1984

Analytic Studies In Econometric Inference With Small Samples

Lonnie John Magee

ANALYTIC STUDIES IN ECONOMETRIC INFERENCE WITH SMALL SAMPLES

by

Lonnie Magee

Department of Economics

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

June, 1984

ABSTRACT

In this thesis the analytic techniques of Edgeworth and Nagar's expansion and approximate slope are used to examine small sample properties of estimators and test statistics in some common econometric models. The chapters are largely self-contained.

Chapter one provides some definitions, hypothesis test construction methods, and a description of techniques to be used and models to be studied.

In chapter two, Edgeworth expansions are used to examine the properties of LR, W and LM tests for linear restrictions on regression parameters in the standard one-equation model with Student's t errors. Edgeworth size-correction factors are found to be more effective than degrees-of-freedom-based corrections.

Chapter three examines three aspects of the regression model with first order autoregressive errors. First, it is shown that the LM test statistic for the existence of this kind of autocorrelation is numerically insensitive to whether the error terms have normal or Student's t distributions. Second, Nagar's expansion techniques are used to compare the efficiency of various estimators of the regression coefficients (including

iterative). The results largely support the results of previous Monte Carlo studies. Finally, an Edgeworth expansion is used to provide a size correction factor for the Wald test for a zero coefficient restriction.

Chapter four deals with the test for existence of contemporaneous correlation between errors of different regression equations. This is a relevant pre-test for specification of SURE models. A variety of tests are presented, including one based on the Union-Intersection (UI) test construction principle. Relationships between the tests in some special cases are discussed, including a comparison of their approximate slopes. The UI test is exact, easy to use, but may have lower power.

Chapter five deals with the Cox and J tests for choosing between two non-nested single equation models. An Edgeworth expansion for the J test under both models is obtained, as well as a size correction factor. The approximate slopes of Cox and J tests are used to explore situations where the small sample properties of the two tests may differ substantially.

ACKNOWLEDGEMENTS

Aman Ullah provided a great deal of invaluable assistance during all stages of the preparation of this thesis: from the initial step of selecting a topic, to the learning of the required techniques, to the derivations and presentation of the results. I always knew that I could depend on his careful supervision to keep me on the right track.

Robin Carter provided many detailed critical comments on the written work in the later stages which greatly improved the end result. He also offered me some very helpful advice regarding the organization and presentation of the material.

Vicky Zinde Walsh offered a generous amount of time in going over the written drafts. I appreciated both her detailed comments and those of a more general nature.

Many members of the Economics Department at The University of Western Ontario also gave me advice and encouragement during my stay there. Fellow graduate students and friends were supportive. Their friendliness and tolerance helped me to keep the project in perspective and stay in the right frame of mind to enjoy the work. I owe a special thanks in this regard to Chris Towe, Jim Frazer, and Jessica Brennan.

Aloma Munian typed the manuscript. Her diligence, cooperation, patience and good humour made the final stages of this project much less painful than I had feared.

The Economics Department members at McMaster University were very accommodating. My employment there began before the thesis was complete, and they arranged my teaching schedule and other responsibilities with this in mind.

The Social Science and Humanities Research Council of Canada and the Government of Ontario provided direct financial assistance in addition to the Funding from the University of Western Ontario.

TABLE OF CONTENTS

I. INTRODUCTION

Asymptotic properties of statistics used in econometrics, both for parameter estimates and hypothesis test statistics, are generally well known. Less is known about their finite sample behaviour, not to mention the effects of relaxing the distributional assumptions. Since in many situations one must choose from a group of procedures having identical asymptotic properties, a good choice should take finite sample and robustness properties into account. This is especially true as the number of practical estimation methods grows due to improving computer performance, as well as a growing variety of hypothesis tests available due to an increased interest in test construction methods, particularly in the Wald (W), Likelihood Ratio (LR), and Lagrange Multiplier (LM) tests, which have identical asymptotic properties.

This thesis examines small sample properties of estimators and test statistics in some common econometric models. Some attention is also given to robustness and application of testing principles. In the rest of this chapter, definitions and desirable properties of estimators and test statistics are given. A discussion of methods of hypothesis test construction is next, followed by a description of ways of analyzing small sample properties.

Finally, the econometric models which will be the subject of the following chapters are introduced.

I.1 ESTIMATORS AND TEST STATISTICS: DEFINITIONS AND PROPERTIES

The emphasis in this section, as in the thesis, will be on hypothesis tests, although some discussion of parameter estimates is included in Chapter III. A hypothesis test can in fact be interpreted as an estimate of an unknown value of 0 or 1; 0, say, if the null hypothesis is true and 1 if it is false. A parameter estimate, on the other hand, could be seen as the result of a test involving many hypotheses, each one corresponding to a possible point estimate. A collection of articles criticizing the standard hypothesis testing procedure can be found in Morrison and Henkel (1970), while Leamer (1977) emphasizes the importance of the researcher's motivation for performing the test. A theme of this section is that many attempts to derive useful optimality properties for statistical tests have been made, but in practice these properties are either so strict that no test satisfies them, or not restrictive enough, so that more than one test satisfies them. This necessitates the case by case approach of the thesis.

It is assumed that the data, y, is generated by the following probability density function

$$f(y,\theta), \tag{1}$$

where θ is a $k \times 1$ vector of unknown parameters. In econometrics, this density is often conditional on some other data, X, i.e.,

$$f(y,\theta|X), \tag{2}$$

but X will be omitted for now for notational simplicity. The estimation problem involves the use of the data y to estimate the unknown vector θ, or some elements of θ. The hypothesis testing problem involves deciding whether or not θ falls into a pre-specified subset of the set of possible values. More formally, letting Ω be the set of admissible parameter values for θ, with $\Omega \subseteq \mathbb{R}^k$ we are testing H_0, the null hypothesis, against H_1, the alternative, where

$$\begin{aligned} &H_0 \text{ is } \theta \in \omega_0,\ \omega_0 \subset \Omega, \\ &\text{and } H_1 \text{ is } \theta \in \omega_1,\ \omega_1 \subseteq \Omega - \omega_0 \end{aligned} \tag{3}$$

The test can be defined by using the decision variable $\delta(y)$, where

$$\delta(y) = 0 \Longrightarrow \text{accept } H_0 \qquad (4)$$

$$= 1 \Longrightarrow \text{reject } H_0$$

The framework so far is very general, and could be used to describe both nested hypotheses (where H_0 involves a set of p restrictions ($p \leq k$) on θ, and $\omega_1 = \Omega - \omega_0$), and some non-nested hypotheses, where the null parameter space is not formed by placing restrictions on the alternative parameter space. In many econometric testing situations, including those discussed in this thesis, the density function will satisfy the assumptions of Wald (1943, pp. 428-9). These are standard, commonly made regularity assumptions and will not be given here.

Some definitions and properties follow.

(i) <u>Size</u> of a test: The maximum probability, over all θ, of rejecting H_0 when H_0 is true. For a test δ of size α, then,

$$\max_{\theta} \text{prob } \{\delta(y) = 1 | H_0\} = \alpha \qquad (5)$$

(ii) <u>Power</u> of a test: The probability of rejecting H_0 when H_0 is false.

Since the power will depend on what particular value θ takes on in the set ω_1 of alternatives, the power function of δ is

$$\beta(\theta^*, \delta) = \text{prob}\ \{\delta(y) = 1 | \theta = \theta^* \in \omega_1\} \qquad (6)$$

(iii) Uniformly most powerful (UMP) test. If, among all tests having a given size, there exists one test which has at least as high a power as all others over the entire alternative space ω_1, that test is UMP.

When H_0 and H_1 are simple (that is, there is only one member of ω_0 and of ω_1) a UMP test for any given size can be constructed using the Neyman-Pearson lemma. That is, for testing

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1 \qquad (7)$$

the UMP test of size α is given by

$$\delta(y) = 0 \text{ if } f(y,\theta_1) \leq c \cdot f(y,\theta_0) \qquad (8)$$

$$= 1 \text{ if } f(y,\theta_1) > c \cdot f(y,\theta_0)$$

where c is a constant such that $\text{prob}\{\delta(y) = 1 | \theta = \theta_0\} = \alpha$ There are some other situations for which UMP tests can be constructed; however, for many hypotheses, a UMP test does not exist. (See for example King (1980), Kariya and Eaton (1977)).

(iv) Consistency. An estimator is consistent if it approaches the true parameter value as the sample size T grows large; that is, $\operatorname{plim}_{T\to\infty} \hat{\theta} = \theta$ for a consistent estimator $\hat{\theta}$. Similarly, a test is consistent if its power approaches one against any fixed alternative as the sample size grows large.

(v) Unbiasedness. An estimator is unbiased if its expected value equals the true parameter value θ over all $\theta \varepsilon \Omega$. A test is unbiased if

$$\begin{aligned} &\operatorname{prob}\{\delta(y) = 0 \mid \theta = \theta_0 \varepsilon \omega_0\} \geq \\ &\operatorname{prob}\{\delta(y) = 0 \mid \theta = \theta_1 \varepsilon\ \omega_1\} \end{aligned} \tag{9}$$

for all $\theta_0 \varepsilon\ \omega_0,\ \theta_1 \varepsilon\ \omega_1$

An unbiased test, then, has a power against any alternative $\theta \varepsilon \omega_1$ greater than or equal to its size.

(vi) Invariance and similarity (of a test). It is often seen as desirable that a test be invariant to (i.e.. not depend on) certain transformations of the data. For example, tests concerning parameters of the linear regression model are often invariant to linear transformations of the independent variables and corresponding linear transformations of the regression coefficients. A test is similar when the test outcome is never affected by changes in the nuisance parameters in the model, i.e., the test is conditioned on sufficient statistics for the nuisance parameters. For a brief discussion as well as an example where nonsimilar tests are also considered, see Evans and Savin (1983).

(vii) Locally best unbiased tests. This is a less strict optimality property than UMP, and grew out of attempts by Neyman and Pearson (1936, 1938), Isaacson (1951) and others, to create more widely applicable properties than UMP. The general idea is to find an unbiased test whose power function is greater than that of all other unbiased tests having equal size, for parameters in the alternative space infinitely close to the null space ω_0 of (4). Application even of this less restrictive property has proven difficult, especially when more than one restriction is being tested in a multivariate problem.

(vii) Wald's asymptotic optimality properties. Wald (1943) defines three optimality properties and their asymptotic equivalents, and then defines a generally applicable testing procedure (the Wald (W) test discussed in I.2) for which he demonstrates, along with the likelihood ratio (LR) test, that these asymptotic properties are satisfied. A brief description of these three properties follows.

(viiia) Asymptotically best average power. Defining a family of surfaces in the parameter space by K(θ) and a weight function w(θ), a test has asymptotically best average power with respect to K(θ) and w(θ) if, as the sample size approaches infinity, the average power over any member of K(θ) with the average taken according to w(θ) is equal to or larger than any other test having equal size.

(viiib) Asymptotically best constant power. A test has asymptotically best constant power on the surfaces K(θ) if its power is constant over each member of K(θ) and that power is equal to or larger than the power of any other test having equal size and the constant power property over the same family K(θ), as the sample size approaches infinity.

(viiic) Asymptotically most stringent. A test is an asymptotically most stringent test if, as the sample size approaches infinity, it minimizes the maximum difference between its own power at θ and the power of the most powerful test of equal size for that particular θ, maximized over $\theta\varepsilon\omega_1$.

Wald is able to find a family of surfaces $K(\theta)$ and a weighting function $w(\theta)$ for which the W and LR tests have these three properties.

I.2 TEST CONSTRUCTION METHODS

Hypothesis tests need not be constructed using a formal method. One could construct a test based on intuitive grounds which performs acceptably. An advantage of using a method such as LR, W, or LM is that the resulting test will be known immediately to have certain asymptotic properties (see previous section). These tests also have attractive intuitive interpretations. In fact, many tests that were formed originally without the use of formal methods are LR, W, or LM tests, or are at least very similar. In this section the LR, W, and LM tests are described.

In addition, the union-intersection (UI) method, introduced by S.N. Roy (1953), is discussed. This is not a precise method of test construction, but is a framework for specifying a series of sub-hypotheses which together imply the null hypothesis. The null hypothesis is then accepted only if each of the sub-hypotheses are accepted.

The hypotheses to be considered here are parameter restrictions (that is, nested) in the form of p independent restrictions on the k parameters:

$$H_0 : g(\theta) = 0 \quad \text{vs.} \quad H_1 : g(\theta) \neq 0 \tag{10}$$

where $g(\theta)$ is a $p \times 1$ vector. The definitions will also use the likelihood function

$$L(\theta|y,X) = f(y|\theta, X) \tag{11}$$

where f is the density function as in (2), assumed to satisfy Wald's conditions mentioned earlier.

(i) <u>LR test</u>. The LR test determines if the reduction in maximum likelihood due to the restrictions imposed on θ is significantly large. The likelihood ratio is given by

$$\lambda = \max_{\theta \varepsilon \omega_0} L(\theta|y,X) / \max_{\theta \varepsilon \Omega} L(\theta|y,X)$$

$$= L(\hat{\theta}_R|y,X)/L(\hat{\theta}|y,X) \tag{12}$$

where ω_0 and Ω are defined in (4), and $\hat{\theta}$ and $\hat{\theta}_R$ are termed the unrestricted and restricted maximum likelihood (ML) estimators, respectively. The LR test statistic is then

$$LR = -2 \log \lambda \qquad (13)$$

Under H_0, $LR \sim \chi_p^2$ asymptotically (see Silvey (1970, pp. 113-14)). The likelihood ratio test is

$$\text{Accept } H_0 \text{ if } LR \leq c(p,\alpha) \qquad (14)$$

$$\text{Reject } H_0 \text{ otherwise}$$

where $\text{prob}\{\chi_p^2 > c(p,\alpha)\} = \alpha$, α is the pre-specified nominal size of the test, p is the number of restrictions and $c(p,\alpha)$ is the critical value.

The nominal size of the test generally does not equal the actual size since $LR \sim \chi_p^2$ asymptotically but not necessarily in small samples.

(ii). W test. The W test first proposed by Wald (1943) determines if the restriction vector $g(\theta)$ differs significantly from zero when evaluated at the unrestricted ML estimates, $\hat{\theta}$ of (12). It does not require calculation of the restricted ML estimates, $\hat{\theta}_R$. The test is based on the asymptotic result

$$\sqrt{T}(\hat{\theta} - \theta) \sim N(0, I_L^{-1}(\theta)) \quad \text{as } T \to \infty \qquad (15)$$

where T is sample size and

$$I_L(\theta) = \lim_{T\to\infty} I(\theta)/T, \tag{16}$$

$$I(\theta) = -E_y\{\partial^2 \log L(\theta|y;X)/\partial\theta\partial\theta'\} \tag{17}$$

i.e., $I(\theta)$ is the information matrix.

The W test statistic is

$$W = g(\hat{\theta})'[\hat{G}\, I(\hat{\theta})^{-1}\, \hat{G}']^{-1} g(\hat{\theta}) \tag{18}$$

where $\hat{G} = \partial g(\theta)/\partial\theta'|_{\theta=\hat{\theta}}$, a $p \times k$ matrix, and $I(\hat{\theta})$ is the negative of the expected value of the matrix of second derivatives of the log of the likelihood function evaluated at $\hat{\theta}$ with the expectation taken over y assuming its density to be $f(y,\hat{\theta}|X)$. Under H_0, $W \sim \chi^2_p$ asymptotically (see Silvey (1970, pg. 116)).

The W test of nominal size α is

$$\text{Accept } H_0 \text{ if } W \leq c(p,\alpha) \tag{19}$$

$$\text{Reject } H_0 \text{ otherwise}$$

where $c(p,\alpha)$ is defined as in (14).

(iii) <u>LM test</u>. The LM test uses the restricted ML estimates $\hat{\theta}_R$ but not $\hat{\theta}$, and has two interpretations. In its "score" form first proposed by Rao (1948) it determines if the first derivatives of the likelihood function at $\hat{\theta}_R$ are significantly different from zero. In its Lagrange multiplier form of Aitchison and Silvey (1958) and Silvey (1959), it determines if the Lagrange multipliers or "costs" of the restrictions are significantly different from zero. The LM statistic in its score form is

$$LM = \tilde{d}' I(\hat{\theta}_R)^{-1} \tilde{d} \qquad (20)$$

where $\tilde{d} = \partial \log L/\partial\theta|_{\theta=\hat{\theta}_R}$ is a $k \times 1$ vector. Under H_0, $LM \sim \chi^2_p$ asymptotically (see Silvey (1970, pp. 118-19)).

By writing the restricted ML procedure as a Lagrangian

$$h(\theta,\gamma|y,X) = \log L(\theta|y,X) + \gamma' g(\theta) \qquad (21)$$

where γ is a $p \times 1$ vector of Lagrange multipliers, it is seen that k of the first order conditions for maximization are

$$\tilde{d} + \tilde{G}'\tilde{\gamma} = 0 \qquad (22)$$

where $\tilde{G} = \partial g(\theta)/\partial\theta'|_{\theta=\hat{\theta}_R}$ is a $p \times k$ matrix, and $\tilde{\gamma}$ is a vector of estimated Lagrange multipliers. Using (22) the LM statistic of (20) can be written in its Lagrange multiplier form:

$$LM = \tilde{\gamma}' \tilde{G} I(\hat{\theta}_R)^{-1} \tilde{G}' \tilde{\gamma} \qquad (23)$$

The LM test of nominal size α is

$$\text{Accept } H_0 \text{ if } LM \leq c(p,\alpha) \qquad (24)$$

Reject H_0 otherwise.

(iv) Issues concerning LR, W, and LM tests. The above definitions reveal that the three tests examine the behaviour of the likelihood function in three different ways. An intuitive description of their differences in the general single parameter case is given by Buse (1982). In that paper (see also Engle (1982)) he shows that in the common case of linear parameter restrictions, the three tests are identical when the log-likelihood is quadratic in the parameters.

Computational considerations may help in choosing a test. The LR test requires both restricted and unrestricted ML estimation while W requires restricted only and LM requires unrestricted only. On the other hand, LR does not require the calculation of $I(\hat{\theta})$ or $I(\hat{\theta}_R)$, while W requires the former and LM the latter.

Another issue concerning $I(\theta)$ is the choice of its consistent estimate. Alternatives to $I(\hat{\theta})$ and $I(\hat{\theta}_R)$ in

W and LM include the negative Hessian of the log-likelihood

$$-\partial^2 \log L(\theta|y, X)/\partial\theta\partial\theta' \tag{25}$$

evaluated at either $\hat{\theta}$ or $\hat{\theta}_R$,

$$I^{*-1}(\hat{\theta}) = T^{-1} \{\lim_{T\to\infty} T^{-1} I(\hat{\theta})\}^{-1} \tag{26}$$

replacing $I^{-1}(\hat{\theta})$ in the Wald formula (18), or an equivalent for LM, and finally, $V^{-1}(\hat{\theta})$ and $V^{-1}(\hat{\theta}_R)$ which are inverses of some variance estimates and which may be particularly appropriate in non-normal cases where the variances of the estimates may exceed the lower bound, given by the inverse of the information matrix, by a substantial amount. Breusch and Pagan (1979), Buse (1982) and Efron and Hinckley (1978) have discussed the relative merits of (25) vs. $I(\hat{\theta})$ or $I(\hat{\theta}_R)$. In some cases (26) results in easier computations than $I(\hat{\theta})$ or $I(\hat{\theta}_R)$ since some off-diagonal terms in $I(\hat{\theta})$ have $O(T^{-1})$ and so will go to zero in the limit. Some other variations on these testing principles are outlined in Engle (1982, section 10), and Davidson and MacKinnon (1981).

The following asymptotic results may be used to examine individual restrictions and their effects on each parameter. Under H_0,

$$Tg(\hat{\theta}) \sim N(0, T\,\hat{G}\,I(\hat{\theta})^{-1}\hat{G}') \tag{27}$$

$$\tilde{d}/T \sim N(0,\ I(\hat{\theta}_R)/T) \qquad (28)$$

$$\tilde{\gamma}/T \overset{a}{\sim} N(0,(\tilde{G}I(\hat{\theta}_R)^{-1}\tilde{G}')^{-1}/T) \qquad (29)$$

all as $T \to \infty$. If the test rejects, then (27) can be used to examine the individual restrictions, and similarly with (29) for the LM test. The effects of the restrictions on each parameter can be studied by using (28). The use of formal hypothesis tests on individual parameters or restrictions after rejection of H_0 using these relations is not a good idea since the above distributions under H_0 <u>given rejection of H_0</u> will be quite different. Nevertheless, (27)-(29) may help in forming suspicions as to the cause of rejection of H_0.

There are not many general results about the powers of the three tests in small samples. Some results have been obtained by Lee (1971) and Peers (1971), for example, but the criteria for one test having greater power than another usually involve the parameters themselves, which are unknown.

The effects of distributional assumptions on the performance of these tests are also largely unknown. Savin (1982, p. 76) points out that when testing for restrictions on β in the non-linear regression model

$$y_t = f(x_t,\beta) + u_t, \qquad t = 1,\ldots,T \qquad (30)$$

normality of u_t is not required for LR, W, and LM tests based on the normality assumption to retain their asymptotic properties. This is not true in general. For example, an LM test for heteroscedasticity proposed by Breusch and Pagan (1979) in a regression model with a normality assumption has been shown to have incorrect asymptotic significance levels when the normality assumption is incorrect (see Koenker and Bassett (1982)).

The small sample properties of these tests generally are sensitive to the distributional assumption. Ullah and Zinde Walsh (1984) find that W and LM tests for linear restrictions on regression coefficients in the standard one-equation model assuming a t-distribution for the error terms differ numerically from the corresponding test for normally distributed errors cases. The LR test is the same in both cases.

(v) UI test. Consider again the hypotheses in (3)

$$H_0 : \theta \varepsilon \omega_0 \quad \text{vs.} \quad H_1 : \theta \varepsilon \omega_1 \tag{31}$$

The union-intersection (UI) test, first suggested by S.N. Roy (1953), tests (31) by testing a finite or infinite number of sub-hypotheses:

$$H_0 : \theta\varepsilon\omega_0 \quad \text{vs.} \quad H_{1i} : \theta\varepsilon\omega_{1i},\ \omega_{1i} \subset \omega_1 \tag{32}$$

A UI test for H_0 vs. H_1 of size α is formed by testing each sub-hypothesis with an LR test of size α_{SUB}. The test rejects H_0 if and only if H_0 is rejected in any of the tests of the sub-hypotheses, or sub-tests. This implies that the critical region, or set of observations which lead to rejection of H_0, equals the union of the critical regions of the sub-tests, and the acceptance region of the test equals the intersection of the acceptance regions of the sub-tests, hence the name "UI test". The size of the sub-tests, α_{SUB}, is determined by the size of the test, α.

There are several ways in which Roy's original idea could be generalized. Tests other than LR could be performed on the sub-hypotheses. Sub-tests of different sizes could be used in order to place more emphasis on certain alternatives.

A specific <u>infinite UI test</u> for the restrictions hypothesis (10) can be formed by defining the sub-hypotheses:

$$H_{0,a} : a'g(\theta) = 0 \quad \text{vs.} \quad H_{1,a} : a'g(\theta) \neq 0 \quad . \qquad (33)$$

where a is any non-null p-element vector with Euclidean length equal to one. This results in an infinite number of sub-hypotheses as long as $r > 1$. Each $H_{0,a}$ is less restrictive than H_0 since H_0 implies $H_{0,a}$ for each "a" while a single $H_{0,a}$ does not imply H_0. In some cases it is possible to define test statistics for the sub-hypotheses which have identical distributions, and hence critical values, under H_0. If it is a simple matter to find the maximum of these statistics over a, then it is possible to perform the infinite UI test, since this maximum test statistic compared with the critical value of the sub-tests will determine whether or not each of the sub-tests accepts H_0. The F-test for the linear hypothesis on the coefficients of a simple regression model is an infinite UI test (see Morrison (1976, pp. 176-7)). Another example is given in Chapter IV of this thesis for testing the independence of equations in a SURE model.

A finite UI test may be a good procedure when the alternative hypothesis has one or more of the following features: (a) only a finite number of the possible alternatives may have any useful economic interpretation (see Savin (1980, p. 271) for an example); (b) certain alternatives are seen as more likely than others a priori; (c) the consequences of wrongly rejecting certain alternatives are more serious than others. By specifying only a few alternatives in the sub-tests, the critical region can be altered so as to increase the power of the test against some alternatives and reduce it against others (for a given size).

I.3 METHODS OF ANALYZING SMALL SAMPLE PROPERTIES

In this section, methods of examining small sample properties of test statistics and estimators are discussed.

(i) Monte Carlo method. The small sample distribution of a test statistic or estimator given particular values for nuisance parameters and exogenous variables can be approximated by using a random number generator to simulate a large number of sets of data produced by a predetermined parameter set and exogenous data matrix. The researcher then acts as if the true hypothesis or parameter value is unknown and performs the hypothesis test or estimation on each of the artificially produced data sets. The cumulative probability distribution of the statistic can then be approximated by the observed cumulative frequency. In the case of a test statistic one may wish to record only the proportion of statistics less than one or two critical values.

An advantage of this approach is its simplicity, both for the researcher to perform and for the reader to interpret. Its accuracy is better known than that of other methods since the approximation is the result of a randomized experiment.

If one wants a very accurate approximation to the true density, particularly in the tails, the number of replications required grows so large that it may become prohibitively expensive. This is especially true if approximate distributions are desired over a wide variety of conditions (values of parameters of interest, nuisance parameters, and exogenous data). This last point reflects another problem with the Monte Carlo approach; each individual experiment reflects only a particular set of conditions. It is difficult to make general conclusions using this approach. If the test or estimator can be shown to have invariance properties with respect to nuisance parameters or even parameters of interest, the results can be generalized somewhat. (For some invariance properties concerning regression models, see Breusch (1980)).

(ii) Expansion methods. Several methods of approximating unknown distributions have been proposed (see Springer (1979) for a discussion). The Edgeworth expansion has attracted some interest in econometrics, both for estimators and test statistics. In the case of test statistics, local power properties can be examined by defining the true parameter so that it approaches the null at $0(T^{-1/2})$, where T is sample size. This enables the use of asymptotic properties since if a fixed alternative is considered, the test statistics of the thesis would each have $O_p(T)$ since they are consistent. The local alternative on the other hand yields a test statistic of $O_p(1)$ for these statistics, and so is of interest asymptotically since its power will not simply approach one.

Next, the statistic is represented as a series of terms having descending orders of probability with respect to T as defined by Mann and Wald (1943) (see the appendix). Usually, only the $O_p(1)$, $O_p(T^{-1/2})$, and $O_p(T^{-1})$ terms are retained. The moment generating function for this truncated statistic is then obtained, and the inversion theorem is used to get an approximation, to $O(T^{-1})$, of the cumulative density function of the statistic. In the case of LR, W and LM tests, this result takes the form of a weighted sum of non-central χ^2 c.d.f.'s (or central χ^2 if evaluated under H_0). Formulas for the expansion of the LR and W tests to $O_p(T^{-1/2})$ for any likelihood function satisfying certain regularity conditions are given by Hayakawa (1975), and for the LM test by Harris and Peers (1980). These results are not useful in many situations, however, since the $O(T^{-1/2})$ term in the final approximated distribution is often equal to zero. Extension of their results to $O(T^{-1})$ in the general case is difficult, so that these general formulas are not used in this thesis.

Sargan (1975) proves a theorem which establishes the validity of Edgeworth expansions for many tests and estimators met in econometrics. His result is extended to include non-normal disturbances and random exogenous variables by Phillips (1977) and Sargan (1976). Conditions which ensure the validity of this expansion for asymptotically χ^2 tests are given in Sargan (1980).

A criticism of this method, mentioned in Kendall and Stuart (1969), is that it tends to be inaccurate at the tails, sometimes giving negative densities. This is not a serious matter here, since we are most interested in the behaviour of test statistics over the middle range of their distributions. It is not as important a mistake, for example, to think that a test has a power of .99 when it actually is .98 than to think that it has a power of .6, when it actually is .4. In addition, the results of Evans and Savin (1982), in which size correction factors derived from Edgeworth expansions by Rothenberg (1977) are used to make the distributions of LR, W, and LM tests for linear restrictions on regression coefficients in the standard model closer to the assumed asymptotic distribution, are very encouraging. Other studies which have found the Edgeworth expansion method to work well include Fang and Krishnaiah (1982) and Tanaka (1983).

An advantage of this and other analytic approximation methods over Monte Carlo methods is the generality of the result. Any desired value for the parameters and exogenous data can be substituted into the resulting formula giving an approximate distribution. Another constructive result is the size correction factor mentioned above.

The derivation of these expansions, however, is often very tedious and the analytic results hard to interpret unless numerical examples are tabulated. The Edgeworth expansion method when applied to test statistics is limited to local alternatives, that is, alternatives which approach the null asymptotically (for example (42) of Chapter II and (175) of Chapter III), which may seem unattractive on intuitive grounds, yet seems to work very well as an approximation method.

Finally, the method is less accurate when the sample size is small when expanding in orders of sample size T as is done here, since the truncated terms from the expansion are more important in that case than in larger samples.

A simpler expansion method which can be used to evaluate the small sample performance of an estimator without considering its distribution is done by concentrating solely on its mean square error, or MSE. The MSE is such a widespread choice of criterion for estimator selection that a comparison of MSE's, or approximate MSE's, is often seen as satisfactory grounds for estimator choice. (Such is not the case, of course, for choice of test statistic.)

Asymptotic MSE approximations, first developed by Nagar (1959), follow roughly the same principle as Edgeworth expansions. The estimator is expanded by orders of probability to, say, $0_p(T^{-3/2})$. If the estimator $\hat{\theta}$ is unbiased, then $\hat{\theta} - \theta$ has an $0_p(1)$ term equal to zero. This means that terms of $0_p(T^{-1})$, $0_p(T^{-3/2})$ and $0_p(T^{-2})$ of $(\hat{\theta} - \theta)(\hat{\theta} - \theta)'$ are obtainable. For example, if

$$\hat{\theta} = a/b \tag{34}$$

where a and b are both random and have $0_p(1)$, and it is known that plim a = α and plim b = β such that $(a - \alpha)$ and $(b - \beta)$ each have $0_p(T^{-1/2})$, then we can write

$$\begin{aligned} \hat{\theta} &= \{\alpha + (a-\alpha)\}/\{\beta + (b-\beta)\} \\ &= \{\alpha + (a-\alpha)\}\{1-(b-\beta)/\beta + (b-\beta)^2/\beta^2\}/\beta + o_p(T^{-1}) \\ &= (\alpha/\beta) + \{a-\alpha-\alpha(b-\beta)/\beta\}/\beta + \{\alpha(b-\beta)^2/\beta^3 \\ &\quad - (a-\alpha)(b-\beta)/\beta^2\} + o_p(T^{-1}) \end{aligned} \tag{35}$$

where the terms in (35) have $0_p(1)$, $0_p(T^{-1/2})$, and $0_p(T^{-1})$

respectively, and the expansion could be continued to any desired order. If $\hat{\theta}$ is a consistent estimator of θ, then $\alpha/\beta = \theta$.

These expansion terms are typically products of quadratic forms consisting of the vector of disturbances multiplied into a constant matrix. If normality of the disturbances is assumed, then standard expectation formulas can be applied to get $E(\hat{\theta} - \theta)(\hat{\theta} - \theta)'$ to $0(T^{-2})$. Comparison of estimators is then a simple matter if θ is scalar, otherwise the trace, determinant, or some other function of this approximate MSE matrix can be used. For an application, see Chapter III. Sargan (1974) provides conditions for the validity of this approximation method.

(iii) Approximate slope. Another method of examining test statistics was first proposed by Bahadur (1960). For a fixed alternative, a given test statistic, and a given sample size, consider the smallest possible significance level (or size) corresponding to a fixed power against that alternative. Since the power of a consistent test against a fixed alternative approaches one as the sample grows large when its size is fixed, then the size of the test approaches zero when the power against that alternative is fixed. The approximate slope is defined as the limit of the ratio of minus twice the logarithm of this significance level to the sample size as the sample size approaches infinity when the power is held constant against a particular fixed alternative. Geweke (1981a) gives a more rigorous definition.

In that paper it is proven that if the asymptotic distribution under the null is a central χ^2, then its approximate slope equals the limit of the ratio of the test statistic itself to the sample size under a fixed alternative as the sample size (and hence the test statistic) approaches infinity. The approximate slope in this χ^2 case can then be roughly interpreted as the rate that the statistic approaches infinity with sample size. For applications, see theorems 4 and 9 of Chapter IV with discussion, and section V.4.

The main use of approximate slopes is in the comparison of tests having identical asymptotic distributions. If the ratio of their approximate slopes is substantially different than one, the tests probably have substantially different small sample distributions against that alternative. (See Geweke (1981b)).

This method, then, can be used to find situations worthy of further study by, say, Monte Carlo analysis, but its results should not be taken any further than that, as it has some undesirable properties. For example, Geweke (1981a, p. 1432) shows that "if one test with known asymptotic size exists then tests with the same asymptotic size but arbitrarily great approximate slope may be constructed".

This approach is most useful when the test statistics are hard to analyze analytically. Since approximate slopes are usually easily derived, it is a simple matter to pick out alternatives in a complex testing problem for which the tests may behave in a strange manner. Because of its non-local nature, this method can examine a wider range of alternatives than local approaches, particularly in non-nested hypothesis testing problems (see Chapter V).

I.4 ECONOMETRIC MODELS AND HYPOTHESES CONSIDERED IN THE THESIS

In this section the topics covered in the remainder of the thesis are outlined. The econometric models are introduced, the estimation and testing problems are described, and the results are briefly outlined.

(i) Standard Regression Model

Consider the one-equation regression model

$$y = X\beta + u \qquad (36)$$

where X is a known $T \times k$ matrix, β is an unknown $k \times 1$ vector, y is a known $T \times 1$ vector, u is a random $T \times 1$ vector assumed to follow the multivariate normal distribution, $u \sim N(0, \sigma^2 I_T)$, where σ^2 is unknown and u is assumed to be independent of X.

A commonly tested hypothesis in this model is

$$H_0 : R\beta = r \quad \text{vs.} \quad H_1 : R\beta \neq r \qquad (37)$$

often referred to as the linear hypothesis. This is perhaps the most thoroughly studied testing problem in econometrics. In Chapter II, the robustness of the LR, W, and LM tests to the normality assumption is examined by approximating their small sample distributions both under normality and student's t errors by an Edgeworth expansion.

(ii) Standard Regression Model with AR(1) Errors

Another model commonly used in econometrics is similar to (36) except the error terms are assumed to follow a first-order autoregressive process. The difference from (36), then, is that now

$$u \sim N(0, \sigma^2 \Sigma) \tag{38}$$

where

$$\Sigma = ((\sigma_{ij})), \quad \sigma_{ij} = \rho^{|i-j|}, \quad |\rho| < 1,$$

$$\rho \text{ unknown}, \quad i,j = 1,\ldots,T$$

An important hypothesis to be considered here is

$$H_0 : \rho = 0 \quad \text{vs.} \quad H_1 : \rho \neq 0, \tag{39}$$

and this hypothesis too has been the subject of much research (see King (1984) for a survey), with the Durbin-Watson test being the most popular. It is shown in Chapter III that the test based on the LM principle for (39) in this model is the same as for the case where the errors have a multivariate student's t distribution with the same covariance structure, and that this test is slightly different from the LM test as seen in the literature (e.g., in Engle (1982, section 8.2)).

Estimation methods are also an issue in this model since there is in general no best linear unbiased estimator for β because ρ is unknown. A brief discussion of some previous studies is followed by an evaluation of several methods using asymptotic MSE approximations.

Finally, a special case of the linear hypothesis,

$$H_0 : \beta_i = 0 \quad \text{vs.} \quad H_1 : \beta_i \neq 0 \tag{40}$$

where β_i is some element of β, is examined. The t-test, which is a Wald test in model (36), also has Wald test characteristics in this model for any consistent asymptotically normal estimators (see Stroud (1971)). In Chapter III, the small sample distribution of the Wald test using estimates from a Prais-Winsten procedure is derived by an Edgeworth expansion, and the size correction factor is derived.

(iii) Seemingly Unrelated Regression Equations (SURE) Model

Consider the SURE model associated with Zellner (1962)

$$\begin{aligned} y_1 &= X_1\beta_1 + u_{1.} \\ y_2 &= X_2\beta_2 + u_{2.} \\ &\vdots \\ y_N &= X_N\beta_N + u_{N.} \end{aligned} \tag{41}$$

where X_i is a known $T \times k_i$ matrix, β_i is an unknown $k_i \times 1$ vector, y_i is a known $T \times 1$ vector and $u_{i.}$ is an unknown $T \times 1$ vector with $u_{i.} \sim N(0, \omega_{ii} I_T)$, $i = 1, \ldots, N$.

In addition, it is assumed that the $N \times 1$ vector $u_{\cdot t} = (u_{1t}, u_{2t}, \ldots, u_{Nt})$ formed by taking the t^{th} element from each disturbance vector, has the following distribution

$$u_{\cdot t} \sim N(0, \Omega) \tag{42}$$

for each $t = 1, \ldots, T$. Ω is an unknown positive definite symmetric matrix with dimension $N \times N$.

The hypothesis considered in Chapter IV is that a particular $u_{i\cdot}$ is not correlated with any other $u_{j\cdot}$, $i \neq j$. Several tests are presented, including one which uses the infinite UI principle defined in section I.2. These tests are shown to be monotonic functions of a single statistic in the case where $X_i = X$, $i = 1, \ldots, N$, which corresponds to the reduced form system of a simultaneous equations model, for example. This means that these tests are equivalent when their exact distributions are used in this special case. In the case where the X_i's are not all the same, the approximate slopes of the tests are compared.

(iv) Tests for Non-Nested Regression Models

Suppose that there are two competing linear regression models

$$H_0 : y = X\beta_0 + u_0$$

$$\text{vs.} \qquad (43)$$

$$H_1 : y = Z\beta_1 + u_1$$

where X and Z are $T \times k_0$ and $T \times k_1$ known matrices, y is a $T \times 1$ known vector, β_0 and β_1 are unknown $k_0 \times 1$ and $k_1 \times 1$ vectors, and u_0 and u_1 are $T \times 1$ unknown vectors, with $u_0 \sim N(0,\sigma_0^2 I_T)$ and $u_1 \sim N(0,\sigma_1^2 I_T)$, σ_i^2 unknown; $i = 0,1$.

Suppose also that the hypothesis (43) is non-nested in the sense that no restrictions on β_0 or β_1 exist which would cause the two models to be identical. Several tests for this hypothesis have been proposed. One which seems to have good power properties but whose small sample distribution often differs from its asymptotic distribution under the null is the J-test proposed by Davidson and MacKinnon (1981b). In chapter V Edgeworth expansions for the distribution of J under both H_0 and H_1 are derived, along with the size correction factor, under the local alternative as defined in Pesaran (1982).

Edgeworth expansions for these non-nested hypotheses are not valid for all possible alternatives encountered in practice. (See the inequality imposed on the number of regressors in the competing models after equation (12) of chapter V.) These can be explored using a non-local method such as approximate slope. The J-test is compared in this way with the Cox test as modified for this application by Pesaran (1974).

APPENDIX I.

ORDERS IN PROBABILITY

The definitions for order in probability were formalized by Mann and Wald (1943). For the specific case of orders in probability in powers of sample size T for a random variable x(T) which is a function of the sample of size T, they become

(1) $x(T)$ has $o_p(T^i)$ if $\operatorname{plim} \{x(T)/T^i\} = 0$, and (44)

(2) $x(T)$ has $O_p(T^i)$ if $\Pr\{|x(T)/T^i| < A_\varepsilon\} = 1-\varepsilon$ for any arbitrarily small $\varepsilon > 0$ and corresponding finite A_ε.

In this thesis, i is some positive or negative multiple of 1/2.

The corresponding order definitions for non-stochastic functions of the sample data, say z(T), are:

(1) $z(T)$ has $o(T^i)$ if $\lim_{T\to\infty} \{z(T)/T^i\} = 0$, and (45)

(2) $z(T)$ has $O(T^i)$ if $\lim_{T\to\infty} \{z(T)/T^i\}$ is finite but not equal to zero for all possible sample values.

Sufficient conditions for orders in probability involving means and variances are very useful in practice and are used often in the thesis. These are:

(1) $x(T)$ has $o_p(T^i)$ if $\lim_{T\to\infty} E(x(T)/T^i)$

$$= \lim_{T\to\infty} \mathrm{Var}(x(T)/T^i) = 0 \qquad (46)$$

(2) $x(T)$ has $O_p(T^i)$ if $\lim_{T\to\infty} E(x(T)/T^i)$

and $\lim_{T\to\infty} \mathrm{Var}(x(T)/T^i)$

are both finite and at least one of these two limits does not equal zero for all possible sample values.

II. THE LINEAR HYPOTHESIS IN A SINGLE REGRESSION EQUATION

II.1 INTRODUCTION

The model considered in this chapter is the simple one equation regression model with disturbance terms that are uncorrelated and identically distributed across observations. The hypothesis to be tested is the validity of one or more linear restrictions on the unknown regression coefficients. This is a well studied problem, particularly in the common case where the distrubances are assumed to be normally distributed.

In the next section, the model and competing hypotheses are presented, along with the LR, W, and LM tests for the normal disturbances case as well as for the case where the disturbances have a Student's t distribution. In the following section the small sample distributions of the three tests under Student's t are approximated by Edgeworth expansion. The normal case, already examined by Rothenberg (1977), then falls out as a special case. These results are then discussed.

II.2 THE MODEL, THE LINEAR HYPOTHESIS, AND THE TEST STATISTICS

(i) The Model

Consider the one equation regression model,

$$y = X\beta + u \quad (1)$$

where y is a $T \times 1$ endogenous variable vector, X is a $T \times k$ matrix of exogenous variables, β is an unknown $k \times 1$ vector of regression coefficients, and u is an unknown $T \times 1$ disturbance vector. The distributional assumption concerning the disturbance vector u is of central interest in this chapter.

Assumption. The $T \times 1$ disturbance vector u is distributed as

$$u/\sigma \sim MSt_\gamma \quad (2)$$

where MSt_γ refers to a multivariate Student's t distribution[1] with γ degrees of freedom (d.f.), and σ is the dispersion parameter, with a spherical covariance matrix.

It will be useful to represent this multivariate Student's t vector u in the following way, as in Srivastava and Khatri (1979, p. 70, 2.21(i)):

$$u = \sigma a\, \gamma^{1/2}/q \qquad i = 1,\ldots,T \tag{3}$$

where

$$a \sim N(0,I) \tag{4}$$

and

$$q^2 \sim \chi^2_\gamma \tag{5}$$

where a and q^2 are independently distributed.

Some general remarks concerning this distribution follow:

1) The limiting distribution of u in (2) is

$$u \sim N(0, \sigma^2 I) \quad \text{as } \gamma \to \infty \tag{6}$$

One way of seeing this well known result is by noting that

$$1 - \gamma^{1/2}/q \quad \text{has} \quad 0_p(\gamma^{-1/2}) \tag{7}$$

using the order in probability notation of Mann and Wald (1943). This result (7) is discussed in more detail in section 3(i). Therefore, from (3) we see that

$$\operatorname*{plim}_{\gamma\to\infty} u = \sigma a \tag{8}$$

which along with (4) implies (6). This property allows us to obtain results for the more familiar normal distribution assumption by taking the limit of any result as γ approaches infinity.

2) It can be shown that the marginal distribution of each element u_i, $i = 1,\dots,T$ of u has a distribution

$$u_i/\sigma \sim t_\gamma \tag{9}$$

where t_γ is a univariate Student's t distribution with γ d.f.[2]. Although

$$E\, u_i u_j = 0 \quad \text{for all } i \neq j,\ i,j = 1,\dots,T \tag{10}$$

holds for all positive integer values of γ, the elements of u are not distributed independently of each other except for the special (normal) case where γ goes to infinity.

