# ASYMPTOTIC EXPANSION APPROXIMATIONS AND THE DISTRIBUTIONS OF VARIOUS TEST STATISTICS IN DYNAMIC ECONOMETRIC MODELS

Gordon Christopher Russell Kemp

Thesis submitted to the University of Warwick in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

The research reported in this thesis was conducted in the Department of Economics at the University of Warwick.

July, 1987

TABLE OF CONTENTS

| | pages |
|---|---|
| LIST OF TABLES AND FIGURES | (vi) |
| LIST OF ABBREVIATIONS | (vii) |
| ACKNOWLEDGEMENTS | (viii) |
| DECLARATION | (ix) |
| SUMMARY | (x) |
| CHAPTER 1 Introduction | |
| 1.1 Hypothesis Testing in Econometrics | 1 |
| 1.2 The Wald Testing Principle | 5 |
| CHAPTER 2 A Review of Exact and Approximate Finite Sample Theory | |
| 2.1 Introduction | 10 |
| 2.2 Exact Finite Sample Theory | 11 |
| 2.2.1 Exact Finite Sample Theory in the SEM | 12 |
| 2.2.2 Exact Finite Sample Theory in Dynamic Single Equation Models | 16 |
| 2.3 First Order Asymptotic Theory | 18 |
| 2.4 Edgeworth Expansions | 23 |
| 2.4.1 Obtaining Edgeworth Expansions | 25 |
| 2.4.2 The Validity of Edgeworth Expansions | 29 |
| 2.4.3 The Uses and Performance of Edgeworth Expansion Approximations | 31 |
| 2.5 Alternative Approximation Methods | 36 |
| 2.5.1 Asymptotic Expansions in Parameters other than the Sample Size | 36 |
| 2.5.2 Nagar Moments | 37 |
| 2.5.3 Saddlepoint Approximations | 39 |
| 2.5.4 Large Deviation Expansions | 43 |

2.5.5 Curve Fitting Techniques and Transformation Methods 44

2.5.6 Monte Carlo Methods 46

2.A Appendix to Chapter 2 48

2.A.1 Orders of Magnitude Notation 48

2.A.2 Asymptotic Expansions 49

2.A.3 Transformation from the Edgeworth-B to the Edgeworth-A Form 51

CHAPTER 3 An Approach for Obtaining Asymptotic Expansions for the CDF's of Asymptotically Chi-Square Econometric Test Criteria

3.1 Introduction 53

3.2 The Exact Inner Product Case under the Null Hypothesis 54

3.2.1 Multivariate Edgeworth Expansions 54

3.2.2 The Joint Characteristic Function of the Squares 65

3.2.3 The Characteristic Function of the Sum of Squares 70

3.2.4 The Density and Distribution Functions of the Sum of Squares 72

3.3 The Approximate Inner Product Case 74

3.3.1 The Expansion of the Statistic 74

3.3.2 The Characteristic Function of the Statistic 77

3.3.3 The Distribution Function of the Statistic 83

3.4 The Joint Distribution Case 94

3.4.1 The Jointly Independent Case 94

3.4.2 The Jointly Dependent Case 104

3.5 CDF Expansions under Local Alternative Hypothesis Sequences 106

3.A Appendix to Chapter 3 116

3.A.1 Relationship with the Approach in Mauleon-Torres (1983) 116

3.A.2 The Covariance Matrix Assumption 117

CHAPTER 4 Alternative Test Statistics and the Use of the Wald Testing Principle with Reformulations of Parameter Restrictions

