0-315-20326-9

68233

NAME OF AUTHOR/*NOM DE L'AUTEUR* Chin, Chee Fung

TITLE OF THESIS/*TITRE DE LA THÈSE* Some Aspects of Non-nested Model Testing

UNIVERSITY/*UNIVERSITÉ* Simon Fraser University

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* M.A.

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1984

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Peter E. Kennedy

DATED/*DATÉ* May 24, 1984 SIGNED/*SIGNÉ*

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ*

SOME ASPECTS OF NON-NESTED MODEL TESTING.

by

Chin, Chee Fung

B.A. (Hons.) Simon Fraser University, 1981.

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Economics

SIMON FRASER UNIVERSITY

May, 1984

APPROVAL

Name: Chee Fung Chin

Degree: Master of Arts

Title of Thesis: Some Aspects of Non-Nested Model Testing

Examining Committee:

Chairman: Kenji Okuda

Peter E. Kennedy
Senior Supervisor

Terence M. Heaps

Richard D. Routledge
External Examiner
Associate Professor
Department of Mathematics
Simon Fraser University

Date Approved: May 23, 1984

PARTIAL COPYRIGHT LICENSE

Title of Thesis/Project/Extended Essay

Some Aspects of Non-nested Model Testing

Author:

(signature)

Chin, Chee Fung

(name)

May 18, 1984

(date)

ABSTRACT

This thesis examines non-nested hypothesis testing with particular emphasis on the Cox statistic. Intuitive interpretations of the existing non-nested model tests are provided and then five selected topics are examined:

a. An alternative derivation of the Cox statistic is provided for the case in which the competing models fall into the classical linear regression mould.

b. The Cox statistic is shown to have asymptotic properties different from those currently employed.

c. The Atkinson test is shown to be inconsistent in contrast to current belief.

d. The concept of the "direction" of a model, abandoned by the literature, is revived.

e. An alternative method for deriving the asymptotic distribution of the Cox statistic under local alternatives is presented.

TABLE OF CONTENTS

LIST OF TABLES

Introduction and Conclusion.

When testing several models, none of which can be expressed as a general case of another, a major problem which can arise is that there is no meaningful way to nest them together. In this case the usual model-testing method of first forming a general model and then applying the orthodox subset model selection tests is not applicable (Judge et al, 1980). One approach to resolve this difficulty is to employ classical misspecification tests so that any false model can be detected through misspecification problems such as serial correlation or heteroskedasticity. The main defect of applying this method is that the information of the existence of other competing models is not used.

An alternative way of addressing this problem is through the so-called non-nested model tests. These tests are designed to make use of the knowledge of the existence of the competing models to determine the "truth" of a model (as opposed merely to permit a choice between competing models). In the literature, there are several non-nested model (NNM) testing techniques which make use of this information. The most prominent are the Cox test (Cox, 1961,1962) and its variants, Dastoor's R test statistic (Dastoor, 1983), and the test statistics suggested by Davidson and MacKinnon (1981) and their associated variants. In general, these tests are concerned with the ability of a (temporary) null model to predict the behaviour of an estimator formulated on the basis of a given alternative model.[1] Their main advantage is the possibility of rejecting all the competing models. Their main disadvantages are that they may yield inconclusive results and that they are

[1] In contrast, the classical subset model selection technique examines the performance of the null against the alternatives.

large sample tests with unknown small sample properties.

This thesis examines the testing of non-nested models with particular emphasis on the Cox test. There are three main parts. Part I introduces some concepts and definitions that will be relevant to Part II and Part III. Part II consists of an intuitive interpretation of the existing non-nested model tests. No mathematics is involved here since all proofs of the results can either be found in some of the referred articles or in Part III. Nevertheless, an understanding of the "mathematics" in Part I will help. Part III has five sections, each attempting to extend the literature relevant to the Cox test in a particular area. Section 1 provides an alternative derivation of the Cox test statistic (CTS) and the Atkinson test statistic (ATS) when the competing models are classical normal linear regression models. (The same method is also applicable when these models are non-linear.) Section 2 shows that the CTS under the null hypothesis does not have a zero mean, and has an asymptotic variance larger than that usually employed. Subsequently, it is shown that if this non-zero bias is corrected, the resulting standardized test statistic (TS) under certain conditions yields a known non-nested TS, the W- TS. Section 3 shows that the Atkinson test is an inconsistent test, in contrast to current belief. Section 4 applies the concept of the "direction" of a model, defined in Part I, to the Cox test. It is shown that in certain cases the direction of the true model can be inferred from the sign of the CTS. Section 5 provides an alternative method for deriving the asymptotic distribution of the CTS under local alternatives. This section also derives the asymptotic distribution of the ATS under local alternatives, and shows that the asymptotic distributions of the ATS and CTS are equivalent.

Part I.

Preliminaries

This section briefly reviews some concepts in econometrics/statistics that are essential to an understanding of those points developed later in this thesis. In particular, the distinction between separate families of probability density function (pdf) and non-nested models is discussed, the concept of the direction of a model is developed, some comments on likelihood functions are offered, and the notions of information and entropy are developed.

1.1. Separate families of pdf and non-nested models.

A statistical model is a characterization of the probability distribution of a variable in terms of some parameters and, usually, some exogenous variables. For example, if MM is the parameter space of a model, and X is a set (matrix) of exogenous variables, then a statistical model is the probability density function of a variable y given $f(X, \theta)$, where $\theta \in MM$, and $f(X, \theta)$ is a function of X given θ. It is denoted here as: $p(y|f(X, \theta))$. Two statistical models are said to be of <u>separate families</u>, if one pdf cannot be approximated arbitrarily closely by another. That is, if y is the variable, and two statistical models are:

$$p_1(y|f(X_1, \theta_1)), \quad p_2(y|f(X_2, \theta_2))$$

where for $i = 1, 2$, the X_i's are matrices of exogenous variables, and θ_i's are the parameters for each of the two models with MM(i) as the parameter space, then if for an arbitrary value of θ_1, say $\theta_1{}^*$, the pdf $p_1(y|f(X_1,\theta_1{}^*))$

cannot be approximated arbitrarily closely by $p_2(y|f(X_2, \theta_2))$. An example of this is when one distribution is log normal while the other is exponential.

Although the Cox test was originally proposed for the testing of separate families of statistical models, it was later adopted by econometricians to test in the context of non-nested classical normal regression model (CNRM). A CNRM is a class of pdf which has its variable normally distributed and satisfies the classical assumptions. Two CNRM are said to be nested if one is a special case of the other, and they are non-nested, if otherwise. In the econometrics literature the Cox test has been adopted for testing different models which have the same families of pdf's, as opposed to testing separate families of statistical models each of which pertains to different families of pdf's.

This thesis will deal mainly with CNRM. Unless otherwise specified, a model henceforth will mean a CNRM. Since the distinguishing element of one CNRM from another is the vector of the mean value of y, say $f(X_1, b_1)$, it is thus proper to refer to a model as a vector of $f(X_1, b_1)$.

In the case of a linear model, i.e. $f(X_1, b_1) = X_1 b_1$, one model differs from another model when their vector spaces are different. A model is a subset model if its parameter space is contained in another model's parameter subspace.

1.2. Direction of a model/vector.

In this section, an attempt will be made to define the "direction" of a model. This concept of direction has never been expounded in the non-nested model testing literature even though the term had once been used (but was later discarded).[2] The directional concept defined here is a natural extension of the existing geometrical interpretation of regression analysis.

It is well-known that all vectors have lengths and directions. If x is a vector, then its length is simply

$$|x| = (x'x)^{1/2}.$$

Since direction is a relative term, the direction of a vector has to be defined relative to something, say another vector. A natural measure of the direction of one vector relative to another is the cosine of the angle, r, between them. So, if x and z are two vectors, and their "inner product" is

$$x'z = |x|.|z|.\cos(r),$$

then

$$\cos(r) = x'z/\{|x|.|z|\}.$$

This definition of the direction of one vector relative to another is intuitively appealing, because

$$-1 \leq \cos(r) \leq 1,$$

and, when cos (r) = 1, x and z are parallel; when cos (r) = 0, x and z are orthogonal; and when cos (r) = -1, x and z are parallel but in opposite direction. It is to be noted that this definition bears close similarity to Pesaran's (1982, p.129) definition of "distance", as cos (r) is the correlation coefficient of x and z, if these vectors are centered, i.e. if both x and z are orthogonal to the vector of unity.

Often, one may want to know the direction of a vector, say z, relative to a hyperplane in a vector space generated by (the column vectors of) the matrix X. An extension of the above idea to this case is quite simple. The

2 See for eample, Pesaran, 1974, p.158; Fisher and McAleer, 1979, and Dastoor, 1981, p.114.

direction of z relative to this hyperplane can be defined in terms of the cosine of the smallest angle, r, between z and this hyperplane. In other words, it is the cosine of the angle between z and its closest vector in this hyperplane. This closest vector is the orthogonal projection of z onto the hyperplane, say Pz, where $P = X(X'X)^{-1}X'$. It represents the best approximation to z in this hyperplane. Thus, given that the inner product of the two vectors z and Pz is

$$z'Pz = |z|.|Pz|.\cos(r)$$

then $$\cos(r) = |Pz|/|z| = (z'Pz)^{1/2}/(z'z)^{1/2}. \qquad (1)$$

It appears that the following application of the above results to linear models has been overlooked in the literature. Suppose there are two linear models, X_1b_1 and X_2b_2, and if the latter is the true model, then its direction relative to the former is the cosine of the smallest angle between it and its vector projection onto X_1, i.e.

$$\cos(r) = (u'P_1u/u'u)^{1/2},$$

where $u = X_2b_2$, and $P_1 = X_1(X_1'X_1)^{-1}X_1'$. Notice that b_1 does not appear in the formula above because, by assumption, X_1b_1 is not the true model, and hence b_1 does not exist. Since X_2b_2 is the true model, its best approximating model in the vector space generated by X_1 is P_1u. This is the least squares estimate of X_2b_2 in the vector space generated by the columns of X_1.