3) Since the variance of u_i in (9) does not equal σ^2, but rather

$$E(u_i^2) = \{\gamma/(\gamma-2)\}\sigma^2, \qquad \gamma > 2 \tag{11}$$

we refer to σ (or σ^2) as the dispersion parameter rather than standard deviation (or variance). Using the independence of a and q in (3), it follows that

$$E(u) = 0 \tag{12}$$

and

$$E(uu') = \sigma^2\{\gamma/(\gamma-2)\}I, \qquad \gamma > 2 \tag{13}$$

(ii) The Linear Hypothesis

The hypothesis of interest in this chapter is the familiar "linear hypothesis", a set of p independent linear restrictions on β of (1), where $p \leq k$. Stated formally, the competing hypotheses are

$$H_0 : R\beta = r \quad \text{vs.} \quad H_1 : R\beta \neq r \tag{14}$$

where R is a $p \times k$ matrix of constants and has rank p, and r is a fixed $p \times 1$ vector. It will be useful from a notational viewpoint to transform model (1) so that the restrictions can be expressed in a simpler form. Let

$$J = \begin{bmatrix} S \\ \hline R \end{bmatrix} \tag{15}$$

be a $k \times k$ matrix where R is from (14) and S is any $(k-p) \times k$ matrix such that J is non-singular (i.e., has full rank). Then define

$$\theta = J\beta \tag{16}$$

and

$$Z = XJ^{-1} \tag{17}$$

Then we can write model (1) as

$$y = Z\theta + u \tag{18}$$

and by partitioning θ as

$$\theta = [\theta_1' \mid \theta_2']' \tag{19}$$

where θ_1 is $(k - p) \times 1$ and θ_2 is $p \times 1$, and noting that

$$\theta_2 = R\beta, \tag{20}$$

we can rewrite hypotheses (14) in the simpler form

$$H_0 : \theta_2 = r \quad \text{vs.} \quad H_1 : \theta_2 \neq r \tag{21}$$

(iii) The Test Statistics: Normal Distrubances

The problem of testing (14) in model (1) or, equivalently, (21) in (18) under the normality assumption (taking the limit as γ approaches infinity in (2)) has been studied fairly extensively in the econometrics literature. The most commonly used test is the F-test where the test statistic is

$$F = \{y'(M_1-M)y/p\}/(y'My/(T-k)) \tag{22}$$

where

$$M = I - Z(Z'Z)^{-1}Z' \tag{23}$$

$$M_1 = I - Z_1(Z_1' Z_1)^{-1}Z_1' \qquad (24)$$

and we have partitioned Z so that

$$Z = [Z_1 \; Z_2] \qquad (25)$$

where Z_1 is $T \times (k - p)$ and Z_2 is $T \times p$.

Under the null hypothesis, F of (22) follows

$$F \sim F_{p,T-k} \qquad (26)$$

so that exact critical values can be selected from tables of the $F_{a,b}$ distribution.

The LR, W, and LM tests for (21) under the normality assumption (see e.g., Breusch (1979)) are

$$LR = T \log (y'M_1y/y'My) \qquad (27)$$

$$W = T(\hat{\theta}_2-r)'(Z^{22})^{-1}(\hat{\theta}_2-r)/y'My \qquad (28)$$

$$LM = T(\hat{\theta}_2-r)'(Z^{22})^{-1}(\hat{\theta}_2-r)/y'M_1Y \qquad (29)$$

where

$$\hat{\theta} = (Z'Z)^{-1}Z'y \qquad (30)$$

and is partitioned as

$$\hat{\theta} = [\,\hat{\theta}_1' \;\; \hat{\theta}_2'\,]' \qquad (31)$$

where $\hat{\theta}_1$ is $(k - p) \times 1$ and $\hat{\theta}_2$ is $p \times 1$, Z^{22} refers to the lower right hand $p \times p$ matrix of $(Z'Z)^{-1}$ which is partitioned as

$$(Z'Z)^{-1} = \begin{bmatrix} Z^{11} & Z^{12} \\ Z^{21} & Z^{22} \end{bmatrix}, \qquad (32)$$

so that Z^{11} is $(k - p) \times (k - p)$, Z^{12} is $(k - p) \times p$, and $Z^{21} = (Z^{12})'$, and M and M_1 are from (23) and (24), respectively.

These tests are all asymptotically distributed as χ^2_p under the null hypothesis. They have all been shown to be monotonic functions of the F statistic (22) by Vandaele (1981) and Fisher and McAleer (1980, pp. 7-9). These expressions are not given here, but are special cases of similar relationships (38), (39), and (40) given for the Student's t case. This property implies that when exact critical values are used, the F, LR, W, and LM tests will give the identical decision, i.e., they are equivalent. If asymptotic critical values are used for LR, W, and LM, however, they won't always give the same decision. In fact, an inequality exists between these tests in this case, given by

$$LM \leq LR \leq W \qquad (33)$$

with strict inequalities when the three test statistics do not equal zero. This result (33) has been shown in various ways by Savin (1976), Berndt and Savin (1977), Breusch (1979), and Vandaele (1981).

(iv) The Test Statistics: Student's t Disturbances

When assumption (2) is made, allowing for finite γ, the results of the previous section are slightly altered. It is shown by Ullah and Zinde-Walsh (1984) that the statistics of (27), (28), and (29) become

$$LR_t = LR \quad (34)$$

$$LM_t = \lambda^{-1} LM \quad (35)$$

$$W_t = \lambda W \quad (36)$$

where

$$\lambda = (T + \gamma)/(T + \gamma + 2) \quad (37)$$

In that paper it is also shown that these test statistics are monotonic functions of the F statistic (22), where

$$LR_t = T \log [1 + \{pF/(T-k)\}] \tag{38}$$

$$W_t = T \lambda p F/(T - k) \tag{39}$$

and

$$LM_t = T p F/\lambda(T - k + pF) \tag{40}$$

Hence, in this t disturbance case F, LR_t, W_t, and LM_t are still equivalent, in the sense that when exact critical values are used, they each give the same decision. It has been shown (see, e.g., Zellner (1976, pg. 401) and King (1980)) that the exact distributional result for F under the null given in (26) for the normal case also holds for the t-disturbance case, therefore the F test is still exact. Note, however, that the inequality result (33) does not carry over to the t-disturbance case (see Ullah and Zinde Walsh (1984)).

II.3 EDGEWORTH APPROXIMATIONS[3] TO THE DISTRIBUTIONS OF LR_t, W_t, and LM_t

(i) Large T and γ Expansions of LR_t, W_t, and LM_t

Large T and γ approximations to the distributions of LR_t, W_t, and LM_t, are derived here by expanding these statistics in orders of probability of T and γ. The following notation will be useful:

$0_p(i)$ refers to a term of $0_p(T^j \gamma^\ell (T + \gamma)^m)$

and $i = j + \ell + m$ (41)

The expansions will be undertaken under a local alternative, where the true parameters approach parameter values which satisfy the null hypothesis with $0(T^{-1/2})$. Specifically,

$$\theta = \begin{bmatrix} \theta_1 \\ r + \varepsilon/T^{1/2} \end{bmatrix} \tag{42}$$

where θ of (18) has been partitioned as in (19), and ε is a $p \times 1$ non-null preselected vector of constants. This sequence of values for θ ensures that the test statistics will not approach infinity as T grows large, while still allowing for the examination of distributions under an alternative using asymptotic properties. We can use the fact that LR_t, W_t, and LM_t have $0_p(0)$ under (42) to express the expansions of LR_t and LM_t to $0_p(-1)$ as functions of the expansion of W_t to $0_p(-1)$ using the relations (38), (39), and (40), where the order in probability notation is defined in (41).

<u>Lemma 1.</u> <u>If we denote the expansion of</u> W_t <u>of</u> (36) <u>under</u> (42) <u>in orders of probability of</u> (41) <u>by</u>

$$W_t = \eta_0 + \eta_{-1/2} + \eta_{-1} + o_p(-1) \tag{43}$$

<u>where</u> η_i <u>refers to a term having</u> $0_p(i)$, <u>then</u>

$$LR_t = \eta_0 + \eta_{-1/2} + (\eta_{-1} + 2\eta_0/(T + \gamma) - \eta_0^2/2T)$$

$$+ o_p(-1) \tag{44}$$

<u>and</u>

$$LM_t = \eta_0 + \eta_{-1/2} + (\eta_{-1} + 4\eta_0/(T + \gamma) - \eta_0^2/T)$$

$$+ o_p(-1) \tag{45}$$

<u>Proof.</u> Substituting for F from (39) into (38) and (40) and simplifying, we have

$$LR_t = T \log \{1 + (W_t/T\lambda)\} \tag{46}$$

and

$$LM_t = (T-k)W_t/\lambda[(T-k)\lambda + \{1-(k/T)\}W_t] \tag{47}$$

Noting from (37) that

$$\lambda = 1 - 2/(T + \gamma) + o(-1) \tag{48}$$

and

$$\lambda^{-1} = 1 + 2/(T + \gamma) \tag{49}$$

and using the following Taylor series expansion results:

$$\log (1 + \xi_{-1}) = \xi_{-1} - \xi_{-1}^2/2 + o_p(-2) \tag{50}$$

and

$$(1 - \xi_{-1})^{-1} = 1 + \xi_{-1} + o_p(-1) \tag{51}$$

where ξ_{-1} is any term having $0_p(-1)$, we have

$$LR_t = W_t + 2W_t/(T+\gamma) - W_t^2/2T + o_p(-1) \tag{52}$$

and

$$LM_t = W_t + 4W_t/(T+\gamma) - W_t^2/T + o_p(-1) \tag{53}$$

Substituting (43) in for W_t in (52) and (53) yields (44) and (45). Q.E.D.

Using lemma 1, an expansion of W_t will lead easily to expansions for LM_t and LR_t.

<u>Theorem 1.</u> <u>The large</u> T <u>and</u> γ <u>expansions of</u> W_t, LR_t <u>and</u> LM_t <u>of</u> (34), (35), <u>and</u> (36) <u>for testing the linear</u>

hypothesis (21) when the true parameter values are given by the local alternative (42) under the student's t distribution assumption (2) in orders of probability defined in (41) are given to 0(-1) by (43), (44), and (45), respectively, where

$$\eta_0 = \omega_1 - 2\omega_2 + \omega_3 \tag{54}$$

$$\eta_{-1/2} = \delta_* \eta_0 - \delta_q(\omega_1 - \omega_2) \tag{55}$$

$$\eta_{-1} = \{\delta_*^2 - 2(T + \gamma)^{-1}\}\eta_0 - \delta_q \delta_* (\omega_1 - \omega_2) + \delta_q^2 \omega_2 / 4 \tag{56}$$

and

$$\omega_1 = \varepsilon' (Z^{22})^{-1} \varepsilon / T\sigma^2 \tag{57}$$

which is a constant

$$\omega_2 = \varepsilon' B' Z' a / T^{1/2} \sigma \tag{58}$$

$$\omega_3 = a' Z B Z^{22} B' Z' a \tag{59}$$

$$B = \begin{bmatrix} Z_{11}^{-1} & Z_{12} \\ \hline -I_p & \end{bmatrix} \tag{60}$$

which is a $k \times p$ matrix

$$\delta_q \equiv 1 - q^2/\gamma \tag{61}$$

$$\delta_* = 1 - a'Ma/T \tag{62}$$

<u>with the distributions of the random variables given by</u>

$$a \sim N(0, I) \text{ as in (4),} \tag{63}$$

$$q^2 \sim \chi^2_\gamma \tag{64}$$

$$a'Ma = q_*^2 \sim \chi^2_{T-k} \tag{65}$$

<u>and</u> a, q^2, <u>and</u> q_*^2 <u>are distributed independently of each other</u>.

<u>Proof</u>. We need an expansion of W_t to $0_p(-1)$. The results for LR_t and LM_t will then follow from lemma 1.

From (36) and (48) we see that

$$W_t = W - 2(T + \gamma)^{-1}W + o_p(-1) \tag{66}$$

so an expansion of W to 0(-1) will give the result.

From (18) and (30),

$$\hat{\theta} = \theta + (Z'Z)^{-1}Z'u. \tag{67}$$

Therefore

$$\hat{\theta}_2 = \theta_2 + Z^{21} Z_1' u + Z^{22} Z_2' u \tag{68}$$

using the partitionings of Z in (25) and $(Z'Z)^{-1}$ in (32), and so under alternative (42) we have

$$\hat{\theta}_2 - r = \varepsilon/T^{1/2} + (Z^{21} Z_1' + Z^{22} Z_2')\ u \tag{69}$$

By partitioning $Z'Z$ in the same way as $(Z'Z)^{-1}$, so that

$$Z'Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \tag{70}$$

and using the partitioned inverse formula (Theil (1971, p. 18)) to show

$$Z^{21} = -Z^{22} Z_{21} Z_{11}^{-1} \tag{71}$$

we have

$$\hat{\theta}_2 - r = \varepsilon/T^{1/2} - Z^{22} B' Z' u \tag{72}$$

where B is defined in (60).

Let

$$\delta_\sigma = 1 - u'Mu/T\sigma^2, \tag{73}$$

which has $0(T^{-1/2})$ so that we can write

$$(y'My/T)^{-1} = (u'Mu/T)^{-1}$$

$$= \{\sigma^2(1 - \delta_\sigma)\}^{-1}$$

$$= (1 + \delta_\sigma + \delta_\sigma^2)/\sigma^2 + o_p(-1) \qquad (74)$$

Substituting (72) and (74) into (28), we have

$$W = \{(\varepsilon/T^{1/2} - Z^{22}B'Z'u)'(Z^{22})^{-1}(\varepsilon/T^{1/2}$$

$$- Z^{22}B'Z'u)\}(1 + \delta_\sigma + \delta_\sigma^2)/\sigma^2$$

$$+ o_p(-1) \qquad (75)$$

Using (3) we have

$$W = \{\varepsilon'(Z^{22})^{-1}\varepsilon/T\sigma^2 - 2\gamma^{1/2}\varepsilon'B'Z'a/T^{1/2}q\sigma$$

$$+ \gamma a'Z\,BZ^{22}B'Z'a/q^2\}(1 + \delta_\sigma + \delta_\sigma^2)$$

$$+ o_p(-1) \qquad (76)$$

Now consider δ_q of (61), which has $0_p(\gamma^{-1/2})$ since

$$E(q^2) = \gamma \text{ and } Var\,(q^2) = 2\gamma \qquad (77)$$

so that

$$E.\,\delta_q = 0 \qquad \text{and} \qquad Var\,\delta_q = 2/\gamma \qquad (78)$$

We can then expand γ/q^2 in powers of $(q^2/\gamma)-1$ by a Taylor series expansion:

$$\gamma/q^2 = 1 + \delta_q + \delta_q^2 + o_p(-1) \qquad (79)$$

Since q follows a chi distribution which is not as simple to work with as a chi-square, we make the following conversion. From (61) we have

$$\delta_q = \{1-(q/\gamma^{1/2})\}\ \{1 + (q/\gamma^{1/2})\}$$

which yields

$$1-(q/\gamma^{1/2}) = \delta_q/\{1 + (q/\gamma^{1/2})\}$$

$$= \delta_q/2[1-\{1-(q/\gamma^{1/2})\}/2]$$

$$= \delta_q/2 + \delta_q^2/8 + o_p(-1) \qquad (80)$$

by a Taylor series expansion about $\delta_q/2$. Therefore,

$$\gamma^{1/2}/q = 1 + \{1-(q/\gamma^{1/2})\} + \{1-(q/\gamma^{1/2})\}^2$$

$$+ o_p(-1) = 1 + \delta\ /2 + 3\delta_q^2/8$$

$$+ o_p(-1) \qquad (81)$$

Also, by substituting (3) and (79) in (73), we have

$$\delta_\sigma = 1 - \sigma^2\gamma\, a'Ma/Tq^2\sigma^2$$

$$= 1 - \gamma a'Ma/Tq^2$$

$$= 1 - (a'Ma/T)(1 + \delta_q + \delta_q^2) + o_p(-1) \qquad (82)$$

We can use the well known result that $a'Ma$ is distributed independently of $Z'a$ and that

$$a'Ma \sim \chi^2_{T-k} \qquad (83)$$

to define

$$q_*^2 = a'Ma \qquad (84)$$

and

$$\delta_* = 1 - q_*^2/T \qquad (85)$$

where δ_* has $O_p(-1/2)$ and is distributed independently of δ_q and $Z'a$.

Using (82), (84), and (85), we can show that

$$\delta_\sigma = (\delta_* - \delta_q) + \delta_q(\delta_* - \delta_q) + o_p(-1) \qquad (86)$$

Substituting (79), (81), and (86) in (76) and then substituting (76) into (66) and simplifying, we have

$$W_t = \eta_0 + \eta_{-1/2} + \eta_{-1} + o_p(-1), \qquad (87)$$

where η_0, $\eta_{-1/2}$ and η_{-1} are as in (54), (55), and (56), respectively. Now, using results (35) and (36) we can easily obtain similar expansions for LR_t and LM_t.

Q.E.D.

(ii) Edgeworth Expansions

Given the approximations of LR_t, W_t, and LM_t to $O(-1)$ of Theorem 1, we can derive Edgeworth approximations to their distributions. In this section the results are stated in theorem 2, followed by their derivation.

Theorem 2. The Edgeworth expansions for the distributions of W_t, LR_t, and LM_t of (34), (35), and (36) for the hypothesis, parameter values, distributional assumption, and orders given by theorem 1, are given by

$$
\begin{aligned}
pr(W_t \leq x) = {} & pr(\chi_p^2(\omega) \leq x) \\
& + T^{-1} \sum_{i=0}^{4} \tau_{Wi} pr(\chi_{p+2i}^2(\omega) \leq x) \\
& + \gamma^{-1} \sum_{i=0}^{2} \phi_i pr(\chi_{p+2i}^2(\omega) \leq x) \\
& + (T+\gamma)^{-1} \sum_{i=0}^{2} \alpha_{Wi} pr(\chi_{p+2i}^2(\omega) \leq x) \\
& + o(-1) \qquad (88)
\end{aligned}
$$

$$\mathrm{pr}(LR_t \leq x) = \mathrm{pr}(\chi_p^2(\omega) \leq x) + T^{-1} \sum_{i=0}^{3} \tau_{LRi}\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x)$$

$$+ \gamma^{-1} \sum_{i=0}^{2} \phi_i\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x) + o(-1) \qquad (89)$$

$$\mathrm{pr}(LM_t \leq x) = \mathrm{pr}(\chi_p^2(\omega) \leq x) + T^{-1} \sum_{i=0}^{4} \tau_{LMi}\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x)$$

$$+ \gamma^{-1} \sum_{i=0}^{2} \phi_i\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x)$$

$$+ (T + \gamma)^{-1} \sum_{i=0}^{2} \alpha_{LMi}\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x)$$

$$+ o(-1) \qquad (90)$$

where

$\chi_p^2(\omega)$ refers to a non-central χ^2 variable with p d.f. and non-centrality parameter ω,

$$\omega = \omega_1 \qquad (91)$$

where ω_1 is given by (57),

$$\tau_{W0} = -p(2k + 2 - p)/4\,, \qquad \tau_{W1} = p(k-p)/2 - \omega(k-p)/2$$

$$\tau_{W2} = p(p + 2)/4 + \omega(2k + \omega - 4p - 4)/4,$$

$$\tau_{W3} = \omega(p+2-\omega)/2, \qquad \tau_{W4} = \omega^2/4 \qquad (92)$$

$$\tau_{LR0} = -p(2k + 2-p)/4, \qquad \tau_{LR1} = p(2k+2-p)/4 - \omega(k-p)/2$$

$$\tau_{LR2} = \omega(2k + \omega - 2p)/4, \qquad \tau_{LR3} = -\omega^2/4 \tag{93}$$

$$\tau_{LM0} = -p(2k + 2-p)/4, \qquad \tau_{LM1} = p(k+2)/2 - \omega(k-p)/2$$

$$\tau_{LM2} = -p(p+2)/4 + \omega(2k + \omega + 4)/4,$$

$$\tau_{LM3} = -\omega(p + 2)/2, \qquad \tau_{LM4} = -\omega^2/4 \tag{94}$$

$$\phi_0 = \omega^2/4, \quad \phi_1 = -\omega^2/2, \quad \phi_2 = \omega^2/4, \tag{95}$$

$$\alpha_{W0} = p, \quad \alpha_{W1} = \omega - p, \quad \alpha_{W2} = -\omega \tag{96}$$

$$\alpha_{LM0} = -p, \quad \alpha_{LM1} = p - \omega, \quad \alpha_{LM2} = \omega \tag{97}$$

<u>where</u> p <u>is the number of linear restrictions, and</u> k <u>is the number of exogenous variables</u>.

<u>Note</u>: These results can be put back into the original notation of model (1) and the restriction (14) by noting from (17) that

$$\begin{aligned}(Z'Z)^{-1} &= \{(J^{-1})' X'X(J)^{-1}\}^{-1} \\ &= J(X'X)^{-1}J'\end{aligned} \tag{98}$$

therefore, the partitionings of (15) and (32) give

$$Z^{22} = R(X'X)^{-1}R' \qquad (99)$$

and so using (57) we get

$$\omega = \varepsilon'\{R(X'X)^{-1}R'\}^{-1}\varepsilon/T\sigma^2 \qquad (100)$$

Proof. The expansion is derived by obtaining the moment generating function (m.g.f.) to O(-1) and then inverting. Beginning with W_t, the m.g.f. is given by

$$\begin{aligned} M_W(t) &= E\{\exp(t\, W_t)\} \\ &= E[\exp\{t(\eta_0 + \eta_{-1/2} + \eta_{-1})\}] + o(-1) \\ &= E[\{\exp(t\,\eta_0)\}(1 + t\eta_{-1/2} + t\eta_{-1} \\ &\quad + t^2\eta^2_{-1/2}/2)] + o(-1) \end{aligned} \qquad (101)$$

where we have used a Taylor series expansion about $\exp(t\,\eta_0)$ to simplify the calculation, and t is used here as the m.g.f. argument.

It will be useful to evaluate $E \exp(t\,\eta_0)$ first. This will involve a transformation of the random vector $Z'a$ which will greatly simplify the integration of the remaining terms. From (54), (57), (58), and (59), and letting

$$b = Z' a \tag{102}$$

which implies that b, a $k \times 1$ vector, satisfies

$$b \sim N(0,\ Z' Z) \tag{103}$$

due to (63), we can write

$$\eta_0 = \varepsilon' (Z^{22})^{-1} \varepsilon/T\sigma^2 - 2\varepsilon' B' b/T^{1/2}\sigma$$

$$+ b' B Z^{22} B' b \tag{104}$$

Noting that η_0 is not a function of s^2 or s^{*2}, we have

$$E \exp(t\ \eta_0) = \int_b \exp(t\ \eta_0) f_b(b)\, db \tag{105}$$

where $f_b(b)$ is the normal density function for b implied by (103), with η_0 as in (104).

By completing the square in the exponent term of (105) and making the transformation

$$z = D^{-1/2} b + 2t\ D^{1/2} B\varepsilon/T^{1/2}\sigma \tag{106}$$

where

$$D^{-1} = (Z' Z)^{-1} - 2t\ BZ^{22}B' \tag{107}$$

and

$$D = (1-2t)^{-1}[Z' Z - 2t\ Z' ZAZ' Z] \tag{108}$$

where

$$A = \begin{bmatrix} Z_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \text{ a } k \times k \text{ matrix,} \quad (109)$$

the integral (105) becomes

$$E \exp(t\,\eta_0) = \exp\{t\omega + 2t^2 \varepsilon' B' DB\varepsilon / T\sigma^2\} |D|^{1/2} |Z'Z|^{-1/2} \int_z f_z(z)\,dz \quad (110)$$

where $|D|^{1/2}$ is the Jacobean of the transformation, and

$$f_z(z) = (\pi\sigma^2)^{-k/2} \exp(-z'z/2\sigma^2) \quad (111)$$

This is the p.d.f. of a set of independent normal variables, each with mean zero and variance σ^2, which enables us to treat z in the integrations _as if_ $z \sim N(0,\sigma^2 I)$.

Using

$$B'Z'ZB = (Z^{22})^{-1} \text{ and } B'Z'ZA = 0 \quad (112)$$

we have

$$\varepsilon' B' DB\varepsilon = (1 - 2t)^{-1}\, \varepsilon' (Z^{22})^{-1} \varepsilon \quad (113)$$

Also,

$$|D| = (1 - 2t)^{-k}\, |Z'Z| \, |I - 2t\, AZ'Z| \quad (114)$$

Since $AZ'Z$ is upper triangular, we see that

$$|D| = (1-2t)^{-k}|Z'Z|(1-2t)^{(k-p)}$$

$$= |Z'Z|(1-2t)^{-p} \qquad (115)$$

Now, substituting (113) and (115) into (110) and simplifying, using the fact that $f_z(z)$ of (111) can be interpreted as a p.d.f., we have

$$E\exp(t\,\eta_0) = \exp\{t(1-2t)^{-1}\omega\}(1-2t)^{-p/2}\int_z f_z(z)\,dz$$

$$= \exp\{t(1-2t)^{-1}\omega\}(1-2t)^{-p/2} \qquad (116)$$

which is the m.g.f. for a non-central χ^2_p with non-centrality parameter ω given in (91)[4,5].

For evaluation of the higher order terms in (101), note from (116) that, for example,

$$E\{\exp(t\,\eta_0)\}\eta_{-1/2} = E\exp(t\,\eta_0)\int_{q^2}\int_{q_*^2}\int_z \eta_{-1/2} f_z(z) f_{q_*^2}(q_*^2) f_{q^2}(q^2)\,dz\,dq_*^2\,dq^2 \qquad (117)$$

where $f_{q^2}(q^2)$ and $f_{q_*^2}(q_*^2)$ are the χ^2 p.d.f.'s for q^2 and q_*^2 implied by (64) and (65). All that is required for these higher order terms, then, is the triple integral in (117), which can be interpreted as an expectation taken over z (as if $z \sim N(0,1)$), q^2, and q_*^2.

For this "expectation" of $\eta_{-1/2}$ of (55), we note that η_0 and ω_2 are functions of Z'a (hence of z) only, not of q^2 or q_*^2, and we can take expectations over the three stochastic terms separately due to their independence. From (102) and (106) we have

$$Z'a = b = D^{1/2}z - 2t\ DB\varepsilon/T^{1/2}\sigma \tag{118}$$

Expressing (54) as a quadratic form, substituting (118) and simplifying using (112), we get

$$\begin{aligned}\eta_0 &= (Z^{22}B'b-\varepsilon/T^{1/2}\sigma)'(Z^{22})^{-1}(Z^{22}B'b-\varepsilon/T^{1/2}\sigma)\\ &= c'(Z^{22})^{-1}c\end{aligned} \tag{119}$$

where

$$c = Z^{22}B'D^{1/2}z - (1-2t)^{-1}\varepsilon/T^{1/2}\sigma \tag{120}$$

Denoting "expectation" in the sense of the discussion following (117) by $\tilde{E}$, i.e.,[6]

$$\tilde{E}\ g(z,q^2,q_*^2) = \int_z \int_{q^2} \int_{q_*^2} g(z,q^2,q_*^2) f_z(z) f_{q^2}(q^2) f_{q_*^2}(q_*^2)\ dq_*^2\ dq^2\ dz \tag{121}$$

for some function g, we have

$$\tilde{E}\ \eta_0 = \text{tr } D^{1/2} B Z^{22} B' D^{1/2} + (1-2t)^{-2}\omega \tag{122}$$

where we have used

$$E\ v' Q v = \text{tr } Q \tag{123}$$

when

$$v \sim N(0,I) \tag{124}$$

Since using (112) we have

$$B' D\ B Z^{22} = (1-2t)^{-1} I_p, \tag{125}$$

equation (122) becomes

$$\tilde{E}\ \eta_0 = (1-2t)^{-1} p + (1-2t)^{-2}\omega \tag{126}$$

Similarly, we have

$$\tilde{E}\ \omega_2 = -2t(1-2t)^{-1}\omega \tag{127}$$

Using (77) and (61), we have

$$\tilde{E}\ \delta_q = 0, \tag{128}$$

and since, from (62) and (65),

$$\tilde{E}\ \delta_* = k/T, \tag{129}$$

we can apply (126), (128) and (129) to (55) to give

$$\tilde{E}\ \eta_{-1/2} := (1-2t)^{-1}kp/T + (1-2t)^{-2}k\omega/T \tag{130}$$

To complete our calculation of $M_W(t)$ in (101), we still need $\tilde{E}\ \eta_{-1}$ and $\tilde{E}\ \eta_{-1/2}$.

For $\tilde{E}\ \eta_{-1}$, using the fact that for any $x \sim \chi^2_n$,

$$E(x^2) = n^2 + 2n, \tag{131}$$

we can show that from (62) and (65),

$$\tilde{E}\ \delta_*^2 = 2/T + o(-1) \tag{132}$$

and from (61) and (64),

$$\tilde{E}\ \delta_q^2 = 2/\gamma \tag{133}$$

We can use (126) to (129), (132) and (133) along with (56) to show that

$$\tilde{E}\ \eta_{-1} = 2(T^{-1} - (T+\gamma)^{-1})\{(1-2t)^{-1}p + (1-2t)^{-2}\omega\} - t(1-2t)^{-1}\omega/\gamma \tag{134}$$

For $\tilde{E}\ \eta^2_{-1/2}$ we need $\tilde{E}\ \eta_0^2$ and $\tilde{E}(\omega_1-\omega_2)^2$. Note from (119) that

$$n_0^2 = c'(Z^{22})^{-1} c \cdot c'(Z^{22})^{-1} c, \tag{135}$$

with c as in (120). To evaluate the expectation of this, we can use the result

$$E(v'Qv)^2 = (\text{tr } Q)^2 + 2\text{tr } Q^2 \tag{136}$$

when $v \sim N(0,I)$ and Q is symmetric, along with (125) to show that

$$\tilde{E}(z'D^{1/2}BZ^{22}B'D^{1/2}z)^2 = (1-2t)^{-2}p^2 + 2(1-2t)^{-2}p \tag{137}$$

Using (137) along with (122) and noting that "expectations" of odd powers of z are zero, we can show that

$$\tilde{E}\, n_0^2 = (1-2t)^{-2}p(p+2) + 2(1-2t)^{-3}\,\omega(p+2) + (1-2t)^{-4}\,\omega^2 \tag{138}$$

Similarly, we have

$$\tilde{E}(\omega_1-\omega_2)^2 = (1-2t)^{-2}\omega^2 + (1-2t)^{-1}\omega \tag{139}$$

Using (128), (132), (133), (138), and (139) along with (55), we get

$$\tilde{E}\,\eta^2_{-1/2} = 2T^{-1}\{(1-2t)^{-2}p(p+2) + 2(1-2t)^{-3}\omega(p+2)$$

$$+ (1-2t)^{-4}\omega^2\} + 2\gamma^{-1}\{(1-2t)^{-2}\omega^2$$

$$+ (1-2t)^{-1}\omega\} + o(-1) \qquad (140)$$

From (101), using (130), (134), and (140), and simplifying, we see that

$$M_W(t) = \{E\exp(t\eta_0)\}(1 + t\tilde{E}\,\eta_{-1/2} + t\tilde{E}\,\eta_{-1}$$

$$+ t^2\tilde{E}\,\eta^2_{-1/2}/2) + o(-1)$$

$$= \exp\{t(1-2t)^{-1}\omega\}(1-2t)^{-p/2}\{1+T^{-1}\sum_{i=0}^{4}$$

$$\tau_{wi}(1-2t)^{-i} + \gamma^{-1}\sum_{i=0}^{2}\phi_i(1-2t)^{-i}$$

$$+ (T+\gamma)^{-1}\sum_{i=0}^{2}\alpha_{wi}(1-2t)^{-i}\}$$

$$+ o(-1) \qquad (141)$$

where the τ_{wi}'s are given in (92), the ϕ_i's in (95), and the α_{wi}'s in (96).

From (44), we observe that the m.g.f. for LR_t

is

$$M_{LR}(t) = M_w(t) + \{E \exp(t\eta_0)\} \quad (2t\,\tilde{E}\eta_0/(T+\eta)$$

$$-t\tilde{E}\,\eta_0^2/2T) + o(-1) \qquad (142)$$

Using (126), (138), and (141), we have

$$M_{LR}(t) = \exp\{t(1-2t)^{-1}\omega\}(1-2t)^{-p/2}\{1$$

$$+ T^{-1}\sum_{i=0}^{3}\tau_{LRi}(1-2t)^{-i}$$

$$+ \gamma^{-1}\sum_{i=0}^{2}\phi_i(1-2t)^{-i}\} + o(-1) \qquad (143)$$

where the ϕ_i's are given in (95) and the τ_{LRi}'s in (93).

For the m.g.f. of LM_t, $M_{LM}(t)$, we use (45) to show that

$$M_{LM}(t) = M_w(t) + \{E \exp(t\eta_0)\}(4t\tilde{E}\eta_0/(T+\gamma)$$

$$- t\tilde{E}\,\eta_0^2/T) + o(-1) \qquad (145)$$

This, along with (126), (138) and (141) yields

$$M_{LM}(t) = \exp\{t(1-2t)^{-1}\omega\}(1-2t)^{-p/2}\{1+T^{-1}\sum_{i=0}^{4}\tau_{LMi}(1-2t)^{-i}$$

$$+ \gamma^{-1}\sum_{i=0}^{2}\phi_i(1-2t)^{-i} + (T+\gamma)^{-1}\sum_{i=0}^{2}\alpha_{LMi}(1-2t)^{-i}$$

$$+ o(-1) \qquad (146)$$

where the τ_{LMi}'s are given in (94), the ϕ_i's in (95), and the α_{LMi}'s in (97).

Given the m.g.f.'s of (141), (143), and (146), and noting the non-central χ^2 m.g.f. result of (116) and footnote 4, we can apply the Inversion theorem (see Kendall and Stuart (1969, pp. 94-95)), to obtain the approximate cumulative density functions (c.d.f.'s) of W_t, LR_t, and LM_t, which are given in (88), (89), and (90). This is a simple matter since the Inversion theorem involves integrating the m.g.f. over t to obtain the c.d.f., so that this can be done term by term.

Q.E.D.

The Edgeworth approximations (88), (89), and (90), are expressed as a weighted sum of c.d.f.'s, as in Peers (1971), Hayakawa (1975), and Harris and Peers (1980), for example. It is also possible to express these approximations as a sum of the asymptotic c.d.f. (the $O(1)$ term) and a weighted sum of probability density functions (p.d.f.'s) as is done in Rothenberg (1977). For an example of this in the W_t case, see the appendix to this chapter.

Corollary (to theorem 2). When the disturbance vector u of model (1) is normally distributed with a spherical covariance matrix as in (6), the test statistics of theorem 2 are approximately distributed as

$$\mathrm{pr}(W_t \leq x) = \mathrm{pr}(\chi_p^2(\omega) \leq x) + T^{-1} \sum_{i=0}^{4} \tau_{Wi}\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x) + o(-1) \quad (147)$$

$$\mathrm{pr}(LR_t \leq x) = \mathrm{pr}(\chi_p^2(\omega) \leq x) + T^{-1} \sum_{i=0}^{3} \tau_{LRi}\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x) + o(-1) \quad (148)$$

$$\mathrm{pr}(LM_t \leq x) = \mathrm{pr}(\chi_p^2(\omega) \leq x) + T^{-1} \sum_{i=0}^{4} \tau_{LMi}\, \mathrm{pr}(\chi_{p+2i}^2(\omega) \leq x) + o(-1) \quad (149)$$

where the τ_{Wi}'s, τ_{LRi}'s, and τ_{LMi}'s are given by (92), (93), and (94), respectively.

Proof. This result follows directly from theorem 2 by noting that distribution (6) is a limiting case of the assumption (2) in that theorem, as $\gamma \to \infty$, which causes certain terms in equations (88), (89), and (90), to disappear.

Q.E.D.

II.4 COMPARISON OF ACTUAL SIZE AND ASYMPTOTIC SIZE

(i) Introduction

In this section the behaviour of the test statistics W_t, LR_t and LM_t of (34), (35), and (36) under the null hypothesis H_0 of (14), or equivalently, (21), is examined. Under the null, the F test statistic (22) satisfies (26) even when the disturbances have a t distribution (see Zellner (1976, pg. 401) and King (1980)). This fact allows us to calculate the exact rejection probabilities under the null (i.e., the actual size) of these test statistics as well as some altered versions of these tests which are attempts to get the actual size of a test closer to its asymptotic size (which we will take to be 5%). These "size-corrected" tests are considered in the next section, followed by a numerical comparison.

(ii) Size-Corrected Tests

We know from general statistical theory that under the null hypothesis the W_t, LR_t and LM_t tests are asymptotically distributed as χ^2_p where p is the number of restrictions (see e.g., Harvey (1981, pp. 159-175), or Silvey (1970)). The goal of correction factors is to adjust the statistic or critical value by a monotonic transformation so that its small sample size will correspond more closely to its asymptotic size. This should make the actual size of the test closer to the assumed or nominal size which is fixed by selecting a critical value from the asymptotic distribution of the test statistic.

(iia) Correction Based on Degrees of Freedom.

A common and intuitively plausible way of correcting the W and LM statistics in the normal case, (28) and (29), is to correct for the bias in the estimate of the error variance which appears in the denominator when they are written in the following way:

$$W = (\hat{\theta}_2 - r)'(Z^{22})^{-1}(\hat{\theta}_2 - r)/\hat{\sigma}^2 \quad (150)$$

$$LM = (\hat{\theta}_2 - r)'(Z^{22})^{-1}(\hat{\theta}_2 - r)/\hat{\sigma}_R^2 \quad (151)$$

and

$$\hat{\sigma}^2 = y'My/T \quad (152)$$

$$\hat{\sigma}_R^2 = y'M_1y/T \quad (153)$$

where the notation is described in section 2(iii). While (152) and (153) are the unrestricted and restricted ML estimates of the error variance, respectively, they are biased since

$$E\,\hat{\sigma}^2 = (T - k)\sigma^2/T \quad (154)$$

and

$$E\,\hat{\sigma}_R^2 = (T - k + p)\sigma^2/T \text{ under } H_0 \quad (155)$$

Replacing $\hat{\sigma}^2$ and $\hat{\sigma}_R^2$ by unbiased estimates in (150) and (151) would yield the following size-corrected tests in the normal case:

$$W^{df} = (1 - k/T)W \quad (156)$$

$$LM^{df} = (1 - (k-p)/T)LM \quad (157)$$

Due to the close linear relationship between W and LM tests in the normal and t cases given by the definitions of W_t and LM_t in (35) and (36), one could extend this argument to those tests by defining

$$W_t^{df} = (1 - k/T)W_t \tag{158}$$

$$LM_t^{df} = (1-(k-p)/T)LM_t \tag{159}$$

(iib) A Revised Degrees of Freedom Correction.

It could be argued that it is not appropriate to correct for the bias in $\hat{\sigma}^2$ and $\hat{\sigma}_R^2$ in (150) and (151) because they appear in the denominators of W and LM respectively. Rather, one might wish to correct for the bias in $(\hat{\sigma}^2)^{-1}$ and $(\hat{\sigma}_R^2)^{-1}$ directly. Since $\hat{\sigma}^2$ and $\hat{\sigma}_R^2$ under the null follow χ^2 distributions with T - k and T - k + p d.f. respectively, we can apply the result:

$$E[(q^2)^{-1}] = (\gamma - 2)^{-1} \text{ when } q^2 \sim \chi_\gamma^2 \tag{160}$$

to show that

$$E(\hat{\sigma}^2)^{-1} = T/\sigma^2(T-k-2) \tag{161}$$

and

$$E(\hat{\sigma}_R^2)^{-1} = T/\sigma^2(T-k+p-2) \text{ under } H_0 \qquad (162)$$

These results suggest the following size-corrected tests which are slightly different from (156) and (157):

$$W^{df_*} = (1-(k+2)/T)W \qquad (163)$$

$$LM^{df_*} = (1 - (k+2-p)/T)LM \qquad (164)$$

and for the t case,

$$W_t^{df_*} = (1 - (k + 2)/T)W_t \qquad (165)$$

$$LM_t^{df_*} = (1-(k+2-p)/T)LM_t. \qquad (166)$$

It can be seen immediately that $W_t^{df_*}$ and LM^{df_*} "shrink" their respective original statistics more than do W_t^{df} and LM_t^{df}, and so the latter tests will have larger sizes than the corresponding former tests.

(iiic) Edgeworth Correction (Test Statistic)

The Edgeworth correction factors are chosen so that the distributions of the statistics are the same as their asymptotic distribution (which is χ^2_p here) to 0(-1). They are presented and derived in the following corollary.

Corrollary (to theorem 2). The Edgeworth size-corrected test statistics based on the approximate distributions of (88), (89), and (90) are

$$W^e_t = \{1-T^{-1}(k+1-(p/2)+W_t/2) + 2(T+\gamma)^{-1}\}W_t \qquad (167)$$

$$LR^e_t = \{1-(k+1-(p/2))/T\}LR_t \qquad (168)$$

$$LM^e_t = \{1-T^{-1}(k+1-(p/2)-LM_t/2)-2(T+\gamma)^{-1}\}LM_t \qquad (169)$$

so that the distribution of each of these statistics under the null hypothesis (14) is $\chi^2_p + o(-1)$.[7]

Proof: Under the null we have $\omega \doteq 0$ where ω of (91) is the non-centrality parameter of the χ^2 distributions in the expansions (88), (89), and (90). This also simplifies the coefficients of (92) to (97) and consequently the m.g.f.'s of (141), (143) and (146).

By setting $\omega = 0$ in (141) and noting from (126) and (137) that

$$\tilde{E}\ \eta_0 = p(1-2t)^{-1} \text{ and}$$
$$\tilde{E}\ \eta_0^2 = p(p+2)(1-2t)^{-2} \text{ under } H_0 \qquad (170)$$

we can write the m.g.f. of W_t under the null, setting $\omega = 0$ in (130), (134), and (140), as

$$M_W(t) = (1-2t)^{-p/2}\{1+T^{-1}[-t(p-2k-2)\tilde{E}\ \eta_0/2$$
$$+\ t\ \tilde{E}\ \eta_0^2/2]-2(T+\gamma)^{-1}\ \tilde{E}\ \eta_0\} + o(-1) \qquad (171)$$

Seen in this form, it is a simple matter to construct a revised statistic for which the $O(-1)$ terms are zero by noting that

$$W_t/T = \eta_0/T + o_p(-1) \text{ and } W_t^2/T = \eta_0^2/T + o_p(-1) \qquad (172)$$

Using (170) and (171), we see that by adding appropriate multiples of W_t/T and W_t^2/T to W_t we can eliminate the $O(-1)$ part of (171), and this method yields (167).

We can write the m.g.f.'s of LR_t and LM_t under the null in a similar way to (171) as

$$M_{LR}(t) = (1-2t)^{-p/2}\{1+T^{-1}[-t(p-2k-2)\tilde{E}\ \eta_0/2]+ o(-1) \quad (173)$$

$$M_{LM}(t) = (1-2t)^{-p/2}\{1+T^{-1}[-t(p-2k-2)\tilde{E}\ \eta_0/2$$

$$-t\tilde{E}\ \eta_0^2/2] + 2(T+\gamma)^{-1}\tilde{E}\ \eta_0]\} + o(-1), \quad (174)$$

from which results (168) and (169) follow.

The adjustments (167), (168), and (169) remove the 0(-1) terms from the above m.g.f.'s, so that the m.g.f. for W_t^e, LR_t^e and LM_t^e is

$$M^e(t) = (1-2t)^{-p/2} + o(-1) \quad (175)$$

which is the m.g.f. to 0(-1) of a central χ_p^2 variable, that is, it is the same m.g.f. to 0(-1) as that of the corresponding asymptotic distribution of the test statistic.

Q.E.D.

A test having a critical value based on its asymptotic χ_p^2 distribution should have a small-sample size that is closer to its chosen asymptotic size if it has been corrected in the above manner. Unfortunately, W_t^e is not a monotonic nondecreasing function of W_t. As a result, there can be situations where W_t^e will never reject the null, i.e., it has a size of zero. This can be seen from (167) by noting that

$$\partial W_t^e / \partial W_t = 0 \text{ when}$$

$$W_t = W_t^* = T-k + 1-p/2 + 2T/(T + \gamma) \quad (176)$$

Substituting W_t of (176) into (167) and verifying the second-order condition yields

$$\max_{W_t} W_t^e = (T/2)\{1-(k+1-p/2)/T + 2/(T+\gamma)\}^2 \quad (177)$$

Thus a test using W_t^e and critical value x will never reject the null when

$$\max W_t^e = \leq x \quad (178)$$

The non-monotonicity property of W_t^e can be corrected by re-defining it as

$$W_t^e = \begin{cases} \{1-(k+1-(p/2) + W_t/2)/T + 2/(T+\gamma)\}W_t, \\ \text{if } W_t \leq W_t^* \\ \max W_t^e \quad \text{if } W_t > W_T^* \end{cases} \quad (179)$$

This revised W_t^e, however, does not solve the non-rejection problem implied by (178).

(iid) Revised Edgeworth Correction (Critical Value)

The Edgeworth expansions can be used to construct size-corrected critical levels rather than size-corrected test statistics. In other words, the original statistics as defined in (34), (35), and (36) are still used, but the asymptotic critical values are adjusted in an attempt to get the actual size of the test closer to its asymptotic size. This procedure does not result in a non-rejection problem for W_t; however, it can result in a non-acceptance problem for LM_t (i.e., its adjusted critical value can be zero or negative, causing the actual size to be 100%). The condition for the existence of this problem, however, is less strict than condition (178), as we shall see. (These critical value corrections are derived from a method outlined in Rothenberg (1977, pp. 11-12)).

