



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

AMSTER

AE 18/85

WITHDRAWN

INTERFACULTEIT DER
ACTUARIELE WETENSCHAPPEN
EN ECONOMETRIE

A & E RAPPORT

REPORT AE 18/85

TESTING STRATEGIES FOR MODEL SPECIFICATION

J.F. Kiviet and G.D.A. Phillips



Universiteit van Amsterdam

University of Amsterdam, Faculty of actuarial science & econometrics

Title: TESTING STRATEGIES FOR MODEL SPECIFICATION

Authors: Jan F. Kiviet (University of Amsterdam) and
Garry D.A. Phillips (University of Leeds)

Address: Faculty of Actuarial Science and Econometrics
University of Amsterdam
Jodenbreestraat 23
1011 NH Amsterdam

Date: November 1985

Series and Number: Report AE 18/85

Pages: 44

Price: No charge

JEL Subject Classification: 210

Keywords: Neyman-Pearson, sequential testing, specification tests,
misspecification tests, independence of test statistics

Abstract:

A crucial element in the development of econometric methodology during the past decade has been the concern with testing as opposed to estimating econometric models. In this paper we discuss - especially for the econometric analysis of time series - the main types of test procedures, and we also investigate the opportunities to uphold the Neyman-Pearson theory in the context of thorough model specification testing.

TESTING STRATEGIES FOR MODEL SPECIFICATION

by

Jan F. Kiviet

Actuarial Science & Econometrics
University of Amsterdam

Garry D.A. Phillips

School of Economic Studies
University of Leeds

Abstract

A crucial element in the development of econometric methodology during the past decade has been the concern with testing as opposed to estimating econometric models. In this paper we discuss - especially for the econometric analysis of time series - the main types of test procedures, and we also investigate the opportunities to uphold the Neyman-Pearson theory in the context of thorough model specification testing.

In applied work it is quite usual to carry out several tests on the same set of sample data. We consider an extension of the Neyman-Pearson framework to the case of such repeated testing, and examine situations where the various hypotheses under test have a particular nesting structure. For the case where a sequence of superposed alternatives is tested by so-called marginal tests we prove that the various test statistics are asymptotically independent under a common null hypothesis if the statistics are based on either the likelihood ratio, or the Wald, or the Lagrange-multiplier approach. Testing a particular null hypothesis against a series of juxtaposed alternatives appears to lead to independent test statistics only in specific circumstances. It is shown how independence of test statistics enables the control over the overall type I error probability, which is an essential element in the Neyman-Pearson theory.

Using the notions of constructive hypotheses and auxiliary hypotheses we can draw a clear distinction between specification tests and misspecification tests. Next an overview is given of approaches to and examples of specification and misspecification testing. With respect to the former attention is paid to the problem determining the order of dynamics and discriminating between system dynamics and error dynamics. Then misspecification testing is reviewed for: specification error, non-constancy of coefficients, heteroscedasticity, serial dependence, and non-normality of disturbances. Also the problem of testing for several misspecifications jointly or sequentially is considered.

Finally we discuss the options and associated difficulties in implementing the various tests in an overall testing strategy.

TESTING STRATEGIES FOR MODEL SPECIFICATION*

1. Introduction

An econometric model is considered to be an analytical representation of theories of economic behaviour. This representation is dependent upon a statistical implementation for purposes of hypothesis testing and parameter estimation, or for use in prediction or simulation circumstances. A model in this sense may range from a single linear equation to a complicated set of non-linear simultaneous equations.

A crucial element in the development of econometric methodology during the past decade has been the concern with testing as opposed to estimating econometric models. By testing a model we mean the procedure through which a model is compared with actual economic data to determine whether the model is, in fact, a reasonable representation of the process which actually generated the data, i.e. the data generation process (DGP).

The interpretation of test statistics employed in any given situation depends on the model building strategy that has been adopted to the point when testing takes place. It may be that the model under test has been fully articulated on the basis of theoretical or a priori considerations and/or other independent empirical studies and, in such a case, the employment of the test statistics constitute genuine tests in the Neyman Pearson sense. Many of the tests used in econometrics are discussed in this context. What is much more likely is a situation in which the model itself is selected to some degree as a result of a data based strategy in which the test outcomes themselves act as a guide to specification. In the extreme case the model may be heavily dependent on test outcomes through constant revision to eliminate misspecifications so that all meaningful connection with the Neyman Pearson theory is lost and the calculated test statistics are useful for descriptive purposes only.

* The authors wish to thank J.S. Cramer and B.P.M. McCabe for helpful comments.

An approach which preserves the Neyman-Pearson theory to some degree, proceeds by specifying a general or maintained model, chosen to be the most general given existing theoretical knowledge, empirical results from other studies and the available data etc. The appropriateness of this initial choice for the maintained hypothesis is checked through the use of misspecification tests and if misspecification is not indicated, it is assumed that a suitable specification, i.e. description of the DGP, can be found within the general model by imposing restrictions. A restricted version of the general model may subsequently be selected following the sequential testing of restrictions whereby restrictions that cannot be rejected are imposed until an appropriate parsimonious representation is obtained. Of course, following the selection of an appropriate model, we may still wish to conduct further testing and, in particular, to compare its performance against other, possibly non-nested, competing models. However, our primary concern here is with approaches or strategies for testing econometric models either to the point where the model is rejected as being inadequate or to the point where an appropriate specification is chosen, possibly only tentatively, as an adequate representation of the DGP.

An attempt at developing a framework for systematically evaluating econometric models was presented by Dhrymes et al (1972) in an important paper which presaged many of the recent developments in econometric test methodology. More recently, attention has focussed on an overall strategy for model selection, see for example, Harvey (1981, Section 6) and this is the approach we adopt in this paper, where we discuss the main features of model testing in econometrics and investigate the opportunities to uphold the Neyman Pearson theory in practice. In Section 2 we present a general econometric model and in Section 3 we draw a distinction between specification tests and misspecification tests. In Section 4 parametric tests in a maximum likelihood framework are reviewed and then in Section 5

we examine the options to control the overall significance level when the same set of data is used repeatedly for testing purposes. After that we give examples of procedures for specification testing in Section 6 and in Section 7 we do the same for misspecification testing. In Section 8 we consider methods for testing a number of misspecifications simultaneously. Finally, in Section 9 we indicate how a general testing strategy for model specification can be devised and mention the many problems that hamper the full preservation of the Neyman Pearson framework in the practice of model testing in econometrics.

2. Econometric Models

We suppose that the problem being investigated involves a set of observable variables. An econometric model typically consists of a hypothetical description of the process which has generated the particular set of variables which are of interest. We shall distinguish between this unknown *data generation process*, the DGP, and the *econometric model* which attempts to characterise it. The econometric model itself will typically be characterised by (a) a choice of jointly dependent or endogenous variables, y_t , (b) a choice of conditioning or exogenous variables, z_t , and (c) a set of hypothesised density functions

$$f(y_t | z_t, Z_{t-1}, Y_{t-1}, \beta), \quad t = 1, 2, \dots, T, \quad (1)$$

where β is an $m \times 1$ vector of unknown parameters and Z_{t-1} and Y_{t-1} are past observations on z and y . We shall write these density functions as

$$f(y_t | x_t, \beta) \quad (2)$$

where x_t is the set of all conditioning variables and this form implies that, conditional on x_t , the endogenous variables y_t are uncorrelated with any variables from $\{z_t, Z_{t-1}, Y_{t-1}\}$ not included in x_t .

The density functions in (2) are hypothesised as representations of the DGPs. However, in the nature of things they will differ from the DGPs since we can never hope to model a DGP completely. In testing econometric models within the Neyman Pearson framework we are obliged to assume that an econometric model which will adequately mimic the DGP is, in principle, achievable and in carrying out a test the maintained hypothesis is not questioned, i.e. it is assumed to represent the DGP.

Further discussion of DGPs and their relationship to empirical models may be found in Richard (1980) and Hendry and Richard (1982) and Hendry et al (1983).

3. Specification and Misspecification tests

Most of the tests that are used in econometrics are based implicitly or explicitly, on either the Likelihood Ratio, the Wald or Lagrange Multiplier principles. A likelihood ratio (LR) test compares the values of the maximised likelihood on both null and alternative hypotheses and so it involves estimating the parameters of the models corresponding to both hypotheses. The Wald (W) test only requires the estimation of the maintained hypothesis whereupon it tests whether these unrestricted estimates satisfy the restrictions of the null hypothesis. Finally, the Lagrange Multiplier (LM) test requires only the estimation of the restricted model and it checks the need for a more general model by testing the significance of the Lagrange multipliers which are used to impose the restrictions implicit in the estimated model relative to the maintained hypothesis. We shall examine some relationships between and properties of these tests in the next section. Further discussion of the principles involved and the characteristics of the tests is given in Seber (1966) and Silvey (1975).

Econometricians are interested in two basic types of tests. Firstly there are tests which relate to the parameters of the maintained hypothesis i.e. tests which concern hypotheses involving the parameter vector β in the model given in (2). These tests will check to see if a more restricted model is appropriate and typically the LR or W tests will be applicable. In such tests the null and alternative hypotheses are clearly stated; we call them specification tests. They are used to test whether the maintained hypothesis can be simplified and each hypothesis tested can be considered as a potentially adequate representation of the DGP. We shall refer to each such hypothesis as a *constructive hypothesis*. Thus the most general

of the constructive hypotheses considered is the *maintained hypothesis*.