1.3. Likelihood function.

By definition, the likelihood of a model given data y, with parameter space, MM, and parameter, θ, is a constant multiplier of the probability of y given θ, i.e.

$$L(\theta|y) = c.p(y|\theta),$$

where c is an arbitrary positive constant. This arbitrary constant, c, in the likelihood function "generates an equivalence class of similarly shaped functions" (Fraser, 1976,p.312). This arbitrary constant can be a nuisance under some circumstances, in particular, for non-nested model testing. One way to avoid it is to use "relative likelihood" instead. Relative likelihood is formed by arbitrarily setting one of the likelihood functions to unity first and then define other likelihood functions relative to it. For example, let θ be a parameter, and let θ^* be any particular value of θ, then by setting

$$L(\theta^*|y) = c.p(y|\theta^*) = 1,$$

$$\Rightarrow \qquad c = 1/p(y|\theta^*)$$

we can then get $L(\theta^{**}|y) = c.p(y|\theta^{**})$

$$= p(y|\theta^{**})/p(y|\theta^*)$$

where θ^{**} is any value of θ. Sometimes, it might be convenient to deal with this constant systematically by setting the maximum likelihood (ML) of θ equal to one, instead of the likelihood of any θ (Fraser, Edwards). In this case, the likelihood of the estimate itself is the likelihood ratio (LR) of the estimate to its maximum likelihood estimate (MLE).

The above method of eliminating the nuisance parameters is applicable only to nested models. If non-nested models are involved, these nuisance constants cannot be eliminated by using the above method. To see this, let the parameter spaces of two non-nested models be MM(1) and MM(2), with parameters θ_1 and θ_2 respectively, then

$$LR = L(\theta_1 |y)/L(\theta_2 |y))$$

$$= c_1.p(y|\theta_1)/(c_2.p(y|\theta_2)).$$

where c_1 and c_2 are some arbitrary positive constants. Since $c_1 \neq c_2$, this LR is not the likelihood of θ_1 itself. It is because of this that LR is not applicable to NNM testing. When this happens, a "centered LR" is used instead. This will be discussed in Part II.

1.3.1. Information and entropy.

Often, it is convenient to use the log-likelihood function instead of the likelihood function because $-\ln p(y|.)$ is a nice measure of "information" —— a term from Information Theory.[3] By information, we mean the 'degree' of uncertainty associated with an event. That is, a rare event implies a high level of information, as its occurrence will provide more information than the occurrence of a frequent event. Thus for a rare event, $-\ln p(y|.)$ reflects this by being a very large positive number, as opposed to zero for a certainty.[4]

$(-\ln L)$ can also be interpreted as a measure of information because the log of the arbitrary constant is a constant. Consider the LR of two models,

$$\begin{aligned} \ln LR &= -\ln\{(L_1)/(L_2)\} \\ &= -\ln\{c_1 p_1/(c_2 p_2)\} \\ &= -\ln (p_1) - (-\ln (p_2)) - \{\ln c_1 - \ln c_2\}. \end{aligned}$$

Thus ln LR is just the difference in the information from two models; when the models are nested, the last term will disappear.

Another concept from Information Theory that will be useful later is entropy, (E). Entropy is the expected information from an event, i.e.

---

[3] See for example, Jones (1979).

[4] This measure of information has the additivity property, i.e. $-\ln((p_1)(p_2)) = -\ln (p_1) + (-\ln p_2)$.

$$(E) = -\int p \ln p \, dy$$
$$= E(-\ln p).$$

For later discussion, the following definition of entropy, (Eij), will be useful. If p_i is the pdf of model i, and p_j is the pdf of model j, then the expected information from model j, assuming that model i is true, is:

$$(E_{ij}) = -\int (p_i) \cdot \ln (p_j) \, dy$$
$$= E_i(-\ln (p_j)).$$

Part II.

Intuitive Interpretation of Existing NNM Tests.

When non-nested models are involved, the orthodox subset model selection methods are not applicable because of the non-existence of a comprehensive model. The inapplicability of the LR test is due to the presence of the nuisance parameters, c_1 and c_2. One way to solve the problem is to create an artificial nesting model (ANM), and use it as a comprehensive model. Another method is to use a "centered" or "modified" LR, in which the nuisance parameters have disappeared. In this paper, the former will be called the "regression approach", and the latter will be called the "centered likelihood ratio approach".

2.1. Centered likelihood ratio approach.

Because of the nuisance parameters, c_1 and c_2, the LR test is not applicable for non-nested model testing. In this situation, the centered likelihood ratio approach can be used to get rid of these nuisance parameters by subtracting from the LR another term which also contains c_1 and c_2. The Cox test makes use of this principle. The statistic that is used to test one model, say M(1), against a specific alternative model, say M(2), is derived from comparing the difference of the log-likelihood of the two models to its expected difference under the null. In other words, the Cox test compares the observed information change (ic), i.e. information using M(2) less information using M(1), to the expected ic assuming M(1) is true. If this value is "near" zero, then the ic is consistent with the truth of M(1), as the ic should not be too different from its expected ic under the null. Hence to reject the null when the statistic is near zero would not be appropriate. If, on the other hand, the observed value is "far" from zero in absolute value, then this ic is not consistent with the truth of M(1), and consequently, M(1) should be rejected.

2.1.2.Notes.

1. When this observed value (i.e. the CTS) is a large positive number, a case which arises when the observed ic is larger than the expected ic assuming M(1) is true, it means that the use of M(2) has produced more information than expected when M(1) is true. But this means, from the definition of information in Part I, that M(2) is highly unlikely. This suggests that a better model is not in the direction of M(2) (from M(1)). If this observed value is a large negative number instead, then it suggests that the use of M(2) has resulted in a reduction of information. Hence, a better model is in the direction of M(2) (away from M(1)). Note that in both cases, M(1) should be rejected.
2. This analysis, in note #1 above, which is in terms of the direction of the models, has been discarded by Dastoor (1981) as nonsensical because, as he argues, it is not clear what the direction of a model means. In his paper, he provides a case consisting of three non-nested classical normal linear regression models. He shows that if neither of the models specified in the test is the true model, then there is no means to tell the direction of the true model when both of the models are rejected. A moment's reflection shows that he could not make any conclusive statement on the direction of the true model because "direction" was

never defined in the first place. In Part III, it will be shown that by using the concept of direction as defined in the Preliminaries one can infer the direction of the true model from the Cox test statistic.

3. It is obvious that this testing procedure can only test two non-nested models at a time. As a result, if there are more than two competing non-nested models, the testing procedure will have to be repeated a number of times for different combinations of models.

4. This test is not symmetric because the test is conditional on the truth of the null model, say M(1), and not on the truth of the alternative model, say M(2). This means that when the null model is rejected, the alternative model is not automatically accepted. If one wishes to test the truth of M(2) (against the same specific M(1)), the roles of the previous null and alternative models can be reversed for a second-round testing. For this reason, this testing procedure may yield inconsistent test results, e.g. accept both models. Herein lies its disadvantage. However, an advantage is the possibility of rejecting both models (and all models under consideration as well).

5. When comparing two models, there is a total of nine conceivable outcomes. This is shown in Table 1 below.[5]

6. Sawyer (1983) has suggested a different test statistic, called here the Sawyer test statistic (STS), which is based on the same idea as that of Cox's except that it uses entropy change (ec) instead of the ic.[6]

---

[5] This table is similar to the one provided by Fisher and McAleer (1979, p.148).

[6] The ec, from using M(1) to using M(2) when M(2) is assumed true, is a measure on the directed divergence from M(1) in the direction of M(2). The larger the ec (i.e. the larger is the difference in the expected ic resulting

CTS_2 / CTS_1	Significantly positive. Reject M(2). a. M(1)	Insignificant. Do not reject M(2)	Significantly negative. Reject M(2). t. M(1)
Significantly positive. Reject M(1). a. M(2)	Case a. Reject both. Look elsewhere	Case b. Reject M(1). "Qualified" acceptance of M(2). Further testing is needed.	Case c. Reject both. t. M(1)
Insigficant. Do not reject M(1)	Case d. "Qualified" acceptance of M(1). Further testing is needed.	Case e. Do not reject either. Insufficient information to choose between them.	Case f. Reject M(2). Do not reject M(1).
Significantly negative. Reject M(1). t. M(2)	Case g. Reject both. t. M(2)	Case h. Reject M(1). Do not reject M(2).	Case i. Reject both. A combination of the two models may be useful.

Table 1. Cox test.

* CTS_i is the CTS resulted from testing M(i) agianst M(j) where i, j = 1, 2, and i $\neq$ j.

* "a. M(i)." denotes "the true model is away from M(i)".

* "t. M(i)." denotes "the true model is towards M(i)".

However the Sawyer's TS appears to complicate matters. It has never been shown that it will provide a better solution than the CTS.

2.2. Regression approach.

The regression approach uses an artificial nesting model (ANM), which contains each of the competing models as special cases. In every ANM, there is a parameter associated with each of the competing models. Let this parameter be called an artificial parameter (ap). The sum of this set of ap is usually restricted to one; hence, each of the models is a special case of this ANM when the artificial parameter associated with it is one while the others are zeros.[7] Thus, the ANM can perform the same function as a comprehensive model, i.e. for all the competing models to be compared to. The classical subset model selection tests can thus be used to test if any of the models is insignificantly different from the ANM. If it is, then it is said to be accepted as being consistent with the data, otherwise it is not. There are a few problems with this approach.

---

[6](cont'd) from the use of one model to another) the higher the preference is for M(1) if its value is positive; the smaller the ec (in negative value), the higher the preference is for M(2). (Sawyer, 1983.)