First, we require the Edgeworth expansions of W_t, LR_t, and LM_t in (88), (89), and (90), expressed in density function form. This conversion is done for W_t in the appendix, with the result given by equations (A10) to (A13). By similar methods the following expressions for LR_t and LM_t can be derived (using the notation of the appendix):

$$\mathrm{pr}(LR_t \le x) = F(x,p,\omega) + T^{-1} \sum_{j=1}^{3} \tau^{*}_{LRj} f(x,p+2j,\omega)$$

$$+ \gamma^{-1} \sum_{j=1}^{2} \phi^{*}_{j} \; f(x,p+2j,\omega) + o(-1) \qquad (180)$$

$$\mathrm{pr}(LM_t \le x) = F(x,p,\omega) + T^{-1} \sum_{j=1}^{4} \tau^{*}_{LMj} \; f(x,p+2j,\omega)$$

$$+ \gamma^{-1} \sum_{j=1}^{2} \phi^{*}_{j} \; f(x,p+2j,\omega)$$

$$+ (T+\gamma)^{-1} \sum_{j=1}^{2} \alpha^{*}_{LMj} f(x,p+2j,\omega) + o(-1) \qquad (181)$$

where

$$\tau^{*}_{LR1} = -p(2k+2-p)/2, \quad \tau^{*}_{LR2} = \mu(k-p)$$

$$\tau^{*}_{LR3} = \omega^2/2 \qquad (182)$$

$$\tau^{*}_{LM1} = -p(2k+2-p)/2, \quad \tau^{*}_{LM2} = p(p+2)/2-\omega(k-p)$$

$$\tau^{*}_{LM3} = \omega(p+2) + \omega^2/2, \quad \tau^{*}_{LM4} = \omega^2/2 \qquad (183)$$

$$\phi^{*}_{1} = \omega^2/2, \quad \phi^{*}_{2} = -\omega^2/2, \qquad (184)$$

$$\alpha^*_{LM1} = -2p, \quad \alpha^*_{LM2} = -2\omega, \tag{185}$$

and F and f represent a χ^2 c.d.f., and p.d.f., respectively (see appendix).

We are interested in the distributions under the null, so that $\omega = 0$. We can then write the c.d.f. of LM_t, for example, as

$$\begin{aligned} pr(LM_t \leq x) &= F(x,p,0) \\ &+ \sum_{j=1}^{2} (\tau^*_{LMj}/T) f(x,p+2j,0) \\ &+ o(-1) \end{aligned} \tag{186}$$

since the τ^* and ϕ^* terms with subscripts greater than two will equal zero, and similarly for LR_t and W_t.

Now using the following result which holds when $\omega = 0$, (Rothenberg, (1977, p. 11)),

$$x\, f(x,p+i,0) = (p+i) f(x,p+i+2,0), \tag{187}$$

we can express the p.d.f.'s of (186) as functions of f(x,p,0) by noting that

$$f(x,p+2,0) = x\, f(x,p,0)/p \tag{188}$$

and

$$f(x,p+4,0) = x^2 f(x,p,0)/p(p+2) \tag{189}$$

Using these relations we can write (186) in the LR_t case as

$$pr(LR_t \le x) = F(x,p,0) + \{x\tau^*_{LR1}/Tp + x^2\tau^*_{LR2}/T\,p(p+2)\}f(x,p,0) + o(-1) \qquad (190)$$

where the τ^*_{LR}'s are from (182).

By another approximation to $0(-1)$ we can express this c.d.f. of (190) in terms of a single χ^2 p.d.f. by the following manipulations.

Note that

$$F(x,p,0) = \int_0^x f(\tilde{x}, p, 0)\,d\tilde{x} \qquad (191)$$

and

$$cf(x,p,0) = f(x,p,0) \int_x^{x+c} d\tilde{x}$$

$$= \int_x^{x+c} f(x,p,0)\,d\tilde{x}$$

$$= \int_x^{x+c} [f(\tilde{x},p,0) - \{f(x,p,0) - \dot{f}(\tilde{x},p,0)\}]\,d\tilde{x} \qquad (192)$$

where

$$c = x\tau^{*}_{LR1}/Tp + x^{2}\ \tau^{*}_{LR2}/Tp(p+2) \qquad (193)$$

Substituting (191) and (192) into (190) gives

$$pr(LR_t \leq x) = \int_0^x f(\tilde{x},p,0)d\tilde{x} + \int_x^{x+c} [f(\tilde{x},p,0) + \{f(x,p,0)$$

$$- f(\tilde{x},p,0)\}]d\tilde{x} + o(-1) \qquad (194)$$

Noting that c has $0(T^{-1})$, we see that

$$\int_x^{x+c} \{f(x,p,0) - f(\tilde{x},p,0)\}d\tilde{x} = 0 + o(-1), \qquad (195)$$

therefore,

$$pr(LR_t \leq x) = \int_0^{x+c} f(\tilde{x},p,0)d\tilde{x} + o(-1)$$

$$= F(x+c,p,0) + o(-1) \qquad (196)$$

keeping in mind that we are assuming that the null hypothesis is true. Now, since c has $0(T^{-1})$ we can write

$$pr(LR_t \leq x-c) = F(x,p,0) + o(-1) \qquad (197)$$

Thus, if x represents the critical value which yields the desired asymptotic size, then (197) suggests the use of the following adjusted critical value for small samples:

$$x^*_{LR} = x - c = x\{1 + (k+1-p/2)/T\} \tag{198}$$

By following the same procedure as above we can derive critical value adjustments for LM_t and W_t, which are

$$x^*_W = x\{1+(k+1+(x-p)/2)/T-2/(T+\gamma)\} \tag{199}$$

and

$$x^*_{LM} = x\{1+(k+1-(x+p)/2)/T + 2/(T+\gamma)\} \tag{200}$$

It is clear from (199) that the non-rejection problem of W^e_t discussed previously will not occur when the critical value adjustment is used. It can be shown from (200), however, that

$$x^*_{LM} \leq 0 \text{ when } x \geq 2T\{1+(k+1-p/2)/T + 2/(T+\gamma)\}, \tag{201}$$

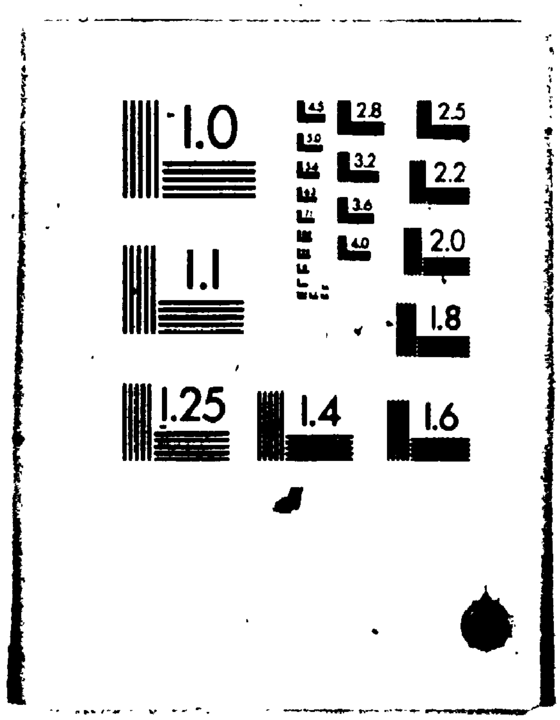
in which case the LM test using LM_t and critical value x^*_{LM} will <u>always</u> reject the null even when it is true[8]. Casual inspection of (177) and (201) would suggest that this non-acceptance problem will occur in a smaller set of T, p, k and γ values than the non-rejection case of W_t.

(iii) Numerical Comparison of Size Corrections

Earlier in this chapter it was noted that the F statistic (22) has an $F_{p,T-k}$ distribution under the null for all values of γ, that is, for the entire class of spherical student's t distributions (including the normal) for the disturbances. This, along with relations (38), (39), and (40) can be used to calculate the exact sizes of the test statistics, LR_t, LM_t, W_t as well as the size-adjusted statistics given in the previous section, for a specified asymptotic size (which in this chapter is the usual 5%[9]). The exact sizes are calculated by converting the critical value for a given test into the critical value for the F test which would have the same exact size. This is done by expressing F as a function of the given statistic. The small-sample knowledge of the F-test in this situation can then be used to obtain this exact size.

Denoting the critical value for one of the tests by x, and its corresponding F-test small sample critical value by $h(x,LR_t)$, say, for the LR_t test case, we can begin by using (38) to (40) to obtain

2



$$h(x, LR_t) = (T-k)\{\exp(x/T)-1\}/p \tag{202}$$

$$h(x, W_t) = (T-k)x/T\lambda p \tag{203}$$

and

$$h(x, LM_t) = (T-k)\lambda x/p(T-\lambda x) \tag{204}$$

Similar functions can be obtained for the statistics W_t^{df}, LM_t^{df}, $W_t^{df_*}$ and $LM_t^{df_*}$ of (158), (159), (165), and (166) by using those relations to express these statistics as functions of the F statistic and then solving for F, giving

$$h(x, W_t^{df}) = x/\lambda p \tag{205}$$

$$h(x, LM_t^{df}) = (T-k)\lambda x/p(T-k+p-\lambda x) \tag{206}$$

$$h(x, W_t^{df_*}) = (T-k)x/(T-k-2)\lambda p \tag{207}$$

and

$$h(x, LM_t^{df_*}) = (T-k)\lambda x/p(T-k+p-2-\lambda x) \tag{208}$$

The desired functions for the Edgeworth size corrected statistics W_t^e, LR_t^e, and LM_t^e of (167), (168), and (169) are solved in the same fashion, although the expressions for W_t^e and LM_t^e are somewhat complicated since they are each nonlinear functions of F, and in each case the derivation of h involves solving for the roots of a quadratic. The results are

$$h(x, LR_t^e) = (T-k)\{\exp((x/(T-m))-1)\}/p \qquad (209)$$

$$h(x, W_t^e) = (T-k)(T-m)[1-\{1-2Tx/(T-m)^2\}^{1/2}]/T\lambda p \qquad (210)$$

$$h(x, LM_t^e) = (T-k)\lambda\phi/p(T-\lambda\phi) \qquad (211)$$

where

$$\phi = (T-m)[\{1 + 2Tx/(T-m)^2\}^{1/2}-1] \qquad (212)$$

and

$$m = k + 1 - p/2 \qquad (213)$$

The appropriate h function for the tests which use the Edgeworth corrected critical values (198), (199), and (200) are found by substituting those adjusted critical values in place of x in the h functions for the unadjusted test statistics given in (202), (203), and (204). Denoting the use of the adjusted critical values with LR_t^e by LR_t^{e*} and so on, we then have

$$h(x, LR_t^{e*}) = (T-k)\{\exp(x_{LR}^*/T)-1\}/p \qquad (214)$$

$$h(x, W_t^{e*}) = (T-k)x_W^*/T\lambda p \qquad (215)$$

and

$$h(x, LM_t^{e*}) = (T-k)\lambda x_{LM}^*/p(T-\lambda x) \qquad (216)$$

where the x^*'s are the adjusted critical values given by (198), (199), and (200).

We can now find the exact size of any of these tests, say LR_t for example, by using

$$pr\{LR_t \leq x\} = pr\{F_{p,T-k} \leq h(x, LR_t)\} \qquad (217)$$

which holds under the null, and similarly for the other tests. Each of these h functions approaches x/p as T approaches infinity, which is to be expected since each of the test statistics is asymptotically χ_p^2 while $pF_{p,T-k}$ also becomes a χ_p^2 variable asymptotically.

The table which follows compares exact sizes of these tests for T,p, and k values taken from Tables I and II of Evans and Savin (1982) with an asymptotic size of 5%. The normality ($\gamma = \infty$) case and the $\gamma = 5$ case are considered. The h functions given above are used to get critical values for the $F_{p,T-k}$ variable as in (214), and the probability of rejection is found by using the subroutine MDFD from the MATHLIB package.

TABLE 1

Exact Sizes of Various Test Statistics When the Disturbances are Normal and Student's t with $\gamma = 5$, and $k = 8$

T	p	γ	Exact Size (Asymptotic Size is 0.05) LR_t	W_t	LM_t	W_t^{df}	LM_t^{df}	$W_t^{df_*}$	$LM_t^{df_*}$
33	1	∞	.091	.100	.082	.061	.048	.052	.039
33	1	5	.091	.092	.090	.055	.054	.046	.044
58	1	∞	.070	.075	.066	.056	.049	.051	.044
58	1	5	.070	.071	.070	.052	.053	.048	.048
33	8	∞	.110	.218	.024	.098	.024	.074	.014
33	8	5	.110	.192	.037	.083	.037	.061	.022
58	8	∞	.078	.129	.036	.075	.036	.063	.029
58	8	5	.078	.116	.044	.066	.044	.055	.036
108	8	∞	.064	.087	.043	.062	.043	.057	.038
108	8	5	.064	.081	.048	.058	.048	.052	.043

T	p	γ	LR_t^e	W_t^e	LM_t^e	LR_t^{e*}	W_t^{e*}	LM_t^{e*}
33	1	∞	.050	.045	.052	.058	.062	.055
33	1	5	.050	.049	.051	.058	.060	.056
58	1	∞	.050	.049	.051	.053	.054	.052
58	1	5	.050	.048	.047	.053	.053	.052
33	8	∞	.062	.000	.041	.059	.083	.048
33	8	5	.062	.000	.044	.059	.080	.046
58	8	∞	.051	.036	.047	.053	.061	.049
58	8	5	.051	.037	.048	.053	.059	.049
108	8	∞	.050	.047	.049	.051	.053	.050
108	8	5	.050	.048	.050	.051	.053	.050

The following remarks concern the results of Table 1.

1) The values of T, p, and k for $\gamma = \infty$ correspond to the cases considered in Evans and Savin (1982) in their Tables I and II when the null is true (in their notation, when the non-centrality parameter d is zero). The only tests tabulated here that are exactly comparable to the tests that they consider are the unadjusted tests, LM_t, LR_t and W_t (in the normal case) and LR_t^e. The results here correspond with theirs except for the $(T,p,\gamma) = (33,8,\infty)$ case for LR_t^e in which they obtain an exact size of .052 while the method described here yields .062.

In their tables they also give power comparisons. In the normal case this involves the non-central F distribution for which they have computational procedures. In the non-normal case considered here $(\gamma = 5)$ the exact distributions are not so simple. Since all of the tests in this chapter involve test statistics which are monotonic nondecreasing functions of $\dot{F}$ of (22) (when the revised W_t^e of (179) is used), then the test having larger size will also have larger power in these approximations under any local alternative.

2) The sizes of the unadjusted tests are generally quite different from .05. The LR_t and W_t tests are larger than .05 in every case while the LM_t test has sometimes a larger size and sometimes smaller. The sizes of three tests when $\gamma = \infty$ follow the same inequality as that of the statistics themselves given by (35) since the same critical value is used for each test ($x = 3.84$ when $p = 1$ and $x = 15.51$ when $p = 8$). These sizes are particularly far from .05 when p is large relative to T; in the table, see the $p = 8$, $T = 33$ case.

The most notable difference between the $\gamma = \infty$ and $\gamma = 5$ cases is that the unadjusted W_t and LM_t tests have sizes that are more similar in the latter case than the former, while the LR_t test is unaffected. The differences between the three test sizes, then, seem to be lessened by departures from normality to student's t. This result is backed up by inspection of the expansion formulas of (88) and (90), where the α's when $\omega = 0$ have signs opposite to the corresponding τ's of the same subscript[10].

3) The degrees of freedom corrections reduce the sizes of the tests, which is clear from the formulas for those statistics since the adjustments make them smaller in every case. The revised d.f. corrections reduce

the sizes by a larger amount than the original d.f. corrections. In the LM_t test cases where the unadjusted size is already less than .05, these d.f. corrections do more harm than good. The W_t^{df*} test sizes tend to be closer to .05 than the W_t^{df} test sizes since the latter adjustment often does not reduce the unadjusted test size enough.

4) Both types of Edgeworth correction are more effective than the d.f. corrections, with the critical value adjustment tending to give a larger test size than the test statistic adjustment, and with no one dominating the other in proximity to .05.

The (T,p) = (33,8) case is an example of a situation where the inequality (178) holds since $\dot{x}$ = 15.51 and max W_t^e = 11.88, thus the W_t^e test has a size of zero. There is no non-acceptance problem, however, and in fact the LM_t^{e*} test has a size which is very close to .05 for all parameter values considered here.

II.5 SUMMARY

In this chapter the Edgeworth expansions for the W, LR and LM tests for the linear hypothesis in the simple regression model in the case of student's t disturbances given by Ullah and Zinde Walsh (1984) are derived. These are then used to obtain size correction factors, and the exact sizes of these and other size corrected tests based on degrees of freedom adjustments are compared in some specific cases.

The sizes of the unadjusted W, LR and LM tests differ more under normality than in the student's t with d.f. $\gamma = 5$ case, and so the same can be said for their powers, since all are monotonic functions of the F statistic (22). The Edgeworth size correction factors work better than the degrees of freedom correction factors. Both Edgeworth factors, one which operates on the test statistic itself and one which operates on the critical value, are about equally effective, although when the number of parameters to be estimated is large relative to the sample size, the critical value adjustment may be preferred.

APPENDIX

In this appendix the approximation to the c.d.f. of W_t given in (88) is expressed in a form involving weighted sums of p.d.f.'s (as in Rothenberg (1977)), rather than c.d.f.'s. We introduce the following notation:

$$F(x,\nu,\delta) = \text{pr}\{\chi^2_\nu(\delta) \leq x\} \qquad x \geq 0 \tag{A1}$$

$$f(x,\nu,\delta) = \partial F(x,\nu,\delta)/\partial x \qquad x \geq 0 \tag{A2}$$

hence $F(x,\nu,\delta)$ is the c.d.f. of a non-central χ^2 variable and $f(x,\nu,\delta)$ is its corresponding p.d.f.

Note that the characteristic function of a non-central χ^2 having a c.d.f. of $F(x,\nu,\delta)$ is

$$C(t) = \exp\{it(1-2it)^{-1}\delta\}(1-2it)^{-\nu/2} \tag{A3}$$

where we have simply replaced t in the m.g.f. formula (110) by it, where i in this context is the complex number $i = (-1)^{1/2}$. Using the Inversion theorem formulae given in Kendall and Stuart (1969, p. 94-95), and our knowledge of the distribution that corresponds to C(it) in (A3), we have:

$$F(x,\nu,\delta) = (2\Pi)^{-1/2}\int_{-\infty}^{\infty} C(t)\{(1-e^{-ixt})/it\}dt \qquad (A4)$$

and

$$f(x,\nu,\delta) = (2\Pi)^{-1/2}\int_{-\infty}^{\infty} C(t)e^{-ixt}dt \qquad (A5)$$

Using the fact that

$$it(1-2it)^{-1} = (-1+(1-2it)^{-1})/2, \qquad (A6)$$

and

$$1 = e^{-i(0)t} \qquad (A7)$$

along with

$$f(0,\nu,\delta) = 0 \qquad (A8)$$

we can manipulate equations (A4) and (A5) to show that

$$F(x,\nu,\delta) - F(x,\nu+2,\delta) = 2f(x,\nu+2,\delta) \qquad (A9)$$

Applying result (A9) to (88), we can write the approximate c.d.f. of W_t in the following numerically equivalent form:

$$pr(W_t \leq x) = F(x,p,\omega) + T^{-1}\sum_{j=1}^{4}\tau^*_{Wj}f(x,p+2j,\omega)$$

$$+ \gamma^{-1}\sum_{j=1}^{2}\phi^*_j f(x,p+2j,\omega)$$

$$+ (T+\gamma)^{-1}\sum_{j=1}^{2}\alpha^*_{Wj}f(x,p+2j,\omega) + o(-1) \qquad (A10)$$

where

$$\tau^*_{W1} = -p(2k+2-p)/2 \qquad \tau^*_{W2} = -p(p+2)/2-\omega(k-p)$$

$$\tau^*_{W3} = \omega(\omega-2p-4)/2, \qquad \tau^*_{W4} = -\omega^2/2 \tag{A11}$$

$$\phi^*_1 = \omega^2/2, \qquad \phi^*_2 = -\omega^2/2 \tag{A12}$$

and

$$\alpha^*_{W1} = 2p, \qquad \alpha^*_{W2} = 2\omega \tag{A13}$$

where the variables p, k, and ω are as in (88).

The choice between (88) and (A10) as representations for our c.d.f. approximation is not crucial, but one may be handier than the other when the formulae are used to approximate sizes and powers in particular numerical examples, as well as for construction of size-corrected tests.

FOOTNOTES

(Chapter LI)

1. The density function of $u/\sigma = v$ is

$$f(v|\sigma) = [\Gamma\{(T+\gamma)/2\}/\Gamma(\gamma/2)(\gamma\Pi)^{T/2}](1+v'v/\gamma)^{-(T+\gamma)/2}$$

It turns out, however, that a different representation of this density given in equation (3) is much easier to work with in this context.

2. There is more than one way to construct a multivariate distribution in which each variable has an identical marginal distribution that is Student's t. (See Johnson and Kotz (1972)). The one specified here in (3), (4), and (5) is the one on which the LR_t, W_t and LM_t tests of Ullah and Zinde Walsh (1984) are based, and is the one most commonly used.

3. Sargan (1980, p. 1108) gives conditions which are sufficient for validity of Edgeworth expansions of asymptotic χ^2 test statistics. These conditions are not strict and are met by the statistics whose distributions are approximated in this way in the thesis.

4. This can be seen, for example, by using the m.g.f. in the form of equation (4.2) of Peers (1971) and noting that

$$e^{\lambda} = \sum_{j=0}^{\infty} \lambda^j/(j!)$$

5. In some sources the non-centrality parameter is given as $\omega/2$ rather than ω, for example, Graybill (1961, pp. 74-76).

6. If $g(z,q^2,q_*^2)$ of (121) is not a function of z, then the $\tilde{E}$ operator is an expectation in the usual sense. However, $\tilde{E}$ will still be used for notational consistency.

7. Since LR_t, W_t and LM_t are each monotonic functions of the F statistic, their power functions would be equal if their sizes were made equal by monotonic correction factors. If the Edgeworth approximations of the power

functions of LR_t^e, W_t^e and LM_t^e are derived, it is found that they are also equal, as would be expected, to $O(-1)$.

8. This point is raised in connection with a similar LM test for this hypothesis in the normal case by Evans and Savin (1982, p. 743).

9. The practice of having a test size chosen independently of sample size, and especially of having a test size as large as 5% for a large sample is criticized in Leamer (1977).

10. For some inequality results for LR_t, W_t and LM_t, see Ullah and Zinde Walsh (1984).

$$\hat{\rho} = \rho + \theta_{-1/2} + \theta_{-1} + o_p(T^{-1}) \tag{120}$$

where a term θ_i has $O(T^{i})$, we can write the corresponding expansion for $\hat{\beta}(\hat{\rho})$, i.e., the Aitken-type estimator of β using (15) with $\hat{\rho}$ replacing ρ, by

$$\hat{\beta} = \beta + (X'\hat{\Sigma}^{-1}X)^{-1} X' \hat{\Sigma}^{-1} u \tag{121}$$

and from (38) we can write

$$(1-\rho^2)\hat{\Sigma}^{-1} = (1-\rho^2)\Sigma^{-1} + \theta_{-1/2} R$$
$$+ (\theta_{-1} R + \theta^2_{-1/2} D)$$
$$+ \text{"}o_p\text{"} (T^{-1})^6 \tag{122}$$

where

$$D = \text{diag}(0,1,\ldots,1,0), \quad \text{a } T \times T \text{ matrix} \tag{123}$$

and

$$R = 2(\rho D - A) \tag{124}$$

where A is defined in (102). Now by expanding $(X'\hat{\Sigma}^{-1}X)^{-1}$

and is slightly different from the usual form. In the third section, the mean square errors (MSE) of various estimators of the regression coefficients are approximated by an expansion method first exploited by Nagar (1959). Included here are MSE expansions of iterative estimators using a general technique which could be applied to iterative estimators in other models such as the SURE model of chapter IV or simultaneous equations models. Finally, the distribution of the t test for significance of a regression coefficient, discussed by Park and Mitchell (1980), is approximated by an Edgeworth expansion, which leads to a size correction which may remove the marked over-rejection of the null noted by those authors.

III.2 THE MODEL AND ESTIMATORS

(i) The Model

The model in this chapter is the same as in chapter II except there is now a first-order autoregressive process in the disturbances, and these will be assumed to be normally distributed except in section III.3. We then have

$$y = X\beta + u \tag{1}$$

where

$$y = [y_1 \ldots y_T]' \tag{2}$$

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \\ x_{T1} & \cdots & x_{Tk} \end{bmatrix} = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix} \tag{3}$$

$$\beta = [\beta_1 \ \ldots \ \beta_k]' \tag{4}$$

$$u = [u_1 \ \ldots \ u_T]' \tag{5}$$

So we see that y and u are $T \times 1$ vectors of dependent variables and disturbances, respectively; β is a $k \times 1$ vector of unknown regression coefficients, and X is a $T \times k$ matrix of observed independent variables. The model becomes different from that of chapter II because now we make the following distributional assumption:

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad t = 1,\ldots,T \tag{6}$$

where

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2), \qquad t = 1,\ldots,T \tag{7}$$

$$E(\varepsilon_t \varepsilon_{t'}) = 0, \qquad t \neq t', \tag{8}$$

and

$$|\rho| < 1 \tag{9}$$

The error process is assumed to be stationary by assuming that :

$$u_0 \sim N(0, \sigma^2) \tag{10}$$

where

$$\sigma^2 = \sigma_\varepsilon^2/(1 - \rho^2) \tag{11}$$

so that

$$u_t = N(0, \sigma^2), \qquad t = 1,...,T \tag{12}$$

We can express these assumptions compactly as:

<u>ASSUMPTION</u>

$$u \sim N(0, \sigma^2 \textstyle\sum) \tag{13}$$

<u>where</u>

$$\textstyle\sum = \begin{bmatrix} 1 & \rho & \ldots & \rho^{T-1} \\ \rho & 1 & \rho \ldots & \rho^{T-2} \\ \vdots & & & \vdots \\ & & & \rho \\ \rho^{T-1} & & \rho & 1 \end{bmatrix} \tag{14}$$

i.e., the $(i,j)^{th}$ element of Σ equals $\rho^{|i-j|}$.

(ii) The Estimators

The fact that we may have $\rho \neq 0$ has implications for estimation of β and hence for significance tests. The value of ρ is not known. If it were, we would have the best linear unbiased estimator at our disposal by applying Aitken's theorem (see e.g., Theil (1971, p. 238)) which yields the generalized least squares (GLS) estimator

$$\hat{\beta}(\rho) = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y \qquad (15)$$

The variance-covariance (and MSE) matrix of $\hat{\beta}(\rho)$ is given by

$$E\{\hat{\beta}(\rho) - \beta\}\{\hat{\beta}(\rho) - \beta\}' = \sigma^2 \Omega \qquad (16)$$

where

$$\Omega = (X' \Sigma^{-1} X)^{-1} \qquad (17)$$

This estimator is not operational since ρ is unknown, but we can define a class of estimators for β based on result (15) by replacing ρ by an estimate. Some of these estimates of ρ are summarized by Judge et. al. (1980, pp. 182-3), and are given below:

1) The "standard" estimate,

$$\hat{\rho}_s = \sum_{t=2}^{T} e_t \, e_{t-1} / \sum_{t=1}^{T} e_t^2 \tag{18}$$

where

$$e = [e_1 \ldots e_t \ldots e_T]' = My = Mu, \tag{19}$$

$$M = I - X(X'X)^{-1} X' \tag{20}$$

so that the e_i's are the residuals obtained from OLS regression on (1). There are some slight modifications of $\hat{\rho}_s$ also in use. For example, Park and Mitchell (1980) use a summation from 2 to T-1 in the denominator of (15) instead of 1 to T. This changes the $O_p(T^{-1})$ term of the asymptotic expansion of $\hat{\rho}_s$ and leaves the $O_p(1)$ and $O_p(T^{-1/2})$ terms the same, and as we shall see, this has no effect to $O(T^{-1})$ on the MSE approximation of the resulting estimator, which is

$$\hat{\beta}(\hat{\rho}_s) = \hat{\beta}_s = (X' \hat{\Sigma}_s^{-1} X)^{-1} X' \hat{\Sigma}_s^{-1} y \tag{21}$$

where $\hat{\Sigma}_s$ is simply Σ of (14) with ρ replaced by $\hat{\rho}_s$, and we denote $\hat{\beta}(\hat{\rho}_s)$ by $\hat{\beta}_s$ for notational convenience.

2) The Theil (1971) estimate,

$$\hat{\rho}_{TH} = (T - k)\hat{\rho}_S/(T - 1), \tag{22}$$

where $\hat{\rho}_S$ is given in (18)

3) The estimate based on the Durbin-Watson (1950) statistic,

$$\hat{\rho}_{DW} = 1 - d/2 \tag{23}$$

where

$$d = \sum_{t=2}^{T} (e_t - e_{t-1})^2 / \sum_{t=1}^{T} e_t^2 \tag{24}$$

and d is the familiar Durbin-Watson statistic used to test whether ρ equals zero.

4) The Theil-Nagar (1961) estimate,

$$\hat{\rho}_{TN} = (T^2 \hat{\rho}_{DW} + k^2)/(T^2 - k^2) \tag{25}$$

5) The Durbin (1960) estimate, which is given by the OLS estimate of ρ in the following equation:

$$y_t = \rho y_{t-1} + \beta_1(1-\rho) + x_t^{0'}\beta^0 - \rho x_{t-1}^{0'}\beta_0$$

$$+ \varepsilon_t, \qquad t = 2,\ldots,T \qquad (26)$$

where β^0 and β_0 are coefficient vectors compatible with $x_t^{0'}$ and $x_{t-1}^{0'}$, and $x_t^{0'}$ is the t^{th} row of x with the constant term deleted if it was originally in X, otherwise the $\beta_1(1-\rho)$ term is deleted. The estimated coefficient for the y_{t-1} term is $\hat{\rho}_D$, the Durbin estimate. Equation (26) is derived by subtracting ρ times the $(t-1)^{th}$ row of model (1) from its t^{th} row, and it can be seen that OLS estimation of (26) ignores some implied parameter restrictions.

6) The maximum likelihood estimate of ρ, $\hat{\rho}_{ML}$, is appropriate here because the maximum likelihood estimate of β given $\hat{\rho}_{ML}$ is

$$\hat{\beta}(\hat{\rho}_{ML}) = \hat{\beta}_{ML} = (X' \hat{\Sigma}_{ML}^{-1} X)^{-1} X' \hat{\Sigma}_{ML}^{-1} y \qquad (27)$$

where $\hat{\Sigma}_{ML}$ is Σ of (14) with ρ replaced by $\hat{\rho}_{ML}$. Thus, the ML estimate of β is also an Aitken-type estimate. The estimate $\hat{\rho}_{ML}$ is a function of $\hat{\beta}_{ML}$, however, so that a closed form solution of $\hat{\beta}_{ML}$ and $\hat{\rho}_{ML}$ is unavailable. Beach and MacKinnon (1978) point out that $\hat{\rho}_{ML}$ is given by the solution to the following cubic polynomial equation:

$$\hat{\rho}_{ML}^3 + a\hat{\rho}_{ML}^2 + b\hat{\rho}_{ML} + c = 0 \tag{28}$$

where

$$a = -(T-2)\sum_{t=2}^{T}\hat{u}_t\,\hat{u}_{t-1}/[(T-1)(\sum_{t=2}^{T}\hat{u}_{t-1}^2 - \hat{u}_1^2)] \tag{29}$$

$$b = [(T-1)\hat{u}_1^2 - T\sum_{t=2}^{T}\hat{u}_{t-1}^2 - \sum_{t=2}^{T}\hat{u}_t^2]/[(T-1)(\sum_{t=2}^{T}\hat{u}_{t-1}^2 - \hat{u}_1^2)] \tag{30}$$

$$c = T\sum_{t=2}^{T}\hat{u}_t\,\hat{u}_{t-1}/[(T-1)(\sum_{t=2}^{T}\hat{u}_{t-1}^2 - \hat{u}_1^2)] \tag{31}$$

and $$\hat{u}_t = y_t - x_t'\hat{\beta}_{ML} \tag{32}$$

From (32) we see that $\hat{\rho}_{ML}$ and $\hat{\beta}_{ML}$ must be solved iteratively, but it is found by those authors that such a procedure converges satisfactorily in only a few iterations (see also Harvey and McAvinchey (1978)).

Any one of the above six estimators of ρ can be used to form an estimate of $\sum$ in (15) and hence an estimate of β.[1] This approach was first suggested by Prais and Winsten (1954) using $\hat{\rho}_s$, and the estimator $\hat{\beta}_s$ of (21) is termed the Prais-Winsten estimator.

Another class of estimator, first suggested by Cochrane and Orcutt (1949), was based on a different line of thought but is very closely related to the Aitken-type estimators described above. In Cochrane-Orcutt type estimation the model is transformed in an attempt to remove the autocorrelation from the disturbance term. If ρ were known, (6) suggests subtracting ρ times the $(t - 1)^{th}$ row from the t^{th} row of model (1) so that the transformed model consists of the t-1 equations:

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})' \beta + \varepsilon_t, \qquad (33)$$

$$t = 2,\ldots,T$$

The disturbances in this transformed model are independent so that it is appropriate to estimate β in (33) by OLS. The resulting estimator, $\tilde{\beta}(\rho)$, is given by

$$\tilde{\beta}(\rho) = (X^{*\prime} X^{*})^{-1} X^{*\prime} y^{*} \qquad (34)$$

where

$$X^* = Q^* X, \tag{35}$$

$$y^* = Q^* y \tag{36}$$

and

$$Q^* = \begin{bmatrix} -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ 0 & & \ddots & & \\ \vdots & & & \ddots & \\ 0 & & & -\rho & 1 \end{bmatrix} \tag{37}$$

a (T-1) × T matrix. The estimator (34) can be made operational in the same way as the Aitken-type estimator (15) by replacing ρ in Q^* by an estimate, such as one of the six described above.

The similarity between Aitken-type estimators and Cochrane-Orcutt types can be seen by noting that Σ^{-1} of (15) is

$$\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & & \vdots \\ \vdots & & & \ddots & \vdots \\ \vdots & & & 1+\rho^2 & -\rho \\ 0 & & & -\rho & 1 \end{bmatrix} \tag{38}$$

This can be written as

$$\Sigma^{-1} = (1/(1 - \rho^2))Q^{**'}Q^{**} \tag{39}$$

where

$$Q^{**} = \begin{bmatrix} (1 - \rho^2)^{1/2} \; 0 \; \ldots \; 0 \\ \hline Q^* \end{bmatrix}, \text{ a } T \times T \text{ matrix} \tag{40}$$

Now $\hat{\beta}(\rho)$ of (15) can be written as

$$\hat{\beta}(\rho) = (X^{**'}X^{**})^{-1}X^{**'}y^{**} \tag{41}$$

where

$$X^{**} = Q^{**}X = \begin{bmatrix} (1-\rho^2)^{1/2}x_1 \\ \hline X^* \end{bmatrix} \tag{42}$$

$$y^{**} = Q^{**}y = \begin{bmatrix} (1-\rho^2)^{1/2}y_1 \\ \hline y^* \end{bmatrix} \tag{43}$$

and the similarity of (41) and (34) is clear. We can also express (34) in a form similar to (15) by noting that

$$\frac{1}{1-\rho^2} Q^{*'} Q^{*} = \Sigma^{-1} - D_0 \tag{44}$$

where D_0 is a $T \times T$ matrix of zeroes with a one in the upper left corner. Thus, we can write $\tilde{\beta}(\rho)$ of (34) as

$$\tilde{\beta}(\rho) = (X'(\Sigma^{-1} - D_0)X)^{-1} X'(\Sigma^{-1} - D_0)y, \tag{45}$$

from which the similarity with $\hat{\beta}(\rho)$ of (15) is evident.

These relations will be useful in section 4 where approximate mean square error matrices of some of these estimators are compared.

The maximum likelihood estimate of σ^2 is given by

$$\hat{\sigma}^2_{ML} = \hat{u}' \hat{\Sigma}^{-1} \hat{u}/T \tag{46}$$

where

$$\hat{u} = y - X \hat{\beta}_{ML} \tag{47}$$

and $\hat{\Sigma}_{ML}$ is as in (27) with $\hat{\rho}_{ML}$ as in (28). The more commonly used estimate, however, is the mean square of the estimated residuals,

$$\hat{\sigma}_1^2 = \hat{u}'\ \hat{u}/T \tag{48}$$

where the $\hat{u}$'s are residuals resulting from whatever estimate is used for β. This estimate can be adjusted for degrees of freedom, giving

$$\hat{\sigma}_2^2 = \hat{u}'\ \hat{u}/(T - k) \tag{49}$$

In the next section it is shown that the LM test for the existence of autocorrelation is a function of $\hat{\rho}_s$ of (18) both in the case of normal disturbances and the case of multivariate student's t disturbances.

III.3 THE LM TEST FOR $\rho = 0$ WITH STUDENT'S t DISTURBANCES

In this section the normality assumption is temporarily relaxed, and we consider model (1) when the disturbances follow a multivariate student's t distribution with the following density function:

$$f(u) = (\text{const.})(\sigma^2)^{-T/2}|\Sigma|^{-1/2}(1 + u'\Sigma^{-1}u/\gamma\sigma^2)^{-(\gamma+T)/2} \tag{50}$$

where (const.) is a normalizing constant, γ is a degrees of freedom parameter, σ^2 is a dispersion parameter, and Σ is a $T \times T$ variance-covariance parameter matrix as in the normal case. As in chapter II, the normal case of (13) is obtained in the limit as γ approaches infinity. The u vector still has zero mean and its variance covariance matrix is

$$Euu'/T = [\gamma/(\gamma - 2)]\sigma^2\Sigma \tag{51}$$

We consider the hypothesis that the elements of u are not correlated[2], that is

$$H_0 : \rho = 0 \quad \text{vs.} \quad H_1 : \rho \neq 0 \tag{52}$$

Theorem 1. The LM test for (52) in model (1) when the disturbances are distributed as multivariate student's t with density (50) is given by

$$LM = T^2 \hat{\rho}_S^2/(T - 1) \tag{53}$$

where $\hat{\rho}_S$ is given in (18). This statistic does not depend on the degrees of freedom parameter γ, and so the LM statistic (53) is also appropriate for the normal case of (13).

Proof. From (50) we can write the log likelihood function for the unknown parameters β, σ^2 and ρ as

$$\ell = \log L(\rho, \sigma^2, \beta | y, X) = (\text{const.}) - (T/2) \log \sigma^2$$

$$- (1/2) \log |\Sigma| - ((\gamma + T)/2) \log \{1$$

$$+ g(\rho, \beta)/\gamma\sigma^2(1 - \rho^2)\} \tag{54}$$

where

$$g(\rho, \beta) = (1 - \rho^2) u' \Sigma^{-1} u$$

$$= \sum_{t=1}^{T} u_t^2 + \rho^2 \sum_{t=2}^{T-1} u_t^2 - 2\rho \sum_{t=2}^{T} u_t u_{t-1} \tag{55}$$

$$= g$$

and

$$u_t = y_t - x_t \beta \tag{56}$$

The LM test statistic, from (33) of chapter I, is

$$LM = \tilde{d}' I(\hat{\theta}_R)^{-1} \tilde{d} \tag{57}$$

where

$$\hat{d} = \partial \ell / \partial \theta |_{\hat{\theta}_R} \qquad (58)$$

$$I(\hat{\theta}_R) = -E\ \partial^2 \ell / \partial \theta \partial \theta' |_{\hat{\theta}_R} \qquad (59)$$

and

$$\theta = (\beta', \rho, \sigma^2), \qquad (60)$$

with $\hat{\theta}_R$ referring to the restricted maximum likelihood (ML) estimates.

Taking first derivatives of ℓ of (54) we have

$$\partial \ell / \partial \rho = \rho(T - 1)/(1 - \rho^2) - (\gamma + T)(2\rho\ g/(1 - \rho^2)$$

$$+ \partial g / \partial \rho)/2(\gamma\sigma^2(1 - \rho^2) + g) \qquad (61)$$

$$\partial \ell / \partial \sigma^2 = -T/2\sigma^2 + g(\gamma + T)/2\sigma^2(\gamma\sigma^2(1 - \rho^2) + g) \qquad (62)$$

and

$$\partial \ell / \partial \beta = -(\gamma + T)(\partial g / \partial \beta)/2(\gamma\sigma^2(1 - \rho^2) + g) \qquad (63)$$

We can now calculate $\hat{\theta}_R$ by noting first that

$$\hat{\rho}_R = 0 \qquad (64)$$

due to the restriction of H_0 in (52). The first order conditions then become

$$-T/2\hat{\sigma}_R^2 + \hat{g}_R(\gamma + T)/2\hat{\sigma}_R^2(\gamma\hat{\sigma}_R^2 + \hat{g}_R) = 0 \qquad (65)$$

and

$$\partial g/\partial \beta|_{\hat{\theta}_R} = X'(y - X\hat{\beta}_R) = 0 \qquad (66)$$

Equation (66) yields the familiar OLS estimator

$$\hat{\beta}_R = b = (X'X)^{-1}X'y, \qquad (67)$$

and (64) and (67) substituted in (55) give

$$\hat{g}_R = \sum_{t=1}^{T} e_t^2 \qquad (68)$$

where

$$e_t = y_t - x_t b \qquad (69)$$

as in (19), so that substitution of (68) into (65) and simplifying yields

$$\hat{\sigma}_R^2 = e'e/T, \qquad (70)$$

so that $\hat{\sigma}_R^2$ is simply the mean square of the OLS residual vector.

We then have

$$\tilde{d}' = (0,\ldots,0, -T\, e_{-1}'\, e_{+1}/e'e, 0) \qquad (72)$$

where

$$e_{-1} = (e_1,\ldots,e_{T-1})' \qquad (73)$$

and

$$e_{+1} = (e_2,\ldots,e_T)', \qquad (74)$$

since the first derivatives with respect to the unrestricted parameters are zero at their ML values and the first derivative with respect to ρ is found by substituting the restricted ML estimates in (61) and simplifying.

Next, we show that $I(\hat{\theta}_R)$ is block diagonal such that we need only consider the diagonal element corresponding to ρ for the calculation of LM. Taking second partial derivatives, we find

$$\partial^2\ell/\partial\sigma^2\partial\rho = (\gamma + T)\,\gamma(1 - \rho^2)\{2g\rho/(1 - \rho^2) + \partial g/\partial\rho\}/2\{\gamma\sigma^2(1 - \rho^2) + g\}^2 \qquad (75)$$

and

$$\partial^2\ell/\partial\beta\partial\rho = (\gamma + T)\{(\partial g/\partial\rho - 2\rho\gamma\sigma^2)\partial g/\partial\beta - (\gamma\sigma^2(1 - \rho^2) + g)\partial^2 g/\partial\beta\partial\rho\}/2(\gamma\sigma^2(1 - \rho^2) + g)^2 \qquad (76)$$

Evaluating (75) and (76) at $\hat{\theta}_R$ we find that

$$\partial^2 \ell/\partial\sigma^2\partial\rho|_{\hat{\theta}_R} = -\gamma\ e_{-1}'\ e_{+1}/(\gamma + T)(\hat{\sigma}_R^2)^2 \qquad (77)$$

and

$$\partial^2 \ell/\partial\beta\partial\rho|_{\hat{\theta}_R} = (X_{-1}'\ e_{+1} + X_{+1}'\ e_{-1})/2\hat{\sigma}_R^2 \qquad (78)$$

where X_{-1} and X_{+1} are $(T - 1) \times k$ matrices obtained by deleting the last row and the first row respectively from X.

Taking the expectations of (77) and (78) using $\hat{\theta}_R$ as the true parameter values we have

$$E\ \partial^2 \ell/\partial\sigma^2\partial\rho|_{\hat{\theta}_R} = 0 \qquad (79)$$

and

$$E\ \partial^2 \ell/\partial\beta\partial\rho|_{\hat{\theta}_R} = 0 \qquad (80)$$

All we need, then, for calculating LM in (57) is the expectation of the following second partial:

$$\begin{aligned}\partial^2 \ell/\partial\rho^2 = {} & (T - 1)(1 + \rho^2)/(1-\rho^2)^2 \\ & - (\gamma + T)\{(2g/(1 - \rho^2))\ (\gamma\sigma^2 \\ & + 3\gamma\rho^2\sigma^2 + g(1 + \rho^2)/(1 - \rho^2)) \\ & + 2\rho(\gamma\sigma^2 - g/(1 - \rho^2))\partial g/\partial\rho\end{aligned}$$

...continued

$$- (\partial g/\partial\rho)^2 + (\gamma\sigma^2(1 - \rho^2) + g)\partial^2 g/\partial\rho^2\}$$

$$/2(\gamma\sigma^2(1 - \rho^2) + g)^2 \qquad (81)$$

Evaluated at $\hat{\theta}_R$ this becomes

$$\partial^2\ell/\partial\rho^2|_{\hat{\theta}_R} = T - 1-(\gamma + T)\{2T(T + \gamma)(\hat{\sigma}_R^2)^2$$

$$+ (\gamma + T)\hat{\sigma}_R^2\, \partial^2 g/\partial\rho^2 - (\partial g/\partial\rho)^2\}/$$

$$2((\gamma + T)\hat{\sigma}_R^2)^2 \qquad (82)$$

Using

$$E\, e_t^2|_{\hat{\theta}_R} = \hat{\sigma}_R^2 \qquad t = 1,\ldots,T \qquad (83)$$

and

$$E\, e_t e_{t'}|_{\hat{\theta}_R} = 0 \qquad t \neq t',\ t,t' = 1,\ldots,T \qquad (84)$$

then we have, after some simplification,

$$E\, \partial^2\ell/\partial\rho^2|_{\hat{\theta}_R} = -(T - 1), \qquad (85)$$

so that the required element of $I(\hat{\theta}_R)$ is equal to T - 1. Using this result along with the block diagonality of $I(\hat{\theta}_R)$ and (72), applied to the LM formula (57) yields the result (53) of the theorem.

