SOME RECENT DEVELOPMENTS
IN ECONOMETRIC TEST METHODOLOGY

Garry D.A. PHILLIPS*

Research Memorandum No. 205

October 1984

*University of Leeds, England

An Invited Lecture presented at the Institute for Advanced Studies, Vienna, on March 27th, 1984.

CONTENTS

SOME RECENT DEVELOPMENTS IN ECONOMETRIC TEST METHODOLOGY

by

Garry D.A. Phillips (University of Leeds)

1. Introduction

A crucial element in the development of econometric methodology over the past decade or so has been the concern with testing, as opposed to estimating, econometric models. This has been brought about partly as the result of the subject reaching a new level of maturity and partly because the lack of success of econometric modelling in the seventies, particularly at the macro level, made clear the need for a deeper concern with model evaluation.

Modern econometric practice advocates that a given specification should be subject to a rigourous testing procedure and it is now becoming routine to test for misspecifications such as omitted variables, serially correlated disturbances and structural change and, in addition, to test for heteroscedasticity and incorrect functional form. This kind of intensive misspecification testing leads to problems of distortions in the inference procedures but leading econometricians believe that the importance of carrying out such tests overrides these problems. Thus Sargan (1975) expressed the view that

> "Despite the problems associated with data mining, I consider that a suggested specification should be tested in all possible ways and only those which survive and correspond to a reasonable model should be used."

Also Hendry (1980) argued that

> "Far more rapid progress could be made if all empirical studies would provide greatly improved test information to all readers to correctly judge plausibility. The three golden rules of econometrics are test, test and test."

Hendry advocated that this should be done notwithstanding the difficulties involved in calculating and controlling Type 1 and Type 2 errors.

While it is important to test econometric models rigourously it is also important to seek to structure the testing procedure in such a way that problems of data mining are minimised. In particular we seek test procedures to test for the presence of, possibly, several misspecifications simultaneously in such a way that: (a) the overall Type 1 error probability is controlled within acceptable limits, and (b) the test procedure while having good power properties provides some opportunity for detecting individual types of misspecification. In this paper I consider some recent advances in test methodology which contribute to the development of such procedures.

2. Multiple Testing for Misspecification: Some General Issues

There are two general approaches to conducting tests for misspecification in econometric models. In the first approach, we obtain some sample statistic whose distribution is known under the null hypothesis, i.e. when the model is assumed correct, and if the statistic assumes a significant value this is taken as evidence that the model is misspecified in some unknown way. The D.W. statistic is sometimes used for such a test because it is relatively sensitive to various departures from the null hypothesis. In this case we do not necessarily regard a significant test result as implying that the disturbances are first order serially correlated and proceed with a Cochrane Orcutt type estimation procedure. The significant test result is simply taken as evidence that something is wrong. Tests of this type are known as pure significance tests and do not require the specification of a particular alternative model.

A second approach involves choosing a more general model from that which is assumed to be correct. The assumed model may then be obtained from the more general model by the imposition of restrictions and a test is conducted to test the validity of those restrictions. These are nested hypothesis tests and they usually reduce to investigating whether it is valid to constrain certain coefficients in the general model to be zero, on the basis of the sample data. A feature of this approach is that the misspecification may be parameterised in the sense that if a misspecification is present, a parameter enters the model with a non-zero coefficient. Clearly, within this approach, we should attempt to isolate the particular misspecifications that are present with a view to eliminating them. In practice it appears that the distinction between the two approaches has become somewhat blurred and the nested approach is often used as a pure significance test. However, it is desirable in conducting misspecification tests when more than one misspecification may be present, to isolate and identify the separate effects. Some discussion of the problems of multiple significance testing may be found in Cox and Hinkley (1974) where they note that when testing a null hypothesis for several different kinds of departure the simplest approach is to set up separate statistics for the different departures and combine them into a function sensitive to departures in some or all of the different directions. This is particularly suitable if there may be several relatively small departures from the null. However, if a significant departure from the null is observed we clearly want to explore in detail the nature of the departure. Cox and Hinkley note that it is a serious criticism of the whole formulation of significance testing that no explicit treatment of this problem is available. Within the above procedure we simply take the most significant of the separate statistics as indicating the way in which the null is inadequate.

We shall later consider ways of conducting significance tests, designed to isolate, to some degree, particular departures from the null. However, one obvious problem with multiple significance tests concerns the significance level of the procedure. Suppose k tests are being conducted for up to k departures from H_0 where a significant result for at least one test leads to a rejection of H_0. If the tests are independent and each is conducted at significance level α^*, then the overall significance level is $1 - (1 - \alpha^*)^k \simeq k\alpha^*$ for small α^*. Thus if 5 independent significance tests are carried out at the 5% level, the overall significance level is $1 - (1 - 0.05)^5 = 0.226$.

If the test statistics are not independent then evaluation of the significance level requires a knowledge of their joint distribution. This is not often possible but upper and lower bounds may readily be calculated. Denoting the overall significance level by α, we have $\alpha^* \lesssim \alpha \lesssim k\alpha^*$, when $\alpha = \alpha^*$ when all the significance tests are exactly correlated. These bounds are relatively wide but offer some guide to α. In testing for misspecification in econometrics, the test statistics are usually dependent in small samples and so determining the overall significance level is difficult. This problem is exacerbated if the significance levels of the individual tests cannot be precisely controlled, i.e. if the tests are approximate rather than exact.

3. Some Recent Work on Joint Testing for Misspecification

Two recent papers by Savin and White (1978) and Lahiri and Egy (1981) both employed the Box and Cox (1964) procedure to test simultaneously for the correct functional form and for autocorrelation and heteroscedasticity respectively. Noting that in a time series regression model seemingly autocorrelated residuals are often the result of either disturbances following an autoregressive process or the estimation of an incorrect functional form or both, Savin and White argue the need to distinguish between these two misspecifications. To do this the Box and Cox procedure is generalised to disturbances which follow a first-order autoregressive process and it is then used to simultaneously test for autocorrelation and for the correct functional form.

The Box and Cox model is

$$\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{u}$$

where $y^{(\lambda)} = (y_1^{(\lambda)}, \ldots, y_T^{(\lambda)})'$ is a T x 1 vector of transformed observations on the dependent variable, X is a T x k matrix of rank (k < T) observations on k fixed regressors and $\underset{\sim}{u} = (u_1, u_2, \ldots, u_T)'$ is a T x 1 vector of random disturbances. The transformation is

$$y_t^{(\lambda)} = (y_t^{\lambda} - 1)/\lambda \qquad \lambda \neq 0$$
$$= \log(y_t) \qquad \lambda = 0.$$

Assuming that $\underset{\sim}{u} \sim NID(\underset{\sim}{0}, \sigma^2 I_T)$, the log likelihood function is given by $L(\lambda, \beta, \sigma^2: \underset{\sim}{y}, X) = -\frac{T}{2}\log(2\pi\sigma^2) = \frac{1}{2\sigma^2}(\underset{\sim}{y}^{(\lambda)} - X\underset{\sim}{\beta})'(\underset{\sim}{y}^{(\lambda)} - X\underset{\sim}{\beta}) + \log J$

where $J = \left| \frac{\partial \underset{\sim}{y}^{(\lambda)}}{\partial \underset{\sim}{y}'} \right| = \prod_{t=1}^{T} y_t^{\lambda-1}$ is the Jacobian of the transformation on the dependent variable.