A second class of tests is concerned, essentially, to test whether the maintained hypothesis is general enough. Such tests check whether the hypothesised density in (2) or some constructive restricted version of it can successfully mimic the hypothetical DGP by extending the parameter space and testing the significance of the extra parameters; we call these misspecification tests. Such tests are typically concerned with parameters that do not form part of the initial maintained hypothesis and they may be concerned with the form of the densities involved, the exogeneity assumptions made and other ways in which the maintained hypothesis may be incorrect. When a constructive hypothesis is extended by reparameterisation a specification might be obtained which can usefully challenge it although the reparameterisation would never be accepted as a satisfactory representation of the DGP. This type of hypothesis will be called an *auxiliary hypothesis*.

We are now in a position to make a distinction between a specification test and a misspecification test, see Mizon (1977). In a *specification test* both the null and alternative hypotheses embody constructive hypotheses while in a *misspecification test* the null is constructive and the alternative is of the auxiliary type. Obviously pure significance tests, i.e. tests in which the alternative is, essentially, unspecified, are necessarily misspecification tests. However, the distinction between the tests should be interpreted with care since the same parameters may be involved in both types of test depending upon the particular maintained hypothesis that is chosen. Nonetheless, vagueness about the alternative is an essential characteristic of the misspecification test. Both types of test have application in a testing strategy and we shall return to this topic in later sections.

4. Some Results for Likelihood Ratio, Wald and Lagrange Multiplier Tests.

We commence by noting some well known characteristics of maximum likelihood (ML) estimators.

Let $L(\beta)$ denote the log likelihood of a sample of T mutually independent observations where $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$ is an $m \times 1$ vector of parameters. We assume that the usual regularity conditions apply and we define the score vector as

$$q(\beta) = \frac{\partial L(\beta)}{\partial \beta} \quad (3)$$

and the $m \times m$ Hessian matrix of $L(\beta)$ to be

$$Q(\beta) = \frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} \quad (4)$$

If β is the parameter vector for the process which generated the data, then

$$H = -E_{\beta} Q(\beta) \quad (5)$$

is the Information Matrix which depends only on the sample data and β and which is assumed to be non singular. The unrestricted ML estimator for β which we designate by $\tilde{\beta}$ is obtained by solving the m equations

$$q(\tilde{\beta}) = 0.$$

For the asymptotic distribution of the score it is well known that

$$T^{-\frac{1}{2}} q(\beta) \overset{d}{\rightarrow} N(0, T^{-1}H) \quad (6)$$

and if we expand $q(\tilde{\beta})$ in a Taylor series expansion about β , we have after some rearrangement

$$(\tilde{\beta} - \beta) \approx -Q^{-1}(\beta) q(\beta) \quad (7)$$

and

$$T^{\frac{1}{2}}(\tilde{\beta} - \beta) \overset{d}{\rightarrow} N(0, T.H^{-1}). \quad (8)$$

Consider next the case where β obeys a set of $r < m$ non-linear restrictions which are represented by an r element vector function

$$\phi(\beta) = 0.$$

Writing the restricted ML estimator as $\hat{\beta}$, the estimator is obtained by solving the first order conditions obtained after differentiating the Lagrangian function given by $L(\beta) - \lambda' \phi(\beta)$ where λ is an $r \times 1$ vector of indeterminate multipliers. The resulting differentiation yields the equations

$$q(\hat{\beta}) = F'(\hat{\beta})\lambda,$$

$$\phi(\hat{\beta}) = 0,$$

where $F(\beta) = \frac{\partial \phi(\beta)}{\partial \beta}$ is an $r \times m$ full row rank matrix of continuous

functions of β .

It is well known, see Silvey (1975, p.80), that

$$\hat{\beta} - \beta \approx \{H^{-1} - H^{-1}F'(\beta)[F(\beta)H^{-1}F'(\beta)]^{-1}F(\beta)H^{-1}\}q(\beta)$$

and on using (6) we may write

$$T^{\frac{1}{2}}(\hat{\beta} - \beta) \stackrel{d}{\rightarrow} N(0, T.H^{-1} - T.H^{-1}F'(\beta)[F(\beta)T.H^{-1}F'(\beta)]^{-1}F(\beta)T.H^{-1}). \quad (9)$$

To test $H_0: \phi(\beta) = 0$ against $H_A: \phi(\beta) \neq 0$, using the LR approach, the test statistic is

$$LR = -2[L(\hat{\beta}) - L(\tilde{\beta})].$$

On expanding $L(\hat{\beta})$ around $\tilde{\beta}$ and ignoring asymptotically negligible terms we have the result that

$$LR \approx (\hat{\beta} - \tilde{\beta})'Q(\tilde{\beta})(\hat{\beta} - \tilde{\beta}) \stackrel{d}{\rightarrow} \chi^2_{(r)}, \text{ under } H_0. \quad (10)$$

If the W approach is followed, the test statistic is

$$W = -\phi'(\tilde{\beta})[F(\tilde{\beta})Q^{-1}(\tilde{\beta})F'(\tilde{\beta})]^{-1}\phi(\tilde{\beta}) \quad (11)$$

and if the LM approach is employed

$$LM = -q'(\tilde{\beta}) Q^{-1}(\tilde{\beta}) q(\tilde{\beta}) \quad (12)$$

is derived. These three test statistics are asymptotically equivalent under H_0 and each is distributed as

$$LR = W = LM = (\hat{\beta} - \tilde{\beta})' H^{-1} (\hat{\beta} - \tilde{\beta}) \stackrel{a}{\sim} \chi^2_{(r)} \quad (13)$$

subject to regularity conditions.

Engle (1984) notes the important characteristics of these tests. Although the test statistics differ in small samples, they have the same limiting distribution when either the null hypothesis or local alternatives of the form $H_A: \phi(\beta) = \delta/T^{1/2}$ for some fixed $r \times 1$ vector δ , are true. In addition they are asymptotically locally most powerful invariant. Power comparisons for non-local alternatives based upon second order approximations are discussed by Rothenberg (1982), who also notes that the actual significance levels of commonly used tests may deviate substantially from nominal levels in finite samples. He advises using modified critical regions based on an examination of the Edgeworth expansion of the distribution functions. He notes that once the tests are size corrected the differences in the test powers appear to be small in most cases but, since power surfaces cross, no statements can be made to suggest that one test is uniformly best.

5. Repeated Testing

When testing an econometric model it is usual to carry out several tests on the same set of sample data. Here we consider the scope for maintaining the Neyman-Pearson (NP) framework in the context of such repeated testing. An essential element of the NP approach to testing is the exercise of control over the probability of committing a Type 1 error i.e. of rejecting the restricted hypothesis when it is true. Therefore, an extension of the NP framework to the case of more than one test is, strictly, possible only when the most restricted hypothesis considered is nested within the set of hypotheses under consideration. In such a case a Type 1 error is committed in repeated testing when the most restricted hypothesis is not accepted when it is true. Before continuing we clarify what is meant by *nested hypotheses*. The hypothesis defined by the probability density $f_0(y_t|x_t, \beta_0)$ is nested in the hypothesis $f_1(y_t|x_t, \beta_1)$ if, for any particular value β_0^* of the parameter vector β_0 , the probability density $f_0(y_t|x_t, \beta_0^*)$ can be approximated arbitrarily closely by $f_1(y_t|x_t, \beta_1)$ for some β_1 . If f_0 is nested in f_1 , then taking f_0 as the null hypothesis and f_1 as the alternative, we may be able to carry out a NP test. The associated test will be indicated by NP_0^1 .

Consider now the case where we have an additional test NP_0^2 , f_0 being nested in f_2 also. We shall distinguish two cases as follows: (i) where f_1 and f_2 are non-nested whereupon we shall say that NP_0^1 and NP_0^2 have *juxtaposed alternatives*, and (ii) where f_1 is nested in f_2 (or vice versa) in which case the tests have *superposed alternatives*. Where two related tests are considered these are the only cases where, in principle, a generalisation of NP testing is possible and this is so because there is a common null hypothesis. If the tests NP_0^1 and NP_3^1 are combined where

f_3 is nested in f_1 but f_0 and f_3 are non-nested, we have an example of a situation in which NP testing does not operate though the theory ^{Mc Aleer and} of testing non-nested hypotheses, see Mackinnon (1983) and Pesaran (1985), may well be applicable.

We shall now examine some results on LR, W and LM tests in the context of the sequential testing of hypotheses and juxtaposed and superposed hypotheses will be dealt with separately. Note that hypotheses which have less nesting structure than these can scarcely be handled within an overall NP framework.

Let f_0 be the density of the null hypothesis which is common to the l juxtaposed alternatives f_j , $j = 1, \dots, l$, then on noting the result in (12), the tests NP_0^j can be approximated under the null by

$$NP_0^j = q_j'(\beta_0) H_j^{-1} q_j(\beta_0) \quad (14)$$

where H_j is the information matrix and $q_j(\beta_0)$ is the score vector of f_j evaluated under the null. For a sequence of tests the *overall significance level* is the probability that one or more tests in the sequence rejects the null when f_0 is correct. Thus if α_j is the size of the individual test NP_0^j then the overall size, α , satisfies the inequality

$$\max \{ \alpha_j | j = 1, \dots, l \} < \alpha \leq \min \left\{ 1, \sum_{j=1}^l \alpha_j \right\}. \quad (15)$$

The lower bound here would only be attained when the tests are completely dependent but this is scarcely conceivable when juxtaposed hypotheses are involved since they are non-nested and thus cannot be equivalent. When the tests are independent we have

$$1 - \alpha = \prod_{j=1}^l (1 - \alpha_j) \quad (16)$$

and the overall significance level can be exactly controlled.

From (6) and (14) we note that the juxtaposed tests NP_0^j and NP_0^i are asymptotically independent if the asymptotically standard normal vectors $H^{-\frac{1}{2}}q_j(\beta_0)$ and $H^{-\frac{1}{2}}q_i(\beta_0)$ have a zero covariance matrix.