Q.E.D.

Since there is one parameter restriction being tested by (53), its asymptotic null distribution is χ_1^2 and so an asymptotic critical value for the LM test can be chosen accordingly. This test statistic is slightly different from that which is usually used in this situation, which is

$$LM^* = T\,\hat{\rho}_S^2 \qquad (86)$$

which results from the discussion in Breusch and Pagan (1980, pg. 244) for example. LM of (53) and LM^* of (86) are asymptotically identical and take on very similar values in small samples as well, since

$$LM = T\,LM^*/(T-1) \qquad (87)$$

This result is also applicable to the LM test for MA(1) errors since Godfrey's (1981) results imply that the LM test for hypotheses (52) which tests against a first order autoregressive process in the disturbances is identical to the test that would result if the alternative was a first order moving average process, i.e., if (6) was replaced by

$$u_t = \rho\varepsilon_{t-1} + \varepsilon_t, \qquad t = 2,\ldots,T \qquad (88)$$

Although the LM test statistic (53) is not a function of the γ parameter which helps to define the distribution of the disturbances, its small sample distribution may very well be, but this issue is not examined here. In the rest of this chapter, we again impose the normality assumption (13).

III.4 COMPARISON OF APPROXIMATE MSE'S OF ESTIMATORS

In section 2, six proposed estimators for ρ and two estimators for β using a ρ estimate were given. Five of the six ρ estimators can be used in two-stage estimation (excluding $\hat{\rho}_{ML}$, which would no longer be an ML estimate if (29) to (31) used e instead of $\hat{u}$). Five of the six ρ estimators can be used in iterative estimation (with $\hat{\rho}_D$ being the exception) as mentioned in footnote 1. This leaves twenty possible estimates for β after accounting for the three choices to be made: first an estimator for ρ, next an estimator for β given $\hat{\rho}$, and finally, whether to iterate or to stop at the first iteration. The most important decision of the three in most cases is the choice of estimator for β given $\hat{\rho}$, despite the analytic similarity between $\hat{\beta}$ of (15) and $\tilde{\beta}$ of (45) discussed earlier. Since $\hat{\beta}$ seems to perform better[3] than $\tilde{\beta}$ in previous studies mentioned below, only one variant of $\tilde{\beta}$ is considered in this section.

The Aitken-type estimators are considered in more detail here, including all of the possible ρ estimators in the two-stage procedures and two iterative estimators. The method used is Nagar's (1959) expansion and a subsequent approximation of the MSE matrix of the β estimate to $0(T^{-2})$. It turns out that this approximation gives an identical result for many of the estimators.

A brief literature review is next followed by the derivation of the MSE approximations and a numerical comparison.

(i) Literature Review

The use of OLS in model (1) can be a good idea in some situations even if $\rho \neq 0$. Kadiyala (1968) and Maeshiro (1976) demonstrate analytically that OLS outperforms Cochrane-Orcutt (CO) with ρ known and certain prespecified values for X.[4] Taylor (1981) argues that the poor performance of CO in those studies is due to the fixed nature of the X matrices. In Monte Carlo studies of Rao and Griliches (1969) and Spitzer (1979) where a separate drawing of X from a prespecified stochastic process is used for each replication, CO performs better than in the fixed X studies, although Spitzer recommends OLS when $|\rho| \leq .2$. Hoque (1980) shows analytically that CO is better than OLS when X contains no constant and a linear or geometric trend.

Tillman (1975) discusses a class of X matrices introduced by Watson (1967) consisting of certain linear combinations of eigenvectors of $\sum$ for which the performance of OLS is especially poor. One might then want to avoid OLS when X is in or close to this class.

In the comparisons between Prais-Winsten (PW) estimators (with either $\hat{\beta}_S$ or $\hat{\beta}_D$ or both being considered) and CO estimators, PW generally has lower MSE. Maeshiro (1979) finds this analytically in the known ρ case (which is to be expected, since PW is then BLUE due to Aitken's theorem), especially for large ρ when X is trended and for small (in magnitude) ρ when X is non-trended, even for sample sizes as large as 100. Park and Mitchell (1980) recommend PW, especially the iterated version, and Spitzer (1979) recommends PW when $.2 \leq |\rho| \leq .5$.

The relative merits of $\hat{\rho}_D$ and $\hat{\rho}_S$ in PW estimation (i.e., $\hat{\beta}_D$ vs. $\hat{\beta}_S$) has also been considered. Rao and Griliches (1969) favour $\hat{\beta}_D$ as they find that $\hat{\rho}_D$ is less biased than $\hat{\rho}_S$. In Monte Carlo studies, however, Harvey and McAvinchey (1978), Spitzer (1979), and Kramer (1980) find that $\hat{\beta}_S$ tends to have a lower MSE than $\hat{\beta}_D$.

The maximum likelihood estimate $\hat{\beta}_{ML}$ is found to be very efficient by Beach and MacKinnon (1978), who also present an efficient algorithm for its calculation. Harvey and McAvinchey (1978) find it especially effective when ρ is close to one, and Spitzer (1979) recommends its use when $|\rho| > .5$, all of the above results being based on Monte Carlo studies.

The studies cited here are based either on Monte Carlo simulations or are anlaytic with ρ assumed known. In the following section these estimators are compared analytically without assuming that ρ is known by means of asymptotic approximations to their MSE matrices using Nagar's expansions.

(ii) MSE Approximations[5]

In this section it is shown that the MSE approximations of the two-stage $\hat{\beta}_S$ of (21), the converged iteration of $\hat{\beta}_S$ (denoted $\hat{\beta}_S^I$), two-stage $\hat{\beta}_{TH}$, $\hat{\beta}_{DW}$, $\hat{\beta}_{TN}$, as well as $\hat{\beta}_{ML}$, are identical to $0(T^{-2})$. In addition, the MSE approximations to $0(T^{-2})$ for $\hat{\beta}_D$ and $\tilde{\beta}_S$ (which is calculated by using (34) with $\hat{\rho}_S$ in place of ρ) are derived.

Since the OLS estimator is also useful in certain cases, its MSE will also be considered. The OLS estimator is

$$b = (X'X)^{-1} X'y \tag{89}$$

Its MSE is given by

$$\begin{aligned} MSE(b) &= E(b-\beta)(b-\beta)' = E(X'X)^{-1}X'uu'X(X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} X'\Sigma X (X'X)^{-1} \end{aligned} \tag{90}$$

due to the normality assumption (13) and its unbiasedness.

Next, we present approximations to $O_p(T^{-1})$ for the various estimators of ρ given in section (ii). These will be required for the MSE approximations which will follow.

<u>Lemma 1.</u> <u>The expansions of the estimators of ρ of section (ii) in orders of probability of $T^{1/2}$ are given by:</u>

$$\hat{\rho}_S = \rho + \theta^S_{-1/2} + \theta^S_{-1} + o_p(T^{-1}) \tag{91}$$

$$\hat{\rho}_{TH} = \hat{\rho}_S - (k - 1)\rho/T + o_p(T^{-1}) \tag{92}$$

$$\hat{\rho}_{DW} = \hat{\rho}_S + (e_1^2 + e_T^2)/2T\sigma^2 + o_p(T^{-1}) \tag{93}$$

$$\hat{\rho}_{TN} = \hat{\rho}_{DW} + o_p(T^{-1}) \tag{94}$$

$$\hat{\rho}_S^I = \hat{\rho}_S + (2e'APQ\Sigma^{-1}u + u'\Sigma^{-1}QPBPQ\Sigma^{-1}u)/T\sigma^2 + o_p(T^{-1}) \tag{95}$$

$$\hat{\rho}_{ML} = \hat{\rho}_S^I + \rho(e_1^2 + e_T^2 - \sigma^2)/T\sigma^2 + o_p(T^{-1}) \tag{96}$$

$$\hat{\rho}_D = \rho + \theta^D_{-1/2} + \theta^D_{-1} + o_p(T^{-1}) \tag{97}$$

where

$$\theta^S_{-1/2} = e'Be/T\sigma^2 \tag{98}$$

$$\theta^S_{-1} = -\theta^S_{-1/2}((e'e/T) - \sigma^2)/\sigma^2 \tag{99}$$

$$\theta^D_{-1/2} = u'C_1'M_Z C_\rho u/T\sigma^2 \tag{100}$$

$$\theta^D_{-1} = -\theta^D_{-1/2}((u'C_1'M_Z C_1 u/T) - \sigma^2)/\sigma^2 \tag{101}$$

$$A = ((a_{ij})) \tag{102}$$

<u>a</u> T × T <u>matrix with</u>

$$a_{ij} = \begin{cases} 1/2 & \text{if } |i - j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$P = X(X'X)^{-1}X' \tag{103}$$

$$Q = \Sigma - X \Omega X' \tag{104}$$

<u>where</u> Σ <u>is from</u> (14) <u>and</u> Ω <u>from</u> (17),

$$B = A - \rho I_T \tag{105}$$

$$C = [0 \vdots I_{T-1}] \tag{106}$$

<u>a</u> (T - 1) × T <u>matrix</u>

$$C_1 = [I_{T-1} \vdots 0] \tag{107}$$

<u>a</u> (T - 1) × T <u>matrix</u>

$$C_\rho = C - \rho C_1 \tag{108}$$

$$M_Z = I - Z(Z'Z)^{-1}Z' \tag{109}$$

<u>a</u> (T - 1) × (T - 1) <u>matrix with</u>

$$z' = [z_2', \ldots, z_T'] \quad (110)$$

and

$$z_t = [1 \quad x_{t-1}^{0'} \quad x_t^{0'}] \quad (111)$$

where the x_t^0's are defined in (26)

when X contains a constant term (e.g., a column of ones), and

$$z_t = [x_{t-1}^{0'} \quad x_t^{0'}] \quad (112)$$

when X does not contain a constant term, so that Z is $(T - 1) \times (2k - 1)$ when (111) holds, and Z is $(T - 1) \times 2k$ when (112) holds, and e is the OLS residual vector defined in (19).

Proof. Considering first $\hat{\rho}_S$ of (18), note that it can be written as

$$\hat{\rho}_S = e'Ae/e'e \quad (113)$$

where A is defined in (102). It can be verified that $(e'Ae/T - \rho\sigma^2)$ and $(e'e/T - \sigma^2)$ have $0_p(T^{-1/2})$, since $e'Ae = T\rho\sigma^2 + 0_p(T^{1/2})$ and $e'e = T\sigma^2 + 0_p(T^{1/2})$ so that

$$\hat{\rho}_S = \{\rho\sigma^2 + (e'Ae/T - \rho\sigma^2)\}/\{\sigma^2 + (e'e/T - \sigma^2)\} \tag{114}$$

and expand the denominator to give the result (91) using Nagar's method.

The result for $\hat{\rho}_{TH}$ of (22) given in (92) follows from (22). For $\hat{\rho}_{DW}$, note from (24) that the Durbin-Watson statistic d can be written

$$d = 2(1 - \hat{\rho}_S) - (e_1^2 + e_T^2)/e'e \tag{115}$$

and combining (115) with (23) yields the result for $\hat{\rho}_{DW}$ of (93). From (25) it is clear that $\hat{\rho}_{TN}$ and $\hat{\rho}_{DW}$ equal each other to $0_p(T^{-1})$.

For $\hat{\rho}_D$ from the regression equation (26), we write those regression equations in matrix form as

$$y_{+1} = \rho y_{-1} + \gamma Z + \varepsilon \tag{116}$$

where y_{+1} and y_{-1} are $(T - 1) \times 1$ vectors equal to y with y_1 and y_T deleted respectively, Z is defined in (110) to (112), and

$$\varepsilon = (\varepsilon_2, \ldots, \varepsilon_T)' \tag{117}$$

where the ε_t's are from (6), (7), and (26). Now we can write the OLS estimate of ρ from (116) as

$$\hat{\rho}_D = y_{-1}' M_Z y_{+1} / y_{-1}' M_Z y_{-1} \tag{118}$$

from which one can show that

$$\begin{aligned}\hat{\rho}_D &= u_{-1}' M_Z u_{+1} / u_{-1}' M_Z u_{-1} \\ &= u' C_1' M_Z C u / u' C_1' M_Z C_1 u \end{aligned} \tag{119}$$

where u_{-1} and u_{+1} are defined in the same way as y_{-1} and y_{+1} of (116), and M_Z, C, and C_1 are from (109), (106), and (107), respectively. Using Nagar's method as in the derivation of the expansion for $\hat{\rho}_S$ of (113) and (114), we arrive at the result (97) for $\hat{\rho}_D$.

Since the iterated $\hat{\rho}$ estimates of the theorem, $\hat{\rho}_{ML}$ and $\hat{\rho}_D^I$, are functions of their respective iterated estimates of β, $\hat{\beta}_{ML}$ and $\hat{\beta}_D^I$, their expansions require expansions of the corresponding β estimate. In this instance we proceed by:

1) expressing the Aitken-type estimator $\hat{\beta}$ using <u>any</u> consistent estimator of ρ as a function of the terms of the expansion of that estimate of ρ (given in (120) for the general case) so that the $\hat{\beta}$ expansion in this general case is given by (125) to (128);

2) expressing the <u>particular</u> estimate of ρ in question as a function of the corresponding $\hat{\beta}$ from step (1) with which it is iterating, namely, $\hat{\rho}_S^I$ of (130) and (131) and $\hat{\rho}_{ML}$ of (133) and (134);

3) using the fact that we are considering a converged solution to equate the general expression for $\hat{\rho}$ in (120), which can be thought of as the ρ estimate after, say, i iterations, with the $\hat{\rho}$ expression of step (2), which can be thought of as the ρ estimate of the particular method after i + 1 iterations having used the $\hat{\rho}$ of (120) from the i^{th} iteration;

4) solving for the terms of the expansion of the converged $\hat{\rho}$ using the result from step (3).

Given <u>any</u> consistent estimate of ρ, say $\hat{\rho}$, and its expansion as

$$\hat{\rho} = \rho + \theta_{-1/2} + \theta_{-1} + o_p(T^{-1}) \tag{120}$$

where a term θ_i has $O(T^i)$, we can write the corresponding expansion for $\hat{\beta}(\hat{\rho})$, i.e., the Aitken-type estimator of β using (15) with $\hat{\rho}$ replacing ρ, by

$$\hat{\beta} = \beta + (X' \hat{\Sigma}^{-1} X)^{-1} X' \hat{\Sigma}^{-1} u \tag{121}$$

and from (38) we can write

$$\begin{aligned}(1-\rho^2)\hat{\Sigma}^{-1} = (1-\rho^2)\Sigma^{-1} &+ \theta_{-1/2} R \\ &+ (\theta_{-1} R + \theta^2_{-1/2} D) \\ &+ \text{"}o_p\text{"} (T^{-1})^6\end{aligned} \tag{122}$$

where

$$D = \text{diag}(0,1,\ldots,1,0), \quad \text{a } T \times T \text{ matrix} \tag{123}$$

and

$$R = 2(\rho D - A) \tag{124}$$

where A is defined in (102). Now by expanding $(X' \hat{\Sigma}^{-1} X)^{-1}$

around Ω of (17) we have a general expansion for some $\hat{\beta}$ based on a consistent ρ estimate from (120):

$$\hat{\beta} = \beta + \zeta_{-1/2} + \zeta_{-1} + \zeta_{-3/2} + o_p(T^{-3/2}) \quad (125)$$

where

$$\zeta_{-1/2} = \Omega X' \Sigma^{-1} u \quad (126)$$

$$\zeta_{-1} = \theta_{-1/2}\, \Omega X' R Q \Sigma^{-1} u/(1 - \rho^2) \quad (127)$$

$$\zeta_{-3/2} = \Omega X'(\theta_{-1} R + \theta^2_{-1/2} D) Q \Sigma^{-1} u/(1 - \rho^2)$$
$$- \theta^2_{-1/2}\, \Omega X' R X \Omega X' R Q \Sigma^{-1} u/(1-\rho^2)^2 \quad (128)$$

and ζ_i has $O_p(T^i)$.

Now for $\hat{\rho}^I_S$, expressed in expanded form as

$$\hat{\rho}^I_S = \rho + \theta^{SI}_{-1/2} + \theta^{SI}_{-1} + o_p(T^{-1}) \quad (129)$$

we have

$$\hat{\rho}^I_S = \tilde{u}' A \tilde{u} / \tilde{u}' \tilde{u} \quad (130)$$

where

$$\tilde{u} = y - X \hat{\beta}^I_S \quad (131)$$

and $\hat{\beta}_S^I$ is given to $0(T^{-3/2})$ by (125) to (128) with the θ's replaced by the θ^{SI}'s. We can now solve for the θ^{SI}'s by expanding $\hat{\rho}_S^I$ of (130) about ρ in a Taylor series and express it as a function of the θ^{SI}'s by using (131) along with the expansion for $\hat{\beta}_S^I$, which when applied to (131) gives

$$\tilde{u} = Mu + PQ\Sigma^{-1}u - \theta_{-1/2}^{SI} X\Omega X'RQ\Sigma^{-1}u/(1-\rho^2) \tag{132}$$

By substituting (132) into (130), expanding in a Taylor series about ρ, and equating the terms of appropriate order of (129) and (130), we arrive at the result for $\hat{\rho}_S^I$ of (95).

For $\hat{\rho}_{ML}$ we use (28) along with the expansion (125) for $\hat{\beta}_{ML}$ with the θ's replaced by θ^{ML}'s, which are the terms for which we need explicit solutions. Thus, we replace a, b, and c in (28) by their expressions of (29), (30), and (31) and with the $\hat{u}_t$'s replaced by the expansion as (132) with $\theta_{-1/2}^{SI}$ replaced by $\theta_{-1/2}^{ML}$, and replace $\hat{\rho}_{ML}$ in (28) by its expansion

$$\hat{\rho}_{ML} = \rho + \theta^{ML}_{-1/2} + \theta^{ML}_{-1} + o_p(T^{-1}) \tag{133}$$

Equation (28) then becomes, to $0_p(T)$, after multiplication by the common denominator of a, b, and c,

$$\begin{aligned} &\{\rho^3 + 3\rho^2\theta^{ML}_{-1/2} + (3\rho^2\theta^{ML}_{-1} + 3\rho(\theta^{ML}_{-1/2})^2\} \\ &\quad \{(T-1)\hat{u}'\hat{u} - T(\hat{u}_1^2 + \hat{u}_T^2)\} - \{\rho^2 + 2\rho\theta^{ML}_{-1/2} \\ &\quad + (2\rho\theta^{ML}_{-1} + (\theta^{ML}_{-1/2})^2\}(T-2)\hat{u}'A\hat{u} \\ &\quad - \{\rho + \theta^{ML}_{-1/2} + \theta^{ML}_{-1}\}\{(T+1)\hat{u}'\hat{u} - T(\hat{u}_1^2 + \hat{u}_T^2)\} \\ &\quad + T\hat{u}'A\hat{u} = 0 \end{aligned} \tag{134}$$

By setting $\theta^{ML}_{-1/2}$ and θ^{ML}_{-1} to satisfy (134) to $0_p(T)$ we obtain the result (96).

Q.E.D.

The results of Lemma 1 can be used to derive the MSE approximations, which are given in the following theorems.

Theorem 2. If the disturbances are symmetrically distributed then the Aitken-type estimators $\hat{\beta}_S$, $\hat{\beta}_{TH}$, $\hat{\beta}_{DW}$, $\hat{\beta}_{TN}$, $\hat{\beta}_S^I$, $\hat{\beta}_{ML}$, and $\hat{\beta}_D$ are unbiased if their mean vectors exist, and their mean square errors (MSE's) to $O(T^{-2})$ are equal under normality assumption (13), and are given by

$$MSE(\hat{\beta}) = \sigma^2\Omega + \{\sigma^2(1 - \rho^2)/T\rho^2\}\Omega X'QX\Omega + o(T^{-2}) \tag{135}$$

when $\rho \neq 0$, where Q is defined in (104).

Proof. Note from (121) and (122) that the stochastic part of the Aitken-type estimator is an odd function of u for the ρ estimates considered here, since the θ's are even functions of u in every case. Thus, when u is symmetrically distributed about zero we have

$$E(\hat{\beta} - \beta) = 0 \tag{136}$$

which proves the first part of the theorem.

For the result (135), we use the expansion (125) to give the general formula for the MSE of an Aitken-type estimator of

$$\text{MSE}(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = E(\zeta_{-1/2}\,\zeta_{-1/2}')$$

$$+ E(\zeta_{-1/2}\,\zeta_{-1}' + \zeta_{-1}\,\zeta_{-1/2}')$$

$$+ E(\zeta_{-1/2}\,\zeta_{-3/2}' + \zeta_{-3/2}\,\zeta_{-1/2}'$$

$$+ \zeta_{-1}\,\zeta_{-1}') + o(T^{-2}) \qquad (137)$$

The MSE for a particular Aitken-type estimator is then found by replacing the θ's in (127) and (128) by the particular expansion terms from the results of lemma 1, then taking the expectations in (137). The following expectations for u under the normality assumption (13) will be useful, and can be derived from results in a paper of Srivastava and Tiwari (1976) for any non-random $T \times T$ symmetric matrix G:

$$E(u'Gu \cdot uu') = (\text{tr}\,\Sigma G)\Sigma + 2\,\Sigma G \Sigma \qquad (138)$$

$$E\{(u'Gu)^2 uu'\} = \{(\text{tr}\,\Sigma G)^2 + 2\text{tr}\,\Sigma G \Sigma G\}\Sigma$$
$$+ 4(\text{tr}\,\Sigma G)\Sigma G \Sigma + 8\,\Sigma G \Sigma G \Sigma \qquad (139)$$

Since $\zeta_{-1/2}$ of (126) is not a function of the θ's we know that for any Aitken-type estimator which uses a consistent ρ estimate,

$$E\,\zeta_{-1/2}\,\zeta_{-1/2}' = \Omega X'\textstyle\sum^{-1}(Euu')\textstyle\sum^{-1}X\Omega$$

$$= \sigma^2\Omega \qquad (140)$$

Next, considering the particular estimate $\hat{\beta}_S$, we use the expansion for $\hat{\rho}_S$ of (91), substitute these θ's into (127) and (128), and calculate the expectations (137). Denoting ζ_{-1} and $\zeta_{-3/2}$ in the $\hat{\beta}_S$ case as ζ^S_{-1} and $\zeta^S_{-3/2}$, and using

$$MX = 0 \text{ and } Q\textstyle\sum^{-1}X = 0 \qquad (141)$$

we find that

$$E(\zeta^S_{-1}\,\zeta^{S'}_{-1/2}) = E(\zeta^S_{-1/2}\,\zeta^{S'}_{-1}) = E(\zeta^S_{-3/2}\,\zeta^{S'}_{-1/2})$$

$$= E(\zeta^S_{-1/2}\,\zeta^{S'}_{-3/2}) = 0 \qquad (142)$$

Using (139) and eliminating higher order terms we have

$$E(\zeta_{-1}^{S}\ \zeta_{-1}^{S'}) = \{2\sigma^2(\mathrm{tr}\Sigma MBM\Sigma MBM)/T^2(1-\rho^2)^2\}$$

$$\Omega X' RQ\Sigma^{-1} QRX\Omega \qquad (143)$$

This can be simplified by noting that

$$\mathrm{tr}\Sigma MBM\Sigma MBM = \mathrm{tr}\Sigma B\Sigma B + o(T) \qquad (144)$$

where B is from (105). Next we note that

$$\mathrm{tr}\Sigma^2 = T(1+\rho^2)/(1-\rho^2) + o(T);$$

$$\mathrm{tr}\Sigma A\Sigma = 2\rho T/(1-\rho^2) + o(T);$$

$$\mathrm{tr}\Sigma A\Sigma A = T\{(1+\rho^2)/(1-\rho^2) - (1-\rho^2)/2\} + o(T)$$

$$(145)$$

which yields

$$\mathrm{tr}\Sigma MBM\Sigma MBM = T(1-\rho^2)/2 + o(T) \qquad (146)$$

In addition, we can use (141) and

$$Q\Sigma^{-1}Q = Q \qquad (147)$$

$$R = \{(1 - \rho^2)\textstyle\sum^{-1} + J\}/\rho \qquad (148)$$

$$J = I - \rho^2 D \qquad (149)$$

to show that

$$\Omega X' RQ\textstyle\sum^{-1} QRX\Omega = \{(1 - \rho^2)^2/\rho^2\}\Omega X' QX\Omega$$

$$+ o(T^{-1}) \qquad (150)$$

Combining (146), (150), and (143) to simplify $E\zeta^S_{-1}\zeta^{S'}_{-1}$, and substituting this along with (142) and (140) into the general MSE formula (137), we obtain the result (135) for $\hat{\beta}_S$.

For $\hat{\beta}_{TH}$, $\hat{\beta}_{DW}$, $\hat{\beta}_{TN}$, $\hat{\beta}^I_S$, and $\hat{\beta}_{ML}$, observe from (92) to (96) that the corresponding $\hat{\rho}$ estimate differs from $\hat{\rho}_S$ by a term of $0_p(T^{-1})$. From (125) it is seen that the resulting $\hat{\beta}$ estimate is only affected at $0_p(T^{-3/2})$ and higher. Thus, the resulting MSE of any of these $\hat{\beta}$'s can only differ from $\hat{\beta}_S$ in the $E(\zeta_{-1/2}\zeta'_{-3/2})$ or $E(\zeta_{-3/2}\zeta'_{-1/2})$ terms, particularly in the term of $\zeta_{-3/2}$ involving θ_{-1}.

TABLE 2.

Efficiency Measures of Estimators for M2

SAMPLE SIZE: 10

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.007	1.286	1.448	1.003	1.130	1.185
.10	1.026	1.297	1.462	1.013	1.123	1.175
.20	1.113	1.325	1.492	1.054	1.111	1.156
.30	1.284	1.356	1.522	1.125	1.106	1.146
.40	1.594	1.381	1.539	1.241	1.108	1.144
.50	2.170	1.383	1.528	1.428	1.112	1.143
.60	3.322	1.348	1.472	1.745	1.107	1.135
.70	5.923	1.266	1.363	2.329	1.087	1.110
.80	13.064	1.154	1.219	3.582	1.053	1.069
.90	42.226	1.048	1.080	7.475	1.017	1.026
.95	110.978	1.013	1.029	15.309	1.005	1.009

SAMPLE SIZE: 40

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.003	1.033	1.072	1.002	1.023	1.038
.10	1.012	1.039	1.081	1.008	1.026	1.043
.20	1.053	1.054	1.102	1.036	1.034	1.052
.30	1.137	1.072	1.127	1.089	1.042	1.061
.40	1.292	1.093	1.155	1,180	1.048	1.067
.50	1.584	1.117	1.184	1.329	1.052	1.071
.60	2.180	1.144	1.214	1.576	1.054	1.071
.70	3.626	1.168	1.235	2.017	1.056	1.072
.80	8.519	1.162	1.217	2.978	1.054	1.067
.90	42.269	1.091	1.121	6.381	1.032	1.040
.95	181.808	1.035	1.051	14.026	1.012	1.017

the case where G is non-symmetric:

$$E(u'Gu \cdot uu') = (\mathrm{tr}\,\Sigma G)\Sigma + \Sigma(G + G')\Sigma \qquad (154)$$

and

$$E\{(u'Gu)^2 uu'\} = \{(\mathrm{tr}\,\Sigma G)^2 + \mathrm{tr}\,\Sigma G\Sigma(G + G')\}\Sigma$$

$$+ 2(\mathrm{tr}\,\Sigma G)\Sigma(G + G')\Sigma$$

$$+ 2\Sigma(G + G')\Sigma(G + G')\Sigma \qquad (155)$$

The result (151) can be simplified by dropping terms of higher order. In addition, it can be shown that

$$\mathrm{tr}\,N = 0 + o(T) \qquad (156)$$

$$\mathrm{tr}\,N^2 + \mathrm{tr}\,\Sigma N_1' N = 2\mathrm{tr}\,\Sigma B \Sigma B + o(T)$$

$$= T(1 - \rho^2) + o(T) \qquad (157)$$

and

$$\Omega X' JQJX\Omega = (1 - \rho^2)^2 \Omega X' QX\Omega + o(T^{-1}) \qquad (158)$$

so we can simplify (151) to yield (135) for $\hat{\beta}_D$.

Q.E.D.

Corollary. The MSE of $\hat{\beta}_S$ and the other estimators of theorem 2 is greater than or equal to the MSE of the GLS estimator $\beta(\rho)$ of (15) in the sense that the former exceeds the latter by a positive semidefinite matrix to $O(T^{-2})$.

Proof. The MSE of $\hat{\beta}_S$ exceeds the MSE of $\hat{\beta}(\rho)$ to $O(T^{-2})$ by a positive scalar times $\Omega X' Q X \Omega$, thus the corollary is proved if Q is positive semidefinite. To see this, note that $Q = Q\Sigma^{-1}Q$ from (147), and that $\Sigma^{-1} = Q^{**'}Q^{**}/(1 - \rho^2)$ from (39) where Q^{**} is from (40). We then have

$$Q = (Q^{**}Q)'(Q^{**}Q)/(1 - \rho^2) \tag{159}$$

from which it is clear that Q is positive semidefinite

Q.E.D.

Next, the MSE of the two-step Cochrane-Orcutt type estimator $\tilde{\beta}_S$ which results from using (34) with $\hat{\rho}_S$ of (48) replacing ρ will be considered.

Theorem 3. *If the disturbances are symmetrically distributed then the Cochrane-Orcutt type estimator* $\tilde{\beta}_S$ *is unbiased if its mean vector exists, and its* MSE *to* $0(T^{-2})$ *under normality assumption* (13) *is given by*

$$MSE(\tilde{\beta}_S) = MSE(\hat{\beta}_S) + \sigma^2 \Omega X' D_0 X \Omega + o(T^{-2}) \qquad (160)$$

where $MSE(\hat{\beta}_S)$ *is given by* (135) *and* D_0 *is defined in* (44).

Proof. We can expand $\tilde{\beta}_S$ to $0(T^{-3/2})$ in the same way as was done for the general $\hat{\beta}$ in (125) by using $\tilde{\beta}_S$ as in (45) and the expansion for $\hat{\Sigma}^{-1}$ of (122). This yields

$$\tilde{\beta}_S \doteq \beta + \eta_{-1/2} + \eta_{-1} + \eta_{-3/2} + o_p(T^{-3/2}) \qquad (161)$$

where

$$\eta_{-1/2} = \zeta_{-1/2} \qquad (162)$$

$$\eta_{-1} \doteq \zeta_{-1} - \Omega X' D_0 u \qquad (163)$$

$$\eta_{-3/2} = \zeta_{-3/2} + \Omega X'[\theta_{-1/2}\{2\rho + RX\Omega X'\}D_0 u/(1-\rho^2)$$

$$+ D_0 X\Omega X' \Sigma^{-1} u] \qquad (164)$$

and the ζ's are defined in (126) to (128). Now, taking the expectation of $(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'$ as in (137) and deleting higher order terms, we obtain the result (160).

Q.E.D.

Corollary. The MSE of $\hat{\beta}_S$ to $0(T^{-2})$ is less than or equal to the MSE of $\tilde{\beta}_S$ to $0(T^{-2})$ in the sense that the latter exceeds the former by a positive semidefinite matrix.

Proof. This follows from comparing (135) and (160) and noting that $\Omega X' D_0 X\Omega$ is positive semidefinite since D_0 is positive semidefinite.

Q.E.D.

(iii) Numerical Experiment

In this section the performances of four estimators are compared by evaluating the approximate MSE's of the last section at particular ρ and X values[7]. The estimators are

1. the OLS estimator b of (89) whose exact MSE is given in (90),

2. the GLS estimator $\hat{\beta}(\rho)$ of (15) whose exact MSE is given by (16),

3. the two-step Prais-Winsten (PW) estimator $\hat{\beta}_S$ of (21) whose approximate MSE to $0(T^{-2})$ is given by (135) of theorem 2, and

4. the two-step Cochrane-Orcutt (CO) estimator $\tilde{\beta}_S$ of (34) where $\hat{\rho}_S$ replaces ρ whose approximate MSE to $0(T^{-2})$ is given by (160) of theorem 3.

The GLS estimator is not operational since ρ is unknown in practice, but it will be used as a benchmark here, since it is known to be BLUE. The comparisons are done by taking the traces and determinants of the MSE matrices of the operational estimators b, $\hat{\beta}_S$ and $\tilde{\beta}_S$ and dividing each by the trace or determinant of the MSE matrix of $\hat{\beta}(\rho)$.

Since $\hat{\beta}(\rho)$ is BLUE, we already know that the resulting ratio corresponding to b will exceed one due to properties of traces and determinants. It follows from the corollaries to theorems 2 and 3 that the MSE ratio for $\hat{\beta}_S$ will exceed one, but will be less than the MSE ratio of $\tilde{\beta}_S$. Nevertheless, the numerical comparison will shed some light on the relative merits of b vs. $\hat{\beta}_S$ and $\tilde{\beta}_S$, and also on situations where $\tilde{\beta}_S$ will perform particularly poorly.

Since many other estimators were found to have MSE's equal to that of $\hat{\beta}_S$ to $0(T^{-2})$ in theorem 2, the results given below for $\hat{\beta}_S$ can be taken as results for those estimators as well.

Four models or specifications of X, are considered, with sample sizes of 10 and 40 for each model and a range of positive ρ values considered for each model and sample size. The models for the T = 10 case are

$$M_1 : X = [1 \quad x_1 \quad x_2] \tag{165}$$

$$M_2 : X = [x_1 \quad x_2] \tag{166}$$

$$M_3 : X = [1 \quad x_3] \tag{167}$$

$$M_4 : X = [x_4 \quad x_5 \quad x_6] \tag{168}$$

where X is the matrix of exogenous variables of model (1) and

$$x_1' = [1.723, \ .022, \ 1.157, \ .504, \ 2.832, \ .902, \ .853, \ 1.816, \ 2.898, \ 1.019]$$

$$x_2' = [.482, \ 1.376, \ 1.01, \ .005, \ 1.393, \ 1.787, \ .105, \ 1.339, \ 1.041, \ .279]$$

$$x_3' = [1.809, \ 2.309, \ 2.691, \ 3.191, \ 4.0, \ 5.191, \ 6.691, \ 8.309, \ 9.809, \ 11.0]$$

and

$$x_4 = (\lambda_1 + \lambda_T)/2^{1/2}, \quad x_5 = (\lambda_2 + \lambda_{T-1})/2^{1/2},$$

$$x_6 = (\lambda_3 + \lambda_{T-2})/2^{1/2}$$

and λ_i refers to the normalized characteristic vector corresponding to the i^{th} largest characteristic root of a certain approximation to $\sum$, call it $\sum^*$, where

$$\sum\nolimits^{*-1} = \sum\nolimits^{-1} - \{\rho(1 - \rho)/(1 - \rho^2)\}D_1 \tag{169}$$

and D_1 is a T × T matrix of zeroes with a one in the upper left and lower right hand corner (see Tillman (1975, p. 960)).

M4 is chosen because the X matrix of (167) is known as "Watson's X" (Watson (1967)) and is known to be a model in which OLS is particularly poor. The characteristic vectors λ_i can be easily derived using formulas given in Tillman (1975, p. 965)[8]. M1, M2, and M3 are chosen to correspond with the vectors used in a study by Raj, Srivastava and Upadhyaya (1980). Note that x_1 and x_2 are non-trended while x_3 is strongly trended.

For the T = 40 case in M1, M2 and M3, the models given above were modified by using each observation from the x vectors four consecutive times. For example, when T = 40,

$$x_1' = [1.723, 1.723, 1.723, 1.723, .022, .022, \ldots, 1.019, 1.019]$$

This approach was used rather than appending complete vectors so that the "shape" of each vector (particularly the trend in x_3) was preserved while sample size was increased. For M4 the λ_i's corresponding to Σ^* when T = 40 are used.

The notation for the tables is:

$$e_1 = \det(b)/\det(\hat{\beta}(\rho))$$

$$e_2 = \det(\hat{\beta}_S)/\det(\hat{\beta}(\rho))$$

$$e_3 = \det(\tilde{\beta}_S)/\det(\hat{\beta}(\rho))$$

$$e_1^* = \mathrm{tr}(b)/\mathrm{tr}(\hat{\beta}(\rho))$$

$$e_2^* = \mathrm{tr}(\hat{\beta}_S)/\mathrm{tr}(\hat{\beta}(\rho))$$

$$e_3^* = \mathrm{tr}(\tilde{\beta}_S)/\mathrm{tr}(\hat{\beta}(\rho)) \qquad (170)$$

where det(b) and tr(b), for example, represent the determinant and trace, respectively, of the variance-covariance (MSE) matrix of b, the OLS estimate of β.

TABLE 1.

Efficiency Measures of Estimators for M1

SAMPLE SIZE: 10

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.005	1.215	1.431	1.002	1.066	1.112
.10	1.021	1.199	1.419	1.007	1.060	1.106
.20	1.084	1.167	1.397	1.026	1.047	1.097
.30	1.193	1.139	1.384	1.055	1.036	1.095
.40	1.360	1.121	1.386	1.093	1.030	1.105
.50	1.607	1.113	1.408	1.134	1.032	1.138
.60	1.971	1.115	1.455	1.170	1.045	1.204
.70	2.505	1.122	1.528	1.192	1.068	1.315
.80	3.269	1.118	1.628	1.182	1.088	1.477
.90	4.292	1.075	1.756	1.124	1.067	1.684
.95	4.874	1.032	1.845	1.072	1.030	1.811

SAMPLE SIZE: 40

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.004	1.052	1.102	1.002	1.022	1.035
.10	1.019	1.060	1.115	1.008	1.026	1.039
.20	1.082	1.077	1.142	1.034	1.032	1.047
.30	1.209	1.093	1.170	1.083	1.037	1.055
.40	1.440	1.104	1.193	1.162	1.039	1.060
.50	1.857	1.105	1.209	1.275	1.036	1.062
.60	2.631	1.094	1.217	1.423	1.028	1.063
.70	4.162	1.078	1.228	1.581	1.022	1.078
.80	7.603	1.075	1.278	1.662	1.031	1.145
.90	17.807	1.116	1.450	1.511	1.091	1.365
.95	31.244	1.132	1.634	1.313	1.122	1.591

TABLE 2.

Efficiency Measures of Estimators for M2

SAMPLE SIZE: 10

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.007	1.286	1.448	1.003	1.130	1.185
.10	1.026	1.297	1.462	1.013	1.123	1.175
.20	1.113	1.325	1.492	1.054	1.111	1.156
.30	1.284	1.356	1.522	1.125	1.106	1.146
.40	1.594	1.381	1.539	1.241	1.108	1.144
.50	2.170	1.383	1.528	1.428	1.112	1.143
.60	3.322	1.348	1.472	1.745	1.107	1.135
.70	5.923	1.266	1.363	2.329	1.087	1.110
.80	13.064	1.154	1.219	3.582	1.053	1.069
.90	42.226	1.048	1.080	7.475	1.017	1.026
.95	110.978	1.013	1.029	15.309	1.005	1.009

SAMPLE SIZE: 40

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.003	1.033	1.072	1.002	1.023	1.038
.10	1.012	1.039	1.081	1.008	1.026	1.043
.20	1.053	1.054	1.102	1.036	1.034	1.052
.30	1.137	1.072	1.127	1.089	1.042	1.061
.40	1.292	1.093	1.155	1,180	1.048	1.067
.50	1.584	1.117	1.184	1.329	1.052	1.071
.60	2.180	1.144	1.214	1.576	1.054	1.071
.70	3.626	1.168	1.235	2.017	1.056	1.072
.80	8.519	1.162	1.217	2.978	1.054	1.067
.90	42.269	1.091	1.121	6.381	1.032	1.040
.95	181.808	1.035	1.051	14.026	1.012	1.017

TABLE 3.

Efficiency Measures of Estimators for M3

SAMPLE SIZE: 10

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.001	1.055	1.323	1.001	1.028	1.277
.10	1.005	1.063	1.354	1.002	1.032	1.301
.20	1.018	1.079	1.423	1.009	1.039	1.354
.30	1.040	1.098	1.504	1.019	1.046	1.413
.40	1.069	1.118	1.595	1.032	1.051	1.479
.50	1.106	1.137	1.697	1.046	1.055	1.552
.60	1.146	1.150	1.806	1.057	1.055	1.633
.70	1.184	1.152	1.910	1.063	1.053	1.722
.80	1.209	1.133	1.988	1.058	1.050	1.821
.90	1.201	1.075	2.007	1.039	1.040	1.918
.95	1.176	1.030	1.994	1.023	1.021	1.956

SAMPLE SIZE: 40

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.000	1.006	1.072	1.000	1.004	1.067
.10	1.002	1.007	1.080	1.001	1.004	1.074
.20	1.009	1.011	1.099	1.005	1.006	1.090
.30	1.021	1.016	1.124	1.012	1.010	1.112
.40	1.041	1.025	1.158	1.025	1.015	1.140
.50	1.075	1.039	1.205	1.045	1.023	1.179
.60	1.133	1.062	1.274	1.080	1.037	1.234
.70	1.242	1.102	1.382	1.142	1.059	1.314
.80	1.482	1.166	1.550	1.267	1.090	1.430
.90	2.165	1.216	1.757	1.512	1.110	1.583
.95	3.016	1.169	1.820	1.584	1.105	1.706

TABLE 4.

Efficiency of Estimators for M4

SAMPLE SIZE: 10

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.020	1.980	2.686	1.007	1.258	1.401
.10	1.082	1.950	2.635	1.027	1.252	1.390
.20	1.370	1.839	2.453	1.111	1.228	1.354
.30	2.025	1.681	2.203	1.268	1.191	1.304
.40	3.472	1.509	1.931	1.527	1.149	1.247
.50	6.859	1.349	1.674	1.942	1.106	1.188
.60	15.542	1.215	1.455	2.626	1.067	1.134
.70	40.626	1.115	1.282	3.838	1.037	1.086
.80	126.944	1.048	1.152	6.354	1.015	1.048
.90	552.844	1.011	1.061	14.049	1.004	1.020
.95	(*)	1.003	1.027	29.526	1.001	1.009

SAMPLE SIZE: 40

ρ	e_1	e_2	e_3	e_1^*	e_2^*	e_3^*
.05	1.029	1.309	1.397	1.010	1.094	1.118
.10	1.121	1.300	1.387	1.039	1.092	1.115
.20	1.576	1.270	1.350	1.164	1.083	1.105
.30	2.785	1.228	1.299	1.407	1.071	1.091
.40	6.209	1.181	1.241	1.839	1.057	1.075
.50	17.733	1.135	1.183	2.610	1.043	1.058
.60	67.173	1.093	1.130	4.075	1.030	1.042
.70	362.582	1.057	1.084	7.180	1.019	1.027
.80	(*)	1.029	1.046	15.105	1.010	1.015
.90	(*)	1.009	1.016	44.708	1.003	1.005
.95	(*)	1.002	1.006	109.573	1.001	1.002

(*) indicates a number exceeding 1,000

Some remarks follow:

1) OLS is better than PW for small values of ρ; roughly for ρ values less than 0.2 or 0.3, while PW is better for higher ρ values. This agrees with Monte Carlo findings of Griliches and Rao (1968) and Spitzer (1979).

2) The performance of OLS for large ρ values is particularly bad when there is no constant term (M2 and M4) while PW and CO become almost as good as GLS in those cases. These remarks hold true especially in the Watson's X (M4) case. The OLS result agrees with an analytical finding of Krämer (1980).

3) CO performs worse in models where there is a constant than in those where there is not. It is particularly poor in M3 where there is a constant and a trended variable which supports a result of Maeshiro (1976).

4) The fact that the approximate MSE method used here cannot distinguish between the two stage PW, iterative PW, or ML estimators is not a severe shortcoming since these estimators typically give similar results in Monte Carlo studies. The case in which they tend to differ most in those studies is when ρ is close to one (e.g., Harvey and McAvinchey (1978)). This is also the case in

which the MSE approximation method used here is likely to be least effective since as $\sum$ approaches singularity the deleted terms of the expansions become more important.