[7] For example, suppose there are two statistical models, p_1 and p_2, then the ANM is $p = k.p_1^{\lambda_1}p_2^{\lambda_2}$, where $K^{-1} = \int_R p_1^{\lambda_1}p_2^{\lambda_2}dy$, $\lambda_1 + \lambda_2 = 1$, and R is the appropriate region of integration. Here λ_1 and λ_2 are the ap. If the two models are classical normal linear regression models with pdf $N(X_i b_i,\ s_i^2 I)$ as defined in equation (5) (see Part III), then the ANM is $N(\alpha_1 X_1 b_1 + \alpha_2 X_2 b_2,\ s^2 I)$ where α_i's are functions of the ap, and s^2 is a function of the ap and s_i^2. In this case, the test of $\alpha_i = 0$ is equivalent to the test of $\lambda_i = 0$.

First, since the "closeness" of one model to the ANM would not preclude the possible closeness of other models to it, more than one model can possibly be found to be accepted. Secondly, when none of the competing models is found to be true, we will be forced to conclude that the ANM is the true model, even though it is not meaningful in explaining the distribution of the dependent variable. (Though this is not a good sign, it can be taken as a sign that the true model is closer to the ANM than any of the competing models.) Thirdly, there are many ways to construct an ANM. It is not clear if the results obtained from using one particular form of ANM will be compatible with the results obtained from another form of ANM. Since it is very possible that they are not, one can never know if the existing non-nested tests are optimum tests. (This could be an interesting area for future research.) Lastly, the unknown parameters within each of the models are usually not identifiable from those artificial parameters; this problem is circumvented by employing estimates of these unknown parameters.

In the regression approach, the TS used to test the set of competing models can be derived using the classical regression method or the maximum likelihood approach. Two methods in this category will be discussed. One is suggested by Dastoor (1983), the other by Davidson and MacKinnon (1981).

2.2.1. Dastoor's R test statistics.

The test statistic suggested by Dastoor, called Dastoor's R TS, compares the best (maximum likelihood) estimate of the parameters of an alternative model, say M(2), assuming M(2) is true, to its best estimate when the null model, say M(1), is assumed true. The difference between these two estimates should have an expected value of zero if M(1) is in fact the true model. Hence, if this difference is insignificant, then M(2) cannot be rejected; otherwise, it is rejected.

This TS is obviously asymmetric. In this case, however, there are only two possible outcomes (i.e. reject or not-to-reject a model) when testing the truth of one model against another. Two more possible outcomes could come about upon reversing the roles of these two models; thus giving rise to a total of four possible outcomes. This is shown in Table 2 below.

2.2.2. Davidson and MacKinnon's test statistics.

The method suggested by Davidson and MacKinnon (D&M) uses an ANM that is formed by combining linearly all the competing models in a regression model. The idea of D&M's test is that it attempts to test if the alternative model, say M(2), can significantly explain the variation of the dependent variable that is not explained by the null model, say M(1). (It is assumed here that there are only two competing models.) If it cannot, then M(1) cannot be rejected as this is consistent with the truth of M(1). If, on the other hand, it can, then M(1) has to be rejected. This is because the occurence of this latter case is not consistent with the truth of M(1), for if M(1) is in fact the true model, then the variation of y that cannot be explained by M(1) should be random.

This test is obviously asymmetric, as it only tests the "truth" of the null model, and no inference can be made about the the truth of the alternative model. Hence, to test the truth of the alternative model, the role of the previous null and alternative models can be reversed for a second

D's R_1 \ D's R_2	Insignificant. Do not reject M(1).	Significant. Reject M(1).
Significant. Do not reject M(2).	Do not reject either model.	Reject M(1) Do not reject M(2)
Significant. Reject M(2).	Do not reject M(1). Reject M(2).	Reject both.

Table 2. Dastoor's R test.

* D's Ri refers to Dastoor's R TS in the test of the truth of M(i), i = 1 or 2.

round testing. So, as with the Cox test, this may yield inconsistent test results. In the case of only two competing models, there is a total of four conceivable outcomes (instead of the nine possible outcomes as suggested by Fisher and McAleer (F&M) (1979)).[8] This is shown in Table 3 below.

Fisher and McAleer suggest that the sign and magnitude of the estimates of the ap can be employed to determine the "direction" of the true model. Unfortunately, this only makes sense in a context in which the true model is

[8] Note that the reason for having only four possible outcomes here is different from that of Dastoor's, whose arguments are based on the presumption that the direction of a model does not make sense.

J_1 \ J_2	Significant. Reject M(1).	Insignificant. Do not reject M(1)
Significant. Reject M(2).	Reject both.	Do not reject M(1) Reject M(2).
Insignificant. Do not reject M(2).	Reject M(1). Do not reject M(2).	Reject neither.

Table 3. J- test.

* J_i refers to the test of the truth of M(i), i = 1 or 2.

....

a linear combination of the competing models. Consequently, this means of assessing the true direction of a model is of little value; the exposition of section 3.4 utilizing the definition of direction given in section 1.2. earlier is able to overcome this problem only through use of the CTS.

One advantage of this test over the Cox test is that it can test against several alternative models at the same time, by using an ANM which has as its regressors the regressors of all competing models. In this case, a "sequence" of tests on the set of alternative models is needed. They are tested to see if the vector of ap's is significantly different from the zero vector. If they are, then the null model can be rejected, otherwise it cannot be rejected. The subsequent analysis of this case can be generalized from the above discussion.

2.3. Properties of the test statistics.

This section discusses the properties of the TS with the classical normal linear regression model as a special case.

2.3.1. The Cox and Sawyer procedures.

The CTS is computed by subtracting an estimate of the expected log-likelihood ratio, under the null, from the actual log-likelihood ratio (Cox, 1961, 1962). The Sawyer test statistic (STS) can be computed by subtracting an estimate of the expected ec (under the null) from the actual ec.[9] These TS can be shown to have an asymptotic normal distribution with zero mean and finite variance under the null hypothesis (Sawyer, 1983). Hence, one can test if they are significantly different from zero by using the asymptotic normal test.

Since the CTS is of probability order[10] $o_p(T)$ under the null hypothesis (where T is the sample size), we can determine its consistency by checking

[9] The ec used here is computed under the assumption that the alternative model, M(2), is true.

[10] Four types of "order" notations will be used in this thesis: $O(.)$, $o(.)$, $O_p(.)$, and $o_p(.)$. Suppose $u(T)$ and $v(T)$ are two different functions of T, then as T tends to infinity, the notation $u(T) = O(v(T))$ denotes that $|u(T)/v(T)|$ remains bounded as T tends to infinity, and the notation $u(T) = o(v(T))$ denotes that $\lim \{u(T)/v(T)\} = 0$. Suppose $X(T)$, $T = 1,2,\ldots$, is a sequence of random variable, and suppose further that it converges in distribution to X, then we write $X(T) = O_p(1)$. If there is another sequence of random variable, $Y(T)$, then the notation $X(T) = O_p(Y(T))$ denotes that the sequence $X(T)/Y(T)$ is $O_p(1)$, and the notation $X(T) = o_p(Y(T))$ denotes that the sequence $X(T)/Y(T)$ converges to zero in probability. Note that $X(T) = o_p(Y(T))$ implies that $X(T) = O_p(Y(T))$.

the value of the probability limit of (1/T) times the TS under both the null and alternative hypotheses. This value can be shown to be zero under the null, and negative under the alternative. Therefore, the test is consistent (Pereira, 1977). Likewise, the consistency of the STS can be determined as it is also of probability order $o_p(T)$. In this case, the value of this probability limit can be shown to be zero under the null, and non-zero under the alternative. Therefore this test is also consistent.[11]

2.3.1.1. Atkinson test statistic.

One variant of the CTS that has often been discussed in the literature is the Atkinson test statistic (ATS). This ATS employs estimators for the parameters in the TS which are consistent when evaluated under the null hypothesis, whereas in the Cox test, the estimates used are consistent when evaluated under their respective hypotheses.[12] This is an advantage over the CTS, because the best performance of the TS has been allowed under the null. Consequently, the null hypothesis is harder to reject when the alternative hypothesis is in fact true. Obviously, when compared to the CTS, this ATS will have a smaller size.

The main disadvantage of using the ATS is that the test is inconsistent, as has been proven by Pereira (1977). Pereira's result has been

---

[11] Although this has not been shown in the literature, it can be shown quite easily by using equation (18) of Sawyer (1983).

[12] In other words, the difference between the ATS and the CTS is the ic from using the estimated M(2) under H2 to using the estimated M(2) under H1. Since the former is less than the latter numerically, the ATS is thus larger than the CTS.

contested by Fisher and McAleer (1981), who argue that Pereira's conclusion was incorrect because of his inability to show that the ATS has a negative mean under the alternative hypothesis. They attempt to prove their case by showing that, for the classical normal regression model, the ATS converges to a negative value with probability one under the alternative hypothesis. However, as we will show in part III, the ATS is in fact an inconsistent test, and Fisher and McAleer have erred in their analysis. Even if Fisher and McAleer's result is correct, there is no contradiction between theirs and Pereira's results, since their studies do not have a common basis for comparison. Pereira discusses separate families of statistical models, whereas Fisher and McAleer discuss non-nested CNRM. Non-nested CNRM is only a class of statistical models.

2.3.1.2. **Small sample adjustment of the Cox test statistic.**

The available results from Monte Carlo studies indicate that the Cox test over-rejects the true null model in small samples, and the frequency of this over-rejection decreases as the sample size increases (Pesaran, 1974; and D&M, 1982). According to Godfrey and Pesaran (1983), this may be caused by the small sample bias of the CTS, as $T^{-1/2}$(CTS) has zero mean only asymptotically. (This is because the large sample property of the test statistic is used when there is only a small sample size, as the exact distribution of the CTS is unknown in small samples.)