4.1 Introduction 123

4.2 Comparing Alternative Test Statistics 124

4.3 Wald Statistics for Reformulations of a Set of Parameter Restrictions 135

4.4 Testing for a Non-Linear Restriction in a Simple Model using the Wald Principle 142

4.5 Asymptotic Expansion Approximations to the Distributions of the Alternative Wald Statistics 152

4.6 Conclusions 158

4.A Appendix to Chapter 4 160

CHAPTER 5 Predictive Failure Tests in the AR(1) Model

5.1 Introduction 163

5.2 The Use of Predictive Failure Tests 164

5.2.1 Parameter Constancy and Predictive Failure Tests 164

5.2.2 Predictive Failure Tests in the AR(1) Model 169

5.3 Properties of Forecast Errors and Predictive Failure Tests 171

5.3.1 Properties Relevant to One-Step Ahead Tests 172

LIST OF TABLES AND FIGURES


| | | pages |
|---|---|---|
| Figure 3.1 | Conditional Distribution of a Statistic with a Minimum Not Equal to Zero | 87 |
| Figure 4.1 | Alternative Formulations of the Restriction Function | 138 |
| Table 4.1 | Test Statistic Rejection Probabilities at the 5% Critical Value for a $\chi^2(1)$ Distribution | 155 |
| Table 4.2 | Test Statistic Rejection Probabilities at the 5% Critical Values for F(1,T-3) Distributions | 156 |
| Table 4.3 | Complete Approximate Rejection Probability Results Expressed as %'s | 157 |
| Table 5.1 | Approximate Rejection Probabilities Expressed as %'s for $F^h_{1\|n}$ and $F^c_{1\|n}$ when n = 10 | 189 |
| Table 5.2 | Asymptotic Expansion Coefficients for the Joint CDF's of $(F^h_{1\|n}, F^{h+}_{2\|n})$ and of $(F^c_{1\|n}, F^{c+}_{2\|n})$ | 192 |
| Table 5.3 | Asymptotic Expansion Coefficients for the CDF's of $F^h_{2\|n}$ and $F^c_{2\|n}$ | 193 |
| Table 5.4 | Approximate Rejection Probabilities Expressed as %'s for $F^h_{2\|n}$ and $F^c_{2\|n}$ when n = 10 | 196 |
| Table 6.1 | First and Second Order Sample Autocorrelation Coefficients for the 3 Exogenous Data Sets | 223 |
| Table 6.2 | Asymptotic Expansion Approximation Coefficients Multiplied by the Sample Size | 224 |
| Table 6.3 | Qualtitative Behaviour of Expansion Coefficients for Different Data Sets | 226 |
| Table 6.4 | Evaluated Approximate Rejection Probabilities expressed as %'s for $W_1$ | 227-228 |
| Table 6.5 | Evaluated Approximate Rejection Probabilities expressed as %'s for $W_{(2)}$ | 230-231 |
| Table 6.6 | Evaluated Approximate Rejection Probabilities expressed as %'s for $W_2$ | 233-234 |
| Table 6.7 | Evaluated Approximate Rejection Probabilities expressed as %'s for Sequential Tests | 235-238 |

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AD | Autoregressive Distributed lag |
| AR(1) | First-Order Autoregression |
| ARMA | Autoregressive Moving Average |
| cdf | cumulative distribution function |
| cf | characteristic function |
| cgf | cumulant generating function |
| CLR | Classical Linear Regression |
| COMFAC | Common Factor |
| DLR | Double Length Regression |
| DSEM | Dynamic Simultaneous Equations Model |
| ESSACS | Exact Sum of Squares Asymptotically Chi-Square Statistics |
| FIML | Full Information Maximum Likelihood |
| iid | idependently identically distributed |
| IV | Instrumental Variables |
| LIML | Limited Information Maximum Likelihood |
| LM | Lagrange Multiplier |
| LR | Likelihood Ratio |
| ML | Maximum Likelihood |
| MLE | Maximum Likelihood Estimator |
| MSE | Mean Square Error |
| OLS | Ordinary Least Squares |
| OPG | Outer Product of the Gradient |
| pdf | probability density function |
| plim | probability limit |
| SEM | Simultaneous Equations Model |
| 3SLS | Three Stage Least Squares |
| 2SLS | Two Stage Least Squares |
| W | Wald |

## ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my thesis supervisor, Professor Kenneth Wallis, for his guidance, criticism and consistent encouragement.

Many others in the Department of Economics at the University of Warwick have contributed in some way to the environment in which I have worked; in particular I would like to thank Professor Marcus Miller and Dr Mark Stewart.

My thanks also go to Elizabeth Blundell, Cindy Cudmore, Penny Martin, Maureen Smith and Pamela Spring for having typed the manuscript with skill and care and great patience.

DECLARATION

The contents of this thesis are my own work, except where otherwise indicated.