If it is assumed that the disturbances, u_t, follow a stationary AR process,

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad |\rho| < 1,$$

then $E(\underset{\sim}{u}\,\underset{\sim}{u}') = V$ where the form of V is well known, see Theil (1971, p.252). The log-likelihood function is

$$L(\lambda, \rho, \beta, \sigma^2 : \underset{\sim}{y}, X) = -\frac{T}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(1-\rho^2) - \frac{1}{2\sigma^2}(\underset{\sim}{y}^{(\lambda)} - X\underset{\sim}{\beta})'V^{-1}(\underset{\sim}{y}^{(\lambda)} - X\underset{\sim}{\beta}) + \log(J).$$

Maximising this log likelihood with respect to $\underset{\sim}{\beta}$ and σ^2, given λ and ρ, we obtain the estimators

$$\underset{\sim}{\beta}(\lambda,\rho) = (X'V^{-1}X)^{-1}X'V^{-1}\underset{\sim}{y}^{(\lambda)},$$

$$\sigma^2(\lambda,\rho) = \frac{1}{T}(\underset{\sim}{y}^{(\lambda)} - X\hat{\underset{\sim}{\beta}}(\lambda,\rho))'V^{-1}(\underset{\sim}{y}^{(\lambda)} - X\hat{\underset{\sim}{\beta}}(\lambda,\rho)).$$

Substituting these terms into the log-likelihood function yields the concentrated likelihood function;

$$L(\lambda, \rho : \underset{\sim}{y}, X) = -\frac{T}{2}\{\log(2\pi) + 1\} - \frac{T}{2}\log(\hat{\sigma}^2(\lambda,\rho)) + \frac{1}{2}\log(1-\rho^2) + (\lambda - 1)\sum_{t=1}^{T}\log(y_t).$$

A test that the functional form is linear without the constraint $\rho = 0$, may be based upon the likelihood ratio. The critical region for $H_0: \lambda = 1$ against $H_1: \lambda \neq 1$ is given by

$$2\left[L(\lambda,\rho) - L(\lambda = 1, \rho)\right] > \chi^2_\alpha(1).$$

Similarly, one can test $H_0: \rho = 0$ given $\lambda = 1$ and the joint hypothesis $H_0: \lambda = 1, \rho = 0$ against $H_1: \lambda \neq 1, \rho \neq 0$, based upon the critical region

$$2\left[L(\lambda,\rho) - L(\lambda = 1, \rho = 0)\right] > \chi^2_\alpha(2).$$

Savin and White show through examples and through a Monte Carlo study, how one may make different inferences if one ignores the presence of one misspecification when testing the other. In particular, it is demonstrated how functional form misspecification may easily be misinterpreted as an autocorrelation problem.

Lahiri and Egy consider the joint testing of functional form and heteroscedasticity using essentially the same approach. However, heteroscedasticity of disturbances is assumed but no serial dependence. In particular, it is assumed that the disturbance variances are given by

$$\sigma_t^2 = \sigma^2 z_t^{\delta}$$

where z is known but δ is to be determined.

The concentrated log likelihood function in this case is

$$L(\lambda, \delta_i : \underset{\sim}{y}, X) = C - \frac{\delta}{2} \sum_{i=1}^{T} \log z_i + (\lambda - 1) \sum_{t=1}^{T} \log y_t - \frac{T}{2} \log \sigma^2(\lambda, \delta).$$

A likelihood ratio test may be carried out as before. Here the most general hypothesis allows for both non-linearity ($\lambda \neq 1$) and heteroscedasticity ($\delta \neq 0$). Lahiri and Egy show, through an example, how a wrong functional form may lead to an incorrect choice of heteroscedasticity parameter. Again joint testing is advocated.

In both of these studies we may note the following:

(a) The tests require maximum likelihood estimation under both the null and alternative hypotheses i.e. in both the restricted and unrestricted cases, and

(b) Several tests are applied to the same data ignoring any implications for type 1 and type 2 errors.

Bera and Jarque (1982) sought to overcome some of the problems of earlier studies by developing test procedures based upon the Lagrange multiplier principle. They noted that common misspecification test procedures are not robust in the presence of other misspecifications so that applying tests one at a time will often result in misleading conclusions. They proposed a nested hypothesis testing procedure based upon a general model, which permits the simultaneous testing of non-normality of disturbances, heteroscedasticity, autocorrelation and incorrect functional form. The actual model studied was

$$y_t = \sum_{j=1}^{K} x_{tj}^{(\lambda_j)} \beta_j + \sum_{j=1}^{M} d_{tj}\mu_j + \sum_{j=1}^{L} w_{tj}\,\delta_j + u_t, \qquad t = 1, 2, \ldots, T,$$

$$x_{tj}^{(\lambda_j)} = (x_{tj}^{\lambda_j} - 1)/\lambda_j \qquad \lambda_j \neq 0,$$

$$= \log\,(x_{tj}) \qquad \lambda_j = 0.$$

The $x_{tj}^{(\lambda)}$ represent transformed observations on K fixed regressor variables while the d_{tj} include the constant term and any dummy variables. The w_{tj} are observations on another set of fixed regressors and the disturbances, u_t, follow an autoregressive process

$$u_t = \gamma_1 u_{t-1} + \gamma_2 u_{t-2} + \cdots + \gamma_p u_{t-p} + \varepsilon_t$$

where the ε_t are independently distributed.

The density of ε_t is assumed to be a member of the Pearson family of distributions so that

$$g(\varepsilon_t) \propto \exp\left[\chi(\varepsilon_t)\right] \Big/ \int_{-\infty}^{\infty} \exp\left[\chi(\varepsilon_t)\right] d\varepsilon_t, \qquad -\infty < \varepsilon_t < \infty,$$
$$t = 1, 2, \ldots, T,$$

and

$$\chi(\varepsilon_t) = \int |(c_{1t} - \varepsilon_t)| / (c_{0t} - c_{1t}\varepsilon_t + c_2\varepsilon_t^2) d\varepsilon_t.$$

It is assumed that $c_{1t} = c_1$ for all t which implies that the mode of the distribution is the same for all t. Also $E(\varepsilon_t^2) = c_{0t}/(1-3c_{2t})$ and $g(\varepsilon_t) \sim N(0, c_{0t})$ when $c_1 = c_{2t} = 0$. The model is parameterised with $c_{2t} = c_2$ and the possibility of additive heteroscedasticity is introduced by putting $c_{0t} = \sigma^2 + \underset{\sim}{z}_t'\underset{\sim}{\alpha}$ where $\underset{\sim}{z}_t'$ is a $1 \times q$ vector of fixed variables.