Thus they suggest an adjusted TS, which involves two types of adjustments: the use of unbiased estimators of the parameters in the CTS, and an adjustment for the loss in the degrees of freedom in the statistic. This adjusted CTS, denoted here as ACTS, can be shown to have the same asymptotic properties as the CTS. Hence the usual testing procedure as applied to the CTS can be used here.

These adjustments in the CTS are only approximate. Simulation studies have indicated that when the adjusted CTS is used with the "old" variance --- this standarized TS is denoted here as ACTS* --- it still over-rejects the true null hypothesis, and this over-rejection decreases as the sample size increases. It has also been shown that CTS and ACTS* perform about the same by type I error criterion (D&M, 1982). However, in another simulation study where this ACTS was used with a new variance, it has significantly reduced the frequency of rejecting the true null hypothesis, relative to the use of CTS (Godfrey and Pesaran, 1983). The new variance that is used along with the ACTS is in fact the variance of a new TS, called the W- TS. (This W- TS will be discussed later.) This new variance is larger than the "old" variance of CTS by one positive term.

There are two problems that need to be considered:

a. why the use of the ACTS with the smaller old variance will not reduce the over-rejection of the true null hypothesis, and why this over-rejection decreases as the sample size increases?

b. why the use of ACTS with a larger variance will decrease the over-rejection of the true null hypothesis in small samples?

One possible explanation for these is that the usual suggested estimated variance is not appropriate for finite sample testing, and a more appropriate variance is one with a larger size, but their size difference will vanish as the sample size increases. Later, in Part III, it will be shown that a more appropriate variance to be used should have one more positive term, and this extra term will converge to zero as the sample size increases.

2.3.1.3.W-TS.

Godfrey and Pesaran (1983) have suggested a new TS, called W- TS. This TS was derived in an effort to search for a TS that can overcome the small sample bias of the CTS. They look at the numerator of the linearized CTS, and find that it has a non-zero mean under the null. A statistic is thus formed by subtracting from it an unbiased estimate of its expected value under the null. This statistic when divided by its standard deviation is the W- TS. Since the denominator of the linearized CTS converges to a finite constant as the sample size increases, and its numerator is of probability order unity, the W- TS is thus a variant of CTS. It has the same asymptotic distribution as that of the linearized CTS, and hence it is used as a CTS. A simulation study has shown that this W- TS has reduced significantly the type I error of the CTS (Godfrey and Pesaran, 1983).

2.3.2.Dastoor's R TS.

As mentioned in section 2.2.1. earlier, Dastoor's R TS examines the difference between two estimates of the parameters of the alternative model, one estimated under the null and the other estimated under the alternative hypothesis. This difference can be shown to have a normal distribution, and hence can be exploited, along with its variance-covariance matrix, to produce a statistic with a chi-square distribution.

The variance of the residual error, which appears in the variance-covariance matrix is generally unknown, and hence have to be estimated. Two commonly used estimated variances are: the variance estimated when the null hypothesis is assumed to be true, and the variance estimated when the ANM is assumed to be true. The ANM used has as its regressors the combined regressors of both models, M(1) and M(2). The use of the former estimated variance will yield a test that is equivalent to the usual Lagrangian-multiplier (LM) TS for testing the null hypothesis of zero restrictions on the parameters of M(2) in the ANM, while the use of the latter would yield a test that is equivalent to the usual Wald test.

Dastoor's R TS are equivalent to the orthodox LM and Wald TS. Furthermore, they are distributed asymptotically as central chi-square variates under the null hypothesis, and as non-central chi-square variates under the alternative hypothesis with a non-centrality parameter of order O(T). Consequently, they are consistent tests.

Note that this method looks at the problem indirectly. When the expected difference between the two estimates is zero, it does not necessarily mean that the first model is true. An example is when the true model is M(3), which is orthogonal to both M(1) and M(2).

2.3.3.Davidson and MacKinnon test statistics.

The D&M test can be done by regressing the dependent variable on the set of competing models, with the restriction that the sum of all the artificial parameters associated with each of the competing models must equal one. These parameters, if they are identifiable, are then tested to see if they are significantly different from zero. If they are not, then the null model is to be rejected, otherwise not.

Since these artificial parameters are usually not identifiable, as the set of competing models contain unknown parameters, at least (N-1) of the models have to be estimated in order for the test to be carried out[13] (where N is the number of competing models to be considered in the test). For this reason, there are two basic versions of the test. The first uses all estimated models in the regression and the second uses (N-1) estimated models for those models whose associated parameters needed to be identified in the regression. The first test is called the <u>C-test</u>, while the second is called the <u>J-test</u>.

These two TS can be shown to have an asymptotic normal distribution. For the J-test, the TS has zero mean and finite variance. Hence it can be tested to see if it is significantly different from zero by using asymptotic normal test. As for the C-test, the TS has zero mean only when certain consistent estimators are used, and it has a finite variance, which is biased as shown by Davidson and MacKinnon (1981, p787). An unbiased estimated variance for this C-test can be computed by doing an auxiliary regression and some other calculations. However, D&M note that a simpler method is to use another regression model which has the null model linearized to the first degree, while keeping all the other competing models as they are. This is called the <u>P-test</u>. Notice that this P-test will be exactly the same as the J-test if the null model is a linear model, because a linearized model of a linear model is a linear model.

---

[13] An example might clarify this. Suppose the competing models are the two classical normal linear regression models as mentioned in footnote #7, then the test of α_2 equals zero is accomplished by first regressing y against X_1 and the composite variable, $X_2\hat{b}_2$ (where $\hat{b}_2$ is a consistent estimate of b_2), and subsequently use a t- statistic to effect the required test.

Since there exist many possible estimated models, there is the problem of determining the best consistent estimators to be used. D&M have suggested estimating the models by replacing their parameters with their MLE under their respective hypotheses. The resulting tests can be shown to be consistent tests.

For the J-test, F&M have suggested estimating the alternative models by using their consistent estimators under the null. This bears close resemblance to the ATS, and hence it is called the JA-test. In the case where there are only two competing linear models, the distribution of the resulting TS can be shown to have a t-distribution in finite samples under the null hypothesis. For this reason, this JA-test has been suggested to be preferable to the J-test. Nevertheless, because of its unknown distribution under the alternative hypothesis, the power function of the TS cannot be derived. It can be shown that there is in fact a class of consistent estimators which, when used in the J-test, will yield a TS that has a t-distribution under the null. Subsumed in this class is another set of consistent estimators that will yield a TS with a t-distribution under both null and alternative hypotheses, in finite samples. However, the gist of the problem still remains, namely that there is no unique "best" estimator within this class of estimators.

2.4.Comparison of the TS.

Since all the above-discussed TS essentially do the same work, using all of them will be a waste of resources and time. For this reason, it is necessary to compare these TS so as to determine the optimum one under various conditions, and to evaluate their relative efficiency, so that the loss in efficiency incurred in using any other test than the optimum can be determined, and thus allow us to weigh the costs and benefits from using these sub-optimum TS.

There are many ways to define an optimum TS. For example, it can be defined as one that maximizes the power of the test, or one that maximizes the "local power" of the test, or one that involves the least computational costs, or some other criteria. In the following, we will discuss only a few of these. Let us begin with relative efficiency.

An "efficient" test is the most powerful test in the class of tests being considered. And if an efficient test of size-α requires T_1 observations to attain a certain power, and a second size-α test requires T_2 observations to attain the same power against the same alternative, then the relative efficiency of the second test in attaining that power against that alternative is defined as T_1/T_2. This definition of the relative efficiency is not asymptotic, and it imposes no restriction upon the forms of the sampling distribution of the test being considered. It is possible to compare any two tests in this way because the power functions of the tests, from which the relative efficiency is calculated, take comprehensive account of the distributions of the TS; the power functions contain all the relevant information for our comparison. Therefore, in order to determine the relative efficiency of a TS, knowledge of its power function is required.

2.4.1. Comparison of the power functions of the test.

The power function of a test is a function that shows (a) the size-α of the test (when the null hypothesis is true), and (b) the probability of rejecting the false null hypothesis. It is determined by three factors: the size-α of the test, the "distance" between the hypotheses tested, and the sample size required by the efficient test. This means that three entries are needed, one for each of the three factors, in order to have a useful comparison of the power functions of different tests. Notice also that it is also necessary to define the "distance between the hypotheses tested" before we can have a fruitful discussion.

In the nested model case, the distance between the hypotheses can be defined in terms of the parameters' values. However, in the non-nested models case, this is not possible because, by definition, there can be no connection between them in terms of the parameters of the models. In the non-nested model testing literature, this distance is usually not explicitly defined, and not even mentioned in some papers. Only Pesaran (1982) has attempted to define this concept of distance. In the case of two non-nested models with one independent variable for each model, the distance between the two is defined in terms of the correlation coefficient between the two variables. Where there are more than one independent variable for the competing models, the multicollinearity of the columns of the two matrices is used as a measure of distance instead. This multicollinearity measure is possible because he assumes that there is a specific relationship between the two matrices.

Apart from this distance problem, there is also another problem. The theoretical power functions of all the TS are unknown in finite samples because their finite sample distributions are unknown. Furthermore, even though these tests are consistent and have correct sizes asymptotically, their power converges to one as the sample size increases. Thus, this is not useful for our analysis. There are two obvious ways to get around this problem. One is to use simulation to get an approximate power function of the test, and the other is to first impose certain restrictions to the effect that the asymptotic power of these tests do not converge to one, and then compare their asymptotic power. In the literature, the latter is analysed in terms of local alternatives.

2.4.2. Local alternatives.

Pesaran (1982) has defined a sequence of alternative hypotheses which approaches the null hypothesis in some well-defined manner. This sequence of alternative hypotheses, which approaches the null as the sample size increases, is called a set of local alternatives. It is used to compare the asymptotic efficiency of the tests as sample size increases.