Gordon C. R. Kemp

## SUMMARY

In this thesis we examine the derivation of asymptotic expansion approximations to the cumulative distribution functions of asymptotically chi-square test statistics under the null hypothesis being tested and the use of such approximations in the investigation of the properties of testing procedures. We are particularly concerned with how the structure of various test statistics may simplify the derivation of asymptotic expansion approximations to their cumulative distribution functions and also how these approximations can be used in conjunction with other small sample techniques to investigate the properties of testing procedures. In Chapter 1 we briefly review the construction of test statistics based on the Wald testing principle and in Chapter 2 we review the various approaches to finite sample theory which have been adopted in econometrics including asymptotic expansion methods. In Chapter 3 we derive asymptotic expansion approximations to the joint cumulative distribution functions of asymptotically chi-square test statistics making explicit use of certain aspects of the structure of such test statistics. In Chapters 4, 5 and 6 we apply these asymptotic expansion approximations under the null hypothesis, in conjunction with other small sample techniques, to a number of specific testing problems. The test statistics considered in Chapters 4 and 6 are Wald test statistics and those considered in Chapter 5 are predictive failure test statistics. The asymptotic expansion approximations to the cumulative distribution functions of the test statistics under the null hypothesis are evaluated numerically; the implementation of the algorithm for obtaining asymptotic expansion approximations to the cumulative distribution functions of test statistics is discussed in an Appendix on Computing Finally, in Chapter 7 we draw overall conclusions from the earlier chapters of the thesis and discuss briefly directions for possible future research.

# CHAPTER 1: Introduction

## 1.1 Hypothesis Testing in Econometrics

The modelling of relationships between economic variables in applied econometric studies presents a number of problems. These include the choice of model specification in the light of the data, the assessment of the validity of theories about the nature of the economy, and the forecasting of the future behaviour of economic variables of interest. In addressing these problems, applied econometric studies often draw heavily upon the machinery of statistical hypothesis testing. Thus tests are used to determine the adequacy of a given model and of possible simplifications of it, the validity of actual hypotheses of interest in economics, and the accuracy of forecasts based upon the resulting estimated model.

Most research into the statistical methodology of econometric testing procedures has been focussed on the construction of tests for given hypotheses in given situations. Three general principles have been formulated for the construction of tests: the Wald principle which is discussed in more detail in Section 1.2, the technical summary at the end of this chapter; the Likelihood Ratio principle; and the Lagrange Multiplier principle. These provide bases for the construction of tests in a wide range of circumstances.

The use of tests based on these principles has been justified primarily on the basis of their large sample asymptotic properties, i.e. from the limiting (or asymptotic) distributions of the test statistics involved as the number of available observations tends to infinity. This emphasis on the asymptotic properties of tests has arisen because of the relative difficulty of deriving and computing the exact finite sample distributions of the test statistics involved and the relative ease of deriving and computing the asymptotic distributions of these test statistics as approximations to their exact finite sample distributions. The choice in a given situation between alternative tests with identical asymptotic properties has then been based in general on comparisons of their ease of computation in that situation together with knowledge about specific tests from past experience and Monte Carlo studies; e.g. Summers (1965) on the behaviour of various simultaneous equations estimators.

In this thesis we are concerned with extending the analysis of the properties of testing procedures beyond the derivation and computation of the asymptotic distributions of the test statistics involved. The main approach we adopt is that of using asymptotic expansion approximations to the cumulative distribution functions (cdf's) of the test statistics involved. The approximation error from this type of approximation to the cdf tends to zero more rapidly than that from the asymptotic cdf approximation as the number of observations increases. Therefore this type of approximation can be regarded as a refinement of the asymptotic cdf. Such approximations to the cdf's of estimators and test statistics have been developed for some cases in econometrics; see Sargan (1976) and Rothenberg (1984b) on the derivation and use of such approximations in econometrics .

In Chapter 2 we review the available finite sample results in econometrics together with various approaches to approximate finite sample theory including the use of asymptotic expansions. In Chapter 3 we develop a new algorithm for obtaining asymptotic expansion approximations to the cdf's of asymptotically chi-square test statistics under the null hypotheses being tested. This method utilises certain aspects of the structure of test statistics based on the Wald, Likelihood Ratio and Lagrange Multiplier testing principles. We then extend this algorithm so as to obtain such approximations under local hypothesis sequences. This enables us to obtain more refined approximations to the properties of testing procedures than are available simply from the asymptotic cdf approximation. In the Appendix on Computing we briefly discuss a computer program, ESSACS, which we have written to implement the algorithm of Chapter 3 for the null hypothesis case and its use in the evaluation studies of Chapters 4 to 6.