When $\underset{\sim}{\lambda}' = (\lambda_i \ldots, \lambda_2, \ldots, \lambda_K) = (1, 1, \ldots, 1)$, $\underset{\sim}{\delta}' = (\delta_1, \delta_2, \ldots \delta_L) = \underset{\sim}{0}'$, $\underset{\sim}{\gamma}' = (\gamma_1, \gamma_2, \ldots, \gamma_p) = \underset{\sim}{0}'$, $\underset{\sim}{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_q) \doteq \underset{\sim}{0}'$ and $c_1 = c_2 = 0$, the model becomes

$$y_t = \bar{\underset{\sim}{x}}_t'\underset{\sim}{\beta} + \underset{\sim}{d}_t'\underset{\sim}{\mu} + u_t \quad , \quad t = 1, 2, \ldots, T,$$

where $\bar{\underset{\sim}{x}}_t' = \underset{\sim}{x}_t' - (1, 1, \ldots, 1)$ and the u_t are i, i, d, $N(0, \sigma^2)$. Within this framework the following tests may be carried out:

Normality: $H_0(N) : c_1 = c_2 = 0.$

Homoscedasticity: $H_0(H): \underset{\sim}{\alpha} = \underset{\sim}{0}.$

Serial independence: $H_0(I): \underset{\sim}{\gamma} = \underset{\sim}{0}.$

Correct functional form: $H_0(F): \underset{\sim}{\lambda} = \underset{\sim}{i}, \underset{\sim}{\delta} = \underset{\sim}{0}.$

Bera and Jarque show that if LM_{NHIF} is the LM test for testing, simultaneously, all the misspecifications, then under H_0: $c_1 = c_2 = 0$, $\underset{\sim}{\alpha} = 0$, $\underset{\sim}{\gamma} = 0$, $\underset{\sim}{\gamma} = i$, $\underset{\sim}{\delta} = 0$, LM_{NHIF} is asymptotically distributed as $\chi^2_{2+q+p+K+L}$ and $LM_{NHIF} = LM_N + LM_H + LM_I + LM_F$.

Thus the simultaneous test is the sum of the individual or one-directional LM tests. Hence tests can be carried out to test, simultaneously, for any combination of misspecifications. The additive property of the LM tests shows that for large samples the one-directional tests are independent and so overall significance levels can be controlled at least asymptotically.

One interesting feature of the model analysed is that a locally equivalent alternative model can be found which explains the additivity of the LM tests. Consider the following models

$$(1)\quad y_t = \sum_{j=1}^{K} x_{tj}^{(\lambda_i)} \beta_j + \sum_{j=1}^{M} d_{tj}\mu_j + \sum_{j=1}^{L} w_{tj}\,\delta_j + u_t, \qquad t = 1, 2, \ldots, T,$$

$$(2)\quad y_t = \bar{\underset{\sim}{x}}_t'\beta + \underset{\sim}{d}_t'\underset{\sim}{\mu} + \sum_{j=1}^{K} (\lambda_j - 1)\beta_j\, x_{tj} + \sum_{j=1}^{L} w_{tj}\delta_j$$

$$+ \sum_{j=1}^{2} c_j P_j(u_t) + \sum_{j=1}^{q} \alpha_j (z_{tj} u_t / 2\sigma^2) + \sum_{j=1}^{p} \gamma_j u_{t-j} + \eta_t, \quad t = 1, 2, \ldots, T$$

$$(3)\quad y_t = \bar{\underset{\sim}{x}}_t'\underset{\sim}{\beta} + \underset{\sim}{d}_t'\underset{\sim}{\mu} + u_t \qquad t = 1, 2, \ldots, T,$$

where in (2) the z_{tj} are measured from sample means, the η_t are NID(0, σ^2), $P_1(u_t) = u_t^2/3\sigma^2 + 1$ and $P_2(u_t) = u_t^3/4\sigma^4$.

Notice that both (1) and (2) reduce to (3) under the null hypothesis NHIF. It can be shown that the LM test for testing NHIF in (2) is the same as for (1). It is interesting to note that the one directional tests are, in fact, tests of the regression coefficients in (2) where the regressors associated with c_1, c_2, $\underset{\sim}{\alpha}$, $\underset{\sim}{\gamma}$, and $\underset{\sim}{\delta}$ are all asymptotically orthogonal except for regressors associated with $\underset{\sim}{\lambda}$ and $\underset{\sim}{\delta}$. This indicates why the one-directional tests are asymptotically independent.

Bera and Jarque carry out a Monte Carlo experiment using a simple version of the model discussed above, namely:

$$y_t = \underset{\sim}{d}_t'\underset{\sim}{\mu} + u_t \ , \ \underset{\sim}{d}_t' = (d_{t_1}, d_{t_2}, d_{t_3}, d_{t_4}), \quad t = 1, 2, \ldots, T.$$

Departures from the null arise when $w_{t1} = d_{t_2}^2$ is added into the regression or when $u_t = \rho u_{t-1} + \varepsilon_t$ (ρ = 0.3 or 0.7), $E(\varepsilon_t^2) = \sigma_t^2 = 25 + \alpha z_t$ ($\alpha \approxeq$ 0.25 or 0.85) or the disturbance is not normally distributed but follows a students $t_{(5)}$ or log normal distribution. Significance points for conducting the one-directional tests at the 10% level were obtained empirically. The results indicated that the one directional tests had low power against alternatives other than the immediate ones but violations of the maintained assumptions e.g. testing for a misspecification in the context of others, may lead to a considerable loss of power. With joint tests, overtesting results in little loss of power e.g. testing NHIF when $\overline{\text{NHIF}}$ holds, while undertesting, e.g. testing NH when $\overline{\text{NHIF}}$ holds, led to a considerable loss of power. In terms of power considerations alone, the joint tests were appealing but when the null was rejected there was uncertainty about the nature of the departure unless the individual tests were examined.

To help identify possible sources of departure from H_0, Bera and Jarque used a Multiple Comparisons procedure based on the four one directional tests. Denoting the significance levels for LM_N, LM_H, LM_I and LM_F by α_1, α_2, α_3 and α_4, respectively, the overall significance level i.e. the probability of rejecting H_0 when true on the basis of at least one significant test result, is given by

$$\alpha = \alpha_1+\alpha_2+\alpha_3+\alpha_4-\alpha_1\alpha_2-\alpha_1\alpha_3-\alpha_1\alpha_4-\alpha_2\alpha_3-\alpha_2\alpha_4-\alpha_3\alpha_4+\alpha_1\alpha_2\alpha_3+\alpha_2\alpha_3\alpha_4$$
$$+\alpha_1\alpha_3\alpha_4-\alpha_1\alpha_2\alpha_3\alpha_4.$$

Thus α_i were chosen, empirically, to get an overall α of 10% for comparison with the joint test LM_{NHIF}. The approach was found to have good power properties, comparable to LM_{NHIF}, and it was reasonably efficient at identifying departures in the N, H and I directions. It seems that the one directional tests were unstructured and all four tests were carried out at each iteration of the experiment.