In the special case where the sequence of l juxtaposed tests NP_0^j , $j = 1, \dots, l$, are all mutually independent (examples are given in

section 8) the overall test of f_0 may be based on the combined statistic $\sum_{j=1}^l NP_0^j$ with the significance level chosen as α . The tests are then said to be *additive*. However inferences based on this combined statistic may differ from those based on sequential testing depending on the actual density of the DGP and the choice of the α_j .

In the testing of superposed hypotheses we have, in addition to f_0 , a sequence of alternative hypotheses f_i , $i = 1, \dots, l$, where f_{i-1} is nested in f_i . Three different types of sequences of tests emerge, namely NP_0^i , NP_{i-1}^l and NP_{i-1}^i for $i = 1, \dots, l$ respectively. The first two concern the testing of *overlapping hypotheses* which will always involve dependent test statistics. However the third sequence involves what we shall call *marginal tests* and Hogg (1961) notes that if the marginal tests are based on the LR test approach then often they are mutually stochastically independent under the overall null hypothesis. The argument for this assumes that the LR tests at each stage are functions of complete sufficient statistics so that the theorem of Basu (1955) may be invoked. This is not justified in many cases of interest to econometricians since often complete sufficient statistics cannot be found. However, we now show, following Kiviet (1982), that the marginal tests are asymptotically mutually independent for general superposed hypotheses. This result is noted in Sargan (1980) in respect of W tests though a proof is not given. We commence our proof by ordering the restrictions to be tested in increasing order of restrictiveness.

Let $\phi' = (\phi'_1, \phi'_2, \dots, \phi'_\ell)$ be a partition of the r element vector function ϕ where we suppose that there is a natural ordering of the hypotheses so that we should not wish to accept the s^{th} restriction $\phi_s = 0$ if the previous restriction $\phi_{s-1} = 0$ had been rejected and let $\phi^{(j)}$ denote the vector

$$\phi^{(j)}(\beta) = \begin{bmatrix} \phi_{(1)}(\beta) \\ \phi_{(2)}(\beta) \\ \vdots \\ \phi_{(j)}(\beta) \end{bmatrix}, \quad j = 1, \dots, \ell. \quad (17)$$

Note that $\phi^{(\ell)}(\beta) = \phi(\beta)$ while $\phi^{(1)}(\beta) = \phi_{(1)}(\beta)$. We are concerned with an ordered set of nested hypotheses where each successive null can be written as

$$H_0^{(j)} : \phi^{(j)}(\beta) = 0, \quad j = 0, \dots, \ell. \quad (18)$$

Here $H_0^{(0)}$ indicates the general maintained hypothesis / within which β is unconstrained.

Hence in terms of the foregoing notation $H_0^{(j)}$ corresponds to $f_{\ell-j}$.

The sequence of marginal tests now consists of testing $H_0^{(j)}$ against $H_A^{(j)}$, where $H_A^{(j)}$ is $H_0^{(j-1)} \cap \phi_{(j)}(\beta) \neq 0$, $j = 1, \dots, \ell$. At each stage of the procedure new restrictions $\phi_{(j)}(\beta) = 0$ are tested against $\phi^{(j-1)}(\beta) \cap \phi_{(j)}(\beta) \neq 0$.

To show the asymptotic independence of the marginal tests, let $\dot{\beta}^{(j)}$ denote the restricted ML estimator under the restrictions imposed by $H_0^{(j)}$, then if $\phi^{(j)}(\beta) = 0$, in large samples we have

$$H \approx -Q(\dot{\beta}^{(j)})$$

and so the marginal test $NP_{\ell-j}^{\ell-j+1}$ can be approximated by

$$(\dot{\beta}^{(j)} - \dot{\beta}^{(j-1)})' H^{-1} (\dot{\beta}^{(j)} - \dot{\beta}^{(j-1)}), \quad j = 1, \dots, \ell \quad (19)$$

on noting the result in (13).

Now

$$\begin{aligned} (\dot{\beta}^{(j)} - \beta) &\approx \{H^{-1} - H^{-1}F_j'(B)[F_j(B)H^{-1}F_j'(B)]^{-1}F_j(B)H^{-1}\}q(B) \\ &= H^{-\frac{1}{2}}M_jH^{-\frac{1}{2}}q(B) \end{aligned} \quad (20)$$

where $M_j = I_m - X_j(X_j'X_j)^{-1}X_j'$, $X_j = H^{-\frac{1}{2}}F_j'(B)$ and $F_j(B)$ is the matrix $\frac{\partial \phi^{(j)}}{\partial \beta}(B)$, evaluated at B , and so it follows that

$$(\dot{\beta}^{(j)} - \dot{\beta}^{(j-1)}) \approx H^{-\frac{1}{2}}[M_j - M_{j-1}]H^{-\frac{1}{2}}q(B). \quad (21)$$

Noting that $M_k M_j = M_k$ for $j \leq k$ and noting the result in (6), it may be shown directly that the asymptotically normal vectors $H^{-\frac{1}{2}}(\dot{\beta}^{(j)} - \dot{\beta}^{(j-1)})$ and $H^{-\frac{1}{2}}(\dot{\beta}^{(k)} - \dot{\beta}^{(k-1)})$ have a zero covariance matrix if $j \neq k$ and so under the overall null $H_0^{(\ell)}$ the sequence of marginal tests will be asymptotically mutually independent. Thus we can approximate the overall significance of the overlapping tests NP_0^i and NP_{i-1}^{ℓ} for $i = 1, \dots, \ell$ in view of the fact that

$$NP_j^k \approx \sum_{i=j+1}^k NP_{i-1}^i$$

under the validity of f_j and given that the marginal tests are additive asymptotically. Hence the marginal tests imply a sequence of overlapping tests. The fact that the NP approach to testing is preserved when superposed alternatives are tested by marginal tests can also be exploited when one is faced with a number of juxtaposed alternatives that involve dependent test statistics. One can form a sequence of comprehensive models in which an increasing number of juxtaposed alternatives are nested and these superposed comprehensive models may be tested using marginal tests so that the significance level may be controlled asymptotically. However a problem arises with marginal testing when there is no unique (natural) ordering of the restrictions and this is examined in the next section.

6. Specification Testing

Modern econometric practice advocates the need for the most thorough testing of any econometric model when it is desired to check whether an adequate representation of the DGP has been achieved and hypothesis testing is a natural vehicle for doing this. While there has been a great deal of work done on developing misspecification tests in econometric models there has been correspondingly less attention given to the area of specification testing. In particular there is no generally accepted approach to specification testing. Until recently the most common approach involved the application of numerous tests of significance, often in an ad hoc way, until a model was obtained in which there were a set of variables with statistically significant coefficients of the correct a priori sign. Such an approach was very often accompanied by little or no misspecification testing. This practice has been the subject of much criticism for the pre test bias that is typically introduced and the associated problems of data mining. Even so, much applied econometric work decides upon a particular model as an adequate representation of the DGP only after a good deal of experimentation has been carried out. While one can hardly criticise experimentation as a means of developing and refining a model one can certainly criticise the practice of presenting empirical findings with accompanying test results as though the tests were conducted in a Neyman Pearson framework. The adage "one cannot use the same data to both suggest and test a model" is constantly ignored. The problem has long been recognised, see Theil (1961 p.206) and, more recently, the problem has been addressed again, see Leamer (1978), Theil (1978) and Hendry (1983).

Recently the use of a search process has been advocated within which tests of specification are carried out in a structured way commencing from an overall maintained hypothesis which is carefully chosen to be the most general hypothesis likely to be relevant while taking advantage of any

nesting and ordering of relevant hypotheses which may be present. In addition the problems associated with non nesting are taken into account. This approach to testing which Anderson (1971, ch. 6) showed may lead to tests with optimal power properties, is discussed by Mizon (1977) and other applications are examined by Hendry/^{and Anderson} (1977) and Sargan (1980). A set of uniquely ordered hypotheses has the property that if any hypothesis is rejected then all succeeding hypotheses will also be rejected and so need not be tested. Thus if the sequential procedure begins with the maintained hypothesis, hypotheses are systematically tested in increasing order of restrictiveness until a significant test is encountered or the most restricted hypothesis is reached. The hypothesis accepted is the one immediately prior to the one which produced the significant result. Hypotheses are always tested against the immediately preceding hypothesis, hence (in our terminology) in this approach a sequence of marginal tests/^{of superposed hypotheses} is performed. As noted by Mizon (1977, p. 1225) the decision problem involves balancing the costs of accepting a less restricted hypothesis than the true one against those of accepting a more restricted hypothesis than the true one. If the maintained hypothesis is chosen to be very general and the significance levels for tests of hypotheses close to the maintained are chosen to be small then if a very general model is required there is a good chance of learning this whereas, if a general model is not needed, the probability that one such model is accepted will be low. Underlying this is the important result 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 less restricted hypotheses in the sequence but not on that of more restricted hypotheses. In addition, as noted in the previous section, the series of test statistics will be asymptotically mutually independent under the null of the last test performed. One problem is the choice of Type 1 error probabilities for each test. If α_j is the significance level of the j^{th} test in the sequence, then the significance level of the implicit integral test of the i^{th} hypothesis against the maintained is asymptotically

$1 - \prod_{j=1}^i (1 - \alpha_j)$, hence the significance levels of the implicit tests form a monotonically nondecreasing sequence. If there are l hypotheses in the ordered nest then the maximum significance level for the whole procedure can be controlled, at least, approximately. For example, if this maximum significance level is α , the α_j are determined so that

$$1 - \prod_{j=1}^l (1 - \alpha_j) = \alpha.$$

The application of this approach to testing is relatively straightforward when hypotheses are uniquely ordered but if this is not the case it may still be possible to obtain an appropriate ordering of the hypotheses through imposing some structure on the problem. In some cases, although a unique ordering cannot be found, a particular ordering can be imposed but the lack of uniqueness means that the hypotheses that follow a rejected hypothesis remain untested.