In Chapters 4 to 6 we consider the application of asymptotic expansion approximations to the cdf's of test statistics in a number of cases. The example we consider in Chapter 4 concerns the choice between alternative tests of a given hypothesis. Firstly we review the background to the problem of choosing a test, and then we examine in more detail a specific case involving choosing between Wald tests based on various formulations of a non-linear restriction in a simple Classical Linear Regression (CLR) model. We show that the main features of the qualitative behaviour of the tests across the parameter space, as found by previous Monte Carlo studies, can be explained by their limiting behaviour for various parameter sequences with the sample size fixed. We also conduct a numerical evaluation study using the ESSACS program which broadly supports the earlier conclusions about the behaviour of the tests across the parameter space.

In Chapter 5 we consider the use of predictive failure tests. After reviewing the background to predictive failure tests, we examine in more detail the behaviour of forecast errors and predictive failure tests in the first-order autoregression (AR(1)) model. We show that there is a term in the covariance matrix of the forecast errors which in the CLR model is equal to zero but in the AR(1) model is generally non-zero. This suggests that the conventional predictive failure tests which use an estimate of the covariance matrix of the forecast errors based on the CLR formula are inappropriate in the AR(1) model and in more general dynamic models. We also conduct a numerical evaluation study using the ESSACS program. The results of this study indicate that the conventional predictive failure tests still perform reasonably well in spite of the earlier argument.

In Chapter 6 we consider the use of sequential testing procedures. Firstly, we review the background to sequential testing procedures and in particular the use of Wald tests in such procedures. Then we examine in more detail a specific sequential testing procedure for a sequence of two nested hypotheses in the autoregressive-distributed lag model with one lag on the endogenous variable and on a single exogenous variable (AD(1, 1) model). In the Appendix to Chapter 6 we prove a theorem on the construction of invariant Wald tests. We apply this theorem to show that the Wald tests we use for the testing sequence being considered in the AD(1,1) model have certain invariance properties. Then we perform a numerical evaluation study using the ESSACS program to investigate the properties of the sequential testing procedure under the null hypothesis. The results of this study indicate that the dynamic structure of the exogenous data is very important and, in particular, that the more powerful are the dynamics of the exogenous data the stronger is the interaction between the two tests in the sequence.

Finally, in Chapter 7 we draw overall conclusions from the earlier chapters about the use and usefulness of asymptotic expansion approximations to the cdf's of test statistics in the analysis of the properties of testing procedures and discuss possible directions for future research in this area.

## 1.2 The Wald Testing Principle

In econometrics one of the most commonly used bases for constructing testing procedures is the Wald testing principle; Wald (1943). Suppose that Y is a random variable (possibly a vector) which is distributed with a density function $f(y;\theta)$ where the parameter $\theta$ is a k-vector belonging to $\Theta$ . If we wish to test the null hypothesis $H_0$: $\phi(\theta) = 0$, where $\phi(.)$ is a p-vector, then providing that $\phi(.)$ is appropriately differentiable and that the family of density functions $f(y;.)$ meets appropriate regularity conditions (see Wilks (1962)) we can construct the following test statistic for $H_o$ :

$$W = T\phi(\hat{\theta})'[\hat{F}\hat{\Omega}\hat{F}']^{-1}\phi(\hat{\theta})$$

where: T is an indexing parameter;

$\hat{\theta}$ is a consistent asymptotically normal estimator of $\theta$ such that $T^{1/2}[\hat{\theta}-\theta]\xrightarrow{D}N[0,\Omega]$ as $T \to \infty$ ; $\hat{F} = [\partial\phi/\partial\theta']$ at $\hat{\theta}$ ; and

$\hat{\Omega}$ is a consistent estimator of $\Omega$ which is symmetric positive definite.

This is the Wald test statistic for $H_o : \phi(\theta) = 0$ using $\hat{\theta}$ and $\hat{\Omega}$, and under the null hypothesis it is asymptotically distributed as a central chi-square variate with p degrees of freedom provided that $\hat{F}$ is of full rank in a neighbourhood of the true parameter value $\theta$. Under the local alternative hypothesis sequence $H_A$: $\theta_T = \theta_o + T^{-1/2}\psi$ where $\phi(\theta_o) = 0$ then W is asymptotically distributed as a non-central chi-square variate with p degrees of freedom and non-centrality parameter $\mu^2$:

$$\mu^2 = \psi'F_o'[F_o\Omega F_o']^{-1}F\psi \quad ,$$

where $F_o$ is $\hat{F}$ evaluated at $\theta_o$.