A drawback to practical application of the approach, however, is that the critical values of the one directional tests have to be obtained through artificial data generation. The asymptotic critical values were not used presumably because they were substantially different from the correct ones.

4. Locally Equivalent Alternative Models

Godfrey and Wickens (1982) use a nested hypothesis test approach to derive misspecification tests for common misspecifications. They note that,with complex alternative models, it is often easier to compute an LM test rather than tests based upon the likelihood ratio and Wald principles and in many cases a further simplification is possible using a locally equivalent alternative model rather than the original alternative model. These locally equivalent alternatives are such that they provide statistics with the same asymptotic distribution as the LM test based on the original alternative model c.f. Godfrey (1981). Often the test statistic for the null model will be the same for a set of locally equivalent alternatives and with this approach it is possible to show that each type of misspecification can be incorporated in the locally equivalent alternative model as additional regressors.

In the regression model case, suppose we have the models:

(1) $f_t(y_t, \underset{\sim}{x}'_t : \underset{\sim}{\theta}) = u_t$,

(2) $f^*_t(y_t, \underset{\sim}{x}'_t : \underset{\sim}{\theta}) = u_t$

where (1) and (2) are different except when $\underset{\sim}{\theta}_2 = \underset{\sim}{0}$ and where $\underset{\sim}{\theta}' = (\underset{\sim}{\theta}'_1 : \underset{\sim}{\theta}'_2)$. Let $L_T(\underset{\sim}{\theta})$ and $L^*_T(\underset{\sim}{\theta})$ be, respectively, the log likelihoods for (1) and (2) where $L_T(\underset{\sim}{\theta}) = \sum_{t=1}^{T} l_t(\underset{\sim}{\theta})$ and $L^*_T(\underset{\sim}{\theta}) = \sum_{t=1}^{T} l^*_t(\underset{\sim}{\theta})$

and suppose that $\hat{\underset{\sim}{\theta}}$ maximises $L_T(\underset{\sim}{\theta})$ subject to H_0: $\underset{\sim}{\theta}_2 = \underset{\sim}{0}$. Then if model (2) is such that

(i) $\hat{\underset{\sim}{\theta}}$ maximises $L^*_T(\underset{\sim}{\theta})$ subject to H_0: $\underset{\sim}{\theta}_2 = \underset{\sim}{0}$, and

(ii) $\partial l^*_t(\hat{\underset{\sim}{\theta}})/\partial\underset{\sim}{\theta}_2 = \partial l_t(\hat{\underset{\sim}{\theta}})/\partial\underset{\sim}{\theta}_2$,

then, obviously, an appropriate LM test of H_0: $\underset{\sim}{\theta}_2 = \underset{\sim}{0}$ in both models may be based upon

$$LM = \left(\sum_{t=1}^{T} \partial l_t(\hat{\underset{\sim}{\theta}})/\partial\underset{\sim}{\theta}\right)' \left[\sum_{t=1}^{T} (\partial l_t(\hat{\underset{\sim}{\theta}})/\partial\underset{\sim}{\theta})(\partial l_t(\hat{\underset{\sim}{\theta}})/\partial\underset{\sim}{\theta})'\right]^{-1} \left(\sum_{t=1}^{T} \partial l_t(\hat{\underset{\sim}{\theta}})/\partial\underset{\sim}{\theta}\right).$$

To derive an appropriate model (2), Godfrey and Wickens show by expanding $l_t(\underset{\sim}{\theta})$ in a Taylor series about $\underset{\sim}{\theta}_2 = 0$, that

$l_t(\underset{\sim}{\theta}) = l^*_t(\underset{\sim}{\theta}) + R$

where R is a remainder term, not affecting $\partial l_t(\underset{\sim}{\theta}_1 : \underset{\sim}{0})/\partial\underset{\sim}{\theta}$, and

$f^*_t(y_t, \underset{\sim}{x}'_t : \underset{\sim}{\theta}) = f_t(y_t, \underset{\sim}{x}'_t, (\underset{\sim}{\theta}'_1, \underset{\sim}{0}')') + [\partial f_t(\underset{\sim}{\theta}'_1 : \underset{\sim}{0}')'/\partial\underset{\sim}{\theta}_2]\underset{\sim}{\theta}_2$.

A further approach yielding an alternative approximation is to take

$$f_t^{+}(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta}) = f_t(y_t, \underset{\sim}{x}_t', (\hat{\underset{\sim}{\theta}}_1' : \underset{\sim}{0}')') + [\partial f_t(\hat{\underset{\sim}{\theta}}_1' : \underset{\sim}{0}')'/\partial \underset{\sim}{\theta}_2] \underset{\sim}{\theta}_2$$

whereupon an asymptotically equivalent test procedure is obtained. Godfrey and Wickens note that this approximation yields a crucial simplification in that the test of H_0 is now formulated as one of testing the joint significance of a subset of regressors which are constructed from $\partial f_t(\hat{\underset{\sim}{\theta}})/\partial \underset{\sim}{\theta}_2$ and which enter $f_t^{+}(\underset{\sim}{\theta})$ linearly. Thus combinations of misspecifications may be tested, although, given that the tests involve testing the validity of omitting $\partial f_t(\hat{\underset{\sim}{\theta}})/\partial \underset{\sim}{\theta}_2$ from the specification $f_t^{+}(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta})$, it may be more convenient to use the LR or Wald tests rather than the LM test.