III.5 AN EDGEWORTH EXPANSION FOR DISTRIBUTION OF A TEST FOR $\beta_1 = 0$

Park and Mitchell (1980) note that the conventional tests for the significance of an element of the coefficient vector β of (4) tend to reject the null when it is true far too often (that is, the type I error is actually much larger than the stated value which is based on asymptotic results). The test statistic which will be considered here is the following:

$$W = (\hat{\beta}_S)_1^2 / s^2 \, [(X' \hat{\Sigma}_S^{-1} X)^{-1}]_{11} \tag{171}$$

where

$$s^2 = e'e/(T - k) \tag{172}$$

where e is the OLS residual vector from (19), $(\hat{\beta}_S)_1$ refers to the first element of $\hat{\beta}_S$, and $[(X' \hat{\Sigma}_S^{-1} X)^{-1}]_{11}$ refers to the upper-left hand corner element of $(X' \hat{\Sigma}_S^{-1} X)^{-1}$, where $\hat{\Sigma}_S$ is formed by replacing ρ in Σ of (14) by $\hat{\rho}_S$ of (18). The statistic is denoted by W since it is a variant of the Wald statistic and in fact, is a Wald statistic, using a more general definition given by Stroud (1971), and shares it asymptotic properties.

The W statistic (171) uses the two-stage Prais-Winsten estimator $\hat{\beta}_S$. Park and Mitchell (1980) also consider the iterated Prais-Winsten estimator $\hat{\beta}_S^I$ and the maximum likelihood estimator $\hat{\beta}_{ML}$ and decide that $\hat{\beta}_S^I$ is the best, although their Table 6 indicates the three resulting statistics give very similar results for reasonably low ρ values in a Monte Carlo study. We have decided to use the $\hat{\beta}_S$ variant here due to its analytic simplicity as well as the similarity of the Monte Carlo results mentioned above and the equivalence to $0(T^{-2})$ of the MSE's of the three estimators demonstrated in theorem 2. Additionally, the variance estimator s^2 in (171) has replaced the corresponding estimate using the second-stage Prais-Winsten residual vector. This substitution is made because the expansion resulting from W as in (171) is even more cumbersome than that of theorem 4 which follows. Since the motivation behind this expansion is to derive the Edgeworth correction factor for testing (174), any additional bias in the test resulting from the use of s^2

instead of σ_S^2 should be erased to $0(T^{-1})$ by the correction factor. It seems unlikely that this substitution of variance estimators would have much effect on the powers of the size corrected test. For notational simplicity the S subscripts of (171) will be dropped for the remainder of this section, so that (171), along with the variance estimator substitution, can be re-written as

$$W = \hat{\beta}_1^2/s^2[(X' \hat{\Sigma}^{-1} X)^{-1}]_{11} \qquad (173)$$

(i) <u>The Expansion</u>

In this section we present an Edgeworth expansion for the distribution of W in (173) to $0(T^{-1})$ for testing the hypotheses:

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0 \qquad (174)$$

under the distributional assumption (13) and the alternative:

$$\beta_1 = \varepsilon/T^{1/2} \qquad (175)$$

Theorem 4. The Edgeworth expansion for the distribution of W of (173) to $O(T^{-1})$ under local alternative (175) is given by

$$pr(W \leq x) = pr(\chi_1^2(\delta) \leq x) + T^{-1} \sum_{i=0}^{4} \tau_i \, pr(\chi_{1+2i}^2(\delta) \leq x) + o(T^{-1}) \qquad (176)$$

where

$$\delta = \omega_{11}^{-1} \gamma \qquad (177)$$

is the non-centrality parameter, ω_{11} is the $(1,1)^{th}$ element of Ω of (17),

$$\gamma = \varepsilon^2/T\sigma^2 \qquad (178)$$

where ε is a non-stochastic pre-specified scalar,

$$\tau_0 = -\omega_{11}\phi_1/2 + 3\omega_{11}\phi_2/4 - (\omega_{11} - \gamma)\phi_3/2$$

$$\tau_1 = (\omega_{11} - \gamma)\phi_1/2 - 3(\omega_{11} - \gamma)\phi_2/2$$

$$+ (\omega_{11} - 2\gamma)\phi_3/2,$$

$$\tau_2 = \gamma\phi_1/2 + (3\omega_{11} - 12\gamma + \omega_{11}^{-1}\gamma^2)\phi_2/4 + \gamma\phi_3/2,$$

$$\tau_3 = (3\gamma - \omega_{11}^{-1}\gamma^2)\phi_2/2,$$

$$\tau_4 = \omega_{11}^{-1}\gamma^2\phi_2/4 \tag{179}$$

$$\phi_1 = T^{-1}\{\omega_{11}^{-1}(\mathrm{tr}\textstyle\sum P + (6\rho^2 - 2)/(1 - \rho^2) - k$$

$$+ \mathrm{tr}\textstyle\sum MBM/\rho) + \omega_{11}^{-2}\,\Omega_1.X'X\Omega._1(5 - \mathrm{tr}\textstyle\sum MBM/\rho)$$

$$+ \omega_{11}^{-2}(1-\rho^2)\Omega_1.X'X_G(X_G'\textstyle\sum^{-1}X_G)^{-1}X_G'\,X\Omega._1/\rho^2\} \tag{180}$$

$$\phi_2 = T^{-1}[\omega_{11}^{-1}\{(-3 + 4\rho^2 + 3\rho^4)/\rho^2(1 - \rho^2)\}$$

$$+ \omega_{11}^{-2}\,\Omega_1.X'X\Omega._1(6\rho^2 - 2)/\rho^2$$

$$+ \omega_{11}^{-3}(\Omega_1.X'X\Omega._1)^2(1 - \rho^2)/\rho^2] \tag{181}$$

$$\phi_3 = T^{-1} \omega_{11}^{-2} \Omega_{1}.X' RQRX\Omega._{1}/(1 - \rho^2) \tag{182}$$

$$P = X(X'X)^{-1}X' \tag{183}$$

X_G _is the_ $T \times (k - 1)$ _matrix consisting of_ X _with the first column deleted,_ $\Omega_1.$ _is the first row of_ Ω _and_ $\Omega._1$ _its first column,_ Q _is defined in_ (104), B _in_ (105), R _in_ (124), _and_ M _in_ (20).

Proof. This proof uses the same expansion technique as the proof of theorem 2 of chapter II. First, we require an expansion to $0(T^{-1})$ of W of (173). The approach used here will be to express the expansion of W as a product of the expansions of $\hat{\beta}_1^2$, $(s^2)^{-1}$, and $[(X' \hat{\Sigma}^{-1} X)^{-1}]_{11}^{-1}$. For $\hat{\beta}_1^2$ we can use the expansion for the $\hat{\beta}$ vector given in (125), substituting in $\theta_{-1/2}^{S}$ and θ_{-1}^{S} of (98) and (99) to give

$$\hat{\beta}_1^2 = b_{-1} + b_{-3/2} + b_{-2} + o_p(T^{-2}) \tag{184}$$

where

$$b_{-1} = (\zeta_{-1/2,1} + \varepsilon/T^{1/2})^2,$$

$$b_{-3/2} = 2\zeta_{-1,1}(\zeta_{-1/2,1} + \varepsilon/T^{1/2}),$$

$$b_{-2} = 2\zeta_{-3/2,1}(\zeta_{-1/2,1} + \varepsilon/T^{1/2}) + \zeta^2_{-1,1}$$

(185)

and

$$\zeta_{-1/2,1} = \Omega_{1}.X'\Sigma^{-1}u;$$

$$\zeta_{-1,1} = \theta_{-1/2}\Omega_{1}.X'RQ\Sigma^{-1}u/(1 - \rho^2),$$

$$\zeta_{-3/2,1} = \Omega_{1}.X'(\theta_{-1}R + \theta^2_{-1/2}D)Q\Sigma^{-1}u/(1 - \rho^2)$$
$$- \theta^2_{-1/2}\Omega_{1}.X'RX\Omega X'RQ\Sigma^{-1}u/(1 - \rho^2)^2$$

(186)

where the $\zeta_{i,1}$'s are the first elements of the ζ vectors of (126), (127), and (128), and the S subscripts have been dropped from the θ's

For expanding $(s^2)^{-1}$, we can use

$$s^2 = \sigma^2 + (u'Mu/T - \sigma^2) + k\,u'Mu/T^2 + o_p(T^{-1}),$$

(187)

and expand $(s^2)^{-1}$ about $(\sigma^2)^{-1}$ using Nagar's method to give

$$(s^2)^{-1} = c_0 + c_{-1/2} + c_{-1} + o_p(T^{-1}) \tag{188}$$

where

$$c_0 = 1/\sigma^2;\; c_{-1/2} = -(u'Mu/T\sigma^2 - 1)/\sigma^2$$

$$c_{-1} = (u'Mu/T\sigma^2 - 1)^2/\sigma^2 - ku'Mu/T^2\sigma^4 \tag{189}$$

Finally, we note that

$$[(X'\hat{\Sigma}^{-1}X)^{-1}]_{11}^{-1} = X_1'\hat{\Sigma}^{-1}X_1 - X_1'\hat{\Sigma}^{-1}X_G(X_G'\hat{\Sigma}^{-1}X_G)^{-1} X_G'\hat{\Sigma}^{-1}X_1 \tag{190}$$

having partitioned X into T × 1 and T × (k-1) submatrices X_1 and X_G, respectively. Now we can use the expansion (122) to expand (190) which after some simplification yields

$$[(X'\hat{\Sigma}^{-1}X)^{-1}]_{11}^{-1} = a_1 + a_{1/2} + a_0 + o_p(T^0) \tag{191}$$

where

$$a_1 = \omega_{11}^{-1}; \quad a_{1/2} = \theta_{-1/2}X_{1.G}'RX_{1.G}/(1 - \rho^2)$$

$$a_0 = \theta_{-1/2}^2 h/(1 - \rho^2) + \theta_{-1}X_{1.G}'RX_{1.G} \tag{192}$$

and

$$X_{1.G} = X_1 - X_G(X_G'\Sigma^{-1} X_G)^{-1}X_G'\Sigma^{-1}X_1 \tag{193}$$

$$h = X_{1.G}'DX_{1.G} - X_{1.G}' RX_G(X_G'\Sigma^{-1}X_G)^{-1}X_G'RX_{1.G} \tag{194}$$

Using these expansions (184), (188) and (191) we can write the expansion for W as

$$W = \eta_0 + \eta_{-1/2} + \eta_{-1} + o_p(T^{-1}) \tag{195}$$

where

$$\eta_0 = a_1 b_{-1}c_0; \quad \eta_{-1/2} = a_1(b_{-1}c_{-1/2} + b_{-3/2}c_0) + a_{1/2}b_{-1}c_0$$

$$\eta_{-1} = a_1(b_{-1}c_{-1} + b_{-3/2}c_{-1/2} + b_{-2}c_0) + a_{1/2}(b_{-1}c_{-1/2} + b_{-3/2}c_0) + a_0 b_{-1}c_0 \tag{196}$$

As in (93) of chapter II, we require the moment generating function (m.g.f.) of W to $0(T^{-1})$ which is given by

$$M_W(t) = E\{\exp(tW)\}$$

$$= E[\{\exp(t\eta_0)\}(1 + t\eta_{-1/2} + t\eta_{-1} + t^2\eta^2_{-1/2}/2)] + o(T^{-1}) \quad (197)$$

We first require

$$E \exp t\eta_0 = \int_u \exp\{t\ \omega_{11}^{-1}(\zeta_{-1/2,1} + \varepsilon/T^{1/2})^2/\sigma^2\}f(u)du \quad (198)$$

where f(u) is the density function for the disturbances resulting from the distributional assumption (13), which is

$$f(u) = (2\Pi\sigma^2)^{-T/2}|\Sigma|^{-1/2} \exp(-u'\Sigma^{-1}u/2\sigma^2) \quad (199)$$

Make the following transformation of u:

$$H^{1/2}z = X'\Sigma^{-1}u - \{2t\ \varepsilon\omega_{11}^{-1}/T^{1/2}(1 - 2t)\}i^{*}_{1,k} \tag{200}$$

$$x = Cu \tag{201}$$

where

$$H^{-1} = \Omega - 2t\ \omega_{11}^{-1}\ \Omega_{\cdot 1}\Omega_{1\cdot}\ ;$$

$$H = \Omega^{-1} + (2t/(1 - 2t))\omega_{11}^{-1}\ I^{*}_{1,k} \tag{202}$$

C is a $(T - k) \times T$ matrix which satisfies

$$C'C = M; \quad CC' = I_{T-k}; \quad CX = 0 \tag{203}$$

(for example, the matrix used by Theil (1971, p.206) in the construction of BLUS residuals), $i^{*}_{1,k}$ is a $k \times 1$ vector of zeroes with a one in the first element, and $I^{*}_{1,k}$ is a $k \times k$ matrix of zeroes with a one in the upper left hand corner, and z and x are $k \times 1$ and $(T - k) \times 1$ stochastic vectors, respectively. Furthermore, x and z are independently distributed.

By completing the square in (198) as was done in the proof of theorem 2, chapter II we find that

$$E \exp t\eta_0 = \exp\{t(1 - 2t)^{-1}\delta\}(1 - 2t)^{-1/2}$$

$$\int_z f_z(z)\,dz \int_x f_x(x)\,dx \qquad (204)$$

where $f_z(z)$ and $f_x(x)$ are p.d.f.'s as if

$$z \sim N(0, \sigma^2 I) \text{ and } x \sim N(0, \sigma^2 C \Sigma C') \qquad (205)$$

so we have

$$E \exp t\eta_0 = \exp\{t(1 - 2t)^{-1}\delta\}(1 - 2t)^{-1/2} \qquad (206)$$

where δ is from (177). This is the m.g.f. for a non-central $\chi_1^2(\delta)$ which is the asymptotic distribution of W under local alternative (175). In obtaining (205) we have used

$$|H^{1/2}| = |\Omega|^{-1/2}(1 - 2t)^{-1/2} \qquad (207)$$

The higher-order terms in the m.g.f. (197) are now obtained by transforming the η's from functions of u to functions of x and z and then integrating over those variables weighted by the f_x and f_z functions. We will again use $\tilde{E}$ to denote this integration as in chapter II. The transformations use the following relations:

$$u = X\Omega H^{1/2}z + \Sigma C'(C\Sigma C')^{-1}x$$

$$+ \{2\varepsilon\ t\omega_{11}^{-1}/T^{1/2}(1 - 2t)\}X\Omega_{.1}$$

$$\zeta_{-1/2,1} + \varepsilon/T^{1/2} = \Omega_{1}.H^{1/2}z + (1-2t)^{-1}\varepsilon/T^{1/2}$$

$$Q\Sigma^{-1}u = \Sigma C'(C\Sigma C')^{-1}x \qquad (208)$$

$$u'Mu = x'x, \quad u'MBMu = x'CBC'x$$

The expectations will use the following formulas for a random $N \times 1$ vector $v \sim N(0,V)$ and symmetric nonstochastic $N \times N$ matrices, F_1, F_2, and F_3:

$$E\ v'F_1v = \text{tr}\ VF_1$$

$$E\ v'F_1v.v'F_2v = \text{tr}\ VF_1\ \text{tr}\ VF_2 + 2\ \text{tr}\ VF_1VF_2$$

$$E\ v'F_1v \cdot v'F_2v \cdot v'F_3v = (\prod_{i=1}^{3}\ \text{tr}\ VF_i)$$

(continued..)

$$+ 2\{\mathrm{tr}\ VF_1 VF_2\ \mathrm{tr}\ VF_3 + \mathrm{tr}\ VF_1 VF_3\ \mathrm{tr}\ VF_2$$

$$+ \mathrm{tr}\ VF_2\ VF_3\ \mathrm{tr}\ VF_1\}$$

$$+ 4\ \mathrm{tr}\ VF_1\ VF_2\ VF_3 + 4\ \mathrm{tr}\ VF_1\ VF_3\ VF_2 \qquad (209)$$

Keeping in mind that

$$\tilde{E}\ g(x,z) = \int_x \int_z g(x,z)\ f_z(z)\ f_x(x)\ dx\ dz \qquad (210)$$

for any function $g(x,z)$, and using the independence of z and x, the zero means of the distributions in (205), and the following table

expansion term	powers of x	powers of z
a_1	0	0
$a_{1/2}$	even	0
a_0	even	0
b_{-1}	0	even, odd
$b_{-3/2}$	odd	even, odd
b_{-2}	even, odd	even, odd
c_0	0	0
$c_{-1/2}$	even	0
c_{-1}	even	0

we evaluate the expectations of (197) required for the approximate m.g.f. of W.

For $\tilde{E}\,\eta_{-1/2}$, we can use the above results to obtain

$$\tilde{E}\, a_1 b_{-1}\, c_{-1/2} = q_1\, \omega_{11}^{-1}\, \mathrm{tr}\, \Sigma P/T;\ \tilde{E}\, a_1\, b_{-3/2}\, c_0 = 0$$

$$\tilde{E}\, a_{1/2} b_{-1}\, c_0 = q_1\, X'_{1.G}\, RX_{1.G}\, \mathrm{tr}\, \Sigma MBM/T(1 - \rho^2) \tag{211}$$

where

$$q_1 = \omega_{11}(1 - 2t)^{-1} + \gamma(1 - 2t)^{-2} \tag{212}$$

so that

$$\tilde{E}\, \eta_{-1/2} = (q_1/T)\{\omega_{11}^{-1}\, \mathrm{tr}\, \Sigma P + X'_{1.G} RX_{1.G}\ \mathrm{tr}\, \Sigma MBM/(1 - \rho^2)\} \tag{213}$$

The expectations required for $\tilde{E}\,\eta_{-1}$ are

$$\tilde{E}\, a_1\, b_{-1}\, c_{-1} = (q_1\, \omega_{11}^{-1}/T)\{2(1 + \rho^2)/(1 - \rho^2) - k\} \tag{214}$$

where the tr Σ^2 formula of (145) has been used,

$$\tilde{E}\, a_1\, b_{-3/2}\, c_{-1/2} = 0,$$

$$\tilde{E} a_1 b_{-2} c_0 = \{\omega_{11}^{-1}/T(1 - \rho^2)\}\Omega_1.X'RQRX\Omega._1 \quad (215)$$

which has used the tr $\sum MBM \sum MBM$ formula of (146),

$$\tilde{E}\, a_{1/2}\, b_{-1}\, c_{-1/2} = (-2q_1/T)\,\rho\, X'_{1.G}RX_{1.G}/(1 - \rho^2) \quad (216)$$

which follows from application of (145) to show that

$$\text{tr} \sum MBM \sum M = T\rho + o(T) \quad (217)$$

$$\tilde{E}\, a_{1/2}\, b_{-3/2}\, c_0 = 0 \quad (218)$$

$$\tilde{E}\, a_0\, b_{-1}\, c_0 = (q_1/T)\{h - 2\rho\, X'_{1.G}RX_{1.G}/(1 - \rho^2)\} \quad (219)$$

Using (214) to (219), we have

$$\tilde{E}\, \eta_{-1} = (q_1/T)\{\omega_{11}^{-1}(2(1 + \rho^2)/(1 - \rho^2) - k) + h - 4\rho\, X'_{1.G}RX_{1.G}/(1 - \rho^2)\} + \{\omega_{11}^{-1}/T(1 - \rho^2)\}\Omega_1.X'RQRX\Omega._1 \quad (220)$$

We also require $\tilde{E}\,\eta^2_{-1/2}$ which involves the following terms

$$\tilde{E}\, a_1^2\, b_{-1}^2\, c_{-1/2}^2 = \{2\omega_{11}(1+\rho^2)/T(1-\rho^2)\}q_2$$

$$\tilde{E}\, 2a_1^2\, b_{-1}\, b_{-3/2}\, c_0\, c_{-1/2} =$$

$$\tilde{E}\, 2a_1 a_{1/2} b_{-1} b_{-3/2} c_0^2 = 0,$$

$$\tilde{E}\, a_1^2\, b_{-3/2}^2\, c_0^2 = 4(q_1/T)\omega_{11}^{-2}\, \Omega_{1.}\, X'RQRX\Omega_{.1}/(1-\rho^2)$$

$$\tilde{E}\, 2a_1 a_{1/2}\, b_{-1}^2\, c_0\, c_{-1/2} = \{-4\rho\, X_{1.G}'RX_{1.G}$$

$$/T(1-\rho^2)\}q_2$$

$$\tilde{E}\, a_{1/2}^2\, b_{-1}^2\, c_0^2 = \{\omega_{11}^{-1}(X_{1.G}'RX_{1.G})^2/T(1-\rho^2)\}q_2 \qquad (221)$$

where

$$q_2 = 3\omega_{11}(1-2t)^{-2} + 6\gamma(1-2t)^{-3} + \omega_{11}^{-1}\gamma^2(1-2t)^{-4} \qquad (222)$$

so that

$$\tilde{E}\,\eta^2_{-1/2} = \{4\,\omega_{11}^{-2}\,\Omega_{1}.X'RQRX\Omega._{1}/T(1-\rho^2)\}q_1$$

$$+ [\{2\omega_{11}(1+\rho^2) - 4\rho X'_{1.G}RX_{1.G}$$

$$+ \omega_{11}^{-1}(X'_{1.G}RX_{1.G})^2\}/T(1-\rho^2)]q_2 \tag{223}$$

For the m.g.f. of (197) we use (213), (220) and (223) along with

$$X'_{1.G}RX_{1.G} = (1-\rho^2)(\omega_{11}^{-1} - \omega_{11}^{-2}\,\Omega_{1}.X'X\Omega._{1})/\rho$$

$$+ o(T) \tag{224}$$

$$h = \omega_{11}^{-2}\,\Omega_{1}.X'X\Omega._{1} - \omega_{11}^{2}(1-\rho^2)\Omega_{1}.X'X_G(X'_G\,\Sigma^{-1}X_G)^{-1}$$

$$X'_G\,X\Omega._{1}/\rho^2 + o(T) \tag{225}$$

$$\Omega_{1}.X'RQRX\Omega._{1} = (1-\rho^2)^2\,\Omega_{1}.X'QX\Omega._{1}/\rho^2 + o(T^{-1}) \tag{226}$$

After some simplification we have

$$t\,\tilde{E}(\eta_{-1/2} + \eta_{-1}) + t^2\tilde{E}\,\eta^2_{-1/2}/2$$

$$= T^{-1}\sum_{i=0}^{4}\tau_i(1-2t)^{-i} + o(T^{-1}) \qquad (227)$$

where the τ's are from (179). Therefore,

$$M_W(t) = \exp\{t(1-2t)^{-1}\gamma\}(1-2t)^{-1/2}\{1$$

$$+ T^{-1}\sum_{i=0}^{4}\tau_i(1-2t)^{-i}\} + o(T^{-1})$$

$$(228)$$

which can be inverted to give the approximation of (176).

(ii) A Correction Factor

The result (176) can be used to derive an Edgeworth size correction factor for W. Because there can be a non-monotonicity problem here as for W_t of section 3(iic) of chapter II, the critical value correction method of section 3(iid) of chapter II will be used instead. First, we set $\gamma = \delta = 0$ since we are concerned here with the null distribution of W. The expansion (176) must be written in terms of χ^2 p.d.f.'s as in the appendix of chapter II. Using the notation of that appendix, with F and f representing a χ^2 c.d.f. and p.d.f. respectively, under H_0

$$pr(W \leq x) = F(x,1,0) + T^{-1} \sum_{i=1}^{2} \tau_i^* f(x, 1+2i, 0) + o(T^{-1}) \qquad (229)$$

where

$$\tau_1^* = \omega_{11}(-2\phi_1 + 3\phi_2 - 2\phi_3)/2$$

$$\tau_2^* = -3\omega_{11}\phi_2/2 \qquad (230)$$

and the ϕ's are from (180), (181), and (182).

By following the same procedure as in section 3(iid) of chapter II, we arrive at the following critical value adjustment given the initial critical value x which is based on the asymptotic distribution of W_0 (so that x = 3.841 for a test of asymptotic size 5%):

$$x^* \doteq x\{1-(\hat{\tau}_1^* + x\hat{\tau}_2^*/3)/T\} \qquad (231)$$

and $\hat{\tau}_1^*$, $\hat{\tau}_2^*$ are constructed using any consistent estimate of ρ. For large ρ values when the over-rejection problem is most severe, one would hope that $x^* > x$ so that the adjustment reduces the nominal size of the test.

The test statistic (173) is not often used; However, the results of this section are potentially valuable for two reasons:

1) The application of the Edgeworth expansion method to a regression problem with autoregressive errors is new. The step which makes this possible, and which could probably be used in other cases with similar error structure, is the transformation of the error vector u of (200) and (201). Extension of this method to expansions in cases of more general covariance structures could be done by replacing $\sum$ and Ω in (200) and (202) with the appropriate error covariance matrix and GLS estimator covariance matrix, respectively.

2) This test statistic (173) is not used since it is constructed using an inefficient variance estimator. However, this enables a size correction factor to be derived which, while algebraically lengthy,

is considerably less cumbersome than it would be when the efficient variance estimator is used, as in Park and Mitchell (1980). With more attention being paid to small sample sizes, tests such as (173) for which some small sample information is known, may be considered more often. The issue becomes choosing between "inefficiency" along with (presumably) a loss in power, and knowledge of small sample size. A similar kind of choice is involved when choosing between the K test of (128) in chapter IV with its local optimality property, and the EFT test of (132) of that chapter with its exact small sample distributional property. Another example is the non-nested situation of chapter V for which an exact test could be performed by artificially constructing a nested hypothesis and using the F statistic of (22) in chapter II, which is strictly less powerful than the Cox and J tests of chapter V in the asymptotic local sense of Pesaran (1982), but which possesses a much simpler small sample distribution.

III.6 SUMMARY

In this chapter some estimation and testing problems in the regression model with autocorrelated disturbances are considered. First, it is shown that the LM test for the existence of the autocorrelation is numerically robust to the t distribution in the errors. Next, the performance of various estimators of the regression coefficients is compared by approximating their MSE matrices using Nagar's expansion. The results largely support previous analytical and Monte Carlo studies. Finally, the distribution of a test for significance of a single regression coefficient is approximated by Edgeworth expansion, and a size-corrected critical value is offered.

FOOTNOTES

Chapter III

1. The use of $\hat{\rho}_{ML}$ requires iteration since it is a function of the final β estimate. Four of the other five estimates, $\hat{\rho}_D$ excluded, could also be used in an iterative fashion, with the ρ estimate from each iteration being formed by replacing e in the appropriate $\hat{\rho}$ formula by the residual vector formed using the β estimate from the previous iteration. This new ρ estimate can be used to form a new β estimate, and so on. The iterative $\hat{\rho}_S$ estimator is considered in section 4.

2. In the student t case, $\sum = I$ does not imply that the elements of u are independently distributed.

3. "Better" in this context means "lower MSE".

4. It is not always optimal to use the correct ρ value when estimating by CO. An extreme counter-example is when X is a column of ones as in Kadiyala (1968). See Magee (1982, section 5).

5. Some of these results can be found in Magee, Ullah, and Srivastava (1984).

6. The "orders" of T here refer to the orders of the individual elements of the $T \times T$ matrices.

7. Since all MSE's and approximated MSE's considered here involve σ^2 only as a scalar multiple, the resulting comparisons will be invariant to the true σ^2 value as long as it is greater than zero.

8. Tillman (1975, p. 970) defines Watson's X in a slightly different way. His definition includes a constant term while Watson's own definition does not. (λ_1 is a constant vector.) Consequently, the markedly poor performance of OLS in M4 reported here is not nearly as severe when Tillman's definition is used.

IV. TESTS FOR INDEPENDENCE IN SURE MODELS

IV.1 INTRODUCTION

The Seemingly Unrelated Regression Equation (SURE) model developed by Zellner (1962) is useful when a set of regression equations have error terms which could be contemporaneously correlated and the right-hand side variables are distributed independently of their corresponding error terms. An example of such a situation is a series of aggregate factor demand equations, where each factor demand is a function of input and output prices and output levels (Kokkelenberg (1983)).

The problem considered in this chapter is testing for correlation between the residuals of one equation and those of the remaining (N-1) equations. If the hypothesis of no correlation is rejected, then the equation in question should probably be included with the others in a SURE system. Tests are developed and discussed for the two equation and general N equation cases, and the approximate slopes of some of these tests are compared.

IV.2 THE MODEL AND ESTIMATORS

(i) The Model

The model is

$$
\begin{aligned}
y_1 &= X_1\beta_1 + u_{.1} \\
&\vdots \\
y_i &= X_i\beta_i + u_{.i} \\
&\vdots \\
y_N &= X_N\beta_N + u_{.N}
\end{aligned}
\qquad (1)
$$

where X_i is a $T \times k_i$ matrix of exogenous variables, β_i is an unknown $k_i \times 1$ vector of regression coefficients, y_i is a $T \times 1$ endogenous variable vector, and $u_{.i}$ is an unknown $T \times 1$ disturbance vector, with

$$
X_i = \begin{bmatrix} x_{i11} & \cdots & x_{i1k_i} \\ & & \\ x_{iT1} & \cdots & x_{iTk_i} \end{bmatrix} = \begin{bmatrix} x'_{i1.} \\ \vdots \\ x'_{iT.} \end{bmatrix} \qquad (2)
$$

$$
\beta_i = \begin{bmatrix} \beta_{i1} \\ \vdots \\ \beta_{ik_i} \end{bmatrix}, \; y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} \text{ and } u_{.i} = \begin{bmatrix} u_{1i} \\ \vdots \\ u_{Ti} \end{bmatrix} \qquad (3)
$$

where x_{it}', $t = 1, \ldots, T$ in (2) represents the t^{th} row vector of X_i.

Consider a $T \times N$ stochastic matrix U whose i^{th} column is given by $u_{\cdot i}$, defined in (3). We then have,

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1i} & \cdots & u_{1N} \\ \vdots & & \vdots & & \\ u_{t1} & \cdots & u_{ti} & \cdots & u_{tN} \\ \vdots & & \vdots & & \\ u_{T1} & \cdots & u_{Ti} & \cdots & u_{TN} \end{bmatrix}$$

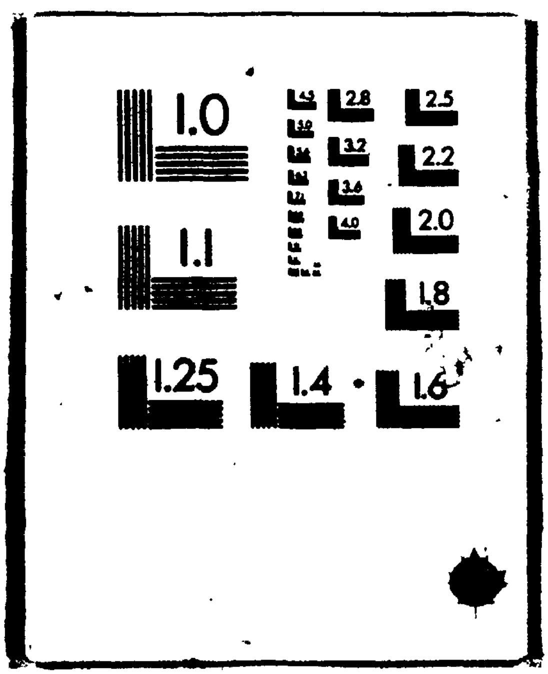
$$= \begin{bmatrix} u_{\cdot 1} & \cdots & u_{\cdot i} & \cdots & u_{\cdot N} \end{bmatrix}$$

$$= \begin{bmatrix} u_{1\cdot} & \cdots & u_{t\cdot} & \cdots & u_{T\cdot} \end{bmatrix}' \qquad (4)$$

where $u_{t\cdot}'$ is the t^{th} row of U.

Notice that $u_{\cdot i}$ is a $T \times 1$ column vector of residuals <u>from</u> the i^{th} equation <u>across</u> the T observations, while $u_{t\cdot}$ is an $N \times 1$ column vector of residuals <u>from</u> the t^{th} observation <u>across each</u> of the N equations. It will be useful in following sections to use both of these vectors, and so their distinction is to be carefully noted at this point.

3



The following distributional assumption is made.

ASSUMPTION

The rows of U are independent identically distributed (i.i.d.) miltivariate normal with

$$u_{t\cdot} \sim N(0,\Omega), \qquad t = 1,\ldots,T, \tag{5}$$

and

$$\Omega = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1j} & \cdots & \omega_{1N} \\ \vdots & & & & \\ \omega_{i1} & \cdots & \omega_{ij} & \cdots & \omega_{iN} \\ \vdots & & & & \\ \omega_{N1} & \cdots & \omega_{Nj} & \cdots & \omega_{NN} \end{bmatrix} \tag{6}$$

where Ω is an $N \times N$ positive definite symmetric matrix.

It is useful to write the set of equations (1) in the following stacked form:

$$y = X\beta + u \tag{7}$$

where

$$y = \left[y_1', \ldots, y_i', \ldots, y_N'\right]', \tag{8}$$

a $TN \times 1$ vector,

$$\beta = \left[\beta_1', \ldots, \beta_i', \ldots, \beta_N'\right]', \tag{9}$$

a $K \times 1$ vector with $K = \sum_{i=1}^{N} k_i$,

$$X = \begin{bmatrix} X_1 0 & \ldots & 0 & \ldots\ldots & 0 \\ \vdots & \ddots & & & \vdots \\ 0 \ldots\ldots & \vdots & 0 X_i 0 & \ldots & 0 \\ \vdots & & & \ddots & \vdots \\ \vdots & & & & .0 \\ 0 & \ldots\ldots\ldots & & & 0 X_N \end{bmatrix} \tag{10}$$

a $TN \times K$ matrix, and

$$u = \left[u_{\cdot 1}', \ldots, u_{\cdot i}', \ldots, u_{\cdot N}'\right]', \tag{11}$$

a $TN \times 1$ vector.

The assumption of (5) implies that

$$u \sim N(0, \Omega \otimes I_T), \tag{12}$$

where $\otimes$ represents the Kroneker product.

(ii) Unrestricted ML Estimation

Given (12), we have the probability density function of u as

$$f(u) = (2\Pi)^{-NT/2}|\Omega \otimes I_T|^{-1/2} \exp\{-u'(\Omega \otimes I_T)u/2\} \qquad (13)$$

We can now write the likelihood function for β and Ω:

$$\log L(\beta,\Omega|y,X) = \ell = -(NT/2) \log 2\Pi$$

$$- (1/2) \log|\Omega \otimes I_T|$$

$$- (1/2)\, u'(\Omega \otimes I_T)^{-1}u \qquad (14)$$

where $u = y - X\beta$ as in (7). Using (11) and the alternate partitionings of U in (4), we can show that

$$u'(\Omega \otimes I_T)^{-1}u = u'(\Omega^{-1} \otimes I_T)u$$

$$= [u'_{.1} \dots u'_{.N}] \begin{bmatrix} \omega^{11}I_T & \dots & \omega^{1j}I_T & \dots & \omega^{1N}I_T \\ \vdots & & & & \vdots \\ \omega^{i1}I_T & \dots & \omega^{ij}I_T & \dots & \omega^{iN}I_T \\ \vdots & & & & \vdots \\ \omega^{N1}I_T & \dots & \omega^{Nj}I_T & \dots & \omega^{NN}I_T \end{bmatrix} \begin{bmatrix} u_{.1} \\ \vdots \\ u_{.N} \end{bmatrix}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} u_{\cdot i}' \, \omega^{ij} \, u_{\cdot j}$$

$$= \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} u_{ti} \, \omega^{ij} u_{tj} \right)$$

$$= \sum_{t=1}^{T} u_{t\cdot}' \, \Omega^{-1} \, u_{t\cdot} \tag{15}$$

where ω^{ij} refers to the $(i,j)^{th}$ element of Ω^{-1}.

The log likelihood (14) can then be written as

$$\ell = -(NT/2) \log 2\Pi - (T/2) \log |\Omega| - (1/2) \sum_{t=1}^{T} u_{t\cdot}' \, \Omega^{-1} u_{t\cdot} \tag{16}$$

where we have used $|\Omega \otimes I_T| = |\Omega|^T |I_T|^N = |\Omega|^T$ (see Theil (1971, pg. 305)) and from the definitions in (2), we have

$$u_{t\cdot}' = [y_{1t} - x_{1t\cdot}\beta_1, \ldots, y_{it} - x_{it\cdot}\beta_i, \ldots, y_{Nt} - x_{Nt\cdot}\beta_N] \tag{17}$$

Noting from (15) that the first derivatives of ℓ in (16) are[1]

$$\partial\ell/\partial\beta_i = X_i' \left(\sum_{j=1}^{N} \omega^{ij} \, u_{\cdot j} \right), \quad i=1,\ldots,N, \tag{18}$$

and that

$$\partial \ell / \partial \omega_{ii} = -T\, \omega^{ii}/2 + (1/2) \sum_{t=1}^{T} u_{t.}' \, \Omega^{.i} \, \Omega^{i.} \, u_{t.} \, , \qquad i = 1,\ldots,N, \tag{19}$$

$$\partial \ell / \partial \omega_{ij} = -T\omega^{ij} + \sum_{t=1}^{T} u_{t.}' \, \Omega^{.i} \, \Omega^{j.} \, u_{t.} \qquad i \neq j, \; i,j = 1,\ldots,N \tag{20}$$

where $\Omega^{.i}$ refers to the i^{th} column of Ω^{-1} and $\Omega^{i.}$ to its i^{th} row.

Setting the derivatives of (18) to (20) to zero and simplifying, we can show that the unrestricted ML estimates, $\hat{\beta}$ and $\hat{\Omega}$, satisfy the following first order conditions:

$$X'[\hat{\Omega}^{-1} \otimes I_T](y - X\hat{\beta}) = 0 \tag{21}$$

$$-T\, \hat{\Omega}^{-1} + \hat{\Omega}^{-1} \sum_{i=1}^{T} \hat{u}_{t.} \hat{u}_{t.}' \, \hat{\Omega}^{-1} = 0 \tag{22}$$

These conditions (21) and (22) can be written as

$$\hat{\beta} = [X'(\hat{\Omega}^{-1} \otimes I_T)X]^{-1} X'(\hat{\Omega}^{-1} \otimes I_T)y \tag{23}$$

$$\hat{\Omega} = \hat{U}'\hat{U}/T \tag{24}$$

where

$$\hat{U} = [\hat{u}_{\cdot 1}, \ldots, \hat{u}_{\cdot i}, \ldots, \hat{u}_{\cdot N}],$$

$$= [\hat{u}_{1.}, \ldots, \hat{u}_{t.}, \ldots, \hat{u}_{T.}]' \quad (25)$$

is a T × N matrix of unrestricted ML estimates of U in (4), and

$$\hat{u}_{\cdot i} = y_i - X_i \hat{\beta}_i, \qquad i = 1, \ldots, N \quad (26)$$

The above solution is not in closed form, and so it is usually evaluated by an iterative procedure. For example, one could start by using OLS estimates:

$$b_i = (X_i' X_i)^{-1} X_i' y_i, \qquad i = 1, \ldots, N \quad (27)$$

Then equation (24) could be used to estimate Ω and the resulting $\hat{\Omega}$ could be used in (23) to get a new estimate of β. This resulting $\hat{\beta}$ is known as Zellner's two-step estimator. If one continues to use (23) and (24) until the parameters converge, then the resulting iterative Zellner estimator can be interpreted as the unrestricted ML estimate[2], and is equivalent to another iterative method proposed by Telser (1964) (see Kmenta and Gilbert (1968)). There is evidence that in some situations the efficiency (in terms of mean square error) of the two-step estimator of β is greater than the iterative estimator (Kmenta and

Gilbert (1968), Conniffe (1982)).

It should be noted here that Rao (1974) has shown that the OLS estimates (27) are more efficient than the SURE estimates (23) when the cross equation correlation of the residuals is due to a common omitted variable rather than due to non-zero off-diagonal elements of Ω in (5).

(iii) Restricted ML Estimation

If it is assumed that the residuals from the first equation are distributed independently of the residuals from the remaining N-1 equations (which is the hypothesis to be tested in the next section), then the implied parameter restriction is

$$\Omega_{S1} = [\omega_{21}, \ldots, \omega_{i1}, \ldots, \omega_{N1}]' = 0 \qquad (28)$$

where the matrix Ω from (5) has been partitioned in the following way:

$$\Omega = \begin{bmatrix} \omega_{11} & \Omega_{1S} \\ \Omega_{S1} & \Omega_{SS} \end{bmatrix}, \qquad (29)$$

where Ω_{S1} is an $(N-1) \times 1$ vector, $\Omega_{1S} = \Omega_{S1}'$ due to the symmetry of Ω, and Ω_{SS} is an $(N-1) \times (N-1)$ positive definite symmetric sub-matrix.

It will be useful to partition the variables from the stacked system (7) to (11) to conform with the partitioning of Ω in (29):

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_S \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ y_S \end{bmatrix},$$

$$X = \begin{bmatrix} X_1 & 0 \\ 0 & X_S \end{bmatrix} \qquad u = \begin{bmatrix} u_{\cdot 1} \\ u_S \end{bmatrix} \tag{30}$$

where the S subscript refers to the variables from equations 2 through N.

In a similar way, partition U in (4) as

$$U = [u_{\cdot 1} \quad U_S] \tag{31}$$

Imposing restriction (28) on the log likelihood function (16), we see that the restricted log likelihood function (of the remaining unrestricted parameters) is

$$\ell_R = -(NT/2) \log 2\Pi - (T/2) \log \omega_{11}|\Omega_{SS}|$$
$$-(1/2)(u'_{\cdot 1}u_{\cdot 1}\omega^{11} + \sum_{t=1}^{T} u'_{St\cdot}\Omega_{SS}^{-1}u_{St\cdot}) \tag{32}$$

where $u'_{St\cdot}$ refers to the t^{th} row of U_S.

Proceeding as in the unrestricted ML case we have from (32),

$$\partial \ell_R / \partial \beta_1 = \omega^{11} X_1' u_1. \tag{33}$$

and

$$\partial \ell_R / \partial \beta_i = X_i' \left(\sum_{j=2}^{N} \omega^{ij} u_{\cdot j} \right), \quad i=2,\ldots,N. \tag{34}$$

From (32) we see that

$$\partial \ell_R / \partial \omega_{11} = -T/2 + (1/2) u_{\cdot 1}' u_{\cdot 1} \tag{35}$$

$$\partial \ell_R / \partial \omega_{ii} = -T\omega^{ii}/2 + (1/2) \sum_{t=1}^{T} u_{St\cdot}' \Omega_{SS}^{\cdot i} \Omega_{SS}^{i\cdot} u_{St\cdot} \tag{36}$$

$$i = 2,\ldots,N,$$

and

$$\partial \ell_R / \partial \omega_{ij} = -T\omega^{ij} + \sum_{t=1}^{T} u_{St\cdot}' \, \Omega_{SS}^{\cdot i} \, \Omega_{SS}^{j\cdot} \, u_{St\cdot} \tag{37}$$

$$i,j = 2,\ldots,N; \; i \neq j,$$

where $\Omega_{SS}^{\cdot i}$ refers to the i^{th} column of Ω_{SS}^{-1} and $\Omega_{SS}^{i\cdot}$ to its i^{th} row.

It can be seen from (33) to (37) that the derivatives with respect to the parameters of the first equation do not depend on any of the other parameters, and vice versa. Thus, the restricted ML estimates for the equation 1 parameters can be written from (33) and (35) as

$$\hat{\beta}_{1R} = (X_1' X_1)^{-1} X_1' y_1 \qquad (38)$$

and

$$\hat{\omega}_{11R} = \hat{u}_{\cdot 1R}' \hat{u}_{\cdot 1R}/T \qquad (39)$$

where

$$\hat{u}_{\cdot 1R} = y_1 - X_1\hat{\beta}_{1R}, \qquad (40)$$

and the R subscript refers to restricted ML estimates. These are simply OLS estimates.

Comparing (34), (36), and (37) with equations (18), (19), and (20), it can be seen that the restricted ML estimates for the parameters of equations two through N are the same as unrestricted ML estimates treating these equations as a SURE system separately from equation one. In the N = 2 case, then, this would result in OLS estimates of the parameters of equation two. Proceeding in the manner of equations (21) to (24), we have

$$\hat{\beta}_{SR} = (X_S'(\hat{\Omega}_{SSR}^{-1} \otimes I_T)X_S)^{-1}X_S'(\hat{\Omega}_{SSR}^{-1} \otimes I_T)y_S, \tag{41}$$

and

$$\hat{\Omega}_{SSR} = \hat{U}_{SR}' \hat{U}_{SR}/T \tag{42}$$

where

$$\hat{U}_{SR} = [\hat{u}_{\cdot 2R}, \ldots, \hat{u}_{\cdot iR}, \ldots, \hat{u}_{\cdot NR}] \tag{43}$$

and

$$\hat{u}_{\cdot iR} = y_i - X_i\hat{\beta}_{iR}, \quad i = 2, \ldots, N. \tag{44}$$

IV.3 TEST STATISTICS IN THE N EQUATION CASE

(i) Introduction and Statement of the Hypotheses

As indicated previously, the hypothesis tests discussed in this chapter are designed to determine whether or not the errors from a particular equation (say equation one without loss of generality) from the set of equations (1) are contemporaneously correlated with the errors of any of the other equations. The hypotheses are then

$$H_0 : \Omega_{S1} = 0 \quad \text{vs.} \quad H_1 : \Omega_{S1} \neq 0 \tag{45}$$

using the partitioning of Ω given in (29).