The underlying approach of the Wald statistic is to measure the extent to which the estimated parameter values fail to meet the hypothesised restrictions. The simplest choice of a measure would be $\phi(\theta)$ which is only equal to zero if the estimates meet the restrictions. However, $\phi(\theta)$ is generally a vector and can be positive, negative or zero. A measure of distance should be a non-negative scalar. The easiest way to obtain a non-negative scalar from $\phi(.)$ is to construct a quadratic form in $\phi(.)$ using a positive definite matrix. The matrix used here is an estimate of the asymptotic covariance matrix of $T^{1/2}\phi(\theta)$. As a consequence, the Wald statistic has the property that it can be decomposed as the inner product of a vector which is asymptotically distributed as a standard normal vector under the null hypothesis. Thus we define:

$$T^{1/2}\eta = T^{1/2}[\hat{F}\hat{\Omega}\hat{F}']^{-1/2}\phi(\hat{\theta}) \quad ,$$

so that $W = T\eta'\eta$ where $[\hat{F}\hat{\Omega}\hat{F}']^{-1/2}$ is some appropriate square root of $[\hat{F}\hat{\Omega}\hat{F}']^{-1}$ such as the lower triangular Cholesky decomposition. Under the local alternative hypothesis sequence $H_A$ above, $T^{1/2}\eta$ is asymptotically distributed as normal with mean $[F_o\Omega F_o']^{-1/2}F_o\psi$ and an identity covariance matrix.

The Wald test can be regarded as a generalisation of the F-test for linear regression coefficient restrictions in the Classical Linear Regression (CLR) model which takes the form:

$$y = X\beta + u \quad ,$$

where y is a T-vector of endogenous random variables, X is a (T×K) matrix of known constants, $\beta$ is a K-vector of unknown parameters, and u is distributed $N[0,\sigma^2 I]$ where $\sigma^2$ is unknown. The test statistic for the hypothesis $H_1$: $R\beta - r = 0$, where R is a (q×K) matrix of known constants and r is a q-vector of known constants, is:

$$F = \frac{(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)/q}{\hat{\sigma}^2}$$

where $\hat{\beta} = (X'X)^{-1}(X'y)$ and $\hat{\sigma}^2 = y'[\ I - X(X'X)^{-1}X'\ ]y\ /\ (T-K)$ are the Ordinary Least Squares (OLS) estimators of $\beta$ and $\sigma^2$; $\hat{\beta}$ is distributed as a jointly normal vector with mean $\beta$ and covariance matrix $\sigma^2(X'X)^{-1}$, and $(T-K)\hat{\sigma}^2/\sigma^2$ is distributed independently of $\hat{\beta}$ as a central chi-square variate with (T-K) degrees of freedom. The statistic F is distributed as a non-central F-distributed variate with (q,T-K) degrees of freedom and non-centrality parameter:

$$\delta^2 = \frac{(R\beta-r)'[R(X'X)^{-1}R']^{-1}(R\beta-r)}{\sigma^2}$$

provided that rank(R) = q. If $\delta^2$ converges to a finite limit $\mu^2$, then qF is asymptotically distributed as a non-central chi-square variate with q degrees of freedom and non-centrality parameter $\mu^2$.

The Wald testing principle is particularly appropriate for constructing tests of hypotheses within a maintained model because it only requires unrestricted estimation of the maintained model. In contrast other testing principles such as the Lagrange Multiplier or Likelihood Ratio testing principles usually require estimation of the maintained model subject to the hypothesised restrictions. Thus the Wald testing principle is in general much more convenient computationally for constructing tests of different hypotheses within the same maintained model than are the Likelihood Ratio or Lagrange Multiplier testing principles.

In contrast, the Wald testing principle is not very convenient for constructing diagnostic tests of a given maintained model, i.e. tests of the given maintained model against various more general models. This is because it usually requires that each of the more general models considered is separately estimated so that there is typically a heavy computational burden for each diagnostic test constructed using the Wald principle. There is a wide variety of diagnostic tests based on various principles such as the Lagrange Multiplier principle, the analysis of residuals and the analysis of forecast errors. The Lagrange Multiplier principle is particularly convenient for constructing diagnostic tests since it only requires estimation of the maintained model.