If one adopts an exhaustive approach, within which all possible orderings are considered, one may obtain a number of non-rejected, non-nested, constructive hypotheses. However, there does not seem to be a practicable approach within the Neyman Pearson framework for discriminating amongst them and one usually resorts to the application of selection criteria such as measures of goodness of fit or one employs more sophisticated techniques for testing non-nested hypotheses, see Pesaran (1985).

see Mizon and Hendry (1980),

As an example of the sequential approach to specification testing, consider the problem of determining from the data the order of the lag polynomials in the model with k explanatory variables z_t of the form

$$\beta(L)y_t = \alpha'(L)z_t + \frac{\mu(L)}{\rho(L)} \epsilon_t \quad (22)$$

where ϵ_t is a white noise process and L is a lag operator with $Ly_t = y_{t-1}$, etc. This model may be written in simpler form as

$$\phi'(L)w_t = \rho(L)\gamma'(L)w_t = \mu(L)\epsilon_t, \quad (23)$$

where $\phi(L)$ is a $k + 1$ element vector of lag polynomials.

To simplify the problem we may suppose that the order of the lag polynomial $\mu(L)$ is known a priori and we wish to use the data to determine the order of the polynomials generating the *system dynamics* $\gamma(L)$ and the *error dynamics* $\rho(L)$. Suppose we can establish a maximum value for the order of the polynomials in $\phi(L)$, \bar{m} say, giving the maintained hypothesis $\phi(L) = \phi_0 + \phi_1 L + \dots + \phi_{\bar{m}} L^{\bar{m}}$, where $\phi_0, \dots, \phi_{\bar{m}}$ denote coefficient vectors of $k+1$ elements. Then hypotheses about the order of $\rho(L)$ and $\gamma(L)$ are nested within this maintained hypothesis but they are not uniquely ordered. However, a unique ordering may be achieved by *imposing* a value for the order of the system dynamics and then the overall problem reduces to determining the order of the autoregression, i.e. the order of $\rho(L)$. Alternatively, the order of the maximal lag, $m \leq \bar{m}$, may be estimated and conditional on this value we may test for the factorisation $\phi(L) = \rho(L) \gamma(L)$ where $\rho(L)$ is of order r and $\gamma(L)$ is of order $m - r$. The first stage is achieved through a sequence of tests for the sequential hypotheses

$$\begin{aligned} H^{(1)}: \quad \phi_{\bar{m}} &= 0 \\ H^{(2)}: \quad \phi_{\bar{m}} &= \phi_{\bar{m}-1} = 0 \\ &\vdots \\ H^{(\bar{m})}: \quad \phi_{\bar{m}} &= \phi_{\bar{m}-1} = \dots = \phi_1 = 0, \end{aligned}$$

and it is interesting to note that the ordering of the hypotheses is unique. At the second stage the hypotheses again form a uniquely ordered nest and so an optimal sequential testing procedure may be followed. This part of the problem consists of testing the sequence of hypotheses

$$\rho_r(L) \gamma_{m-r}(L) = \phi_m(L) \quad r = 0, 1, 2, \dots, m.$$

For this problem the COMFAC algorithm developed by Sargan and Sylvestrowicz (1976) may be used which tests how many common factors there are with m fixed at the first stage. An alternative approach first determines the number of common factors in $\phi(L)$ that are consistent with the data at a chosen significance level

and secondly tests for zero roots among the set of r roots extracted. At the first stage COMFAC may again be used based on unrestricted estimation under the maintained hypothesis. The test for common factors is based upon formulating the vector of restrictions as $f(\phi) = 0$. Given the unrestricted vector estimate $\hat{\phi}$ and estimated covariance matrix, \hat{V} , the Wald statistic $T f'(\hat{\phi}) \hat{\Omega}^{-1} f(\hat{\phi})$ where $\hat{\Omega} = \left(\frac{\partial f}{\partial \phi} \right)' \hat{V} \frac{\partial f}{\partial \phi}$ evaluated at $\hat{\phi}$, is asymptotically distributed as χ^2 with rk degrees of freedom when there are r common factors, i.e. when $f(\phi) = 0$ is valid. This Wald test is preferred to a LR test on computational grounds particularly since $f(\phi) = 0$ can be expressed in the form of a matrix and the resulting test involves a determinantal condition which forms part of the COMFAC algorithm developed for use with the package GIVE.

It is clear that the success of the specification testing procedure examined here largely depends upon making an appropriate initial choice for the general maintained hypothesis. It follows that the appropriateness of the choice that is made should be checked through the use of misspecification tests. These are the subject of the next section.

7. Misspecification Testing

The essential purpose of the model is to represent the process by which the data are generated. This process, the DGP, is inevitably immensely complicated and we may hope that, at best, the model captures its main features. In constructing the model we attempt to systematically account for as much variation in the data as possible, with due regard for the principle of parsimony, and variation that is not accounted for is typically attributed to random factors. Thus the specified model incorporates random disturbances and if the model adequately characterises the DGP the behaviour of the residuals from the fitted model will approximate that specified for the random disturbances. It is not surprising, therefore, that many tests for misspecification are based directly upon the residuals from the fitted model and the justification for such tests is often heuristic rather than deriving from formal testing principles. Frequently, however, a test procedure whose high power can be justified heuristically can be shown to coincide with a test based upon the likelihood principle, e.g. an LM test.

Misspecification tests which check the adequacy of the statistical model are often referred to as *diagnostic checks* and it is particularly appropriate to use the term check rather than test when the intensity of testing is carried to the point where the Neyman Pearson framework is lost. However, the use of the term diagnostic should not be taken to imply that the associated check is necessarily capable of diagnosing a particular misspecification since misspecification tests have, in most cases, a very limited capacity for doing this, the basic difficulty being that a significant test result may be caused by a variety of possible misspecifications. We shall consider this point again later when joint testing procedures are examined.

Pagan (1984) considers approaches to testing an econometric model for misspecification. He highlights the approach which

is characterised by the addition of selected variables to the model under scrutiny which results in overfitting the model in several directions. The tests then have the appearance of significance tests of the extra parameters and they may be based on one of the test principles LR, W or LM. A related approach is to introduce variable transformations in such a way that their impact on the model can be predicted if the model represents an adequate specification. Econometricians have so far been primarily concerned with the first of these approaches, although the relevance of the second is increasingly being recognised, see Plosser et al (1982) for some recent work but see also Mizon and Richard (1982), and Mizon (1984) for their work based on the so-called *encompassing principle* for comparing models.

Pagan and Hall (1983) recast diagnostic tests in terms of *residual analysis* by first considering the testing problem that would arise if the stochastic disturbances were exactly known and then approximating the resulting procedure using the estimated residuals. Engle (1983) argues that proceeding in this way without reliance on the likelihood principle, limits the range of tests that may be developed and it may lead to the development of suboptimal tests. His argument turns on the fact that in the likelihood approach, the first derivative of the log likelihood function with respect to the parameters under test, evaluated under the null, is a sufficient statistic for testing the null against a local alternative and, as a result, the score test will be locally optimal. However, there is no guarantee of local optimality if the likelihood approach is not followed and this is particularly true the more complex the model specified.

We shall examine examples of tests derived by the approaches discussed in the context of the model specified below. Our coverage which will be far from exhaustive, is intended merely to capture the essentials of approaches used to test for major types of misspecification.

We shall consider problems of testing for misspecification in the linear model

$$y = X\beta + \epsilon \quad (24)$$

where y is a $(T \times 1)$ vector of observations on a dependent variable $\{y_t\}$, X is a $T \times k$ matrix of observations on a set of k regressors $\{x_t\}$ and ϵ is a $(T \times 1)$ vector of independent disturbances $\{\epsilon_t\}$, assumed to be normally and identically distributed random variables with $E(\epsilon_t | x_t) = 0$ and finite constant variance σ^2 . X will often be assumed to be non-stochastic but this is mainly for simplicity of exposition. In fact, X may include lagged values of the dependent variable and when (24) is part of a simultaneous equation system it may include endogenous variables as well. Of course, in this last case the assumption $E(\epsilon_t | x_t) = 0$ does not hold.

Following Pagan and Hall (1983), we may note that four important assumptions are made in the specification of the model above,

- (i) $E(\epsilon_t | x_t) = 0$. This reflects the belief that the conditional mean of the relationship has been correctly specified and it covers both the correct functional form, the correct selection of regressors and correct assumptions on the joint-dependence of regressors and regressand.
- (ii) Constancy of parameters. This covers β and σ^2 which are assumed to be fixed over the sample period.
- (iii) Serial independence in the disturbance ϵ_t .
- (iv) A distributional assumption of normality for ϵ_t .

We shall examine tests for departures from the above four assumptions and the associated misspecifications will be referred to as (i) *Specification Error*, (ii) *Non-constancy of coefficients* and *heteroscedasticity* of disturbances, (iii) *Serial dependence* in the disturbances, and (iv) *Non-normality* of disturbances.

(i) Specification Error

Suppose that (24) is a potentially misspecified version of

$$y = X\beta + Z\gamma + u \quad (25)$$

where Z is a $T \times p$ non-stochastic matrix of observations on the $1 \times p$ vector $\{z_t\}$ with $u \sim N(0, \sigma^2 I_T)$ and $E(u_t | x_t, z_t) = 0$.