In this section the tests for (45) which result from application of the LR, LM, and UI principles in the general N equation case are presented along with derivations.

(ii) LR Test

Theorem 1. The LR test statistic for testing hypotheses (45) in the SURE model (1) under assumption (5) is given by

$$LR = T(\log \hat{\omega}_{11R}|\hat{\Omega}_{SSR}| - \log|\hat{\Omega}|) \quad (46)$$

where $\hat{\omega}_{11R}$, $\hat{\Omega}_{SSR}$ and $\hat{\Omega}$ are given in (39), (42), and (24), respectively.

Note: The outcome of the test with asymptotic size α is determined by the following decision rule:

$$\begin{aligned} &\text{Accept } H_0 \text{ if } LR \leq c(N-1,\alpha) \\ &\text{Reject } H_0 \text{ if } LR > c(N-1,\alpha) \end{aligned} \quad (47)$$

where

$c(N-1,\alpha)$ satisfies $\text{prob}\{\chi^2_{N-1} \leq c(N-1,\alpha)\} = 1-\alpha$

This decision rule follows naturally from the discussion in section 2(i) of chapter 1.

Proof. Letting $\theta = \{\beta', (\text{vec } \Omega)'\}$ represent the set of unknown parameters, the LR statistic given in equation (27) of chapter 1 can be used:

$$LR = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}_R)) \tag{48}$$

where $\hat{\theta}$ and $\hat{\theta}_R$ represent the restricted and unrestricted ML estimates, respectively. Thus, the derivation of the LR test statistic simply requires the evaluation of the log likelihood function (14) or (16) at $\hat{\theta}$ and at $\hat{\theta}_R$.

For the evaluation of $\ell(\hat{\theta})$, note that

$$\begin{aligned} \sum_{t=1}^{T} \hat{u}_{t\cdot}' \hat{\Omega}^{-1} \hat{u}_{t\cdot} &= T \text{ tr } \hat{U}[\hat{U}'\hat{U}]^{-1} \hat{U}' \\ &= TN \end{aligned} \tag{49}$$

by using (24), (25), (26), and (4).

Substituting (49) into (16), we have

$$\begin{aligned} \ell(\hat{\theta}) = &-(NT/2) \log 2\Pi - (T/2) \log |\hat{\Omega}| \\ &- (1/2)NT \end{aligned} \tag{50}$$

Similarly, for $\ell(\hat{\theta}_R)$, noting that under the restrictions we can use the restricted log likelihood (32), which gives

$$\hat{u}'_{\cdot 1R}\,\hat{u}_{\cdot 1R}\,\hat{\omega}^{11R} = T \tag{51}$$

and

$$\sum_{t=1}^{T} \hat{u}'_{St.R}\,\hat{\Omega}^{-1}_{SSR}\,\hat{u}_{St.R} = T\ \mathrm{tr}\ \hat{U}_{SR}(\hat{U}'_{SR}\hat{U}_{SR})^{-1}\hat{U}'_{SR}$$

$$= T(N-1) \tag{52}$$

by using (38), (39), and (42).

Substituting (51) and (52) into (32) yields

$$\ell(\hat{\theta}_R) = -(NT/2)\log 2\Pi - (T/2)\log \hat{\omega}_{11R}|\hat{\Omega}_{22R}|$$

$$- (1/2)TN \tag{53}$$

Now substituting (50) and (53) into (48), we obtain result (46).

Q.E.D.

(iii) LM Test

The LM test in this section tests hypothesis (45). The stronger hypothesis of independence of errors in each equation from each other equation can also be tested using an LM procedure as in Breusch and Pagan (1980). When N = 2, these tests become identical.

Theorem 2. The LM test statistic for testing hypothesis (45) in the SURE model (1) under assumption (5) is given by

$$LM = T\,\hat{R}^2 \qquad (54)$$

where

$$\hat{R}^2 = \hat{u}'_{\cdot 1R}\,\hat{U}_{SR}(\hat{U}'_{SR}\,\hat{U}_{SR})^{-1}\,\hat{U}'_{SR}\,\hat{u}_{\cdot 1R}/\hat{u}'_{\cdot 1R}\,\hat{u}_{\cdot 1R} \qquad (55)$$

is the multiple correlation coefficient from the regression of $\hat{u}_{\cdot 1R}$ on $\hat{U}_{SR}$; $\hat{u}_{\cdot 1R}$ and $\hat{U}_{SR}$ are given in (40) and (43) respectively.

Note. The outcome of the test with asymptotic size α is given by the following decision rule:

$$\text{Accept } H_0 \text{ if } LM \le c(N-1,\alpha)$$

$$\text{Reject } H_0 \text{ if } LM > c(N-1,\alpha) \tag{56}$$

where $c(N-1,\alpha)$ satisfies $\text{prob}\{\chi^2_{N-1} \le c(N-1,\alpha)\} = 1 - \alpha$

Proof. From (33) of chapter 1 we have

$$LM = \tilde{d}' \, I(\hat{\theta}_R)^{-1} \, \tilde{d} \tag{57}$$

where

$$\tilde{d} = \partial \ell / \partial \theta |_{\hat{\theta}_R} \tag{58}$$

and

$$I(\hat{\theta}_R) = -E(\partial^2 \ell / \partial \theta \partial \theta') \tag{59}$$

with the expectation taken using $\hat{\theta}_R$ as the parameter values. The elements of $\tilde{d}$ will be zero except for those terms corresponding to the restricted parameters, i.e., the elements of Ω_{S1}. From (20) we have

$$\partial \ell / \partial \omega_{1i} = -T\omega^{1i} + \sum_{t=1}^{T} u'_{t\cdot} \, \Omega^{\cdot 1} \, \Omega^{i\cdot} \, u_{t\cdot} \tag{60}$$

$$i = 2, \ldots, T$$

Evaluated at $\hat{\theta}_R$, noting that $\hat{\Omega}_{S1R} = \hat{\Omega}^{S1}_R = 0$, we have

$$\partial \ell / \partial \omega_{1i} |_{\hat{\theta}_R} = \sum_{t=1}^{T} \hat{u}_{t1R} \, \hat{\omega}_R^{11} \, \hat{\Omega}_R^{i.} \, \hat{u}_{t\cdot R}$$

$$= \hat{\omega}_R^{11} \, \hat{u}'_{\cdot 1R} \, \hat{U}_{\cdot SR} \, \hat{\Omega}_{SSR}^{i.} \qquad (61)$$

$$\therefore \quad \partial \ell / \partial \Omega_{S1} |_{\hat{\theta}_R} = \hat{\omega}_R^{11} \, \hat{\Omega}_{SSR}^{-1} \, \hat{U}'_{\cdot SR} \, \hat{u}_{\cdot 1R}$$

$$= \tilde{d}_p \qquad (62)$$

where $\tilde{d}_p$ defines the $(N-1) \times 1$ vector of non-zero elements of $\tilde{d}$.

We are interested in the block of $I(\hat{\theta}_R)^{-1}$ corresponding to the elements of Ω_{S1}. First, consider the elements of $I(\hat{\theta}_R)$ corresponding to second derivatives of ℓ with respect to one element of Ω_{S1} and another element of θ which is not an element of Ω_{S1}. If these can all be shown to be zero, then we need consider only the block of $I(\hat{\theta}_R)$ corresponding to the elements of Ω_{S1} only in our calculation of LM. This turns out to be the case, and is shown below.

From (20), using the derivative rules of footnote 1, we have

$$\partial^2 \ell / \partial \beta_i \partial \omega_{1j} = -X_i' \sum_{j=1}^{N} (\omega^{i1} \omega^{jj} + \omega^{ij} \omega^{1j}) u_{\cdot j} \qquad (63)$$

$$j = 2, \ldots, N$$

so that

$$-E\partial^2 \ell/\partial\beta_i \partial\omega_{ij}|_{\hat{\theta}_R} = 0, \quad j = 2,\ldots,N \tag{64}$$

Using (20) and the rules of footnote 1,

$$\partial^2 \ell/\partial\omega_{1j}\partial\omega_{k\ell} = T(\omega^{1k}\,\omega^{j\ell} + \omega^{1\ell}\,\omega^{jk})$$

$$- \sum_{t=1}^{T} u_{t\cdot}'\{(\Omega^{\cdot k}\,\omega^{1\ell} + \Omega^{\cdot\ell}\omega^{1k})\Omega^{j\cdot}$$

$$+ \Omega^{\cdot 1}(\omega^{jk}\,\Omega^{\ell\cdot} + \omega^{j\ell}\,\Omega^{k\cdot})\}u_{t\cdot} \tag{65}$$

for $j \neq 1$, and $k \neq \ell$

Since $\hat{\Omega}_R^{S1} = 0$,

$$\partial^2 \ell/\partial\omega_{1j}\partial\omega_{k\ell}|_{\hat{\theta}_R} = - \sum_{t=1}^{T} u_{t1}\,\hat{\omega}_R^{11}(\hat{\omega}_R^{jk}\,\hat{\Omega}_R^{\ell\cdot}$$

$$+ \hat{\omega}_R^{j\ell}\,\Omega_R^{k\cdot})u_{t\cdot}' \tag{66}$$

for $j,k,\ell \neq 1$ and $k \neq \ell$.

Noting that $\hat{\omega}_R^{\ell 1} = \hat{\omega}_R^{k1} = 0$, and that, taking expectations assuming parameter values of $\hat{\theta}_R$,

$$E\, u_{t1}\, u_{ti} = 0, \qquad i = 2,\ldots,N, \tag{67}$$

we have

$$-E\, \partial^2 \ell / \partial\omega_{1j} \partial\omega_{k\ell} |_{\hat{\theta}_R} = 0, \tag{68}$$

for $j,k,\ell \neq 1$, and $k \neq \ell$

By procedures very similar to the above, it can be easily shown that

$$-E \partial^2 \ell / \partial\omega_{1j} \partial\omega_{kk} |_{\hat{\theta}_R} = 0 \quad \text{for } j,k \neq 1, \tag{69}$$

and

$$-E \partial^2 \ell / \partial\omega_{1j} \partial\omega_{11} |_{\hat{\theta}_R} = 0 \quad \text{for } j \neq 1. \tag{70}$$

Using (64), (68), (69) and (70), and denoting the block of $I(\hat{\theta}_R)$ corresponding to elements of Ω_{S1} by $I(\hat{\theta}_R)_p$, we now have:

$$\{I(\hat{\theta}_R)^{-1}\}_p = \{I(\hat{\theta}_R)_p\}^{-1} \tag{71}$$

We can then write

$$LM = \tilde{d}_p' \{I(\hat{\theta}_R)_p\}^{-1} \tilde{d}_p \tag{72}$$

where $\tilde{d}_p$ is defined in (62) and

$$I(\hat{\theta}_R)_p = -E\partial^2 \ell/\partial\Omega_{S1}\partial\Omega'_{S1}|_{\hat{\theta}_R} \tag{73}$$

To evaluate $I(\hat{\theta}_R)_p$, we have from (65),

$$\partial^2\ell/\partial\omega_{1i}\partial\omega_{1j} = T(\omega^{11}\omega^{ij} + \omega^{1i}\,\omega^{1j})$$
$$- \sum_{t=1}^{T} u'_{t\cdot}\{(\omega^{\cdot 1}\,\omega^{1j} + \omega^{\cdot j}\,\omega^{11})\omega^{i\cdot}$$
$$+ \omega^{\cdot 1}(\omega^{1i}\,\omega^{j\cdot} + \omega^{ij}\,\omega^{1\cdot})\}u_{t\cdot}$$
$$i\,,j \neq 1, \tag{74}$$

so that at $\hat{\theta}_R$ we have

$$\partial^2\ell/\partial\omega_{1i}\partial\omega_{1j}|_{\hat{\theta}_R} = T\hat{\omega}_R^{11}\,\hat{\omega}_R^{ij} - \hat{\omega}_R^{11}\sum_{t=1}^{T} u'_{t\cdot}\,\hat{\omega}_R^{\cdot j}\,\hat{\omega}_R^{i\cdot}u_{t\cdot}$$
$$- \hat{\omega}_R^{11}\sum_{t=1}^{T} u_{t1}\{(\hat{\omega}_R^{1j}\hat{\omega}_R^{i\cdot} + \hat{\omega}_R^{1i}\hat{\omega}_R^{j\cdot})u_{t\cdot}$$
$$+ \hat{\omega}_R^{ij}\,\hat{\omega}_R^{11}\,u_{t\cdot}\} \quad i,j \neq 1 \tag{75}$$

Evaluating the expectation of (75) using $\hat{\theta}_R$ as the parameter values and noting (67), we get

$$E\partial^2\ell/\partial\omega_{1i}\partial\omega_{1j}|_{\hat{\theta}_R} = T\,\hat{\omega}_R^{11}\,\hat{\omega}_R^{ij}$$

$$-\,\hat{\omega}_R^{11}\,T\,\hat{\omega}_R^{i.}\hat{\Omega}_R\,\hat{\omega}_R^{.j}$$

$$-\,(\hat{\omega}_R^{11})^2\,\hat{\omega}_R^{ij}(T\hat{\omega}_{11R})$$

$$= -T\,\hat{\omega}_R^{11}\,\hat{\omega}_R^{ij}, \quad i,j \neq 1 \qquad (76)$$

This gives the result

$$I(\hat{\theta}_R)_p = T\,\hat{\omega}_R^{11}\,\hat{\Omega}_{SSR}^{-1} \qquad (77)$$

Substituting (62) and (77) into (72) and using (39) and (42), we have the results (54) and (55). The decision rule (56) follows from the discussion of chapter 1.

Q.E.D.

(iv) UI Test

A test for hypotheses (45) can be constructed by using the infinite UI procedure of chapter 1, section 2(v). This test will have a known small sample distribution when it is based on the OLS residuals from the following regression equations:

$$y_i = X^* \gamma_i + v_i, \qquad i = 1,\ldots,N \tag{78}$$

where X^* is a $T \times K^*$ matrix which includes all of the columns of $X_1, X_2, \ldots, X_N$ with no column appearing twice. We assume that $X^{*'}X^*$ is non-singular and that $K^* < T$, which will be the case in most econometric applications. The OLS residuals from the equations (78) are:

$$\hat{v}_i = M^* y_i, \qquad i = 1,\ldots,N \tag{79}$$

where

$$M^* = I - X^*(X^{*'} X^*)^{-1}X^{*'} \tag{80}$$

The advantage of using the $\hat{v}_i$'s is that the resulting variance-covariance matrix has the following exact distributional property under assumption (5):

$$T\tilde{\Omega} \sim W_i(\Omega,\ T-K^*) \tag{81}$$

where

$$\tilde{\Omega} = [\hat{v}_1 \ldots \hat{v}_i \ldots \hat{v}_N]'[\hat{v}_1 \ldots \hat{v}_i \ldots \hat{v}_N]/T$$

$$= \hat{V}'\hat{V}/T \tag{82}$$

and W_i refers to the Wishart distribution, the multivariate analog to χ^2, (for properties, see Muirhead (1982)), with $T-K^*$ degrees of freedom, $\tilde{\Omega}$ has a mean of Ω, and

$$\hat{V} = [\hat{v}_1 \ldots \hat{v}_i \ldots \hat{v}_N]$$

$$= [\hat{v}_1\ \hat{V}_S] \tag{83}$$

Theorem 3. Applying the infinite UI test procedure of section 2(v) of Chapter 1 to test hypotheses (45) using the matrix $\tilde{\Omega}$ of (82) yields the following test statistic:

$$UI = T\tilde{\Omega}_{S1}'\,\tilde{\Omega}_{SS}^{-1}\,\tilde{\Omega}_{S1}/\tilde{\omega}_{11} \tag{84}$$

where $\tilde{\Omega}$ has been partitioned identically to Ω in (29).

Note: Unlike the LR test statistic of (46) or the LM statistic (54), this UI statistic has a known exact small sample distribution under H_0 (see Srivastava and Khatri (1979, pp. 223-4)). Critical values can be taken from these distributional results or more conveniently from tables and graphs in Morrison (1976).

<u>Proof</u>. As in (46) of chapter 1, form the following sub-hypotheses of the main hypothesis (45):

$$H_{0,a} : a'\Omega_{S1} = 0 \quad \text{vs.} \quad H_{1,a} : a'\Omega_{S1} \neq 0 \qquad (85)$$

where a is any fixed non-null $(N - 1) \times 1$ vector.

To test (85), consider the statistic

$$UI_a = T(a'\tilde{\Omega}_{S1})^2/(v_{11}(a'\tilde{\Omega}_{SS}a)), \qquad (86)$$

which is T times the sample correlation coefficient of $\hat{v}_1$ and $[\tilde{v}_2 \ldots \tilde{v}_i \ldots \tilde{v}_N]a$. Under H_0 we have

$$E\, v_1'([v_2 \ldots v_i \ldots v_N]a) = 0 \qquad (87)$$

where

$$v_i = y_i - X_i^* \gamma_i = y_i - X_i\beta_i = u_{\cdot i} \qquad (88)$$

since, if the equations (1) are correctly specified, then the coefficients of γ_i corresponding to columns of X_i^* which are not in X_i are zero. This fact, along with the normality assumption (5) implies that the test statistics (86) have identical distributions for all a under the null. We can then proceed to find a UI statistic by finding the maximum of the UI_a's in (86) over a. This maximum is

$$UI = T\ \tilde{\Omega}_{S1}'\ \tilde{\Omega}_{SS}^{-1}\ \tilde{\Omega}_{S1}/v_{11} \tag{89}$$

Q.E.D.

(v) Remarks

1) From the discussion in chapter 1, section 2, we see that under H_0

$$LR \sim \chi^2_{N-1} \text{ as } T \to \infty \text{ and } LM \sim \chi^2_{N-1} \text{ as } T \to \infty \tag{90}$$

While this does not provide any criteria for choice between LR, LM on asymptotic grounds, it does allow for a rough comparison of rejection probabilities by comparing approximate slopes. Calculation of the approximate slopes AS_{LR}, AS_{LM} is simple due to the asymptotic χ^2 results (90).

Theorem 4. The approximate slopes of LR in (46) and LM in (54) are, respectively,

$$AS_{LR} = -\log(1-R^2) \tag{91}$$

$$AS_{LM} = R^2 \tag{92}$$

where

$$R^2 = \Omega_{S1}' \Omega_{SS}^{-1} \Omega_{S1}/\omega_{11} \tag{93}$$

and the variables in (93) come from the partitioning of Ω in (29).

Proof. Due to the asymptotic χ^2 distributions of LR and LM, we can use

$$AS_{LR} = \lim_{T\to\infty} LR/T \tag{94}$$

and similarly for LM, using a result of Geweke (1980). Since $\hat{\omega}_{11R}$, $\hat{\Omega}_{SSR}$, and $\hat{\Omega}$ are all consistent estimates, we have

$$AS_{LR} = \lim_{T\to\infty} \{T(\log \hat{\omega}_{11R}|\hat{\Omega}_{SSR}| - \log|\hat{\Omega}|)\}/T$$

$$= \log \omega_{11}|\Omega_{SS}| - \log|\Omega|$$

$$= -\log(|\Omega|/\omega_{11}|\Omega_{SS}|). \tag{95}$$

Using a result in Muirhead (1982, p. 581), we have

$$|\Omega| = |\Omega_{SS}|(\omega_{11} - \Omega_{S1}' \Omega_{SS}^{-1} \Omega_{S1}) \tag{96}$$

By substituting (96) into (95), we get the result (91).

For AS_{LM} we note from the consistency of $\hat{\beta}_{1R}$ in (38) and $\hat{\beta}_{SR}$ in (41) that

$$\hat{U}_R' \hat{U}_R/T = \Omega + o_p(1) \tag{97}$$

where

$$\hat{U}_R = [\hat{u}_{\cdot 1R} \; \hat{U}_{SR}] \tag{98}$$

and $\hat{u}_{\cdot 1R}$ and $\hat{U}_{SR}$ are defined in (40) and (43), respectively. Applying (97) to (55) we have

$$\hat{R}^2 = R^2 + o_p(1) \tag{99}$$

Therefore, using (54),

$$AS_{LM} = \lim_{T\to\infty} T\hat{R}^2/T$$

$$= R^2$$

Q.E.D.

Corollary. Under H_1 (i.e., when $R^2 > 0$),

$$AS_{LM} < AS_{LR}. \tag{100}$$

Proof. This follows from noting results (91), (92), and the following inequality from Korovkin (1961, p. 40, inequality (23)):

$$1/(n + 1) < \log(1 + (1/n)) \tag{101}$$

where one can substitute $(1 - R^2)/R^2$ for n to give (92).

Q.E.D.

This result suggests that when asymptotic critical values are used, the LR test (46) will tend to reject H_0 more often than the LM test (54).

2) The special case $X_1 = \ldots = X_i = \ldots = X_N (=X^*)$ is sometimes of interest. This case corresponds to the reduced form system of a simultaneous equation model[3] as well as occurring in some other contexts (e.g., Abel and Mishkin (1983)). First, we note that in this case the restricted ML, unrestricted ML, and OLS estimates of the β_i's coincide (see Kmenta and Gilbert (1968)). It can also be seen that the regression equations (78) will now be the same as the equations of the original model (1). These observations imply that

$$\hat{V} = \hat{U} = \hat{U}_R \tag{102}$$

where these matrices have been defined in (83), (25), and (98), respectively. The result (102) implies that

$$\tilde{\Omega} = \hat{\Omega}, \ \hat{\omega}_{11R} = \hat{\omega}_{11}, \text{ and } \hat{\Omega}_{SSR} = \hat{\Omega}_{SS} \tag{103}$$

<u>Theorem 5</u>. <u>When</u> $X_1 = \ldots = X_i = \ldots = X_N$, <u>the following results hold</u>:

$$LR = -T \log(1 - \hat{R}^2) \quad (104)$$

$$LM = UI = T\hat{R}^2 \quad (105)$$

$$LR > LM = UI \quad (106)$$

where $\hat{R}^2$ is defined in (55), but can be written in a variety of ways due to (102).

Proof. Noting (103), (55) and the definition of LR in (46), the appropriate substitutions and the manipulation of (95) and (96) give result (104).

Result (105) follows from noting (102) and (103), which imply the equivalence of LM in (54) and UI in (84), and (106) follows from inequality (101)[4].

Q.E.D.

3) The tests given in this section can be extended to the case where the hypothesis concerns the independence of the residuals of one set of equations from those of another set. The hypotheses are now still

$$H_0 : \Omega_{S1} = 0 \quad \text{vs.} \quad H_1 : \Omega_{S1} \neq 0, \quad (107)$$

but Ω_{S1} is now an $M \times (N-M)$ matrix, $1 < M < N$, and so we have repartitioned Ω from (29) as

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{1S} \\ \Omega_{S1} & \Omega_{SS} \end{bmatrix} \qquad (108)$$

where Ω_{11} is an $M \times M$ matrix and Ω_{SS} is an $(N-M) \times (N-M)$ matrix, both being positive definite symmetric. We will use this partitioning for $\hat{\Omega}_R$ and $\tilde{\Omega}$ in the definitions of the test statistics.

The restricted ML estimation will now involve the estimation of two separate SURE systems by unrestricted ML corresponding to the first M and remaining (N - M) equations. The LR test is

$$LR = T(\log|\hat{\Omega}_{11R}||\hat{\Omega}_{SSR}| - \log|\hat{\Omega}|), \qquad (109)$$

which has an asymptotic null distribution of

$$LR \sim \chi^2_{M(N-M)} \text{ under } H_0 \text{ as } T \to \infty \qquad (110)$$

from which critical values can be taken.

The UI statistic for testing (107) is

$$UI = ch_1(\tilde{\Omega}_{S1}' \, \tilde{\Omega}_{SS}^{-1} \, \tilde{\Omega}_{S1} \, \tilde{\Omega}_{11}^{-1}) \qquad (111)$$

where $ch_1(D)$ refers to the largest characteristic root of D. An exact test for (107) using UI can be constructed using the distributional result given in Srivastava and Khatri (1979, p. 224).

4) If $\hat{V}$ of (83) is replaced by $\hat{U}_R$ of (98) in the calculation of the UI statistic (84), then we have instead the LM statistic (54). The advantage gained by using $\hat{V}$ from the artificial regressions, then, is a known small sample distribution for the resulting UI statistic.

5) When $\hat{V}$ is used instead of the first-stage $\hat{U}$ matrix of (25), which is composed of OLS residuals using the b_i estimates of (27), in the calculation of the two-step Zellner estimator of section 2(ii), the change in the mean square error of the estimates of β is smaller than $O(T^{-1})$ (see Srivastava and Upadhyaya (1978)). While this fact concerns an estimation problem rather than a testing problem, it may be an indication that the loss of efficiency in the estimation of Ω caused by the misspecification involved in the calculation of $\tilde{\Omega}$ may be small enough to justify the resulting gain in distributional knowledge of UI over LR and LM.

IV.4 TEST STATISTICS IN THE TWO EQUATION CASE

The SURE model with two equations is of special theoretical interest since it is analytically simpler to deal with, and is of some practical interest as well. In this section, tests are presented for the two-equation version of hypotheses (45), which involves a single restriction

$$H_0 : \omega_{21} = 0 \quad \text{vs.} \quad H_1 : \omega_{21} \neq 0 \tag{112}$$

where the partitioning of Ω in (29) now consists of four scalars:

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \tag{113}$$

In this case, restricted ML estimation simply involves two OLS regressions.

(i) LR Test

Corollary (to theorem 1). When N = 2, the LR test statistic for hypotheses (112) is given by

$$LR = T(\log \hat{\omega}_{11R}\, \hat{\omega}_{22R} - \log |\hat{\Omega}|) \tag{114}$$

Note. To perform a test of asymptotic size α,

$$\begin{aligned} &\text{Accept } H_0 \text{ if } LR \leq c(1,\alpha) \\ &\text{Reject } H_0 \text{ if } LR > c(1,\alpha) \end{aligned} \tag{115}$$

where

$$c(1,\alpha) \text{ satisfies } \text{prob}\{\chi_1^2 \leq c(1,\alpha)\} = 1-\alpha$$

Proof. By setting $N = 2$ in (46) and noting the partitioning (113), we get result (114).

Q.E.D.

(ii) LM Test

Corollary (to theorem 2). When $N = 2$, the LM test statistic for hypothesis (112) is given by

$$LM = T\,\hat{r}^2 \tag{116}$$

where

$$\hat{r}^2 = (\hat{u}_{\cdot 1R}' \hat{u}_{\cdot 2R})^2 / (\hat{u}_{\cdot 1R}' \hat{u}_{\cdot 1R})(\hat{u}_{\cdot 2R}' \hat{u}_{\cdot 2R}) \quad (117)$$

is the squared correlation of $\hat{u}_{\cdot 1R}$ and $\hat{u}_{\cdot 2R}$ of (40) and (44).

Note. For a test of asymptotic size α using LM, follow the decision rule (115), replacing LR with LM.

Proof. Since N = 2, the inverted matrix of (55) is a scalar since

$$\hat{U}_{SR} = \hat{u}_{\cdot 2R} \quad (118)$$

and so results (116) and (117) follow.

Q.E.D.

(iii) UI Test

Corollary (to theorem 3). When N = 2, the UI test statistic for hypothesis (112) is given by

$$UI = T(\hat{v}_1' \hat{v}_2)^2 / (\hat{v}_1' \hat{v}_1)(\hat{v}_2' \hat{v}_2) \quad (119)$$

where the $\hat{v}_i$'s are defined in (79).

Note. The selection of a critical value for the exact test can again be taken from Morrison (1976).

Proof. Setting N = 2 and noting (82) and (83) makes the result clear.

Q.E.D.

(iv) *W Test*

By applying the Wald test principle described in chapter 1(iii) to the N = 2 case of model (1) we arrive at the following theorem:

Theorem 6. *When* N = 2, *the* W *test statistic for hypothesis* (112) *is given by*

$$W = T\,\hat{\omega}_{12}^{2}/(\hat{\omega}_{11}\,\hat{\omega}_{22} + \hat{\omega}_{12}^{2}) \qquad (120)$$

where the $\hat{\omega}$'s *come from the partitioning* (113) *of* $\hat{\Omega}$ *in* (24).

Note. For a test having asymptotic size α, using W, follow decision rule (115) replacing LR with W.

Proof. We wish to calculate the right-hand side elements of the formula for W given in equation (31) of chapter 1, repeated here:

$$W = g(\hat{\theta})' [\hat{G}\ I(\hat{\theta})^{-1}\ \hat{G}']^{-1}\ g(\hat{\theta}) \tag{121}$$

where detailed definitions of these elements can be found in chapter 1, section 2(iii).

Since the restriction of H_0 is $\omega_{12} = 0$, we have

$$g(\hat{\theta}) = \hat{\omega}_{12}, \tag{122}$$

and

$$\hat{G}\ I(\hat{\theta})^{-1}\ \hat{G}' = [I(\hat{\theta})^{-1}]_p, \tag{123}$$

where $[I(\hat{\theta})^{-1}]_p$ refers to the diagonal element of $I(\hat{\theta})^{-1}$ corresponding to ω_{12}.

The calculation of (123) is simplified by noting that

$$E\ \partial^2 \ell / \partial\omega_{21}\ \partial\beta_i = 0 \qquad i = 1,2, \tag{124}$$

where ℓ is given in (16), and the derivative formulas of footnote 1 are used. This result allows us to consider only the 3 × 3 diagonal block of $I(\hat{\theta})$ corresponding to the elements of Ω, which we will denote by $I(\hat{\theta})_{\Omega}$. Taking minus the expectation of the second derivatives of ℓ with respect to the elements of Ω, evaluated at $\hat{\Omega}$, we have

$$I(\hat{\theta})_{\Omega} = T|\hat{\Omega}|^{-2} \begin{bmatrix} \hat{\omega}_{22}^2/2 & \hat{\omega}_{12}^2/2 & -\hat{\omega}_{12}\hat{\omega}_{22} \\ \hat{\omega}_{12}^2/2 & \hat{\omega}_{11}^2/2 & -\hat{\omega}_{12}\hat{\omega}_{11} \\ -\hat{\omega}_{12}\hat{\omega}_{22} & -\hat{\omega}_{12}\hat{\omega}_{11} & (\hat{\omega}_{11}\hat{\omega}_{22}+\hat{\omega}_{12}^2) \end{bmatrix} \quad (125)$$

$[I(\hat{\theta})^{-1}]_p$ is the (3,3) element of $[I(\hat{\theta})_{\Omega}]^{-1}$. Using the partitioned inverse formula (Theil (1971, p. 18)) and noting that

$$\Omega^{-1} = |\Omega|^{-1} \begin{bmatrix} \omega_{22} & -\omega_{12} \\ -\omega_{12} & \omega_{11} \end{bmatrix} \quad (126)$$

it can be shown that

$$[I(\hat{\theta})^{-1}]_p = (\hat{\omega}_{11}\, \hat{\omega}_{22} + \hat{\omega}_{12}^2)/T \tag{127}$$

Substituting (127) and (122) into (121), we arrive at the result of the theorem.

Q.E.D.

(v) K Test

A test proposed by Kariya (1981) for the N = 2 case was shown in that paper to be locally best invariant[5] unbiased, and is based on the statistic[6]

$$K = \{(T-k_1)(T-k_2)/T\}\hat{r}^2 - \{(T-k_1)/T\}(1 - R_1^2) - \{(T - k_2)/T\}\ (1 - R_2^2) \tag{128}$$

where R_1^2 is the R^2 from a regression of $\hat{u}._{1R}$ on X_2, R_2^2 is the R^2 from a regression of $\hat{u}._{2R}$ on X_1, and $\hat{r}^2$ is given in (117).

Kariya (1981) gives some approximations to the null distribution of K. An asymptotic critical level could be selected by noting that

$$\lim_{T\to\infty} R_i^2 = 0 \quad \text{under} \quad H_0, \quad i = 1,2 \tag{129}$$

so that

$$\underset{T\to\infty}{\ell im}\, K = \underset{T\to\infty}{\ell im}\, T\, \hat{r}^2 - 2$$

$$= \underset{T\to\infty}{\ell im}\, LM - 2 \text{ under } H_0 \qquad (130)$$

where LM is defined in (116). We could then use (K + 2) in the same way as LR in the decision rule (115) for a test of given asymptotic size.

(vi) LR* Test

Suppose the likelihood ratio test construction principle is applied to the Wishart density result (81) instead of the overall density (12). We can treat $\tilde{\Omega}$ of (82) as the data and Ω as the unknown parameter, use (81) to obtain a likelihood function for Ω given $\tilde{\Omega}$, and proceed. It is a standard statistical result that the unrestricted ML estimate of Ω in this framework is $\tilde{\Omega}$ while the restricted ML estimate (subject to $\omega_{12} = 0$) is simply $\tilde{\Omega}$ again only with $\tilde{\omega}_{12} = 0$, due to a property of marginal Wishart distributions (see Muirhead (1982, p. 92, theorem 3.2.7)). Application of the LR principle in this setting yields

$$LR^* = T\{\log \tilde{\omega}_{11} \tilde{\omega}_{22} - \log|\tilde{\Omega}|\} \qquad (131)$$

where decision rule (115) is used.

(vii) EFT Test

In a recent paper, Harvey and Phillips (1982) propose a test for (112) which is also based on elements of $\tilde{\Omega}$ of (82). This proposed test is known to have an F distribution under the null, hence is called the "exact F test". The statistic is

$$EFT = (T - K^* - 1)\ \tilde{\omega}_{12}^2/|\tilde{\Omega}| \qquad (132)$$

and it is known that

$$EFT \sim F_{1,T-K^*-1} \text{ under } H_0 \qquad (133)$$

(viii) Remarks

1) It is easily shown that the three tests based on $\tilde{\Omega}$ of (82) given above, UI, LR^*, and EFT, are monotonic functions of each other. From their definitions

in (119), (131), and (132), some algebraic manipulation reveals that

$$LR^{*} = -T \log(1-(UI/T)) \tag{134}$$

$$EFT = (T - K^{*} - 1)UI/(T - UI) \tag{135}$$

This implies that these three tests will result in the same decision if exact critical values are used. Since the exact distribution of EFT is well known and tabulated, it would be the best of the three to use.

2) The asymptotic null distributions of all of the above tests are easily derived from previous results.

Theorem 7. The asymptotic null distribution of each of LR, LM, UI, W, (K + 2), LR^{*}, and EFT of sections (i) through (vii) is χ_1^2.

Proof. This result follows naturally for LR, LM, and W from the discussion in section 2 of chapter 1. The result for K then follows from (130).

The result for EFT follows since we have

$$\lim_{b\to\infty} a\, F_{a,b} \equiv \chi_a^2 \tag{136}$$

(see Muirhead (1982)), so that

$$\lim_{T\to\infty} F_{1,T-K^*-1} \equiv \chi_1^2 \qquad (137)$$

For UI and LR^*, note from (134) and (135) that under H_0, when UI has $0_p(1)$, we have

$$\lim_{T\to\infty} LR^* = \lim_{T\to\infty} EFT = \lim_{T\to\infty} UI \qquad (138)$$

From (137) and (138), we get the result for UI and LR^*.

3) Consider the special case $X_1 = X_2$. The following theorem expresses the seven tests (i) through (vii) as monotonic functions of each other in this case.

Theorem 8. When $N = 2$ and $X_1 = X_2$, which implies $k_1 = k_2 = K^*$, the following results hold:

$$LR = LR^* = -T \log(1 - \hat{r}^2) \qquad (139)$$

$$W = T\,\hat{r}^2/(1 + \hat{r}^2) \qquad (140)$$

$$LM = UI = T\cdot \hat{r}^2 \qquad (141)$$

$$K = \{(T - K^*)^2/T\}\hat{r}^2 \qquad (142)$$

$$EFT = (T - K^* - 1)\hat{r}^2/(1 - \hat{r}^2) \qquad (143)$$

where $\hat{r}^2$ is defined in (117), which yields the following inequalities when $\hat{r}^2 > 0$:

$$W < (LM = UI) < (LR = LR^*) \tag{144}$$

$$K < (LM = UI) \tag{145}$$

$$LM < EFT \text{ if and only if } K^* + 1 < T\,\hat{r}^2 \tag{146}$$

$$W < EFT \text{ if and only if } K^* + 1 < (2T - K^* - 1)\,\hat{r}^2 \tag{147}$$

Proof. Note that, as in the N equation case, $X_1 = X_2$ implies that (103) holds. Thus, using (24) and (117), we see that

$$\hat{r}^2 = \hat{\omega}_{12}/\hat{\omega}_{11}\,\hat{\omega}_{22}$$

$$= (\hat{v}_1'\,\hat{v}_2)^2/(\hat{v}_1'\hat{v}_1)(\hat{v}_2'\,\hat{v}_2) \tag{148}$$

so that simple manipulation of the test statistics yields (139) to (143). Result (144) follows from inequality (101), while (145) to (147) follow from simple manipulation.

Q.E.D.

These inequalities can be used to form inequalities for rejection probabilities when asymptotic critical values based on the result of theorem 7 are used. One would want to use EFT in this situation due to its simple small sample null distribution, since (139) to (143) imply that the exact tests are equivalent.

5) Theorem 7 allows a comparison of the approximate slopes of some of the tests for the general case $X_1 \neq X_2$, which are given below.

<u>Theorem 9</u>. <u>When</u> N = 2, <u>the approximate slopes of the tests</u> (i) <u>to</u> (vii) <u>for hypotheses</u> (112) <u>are given by</u>

$$AS_{LR} = AS_{LR^*} = -\log(1 - r^2) \tag{149}$$

$$AS_{LM} = AS_{UI} = AS_K = r^2 \tag{150}$$

$$AS_W = r^2/(1 + r^2) \tag{151}$$

$$AS_{EFT} = r^2/(1 - r^2) \tag{152}$$

<u>where</u>

$$r^2 = \omega_{12}^2/\omega_{11}\,\omega_{22} \tag{153}$$

from which it follows that

$$AS_W < (AS_{LM} = AS_{UI} = AS_K) < (AS_{LR} = AS_{LR*})$$

$$< AS_{EFT} \tag{154}$$

under H_1.

Proof. Under H_1, $\omega_{12} \neq 0$, and $T \to \infty$, the unrestricted ML estimates approach their true values, as do the restricted ML estimates apart, of course, from $\hat{\omega}_{12R} = 0$. This implies that

$$\lim_{T\to\infty}(\hat{u}'_{\cdot 1R}\, \hat{u}_{\cdot 2R})^2/(\hat{u}'_{\cdot 1R}\, \hat{u}_{\cdot 1R})(\hat{u}'_{\cdot 2R}\, \hat{u}_{\cdot 2R}) = r^2 \tag{155}$$

Also, since $\hat{v}_1$ and $\hat{v}_2$ approach their true values, we have

$$\lim_{T\to\infty} (\hat{v}'_1\, \hat{v}_2)^2/(\hat{v}'_1\, \hat{v}_1)(\hat{v}'_2\, \hat{v}_2) = r^2 \tag{156}$$

Since theorem 7 allows us to calculate the approximate slope of each statistic, take LM for an example, by

$$AS_{LM} = \lim_{T\to\infty} LM/T, \qquad (157)$$

it is simple to verify (149) to (152) using the above results.

Using the inequality (Korovkin (1961, pg. 40, (23)))

$$1/(n + 1) < \log(1 + (1/n)) < 1/n, \; n > 0, \qquad (158)$$

and letting $n = (1/r^2)-1$, inequality (154) is established.

Q.E.D.

Note that (154) also indicates a possible ordering for the rejection probabilities of the various tests in small samples at a given asymptotic critical level, as discussed in Geweke (1980).

FOOTNOTES

Chapter 4

1. The following matrix derivative results were used: $\partial \log|\Omega|/\partial\omega_{ij} = \omega^{ij}$, $\partial\omega^{ij}/\partial\omega^{kk} = -\omega^{ik}\omega^{kj}$ and $\partial\omega^{ij}/\partial\omega^{k\ell} = -\omega^{ik}\omega^{\ell j} - \omega^{i\ell}\omega^{kj}$ when $k \neq \ell$ and Ω is symmetric (see Rogers (1980, pg. 80)).

2. Zellner (1962) originally derived this estimator in a generalized least squares framework, but the iterative Zellner estimator is also ML due to the normality assumption (5).

3. Usually, however, the reduced form coefficients will be subject to overidentifying restrictions.

4. This uses the same procedure as for proving (100), replacing R^2 with $\hat{R}^2$.

5. Invariance here refers to invariance to the group of linear transformations of the dependent variables and corresponding linear transformation of the regression coefficients (see chapter 1, section 1.(vi) for a fuller discussion of invariance).

6. The statistic given in Kariya (1981) is actually TK, but it is presented in a form here that is more comparable to the other statistics.

V. TESTS FOR NON-NESTED REGRESSION MODELS

V.1 INTRODUCTION

The hypotheses considered in earlier chapters are nested in the sense that the null hypothesis is a special (restricted) case of the alternative hypothesis. In this chapter the two hypotheses are each single regression equation models of the simplest kind with the same dependent variable, as considered in chapter II, but neither is a restricted case of the other. The linear hypotheses tests used in that chapter then, cannot be used here without some modification.

Several tests have been proposed for this kind of non-nested hypothesis (see Fisher (1983a)). The three tests that are considered in this chapter are the Cox test, first proposed in this context by Pesaran (1974), the J test of Davidson and MacKinnon (1981), and the F test, which is the same as that of chapter II after the hypotheses have been altered so that they are artificially nested[1].

In the following section the hypotheses and tests are described formally, and the local alternative concept used by Pesaran (1982a) is presented. The

Edgeworth expansions for J^2 are derived under the null and alternative, followed by a correction factor which could alleviate the over-rejection problem noted by Davidson and MacKinnon (1983) and Godfrey and Pesaran (1982). Finally, the small sample behaviour of the Cox and J tests are compared using the approximate slope ratios.

V.2 THE HYPOTHESES AND TESTS

(i) The Hypotheses

The hypotheses to be tested are

$$H_0 : y = X \beta_0 + u_0 \text{ vs. } H_1 : y = Z \beta_1 + u_1 \tag{1}$$

where

$$X = [W \quad X_1], \quad Z = [W \quad Z_1],^2 \tag{2}$$

and X is $T \times k_0$, W is $T \times k_c$, X_1 is $T \times k_X$, Z is $T \times k_1$, Z_1 is $T \times k_Z$, and

$$u_i \sim N(0, \sigma_i^2 I_T), \qquad i = 0,1 \tag{3}$$

For the hypotheses to be non-nested we require that k_X and k_Z exceed zero, that no column of X_1 can be expressed as a linear combination of the columns in Z, and similarly, no column of Z_1 can be expressed as a linear combination of the columns in X.

(ii) The Cox Test

This test was first proposed by Pesaran (1974) and was based on suggested procedures for more general non-nested cases by Cox (1961, 1962). The test statistic is

$$N = T \log (\hat{\sigma}_1^2/\hat{\sigma}_{10}^2)/2\{\hat{\sigma}_0^2 \cdot \hat{\beta}_0' X' M_1 M_0 M_1 X \hat{\beta}_0/\hat{\sigma}_{10}^4\}^{1/2} \tag{4}$$

where

$$\hat{\beta}_0 = (X'X)^{-1}X'y; \quad \hat{\sigma}_1^2 = y'M_1y/T; \quad \hat{\sigma}_0^2 = y'M_0y/T;$$

$$\hat{\sigma}_{10}^2 = \hat{\sigma}_0^2 + (\hat{\beta}_0'X'M_1X\hat{\beta}_0)/T; \quad M_0 = I_T - P_0; \quad M_1 = I_T - P_1$$

$$P_0 = X(X'X)^{-1}X'; \quad P_1 = Z(Z'Z)^{-1}Z' \tag{5}$$

It can be shown that asymptotically under the null, $N \sim N(0,1)$. The asymptotic test procedure, then, is a two-tailed test for N having mean zero or, equivalently, a one-tailedtest for N^2 being an observation from a central χ_1^2, since asymptotically under the null, $N^2 \sim \chi_1^2$. The latter form of this test will be considered here.

(iii) The J Test

This test was first proposed by Davidson and MacKinnon (1981). The test statistic is

$$J = \hat{\beta}_1' Z' M_0 y / \hat{\sigma} (\hat{\beta}_1' Z' M_0 Z \hat{\beta}_1)^{1/2} \qquad (6)$$

where

$$\hat{\beta}_1 = (Z'Z)^{-1} Z' y \qquad (7)$$

and $\hat{\sigma}^2$ is the mean square residual from the following regression:

$$y = X\beta + \lambda(Z \hat{\beta}_1) + u \qquad (8)$$

The J test can be interpreted as a test for the significance of the regression coefficient corresponding to the vector of predicted y from OLS on the alternative model when included along with X in the null model. As in the Cox test, asymptotically under the null, $J \sim N(0,1)$, and so $J^2 \sim \chi_1^2$. The asymptotic test, then, would be the same as for the Cox test with J (or J^2) replacing N (or N^2).