As an example, suppose that

$$f_t(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta}) = y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta} = u_t$$

where $\underset{\sim}{x}_t$ is a K x 1 vector of exogenous variables and the alternative hypothesis is that

$$u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots + u_p u_{t-p} + \varepsilon_t.$$

Then

$$f_t^{*}(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta}) = y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta} - \alpha_1 u_{t-1} - \alpha_2 u_{t-2} \dots - \alpha_p u_{t-p} = \varepsilon_t,$$

and

$$f_t^{+}(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta}) = y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta} = \alpha_1 \hat{u}_{t-1} - \alpha_2 \hat{u}_{t-2} \dots - \alpha_p \hat{u}_{t-p} = \varepsilon_t.$$

Similarly, if the alternative hypothesis is that

$$\sigma_t^2 = \sigma^2(1 + \underset{\sim}{z}_t'\underset{\sim}{\lambda})$$

where $\underset{\sim}{z}_t'$ is a 1 x q vector of fixed variables, then

$$f_t^{*}(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta}) = y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta} - \sum_{i=1}^{q} \lambda_i u_t Z_{ti} = \varepsilon_t,$$

$$f_t^{+}(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta}) = y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta} - \sum_{i=1}^{q} \lambda_i \hat{u}_t Z_{ti} = \varepsilon_t, \qquad t = 1, 2, \dots, T.$$

It is now clear how the use of the f_t approximation reduces the problem of testing for serial correlation and heteroscedasticity to a problem of testing the significance of regression coefficients. Clearly a joint test can be carried out by choosing

$$f_t^+(y_t, \underset{\sim}{x}_t', \underset{\sim}{\theta}) = y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta} - \sum_{i=1}^{p} \alpha_i \hat{u}_{t-i} - \sum_{i=1}^{q} \lambda_i \hat{u}_t z_{ti} = \varepsilon_t$$

and evaluating the joint significance of $\underset{\sim}{\alpha}$ and $\underset{\sim}{\lambda}$.

Godfrey and Wickens point out that if the standard F-test approach is used here rather than the LM test, a correction is required because $\hat{u}_t z_{ti}$ does not satisfy the usual assumptions of regressors in the general linear model in that

$T^{-\frac{1}{2}} \sum_{t=1}^{T} (\hat{u}_t z_{ti})\varepsilon_t$ has the same limiting distribution under H_0 as

$T^{-\frac{1}{2}} \sum_{t=1}^{T} (\varepsilon_t^2 - \sigma^2) z_{ti}$. Under H_0, the $\varepsilon_t^2 - \sigma^2$ are i.i.d. with mean zero and finite variance $2\sigma^4$. The limiting distribution of $T^{-\frac{1}{2}} \sum_{t=1}^{T} (\hat{u}_t z_{ti})\varepsilon_t$ is $N(0, \Delta_i)$ where $\Delta_i = 2\sigma^2 \operatorname{plim} T^{-1} \sum_{t=1}^{T} (\hat{u}_t z_{ti})^2$ which differs from standard formula in that the factor 2 appears here. The required adjustment to the standard formula is easy to make, however.

It is of interest to note that under H_0: $\underset{\sim}{\alpha} = \underset{\sim}{0}$, $\underset{\sim}{\lambda} = \underset{\sim}{0}$, the regressors in the above equation are asymptotically uncorrelated and the test statistics used to test for serial correlation and heteroscedasticity, separately, are asymptotically independent.

This approach to misspecification testing has obvious attractions; it provides a unified framework in terms of the formal F test and a number of possible combinations of misspecifications can be tested at a single step. However, if the overall null is rejected one might still wish to determine which particular misspecifications were causing the trouble. Some writers have expressed reservations about the power of tests based on LEA's, see Schonfeld (1982), particularly under non-local alternatives.

5. Additivity of Diagnostic Tests

In a most recent survey, Pagan and Hall (1983) show that most existing diagnostic tests can be derived from elementary principles of specification analysis without relying upon likelihood theory. They argue that applied researchers are primarily concerned with an examination of residuals as a principal way of determining model inadequacy.

Suppose we consider a standard regression model with T observations, then

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$$

where $\underset{\sim}{u} \sim N(\underset{\sim}{0}, \sigma^2 I_T)$.

Pagan and Hall consider a set of tests in which each diagnostic has an associated regression. This leads to the following regression equations:

Specification (S) $\quad \hat{u}_t = \bar{\underset{\sim}{z}}_t' \underset{\sim}{\gamma} + \varepsilon_t + (\hat{u}_t - u_t)$

Autocorrelation (A) $\quad \hat{u}_t = \rho u^*_{t-j} + \varepsilon_t + (\hat{u}_t - u_t)$

Heteroscedasticity (H) $\quad \hat{u}_t^2 = \sigma^2 + \underset{\sim}{z}_t' \underset{\sim}{\gamma} + (\varepsilon_t^2 - \sigma^2) + (\hat{u}_t^2 - u_t^2)$

Normality(N) (i) $\quad \hat{u}_t^4 - 3\hat{\sigma}^2 \hat{u}_t^2 = \gamma_3 + (\varepsilon_t^4 - \mu_4) - 3\sigma^2(\varepsilon_t^2 - \mu_2) + Op(1)$

(ii) $\quad \hat{u}_t^3 = \gamma_4 + \varepsilon_t^3 + \hat{u}_t^3 - u_t^3.$

When none of the misspecifications are present, $u_t = \varepsilon_t \sim NID(0.\sigma^2)$. The tests are based on significance tests of the coefficients in these equations and a necessary and sufficient condition that joint tests can be performed by adding the individual tests is that the coefficient estimates be uncorrelated, at least asymptotically. By analogy with SUR estimation of systems of regression equations this will hold if and only if either (a) the regressors in different equations have zero covariance or (b) the disturbances of the relevant equations are uncorrelated in large samples.

Applying these criteria, the asymptotic independence properties of these test statistics may be summarised as follows:

Conditions Required for Independence of Diagnostic Test Statistics

	S	A	H	$N_{(i)}$	$N_{(ii)}$
S	*				
A	C_1	*			
H	$\mu_3 = 0$	$\mu_3 = 0$ or C_1	*		
$N_{(i)}$	None	None	None	*	
$N_{(ii)}$	None	None	None	C_2	*

C_1 : y_{t-k} ($k \leq j$) appears in $\underset{\sim}{Z}_t$ of S test or in $\underset{\sim}{x}_t$.

C_2 : $\mu_7 - (\mu_4 - 3\sigma^4)\,\mu_3 - 3\sigma^2\mu_5 = 0$.

Additivity holds between these tests when lagged dependent variables are absent from the model and the disturbances are normal.

Pagan and Hall note the problems involved in testing for one misspecification in the presence of others. They suggest that joint estimation of the equations might be preferable if disturbances are non-normal and that the test for heteroscedasticity might be improved if a correction is made for autocorrelation of disturbances when autocorrelation is present. An overall strategy is not pursued, however.

Some evidence on the robustness of tests for autocorrelation and heteroscedasticity when both misspecifications are present, was provided by Epps and Epps (1977). They found that the Durbin Watson test was relatively robust to moderate heteroscedasticity, supporting the results of Harrison and McCabe (1975), while the Goldfeld-Quandt and Glejser tests were approximately valid if applied to a Cochrane-Orcutt transformed model whenever an initial test for autocorrelation indicates the presence of that problem. What is implicitly suggested by their study is that the misspecification tests should be carried out sequentially with the autocorrelation test being carried out first, followed by the heteroscedasticity test. The implications of their results for a structured approach to misspecification testing were not pursued, however.