Pesaran constructs the set of local alternatives in terms of the regressor matrices. One restriction[14] is that the regressor matrix of the null model, M(1), must be at least as large as that of the alternative model, M(2). It is also assumed that the two exogenous matrices are related in such a way that the regressor matrix of M(2) is a combination of three matrices. The first of these is linearly dependent on the regressors of M(1), and the other two are matrices of constants, one of which is of order $O(T^{-1/2})$, and the other is of order $o(T^{-1/2})$. In addition, he assumes some regularity conditions to keep the power bound away from unity as the alternative hypothesis approaches the null, and to keep the two asymptotic moment matrices of the regressors from exploding as the sample size tends to infinity.

---

14 ~~Because of this restriction, the roles of the two models cannot be~~ reversed for a second round testing.

Pesaran has applied this method to the CTS, J-test and the LM- test, and it is shown that the Cox test and the J-test have the same limiting local power. The square of the CTS and the LM- test both have non-central chi-square distribution, with one and K* degrees of freedom respectively, and they have the same non-centrality parameter. K* is the number of regressors in the alternative model that are not found among the regressors of the null model. This difference in the degrees of freedom suggests that the two non-nested tests have more local power than the LM- test, since the larger the degrees of freedom, the higher is the significance level for the same critical point. Notice that, because of the way Pesaran designed his local alternatives, the non-nested tests have greater power than the orthodox test only when the number of regressors in the null model is at least as large as the number of regressors in the alternative model.

In part III, we will apply the same local alternatives analysis to the ATS. It is shown that the ATS is asymptotically equivalent to the CTS.

2.4.3.Note.

1. Davidson and MacKinnon (1982) say that there are two ways to design local alternatives. One way is to assume that H_1 is fixed, while H_2 approaches H_1 in some well-defined manner. This is the way suggested by Pesaran as discussed above. The other way is to assume that both M(1) and M(2) are fixed, but the true model is known to lie "between" M(1) and M(2), and this true model approaches M(1) in some well-defined manner. They argue that this latter approach has an advantage over the first approach in that neither of the hypotheses needs to be assumed to be true. They thus used this approach in their analysis. In their

investigation of the local power of the P-, "PA-",[15] and the Cox tests, under their definition of local alternatives, it is found that the three tests are asymptotically equivalent.

---

[15] PA- test is the P- test which uses all estimators evaluated under the null hypothesis. It is based on the same idea as the JA- test.

Part III.

An Examination of Selected Aspects of Cox Test Statistic.

In this Part III, five aspects of CTS are examined.

1) An alternative derivation of the CTS (and ATS) when the competing models are classical normal linear regression models is provided.
2) The asymptotic properties of CTS are derived, and it is found that a more suitable variance for the CTS in finite sample test should have a larger magnitude than that usually employed.
3) The ATS is shown to be an inconsistent test, in contrast to current belief.
4) The concept of the "direction" of a model defined in Part I is applied to the Cox test and it is shown that in some cases the direction of the true model can be inferred from the sign of the CTS.
5) An alternative method for deriving the asymptotic distribution of the CTS under local alternatives is provided. The asymptotic properties of the ATS under local alternatives is also provided.

Suppose there are two competing separate families of statistical models:

$$H_i : p_i = p_i(y,\theta_i),\ i = 1,\ 2,$$

where for $i = 1$ or 2, H_i denotes hypothesis i, p_i is the probability density of model i, y is a vector representing an independently distributed dependent variable, and θ_i is a vector of parameters of model i. The CTS and ATS for testing H_1 against H_2 are respectively:[16]

$$CTS = \ln(\hat{p}_1/\hat{p}_2) - \{E_1 \ln(\hat{p}_1/\hat{p}_2)\}_{\hat{\theta}_1}$$
$$ATS = \ln(\hat{p}_1/\hat{p}_{21}) - \{E_1 \ln(\hat{p}_1/\hat{p}_2)\}_{\hat{\theta}_1}$$

where for i= 1, 2, $\hat{p}_i$ is $p(y,\hat{\theta}_i)$ (here y represents the sample value of the dependent variable), p_{21} is $p_2(y,\ \theta_{21})$, $\hat{p}_{21}$ is $p_2(y,\ \hat{\theta}_{21})$, θ_{21} is the probability limit of $\hat{\theta}_2$ under H_1, and $\hat{\theta}_{21}$ is its consistent estimate. $E_i(.)$ denotes expectation under H_i. $\{\ldots\}_{\hat{\theta}_i}$ means evaluated at $\theta_i = \hat{\theta}_i$. Since $\ln(\hat{p}_1/\hat{p}_2)$ is asymptotically equivalent to $\ln(p_1/p_{21})$,[17] the following formulae for the CTS and ATS can also be used.

$$CTS = \ln(\hat{p}_1/\hat{p}_2) - \{E_1 \ln(p_1/p_{21})\}_{\hat{\theta}_1} \tag{2}$$
$$ATS = \ln(\hat{p}_1/\hat{p}_{21}) - \{E_1 \ln(p_1/p_{21})\}_{\hat{\theta}_1} \tag{3}$$

Since these two statistics are asymptotically equivalent under the null, their asymptotic variances are thus equal. It is usually computed by using the following general formula:[18]

$$V_1(CTS) = V_1(S) - \{\partial(E_1S)/\partial\theta_1\}'(I)^{-1}\{\partial(E_1S)/\partial\theta_1\} \tag{4}$$

---

16 See Cox (1961,1962); Atkinson (1970).

17 See Walker (1967).

18 See Walker (1967).

where, $V_1(S)$ is the variance of S under H_1, S is ln (p_1/p_2), and (I) is the information matrix of θ_1.[19]

When non-nested classical normal linear regression models are to be tested, the above two test statistics can be specialized. Suppose there are two competing models, for i = 1, 2:

$$H_i : y \sim N(X_i b_i, s_i^2 I), \text{ given } X_i, \tag{5}$$

$$\Rightarrow H_i : y = X_i b_i + \epsilon_i, \tag{6}$$

$$\epsilon_i \sim N(0, s_i^2 I) \tag{7}$$

where X_i is a matrix of independent variables, b_i is a vector of parameters of model i, ϵ_i is a vector of random disturbances of model i, and s_i^2 is the variance of ϵ_i. It is assumed that the columns of X_1 and X_2 are independent but not orthogonal to each other. The TS are:[20]

$$CTS = -(T/2)\ln \{\hat{s}_{21}^2/\hat{s}_2^2\}, \tag{8}$$

$$ATS = CTS + \ln (\hat{p}_2/\hat{p}_{21}) \tag{9}$$

$$= \hat{\epsilon}_1' M_2 X_1 \hat{b}_1/\hat{s}_{21}^2, \tag{10}$$

where $M_i = I - P_i$, and $P_i = X_i(X_i'X_i)^{-1}X_i'$. Using (4), the asymptotic variance of these TS can be shown to be[21]

$$(s_1^2/s_{21}^4) b_1' X_1' M_2 M_1 M_2 X_1 b_1. \tag{11}$$

A consistent estimate of this asymptotic variance involves the replacement of all unknown parameters in the formula by their respective consistent

---

19 This $(I)^{-1}$ is $-\{E_1[\partial^2(\ln p_1)/(\partial\theta_1\partial\theta_1')]\}$.

20 See Pesaran (1974) for equation (8); and Fisher and McAleer (1981) for equation (9), from which after some algebraic manipulation one can get equation (10).

21 See Pesaran (1974).

estimators.

3.1. An alternative derivation of the CTS.

In the literature, the CTS is derived by using Pesaran's method, which requires knowledge of large sample asymptotic distribution theory.[22] When the competing models are classical normal (linear) regression models, the CTS can be derived without this knowledge.[23] This alternative derivation makes use of the information that each of the p_i's is normal pdf. So,

$$\ln p_i = -(T/2)\ln(2\pi s_i^2)-(y-X_ib_i)'(y-X_ib_i)/(2s_i^2), \quad \text{for } i = 1, 2,$$

$$\Rightarrow \quad \ln \hat{p}_i = -(T/2)\ln(2\pi \hat{s}_i^2) - T/2, \quad \text{for } i = 1, 2, \tag{12}$$

$$\ln p_{21} = -(T/2)\ln(2\pi s_{21}^2)-(y-X_2b_{21})'(y-X_2b_{21})/(2s_{21}^2),$$

$$\text{and } \{E_1(\ln p_1 - \ln p_{21})\} = -(T/2)\ln(s_1^2/s_{21}^2)$$

$$\Rightarrow \quad \{E_1(\ln p_1 - \ln p_{21})\}_{\hat{\theta}_1} = -(T/2)\ln(\hat{s}_1^2/\hat{s}_{21}^2). \tag{13}$$

On substituting (12) and (13) into (2), we get equation (8).

---

[22] Pesaran derived the CTS as follows:

a. Assume that the limits of the products of the regressor matrices exist, and are finite; i.e. (see equation (6)) $\lim X_i'X_j/T = C_{ij}$, for $i,j=1,2$.

b. Replace all unknown expected values of any function by the probability limits (plim) of the function,

c. Replace the plim of the estimates by their consistent estimates; and

d. Use maximum likelihood estimates as the required consistent estimate.

[23] This specialization is useful since econometricians deal mostly with the classical normal regression model.

It is worthwhile to note that the ATS in equation (10) can also be derived in the same manner as above by substituting (12), (13) and

$$\ln \hat{p}_{21} = -(T/2)\ln(2\pi\, \hat{s}_{21}^2) - T/2 - \hat{\epsilon}_1' X_2 \hat{b}_{21}/\hat{s}_{21}^2$$

into (3).

3.2. Asymptotic properties of the CTS.

In this section, the asymptotic properties of the CTS will be derived. The results obtained here are different from those found in the literature in two respects: the asymptotic mean is not zero under the null hypothesis, and the asymptotic variance is larger by an extra term, which is positive in value. The reason for these differences is that the TS used in the literature is truncated up to probability order $o_p(T^{1/2})$. Any term of order less than that is ignored (Cox, 1962, p.409). If only the terms of probability order $o_p(T^{1/2})$ are considered, the results obtained here will be exactly the same as those in the literature.