(iv) The F test

Another approach is ~~to~~ combine X and Z_1 and regress y on them all. The null hypothesis would then imply that the coefficients on the Z_1 vectors are zero. This can be tested by using the F test discussed in chapter II. The test statistic is

$$F = \{y'(M_0 - M)y/k_Z\}/\{y'My/(T - k_0 - k_Z)\} \tag{9}$$

where

$$M = I - \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'; \quad \tilde{Z} = [X \quad Z_1] \tag{10}$$

is a $T \times (k_0 + k_Z)$ matrix

Under the null hypothesis, $F \sim F_{k_Z, T-k_0-k_Z}$.

An advantage of this test, then, is that its exact null distribution is known. For comparison with N^2 and J^2, the asymptotic result $\lim_{\nu \to \infty} q\, F_{q,\nu} \equiv \chi^2_q$ can be used to show that $k_Z\, F \sim \chi^2_{k_Z}$ under the null asymptotically.

(v) Asymptotic Distributions Under a Local Alternative

Pesaran (1982a) derives the asymptotic distributions of these tests under the local alternative[3]

$$Z = XB + T^{-1/2}\Delta, \tag{11}$$

where

$$\Delta = [0 \quad \Delta_Z] \tag{12}$$

is a $T \times k_1$ matrix, B is $k_0 \times k_1$, Δ_Z is $T \times k_Z$, and the inequality $k_1 \leq k_0$ has been imposed. He considers the LM test instead of the F test since they are equivalent here, as in chapter II. The asymptotic distributions of the three statistics under the local alternative (11) are non-central χ^2, with

$$N^2 \sim \chi_1^2(\delta)$$

$$J^2 \sim \chi_1^2(\delta) \tag{13}$$

$$k_Z F \sim \chi_{k_Z}^2(\delta)$$

where

$$\delta = \beta_{1Z}' \Delta_Z' M_0 \Delta_Z \beta_{1Z}/T\sigma_1^2 \tag{14}$$

is the non-centrality parameter, and

$$\beta_1' = (\beta_{1W}', \beta_{1Z}'), \tag{15}$$

where β_1 is a $k_Z \times 1$ vector partiitoned conformably with Δ_Z.

Using (6) it can be shown that when $k_Z > 1$ then the F test is asymptotically less powerful than the Cox and J tests[4].

In the same paper, a Monte Carlo study confirms that the Cox test is more powerful than the F-test in small samples. Davidson and MacKinnon (1983) and Godfrey and Pesaran (1982), find that the J test has good power properties, but rejects too often in small samples.

V.3 APPROXIMATE DISTRIBUTION OF J^2 AND A SIZE CORRECTION

In the case of nested hypotheses (such as the artificially nested hypotheses tested by the F test) it is only necessary to perform one expansion since the null is just a restricted version of the general case. In a non-nested test such as the J test, however, two expansions are required, one for each hypothesis. These two expansion approximations are given below.

(i) Distribution of J^2 Under the Null

Theorem 1. The Edgeworth approximation to the cumulative density function of J^2 of (6) under H_0 when testing against local alternative[5] (11) is given by

$$\mathrm{pr}(J^2 \leq x) = \mathrm{pr}(\chi_1^2 \leq x) + T^{-1} \sum_{i=0}^{2} \tau_i \, \mathrm{pr}(\chi_{1+2i}^2 \leq x) + o(T^{-1}), \tag{16}$$

where

$$\tau_0 = -(2k_0 + 3)/4, \quad \tau_1 = k_0/2, \quad \tau_2 = 3/4 \tag{17}$$

Proof. From (6) we have

$$J^2 = (\hat{\beta}_1' Z' M_0 y)^2 / \hat{\sigma}^2 (\hat{\beta}_1' Z' M_0 Z \hat{\beta}_1) \tag{18}$$

where

$$y = X\beta_0 + u_0; \quad Z = XB + \Delta/T^{1/2};$$

$$\hat{\beta}_1 = (Z'Z)^{-1} Z' y \tag{19}$$

First, the numerator and denominator of (18) will be expanded to $0_p(T^{-1})$. For the numerator we have

$$(y' M_0 Z \hat{\beta}_1)^2 = (u_0' M_0 P_1 y)^2 \tag{20}$$

with P_1 defined in (5).

Expanding P_1, we have

$$P_1 = P_{XB} + T^{-1/2}[M_{XB}Q_{\Delta,XB} + Q_{XB,\Delta}M_{XB}]$$

$$+ T^{-1}[M_{XB}Q_{\Delta,\Delta}M_{XB} - M_{XB}Q^2_{\Delta,XB} - Q^2_{XB,\Delta}M_{XB}$$

$$- Q_{XB,\Delta}M_{XB}Q_{\Delta,XB}] + T^{-3/2}[Q^2_{XB,\Delta}(M_{XB}Q_{\Delta,XB}$$

$$+ Q_{XB,\Delta}M_{XB}) + (M_{XB}Q_{\Delta,XB} + Q_{XB,\Delta}M_{XB})Q^2_{\Delta,XB}$$

$$- Q_{XB,\Delta}M_{XB}Q_{\Delta,\Delta}M_{XB} - M_{XB}Q_{\Delta,\Delta}Q_{XB,\Delta}M_{XB}$$

$$- M_{XB}Q_{\Delta,XB}Q_{\Delta,\Delta}M_{XB} - M_{XB,\Delta}Q_{\Delta,\Delta}M_{XB}Q_{\Delta,XB}]$$

$$+ \text{"}O(T^{-2})\text{"}^{6} \tag{21}$$

where

$$M_{XB} = I - P_{XB}; \quad P_{XB} = XBA^{-1}B'X';$$

$$Q_{\Delta,XB} = \Delta A^{-1}B'X'; \quad Q_{XB,\Delta} = Q'_{\Delta,XB};$$

$$Q_{\Delta,\Delta} = \Delta A^{-1}\Delta'; \quad A = B'X'XB \tag{22}$$

and B and Δ are from (11) and (12). Using

$$M_0 P_{XB} = 0; \quad M_0 Q_{XB,\Delta} = 0; \quad M_0 M_{XB} = M_0 \qquad (23)$$

we can show that

$$M_0 P_1 = T^{-1/2}[M_0 Q_{\Delta,XB}] + T^{-1}[M_0(Q_{\Delta,\Delta} M_{XB} - Q^2_{\Delta,XB})]$$

$$+ T^{-3/2}[M_0(Q^3_{\Delta,XB} - Q_{\Delta,\Delta}(Q_{XB,\Delta} M_{XB}$$

$$+ M_{XB} Q_{\Delta,XB}) - Q_{\Delta,XB} Q_{\Delta,\Delta} M_{XB})] + O(T^{-2})$$[6]

$$(24)$$

Therefore,

$$u_0' M_0 P_1 y = u_0' M_0 P_1 (X\beta_0 + u_0) = \theta_0 + \theta_{-1/2}$$

$$+ \theta_{-1} + o_p(T^{-1}) \qquad (25)$$

where

$$\theta_0 = T^{-1/2} u_0' M_0 Q_{\Delta,XB} X\beta_0$$

$$\theta_{-1/2} = T^{-1/2} u_0' M_0 Q_{\Delta,XB} \tilde{u}_0$$

$$+ T^{-1} u_0' M_0 Q_{\Delta,\Delta} M_{XB} X \beta_0$$

$$\theta_{-1} = T^{-1}u_0'M_0(Q_{\Delta,\Delta}M_{XB} - Q^2_{\Delta,XB})\tilde{u}_0$$

$$+ T^{-3/2}u_0'M_0(Q_{\Delta,\Delta}Q_{XB,\Delta} + Q_{\Delta,XB}Q_{\Delta,\Delta})M_{XB}X\beta_0$$

$$\tilde{u}_0 = u_0 - T^{-1/2}Q_{\Delta,XB}X\ \beta_0 \qquad (27)$$

and θ_i refers to the term having $O_p(T^i)$ so that the numerator can be written

$$(\beta_1'Z'M_0y)^2 = \theta_0^2 + 2\theta_0\theta_{-1/2} + 2\theta_0\theta_{-1}$$

$$+ \theta^2_{-1/2} + o_p(T^{-1}) \qquad (28)$$

For expansion of the denominator of J^2 of (18), we have

$$\hat{\sigma}^2(\hat{\beta}_1'Z'M_0Z\hat{\beta}_1) = \hat{\sigma}^2\hat{\alpha} \qquad (29)$$

where

$$\hat{\sigma}^2 = y'M^*y/T; \quad M^* = I - X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$$

$$X^* = [X \vdots Z\ \hat{\beta}_1]; \quad \hat{\alpha} = \hat{\beta}_1'Z'M_0Z\ \hat{\beta}_1 = y'P_1M_0P_1y$$

(30)

Using (24) we can write

$$\hat{\alpha} = \gamma_0 + \gamma_{-1/2} + \gamma_{-1} + o_p(T^{-1}) \tag{31}$$

where

$$\gamma_0 = T^{-1} \beta_0' X' Q_{XB,\Delta} M_0 Q_{\Delta,XB} X\beta_0 = \alpha$$

$$\gamma_{-1/2} = 2T^{-1}\beta_0' X' Q_{XB,\Delta} M_0 [Q_{\Delta,XB} \tilde{u}_0 + T^{-1/2} Q_{\Delta,\Delta} M_{XB} X \beta_0]$$

$$\begin{aligned}\gamma_{-1} = {} & T^{-1} \tilde{u}_0' Q_{XB,\Delta} M_0 Q_{\Delta,XB} \tilde{u}_0 \\ & + 2T^{-3/2} \beta_0' X' M_{XB} Q_{\Delta,\Delta} M_0 Q_{\Delta,XB} \tilde{u}_0 \\ & + T^{-2} \beta_0' X' M_{XB} Q_{\Delta,\Delta} M_0 Q_{\Delta,\Delta} M_{XB} X\beta_0 \\ & + 2T^{-3/2} \beta_0' X' Q_{XB,\Delta} M_0 (Q_{\Delta,\Delta} M_{XB} - Q^2_{\Delta,XB}) \tilde{u}_0 \\ & + 2T^{-2} \beta_0' X' Q_{XB,\Delta} M_0 (Q_{\Delta,\Delta} Q_{XB,\Delta} \\ & + Q_{\Delta,XB} Q_{\Delta,\Delta}) M_{XB} X \beta_0 \end{aligned} \tag{32}$$

For $\hat{\sigma}^2$ consider

$$X^{*'}X^{*} = \begin{bmatrix} X'X & X'Z\hat{\beta}_1 \\ \hat{\beta}_1'Z'X & \hat{\beta}_1'Z'Z\hat{\beta}_1 \end{bmatrix} \quad (33)$$

and using the formula for a partitioned inverse (Theil (1971, p. 18)), we have

$(X^{*'}X^{*})^{-1} =$

$$\begin{bmatrix} (X'X)^{-1}+(X'X)^{-1}X'Z\hat{\beta}_1\hat{\beta}_1'Z'X(X'X)^{-1}/\hat{\alpha} & -(X'X)^{-1}X'Z\hat{\beta}_1/\hat{\alpha} \\ -\hat{\beta}_1'Z'X(X'X)^{-1}/\hat{\alpha} & 1/\hat{\alpha} \end{bmatrix} \quad (34)$$

Therefore, M^* of (30) can be written as

$$M^* = M_0 - M_0Z\hat{\beta}_1\hat{\beta}_1'Z'M_0/\hat{\alpha} \quad (35)$$

This gives, after some simplification,

$$y'M^*y = u_0'M_0u_0 - (u_0'M_0Z\hat{\beta}_1)^2/\hat{\alpha} \quad (36)$$

and so the denominator of J^2 of (18) can be written as

$$\hat{\sigma}^2\hat{\alpha} = \hat{\alpha}\ u_0'M_0u_0/T - (u_0'M_0Z\hat{\beta}_1)^2/T \tag{37}$$

Now by using

$$u_0'M_0u_0/T = \sigma_0^2 + \sigma_0^2\delta \tag{38}$$

where

$$\delta = (u_0'M_0u_0/T\sigma_0^2) - 1$$

and noting that δ has $0_p(T^{-1/2})$, we have, by substituting (28), (31), and (38), in (37),

$$\hat{\sigma}^2\hat{\alpha} = \sigma_0^2\{\gamma_0 + (\gamma_{-1/2} + \delta\ \gamma_0) + (\gamma_{-1} + \delta\ \gamma_{-1/2} - \theta_0^2/T\sigma_0^2)\} + o_p(T^{-1}) \tag{39}$$

By inverting and using a Taylor series expansion about $\sigma_0^2\ \gamma_0$, we obtain

$$1/\hat{\sigma}^2\hat{\alpha} = (1/\sigma_0^2\gamma_0)\{1 - \gamma_0^{-1}(\gamma_{-1/2} + \delta\ \gamma_0)$$

$$- \gamma_0^{-1}(\gamma_{-1} + \delta\ \gamma_{-1/2} - \theta_0^2/T\sigma_0^2)$$

$$+ \gamma_0^{-2}(\gamma_{-1/2} + \delta\ \gamma_0)^2\} + o_p(T^{-1}) \quad (40)$$

Now the expansion for J^2 can be found by multiplying (28) and (37) to give

$$J^2 = \eta_0 + \eta_{-1/2} + \eta_{-1} + o_p(T^{-1}) \quad (41)$$

where

$$\eta_0 = \theta_0^2/\sigma_0^2\ \gamma_0$$

$$\eta_{-1/2} = (1/\sigma_0^2\gamma_0)\{2\theta_0\theta_{-1/2} - \theta_0^2\gamma_0^{-1}(\gamma_{-1/2} + \delta\ \gamma_0)\}$$

$$\eta_{-1} = (1/\sigma_0^2\gamma_0)\{\theta_{-1/2}^2 + 2\theta_0\theta_{-1} - 2\gamma_0^{-1}\theta_0\theta_{-1/2}(\gamma_{-1/2}$$

$$+ \delta\ \gamma_0) + \theta_0^2\gamma_0^{-2}(\gamma_{-1/2} + \delta\ \gamma_0)^2$$

$$- \theta_0^2\gamma_0^{-1}(\gamma_{-1} + \delta\ \gamma_{-1/2} - \theta_0^2/T\sigma_0^2)\} \quad (42)$$

Next, we want the moment generating function (m.g.f.) of J^2 to $0_p(T^{-1})$. By Taylor series expansion about $t\ \eta_0$ as in (93) of chapter II

$$E \exp(t\ J^2) = E\{(\exp(t\ \eta_0))(1 + t\ \eta_{-1/2} + t\ \eta_{-1} + t^2\ \eta^2_{-1/2}/2)\} + o_p(T^{-1}) \tag{43}$$

First, consider

$$E \exp t\ \eta_0 = \int_{u_0} \exp t\ \eta_0 f(u_0)\, du_0 \tag{44}$$

where

$$f(u_0) = (2\Pi\sigma_0^2)^{-T/2} \exp(-u_0' u_0/2) \tag{45}$$

and

$$\eta_0 = \theta_0^2/\sigma_0^2\gamma_0 = (u_0' M_0 Q_{\Delta,XB} X\hat{\beta})^2/T\sigma_0^2\alpha \tag{46}$$

where α is defined in (32).

The integral (44) is found by making the following transformation of u_0:

$$z = D^{-1/2}Cu_0; \quad v = X'u_0 \tag{47}$$

where v is a $k_0 \times 1$ vector, z is $(T - k_0) \times 1$, C is a $(T - k_0) \times T$ matrix satisfying

$$CX = 0; \quad CC' = I_{T-k_0}; \quad C'C = M_0 \tag{48}$$

similar to the matrix introduced in (201) of chapter III, and

$$D^{-1} = I - 2t\, C\, Q_{\Delta,XB} X\beta_0\beta_0' X' Q_{XB,\Delta} C'/T\alpha \tag{49}$$

is a $(T - k_0) \times (T - k_0)$ matrix.

Since CX = 0, v and z are independently distributed so that

$$E \exp t\,\eta_0 = \int_z (2\Pi\sigma_0^2)^{-(T-k_0)/2} |D^{-1/2}C\,C'D^{-1/2}|^{-1/2}$$

$$\exp\{-z'z/2\sigma_0^2\}dz \int_v (2\Pi\sigma_0^2)^{k_0/2} |X'X|$$

$$\exp\{-v'(X'X)^{-1}v/2\sigma_0^2\}dv$$

$$= |D|^{1/2} \int_z f_z(z)\,dz \int_v f_v(v)\,dv = |D|^{1/2} \tag{50}$$

where, as in (198) of chapter III, $f_z(z)$ and $f_v(v)$ are p.d.f.'s as if

$$z \sim N(0,\sigma_0^2 I) \text{ and } v \sim N(0,\sigma_0^2 X'X) \tag{51}$$

For evaluating $|D|^{1/2}$, using

$$D = I + (2t/(1-2t))(C\, Q_{\Delta,XB} X\beta_0 \beta_0' X' Q_{XB,\Delta} C'/\alpha) \tag{52}$$

it can be seen that D has $T - k_0 - 1$ eigenvalues equal to one, corresponding to $T - k_0 - 1$ linearly independent eigenvectors, say h_i, $i = 1,\dots,T-k_0-1$, satisfying $\beta_0' X' Q_{XB,\Delta} C' h_i = 0$. The other eigenvalue is $(1 - 2t)^{-1}$ corresponding to the eigenvector $h_T = C\, Q_{\Delta,XB} X\beta_0$, therefore

$$|D| = (1 - 2t)^{-1} \tag{53}$$

and so from (50),

$$E \exp t\, \eta_0 = (1 - 2t)^{-1/2} \tag{54}$$

which is the m.g.f. of a variable distributed as central

χ_1^2.

There are other easier ways to show that the asymptotic null distribution of J^2 is χ_1^2, but this method enables us to derive an approximate distribution which is accurate to $0_p(T^{-1})$. For evaluating the remaining terms of (43), it can be seen from (50) that

$$E\{(\exp t\, \eta_0) t\, \eta_{-1/2}\} = (1-2t)^{-1/2} \int_v \int_z t\, \eta_{-1/2} f_v(v) f_z(z)\, dv\, dz$$

$$= (1-2t)^{-1/2}\, t\, \tilde{E}\, \eta_{-1/2} \qquad (55)$$

with the $\tilde{E}$ notation used in the proofs of theorem 2 of chapter II and theorem 4 of chapter III.

After transforming u_0 to v and z in the terms comprising $\eta_{-1/2}$ and η_{-1}, we have

$$\theta_0 = T^{-1/2} z' D^{1/2} C\, Q_{\Delta,XB} X\beta_0$$

$$\theta_{-1/2} = T^{-1/2} z' D^{1/2} C\, \tilde{v} + T^{-1} z' D^{1/2} C Q_{\Delta,\Delta} M_{XB} X\beta_0$$

$$\theta_{-1} = T^{-1}z'D^{1/2}C[Q_{\Delta,\Delta}\{X((X'X)^{-1} - BA^{-1}B')v + C'D^{1/2}z\} - Q_{\Delta,XB}\Delta A^{-1}B'v - T^{-1/2}(Q_{\Delta,\Delta}M_{XB} - Q^2_{\Delta,XB})Q_{\Delta,XB}X\beta_0] + T^{-3/2}z'D^{1/2}C(Q_{\Delta,\Delta}Q_{XB,\Delta} + Q_{\Delta,XB}Q_{\Delta,\Delta})M_{XB}X\beta_0$$

$$\gamma_0 = \alpha = T^{-1}\beta_0'X'Q_{XB,\Delta}M_0Q_{\Delta,XB}X\beta_0 \qquad (56)$$

which is non-stochastic,

$$\gamma_{-1/2} = 2T^{-1}\beta_0'X'Q_{XB,\Delta}M_0\tilde{v} + 2T^{-3/2}\beta_0'X'Q_{XB,\Delta}M_0 Q_{\Delta,\Delta}M_{XB}X\beta_0$$

$$\gamma_{-1} = T^{-1}\tilde{v}'M_0\tilde{v} + 2T^{-3/2}\beta_0'X'M_{XB}Q_{\Delta,\Delta}M_0\tilde{v} + 2T^{-3/2}\beta_0'X'Q_{XB,\Delta}M_0[\{Q_{\Delta,\Delta}X(X'X)^{-1} - (Q_{\Delta,XB}\Delta + Q_{\Delta,\Delta}B)A^{-1}B'\}v + Q_{\Delta,\Delta}C'D^{1/2}z - T^{-1/2}(Q_{\Delta,\Delta}M_{XB} - Q^2_{\Delta,XB})Q_{\Delta,XB}X\beta_0]$$

...continued

$$+ T^{-2}\beta_0' X' M_{XB} Q_{\Delta,\Delta} M_0 Q_{\Delta,\Delta} M_{XB} X\beta_0$$

$$+ 2T^{-2}\beta_0' X' Q_{XB,\Delta} M_0 (Q_{\Delta,\Delta} Q_{XB,\Delta} + Q_{\Delta,XB} Q_{\Delta,\Delta})$$

$$M_{XB} X\beta_0$$

$$\delta = (z' Dz/T\sigma_0^2) - 1$$

$$\tilde{v} = \Delta A^{-1} B' v - T^{-1/2} Q^2_{\Delta,XB} X\beta_0 \qquad (56)$$

The expectation formulas for quadratic forms given by (209) of chapter III are used to evaluate the expectations of the terms in (42) which form $\eta_{-1/2}$ and η_{-1}.

Define the scalars

$$\psi_1 = \omega_1 M_0 (Q_{\Delta,\Delta}\omega_2 - Q_{\Delta,XB}\omega_1);$$

$$\psi_2 = \omega_1' M_0 Q_{\Delta,\Delta} M_0 \omega_1,$$

$$\psi_3 = \omega_1 M_0 \{(Q_{\Delta,\Delta} Q_{XB,\Delta} + Q_{\Delta,XB} Q_{\Delta,\Delta})\omega_2$$

$$+ (Q^2_{\Delta,XB} - Q_{\Delta,\Delta} M_{XB})\omega_1\}$$

... continued

$$\psi_4 = \omega_1' Q_{XB,\Delta} M_0 (Q_{\Delta,XB}\, \omega_1 - 2Q_{\Delta,\Delta}\omega_2)$$

$$+ \omega_2' Q_{\Delta,\Delta} M_0 Q_{\Delta,\Delta}\omega_2$$

$$\phi = \text{tr}\, M_0 Q_{\Delta,\Delta} \tag{57}$$

where

$$\omega_1 = Q_{\Delta,XB} X\beta_0 \text{ and } \omega_2 = M_{XB} X\beta_0 \tag{58}$$

The ψ_i's each have $O(T)$ while ϕ has $O(1)$.

The expectations to $O(T^{-1})$ for the $\eta_{-1/2}$ term are

$$\tilde{E}\ \theta_0\theta_{-1/2} = \sigma_0^2\ \phi_1/T^{3/2}(1 - 2t)$$

$$\tilde{E}\ \theta_0^2\ \gamma_{-1/2} = 2\sigma_0^2\ \alpha\psi_1/T^{3/2}(1 - 2t)$$

$$\tilde{E}\ \theta_0^2\delta = \sigma_0^2\ \alpha\{3 - (k_0+1)(1-2t)\}/T(1-2t)^2 \tag{59}$$

So using (42), we obtain

$$\tilde{E}\ \eta_{-1/2} = (1-2t)^{-1}(k_0+1)/T - 3(1-2t)^{-2}/T + o(T^{-1}) \tag{60}$$

For the η_{-1} term, we need

$$\tilde{E}\ \theta_0\theta_{-1} = \sigma_0^2\psi_3/T^2(1-2t); \quad \tilde{E}\ \sigma_0^2\delta^2 = 2\sigma_0^2\alpha/T(1-2t)$$

$$\tilde{E}\ \theta_0\theta_{-1/2}\gamma_{-1/2} = 2\sigma_0^2(T\sigma_0^2\psi_2 + \psi_1^2)/T^3(1 - 2t)$$

$$\tilde{E}\ \theta_0\theta_{-1/2}\delta = 0; \quad \tilde{E}\ \theta_0^2\gamma_{-1/2}\delta = 0; \quad \tilde{E}\theta_0^4 = 3\sigma_0^4\alpha^2/(1-2t)^2$$

$$\tilde{E}\ \theta_0^2\gamma_{-1/2}^2 = 4\sigma_0^2\alpha(T\sigma_0^2\psi_2 + \psi_1^2)/T^3(1 - 2t)$$

$$\tilde{E}\ \theta_0^2\gamma_{-1} = \sigma_0^2\alpha(T\sigma_0^2\phi_1 + \psi_4 + 2\psi_3)/T^2(1 - 2t) \tag{61}$$

Using (42) we then have

$$\tilde{E}\ \eta_{-1} = 2t(1-2t)^{-1}[T^2\sigma_0^2\alpha\{(\psi_2/T\alpha) - \phi_1\} + (\psi_1^2 + T\alpha\psi_3)]/T^3\alpha^2 + 2(1 - 2t)^{-1}/T + 3(1 - 2t)^{-2}/T + o(T^{-1}) \tag{62}$$

The $\tilde{E}\ \eta_{-1/2}^2$ term is also needed, and from (42),

$$\eta^2_{-1/2} = (\gamma_0^{-2}/\sigma_0^4)\{4\theta_0^2\theta^2_{-1/2} - 4\gamma_0^{-1}\theta_0^3\theta_{-1/2}(\gamma_{-1/2} + \delta\gamma_0) + \gamma_0^{-2}\theta_0^4(\gamma^2_{-1/2} + 2\gamma_0\gamma_{-1/2}\delta + \gamma_0^2\delta^2)\} \tag{63}$$

The expectations required for $\tilde{E}\,\eta^2_{-1/2}$ are

$$\tilde{E}\;\theta_0^2\theta^2_{-1/2} = \sigma_0^6\{T\alpha\phi_1(1-2t) + 2\psi_2(1+t)\}/T^2(1-2t)^2 + \sigma_0^4\{T\alpha\psi_4(1-2t) + 2\psi_1^2(1+t)\}/T^3(1-2t)^2$$

$$\tilde{E}\;\theta_0^3\theta_{-1/2}\gamma_{-1/2} = 6\sigma_0^4\alpha(T\sigma_0^2\psi_2 + \psi_1^2)/T^3(1-2t)^2$$

$$\tilde{E}\;\theta_0^3\theta_{-1/2}\delta = 0; \quad \tilde{E}\;\theta_0^4\gamma_{-1/2}\delta = 0; \quad \tilde{E}\;\theta_0^4\delta^2 = 6\sigma_0^4\alpha^2/T(1-2t)^2$$

$$\tilde{E}\;\theta_0^4\gamma^2_{-1/2} = 12\sigma_0^4\alpha^2(T\sigma_0^2\psi_2 + \psi_1^2)/T^3(1-2t)^2 \tag{64}$$

Now from (63), we have

$$\tilde{E}\;\eta^2_{-1/2} = 4(1-2t)^{-1}\{T\sigma_0^2(T\phi_1\alpha - \psi_2) + T\alpha\psi_4 - \psi_1^2\}/T^3\alpha^2 + 6(1-2t)^{-2}/T + o(T^{-1}) \tag{65}$$

Combining (60), (62), and (65), after many cancellations, we find that

$$\tilde{E}\{t\,\eta_{-1/2} + t\,\eta_{-1} + t^2\,\eta_{-1/2}^2/2\} = t(1-2t)^{-1}(k_0+3)/T$$
$$+ 3t^2(1-2t)^{-2}/T + o(T^{-1}) \qquad (66)$$

This gives the m.g.f. of J^2 under H_0 to $0_p(T^{-1})$:

$$E(\exp t\,J^2) = (1-2t)^{-1/2}\{1 + T^{-1}\sum_{i=0}^{2}\tau_i(1-2t)^{-i}\}$$
$$+ o(T^{-1}) \qquad (67)$$

where the τ_i's are defined in (17).

Inverting the above gives the c.d.f. of J^2 under H_0 to $0_p(T^{-1})$ given in the theorem (equation (16)).

Q.E.D.

(ii) Distribution of J^2 Under a Local Alternative

Theorem 2[7]. The Edgeworth approximation to the c.d.f. of J^2 of (6) under the local alternative (11) is given by

$$\mathrm{pr}(J^2 \le x) = \mathrm{pr}(\chi_1^2(\varepsilon_0) \le x) + T^{-1}\sum_{i=0}^{4}\tau_i \mathrm{pr}(\chi_{1+2i}^2(\varepsilon_0) \le x) + o(T^{-1}) \qquad (68)$$

where

$$\tau_0 = -(2k_0 + 3)/4 - \phi/2 + \varepsilon_1/2\varepsilon_0$$

$$\tau_1 = k_0/2 - k_0\varepsilon_0/2 + \phi/2 - (5/2)\varepsilon_1/\varepsilon_0$$

$$\tau_2 = 3/4 + (k_0 - 3)\varepsilon_0/2 + \varepsilon_0^2/4 + 4\varepsilon_1/\varepsilon_0$$

$$\tau_3 = 3\varepsilon_0/2 - \varepsilon_0^2/2 - 2\varepsilon_1/\varepsilon_0 \qquad \tau_4 = \varepsilon_0^2/4 \qquad (69)$$

and

$$\varepsilon_0 = \beta_1'\Delta' M_0\Delta\beta_1/T\sigma_1^2; \quad \varepsilon_1 = \beta_1'\Delta' M_0\Delta A^{-1}\Delta' M_0\Delta\beta_1$$

$$\phi = \mathrm{tr}\, M_0 Q_{\Delta,\Delta}; \qquad A = B'X'XB \qquad (70)$$

with $Q_{\Delta,\Delta}$ as in (22), and $\chi_j^2(\varepsilon_0)$ refers to a non-central χ_j^2 variable with j d.f. and non-centrality parameter ε_0.

Proof. The proof will be similar to the proof of theorem 1 and so will be outlined in somewhat less detail. We have

$$J^2 = (\hat{\beta}_1' Z' M_0 y)^2 / \hat{\sigma}^2 (\hat{\beta}_1' Z' M_0 Z \hat{\beta}_1) \tag{71}$$

where

$$y = Z\beta_1 + u_1; \quad Z = XB + \Delta/T^{1/2}$$
$$\hat{\beta}_1 = (Z'Z)^{-1} Z' y, \quad \hat{\sigma}^2 = y' M^* y/T \tag{72}$$

as in (30).

For expansion of the numerator, note that

$$(y' M_0 Z \hat{\beta}_1)^2 = (\beta_1' Z' M_0 Z \beta_1 + \beta_1' Z' M_0 u_1 + \beta_1' Z' M_0 P_1 u_1 + u_1' M_0 P_1 u_1)^2 \tag{73}$$

Using (19), (20), and (72), we can show that

$$(y' M_0 Z\hat{\beta}_1)^2 = \theta_0^{*2} + (2\theta_0^* \theta_{-1/2}^*) + (2\theta_0^*\theta_{-1}^* + \theta_{-1/2}^{*2}) + o_p(T^{-1}) \quad (74)$$

where

$$\theta_0^* = T^{-1/2}\beta_1' \Delta' M_0 \tilde{u}_1;$$

$$\theta_{-1/2}^* = T^{-1/2} u_1' Q_{XB,\Delta} M_0 \tilde{u}_1$$

$$\theta_{-1}^* = T^{-1} u_1' (M_{XB} Q_{\Delta,\Delta} - Q_{XB,\Delta}^2) M_0 \tilde{u}_1 \quad (75)$$

and

$$\tilde{u}_1 = u_1 + T^{-1/2} \Delta \beta_1 \quad (76)$$

For the denominator, applying (72) and (35) to (23) we see that

$$\hat{\sigma}^2 \hat{\alpha} = \hat{\alpha} u_1' M_1 M_0 M_1 u_1 / T - (u_1' M_1 P_0 Z \hat{\beta}_1)^2 / T \quad (77)$$

where $\hat{\alpha}$ is from (30).

Expanding $\hat{\alpha}$ using (24) yields

$$\hat{\alpha} = T^{-1}\beta_1'\Delta' M_0\Delta\beta_1 + 2T^{-1}u_1'Q_{XB,\Delta}M_0\Delta\beta_1$$

$$+ 2T^{-3/2}u_1'(M_{XB}Q_{\Delta,\Delta} - Q_{XB,\Delta})M_0\Delta\beta_1$$

$$+ T^{-1}u_1'Q_{XB,\Delta}M_0Q_{\Delta,XB}u_1 + o_p(T^{-1}) \qquad (78)$$

under H_1 of (1) and (11). Using (22) and (23) it can also be shown that

$$u_1'M_1M_0M_1u_1/T = u_1'M_0u_1/T + o_p(T^{-1})$$

$$= \sigma_1^2 + \sigma_1^2\delta^* + o_p(T^{-1}) \qquad (79)$$

where

$$\delta^* = (u_1'M_0u_1/T\sigma_1^2)-1 \qquad (80)$$

and

$$(u_1'M_1P_0Z\hat{\beta}_1)^2/T = (u_1'M_0\Delta\beta_1)^2/T^2 + o_p(T^{-1}) \qquad (81)$$

It will be useful to define

$$\alpha^* = \beta_1'\Delta' M_0\Delta\beta_1/T \qquad (82)$$

Substituting (78), (79), and (81) into (77),

$$\hat{\sigma}^2\hat{\alpha} = \omega_0 + \omega_{-1/2} + \omega_{-1} + o_p(T^{-1}) \tag{83}$$

where

$$\omega_0 = \alpha^*\sigma_1^2; \quad \omega_{-1/2} = 2\sigma_1^2\zeta_{1/2}/T + \sigma_1^2\alpha^*\delta^*$$

$$\omega_{-1} = (\sigma_1^2 u_1' Q_{XB,\Delta} M_0/T)(2\delta^*\Delta\beta_1 + Q_{\Delta,XB}u_1) + 2\sigma_1^2\zeta_0/T - (u_1' M_0 \Delta\beta_1)^2/T^2 \tag{84}$$

and

$$\zeta_{1/2} = u_1' Q_{XB,\Delta} M_0 \Delta\beta_1$$

$$\zeta_0 = T^{-1/2} u_1' (M_{XB} Q_{\Delta,\Delta} - Q^2_{XB,\Delta}) M_0 \Delta\beta_1 \tag{85}$$

Inverting (83), we obtain

$$1/\hat{\sigma}^2\hat{\alpha} = \omega_0^{-1} - \omega_0^{-2}\,\omega_{-1/2} + (\omega_0^{-3}\omega_{-1/2}^2 - \omega_0^{-2}\,\omega_{-1}) + o_p(T^{-1}) = \gamma_0^* + \gamma_{-1/2}^* + \gamma_{-1}^* + o_p(T^{-1}) \tag{86}$$

where

$$\gamma_0^* = 1/\alpha^*\sigma_1^2; \quad \gamma_{-1/2}^* = (1/\alpha^*\sigma_1^2)(-2\zeta_{1/2}/T\alpha^* - \delta^*)$$

$$\gamma_{-1}^* = (1/T\alpha^{*2}\sigma_1^2)(4\zeta_{1/2}^2/T\alpha^* + 2\delta^*\zeta_{1/2}$$

$$- u_1'Q_{XB,\Delta}M_0Q_{\Delta,XB}u_1 + T\delta^{*2}\alpha^*$$

$$+ (u_1'M_0\Delta\beta_1)^2/T\sigma_1^2 - 2\zeta_0) \qquad (87)$$

Multiplying (73) and (86) gives, under H_1

$$J^2 = \eta_0^* + \eta_{-1/2}^* + \eta_{-1}^* + o_p(T^{-1}) \qquad (88)$$

where

$$\eta_0^* = (\beta_1'\Delta'M_0\tilde{u}_1)^2/T\alpha^*\sigma_1^2 = \theta_0^{*2}/\alpha^*\sigma_1^2$$

$$\eta_{-1/2}^* = \{2\theta_0^*\theta_{-1/2}^* - \theta_0^{*2}(\delta^* + 2\zeta_{1/2}/T\alpha^*)\}/\alpha^*\sigma_1^2$$

$$\eta_1^* = [\theta_{-1/2}^{*2} + 2\theta_0^*\theta_{-1}^* - 2\theta_0^*\theta_{-1/2}^*(\delta^* + 2\zeta_{1/2}/T\alpha^*)$$

$$+ \theta_0^{*2}\{4\zeta_{1/2}^2/T\alpha^* + 2\delta^*\zeta_{1/2} - u_1'Q_{XB,\Delta}M_0Q_{\Delta,XB}u_1$$

...continued

$$+ T\delta^{*2}\alpha^{*} + (u_1' M_0 \Delta \beta_1)^2 / T\sigma_1^2 - 2\zeta_0\}/T\alpha^{*}]/\alpha^{*}\sigma_1^2$$

(89)

The m.g.f. of J^2 is as in (43), with these η^{*}'s replacing the η's. The following transformation of u_1 is used:

$$v = X' u_1; \quad z = D^{*-1/2} C u_1 - 2t\, T^{-1/2} D^{*1/2} C \Delta \beta_1$$

(90)

where C is as in (48), and

$$D^{*-1} = I - 2t\, C\Delta\beta_1\beta_1'\Delta' C' / T\alpha^{*} \qquad (91)$$

is a $(T - k_0) \times (T - k_0)$ matrix.

We have assumed that

$$u_1 \sim N(0, \sigma_1^2 I) \qquad (92)$$

so that v and z have the appropriate independent normal distributions.

By completing the square in the exponent term of $E \exp t\, \eta_0^{*}$, we have

$$E \exp t\eta_0^* = |D^*|^{1/2} \exp\{\alpha^* t/\sigma_1^2(1-2t)\}$$

$$\int_z f_z(z)\,dz \int_v f_v(v)\,dv, \tag{93}$$

where $f_z(z)$ and $f_v(v)$ are the same as in (51) and will be used in the same way.

By a similar argument to that between (52) and (53), using

$$D^* = I + [2t/(1-2t)]C\Delta\beta_1\beta_1'\Delta'C'/T\alpha^* \tag{94}$$

we again have

$$|D^*| = (1 - 2t)^{-1} \tag{95}$$

and so

$$E \exp t\eta_0^* = (1-2t)^{-1/2} \exp\{\alpha^* t/\sigma_1^2(1-2t)\} \tag{96}$$

which is the m.g.f. for a non-central χ^2 with non-centrality parameter α^*/σ_1^2. This corresponds with the result of Pesaran (1982a).

The remaining terms of the m.g.f. will be evaluated as in the first proof, by transforming u_1 to v and z and integrating $\eta^*_{-1/2}$, η^*_{-1} and $\eta^{*2}_{-1/2}$ over v and z weighted by $f_v(v)$ and $f_z(z)$ given by the distributions of (51). This integration will again be denoted by $\tilde{E}$.

The expectation formulas (202) of chapter III will be used again here. In addition, formula (2.2) from Srivastava and Tiwari (1976) is used to show that

$$\tilde{E}(z'D^{*1/2}C\Delta\beta_1\beta_1'\Delta'C'D^{*1/2}z)^2(z'D^*z)^2$$

$$= [3T^3\alpha^{*2}\sigma_1^8/(1-2t)^2][T-2k_0 + 10/(1-2t)]$$

$$+ o(T^3) \qquad (97)$$

which is required for the $\tilde{E}\,\theta_0^{*4}\,\delta^{*2}$ term of $\tilde{E}\,\eta^{*2}_{-1/2}$

To transform u_1 to v and z in the η^*'s, (90) can be inverted, giving

$$u_1 = C'D^{*1/2}z + 2tT^{-1/2}M_0\Delta\beta_1'/(1-2t) + X(X'X)^{-1}v \qquad (98)$$

We also use

$$\beta_1'\Delta'C'D^{*i}C\Delta\beta_1 = T\alpha^*/(1-2t)^i \quad i=1,2,3 \tag{99}$$

The expectations to $O(T^{-1})$ required for $\eta^*_{-1/2}$ are

$$\tilde{E}\ \theta_0^*\theta^*_{-1/2} = 0; \quad \tilde{E}\ \theta_0^{*2}\zeta_{1/2} = 0$$

$$E\ \theta_0^{*2}\delta_0^* = [\sigma_1^2\alpha^*\{3(1-2t)^{-2}-(k_0+1)(1-2t)^{-1}\} + \alpha^{*2}\{(1 + 8t + 4t^2)(1 - 2t)^{-3} - (k_0 + 1)(1 -2t)^{-2}\} + (\alpha^{*3}/\sigma_1^2)\{4t^2(1 - 2t)^{-4}\}]/T \tag{100}$$

so that, using (89)

$$\tilde{E}\ \eta^*_{-1/2} = -\{3(1-2t)^{-2} - (k_0 + 1)(1 - 2t)^{-1}\}/T - \alpha^*\{(1 + 8t + 4t^2)(1 - 2t)^{-3} - (k_0 + 1)(1 - 2t)^{-2}\}/T\sigma_1^2 - \alpha^{*2}\{4t^2(1 - 2t)^{-4}\}/T\sigma_1^4 + O(T^{-1}) \tag{101}$$

For $\tilde{E}\ \eta^*_{-1}$ we need

$$\tilde{E}\ \theta^{*2}_{-1/2} = \sigma_1^4\ \phi/T + (\sigma_1^2\varepsilon_1/T(1-2t))\{2t\sigma_1^2/T\alpha^*$$

$$+ 1/T(1 - 2t)\}$$

$$\tilde{E}\ \theta^*_0\theta^*_{-1} = \sigma_1^2\alpha^*\phi/T(1 - 2t) + \sigma_1^2\varepsilon_1(1 + 4t)/T^2(1-2t)^2$$

$$\tilde{E}\ \delta^*\theta^*_0\theta^*_{-1/2} = 0; \quad \tilde{E}\ \theta^{*2}_0\ \delta^*\zeta_{1/2} = 0$$

$$\tilde{E}\ \theta^*_0\theta^*_{-1/2}\zeta_{1/2} = \sigma_1^4\varepsilon_1/T(1-2t) + \sigma_1^2\alpha^*\varepsilon_1/T(1-2t)^2$$

$$\tilde{E}\ \theta^{*2}_0\zeta^2_{1/2} = \sigma_1^4\alpha^*\varepsilon_1/(1 - 2t) + \sigma_1^2\alpha^{*2}\varepsilon_1/(1-2t)^2$$

$$\tilde{E}\ \theta^{*2}_0 u_1' Q_{XB,\Delta} M_0 Q_{\Delta,XB} u_1 = \{\sigma_1^2\alpha^*\phi/(1-2t)\}(\sigma_1^2$$

$$+ \alpha^*/(1 - 2t))$$

$$\tilde{E}\ \theta^{*2}_0\delta^{*2} = \{2\alpha^*/T(1-2t)\}(\sigma_1^2 + \alpha^*/(1-2t))$$

...continued

$$E\theta_0^{*2}(u_1' M_0 \Delta\beta_1)^2 = 3\sigma_1^4 T\alpha^{*2}/(1-2t)^2 + \sigma_1^2 T\alpha^{*3}(1 + 8t + 4t^2)/(1-2t)^3 + 4T\alpha^{*4}t^2/(1-2t)^4$$

$$E\,\theta_0^{*2}\zeta_0 = 2\sigma_1^2\alpha^*\varepsilon_1(1+t)/T(1-2t)^2 \qquad (102)$$

where ε_1 and ϕ are defined in (70). Now using (89) we obtain

$$\tilde{E}\,\eta_{-1}^* = -2t\sigma_1^2\phi/T\alpha^*(1-2t) + \phi(1-4t)/T(1-2t)^2 + 2t\sigma_1^2\varepsilon_1/T^2\alpha^{*2}(1-2t) - \varepsilon_1(1-4t)/T^2\alpha^*(1-2t)^2 + (5-4t)/T(1-2t)^2 + \alpha^*(3+4t+4t^2)/T\,\sigma_1^2(1-2t)^3 + 4t^2\alpha^{*2}/T\,\sigma_1^4(1-2t)^4 \qquad (103)$$

From (89),

$$\eta_{-1/2}^{*2} \doteq \{4\theta_0^{*2}\theta_{-1/2}^{*2} - 4\theta_0^{*3}\theta_{-1/2}^*(\delta_\rho^* + 2\zeta_{1/2}/T\alpha^*)$$

...continued

$$+ \theta_0^{*4}(\delta^{*2} + 4\delta^* \zeta_{1/2}/T\alpha^* + 4\zeta_{1/2}^2/T^2\alpha^{*2})\}$$

$$/\alpha^{*2}\sigma_1^4 .$$

(104)

$E\ \eta_{-1/2}^{*2}$ requires the following expectations:

$$\tilde{E}\ \theta_0^{*2}\theta_{-1/2}^{*2} \doteq \sigma_1^6\alpha^*\phi/T(1-2t) + \sigma_1^4\alpha^{*2}\phi/T(1-2t)^2$$

$$+ \sigma_1^6\varepsilon_1(2t+2)/T^2(1-2t)^2 + \sigma_1^4\alpha^*\varepsilon_1(5$$

$$+ 2t)/T^2(1-2t)^3 + \sigma_1^2\alpha^{*2}\varepsilon_1/T^2(1-2t)^4$$

$$\tilde{E}\ \theta_0^{*3}\theta_{-1/2}^*\delta^* = 0; \quad \tilde{E}\ \theta_0^{*4}\delta^*\zeta_{1/2} = 0$$

$$\tilde{E}\ \theta_0^{*3}\theta_{-1/2}^*\zeta_{1/2} = 3\sigma_1^6\alpha^*\varepsilon_1/T(1-2t)^2$$

$$+ 6\sigma_1^4\alpha^{*2}\varepsilon_1/T^2(1-2t)^3$$

$$+ \sigma_1^2\alpha^{*3}\varepsilon_1/T^2(1-2t)^4$$

$$\tilde{E}\ \theta_0^{*4}\delta^{*2} = 6\sigma_1^4\alpha^{*2}/T(1-2t)^2 + 12\sigma_1^2\alpha^{*3}/T(1-2t)^3$$

$$+ 2\alpha^{*4}/T(1-2t)^4$$

continued

$$\tilde{E}\ \theta_0^{*4}\zeta_{1/2}^2 = \{\sigma_1^2\alpha^{*2}\varepsilon_1/T(1-2t)^2\}\{3\sigma_1^4 + 2\sigma_1^2\alpha^*/(1-2t)+\alpha^{*2}/(1-2t)^2\} \quad (105)$$

where $\tilde{E}\ \theta_0^{*4}\delta^{*2}$ has used the expectation result (97).