Assume for the moment that Z is known to the investigator. This is clearly unrealistic since if it were true Z would be included in the specification and its coefficients could be tested directly using a conventional significance test. However, making this assumption provides a starting point for developing tests in realistic cases. Comparing (24) and (25) leads to the equation

$$\epsilon = Z\gamma + u.$$

A test based on residual analysis is obtained by replacing ϵ with its least squares estimate $\hat{\epsilon} = (I - X(X'X)^{-1}X')y = M_X \epsilon$ so that

$$\hat{\epsilon} = M_X Z\gamma + M_X u. \quad (26)$$

The auxiliary regression of $\hat{\epsilon}$ on $M_X Z$ then provides an exact F test of $H_0: \gamma = 0$ under the assumption that X is non stochastic. This test is exactly the same as the significance test of γ in (25). It can be expressed as

$$F = \frac{T - k - p}{p} \frac{RSS_0 - RSS_1}{RSS_1} \sim F(p, T-k-p) \text{ under } H_0 \quad (27)$$

where RSS_0 denotes the residual sum of squares of the regression (24) and RSS_1 denotes the same for the regression (25).

The LM test statistic is just TR^2 from the regression in (26) which corresponds to

$$LM = T \frac{RSS_0 - RSS_1}{RSS_0} \quad (28)$$

and, further, the Wald and Likelihood Ratio test statistics are, respectively,

$$W = T \frac{RSS_0 - RSS_1}{RSS_1} \quad (29)$$

and

$$LR = T \log \left(\frac{RSS_0}{RSS_1} \right). \quad (30)$$

We thus have four separate tests of the same hypothesis. They are not the same although they are all based on essentially the same statistics. The relationship between them is easily seen to be

$$LR = T \log \left(1 + \frac{W}{T} \right), \quad LM = W / \left(1 + \frac{W}{T} \right), \quad F = \left(\frac{T - k - p}{Tp} \right) W.$$

The LM, W and LR statistics are monotonic functions of the F statistic so that exact tests for each would produce the same critical regions. However, if the asymptotic distribution is used to determine the critical values, the tests will differ in finite samples and their conclusions may be in conflict. Evans and Savin (1982) examine the probability of such occurrences.

The more interesting case arises when Z is unknown and a proxy variable \tilde{Z} is used for the auxiliary hypothesis instead. The regression in (26) is then replaced with

$$\hat{\epsilon} = M_X \tilde{Z} \gamma + M_X u + M_X (Z - \tilde{Z}) \gamma. \quad (31)$$

The regression of $\hat{\epsilon}$ on $M_X \tilde{Z}$ will yield an F test of $H_0: \gamma = 0$ which may have the correct asymptotic size and an asymptotic power of unity under non-local alternatives. The sufficient conditions for this result are that $\text{plim } T^{-1} \tilde{Z}' M_X \tilde{Z} \neq 0$ and $\text{plim } T^{-1} \tilde{Z}' M_X Z \neq 0$; hence \tilde{Z} has to be informative with respect to Z. The F, LM, W and LR tests will have the same form as mentioned above with \tilde{Z} replacing Z.

The RESET test proposed by Ramsey (1969) yields a particular form for \tilde{Z} . Ramsey's assumption was that the effect of omitted variables or

of incorrect functional form in (24) could be expressed as an analytic function of $X\beta$, along similar lines to the interaction effect modelled in a two factor design model by Graybill (1975, p. 596). Ramsey proposed as proxy variables $\hat{y}^2, \hat{y}^3, \dots, \hat{y}^l$ where \hat{y}^i is the vector of i^{th} powers of the terms in the regression function $\hat{y} = X\hat{\beta}$. This test does not require the specification of a precise alternative and its attraction derives from the fact that it has reasonably good power properties against a wide range of alternatives though it will have optimal power only when Ramsey's assumption is approximately correct.

Other specification error tests which employ a particular \tilde{Z} include a test for neglected simultaneity, see Hausman (1978) where \tilde{Z} is formed from the residuals of reduced form regressions.

Note how the approach based on the proxy variable \tilde{Z} has the characteristics we ascribed to misspecification testing in section 3. The tests do not involve the parameters of the maintained hypothesis in (24) and the extended parameterisation which replaces Z with \tilde{Z} in (25) is not considered to be an hypothesised representation of the DGP. Thus the test involves an auxiliary hypothesis not a constructive one.

(ii) Non-Constancy of Coefficients and Heteroscedasticity

Sometimes there are grounds for believing that the relationship has changed at some specified point in the data set and a test for the stability of the regression coefficients is required. If time series data are involved and the change is believed to occur after T_1 observations, we should wish to test $\beta_1 = \beta_2 = \beta$, i.e. $\Delta\beta = 0$, in the equation

$$y = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta \\ \beta + \Delta\beta \end{pmatrix} + u$$

where X_i is $T_i \times k$ with $T_i > k$, $i = 1, 2$, and where $T_1 + T_2 = T$, and $\Delta\beta = \beta_2 - \beta_1$.

This equation may be written

$$y = X\beta + Z\Delta\beta + u \quad (32)$$

where

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} 0 \\ X_2 \end{pmatrix}.$$

A test of the significance of $\Delta\beta$ in this equation based on the conventional F ratio yields the familiar Analysis of Covariance (AOC) test which is also a likelihood ratio test.

Comparing (24) and (32) we may write

$$\epsilon = Z\Delta\beta + u.$$

To derive a test based on residual analysis, we replace ϵ with $\hat{\epsilon} = M_X \epsilon$ to yield

$$\hat{\epsilon} = M_X Z \Delta\beta + M_X u. \quad (33)$$

The exact F test of $\Delta\beta = 0$ from the regression of $\hat{\epsilon}$ on $M_X Z$ yields the same AOC test which may be written as

$$AOC = \frac{T - 2k}{k} \frac{\hat{\epsilon}' \hat{\epsilon} - (\hat{\epsilon}_1' \hat{\epsilon}_1 + \hat{\epsilon}_2' \hat{\epsilon}_2)}{\hat{\epsilon}_1' \hat{\epsilon}_1 + \hat{\epsilon}_2' \hat{\epsilon}_2}$$

where $\hat{\epsilon}_i' \hat{\epsilon}_i$ is the residual sum of squares based on the corresponding T_i observations in the regression (32).

More complicated types of changes in the coefficients could be tested for. Often the change takes place gradually or one is unsure of the specific time point when the change occurred. Then the AOC test involves a proxy variable $\tilde{Z}' = (0: X_2')$ and is clearly a misspecification test based on an auxiliary hypothesis. Sometimes the coefficients will be postulated to behave like random variables and various possible models are considered in Pagan (1980) and in Ullah and Rao (1981).

If it is believed that coefficients have changed due to some shock in the economic environment, it may be unreasonable to assume that the variance has remained constant. A likelihood ratio test of $\sigma_1^2 = \sigma_2^2 = \sigma^2$ may then be based upon the variance ratio statistic

$$VR = \frac{\hat{\epsilon}_2' \hat{\epsilon}_2 / (T_2 - k)}{\hat{\epsilon}_1' \hat{\epsilon}_1 / (T_1 - k)} \quad \text{whose distribution is not affected}$$

by the suspected change in the coefficients. The VR test followed by the AOC test yields a uniformly most powerful invariant test of $H_0: \sigma_1^2 = \sigma_2^2, \beta_1 = \beta_2$ against H_A : at least one of $\sigma_1^2 \neq \sigma_2^2, \beta_1 \neq \beta_2$, see Anderson and Mizon (1983), and the successive tests are independent under H_0 , see Phillips and McCabe (1983).

Harrison and McCabe's (1979) test for heteroscedasticity was shown by Breusch and Pagan (1979, p.1293) to be, essentially, the LM test of $\sigma_1^2 = \sigma_2^2$ given by

$$\sqrt{LM} = \left\{ \frac{T}{2T_1T_2} \right\}^{\frac{1}{2}} \sum_{t=1}^{T_1} \left(\frac{\hat{\epsilon}_t^2}{\hat{\sigma}^2} - T_1 \right) \quad \text{where} \quad \hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{\epsilon}_t^2.$$

The LM test of $\sigma_t^2 = h(z_t'\alpha)$, where z_t' is a $1 \times (p+1)$ vector of variables with first term unity and h is of known form with first and second derivatives, is also derived by Breusch and Pagan. Under $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$ the model reduces to the classical normal regression model. Most types of heteroscedasticity can be incorporated as special cases of this model for generating variances. Let $\theta = (\alpha_1, \alpha_2, \dots, \alpha_p)$ and put $\frac{\partial h}{\partial \theta}|_{\theta=0} = c$ where c is a scalar and Z is a matrix with typical row z_t' , then the score is given by $cf'Z/\hat{\sigma}^2$ where f is a vector with typical element $f_t = \hat{\epsilon}_t^2/\hat{\sigma}^2 - 1$ and $\hat{\epsilon}_t$ and $\hat{\sigma}^2$ are obtained under the null. The LM test is given by $LM = \frac{1}{2} f'Z(Z'Z)^{-1}Z'f$ which is just half the explained sum of squares of a regression of f on Z . The test is the same regardless of the form of h because both the score and the information matrix include only the derivative of h under H_0 and so the overall shape of h doesn't matter. The fact that one test is locally optimal for all h is both a strength and a weakness. It is a strength because in practice it would be difficult to define h in a precise way. We might hypothesise that the vectors z_t are important in generating the disturbance variances in some unknown way so that having a test with good local power properties regardless of the form of h seems valuable. On the other hand when h can be chosen precisely it seems that the LR and Wald tests would be more powerful.