6. A Sequential Approach to Testing Misspecification

In the recent econometric literature, see, especially, Mizon (1977), there has been much concern to develop an appropriate strategy for model selection. The practice of selecting models after applying numerous conventional tests of significance has well-recognised deficiencies. To overcome these problems, a search process has been advocated in which tests of specification are conducted on hypotheses within an overall maintained hypothesis which is carefully chosen to be the most general hypothesis likely to be relevant. If a composite hypothesis, representing the most restricted model, is tested against the maintained and not rejected, then the position is straightforward but when the restricted model is rejected, one does not know which of the constituent hypotheses are responsible. If hypotheses are orthogonal a powerful test of a joint hypothesis is achieved by testing each constituent hypothesis separately but orthogonal hypotheses are not a feature of econometrics. However, if the hypotheses are nested and uniquely ordered, then when any hypothesis is true all preceding hypotheses in the nest must be true and if any hypothesis is false all succeeding ones must be false. This has the advantage of allowing a composite hypothesis to be tested using a sequential procedure which can determine the hypotheses responsible for the rejection. Mizon notes that the sequential approach has certain optimal power properties in the class of procedures that fix the probabilities of accepting a less restricted hypothesis than the true one. Also, this approach, which is outlined in Anderson (1971), may be extended to non-linear models. An important characteristic of the approach is that the asymptotic distribution of the statistic for testing any hypothesis in the ordered sequence against the less restricted hypotheses immediately preceding it, depends on the validity of all less restricted hypotheses in the sequence but not on that of more restricted hypotheses, and each of these test statistics is asymptotically independent. Thus control over the overall Type 1 error probability is possible. If the significance level for each test is chosen at α_i, then the significance level of the implicit test of the r^{th} hypothesis is $1 - \prod_{i=1}^{r} (1 - \alpha_i)$ which is a monotonically non decreasing function of r.

Unfortunately, there are many nests of hypotheses which are not naturally ordered e.g. the determination of the orders of the autoregression and moving average processes in an ARMA model. In this situation it may be possible to impose more structure on the hypotheses to achieve a unique ordering or an exhaustive testing approach is followed in which all possible orderings are examined.

In the discussion of the sequential approach to specification testing the position of misspecification testing is somewhat obscure. Since misspecifications will usually invalidate the specification tests it seems that testing for misspecification is logically prior to testing the specification. Indeed, there seems to be no reason why the possible misspecifications should not be included in the maintained hypothesis. This is very easy to do when, as we have seen, the misspecifications enter the model as added variables.

In fact,Kiviet (1982) examined the overall significance of modelling strategies for simple dynamic models using a sequential procedure which tests a nest of hypotheses incorporating both misspecification and specification tests. The specification tests (or simplification tests) usually involve tests based on the Wald or LR principle whereas the misspecification tests are of the LM or LR type. The misspecification tests actually used were tests for serial correlation (of low and high order) and a post sample predictive performance test which provides an independent check on the chosen specification. When a sequence of tests in a uniquely ordered nest are performed by tests asymptotically equivalent to the LR tests, they are asymptotically independent when the composite hypothesis,representing the most restricted model, is true. Kiviet's Monte Carlo analysis demonstrated that this independence is apparent in relatively small samples, e.g. T = 20, 40, also. Thus the overall significance level may in principle be controlled at some desired level provided that the individual tests have Type 1 error probabilities in small samples close to the nominal significance level. Unless this is so, the overall significance level may be markedly different from the desired level when several tests are carried out. Kiviet's results showed that this may occur when relatively crude asymptotic tests are employed. Despite these difficulties it seems worth while to investigate a sequential approach to misspecification testing, particularly when it is required to test for more than one misspecification, since often the testing of the overall compound hypothesis may be based upon independent test statistics. To proceed, one looks for an ordering of the nested hypotheses such that the asymptotic distribution of the statistic for testing any hypothesis in the ordered sequence against the less restricted hypothesis immediately preceding it depends on the validity of all the less restricted hypotheses in the sequence but not on the validity of the more restricted hypotheses. In some cases there will be a natural choice of ordering whereas, in other cases, the ordering may be such as to permit the testing of more important constituent hypotheses first since the testing procedure stops when any hypothesis is rejected (see Hogg 1961, p. 980). At this point it is important to note an important difference between specification and misspecification testing. Specification tests are such that usually the alternative hypothesis reflects a specification that one might be prepared to accept, at least tentatively. On the other hand, a number of misspecification tests involve a choice of alternative hypotheses which would never be entertained seriously and this arises when a test is required to have reasonable power against a range of possible alternatives e.g. the version of the Reset test proposed by Ramsey (1969) or certain tests for heteroscedasticity and serial correlation. In sequential testing the idea is to allow for the possibility of several misspecifications so that the test statistic used at any particular stage is not affected by the possible presence of misspecifications yet to be tested. In such a case if a misspecification ,which is yet to be tested, is incorrectly modelled, then the validity of the sequential procedure is likely to be undermined to an unknown degree. When the sequential procedure stops because a significant test result is obtained one does not necessarily assume that a particular misspecification has been detected; merely that some misspecification is present. One's interpretation of the test result will be influenced by the nature of the case; in particular, by whether or not the alternative hypothesis is to be tentatively entertained. For example, when a significant heteroscedasticity test result is obtained, one might be unwilling to entertain the assumed alternative hypothesis although the presence of some form of heteroscedasticity might be suspected. On the other hand, when a standard test for structural change at a specified point in time is carried out and a significant result is obtained,one might be much more ready to accept the stated alternative i.e. to assume that a change in the parameters of the relationship has occurred. Clearly,what matters is the justification for the choice of the alternative in the first place.

7. Sequential Misspecification Testing: an Example

To illustrate the sequential approach to testing misspecification, we shall consider an example of a sequence of tests discussed in Phillips and McCabe (1984). To proceed, we assume a linear regression model in which

$$y = X\beta + \varepsilon$$

where y is a T x 1 vector of observations, X is a Txk matrix of rank k containing observations on k non-stochastic regressor variables, β is a kx1 vector of unknown parameters and ε is a Tx1 vector of unobservable normal random variables with mean zero and covariance matrix $\sigma^2 I_T$.

We assume that three types of misspecification are of interest:

(i) Serial correlation

(ii) heteroscedasticity, and

(iii) structural change in one or more components of β.

The serial correlation hypothesis is that the distrubances are generated by the first order autoregressive process:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t , \quad |\rho| < 1, \quad t=0,1, 2, \ldots, T.$$

Let e be the Tx1 vector of least squares residuals which results from the above estimation; then a common test statistic for the null hypothesis H_0: $\rho = 0$, is the Durbin-Watson bounds test statistic

$$d = \frac{e'Ae}{e'e} \quad .$$

An alternative approach is to use an exact test which is based upon transformations of the least squares residuals which have a scalar covariance matrix i.e. the BLUS residuals proposed by Theil (1965) or the recursive residuals proposed by Phillips and Harvey (1974). The test statistic employed is given by

$$d^* = \frac{\underset{\sim}{\varepsilon}'^* A^* \underset{\sim}{\varepsilon}^*}{\underset{\sim}{\varepsilon}'^* \underset{\sim}{\varepsilon}^*}$$

where $\underset{\sim}{\varepsilon}^*$ is a $(T-k)\times 1$ vector of BLUS or recursive residuals and A^* is a $(T-k)\times(T-k)$ version of A. Under H_0, d^* is distributed as the modified von Neumann ratio and its distribution is tabulated, for example, in Theil (1971 p. 726).