The advantage of the following method for deriving the asymptotic properties of the CTS is that it yields a better estimate of the asymptotic variance of the CTS. Furthermore, it allows us to see the relationship between the CTS and the W- TS, and hence enable us to explain the improved performance of the W- TS and the adjusted CTS over the conventional CTS.

Equation (8), upon linearizing to the first order,[24] becomes

---

[24] Another form of linearized CTS is

$$\mathrm{CTS} \approx -(T/2)\{(\hat{s}_{21}^2 - \hat{s}_2^2)/\hat{s}_2^2\}. \qquad (15)$$

Under H_1, this linearized form is asymptotically equivalent to the form in equation (14). Hence, either form can be used. However, equation (14) is used here because it is more convenient for our analysis in this section. Equation (15) is used in a later section.

$$CTS \gtrsim -(T/2)\{(\hat{s}^2_{21} - \hat{s}^2_2)/\hat{s}^2_{21}\} \tag{14}$$

$$= -y'(M_1 + P_1M_2P_1 - M_2)y/(2\hat{s}^2_{21}). \tag{16}$$

Under H_1, equation (16) becomes:[25]

$$CTS = \{b_1'X_1'M_2M_1\epsilon_1 - \epsilon_1'(P_2 - P_1P_2P_1)\epsilon_1/2\}/\hat{s}^2_{21} \tag{17}$$

The two terms in the numerator are of probability order $O_p(T^{1/2})$ and $O_p(1)$ respectively, while the denominator converges in probability to s^2_{21}. The CTS thus has the same asymptotic distribution as its numerator under H_1. Equation (17) shows that it has a mean of

$$-(s^2_1)q/(2s^2_{21}), \tag{18}$$

where $q = (K_2 - tr(P_1P_2))$, and variance

$$\{s^2_1b_1'X_1'M_2M_1M_2X_1b_1 + s^4_1tr(P_2 - P_1P_2P_1)^2/2\}/s^4_{21}.$$

This variance is larger than the one in (11) by its second term. This second term is important because it has a positive non-zero value even asymptotically. Its omission will increase the type I error of the (finite sample) test. Hence, in simulation studies, the over-rejection of the true null hypothesis by the CTS and the reduction of this frequency of

---

25 Under H_2, equation (17) becomes

$$CTS=-\{b_2'X_2'A_1X_2b_2+2b_2'X_2'A_1\epsilon_2+\epsilon_2'(P_2-P_1P_2P_1)\epsilon_2\}/(2\hat{s}^2_{21})$$

where $A_1 = (I - P_1P_2P_1)$. The first term in the numerator is of order $O(T)$, while the other two terms are of probability order $O_p(T^{1/2})$ and $O_p(1)$ respectively. In this case, $\hat{s}^2_{21}$ converges in probability to s^2_{212}. So under H_2 CTS has the same asymptotic distribution as its numerator with mean

$$\{-b_2'X_2'(I-P_1P_2P_1)X_2b_2/2 + s^2_2q\}/s^2_{212},$$

and variance

$$\{s^2_2b_2'X_2'(I-P_1P_2P_1)^2X_2b_2 + s^4_2tr(P_2-P_1P_2P_1)^2/2\}/(s^4_{212}).$$

over-rejection by the adjusted CTS, which is used with a larger variance, are not surprising results.

Since the asymptotic mean of CTS is not zero, it is possible to correct for this "bias" in the test by using a "centered" CTS. This centered CTS is:

$$\mathrm{CCTS} = \mathrm{CTS} - \{E_1(\mathrm{CTS})\}_{\hat{\theta}_1}. \tag{19}$$

This TS obviously has an asymptotic zero mean and finite variance. This TS, if standardized, can be used as the usual CTS.

On substituting equation (14) and (18) into (19), we get

$$\begin{aligned}\mathrm{CCTS} &= -T(\hat{s}^2_{21} - \hat{s}^2_2)/(2\hat{s}^2_{21}) + \hat{s}^2_1 q/(2\hat{s}^2_{21}) \\ &= -\{T(\hat{s}^2_{21} - \hat{s}^2_2) - \hat{s}^2_1 q\}/(2\hat{s}^2_{21}) \\ &= -\{y'(M_1 + P_1M_2P_1 - M_2 - M_1q/T)y\}/(2\hat{s}^2_{21}).\end{aligned} \tag{20}$$

Since $\hat{s}^2_{21}$ converges to s^2_{21} and s^2_{212} with probablity one under H_1 and H_2 respectively, and since the numerator is of probability order $O_p(T^{1/2})$, only the numerator need to be considered if the purpose is to find the asymptotic properties of the TS. Let

$$n = -y'By,$$

where $B = -\{c.M_1 + P_1M_2P_1 - M_2\}$, and $c = 1 - q/T$. This variate, n, under H_1, obviously has an asymptotic zero mean and it can be written as

$$n = 2b_1'X_1'M_2M_1\epsilon_1 + \epsilon_1'B\epsilon_1.$$

Thus under H_1, it has a variance of

$$V_1(n) = 4s_1^2 b_1'X_1'M_2M_1M_2X_1b_1 + 2s_1^4\mathrm{tr}B^2.$$

While under H_2, n can be written as

$$n = b_2'X_2'BX_2b_2 + 2b_2'X_2'B\epsilon_2 + \epsilon_2'B\epsilon_2.$$

This shows that, under H_2, it has a mean of

$$b_2'X_2'(cM_1 + P_1M_2P_1)X_2b_2 + s_2^2\mathrm{tr}B,$$

and variance

$$4s_2^2 b_2' X_2' B^2 X_2 b_2 + 2s_2^4 tr(B^2).$$

Thus the CCTS has an asymptotic zero mean, and a variance of $V_1(n)/(4s_{21}^2)$ under H_1; while under H_2, it has an asymptotic mean of

$$\{b_2' X_2' (cM_1 + P_1 M_2 P_1) X_2 b_2 + s_2^2 trB\}/(2s_{212}^2),$$

and variance of

$$\{4s_2^2 b_2' X_2' B^2 X_2 b_2 + 2s_2^4 trB^2\}/(4s_{212}^4).$$

If all unknowns in the CCTS (equation (20)) are replaced by unbiased estimators and it is appropriately standardized, the W- TS of Godfrey and Pesaran (1983) is produced. In other words, this CCTS without the denominator is the numerator of the W- TS. Godfrey and Pesaran have called this W- TS a correction for the small sample bias of the CTS, since it is assumed that the asymptotic mean of the CTS is zero. However, as mentioned earlier, the asymptotic mean of the CTS is not zero (even asymptotically); the correction for the bias is thus not only for the small sample bias but also for the "non-zero bias".

In essence, the adjusted CTS of Godfrey and Pesaran is the CTS that uses all unbiased estimates. The W- TS is the standardized CCTS of equation (20) when all the unknown estimates are replaced by unbiased estimates.

3.3. Consistency of the Atkinson test statistic.

Pereira (1977) has proved the inconsistency of the Atkinson test, when it was applied to select separate families of statistical models, by showing that the Atkinson test statistic is indeterminate in sign under the alternative hypothesis. However, recently, Fisher and McAleer (1981) "proved" that the Atkinson test is still a consistent test when applied to test non-nested classical normal regression models; they also argued that Pereira is incorrect in his conclusion because of his inability to show that the ATS has a negative mean under the alternative hypothesis.

This section attempts to show that Fisher and McAleer are in error and that for non-nested model testing the Atkinson test is an inconsistent test.

By a consistent test we mean a test that has its power[26] (evaluated at a given fixed alternative hypothesis) converging to one as the sample size tends to infinity. This means that a consistent test will be one which has

$$T.\text{plim}_1 \{ATS/T\} = 0,$$

and

$$T.\text{plim}_2 \{ATS/T\} \neq 0.$$

where plim_i, $i = 1$ or 2, denotes probability limit assuming $M(i)$ is true.

On taking the plim of ATS we get: [27]

$$T.\text{plim}_1 (ATS/T) = 0$$

and

$$T.\text{plim}_2 (ATS/T) = -(T/2)b_{12}'G_1b_{12}/\{s_2^2+b_2'G_2b_2+b_{12}'G_1b_{12}\} \leq 0$$

---

[26] That is, the probability of rejecting a false hypothesis.

[27] The following is useful for the derivations of the plim of the two statistics.

$$\text{plim}_1 \hat{s}_{21}^2 = \text{plim}_1 \hat{s}_2^2 = s_{21}^2 \text{ (by definition)}$$
$$\text{plim}_1 (\hat{e}_1'M_2X_1\hat{b}_1/T) = 0$$
$$\text{plim}_2 (\hat{e}_1'M_2X_1\hat{b}_1/T) = -b_{12}'G_1b_{12}$$
$$\text{plim}_2 \hat{s}_{21}^2 = s_2^2 + b_{12}'G_1b_{12} + b_2'G_2b_2$$
$$\text{plim}_2 \hat{s}_2^2 = s_2^2,$$

where $b_{12} = C_{11}^{-1}C_{12}b_2$, and

$$G_i = \lim (X_i'M_jX_i/T),\ i,j=1,2,\ i\neq j.$$

with the equality sign holding when

$$b_{12}'G_1b_{12} = 0 \tag{21}$$

$$b_{12} = C_{11}^{-1} C_{12}b_2.$$

Consider equation (21). If $r(C_{12}) = K_2 \leq K_1$, then $b_{12} \neq 0$ because $b_2 \neq 0$ under H_2, and all columns of C_{12} are independent.[28] However, if $r(C_{12}) = K_1 < K_2$, then it is possible for b_{12} to be zero because the columns of C_{12} are not independent, meaning that it is possible to have a non-zero vector h such that $C_{12}h = 0$. Since there is a non-zero chance for equation (21) to hold, it is thus possible that the Atkinson test is an inconsistent test. This shows that the conclusion of Fisher and McAleer is incorrect. Their error lies in the last equation of their 1981 paper (p.118), where they argue that (in their symbols):

$$-\theta_1'\{\text{plim}_1\ n^{-1}Z'P_x(I - P_z)P_xZ\}\theta_1 \ ,$$

(which is equivalent to $b_2'C_{21}C_{11}^{-1}G_1C_{11}^{-1}C_{12}b_2$ in the symbols of this paper) is negative. As demonstrated, this is not necessarily true.