Tests can again be based upon residual analysis without reliance upon the LM principle. In the case where h is linear, the heteroscedasticity hypothesis reduces to $\sigma_t^2 = \sigma^2 + \bar{z}_t'\theta$ where $\sigma_t^2 = E(\epsilon_t^2)$. Since disturbances are not available it is natural to put $\epsilon_t^2 = \sigma^2 + \bar{z}_t'\theta + (\epsilon_t^2 - \sigma_t^2)$ which is a regression equation in which the disturbance has a zero mean. Replacing ϵ_t^2 with $\hat{\epsilon}_t^2$, the regression becomes $\hat{\epsilon}_t^2 = \sigma^2 + \bar{z}_t'\theta + (\hat{\epsilon}_t^2 - \sigma_t^2)$ and a test of $\theta = 0$ in this regression will be asymptotically equivalent to the LM test. This is the test given in Koenker (1981) which is robust with respect to normality of the disturbances.

(iii) Serial Dependence of the Disturbance

The most common form of dependence that is tested for is represented by the AR(1) process $\epsilon_t = \rho \epsilon_{t-1} + u_t$ where u_t is white noise. Numerous tests have been proposed and in regression models a test is often based on the Durbin-Watson d statistic. Apart from the omission of a factor $\frac{T}{T-1}$ this is identical to the von Neumann Ratio based upon least squares residuals. The exact distribution of d can be found numerically and exact significance points can be determined. However, the exact test is not commonly used and the associated bounds test is used much more often. Exact tests based on the von Neumann Ratio constructed from BLUS and Recursive residuals can also be used but these tests are less powerful than the exact Durbin Watson test.

Often we may wish to test for higher order dependence and a more general autoregressive process might be tested such as the AR(p) process, $\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \dots + \rho_p \epsilon_{t-p} + u_t$. For these cases there is no generalisation of the Durbin Watson test and so we seek an alternative procedure. The obvious test based on residual analysis is to rewrite the AR(p) as

$$\hat{\epsilon}_t = \rho_1 \hat{\epsilon}_{t-1} + \rho_2 \hat{\epsilon}_{t-2} + \dots + \rho_p \hat{\epsilon}_{t-p} + u_t^+, \quad t = p+1, \dots, T, \quad (34)$$

and then the significance of the $\hat{\epsilon}_{t-i}$, $i = 1, \dots, p$ may be tested in an auxiliary regression. The LM test in this case is shown by Godfrey (1978a) to be based upon TR^2 from the regression

$$\hat{\epsilon}_t = x_t' \beta + \rho_1 \hat{\epsilon}_{t-1} + \rho_2 \hat{\epsilon}_{t-2} + \dots + \rho_p \hat{\epsilon}_{t-p} + u_t^+, \quad t = p+1, \dots, T, \quad (35)$$

which is asymptotically distributed as $\chi^2_{(p)}$ under $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$.

This test is, essentially, the same as a test based upon TR^2 from the auxiliary regression in (34) and so we might as well use this approach. If a test is required for moving average disturbances we consider the MA(p) process

$\epsilon_t = u_t - \rho_1 u_{t-1} - \dots - \rho_p u_{t-p}$ where u_t is white noise.

The score statistic in this case is identical with that in the AR(p) case and the tests are exactly the same for both alternatives. This indicates the dangers of concluding that a specific alternative has generated the data. Godfrey (1978b) shows that if a model is fitted with ARMA (p,q) disturbances, the LM tests against ARMA (p + r, q) and ARMA (p, q + r) are also identical.

Similar results can be obtained in the residual analysis case as shown by Pagan and Hall. For example, in the MA(1) case,

$\epsilon_t = u_t + \rho u_{t-1}$ may be written as $\epsilon_t = \rho \epsilon_{t-1} + u_t + \rho(u_{t-1} - \epsilon_{t-1})$.

In residual form this becomes $\hat{\epsilon}_t = \rho \hat{\epsilon}_{t-1} + u_t^+$ and a suitable test of $H_0: \rho = 0$ may be based upon the t statistic from a regression of $\hat{\epsilon}_t$ on $\hat{\epsilon}_{t-1}$. Thus the test against AR(1) and MA(1) disturbances will be the same.

We see that both the LM and the residual analysis approach lead to a test, in an auxiliary regression, of the significance of the coefficients of \tilde{Z} where \tilde{Z} differs from Z . Having obtained this auxiliary regression we can now decide on which method to use. The $T.R^2$ in (35) version corresponds to the LM significance test in $y = X\beta + \tilde{Z}\gamma + \tilde{u}$. If we use the W, LR or F version the test retains its LM origin because this gave rise to \tilde{Z} . Although the F form of the test will not provide an exact test, Monte Carlo results indicate that it has to be preferred in small samples, see Kiviet (1982).

(iv) Non-Normality of Disturbances.

While small sample inference and prediction proceeds under the assumption of normality for disturbances, tests for non-normality are not often used. There is only a relatively small literature in this area of misspecification testing in econometrics and most text books have little or nothing to say on the subject.

Appropriate tests are available, however, and the main approach builds on the fact that normal disturbances have the property that the third moment, μ_3 , is zero and the fourth moment, μ_4 , is three times the square of the second moment, μ_2 .

Pagan and Hall note the following identities which provide the basis for tests derived from residual analysis:

$$\begin{aligned} \text{(a)} \quad \hat{\epsilon}_t^2 &= \mu_2 + (\epsilon_t^2 - \mu_2) + (\hat{\epsilon}_t^2 - \epsilon_t^2) \\ &= \mu_2 + u_{2t} \end{aligned}$$

$$\begin{aligned} \text{(b)} \quad \hat{\epsilon}_t^3 &= \mu_3 + (\epsilon_t^3 - \mu_3) + (\hat{\epsilon}_t^3 - \epsilon_t^3) \\ &= \mu_3 + u_{3t} \end{aligned}$$

$$\begin{aligned} \text{(c)} \quad \hat{\epsilon}_t^4 &= \mu_4 + (\epsilon_t^4 - \mu_4) + (\hat{\epsilon}_t^4 - \epsilon_t^4) \\ &= \mu_4 + u_{4t} \end{aligned}$$

From (b) and (c), we have

$$\begin{aligned} \text{(d)} \quad \hat{\epsilon}_t^4 - 3\hat{\sigma}^2 \hat{\epsilon}_t^2 &= \mu_4 - 3\sigma^4 + u_{4t} - 3\hat{\sigma}^2 u_{2t} + 3(\sigma^4 - \sigma^2 \mu_2) \\ &= \gamma_4 + u_{5t} \quad \text{where} \quad \gamma_4 = \mu_4 - 3\sigma^4. \end{aligned}$$

The least squares estimates of $\gamma_3 = \mu_3$ and γ_4 from these auxiliary regressions lead to

$$\hat{\gamma}_3 = T^{-1} \sum_{t=1}^T \hat{\epsilon}_t^3, \quad \hat{\gamma}_4 = T^{-1} \sum_{t=1}^T (\hat{\epsilon}_t^4 - 3\hat{\sigma}^2 \hat{\epsilon}_t^2) \text{ which are asymptotically}$$

uncorrelated under the hypothesis of normality.

If the asymptotic variances of $\hat{\gamma}_3$ and $\hat{\gamma}_4$, namely, $6\sigma^6$ and $24\sigma^8$ respectively, are replaced by consistent estimators $6\hat{\sigma}^6$ and $24\hat{\sigma}^8$, the LM test of $H_0: \gamma_3 = \gamma_4 = 0$ is derived as

$$LM = \left(\frac{\hat{\gamma}_3^2}{6\hat{\sigma}^6} + \frac{\hat{\gamma}_4^2}{24\hat{\sigma}^8} \right) = \left\{ \frac{\hat{b}_1}{6} + \frac{\hat{b}_2}{24} \right\} \quad (36)$$

where $\hat{\gamma}_3/\hat{\sigma}^3 = \sqrt{\hat{b}_1}$ and $\hat{\gamma}_4/\hat{\sigma}^4 = \hat{b}_2$ are standard measures for skewness and kurtosis of the regression residuals. This test may also be derived based upon the Information Matrix Test principle which will now be examined.

8. Testing for Several Misspecifications

In practice we shall often wish to test for the presence of several possible misspecifications and we briefly consider approaches to doing this.

Recently White (1982) introduced a general test for misspecification based upon the Information Matrix. Suppose the log likelihood of a random sample is given by $L = \sum \log f(\epsilon_t, \theta)$ where θ is a $p \times 1$ vector of parameters, then if the expectations exist we may define the matrices

$$A(\theta) = \{E(\partial^2 \log f(\epsilon_t, \theta) / \partial \theta_i \partial \theta_j)\},$$

$$B(\theta) = \{E(\partial \log f(\epsilon_t, \theta) / \partial \theta_i \cdot \partial \log f(\epsilon_t, \theta) / \partial \theta_j)\},$$

$$A_T(\theta) = \{T^{-1} \sum_{t=1}^T \partial^2 \log f(\epsilon_t, \theta) / \partial \theta_i \partial \theta_j\},$$

$$B_T(\theta) = \{T^{-1} \sum_{t=1}^T \partial \log f(\epsilon_t, \theta) / \partial \theta_i \cdot \partial \log f(\epsilon_t, \theta) / \partial \theta_j\}.$$

Under the usual regularity conditions and assuming that f is the true probability model, $A(\theta) = -B(\theta)$, and when this equality fails the model is misspecified.