The heteroscedasticity hypothesis we shall examine is that the model is given by

$$\begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \end{pmatrix} = X\underset{\sim}{\beta} + \begin{pmatrix} \underset{\sim}{\varepsilon}_1 \\ \underset{\sim}{\varepsilon}_2 \end{pmatrix}$$

where $\underset{\sim}{\varepsilon}_i \sim N(0, \sigma_i^2 I_{T_i})$, $i = 1, 2$, with $\sigma_1^2 \neq \sigma_2^2$ and $T_1, T_2 > k$. The null hypothesis is chosen as H_0: $\sigma_1^2 = \sigma_2^2$ and the appropriate test statistic is the Variance Ratio (VR) given by

$$F = \frac{RSS_2/(T_2-k)}{RSS_1/(T_1-k)}$$

where RSS_i is the residual sum of squares from a regression carried out on the corresponding T_i observations, i=1, 2. Note that no reordering of the observations is involved and under H_0, $F_1 \sim F(T_2-k, T_1-k)$. Finally, the structural change hypothesis is that the model is given by

$$\begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \underset{\sim}{\beta}_1 \\ \underset{\sim}{\beta}_2 \end{pmatrix} + \begin{pmatrix} \underset{\sim}{\varepsilon}_1 \\ \underset{\sim}{\varepsilon}_2 \end{pmatrix} .$$

The null hypothesis is H_0: $\underset{\sim}{\beta}_1 = \underset{\sim}{\beta}_2 = \underset{\sim}{\beta}$, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, although the the assumption $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is often implicit rather than explicit in discussions of testing for structural change. The appropriate test is the analysis of covariance (AOC) test based upon

$$F_2 = \frac{(RSS - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(T-2k)}$$

where RSS is the residual sum of squares based on all T observations. Under H_0, $F_2 \sim F_{(k, T-2k)}$.

The independence of the VR and AOC statistics under H_0 is demonstrated in Phillips and McCabe (1983).

It is important to note that when testing H_0; $\rho = 0$, using d* (or d), it is assumed that both $\sigma_1^2 = \sigma_2^2$ and $\beta_1 = \beta_2$ hold. On the other hand, when testing H_0: $\sigma_1^2 = \sigma_2^2$, it is assumed that $\rho = 0$ but it is not assumed that $\beta_1 = \beta_2$. Finally, when testing H_0: $\underset{\sim}{\beta}_1 = \underset{\sim}{\beta}_2$, it is assumed that $\rho = 0$ and $\sigma_1^2 = \sigma_2^2$.

The sequence of hypotheses we wish to test commencing with the most general and testing in increasing order of restrictiveness is:

H_3: $\rho \neq 0$, $\sigma_1^2 \neq \sigma_2^2$, $\underset{\sim}{\beta}_1 \neq \underset{\sim}{\beta}_2$,

H_2: $\rho = 0$, $\sigma_1^2 \neq \sigma_2^2$, $\underset{\sim}{\beta}_1 \neq \underset{\sim}{\beta}_2$,

H_1: $\rho = 0$, $\sigma_1^2 = \sigma_2^2$, $\underset{\sim}{\beta}_1 \neq \underset{\sim}{\beta}_2$,

H_0 $\rho = 0$, $\sigma_1^2 = \sigma_2^2$, $\underset{\sim}{\beta}_1 = \underset{\sim}{\beta}_2$.

Alternatively, we wish to decide which (if any) of the following null hypotheses are true:

H_3*: $\rho = 0$,

H_2*: $\rho = 0$, $\sigma_1^2 = \sigma_2^2$,

H_1*: $\rho = 0$, $\sigma_1^2 = \sigma_2^2$, $\underset{\sim}{\beta}_1 = \underset{\sim}{\beta}_2$.

With this ordering, if any null hypothesis is true the preceding hypotheses must be true, and if any hypothesis is false the succeeding ones must be false. The hypotheses are tested in turn starting with H_3^* until either one rejects H_i^* and so decides H_i, i = 1, 2, 3, or one accepts all hypotheses and arrives at H_0.

Now we require that the distribution of the statistic for testing any hypothesis in the ordered sequence against the less restricted hypothesis immediately preceding it should not depend on the validity of more restricted hypotheses. Clearly this requirement rules out a test of H_3^* based on d^*, since the null distribution holds only when H_1^* is true. Since we cannot use d^*, we consider versions of d^* based upon separate sets of T_i-k recursive residuals, i = 1, 2. These are

$$d_i^* = \frac{\underset{\sim}{\varepsilon}_i'^* A_i^* \underset{\sim}{\varepsilon}_i^*}{\underset{\sim}{\varepsilon}_i'^* \underset{\sim}{\varepsilon}_i^*}, \quad i = 1, 2.$$

Under H_3^*, these statistics are independently distributed as the modified von Neumann ratio and their distributions do not depend on the σ_i^2 or $\underset{\sim}{\beta}_i$, i = 1, 2. Additionally, they and the VR and AOC test statistics form a mutually independent set under H_0, as is shown in Phillips and McCabe (1984).

An exact test for serial correlation can be carried out by pooling the results of separate tests based on the d_i^*, i = 1, 2. If we decide to reject the hypothesis of no serial correlation if <u>either</u> test rejects, then choosing a size for each test of 2½% will yield an overall test size of 5%. An alternative approach is to base the test on

$$d^{**} = \frac{\underset{\sim}{\varepsilon}_1'^* A_1 \underset{\sim}{\varepsilon}_1^*}{\underset{\sim}{\varepsilon}_1'^* \underset{\sim}{\varepsilon}_1^*} + \frac{\underset{\sim}{\varepsilon}_2'^* A_2 \underset{\sim}{\varepsilon}_2^*}{\underset{\sim}{\varepsilon}_2'^* \underset{\sim}{\varepsilon}_2^*} = d_1^* + d_2^* .$$

The resulting test is not exact because the distribution of the sum of two independent modified von Neumann ratios is unknown. However the distribution of d** can be approximated by the distribution of a β variate with the same mean and variance and so a test based on d** is close to being exact.