3.4. Interpreting the direction of the true model when neither of the tested models is true.

Ever since Pesaran (1974) specialized the Cox statistic to the testing of non-nested classical normal linear regression models, econometricians have discussed non-nested models as if there were a direction for each model.[29]

---

[28] C_{ij} is the moment matrix ($\lim X_i'X_j/T$), i,j=1,2. Since all the columns of X_1 and X_2 are linearly independent $r(C_{ij}) = K_i < K_j$ tells us that the number of independent columns of X_i is less than that of X_j, $i \neq j$.

[29] For examples, Pesaran (1974), and F&M (1981).

Dastoor (1981) was the first to point out that this concept of direction of a model has never been expounded, and it is not clear exactly what it means. For example, in Pesaran (1974)[30] it is not clear what is meant by "...a significant positive value for N_0 can be interpreted as strong evidence against H_0 in favour of an alternative which differs from H_0 in some sense opposite to that in which H_1 differs from H_0." (The term N_0 refers to the CTS.)

What follows is an attempt to show that, using the concept of the direction of the model defined in the Preliminaries, it is possible to infer the direction of the true model from the sign of the Cox statistic. The following results are needed for this analysis. Suppose the true model is:

$$H_3:\ y = X_3 b_3 + \epsilon_3 \tag{22}$$

$$\epsilon_3 \sim N(0,\ s_3^2 I).$$

In this case equation (8) can be written as:

$$CTS = -(T/2)\ln\left[\frac{b_3{}'X_3{}'A_1X_3b_3/T + s_3^2}{b_3{}'X_3{}'M_2X_3b_3/T + s_3^2} + o_p(1)\right]$$

where $A_1 = M_1 + P_1M_2P_1 = I - P_1P_2P_1$. This shows that CTS is still a consistent test even when both of the hypotheses under test are false. (As it is obvious that $\text{plim}_3\ (CTS/T) \neq 0$.)

We can now proceed to show our case. The following example is given by Dastoor (1981) but the notation follows that used in this paper.

The three models used are as given in (6) and (22). H_3 is assumed to be the true model, while H_1 and H_2 are used as the null and alternative hypotheses. In the first test, H_1 is assumed to be the null hypothesis and H_2

30 Walker (1967) has also written a statement to the same effect.

the alternative hypothesis. Their roles are reversed in the second test. Thus the CTS are:

$$CTS_1 = -(T/2)\ln(\hat{s}^2_{21}/\hat{s}^2_2)$$
$$CTS_2 = -(T/2)\ln(\hat{s}^2_{12}/\hat{s}^2_1)$$

Since H_3 is assumed true, they can be rewritten as:

$$CTS_1 = -(T/2)\ln\left[\frac{b_3'X_3'A_1X_3b_3/T + s_3^2}{b_3'X_3'M_2X_3b_3/T + s_3^2} + o_p(1)\right] \qquad (23)$$

$$CTS_2 = -(T/2)\ln\left[\frac{b_3'X_3'A_2X_3b_3/T + s_3^2}{b_3'X_3'M_1X_3b_3/T + s_3^2} + o_p(1)\right] \qquad (24)$$

where $A_1 = M_1 + P_1M_2P_1 = I - P_1P_2P_1$, and $A_2 = M_2 + P_2M_1P_2 = I - P_2P_1P_2$. From (23) and (24), one can see that the signs of CTS_1 and CTS_2 depend on the sign of $(a0-b0)/T$ and $(a1-b1)/T$ respectively, where

$$a0 = u'P_2u,$$
$$b0 = u'P_1P_2P_1u,$$
$$a1 = u'P_1u,$$
$$b1 = u'P_2P_1P_2u,$$

and $\quad u = X_3b_3.$

u reflects the true model (i.e. equation (22)). It is a vector in the vector space spanned by the column vectors of X_3. It can be shown easily that[31]

$$a0-b0 > 0 \quad => \quad CTS_1 < 0$$

---

[31] Note that this analysis is supposedly based on the limit of $(a0-b0)/T$ and $(a1-b1)/T$. However, since these limits are continuous functions of $(a0-b0)/T$ and $(a1-b1)/T$ respectively, and since the latter set of formulae are easier to work with, they are thus used in this analysis.

$a0-b0 < 0 \quad \Rightarrow \quad CTS_1 > 0$

$a1-b1 > 0 \quad \Rightarrow \quad CTS_2 < 0$

$a1-b1 < 0 \quad \Rightarrow \quad CTS_2 > 0$

Given that the Cox test is a consistent test, and given the assumption that M(3) is the true model, neither (a0-b0) nor (a1-b1) can be expected to be zero, even though in finite samples, it is possible for it to be zero.

The above shows that there are four possible combinations of the signs of the CTS, if the five cases where either or both of the CTS is zero are ruled out. These are shown in Table 4 below. All four entries in the table refer to cases in which the absolute value of the CTS is large enough that both of the models under test are rejected. However, different entries in the table impart different messages regarding the direction of the true model. The purpose here is to determine if these messages are obtainable from knowledge of the occurrence of a particular case. In other words, we wish to see if we can determine the sizes of the cosine of the smallest angle between u and M(1) relative to that between u and M(2) from the signs of CTS_1 and CTS_2.

The cases in the Table 4 are numbered 1 to 4. Cases 1 to 3 are numbered as Dastoor has numbered them. Case 4 is analytically equivalent to case 2, hence it is not discussed. These cases correspond to cases c, i, g and a in Table 1 above.

To begin, let us mould a0-b0 and a1-b1 to some forms useful for our analysis. First, we need the following. Imagine that X_1 and X_2 are both hyperplanes, and u is a vector. Let the smallest angle between u and X_i be r_i, i=1,2. Let the smallest angle between the vector projection of u onto X_1 and its closest vector in X_2 be r_3, and let the smallest angle between the

CTS_1 (a0-b0) \ CTS_2 (a1-b1)	Positive. (Negative)	Negative. (Positive)
Positive (Negative)	Case 4. Look elsewhere.	Case 1. t. M(1).
Negative (Positive)	Case 3. t. M(2).	Case 2. A combination of both.

Table 4. Cox test.

* "a. M(i)." denotes "the true model is away from M(i)".

* "t. M(i)." denotes "the true model is towards M(i)".

....

vector projection of u onto X_2 and its closest vector in X_1 be r_4. Notice that when X_1 and X_2 are both vectors, r_3 equals r_4.

Using these and equation (1) in the Preliminaries, we get

$$a0-b0 = u'P_2u - u'P_1P_2P_1u$$

$$\left[\frac{a0-b0}{u'P_1u}\right] = \left[\frac{u'u}{u'P_1u}\right] \cdot \left[\frac{u'P_2u}{u'u}\right] - \left[\frac{u'P_1P_2P_1u}{u'P_1u}\right]$$

$$= \cos^2(r_2)/\cos^2(r_1) - \cos^2(r_3)$$

$$al-bl = u'P_1u - u'P_2P_1P_2u$$

$$\left[\frac{al-bl}{u'P_2u}\right] = \left[\frac{u'u}{u'P_2u}\right]\left[\frac{u'P_1u}{u'u}\right] - \left[\frac{u'P_2P_1P_2u}{u'P_2u}\right]$$

$$= \cos^2(r_1)/\cos^2(r_2) - \cos^2(r_4)$$

We can now proceed to consider the three cases.

Case 1.

$$a0-b0 < 0 \implies \cos^2(r_2) < \cos^2(r_3).\cos^2(r_1) \tag{25}$$

$$al-bl > 0 \implies \cos^2(r_1) > \cos^2(r_4).\cos^2(r_2) \tag{26}$$

Equation (25) confirms that $r_1 < r_2$, while equation (26) does not tell us anything. (Note: $\cos^2(r_3)$ and $\cos^2(r_4)$ both have values between zero and one.) This shows that M(3) is closer to M(1) than to M(2) by our definition of direction.

Case 2.

$$a0-b0 > 0 \implies \cos^2(r_2) > \cos^2(r_3).\cos^2(r_1) \tag{27}$$

$$al-bl > 0 \implies \cos^2(r_1) > \cos^2(r_4).\cos^2(r_2) \tag{28}$$

Neither equation (27) nor equation(28) reveals the relative sizes of r_1 and r_2, and consequently, no indication is given about the direction of M(3) relative to that of M(1) and M(2).

Case 3.

$$a0-b0 > 0 \implies \cos^2(r_2) > \cos^2(r_3).\cos^2(r_1) \tag{29}$$

$$al-bl < 0 \implies \cos^2(r_1) < \cos^2(r_4).\cos^2(r_2) \tag{30}$$

Equation (29) says nothing about the relative sizes of r_1 and r_2, but equation (30) indicates that $r_1 < r_2$. This means that M(3) is closer to M(2) than to M(1).

All of the cases discussed above refer to cases where both models under test are rejected. By looking at the signs of the CTS under the two tests, in which the roles of the two models are reversed, the above analysis has shown that for cases 1 and 3 the signs of the CTS_1 and CTS_2 can indicate the direction of the true model relative to the models specified under the null and the alternative hypotheses.

3.5. Local alternatives.

In this section, a simpler method of deriving asymptotic properties, in the context of the CTS under local alternatives, is suggested. After that, the asymptotic properties of the ATS under local alternatives are derived.