If we let $\hat{\theta}_T$ be the ML estimator of θ , then a test for misspecification may be based upon the $\frac{1}{2}p(p+1)$ distinct components of the $p \times p$ symmetric matrix

$$\sqrt{T} (A_T(\hat{\theta}) + B_T(\hat{\theta})). \quad (37)$$

If interest centres on q of these components which together form the $q \times 1$ vector $D_T(\hat{\theta})$, then White shows that under certain

conditions, including correct specification,

- (i) $\sqrt{T} D_T(\hat{\theta}_T) \xrightarrow{d} N(0, V(\theta_0))$,
- (ii) $V_T(\hat{\theta}_T) \xrightarrow{as} V(\theta_0)$ where $V(\theta_0)$ is almost surely non singular as $T \rightarrow \infty$.

(iii) The Information Matrix Test statistic is

$$I_T = T(D_T(\hat{\theta}_T))' [V_T(\hat{\theta}_T)]^{-1} D_T(\hat{\theta}_T) \xrightarrow{d} \chi^2_{(q)}. \quad (38)$$

In the general linear model it can be shown that the approach yields the non-normality test given in (36) and the test for heteroscedasticity proposed by White (1980) and that these tests are asymptotically independent when misspecification is absent. A weakness of the approach, however, is that none of the terms in (37) form the basis of a test for departures from independence of the disturbances.

Bera and Jarque (1982) proposed a nested hypothesis testing procedure based upon a general model which permits the simultaneous testing of incorrect functional form, heteroscedasticity, serial correlation and non-normality. The actual model studied was

$$y_t = \sum_{j=1}^k x_{tj}^{(\lambda_j)} \beta_j + \sum_{j=1}^m d_{tj} \nu_j + \sum_{j=k+1}^l x_{tj}^* \beta_j^* + \varepsilon_t, \quad t = 1, 2, \dots, T$$

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

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

The $x_{tj}^{(\lambda)}$ represent transformed observations on k fixed regressor variables while the d_{tj} and x_{tj}^* are observations on other fixed regressors.

The disturbances, ϵ_t , follow an autoregressive process

$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \dots + \rho_p \epsilon_{t-p} + u_t$ where the u_t are independently distributed. The density of u_t is assumed to be a member of the Pearson family of distributions so that

$$g(u_t) \propto \exp[\chi(u_t)] / \int_{-\infty}^{\infty} \exp[\chi(u_t)] du_t, \quad -\infty < u_t < \infty, \quad t = 1, 2, \dots, T,$$

and

$$\chi(u_t) = \int [(c_{1t} - u_t)] / (c_{0t} - c_{1t}u_t + c_{2t}u_t^2) du_t.$$

It is assumed that $E(u_t^2) = c_{0t} / (1 - 3c_{2t})$ and $g(u_t) = N(0, c_{0t})$ when $c_{1t} = c_{2t} = 0$. The model is parameterised with $c_{2t} = c_2$ and the possibility of additive heteroscedasticity is introduced by putting $c_{0t} = \sigma^2 + \bar{z}_t' \alpha$ with \bar{z}_t' a $1 \times q$ vector of fixed variables.

Within this framework the following tests may be carried out:

Correct functional form: $H_0(F) : \lambda = i, \quad \beta^* = 0$.

Homoscedasticity : $H_0(H) : \alpha = 0$.

Serial Independence : $H_0(I) : \rho = 0$.

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

where $\lambda = (\lambda_1, \dots, \lambda_k)'$, $i = (1, 1, \dots, 1)'$, $\beta^* = (\beta_{k+1}^*, \dots, \beta_\ell^*)'$,

$\alpha = (\alpha_1, \dots, \alpha_q)'$, $\rho = (\rho_1, \dots, \rho_p)'$.

Bera and Jarque show that if LM_F , LM_H , LM_I and LM_N are the individual misspecification tests and LM_{FHIN} is the LM test for testing all the misspecifications simultaneously, then here $LM_{FHIN} = LM_F + LM_H + LM_I + LM_N$.

For large samples the individual LM tests are independent when there are no misspecifications. Here we have an example of additive tests of juxtaposed alternatives. Hence any combination of tests can be carried out and overall significance levels can be controlled at least asymptotically. A similar property will apply to the LR and W tests in this framework although these tests will require unrestricted estimation of the model and this is computationally most burdensome.

Godfrey and Wickens (1982) also suggest a nested hypothesis test approach. Noting that when alternatives involve complex models it is often easier to compute an LM rather than a LR or W test, they show that frequently further simplification is possible using a locally equivalent alternative (LEA) model which yields statistics having the same asymptotic distribution as the LM test based on the original alternative model c.f. Godfrey (1981).

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

$$(i) \quad f_t(y_t | x_t, \theta) = \epsilon_t,$$

$$(ii) \quad f_t^*(y_t | x_t, \theta) = \epsilon_t,$$

where (i) and (ii) are different except when $\theta_2 = 0$ with $\theta' = (\theta_1' : \theta_2')$.

Let $L_T(\theta)$ and $L_T^*(\theta)$ be, respectively, the loglikelihoods for (i) and

$$(ii) \text{ where } L_T(\theta) = \sum_{t=1}^T \ell_t(\theta) \text{ and } L_T^*(\theta) = \sum_{t=1}^T \ell_t^*(\theta) \text{ and suppose that } \hat{\theta}$$

maximises $L_T(\theta)$ subject to $H_0: \theta_2 = 0$. Then if model (ii) is such that

$$(a) \quad \hat{\theta} \text{ maximises } L_T^*(\theta) \text{ subject to } H_0: \theta_2 = 0, \text{ and } (b) \quad \partial \ell_t^*(\hat{\theta}) / \partial \theta_2 =$$

$\partial \ell_t(\hat{\theta}) / \partial \theta_2$, then the LM test of $H_0: \theta_2 = 0$ will be the same in both models.

One way of deriving an appropriate approximation to model (ii) is by using

$$(iii) \quad f_t^+(y_t | x_t, \theta) = f_t(y_t | x_t, (\hat{\theta}_1' : 0')) + [\partial f_t(\hat{\theta}_1' : 0') / \partial \theta_2] \theta_2$$

whereupon an asymptotically equivalent procedure to the LM test 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

$\partial f_t(\hat{\theta}) / \partial \theta_2$ and which enter f_t^+ linearly.

As an example, suppose that $y_t = x_t' \beta + \epsilon_t$ is the model where x_t' is a $1 \times k$ vector of exogenous variables, and the alternative hypothesis is that $\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \rho_3 \epsilon_{t-3} + \dots + \rho_p \epsilon_{t-p} + u_t$ and $E(u_t)^2 = \sigma_t^2 = \sigma^2(1 + \bar{z}_t' \alpha)$ where \bar{z}_t' is a $1 \times q$ vector of fixed variables, then a locally equivalent alternative model is given by

$$y_t = x_t' \beta + \sum_{i=1}^p \rho_i \hat{\epsilon}_{t-i} + \sum_{i=1}^q \alpha_i \hat{\epsilon}_t \bar{z}_{ti} + u_t.$$
 The problem of testing for serial correlation and heteroscedasticity then reduces to testing the significance of regression coefficients in this regression. Under $H_0: \alpha = \rho = 0$, the regressors in the above regression are asymptotically uncorrelated and the test statistics used in testing 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 a formal F test and a number of possible misspecifications can be tested in a single step. However, some writers have expressed reservations about the power of tests based on LEA's, see Schonfeld (1982), particularly under non-local alternatives.

Pagan and Hall examined the conditions under which the tests based on residual analysis are asymptotically independent when the overall null hypothesis of no misspecification is true. For the misspecification tests considered under (i) - (iv) ^{in Section 7,} the situation was concisely summarised as follows:

Conditions Required for Asymptotic Independence of Diagnostic Test Statistics

	S	A	H	N
Specification Error (S)	*			
Serial Dependence (A)	C_1	*		
Heteroscedasticity (H)	$\mu_3 = 0$	$\mu_3 = 0$ or C_1	*	
Non-Normality (N)	None	None	None	*

C_1 : No y_{t-k} $k \leq j$ (where j is the order of the serial correlation) appears in z_t of S or in x_t .

Pairwise asymptotic independence holds between all the tests when lagged dependent variables are absent from the model and the disturbances are normal. However, Pagan and Hall did not examine how these results might be used in a structured approach to misspecification testing.

Finally, Phillips and McCabe (1984) examined a sequential approach to testing for serial dependence, non-constancy of coefficients and heteroscedasticity. They showed that it is possible to order the tests so that (a) the test statistics are exact and mutually independent under the overall null hypothesis, and (b) each test statistic in the sequence depends upon the validity of less restricted hypotheses but not on more restricted hypotheses, i.e. those yet to be tested. However, because the ordering of hypotheses is not unique the testing procedure, which stops when a significant test result is obtained, leaves untested the more restricted hypotheses. This is an example of where inherently juxtaposed hypotheses are combined within a comprehensive model to form a sequence of superposed hypotheses, as discussed at the end of section 5.

9. A Testing Strategy for Model Specification

In considering approaches to specification and misspecification testing starting from a general maintained model, our purpose has been to develop elements of an overall strategy for testing econometric models while maintaining the Neyman Pearson framework at least approximately or asymptotically. By the term strategy we mean a decision making procedure within which each successive decision is made dependent on the information available at the time. Thus a testing strategy is a strategy wherein the use or otherwise of each new test depends upon the outcome of its predecessor.

We have, in fact, identified two sub-strategies which together, essentially, form a testing strategy for model specification; namely the misspecification and specification testing sub-strategies.