8. The Sequential Misspecification Testing Procedure

In applying the sequential test procedure the serial correlation hypothesis is tested first and if the null hypothesis is not rejected the heteroscedasticity hypothesis is tested. If again the null hypothesis is not rejected, the structural change hypothesis is tested. If all three tests fail to reject, the overall null hypothesis H_0 is supported. Since under H_0 the test statistics in section 7 are mutually independent, the overall size of the sequential test procedure can be controlled. For example, an overall test size of 5% is achieved if each test is carried out at the 1.7% level although one may not wish to choose the same size for each test. When a significant test result is achieved the sequential procedure stops since it is assumed that a misspecification has been detected i.e. the specification is rejected.

Note that the sequence of tests considered here could easily be augmented to include the Reset test for functional misspecification. Suppose that the version considered is as proposed by Thursby and Schmidt (1977). Then a set of test variables is added and their significance may be tested over the two sets of observations taken separately. An exact test can be obtained as in the case of the exact test for serial correlation in section 7. The serial correlation test is again carried out first and this is done with the test variables included in the regressions. The Reset test is then done if serial correlation is rejected. If the heteroscedasticity and structural change tests are performed, the test variables are excluded from the associated regressions. Thursby (1981, 1982) presents Monte Carlo results which strongly suggest that the

Reset test statistic obtained when all the data is used together is independent of the usual heteroscedasticity and structural change test statistics when they are obtained with the Reset test variables excluded from the regression. In fact, the asymptotic independence of the Reset test and the heteroscedasticity test is demonstrated by Pagan and Hall as indicated on page 17 of this paper.

Finally, note that conventional significance tests on regression coefficients, e.g. the standard F-test, can also be included in the sequential procedure. The independence of the F statistic and the misspecification test statistics (when misspecifications are absent) follows directly from the results in Phillips and McCabe (1984).

Although the application of the sequential approach considered here is in the context of a restricted model, applications in much less restricted models are possible and further work will undoubtedly show this to be the case.

9. Conclusion

This brief look at recent developments in selected areas of econometric testing methodology illustrates the importance that econometricians attach to this aspect of econometric research. A structured approach to misspecification testing within which the overall type 1 error probability is controlled appears to be attainable, at least in the context of single equation models, following several of the approaches examined here. To this writer, two of the approaches seem particularly attractive; the approach in which misspecifications appear in the form of added regressors and the sequential procedure discussed in the previous section (or, possibly, a combination of the two). It may also be possible to use both methods to isolate, to some degree, particular types of misspecifications e.g. structural change, but we should not be too optimistic about doing this in many other cases. However, both methods may lead to more powerful uni-directional tests in cases where more than one misspecification is present since the possibility of additional misspecifications is allowed for.

Finally, it is to be noted that since the sequential approach commences from an overall maintained hypothesis chosen to be the most general likely to be relevant, the problems of pre-test bias and data mining which result when experimentation is used to determine a specification, should be considerably reduced compared with alternative procedures.

Overall, the points made in favour of the sequential approach to testing econometric models suggest that it is a promising area of future research.

REFERENCES

Anderson, T.W. (1971). The Statistical Analysis of Time Series, New York, Wiley.

Bera, A.K. and Jarque, C.M. (1982). "Model Specification Tests: A Simultaneous Approach." Journal of Econometrics, 20, 59-82.

Box, G.E.P. and Cox, D.R. "An Analysis of Transformations." Journal of the Royal Statistical Society B 26, 211-243.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. Chapman and Hall.

Epps, T.W. and Epps, M.L. (1977). "The Robustness of Some Standard Tests for Autocorrelation and Heteroscedasticity when Both Problems are Present." Econometrica, 45, 745-753.

Godfrey, L.G. (1981). "On the Invariance of the Lagrange Multiplier Test With Respect to Certain Changes in the Alternative Hypothesis." Econometrica 49, 1443-55.

Godfrey, L.G. and Wickens, M.R. (1982). "Tests of Misspecification Using Locally Equivalent Alternative Models." Evaluating the Reliability of Macroeconomic Models (G. Chow and P. Corsi, eds.) London: Wiley.

Harrison, M.J. and McCabe, B.P.M. (1975). "Autocorrelation with Heteroskedasticity: A Note on the Robustness of the Durbin-Watson, Geary, and Henshaw Tests", Biometrika, 62, 215-216.

Hendry, D.F. (1980). "Econometrics: Alchemy or Science?" Economica 47, 387-406.

Hogg, Robert B. (1961). "On the Resolution of Statistical Hypotheses." Journal of the American Statistical Association 56, 947-989.

Kiviet, J.F. (1982). "Size, Power and Interdependence of Tests in Sequential Procedures for Modelling Dynamic Relationships." Unpublished paper, University of Amsterdam.

Lahiri, K.and Egy, D. "Joint Estimation and Testing for Functional Form and Heteroscedasticity." Journal of Econometrics 15, 299-307.

Mizon, G.E. (1977). "Inferential Procedures in Nonlinear Models: An Application in a UK Industrial Cross Section Study of Factor Substitution and Returns to Scale." Econometrica 45, 1221-1242.

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

Phillips, G.D.A. and Harvey, A.C. (1974). "A Simple Test for Serial Correlation in Regression Models." Journal of the American Statistical Association 69, 935-939.

Phillips, G.D.A. and McCabe, B.P.M. (1983). "The Independence of Tests for Structural Change in Regression Models." Economic Letters 12, 283-287.

Phillips, G.D.A. and McCabe, B.P.M. (1984). "A Sequential Approach to Testing Econometric Models." Paper presented at the 1984 European Meeting of the Econometric Society, Madrid.

Ramsey, J.B. (1969). "Tests for Specification Errors in Classical Least Squares Regression Analysis." Journal of the Royal Statistical Society, Series B, 31, 350-371.

Sargan, J.D. (1975) in G.A. Renton, ed., Modelling the Economy. Heinemann Educational Books Ltd. Contribution to the discussion of a paper by D.F. Hendry.

Savin, N.E. and White, K.J. (1978). "Estimation and Testing for Functional Form and Autocorrelation: a Simultaneous Approach." Journal of Econometrics 8, 1-12.

Schonfeld, P. (1982) in Evaluating the Reliability of Macroeconomic Models (G. Chow and P. Corsi, Eds.) London, Wiley. Contribution to the discussion of a paper by Godfrey and Wickens.

Theil, H. (1965). "The Analysis of Disturbances in Regression Analysis." Journal of the American Statistical Association 60, 1067-1078.

——— (1971). Principles of Econometrics, Macmillan, New York.

Thursby, J.G. (1981). "A Test Strategy for Discriminating Between Autocorrelation and Misspecification in Regression Analysis." Review of Economics and Statistics, 63, 117-123.

——— (1982). "Misspecification, Heteroscedasticity and the Chow and Goldfeld-Quandt Tests." Review of Economics and Statistics, 64, 314-321.

Thursby, J.G. and Schmidt, P. (1977). "Some Properties of Tests for Specification Error in a Linear Regression Model." Journal of the American Statistical Association 72, 635-641.