in the neighbourhood of $(\overline{\gamma},\overline{\Omega})$ and for small enough r_{ab} and s_μ, unique solutions to these equations exist and are differentiable in the parameters r_{ab} and s_μ. But for any realization of the random functions $r_{ab}(\cdot)$ and $s_\mu(\cdot)$, equations (A.7) and (A.8) will have, for large enough n, a unique solution in the neighbourhood of $(\bar{\gamma},\bar{\Omega})$. Since at this solution r_{ab} and s_μ take on values which are $o_p(1)$, it follows from the remark following (A.11) that this solution is distant from $(\bar{\gamma},\bar{\Omega})$ by an amount which is $o_p(1)$ as $n \to \infty$. Thus both plim $\hat{\gamma}$ and plim $\tilde{\gamma}$ exist and equal $\bar{\gamma}$.

We now wish to derive an explicit expression for $f_{at}(\hat{\beta})$. Using standard results on maximum likelihood estimators, we obtain

$$f_{at}(\hat{\beta}) = f_{at} + (1/n)f_{at,\mu}\, F^{\mu\nu}\, \omega_o^{bc}\, f_{cs,\nu}\, \varepsilon_b^{\ s} + 0_p(n^{-1}) \tag{A.12}$$

where f_{at} denotes $f_{at}(\beta_0)$, and so on. Here $F^{\mu\nu}$ is an element of the inverse of the matrix with typical element $F_{\mu\nu}$ defined by

$$F_{\mu\nu} = (1/n)\, \omega_0^{ab}\, f_{bt,\mu}\, f_{a\ ,\nu}^{\ t} .$$

It is convenient here to introduce the oblique projection associated with GLS regression on the $f_{at,\mu}$ with covariance matrix Ω_0. This is the projection conventionally expressed as

$$I - Z(Z^T\Omega_0^{\ -1}Z)^{-1}Z^T\Omega_0^{\ -1}$$

in textbooks. In our notation it takes the form:

$$N_{i\ t}^{\ j\ s} = \delta_i^{\ j}\delta_t^{\ s} - (1/n)f_{it,\mu}\, F^{\mu\nu}\, f_{k\ ,\nu}^{\ s}\omega_0^{kj} . \tag{A.13}$$

One can easily check that N is idempotent, so that

$$N_{i\ t}^{\ j\ s}\, N_{j\ s}^{\ k\ r} = N_{i\ t}^{\ k\ r} \tag{A.14}$$

and one can verify the useful result that

$$\omega_{ij}^0\, N_{k\ t}^{\ j\ s} = N_{i\ \ t}^{\ \ell s}\, \omega_{\ell k}^0, \tag{A.15}$$

Note that an expression like $N_{i\ \ t}^{\ \ell s}$ corresponds to the transpose of $N_{i\ t}^{\ \ell\ s}$

It follows directly from (A.12) and (A.9) that

$$y_{at} - \hat{f}_{at} = N_a{}^b{}_t{}^s \varepsilon_{bs} + O_p(n^{-1}) \qquad (A.16)$$

Notice further that if one evaluates all the functions of unknown parameters in (A.13) at the maximum likelihood estimates so as to define $\hat{N}$, then it follows from the likelihood equation

$$\hat{f}_{it,\mu} \hat{\omega}_0^{ij} (y_j{}^t - \hat{f}_j{}^t) = 0$$

that

$$\hat{N}_i{}^j{}_t{}^s (y_{js} - \hat{f}_{js}) = y_{it} - \hat{f}_{it}.$$

As in equation (4.16), we may write the various P-test regressions (4.6), (4.7) and (4.8) in the form

$$\hat{N}_i{}^j{}_t{}^s (y_{js} - \hat{f}_{js}) = a_I \hat{N}_i{}^j{}_t{}^s \hat{\omega}^0_{jk} \hat{\omega}_I^{k\ell} (\hat{g}_{\ell s} - \hat{f}_{\ell s}) + u_{it} \qquad (A.17)$$

where $\hat{\omega}_I^{k\ell}$ denotes the $k\ell$-th element of the inverse of $\hat{\Omega}_0$, $\hat{\Omega}_1$ or $Q_1 Q_0{}^T$ for the P_0-, P_1- and P_2-tests respectively, and a_I stands for α, λ or μ. If $\hat{a}_I$ denotes the GLS estimator of a_I from (A.17), then

$$\sqrt{n}\, \hat{a}_I = \hat{A}_I / \hat{V}_I$$

and the GLS estimator of the variance of $\hat{a}_I$ is $(1/n)\hat{V}_I{}^{-1}$, where

$$\hat{A}_I = n^{-1/2} \hat{\omega}_I^{ij} (\hat{g}_i{}^t - \hat{f}_i{}^t) \hat{N}_j{}^k{}_t{}^r (y_{kr} - \hat{f}_{kr})$$

$$\hat{V}_I = (1/n)(\hat{g}_i{}^t - \hat{f}_i{}^t) \hat{\omega}_I^{ij} \hat{\omega}^0_{ka} \hat{N}_j{}^a{}_t{}^s \hat{\omega}_I^{k\ell} (\hat{g}_{\ell s} - \hat{f}_{\ell s}).$$

Since $\hat{\Omega}_I$, $\hat{N}$, $\hat{g}_{it}$ and $\hat{f}_{it}$ have non-stochastic probability limits, we conclude from (A.14), (A.15) and (A.16) that the t-statistic for $\hat{a}_I$ from (A.17) will be asymptotically distributed as N(0,1).

FOOTNOTES

1. As a matter of historical fact, the J-test was not developed as a way to get around the problem of identification in artificial nesting procedures, but perhaps it should have been.

2. Strictly speaking, we do not require that (2.15) have a solution $\bar{\gamma}$ which is locally unique. Even if the parameters of H_1 are not identified under H_0, our subsequent results will hold under the weaker assumption that there exists a plim of $g(\hat{\gamma})$ under H_0; if so, $g(\tilde{\gamma})$ will have the same plim.

3. Our previous assumptions imply that X and Z are respectively $n \times k$ and $n \times \ell$ matrices, each of full column rank, and such that the dimension of the intersection of the linear spans of their columns is strictly less than $\min(\ell,k)$.

4. Because the y_{it}, $\hat{f}_{it}$ and $\hat{g}_{it}$ are bounded, expressions like $y_{it}y_j^t/n$ and $y_{it}\hat{f}_j^t/n$ will be assumed to have well-defined finite limits as $n \to \infty$. On the other hand, expressions like $y_{it}\varepsilon_j^t/n$ have zero mean and variance of order n^{-1}; the expressions themselves are therefore of order $n^{-\frac{1}{2}}$.

REFERENCES

Atkinson, A.C. (1970), "A Method for Discriminating Between Models", *Journal of the Royal Statistical Society,* Series B, 32, 323-353.

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

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

Davidson, R. and MacKinnon, J.G. (1980), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", Queen's University Institute for Economic Research Discussion Paper No. 378, forthcoming, *Econometrica.*

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

Kendall, M.G. and Stuart, A. (1967), *The Advanced Theory of Statistics,* Vol. 2, 2nd Ed., London: Griffin.

Nelson, C.R. (1973), *Applied Time Series Analysis for Managerial Forecasting,* San Francisco: Holden-Day.

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

____________ (1980), "On the Comprehensive Method of Testing Non-nested Regression Models", unpublished manuscript.

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