When we start from a general maintained hypothesis, we first consider the question of whether the model is general enough to adequately represent the DGP; thus the natural starting point is with the misspecification testing sub-strategy. If no misspecifications are revealed then the specification testing sub-strategy is followed. Ideally, we should like to conduct a sequence of independent tests within each sub-strategy while preserving the Neyman Pearson framework in such a way that the test size is controlled and the procedure has a high test power. Within each sub-strategy it is unlikely that this ideal can be realised except in relatively simple cases but it would seem important to approximate it as closely as possible. We have seen that a combination of tests for common misspecifications can be carried out which are asymptotically independent under the common null, that tests are available that have optimal local power properties and that the LM versions are usually very easily calculated in auxiliary regressions. These tests clearly could form the basis of a

misspecification testing sub-strategy particularly if modifications to the tests are employed so as to bring the individual test sizes closer to the nominal level (see Kiviet 1982). The subsequent specification testing sequence might well be independent of the misspecification tests at least asymptotically as the misspecification stage is superposed with respect to the specification stage. We have already noted the asymptotic sequential independence of the LR, W and LM tests both within and across each class of test, but the argument is strengthened by noting that in common situations, i.e. in standard regression models, the general model may produce a set of complete sufficient statistics from which specification testing proceeds. If the misspecification test statistics are invariant with respect to the parameters involved in the specification tests, then they will be exactly independent of the specification tests which are functions of the complete sufficient statistics for these parameters. This follows from Basu's Independence Theorem (1955) noted earlier. Consequently, if the test size within each sub-strategy can be approximately controlled, then the overall significance level can be controlled also.

While an overall strategy can be devised along these lines there will be difficulties of implementation. If the general maintained model is rejected by the misspecification testing sub-strategy then the model must be revised and this will lead to further misspecification testing of the revised model. This may happen several times and after each revision stage the interpretation of the test outcomes becomes increasingly more difficult and the associated probability statements more invalid. Indeed, the point is soon reached at which the degree of data mining reduces the test results to mere descriptive measures of the characteristics of the data. Another problem arises because it is often felt desirable to subject to further misspecification testing a model that resulted from earlier application of the two sub-strategies. Of course these new tests will not be independent of the misspecification tests of the same type performed in earlier stages.

In fact, under the overall null they are completely dependent asymptotically and so will not force up the overall significance level. However, the sample is only finite in practice and, although at each testing stage the power of the individual tests might well have improved, the overall probability of a type I error will have increased.

A possible way out of the problems of data mining was suggested by Theil (1978 p. 273) who argued that when data are plentiful, a sensible approach is to split them into three parts. One set may then be used to specify the relation, a second set should be used to estimate the parameters and a third for predictions based on the estimated equation to verify whether the specification is acceptable. Adopting this approach in the present context, the first data set might be used to derive the maintained hypothesis, including the misspecification checks, while the second set might be used for specification testing. The use of a third (small) data set for a predictive test of the selected model has been widely advocated, see Dhrymes et al (1972), Harvey (1981) and Hendry (1979), and this seems particularly necessary when a model has been arrived at through considerable experimentation. Unfortunately with the data that is typically available it is not often possible to follow the procedure advocated by Theil and some data mining is almost inevitable. Of course, even if Theil's suggestion is followed and the post sample predictive test suggests that the model is inadequate it is difficult to see how best to proceed. Presumably, the model must be modified and the testing process repeated until a specification is found which passes the tests. While this may lead to a good model based on all reasonable criteria, one is still open to the charge of data mining.

REFERENCES

- Anderson, G.J. and Mizon, G.E. (1983). "Parameter Constancy Tests: Old and New." Discussion paper No. 8325. University of Southampton.
- Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. New York. Wiley.
- Basu, D. (1955). "On Statistics Independent of a Complete Sufficient Statistic." *Sankhya* 15, 377-380.
- Bera, A.K. and Jarque, C.M. (1982). "Model Specification Tests: a simultaneous approach." *Journal of Econometrics*, 20, 59-82.
- Breusch, T.S. and Pagan, A.R. (1979). "A Simple test for heteroscedasticity and random coefficient variation." *Econometrica*, 47, 1287-1294.
- Dhrymes, P.J. et al (1972). "Criteria for Evaluation of Econometric Models." *Annals of Economic and Social Measurement* 1/3, 291-324.
- Engle, R.F. (1983). Contribution to discussion in Pagan, A.R. and Hall, A.D. (1983) below.
- Engle, R.F. (1984). "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics." *Handbook in Econometrics*. Z. Griliches and M.D. Intriligator (eds.), Amsterdam, North Holland.
- Engle, R.F. Hendry, D.F. and Richard, J.R. (1983). "Exogeneity" *Econometrica* 51, 277-304.
- Evans, G.B.A. and Savin, N.E. (1982). "Conflict Among the Criteria Revisited: The W, LR and LM Tests." *Econometrica* 50, 737-748.
- Godfrey, L.G. (1978 a). "Testing Against General Autoregressive and Moving Average Error Models when the Regressors include Lagged Dependent Variables." *Econometrica* 46, 1293-1301.
- Godfrey, L.G. (1978 b). "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables." *Econometrica* 46, 1303-1310.
- Godfrey, L.G. (1981). "On the invariance of the Lagrange Multiplier test with respect to certain changes in the alternative hypothesis." *Econometrica* 49, 1443-1455.
- Godfrey, L.G. and Wickens, M.R. (1982). "Tests of misspecification using locally equivalent alternative models." In: *Evaluating the Reliability of Macroeconomic Models*. (G. Chow and P. Corsi Eds.). London. Wiley.
- Graybill, F. (1975). *Theory and Application of the Linear Model*. Duxbury Press.
- Harrison, M. and McCabe, B.P.M. (1979). "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals." *Journal of the American Statistical Association*, 74, 494-499.
- Harvey, A.C. (1981). *The Econometric Analysis of Time Series*. Phillip Allan.

- Hausman, J.A. (1978). "Specification Tests in Econometrics." Econometrica 46, 1251-1272.
- Hendry, D.F. (1979). "Predictive Failure and Econometric Modelling in Macroeconomics. The Transactions Demand for Money," in: Modelling the Economy (editor P. Ormerod). Heineman Educational Books.
- Hendry, D.F. (1983). "Econometric Evaluation of Linear Macro-Econometric Models." Paper presented to Conference on Forecasting. London Business School 1983.
- Hendry, D.F. and Anderson, G.J. (1977). "Testing Dynamic Specification in Small Simultaneous Systems: An application to model of building society behaviour in the U.K." in Frontiers of Quantitative Economics 3A (ed. M.D. Intrilligator) 361-383. Amsterdam, North Holland.
- Hendry, D.F. and J.F. Richard (1982). "On the Formulation of Empirical Models in Dynamic Econometrics." Journal of Econometrics 20, 3-33.
- Hogg, R.B. (1961). "On the Resolution of Statistical Hypotheses." Journal of the American Statistical Association 56, 978-989.
- Kiviet, J.F. (1982). "Size, Power and Interdependence of Tests in Sequential Procedures for Modelling Dynamic Relationships." Unpublished paper, University of Amsterdam, presented at ESEM 1982, Dublin.
- Koenker, R. (1981). "A note on Studentising a test for heteroscedasticity." Journal of Econometrics, 17, 107-111.
- Leamer, E.E. (1982). Specification Searches. New York. Wiley.
- McAleer, M and Pesaran, M H (1985). "Statistical Inference in Non-Nested Econometric Models". This issue p.
- Mackinnon, J.G. (1983). "Model Specification Tests Against Non-nested Alternatives." Econometric Reviews, 2, 85-158.
- Mizon, G.E. (1977). "Inferential procedures in non-linear models: An application in a U.K. industrial cross section study of factor substitution and returns to scale." Econometrica 45, 1221-1242.
- Mizon, G.E. (1984). "The Encompassing Approach in Econometrics" in D.F. Hendry and K.F. Wallis eds. Econometrics and Quantitative Economics. Basil Blackwell.
- Mizon, G.E. and Hendry, D.F. (1980). "An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification." Review of Economic Studies XLVII, 21-45.
- Mizon, G.E. and J.F. Richard (1982). "The encompassing principle and its application to testing non-nested hypotheses." Unpublished paper. Southampton University.
- Pagan, A.R. (1980). "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients", Journal of Econometrics, 13, 341-363.
- Pagan, A.R. (1984). "Model Evaluation by Variable Addition" in D.F. Hendry and K.F. Wallis, eds. Econometrics and Quantitative Economics. Basil Blackwell.

- Pagan, A.R. and Hall, A.D. (1983). "Diagnostic Tests on Residual Analysis." Econometric Reviews, 2 (2). 159-218.
- 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 to ESEM 1984, Madrid.
- Plosser, C.I., Schwert, G.W. and White, H. (1982). "Differencing as a test of specification." International Economic Review, 23, 535-552.
- 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.
- Richard, J.F. (1980). "Models with Several Regimes and Changes in Exogeneity." Review of Economic Studies. 1-20.
- Rothenberg, T.J. (1982). "Comparing Alternative Asymptotically Equivalent Tests," in *Advances in Econometrics*. (editor W. Hildenbrand) Cambridge University Press.
- Sargan, J.D. (1980). "Some Tests of Dynamic Specification for a Single Equation." Econometrica 48, 879-897.
- Sargan, J.D. and Sylvestrowicz, J.D. (1976). "COMFAC: Algorithm for Wald Tests of Common Factors in Lag Polynomials." Users' Manual. London School of Economics.
- Schönfeld, 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.
- Seber, G.A.F. (1966). *The Linear Hypothesis*. London. Griffin.
- Silvey, S.D. (1975). *Statistical Inference*. Chapman and Hall. London.
- Theil, H. (1961). *Economic Forecasts and Policy*. North Holland. Amsterdam.
- Theil, H. (1978). *Introduction to Econometrics*. Prentice Hall Inc. New Jersey.
- (1981)
Ullah, A. and Rao B., / *Econometrics: A Varying Coefficient Approach*. Croom Helm.
- White, H. (1980). "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." Econometrica 48, 817-838.
- White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models." Econometrica 50, 1-26.