# CHAPTER 2: A Review of Exact and Approximate Finite Sample Theory

## 2.1 Introduction

Many approaches have been adopted to obtain the exact and approximate cumulative distribution functions (cdf's) of econometric estimators and test statistics. This chapter outlines the main approaches that have been used, the circumstances in which they have been used and results which have been obtained with them. Particular attention is paid in this chapter to Edgeworth (and related) expansions of cdf's since this is the main approximation method which is used in the subsequent chapters of this thesis. In Section 2.2 we review the exact finite sample results which are available for econometric statistics and the models for which these results are applicable. In Section 2.3 we discuss the first-order (or basic) asymptotic properties of econometric statistics and the problems which arise from using these asymptotic results to provide approximations to the finite sample properties of such econometric statistics. Then in Section 2.4 we examine Edgeworth expansions in more detail and outline the main methods which have been used to obtain these expansions in econometrics. Lastly, in Section 2.5 we briefly discuss a number of other approximation methods including saddlepoint approximations and Monte Carlo techniques which have also been used in econometrics.

2.2 Exact Finite Sample Theory

There are three major areas in econometrics for which an extensive finite sample theory has been developed. Firstly, there is the Classical Linear Regression (CLR) model which is very widely used in statistics. Secondly, there is the static Simultaneous Equations Model (SEM) which has been the main focus of exact finite sample theory in econometrics. Finally, there are a number of very simple dynamic models where the statistics of interest can be expressed as ratios of quadratic forms in normal variates and here the cdf's of the statistics of interest can often be found by numerical inversion of the characteristic function although some exact theoretical results ara also available.

In all three areas the models considered are linear in both variables and parameters, and the disturbances are normally distributed. Johnson and Kotz (1969, 1970a, 1970b, 1972) provide an extensive survey of the distributions typically encountered in statistics and the exact and approximate properties of the variates with these distributions. Anderson (1958) and Mardia, Kent and Bibby (1979) discuss the distributions involved in multivariate analysis. These are of interest in econometrics since they form the main basis of the exact finite sample theory developed in econometrics.

The properties of the standard statistics used in the CLR model, e.g. Ordinary Least Squares (OLS) and Maximum Likelihood (ML) estimators and t and F tests, are very well-known and an excellent discussion of them can be found in Rao (1973). The CLR model is of interest in econometrics because it is both widely used and also forms the main framework for the analysis of single equation models with dynamics and non-linearities and for the limited information approach to the analysis of individual equations from SEM's or Dynamic Simultaneous Equations Models (DSEM's). However, the introduction of simultaneity or dynamics or non-linearities does serve to complicate the analysis of such models and the properties of the statistics used very considerably. Even simple modifications of the CLR model lead to highly complex distribution theory; see Phillips (1984b) on the exact distribution of the James-Stein rule estimator in the CLR.

2.2.1 Exact Finite Sample Theory in the SEM

The theory of estimation in the SEM is one of the main contributions that econometrics has made to statistical analysis and has been the subject of considerable research in econometrics. Consequently, a great deal of attention has been given to the problem of obtaining exact finite sample results for the various estimators used in the SEM. In the early 1960's some results were obtained for the exact distributions of structural form estimators but these were for very small and highly specific models; e.g. see Bergstrom (1962) on the distribution of the estimated Marginal Propensity to Consume in a very simple, two-equation Keynesian model. Later in the 1960's these sorts of results were extended to more general models with arbitrary numbers of exogenous variables but still with specific numbers of endogenous variables. Thus Richardson (1968) and Sawa (1969) obtained the exact distribution of the Two Stage Least Squares (2SLS) estimator of the coefficient on the right hand side endogenous variable in a two-equation model and Mariano and Sawa (1972) obtained that of the Limited Information Maximum

Likelihood (LIML) estimator. Basmann (1974) provides a useful survey of the results available in the early 1970's. Throughout the 1970's and continuing into the 1980's these results have been further extended to yet more general models with arbitrary numbers of endogenous and exogenous variables; e.g. see Phillips (1980a, 1983b) on the exact distributions of Instrumental Variable (IV) estimators and Phillips (1985b) on the exact distributions of LIML estimators.

However, attention has not only been given to the limited information approach to the estimation of structural coefficients in the SEM. McDonald (1972) obtained the exact distribution of the 2SLS estimator of the disturbance variance in a structural equation and Basmann and Richardson (1973) obtained that of the corresponding LIML estimator. These variance estimators are usually expressed as quadratic forms in the estimated structural coefficients (with the quadratic form matrices being stochastic) and are closely related to the test statistics which have been constructed for the over-identifying restrictions usually imposed in econometrics for estimation. The exact distributions of such test statistics have been examined by Basmann (1965) for a particular case which involved 2SLS estimates and by Rhodes (1981) for the general case involving LIML estimates.