Pesaran's (1982) construction of his local alternatives requires the two exogenous matrices, X_1 and X_2, to be related in the following manner:

$$X_2 = X_1 B + T^{-1/2}(t) + o(T^{-1/2}) \tag{31}$$

where, B and (t) are $K_1 \times K_2$ and $T \times K_2$ matrices of non-zero constants, and (t) is such that the term $((t)'M_1(t)/T)$ converges to a finite non-zero matrix as T tends to infinity. The existence of the limit of $((t)'M_1(t)/T)$ is assumed to ensure that the alternative hypothesis approaches the null at a slow enough rate to keep the power bound away from unity, while its non-zero assumption is needed to ensure that the power of the test is strictly larger than the size of type I error. When the limit of $((t)'M_1(t)/T)$ exists, K_1 needs to be at least as large as K_2, otherwise, C_{22} would possibly be singular.[32]

1. **Local alternatives for CTS.**

Pesaran (1982) has derived the CTS under local alternatives in a rather lengthy way. He first rewrites CTS in equation (8) as:

$$\begin{aligned}\text{CTS} &= -(T/2)\ln\{(w_1/\hat{s}_2^2).(w_2/w_1).(\hat{s}_{21}^2/w_3)\}\\ &= -(T/2)\ln(w_1/\hat{s}_2^2) - (T/2)\ln(w_2/w_1) - (T/2)\ln(\hat{s}_{21}^2/w_3),\end{aligned}$$

where w_1, w_2, and w_3 are some functions of y. He then tries to show that the first two terms in the last equation are of probability order $o_p(T^{-1/2})$, while the last term is of probability order $o_p(T^{-3/2})$. Thus asymptotically, this third term can be dropped, leaving the first two terms, which are then linearized to the first order. The resulting equation is the CTS under local alternatives, i.e.:

$$\text{CTS} = \{-T^{-1}b_2'(W)b_2 - T^{-1/2}b_2'(t)'M_1\epsilon_2\}/s_2^2 + o_p(T^{-1/2}). \qquad (32)$$

where $(W) = (t)'M_1(t)$. The first term is of (big) order unity, while the second term is of probability order $o_p(1)$.

A shorter and simpler method to derive the CTS under local alternatives is to first linearize the CTS to the first order, and then derive the asymptotic TS. In other words, one can use the linearized CTS of equation (15). On substituting equation (31) into it, and after some algebraic manipulations, equation (32) above can be obtained.

---

32 This is because

$$\begin{aligned}C_{22} &= \lim (X_2'X_2/T),\\ &= B'C_{11}B + (t)'(t)/T^2,\end{aligned}$$

and $r(C_{22}) < r(C_{11})$ if the second term above converges to zero. If $K_2 > K_1$, and if $(t)'(t)/T^2$ converges to a non-zero constant, then C_{22} is possibly non-singular. However if $K_2 < K_1$ instead, then C_{22} is non-singular.

If in the above derivation equation (14) is used instead, the resulting CTS under local alternatives will be one with s_2^2 in equation (32) being replaced by $(s_2^2 + 2b_2'(W)b_2/T)$. This is because $\hat{s}_{21}^2$ converges to $(s_2^2 + 2b_2'(W)b_2/T)$ under local alternatives. So, the TS is:

$$CTS = \frac{-T^{-1}b_2'(W)b_2 - T^{-1/2}b_2'(t)'M_1\epsilon_2}{s_2^2 + 2b_2'(W)b_2/T} + o_p(T^{-1/2}). \tag{33}$$

2. **Local alternatives for ATS.**

Since no one has derived the ATS under local alternatives, it will be derived here. Using (31) in (10) we can write ATS as:

$$ATS = \frac{-b_2'(W)b_2/T - b_2'(t)'M_1\epsilon_2/T}{s_2^2 + 2b_2'(W)b_2/T} + o_p(T^{-1/2}).$$

Comparing this with equation (33), it shows that the ATS is asymptotically equivalent to the CTS under local alternatives.

BIBLIOGRAPHY

Atkinson, A.C.: "A Method for discriminating Between Models." Journal of the Royal Statistical Society, B, 1970, pp.323-53.

Cox, D.R.: "Tests of Separate Families of Hypotheses." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol.I, Berkeley, University of California Press, 1961, pp.105-123.

__________: "Further Results on the Tests of Separate Families of Hypotheses." Royal Statistical Society, B, 1962, pp.406-424.

Dastoor, N.K. "Some Aspects of Testing Non-nested Hypotheses." Journal of Econometric, Vol.21, 1983, pp.213-218.

__________: "A Note on the Intrpretation of the Cox Procedure for Non-Nested Hypotheses." Economics Letters 8, 1981, pp.113-119.

Davidson, R. and MacKinnon, J.G.: "On a Simple Procedure for Testing Non-Nested Regression Models." Economics Letters 5, 1980, pp.45-48.

__________: Several Tests for Model Specification in the Presence of Alternative Hypotheses. Econometrica, Vol.49, no.3, May 1981, pp.781-793.

__________: "Some Non-Nested Hypothesis Tests and the Relations Among Them." Review of Economic Studies, XLIX, 1982, pp.551-565.

__________: "Testing Specification of Multivariate Models in the Presence of Alternative Hypotheses." Journal of Econometrics, Vol.23, 1983, pp.301-313.

Davies, R.B. "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative." Biometrika, 64, 2, 1977, pp.247-254.

Edwards, A.W.F. Likelihood. Cambridge University Press, 1972.

Fisher,G.R. "Test for Two Separate Regressions." Journal of Econometrics, Vol.21, 1983, pp.117-132.

__________: "A Guide to Testing Separate Regressions." Paper presented in Canadian Economics Association Meeting, Vancouver, B.C., June 3, 1983.

Fisher, G. R. and McAleer, M. "On the interpretation of the Cox test in econometrics." Economics Letters 4, 1979, pp.145-150.

__________: "Alternative Procedures and Associated Tests of Significance for Non-Nested Hypotheses." Journal of Econometrics, 16, 1981, pp.103-119.

__________: "Testing Separate Regression Models Subject to Specification Error." Economic Society of Australia and New Zealand, Tenth Conference of Economists. Australian National University, Canberra, ACT, 24-28, August, 1981.

__________: "On the Interpretation of Cox Test in Econometrics." Economics Letters 4, 1979, pp.145-150.

Fraser, D.A.S. Probability and Statistics: Theory and Applications. North Scituate, Massachusetts: Duxbury Press, 1976.

Godfrey, L.G. "Testing Non-Nested Models after Estimation by Instrumental Variables or Least Squares." Econometrica, Vol.51,No.2, March 1983,pp.355-365.

Godfrey, L.G. and Pesaran, M.H.: "Tests of Non-nested Regression Models:Small sample Adjustment and Monte Carlo Studies." Journal of Econometrics, Vol.21, 1983, pp.133-154.

Graybill, F.A. and Milliken, G.: "Quadratic Forms and Idempotent Matrices with Random Elements." The Annals of Mathematical Statistics, Vol.40, No.4, 1969, pp.1430-1438.

Jones, D.S.: Elementary Information Theory. Oxford: Clarendon Press, 1979.

Judge, G.G., Griffiths, W.E., Hill, R.C. and Lee, T.S.: The Theory and Practice of Econometrics. Toronto: John-Wiley and sons, 1980.

MacKinnon, J.G.: "Model Specification Tests Against Non-Nested Alternatives." Econometric Review, 2(1), 1983, pp.85-110.

MacKinnon, J.G., White, H., and Davidson, R.: "Tests of Model Specification in the Presence of Alternative Hypotheses." Journal of Econometrics, Vol.21, 1983, pp.53-70.

McAleer, M. "A Small Sample Test for Non-Nested Regression Models." Economics Letters, 7, 1981, pp.335-338.

MacKinnon, J.G.: "Specification Tests for Separate Models: A Survey." Faculty of Economics, Australian National University, 1982.

McAleer, M. and Bera, A.: "Model Specification Tests Against Non-Nested Alternatives: Comment." Working Paper, Australian National University, March 1982.

McAleer, M. and Fisher, G.: "Testing Separate Regression Models Subject to Specification Error." Journal of Econometrics, Vol.19, no.1, May 1982, pp.125-145.

McAleer, M. and Fisher, G. and Volker, P.: "Separate Misspecified Regressions and the U.S. Long-Run Demand for Money Function." Review of Economics and Statistics, Vol.64, pp.572-583, 1982.

Pereira, B. de B.: " A Note on the Consistency and on the Finite Sample Comparisons of Some Tests of Separate Families of Hypotheses." Biometrika, 64, 1, 1977, pp.109-113.

Pesaran, M.H.: "On the General Problem of Model Selection." Review of Economic Studies, pp.153-71, vol. 41, 1974.

__________: "Pitfalls of Testing Non-Nested Hypotheses by Lagrage Multiplier Method" Journal of Econometrics, Vol.17, 1981, pp.323-331.

__________: "On the Comprehensive Method of Testing Non-Nested Regression Models." Journal of Econometrics, Vol.18, 1982, pp.263-274.

__________: "Comparison of Local Power of Alternative Tests of Non-Nested Regression Models." Econometrica, Vol.50, no.5, Sept. 1982, pp.1287-1305.

Pesaran, M.H. and Deaton, A.S. "Testing Non-Nested Nonlinear Regression Models." Econometrica, Vol 46, no.3, May 1978, pp.677-694.

Sawyer, K.R.: "Testing Separate Families of Hypotheses: An Information Criterion." Journal of Royal Statistical Society, Series B, Vol.45, No.1, pp.89-99.

Walker, A.M.: "Some Tests of Separate Families of Hypotheses in Time Series Analysis." Biometrika, 54, 1 and 2, pp.39-68, 1967.

White, H.: "Regularity Conditions for Cox's Test of Non-Nested Hypotheses." Journal of Econometrics, Vol.19, 1982, pp.301-318.