Substituting these in (104),

$$\begin{aligned}\tilde{E}\ \eta_{-1/2}^{*2} = {} & 4\sigma_1^2\phi/T\alpha^*(1-2t) + 4\phi/T(1-2t)^2 \\ & - 4\sigma_1^2\varepsilon_1/T^2\alpha^{*2}(1-2t) \\ & + \varepsilon_1(8t-20)/T^2\alpha^*(1-2t)^3 + 6/T(1-2t)^2 \\ & + 12\alpha^*/T\sigma_1^2(1-2t)^3 + 2\alpha^{*2}/T\sigma_1^4(1-2t)^4 \\ & + o(T^{-1}) \end{aligned} \quad (106)$$

Using (101), (103) and (106) and simplifying we have

$$t(\tilde{E}\,\eta^*_{-1/2} + \tilde{E}\eta^*_{-1}) + t^2\tilde{E}\eta^{*2}_{-1/2}/2$$

$$= \{(3t - 3t^2)/(1-2t)^2 + k_0 t/(1-2t)\}/T$$

$$+ \alpha^*\{3t/(1-2t)^3 + k_0 t/(1-2t)^2\}/T\ \sigma_1^2$$

$$+ \alpha^{*2}\{t^2/(1-2t)^4\}/T\ \sigma_1^4 + \phi t/T(1-2t)$$

$$- \varepsilon_1(t+4t^2+4t^3)/T^2\alpha^*(1-2t)^3 + o(T^{-1}) \qquad (107)$$

This, along with (96), yields a m.g.f. for J^2 under H_1 to $0(T^{-1})$ of

$$E \exp t\,J^2 = (1-2t)^{-1/2} \exp\{\delta^* t/\sigma_1^2(1-2t)\}[1 + T^{-1}\sum_{i=0}^{4}\tau_i(1-2t)^{-i}] + o(T^{-1}) \qquad (108)$$

where the τ_i's are defined in (69).

Inverting the above gives the approximate c.d.f. for J^2 of theorem 2 (equation (68)).

Q.E.D.

(iii) Edgeworth Size Corrections for J^2

As noted in the introduction to this chapter, the actual size of J^2 in small samples has been found to exceed its asymptotic size by a substantial amount in many cases in Monte Carlo studies. In this section, the Edgeworth size corrected statistic is presented along with its approximate distribution under the local alternative. The critical value correction used in chapter II, section 4(iid) and chapter III, section 5(ii) is used again here to provide a size corrected test which avoids the non-rejection problem discussed in chapter II section 4(iic).

Corollary (to theorem 1). The Edgeworth size-corrected statistic $\tilde{J}^2$ based on the approximate distribution of J^2 of (16) under H_0 is

$$\tilde{J}^2 = \{1-T^{-1}(k_0 + 3/2 + J^2/2)\}J^2 \qquad (109)$$

so that under H_0, the distribution of $\tilde{J}^2$ is $\chi_1^2 + o(T^{-1})$.

Proof. Under H_0,

$$\tilde{J}^2 = \eta_0 + \eta_{-1/2} + [\eta_{-1} - T^{-1}\{(k_0 + 3/2)\eta_0 + \eta_0^2/2\}] + o_p(T^{-1}) \quad (110)$$

where the η_i's are defined in (42). Using v and z from (47) and f_v and f_z from (51) along with the $\tilde{E}$ notation described in (55), we have,

$$\tilde{E}\,\eta_0 = (1-2t)^{-1} \text{ and } \tilde{E}\,\eta_0^2 = 3(1-2t)^{-2} \quad (111)$$

From (110) the m.g.f. of $\tilde{J}^2$ to $0(T^{-1})$ under H_0 is given by

$$E \exp t\,\tilde{J}^2 = E \exp t\,J^2 - t(1-2t)^{-1/2}T^{-1}\{(k_0 + 3/2)\tilde{E}\,\eta_0 + \tilde{E}\,\eta_0^2/2\} + o(T^{-1}) \quad (112)$$

Substituting (67) and (111) into (112) and simplifying yields

$$E \exp t\,\tilde{J}^2 = (1-2t)^{-1/2} + o(T^{-1}), \quad (113)$$

which the m.g.f. to $O(T^{-1})$ of a χ_1^2 variable.

Q.E.D.

Unfortunately, $\tilde{J}^2$ of (109) is not a monotonic nondecreasing function of J^2 which leads to a possible non-rejection problem similar to that of W_t^e in chapter II. Maximization of (109) with respect to J^2 yields the following necessary and sufficient condition for non-rejection of $\tilde{J}^2$, i.e., $\tilde{J}^2$ has a size of zero, when a critical value x is used:

$$T\{1-(k_0 + 3/2)/T\}^2/2 < x \tag{114}$$

When the asymptotic size is 5%, the critical value is x = 3.84 so that (114) can be adjusted to give a non-rejection condition of

$$k_0 > T - 2.77\,T^{1/2} - 1.5 \tag{115}$$

Some T, k_0 combinations which satisfy (115) are given below:

TABLE 1.

Boundary Points for Non-Rejection of $\tilde{J}^2$

T	$k_0 \geq$
10	0
15	3
20	7
30	13
40	20
100	70

Thus if the sample size is 10, $\tilde{J}^2$ would never reject H_0 regardless of the number of variables in the null model, k_0. Some of these (T, k_0) combinations could arise in practice so that the critical value correction method is a better approach. First, we require the Edgeworth expansion for J^2 under H_0 expressed in terms of central χ_i^2 p.d.f.'s, which are denoted by $f(x,i,0)$ as in the appendix to chapter II. Using the method of that appendix along with the expansion result (16), we can write the expansion as

$$\text{pr}(J^2 \leq x) = \text{pr}(\chi_1^2 \leq x) + T^{-1} \sum_{i=1}^{2} \tau_i^* f(x,1 + 2i,0) + o(T^{-1}) \qquad (116)$$

where

$$\tau_1^* = -k_0 - 3/2 \text{ and } \tau_2^* = -3/2 \tag{117}$$

Using (188) and (189) of chapter II we can rewrite (116) as

$$pr(J^2 \leq x) = pr(\chi_1^2 \leq x) + T^{-1}(x\tau_1^* + x^2\tau_2^*/3)f(x,1,0) \tag{118}$$

Proceeding again as in chapter II equations (191) to (197), the size-corrected critical value is

$$x^* = [1+\{(k_0 + 3/2)+x/2\}/T]x \tag{119}$$

For example, the asymptotic size of 5% is obtained when x = 3.84 which results in a size corrected critical value of

$$x^* = \{1+(k_0+3/2)/T\}3.84 + 7.37/T \tag{120}$$

Since the two size-correction methods are equivalent to $0(T^{-1})$, an expansion of $\tilde{J}^2$ under H_1 can be used to estimate the power of either size-corrected test. This corollary is given below.

Corollary 2. The statistic $\tilde{J}^2$ in (109) has the following approximate c.d.f. under H_1:

$$\mathrm{pr}(\tilde{J}^2 \le x) = \mathrm{pr}(\chi_1^2(\varepsilon_0) \le x) + T^{-1}\sum_{i=0}^{4}\tau_i \mathrm{pr}(\chi_{1+2i}^2(\varepsilon_0) \le x) + o(T^{-1}) \tag{121}$$

where

$$\tau_0 = -\phi/2 + \varepsilon_1/2\varepsilon_0; \quad \tau_1 = -k_0\varepsilon_0/2 + \phi/2 - (5/2)\varepsilon_1/\varepsilon_0$$

$$\tau_2 = (k_0 - 3)\varepsilon_0/2 + \varepsilon_0^2/4 + 4\varepsilon_1/\varepsilon_0$$

$$\tau_3 = 3\varepsilon_0/2 - \varepsilon_0^2/2 - 2\varepsilon_1/\varepsilon_0 \quad \tau_4 = \varepsilon_0^2/4 \tag{122}$$

and $\phi, \varepsilon_0, \varepsilon_1$ are defined in (70).

Proof. The proof is very similar to that of Corollary 1. Under H_1, $\tilde{J}^2$ is as in (110) with the η's being replaced by the η^*'s of (89). Using v and z from (90) along with f_v and f_z from (51) and the $\tilde{E}$ notation we can show that

$$\tilde{E}\ \eta_0^* = (1-2t)^{-1} \quad \text{and} \quad \tilde{E}\ \eta_0^{*2} = 3(1-2t)^{-2} \tag{123}$$

The m.g.f. of $\tilde{J}^2$ under H_1 is given by

$$E \exp t\ \tilde{J}^2 = E \exp t\ J^2 - (1-2t)^{-1/2}$$
$$\exp\{\alpha^* t/\sigma_1^2(1-2t)\}\ t\{(k_0 + 3/2)\tilde{E}\eta_0$$
$$+ \tilde{E}\ \eta_0^2/2\}/T + o(T^{-1}) \tag{124}$$

where α^* is defined in (82). Substituting (108) and (123) into (124) and simplifying yields

$$E \exp t\ \tilde{J}^2 = (1-2t)^{-1/2} \exp\{\alpha^* t/\sigma_1^2(1-2t)\}$$
$$\{1 + T^{-1} \sum_{i=0}^{4} \tau_i (1-2t)^{-i}\} + o(T^{-1}) \tag{125}$$

where the τ_i's are defined in (122). Inverting this yields the approximate c.d.f. of corollary 2.

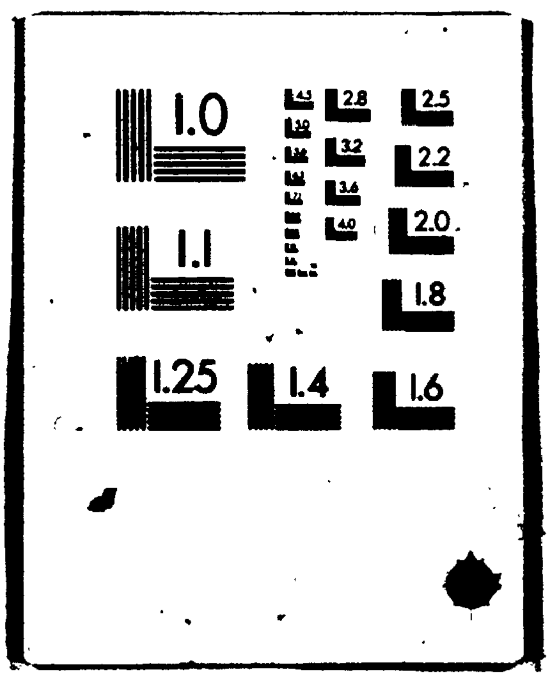
Q.E.D.

This result can be used to approximate the power of both the test statistic adjusted test and the critical value adjusted test by substituting the unadjusted critical value in for x in (121), which will be the subject of future research.

V.4 COMPARISON OF J TEST AND COX TEST BY APPROXIMATE SLOPE

Since Pesaran (1982a) has shown that the Cox and J tests have identical distributions asymptotically under both the null and local alternatives, it may be useful to compare their approximate slopes to look for non-local alternatives for which they differ. An advantage of this non-local method is that there is no restriction required on the alternative unlike the results of the previous section where $k_1 \leq k_0$ was required in defining the local alternative in section 2(v). The general formula for the ratio of approximate slopes of the Cox and J tests (ASR) is derived, and some special cases are discussed.

4 OF/DE 4



(i) Derivation of Approximate Slope Ratio (ASR)

As was mentioned in chapter I, a test having an asymptotic central χ^2 distribution under the null has an approximate slope equal to the limit of the ratio of the test statistic to sample size as the sample approaches infinity (Geweke (1981)). For a fixed alternative and sample of size T, this limit is interpreted as the limiting value of the test statistic when the parameter estimates take on their asymptotic values (which may not be their true values since the estimation is performed under the false hypothesis) divided by T.

The following definitions will condense the notation required for the result in theorem 3. Consider these four regressions which can be performed in sequence:

Regression:

1) $Z_1\beta_Z$ on W; gives prediction $P_W Z_1 \beta_Z$

residual $M_W Z_1 \beta_Z$

2) $M_W Z_1 \beta_Z$ on $M_W X_1$; gives prediction $P_X M_W Z_1 \beta_Z$

$= P_X Z_1 \beta_Z$

residual $M_X M_W Z_1 \beta_Z$

...continued

3) $P_X Z_1 \beta_Z$ on $M_W Z_1$; gives prediction $P_Z P_X Z_1 \beta_Z$

residual $M_Z P_X Z_1 \beta_Z$

4) $P_Z P_X Z_1 \beta_Z$ on $M_W X_1$; gives prediction $P_X P_Z P_X Z_1 \beta_Z$

residual $M_X P_Z P_X Z_1 \beta_Z$

(126)

and P_X is defined in (131), where

$$M_W = I - P_W \text{ and } P_W = W(W'W)^{-1}W' \qquad (126a)$$

Now define

R_i = (sum of squares of prediction vector from regression i)$/T\sigma_1^2$

E_i = (sum of squares of error vector from regression i)$/T\sigma_1^2$

(127)

<u>Theorem 3.</u> <u>The approximate slope ratio</u> (ASR) <u>of the Cox test statistic</u> N^2 <u>and the</u> J <u>test statistic</u> J^2 <u>is given by</u>

$$ASR = \lim_{T\to\infty} N^2/J^2 = \{\log(1 + E_1 - R_3)\}^2 (1 + E_1 - R_3)^2 / 4E_2E_4(1 + E_2) \quad (128)$$

<u>where</u> E_i <u>and</u> R_i <u>are given by</u> (127).

<u>Proof</u>. For the Cox test, we need

$$\lim_{T\to\infty} N^2/T = \lim_{T\to\infty} T[\log(\hat{\sigma}_1^2/\hat{\sigma}_{10}^2)]^2 (\hat{\sigma}_{10}^4/\hat{\sigma}_0^2) (\hat{\beta}_0' X' M_1 M_0 M_1 X \hat{\beta}_0)^{-1}/4 \quad (129)$$

which results from distributional result $N^2 \sim \chi_1^2$ and squaring N of (4).

For notational simplicity, $\lim_{T\to\infty} a = b$ will be represented by $a \to b$.

First, we have

$$X\hat{\beta}_0 = P_0 y \quad \text{and} \quad P_0 = X(X'X)^{-1}X' = P_W + P_X \quad (130)$$

where

$$P_X = M_W X_1 (X_1' M_W X_1)^{-1} X_1' M_W; \quad P_W = W(W'W)^{-1}W' \tag{131}$$

and the partitionings (2) are used, and similarly,

$$P_1 = P_W + P_Z \tag{132}$$

where

$$P_Z = M_W Z_1 (Z_1' M_W Z_1)^{-1} Z_1' M_W \tag{133}$$

Since H_1 is true

$$\hat{\sigma}_1^2 \rightarrow \sigma_1^2 \tag{134}$$

Also,

$$\hat{\sigma}_0^2 = y' M_0 y/T = (\beta_W' W' + \beta_Z' Z_1' + u_1') M_0$$

$$(W\beta_W + Z_1 \beta_Z + u_1)/T = (\beta_Z' Z_1' + u_1') M_0$$

$$(Z_1 \beta_Z + u_1)/T \tag{135}$$

Using

$$M_0 = I - P_W - P_X = (I-P_W)(I-P_X)(I-P_W)$$

$$= M_W M_X M_W \tag{136}$$

which holds since

$$P_W P_X = P_W P_Z = 0 \qquad (137)$$

we have

$$\hat{\sigma}_0^2 = \sigma_1^2 + (\beta_Z' Z_1' M_W M_X M_W Z_1 \beta_Z)/T \qquad (138)$$

Using

$$P_0 M_1 P_0 = P_X M_Z P_X \qquad (139)$$

then we have

$$\hat{\beta}_0' X' M_1 X \hat{\beta}_0/T \rightarrow y' P_0 M_1 P_0 y/T$$

$$\rightarrow \beta_Z' Z_1' P_X M_Z P_X Z_1 \beta_Z/T \qquad (140)$$

and so using $\hat{\sigma}_{10}^2$ of (5) along with (135) and (137) we get

$$\hat{\sigma}_{10}^2 \rightarrow \sigma_1^2 + \beta_Z' Z_1' (M_W M_X M_W + P_X M_Z P_X) Z_1 \beta_Z/T \qquad (141)$$

But since

$$M_W M_X M_W + P_X M_Z P_X = M_W - P_X P_Z P_X \qquad (142)$$

then

$$\hat{\sigma}^2_{10} \to \sigma^2_1 + \beta_Z' Z_1' M_W Z_1 \beta_Z / T$$

$$- \beta_Z' Z_1' P_X P_Z P_X Z_1 \beta_Z / T \qquad (143)$$

Finally,

$$\hat{\beta}_0' X' M_1 M_0 M_1 X \hat{\beta}_0 = y' P_0 M_1 M_0 M_1 P_0 y \qquad (144)$$

and using

$$M_0 M_1 P_0 = M_X P_Z P_X \qquad (145)$$

then

$$(\hat{\beta}_0' X' M_1 M_0 M_1 X \hat{\beta}_0)/T \to \beta_Z' Z_1' P_X P_Z M_X P_Z P_X Z_1 \beta_Z / T \qquad (146)$$

Using the E_i, R_i notation of (127) we can use (138), (143), and (146) to write

$$\hat{\sigma}^2_0 \to \sigma^2_1 (1 + E_2)$$

$$\hat{\sigma}^2_1 \to \sigma^2_1 .$$

$$\hat{\sigma}^2_{10} \to \sigma^2_1 (1 + E_1 - R_3)$$

...continued

$$\hat{\beta}_0' X' M_1 M_0 M_1 X \hat{\beta}_0 / T \rightarrow E_4 \tag{147}$$

Inserting these limits into (129) yields

$$N^2/T \rightarrow \{\log(1 + E_1 - R_3)\}^2 (1 + E_1 - R_3)^2 / 4E_4 (1 + E_2) \tag{148}$$

The approximate slope of J^2 is still required and from (6)

$$J^2/T \rightarrow \lim_{T\rightarrow\infty} (\hat{\beta}_1' Z' M_0 y)^2 / \hat{\sigma}^2 (\hat{\beta}_1' Z' M_0 Z \hat{\beta}_1) \tag{149}$$

Using (132) and (136) we have

$$\begin{aligned} \hat{\beta}_1' Z' M_0 y/T &\Rightarrow y' P_1 M_0 y/T \\ &\rightarrow \beta_Z' Z_1' M_W Z_1 (Z_1' M_W Z_1)^{-1} Z_1' M_W M_X Z_1 \beta_Z / T \\ &= \beta_Z' Z_1' M_W M_X Z_1 \beta_Z / T = \beta_Z' Z_1' M_W M_X M_W Z_1 \beta_Z / T \\ &= \sigma_1^2 E_2 \end{aligned} \tag{150}$$

and

$$\hat{\beta}_1' Z' M_0 Z \hat{\beta}_1/T = y' P_1 M_0 P_1 y/T = y_1' P_1 M_W M_X M_W P_1 y/T$$

$$\rightarrow \beta_Z' Z_1' M_W M_X M_W Z_1 \beta_Z/T = \sigma_1^2 E_2 \tag{151}$$

Also, noting that $\hat{\sigma}^2$ is from regression (8) gives, under H_1

$$\hat{\sigma}^2 \rightarrow \sigma_1^2 \tag{152}$$

Substituting these limits into (149) yields

$$J^2/T \rightarrow E_2 \tag{153}$$

and taking the ratio of (148) and (153) gives the result of the theorem in (128).

Q.E.D.

(ii) ASR Values in Three Special Cases

If ASR exceeds one for particular parameter values and data, one might suspect that the Cox test is more likely to reject the null (i.e., give a correct decision) than the J test when both use the same critical value, and vice versa. This may be due to a difference in power or a difference in actual size, or both. In any case, an ASR which differs substantially from one may indicate a situation where the small sample behaviour of the two tests is substantially different. With this in mind, we examine the behaviour of ASR of (128) in three special cases.

Case 1. Z_1 and X_1 are both vectors. In this case the E_i's and R_i's of (127) reduce to

$$E_i = E_1\rho_1^{2(i-2)}(1-\rho_1^2); \quad R_i = E_1\rho_1^{2(i-1)}, \; i=2,3,4 \tag{154}$$

where

$$\rho_1^2 = (X_1'M_WZ_1)^2/X_1'M_WX_1Z_1'M_WZ_1 \tag{155}$$

is the squared correlation between M_WX_1 and M_WZ_1. Substituting (154) in (128) yields

$$ASR_1 = \{\log(1+E_1(1-\rho_1^4))\}^2(1+E_1(1-\rho_1^4))^2 / 4E_1^2\rho_1^4(1-\rho_1^2)^2(1+E_1(1-\rho_1^2)) \tag{156}$$

from which we can derive the following limiting cases:

$$\text{(i)} \quad \lim_{E_1 \to \infty} ASR_1 = 0$$

$$\text{(ii)} \quad \lim_{E_1 \to 0} ASR_1 = (1 - \rho_1^4)^2$$

$$\text{(iii)} \quad \lim_{\rho_1^2 \to 1} ASR_1 = 1$$

$$\text{(iv)} \quad \lim_{\rho_1^2 \to 0} ASR_1 = \infty \qquad (157)$$

Result (i) indicates that when σ_1^2 is very small compared with the additional explanatory power of Z_1 (given W) in explaining y, J^2 may reject more often than N^2. Result (iv) suggests that when the contributions of X_1 and Z_1 (given W) to explaining y are not very correlated, N^2 may reject more often[8], while (iii) suggests that when their contributions are very correlated (either positively or negatively) then the tests behave similarly.

<u>Case 2</u>. Z_1 is a vector. In this case the results of case 1 still hold, except that now ρ_1^2 is replaced by

$$\rho_2^2 = Z_1'M_WX_1(X_1'M_WX_1)^{-1}X_1'M_WZ_1/Z_1'M_WZ_1. \qquad (158)$$

which is the R^2 of multiple correlation one would obtain from regressing $M_W Z_1$ on $M_W X_1$. Interpretation of this result is similar to that in case 1.

Case 3. X_1 is a vector. Now the E_i's and R_i's of (127) become

$$E_i = E_1 \gamma^{2(i-2)}(1-\rho_3^2); \quad R_i = E_1 \gamma^2 \rho_3^{2(i-2)} \qquad (159)$$
$$i=2,3,4 .$$

where

$$\gamma^2 = (\beta_Z' Z_1' M_W X_1)^2 / X_1' M_W X_1 (\beta_Z' Z_1' M_W Z_1 \beta_Z)$$

$$\rho_3^2 = X_1' M_W Z_1 (Z_1' M_W Z_1)^{-1} Z_1' M_W X_1 / X_1' M_W X_1 \qquad (160)$$

and γ^2 is the squared correlation between $M_W X_1$ and $M_W Z_1 \beta_Z$ while ρ_3^2 is now the R^2 from regressing $M_W X_1$ on $M_W Z_1$.

This yields

$$ASR_3 = \{\log(1 + E_1(1 - \rho_3^2\gamma^2))\}^2 \{1 + E_1(1 - \rho_3^2\gamma^2)\}^2 / 4E_1^2 \rho_3^2 \gamma^2 (1 - \rho_3^2)(1 - \gamma^2) \cdot \{1 + E_1(1 - \gamma^2)\} \qquad (161)$$

Since $\rho_3^2 = 0$ implies that $\gamma^2 = 0$, it can be shown that (157 iv) still holds, with the same interpretation as given in case 1. Other limiting values of ASR_3 require care in interpretation since there is clearly a close relationship between γ^2 and ρ_3^2 of (160).

The importance of ρ_1^2 of (155), ρ_2^2 of (158), and γ^2 and ρ_3^2 of (131) in determining ASR indicates that their values might be of interest when evaluating the two tests in Monte Carlo studies when the hypotheses fall into one of the three cases. For example, Davidson and MacKinnon (1983, table 1) find that the Cox test rejects the null more often than the J test in an experiment which falls into case 3 of this section, and this may be explained by the ρ^2 and γ^2 values of their data.

V.5 CONCLUSION

In this chapter some small sample issues concerning the testing of non-nested hypotheses are considered. The distribution of J^2 under both alternatives is approximated by an Edgeworth expansion, and a size correction factor based on these approximations is proposed with the aim of reducing the over-rejection of the null in small samples. The approximate slope ratio technique is used in order to isolate cases where the small sample behaviour of the Cox and J tests may differ substantially. It is found that in certain cases this ratio depends on the correlation between the contributions of the non-overlapping columns of the two matrices of exogenous variables corresponding to the two models after removing the effect of the overlapping columns (common to both matrices).

__________, Ullah, A., and Srivastava, V.K., "Efficiency of Estimators in the Regression Model with First Order Autoregressive Errors," in M.L. King and D.E.A. Giles, eds., Specification Analysis in the Linear Model: Essays in Honour of Donald Cochrane, Routledge and Kegan Paul, London, forthcoming, 1984.

Mann, H.B., and Wald, A., "On Stochastic Limit and Order Relationships," Annals of Mathematical Statistics, 14 (1943), 217-226.

Morrison, D.E., and Henkel, R.E., The Significance Test Controversy: A Reader, Aldine, Chicago, 1970.

Morrison, D.F., Multivariate Statistical Methods, second edition, McGraw-Hill, New York, 1976.

Muirhead, R.J., Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.

Nagar, A.L., "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," Econometrica, 27 (1959), 575-595.

Neyman, J., and E.S. Pearson, "Contributions to the Theory of Testing Statistical Hypotheses, Part I," Statistical Research Memoirs, 1 (1936), 1-37.

__________, and __________, "Contributions to the Theory of Testing Statistical Hypotheses, Parts II and III," Statistical Research Memoirs, 2 (1938), 25-57.

Park, R.E., and Mitchell, B.M., "Estimating the Autocorrelated Error Model with Trended Data," Journal of Econometrics, 13 (1980), 185-201.

Peers, H.W., "Likelihood Ratio and Associated Test Criteria," Biometrika, 58 (1971), 577-587.

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

__________, "Comparison of Local Power of Alternative Tests of Non-Nested Regression Models," Econometrica, 50, (1982a), 1287-1306.

__________, "A Critique of the Proposed Tests of the Natural Rate-Rational Expectations Hypothesis," Economic Journal, 92 (1982b), 529-554.

Phillips, P.C.B., "Approsimations to Some Finite Sample Distributions Associated with a First Order Stochastic Difference Equation," Econometrica, 45 (1977), 463-485.

__________, "ERA's: A New Approach to Small Sample Theory," Econometrica, 51 (1983), 1505-1526.

8. This result relates to a suggestion of Godfrey and Pesaran (1982) that the Cox test may reject too often, especially when "the explanatory variables of the rival models are not (highly) collinear". Over-rejection is a possible source of a relatively high approximate slope.

SUMMARY

The principal results are summarized along with possible directions for future research.

Chapter II. The small sample distributions of LR, W, and LM tests for linear restrictions on the regression coefficients in a model with multivariate Student's t errors are approximated by Edgeworth expansion. Edgeworth and degrees-of-freedom based size correction factors are examined, and the Edgeworth corrections perform well. Similar results for other error distributions would be of interest.

Chapter III. The LM test statistic for the presence of first-order autocorrelation in the disturbances is identical under normal and multivariate Student's t disturbances. Nagar expansion techniques are used to analyze mean square errors of various estimators when this autocorrelation is present. Many are found to have identical mean square errors to the selected order of approximation. The results correspond largely with previous analytic and Monte Carlo studies. Finally, the distribution of a t-type statistic is approximated

by Edgeworth expansion and its size correction factor derived.

The expansion technique introduced for iterative estimators could be used in many other contexts. A simple size corrected t-test in this model would be very useful.

Chapter IV. Several tests are proposed for determining whether the errors of a regression equation are correlated with errors from N-1 other equations in a SURE model. Their approximate slopes are compared and relationships in special cases are given. Future research could involve analysis of small sample properties of these statistics, and also a study of various pre-test estimators.

Chapter V. The null and alternative distributions of the squared J test statistic for model selection are approximated by Edgeworth expansion and a size correction is proposed. The approximate slopes of J^2 and the squared Cox statistic are compared. The corrected J test should be evaluated by Monte Carlo simulation, and a correction for the Cox statistic would be of use. Power comparisons by simulation could examine cases where the approximate slope ratios indicate that the Cox and J tests may differ greatly.

REFERENCES

Abel, A.B., and Mishkin, F.S., "On the Econometric Testing of Rationality-Market Efficiency," The Review of Economics and Statistics, 65 (1983), 318-323.

Aitchison, J., and Silvey, S.D., "Maximum Likelihood Estimation of Parameters Subject to Restraints," Annals of Mathematical Statistics, 29 (1958), 813-828.

Bahadur, R.R., "Stochastic Comparisons of Tests," Annals of Mathematical Statistics, 31 (1960), 276-295.

Beach, C.M., and MacKinnon, J.G., "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," Econometrica, 46 (1978), 51-58.

Berndt, E.R., and Savin, N.E., "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," Econometrica, 45 (1977), 1263-1278.

Breusch, T.S., "Conflict Among Criteria for Testing Hypotheses: Extension and Comment," Econometrica, 47 (1979), 203-208.

____________, "Useful Invariance Results for Generalized Regression Models," Journal of Econometrics, 13 (1980), 327-340.

____________, and Pagan, A.R., "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," Review of Economic Studies, 47 (1980), 239-254.

Buse, A., "The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note," The American Statistician, 36 (1982), 153-157.

Cochrane, D., and Orcutt, G.H., "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, 44 (1949), 32-61.

Conniffe, D., "A Note on Seemingly Unrelated Regressions," Econometrica, 50 (1982), 229-233.

Cox, D.R., "Tests of Separate Families of Hypotheses," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, U. of California Press, Berkeley, 1961, 105-123.

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

Das Gupta, S., and Perlman, M.D., "Power of the Non-Central F Test: Effect of Additional Variates on Hotelling's T^2-Test," Journal of the American Statistical Association, 69 (1974), 174-180.

Davidson, R., and MacKinnon, J., "Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test," Queen's University Economics Department Discussion Paper #439, 1981a.

____________, and ____________, "Several Tests for Model Specification in the Presence of Alternative Hypotheses," Econometrica, 49 (1981b), 781-793.

____________, and ____________, "Some Non-Nested Hypothesis Tests and the Relations Among Them," Review of Economic Studies, 49 (1982), 551-565.

Durbin, J., "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, Series B, 22 (1960), 139-153.

____________, and Watson, G.S., "Testing for Serial Correlation in Least Squares Regression, I," Biometrika, 37 (1950), 409-428.

Efron, B., and Hinckley, D.V., "Assessing the Accuracy of the Maximum Lieklihood Estimator: Observed versus Expected Information," Biometrika, 65 (1978), 457-487.

Engle, R.F., "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics," in Z. Griliches and M. Intriligator, eds., Handbook of Econometrics, North Holland, Amsterdam, 1981.

Evans, G.B.A., and Savin, N.E., "Conflict Among the Criteria Revisited; The W, LR, and LM Tests," Econometrica, 50 (1982), 737-748.

Evans, G.B.A., and Savin, N.E., "Testing for Unit Roots: 2," mimeograph, Trinity College, Cambridge, April 1983.

Fang, C., and Krishnaiah, P.R., "Asymptotic Distributions of Functions of the Eigenvalues of the Doubly Non-Central Multivariate F Matrix," University of Pittsburgh, Centre for Multivariate Analysis, Technical Report #82-10, 1982.

Fisher, G., "Tests for Two Separate Regressions," Journal of Econometrics, 21 (1983a), 117-132.

Fisher, G., "A Guide to Testing Separate Regressions," paper presented at Canadian Economics Association Meetings, June 1983b.

__________, and McAleer, M., "Principles and Methods in the Testing of Alternative Models," Discussion Paper No. 400, Queen's University, 1980.

Geweke, J., "The Approximate Slopes of Econometric Tests," Econometrica, 49 (1981), 1427-1442.

__________, "A Comparison of Tests of the Independence of Two Covariance-Stationary Time Series," Journal of the American Statistical Association, 76 (1981), 363-373.

Godfrey, L.G., "On the Invariance of the Lagrange Multiplier Test With Respect to Certain Changes in the Alternative Hypothesis," Econometrica, 49 (1981), 1443-1456.

__________, and Pesaran, M.H., "Tests of Non-Nested Regression Models: Small Sample Adjustments and Monte Carlo Evidence," Working Paper in Economics and Econometrics, No. 078, Australian National University, 1982.

Graybill, F.A., An Introduction to Linear Statistical Models, McGraw-Hill, New York, 1961.

Griliches, Z., and Rao, P., "Small-Sample Properties of Several Two Stage Regression Methods in the Context of Auto-Correlated Disturbances," Journal of the American Statistical Association, 64 (1969), 253-272.

Harris, P., and Peers, H.W., "The Local Power of the Efficient Scores Test Statistic," Biometrika, 67 (1980), 525-529.

Harvey, A.C., The Econometric Analysis of Time Series, Philip Alan, Oxford, 1981.

__________, and McAvinchey, D., "The Small Sample Efficiency of Two-Step Estimators in Regression Models with Autoregressive Disturbances," University of British Columbia, Economics Discussion Paper, No. 78-10, 1978.

__________, and Phillips, G.D.A., "Testing for Contemporaneous Correlation of Disturbances in Systems of Regression Equations," Bulletin of Economic Research, 34 (1982), 79-91.

Hayakawa, T., "The Likelihood Ratio Criterion for a Composite Hypothesis Under a Local Alternative," Biometrika, 62 (1975), 451-460.

Hoque, A., The Finite Sample Analysis of Three Dynamic Models, unpublished thesis, University of Western Ontario, 1980.

Isaacson, S.L., "On the Theory of Unbiased Tests of Simple Statistical Hypotheses Specifying the Values of Two or More Parameters," Annals of Mathematical Statistics, 22 (1951), 217-234.

Johnson, N.L., and Kotz, S., Continuous Multivariate Distributions, Wiley, New York, 1972.

Judge, G.G., Griffiths, W.E., Hill, R.C., and Lee, T.C., The Theory and Practice of Econometrics, Wiley, New York, 1980.

Kadane, J., "Comparison of k-Class Estimators When Disturbances are Small," Econometrica, 39 (1971) 723-737.

Kadiyala, K.R., "A Transformation Used to Circumvent the Problem of Autocorrelation," Econometrica, 36 (1968), 93-96.

Kariya, T., "Tests for the Independence Between Two Seemingly Unrelated Regression Equations," The Annals of Statistics, 9 (1981), 381-390.

Kariya, T., and Eaton, M.L., "Robust Tests for Spherical Symmetry," The Annals of Statistics, 5 (1977), 206-215.

Kendall, M.G., and Stuart, A.S., The Advanced Theory of Statistics, Volume 1, third edition, Griffin, London, 1969.

King., M.L., "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," Annals of Statistics, 8 (1980), 1265-1271.

____________, "Testing for Autocorrelation in Linear Regression Models: A Survey," in M.L. King and D.E.A. Giles, eds., Specification Analysis in the Linear Model: Essays in Honour of Donald Cochrane, Routledge and Kegan Paul, London, forthcoming, 1984.

Kmenta, J., and Gilbert, R.F., "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions," Journal of the American Statistical Association, 63 (1968), 1180-1200.

Koenker, R., and Bassett Jr., G., "Robust Tests for Heteroscedasticity Based on Regression Quantiles," Econometrica, 50 (1982), 43-62.

Kokkelenberg, E.C., "Interrelated Factor Demands Revisited," The Review of Economics and Statistics, 65 (1983), 342-347.

Korovkin, P.P., Inequalities, Pergamon Press, New York, 1961.

Kramer, W., "Finite Sample Efficiency of Ordinary Least Squares in the Linear Regression Model with Autocorrelated Errors," Journal of the American Statistical Association, 75 (1980), 1005-1009.

Leamer, E.E., Specification Searcher: Ad Hoc Inference with Nonexperimental Data, Wiley, New York, 1978.

Lee, Y., "Asymptotic Formulae for the Distribution of a Multivariate Test Statistic: Power Comparisons of Certain Multivariate Tests," Biometrika, 58 (1971), 647-651.

Lehmann, E.L., Testing Statistical Hypotheses, Wiley, New York, 1959.

Maeshiro, A., "Autoregressive Transformation, Trended Independent Variables and Autocorrelated Disturbance Terms," Review of Economics and Statistics, 58 (1976), 497-500.

____________, "On the Retention of the First Observations in Serial Correlation Adjustment of Regression Models," International Economic Review, 20 (1979), 259-265.

Magee, L., "Efficiency of Estimators in Regression Model with AR(1) Errors: An Analytical Study," mimeographed, University of Western Ontario, 1982.

________, Ullah, A., and Srivastava, V.K., "Efficiency of Estimators in the Regression Model with First Order Autoregressive Errors," in M.L. King and D.E.A. Giles, eds., Specification Analysis in the Linear Model: Essays in Honour of Donald Cochrane, Routledge and Kegan Paul, London, forthcoming, 1984.

Mann, H.B., and Wald, A., "On Stochastic Limit and Order Relationships," Annals of Mathematical Statistics, 14 (1943), 217-226.

Morrison, D.E., and Henkel, R.E., The Significance Test Controversy: A Reader, Aldine, Chicago, 1970.

Morrison, D.F., Multivariate Statistical Methods, second edition, McGraw-Hill, New York, 1976.

Muirhead, R.J., Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.

Nagar, A.L., "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," Econometrica, 27 (1959), 575-595.

Neyman, J., and E.S. Pearson, "Contributions to the Theory of Testing Statistical Hypotheses, Part I," Statistical Research Memoirs, 1 (1936), 1-37.

________, and ________, "Contributions to the Theory of Testing Statistical Hypotheses, Parts II and III," Statistical Research Memoirs, 2 (1938), 25-57.

Park, R.E., and Mitchell, B.M., "Estimating the Autocorrelated Error Model with Trended Data," Journal of Econometrics, 13 (1980), 185-201.

Peers, H.W., "Likelihood Ratio and Associated Test Criteria," Biometrika, 58 (1971), 577-587.

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

________, "Comparison of Local Power of Alternative Tests of Non-Nested Regression Models," Econometrica, 50, (1982a), 1287-1306.

________, "A Critique of the Proposed Tests of the Natural Rate-Rational Expectations Hypothesis," Economic Journal, 92 (1982b), 529-554.

Phillips, P.C.B., "Approsimations to Some Finite Sample Distributions Associated with a First Order Stochastic Difference Equation," Econometrica, 45 (1977), 463-485.

________, "ERA's: A New Approach to Small Sample Theory," Econometrica, 51 (1983), 1505-1526.

Prais, S.J., and Winsten, C.B., "Trend Estimators and Serial Correlation," unpublished Cowles Dommission Discussion Paper, Chicago, 1954.

Raj, B., Srivastava, V.K., and Upadhyaya, S., "The Efficiency of Estimating a Random Coefficient Model," Journal of Econometrics, (1980), 285-299.

Rao, C.R., "Large Sample Tests Concerning Several Parameters with Applications to Problems of Estimation," Proceedings of the Cambridge Philosophical Society, 44 (1948), 50-57.

Rao, P., "Specification Bias in Seemingly Unrelated Regressions," in W. Selekaerts, ed., Econometrics and Econometric Theory: Essays in Honour of Jan Tinbergen, Macmillan, London, 1974.

Rogers, G.S., Matrix Derivatives, M. Dekker, New York, 1980.

Rothenberg, T.J., "Edgeworth Expansions f[illegible] Multivariate Test Statistics," Working Paper in [illegible]omic Theory and Econometrics IP-255, Center for Research in Management Science, Institute for Business and Economic Research, University of California, Berkeley, 1977.

Roy, S.N., "On a Heuristic Method of Test Construction and its Use in Multivariate Analysis," Annals of Mathematical Statistics, 24 (1953), 220-238.

Salmon, M., "Error Correction Mechanisms," Economic Journal, 92 (1982), 615-629.

Sargan, J.D., "The Validity of Nagar's Expansion for the Moments of Econometric Estimators," Econometrica, 42, (1974), 169-176.

__________, "Gram-Charlier Approximations Applied to t Ratios of k-Class Estimators," Econometrica, 43 (1975), 327-346.

__________, "Econometric Estimators and the Edgeworth Approximation," Econometrica, 44 (1976), 421-448.

__________, "Some Approximations to the Distribution of Econometric Criteria which are Asymptotically Distributed as Chi-Squared," Econometrica, 48 (1980), 1107-1138.

Savin, N.E., "Conflict Among Testing Procedures in a Linear Regression Model with Autoregressive Disturbances," Econometrica, 44 (1976), 1303-1316.

__________, "The Bonferroni and the Scheffé Multiple Comparison Procedures," Review of Economic Studies, 47 (1980), 255-274.

__________, "Multiple Hypothesis Testing," Technical Report No. 370, The Economics Series, Stanford University, 1982.

Silvey, S.D., "The Lagrange Multiplier Test," Annals of Mathematical Statistics, 30 (1959), 389-407.

__________, Statistical Inference, Chapman and Hall, London, 1970.

Siotani, M., "An Asymptotic Expansion of the Non-Null Distribution of Hotelling's Generalized T_0^2-Statistic," Annals of Mathematical Statistics, 42 (1971), 560-571.

Spitzer, J.J., "Small-Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors," Journal of the American Statistical Association, 74 (1979), 41-47.

Springer, M.D., The Algebra of Random Variables, Wiley, New York, 1979.

Srivastava, M.S., and Khatri, C.G., An Introduction to Multivariate Statistics, North-Holland, New York, 1979.

Srivastava, V.K., and Tiwari, R., "Evaluation of Expectation of Products of Stochastic Matrices," Scandinavian Journal of Statistics, 3 (1976), 135-138.

Srivastava, V.K., and Upadhyaya, S., "Large-Sample Approximations in Seemingly Unrelated Regression Equations," Annals of the Institute of Statistical Mathematics, 30 (1978), 89-96.

Stroud, T.W.F., "On Obtaining Large-Sample Tests from Asymptotically Normal Estimators," The Annals of Mathematical Statistics, 42 (1971), 1412-1424.

Tanaka, K., "Asymptotic Expansions Associated with the AR(1) Model with Unknown Mean," Econometrica, 51 (1983), 1221-1232.

Taylor, W.E., "On the Efficiency of the Cochrane-Orcutt Estimator," Journal of Econometrics, 17 (1981), 67-82.

Telser, L.G., "Iterative Estimation of a Set of Linear Regression Equations," Journal of the American Statistical Association, 59 (1964), 845-862.

Theil, H., Principles of Econometrics, Wiley, New York, 1971.

__________, and Nagar, A.L., "Testing the Independence of Regression Disturbances," Journal of the American Statistical Association, 56 (1961), 793-806.

Tillman, J.A., "The Power of the Durbin-Watson Test," Econometrica, 43 (1975), 959-974.

Ullah, A., and Zinde Walsh, V., "On the Robustness of LM, LR, and W Tests in Regression Models," Econometrica, forthcoming (1984).

Vandaele, W., "Wald, Likelihood Ratio, and Lagrange Multiplier Tests as an F Test," Economics Letters, 8 (1981), 361-365.

Watson, G.S., "Linear Least Squares Regression," Annals of Mathematical Statistics, 38 (1967), 1679-99.

Wald, A., "Tests of Hypotheses Concerning Several Parameters When the Number of Observations is Large," Transactions of the American Mathematical Society, 54 (1943), 426-482.

Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association, 57 (1962), 348-368.

__________, "Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms," Journal of the American Statistical Association, 71 (1976), 400-405.

END

27.03.85

FIN