The exact distributions of test statistics for linear restrictions on the structural coefficients of a single equation from an SEM have also been investigated. The simplest case is when the hypothesis completely specifies the vector of coefficients on the right hand side endogenous variables. The restrictions can then be reformulated as a set of linear restrictions on the coefficients of a reduced form equation for an appropriately redefined endogenous variable. Thus they can be tested straightforwardly using F-tests as in the CLR. This case was analysed by Anderson and Rubin (1949). However, most hypotheses in the SEM cannot be expressed in this form and in particular this framework does not allow for the testing of the significance of the coefficient on a single right hand side endogenous variable when there is more than one such variable. This problem was considered by Richardson and Rohr (1971) who obtained the exact distribution of a structural t-statistic. The model they used only had one right hand side endogenous variable which permitted the properties of the statistic to be compared with those of the corresponding Anderson and Rubin statistic.

The techniques used for obtaining exact finite sample results in the SEM generally follow a number of steps. Firstly, the model and the statistic of interest (which may be a vector) are standardised in certain ways so as to simplify the covariance structure of the disturbances and of the exogenous variables. The advantage of doing this when it is possible is that it reduces the number of parameters and functions of the exogenous variables which need to be considered during the remainder of the derivation. Essentially these standardisation transformations reduce the dimensionality of the problem and clarify the important elements of the problem. The standardised statistic is then usually expressed as a function of a number of underlying variates which typically have some known and well-behaved distribution such as the non-central Wishart distribution. These standardisation transformations are discussed in more detail by Mariano (1982) and by Phillips (1983a).

The exact distribution of the standardised statistic is then obtained by an appropriate one-to-one transformation of the underlying variates into a vector which consists of the standardised statistic and a set of other elements which are then integrated out. It is the last two steps of choosing an appropriate one-to-one transformation and then integrating out the remainder that constitute the hardest part of the process. Very few exact results have been obtained for the cdf's of full-information and sub-system estimates and test statistics because the functional forms of the statistics are generally more complicated than their limited information single equation counterparts. This creates greater mathematical difficulties in integrating out the remainder terms using probability density functions (pdf's)

The distributional results obtained in exact finite sample theory for the SEM almost always have extremely complicated functional forms. This arises because the mathematical methods which have proved successful for the extraction of exact cdf's in the SEM usually involve the manipulation of infinite series functions of matrices such as zonal polynomials. These functions are very complex in form and sometimes do not have a general closed functional form and they are often extremely slow to converge when computed numerically. Therefore they usually provide very little aid for understanding the consequences for the properties of the statistics of interest of changing parameter values and using different exogenous variables.

There are also some results available on the existence of moments and on the values of moments which do exist for estimators in the SEM. If the means and variances of different estimators exist then these can be used to compare these estimators on the basis of their location and dispersion properties. If such moments fail to exist, as is the case with Full Information Maximum Likelihood (FIML) estimators, see Sargan (1970), then this indicates that the distributions of such estimators are relatively thick-tailed and that as a result the probability of finding extreme outliers is higher for such estimators than for estimators with finite means and variances. This has prompted the creation of new estimators which will have these moments. Thus Fuller (1977) suggested a modified LIML estimator which is obtained from the standard LIML estimator but will always have mean and variance even if the LIML estimator does not. Greenberg and Webster (1983), Mariano (1982) and Phillips (1983a) all carry out extensive surveys of the results which are available on the moments and cdf's of structural form estimators in the SEM. The articles by Mariano and by Phillips also survey the properties of a number of reduced form estimators, test criteria and other statistics of interest.

### 2.2.2 Exact Finite Sample Theory in Dynamic Single Equation Models

The main type of dynamic single equation model for which an exact finite sample theory has been developed is the first-order autoregression model. The earliest exact result obtained in such a model appears to be the distribution of Fisher's g-statistic under the null hypothesis of independently identically distributed (iid) normal random variables; see Fisher (1929) for the derivation. Later, von Neumann (1941) obtained the distribution of the von-Neumann ratio as a test for first-order autocorrelation. Anderson (1942) obtained the distribution of the sample first-order circular autocorrelation coefficient in a circular autoregression. Anderson's method was extended by Watson

(1956) to obtain the distribution of several sample circular autocorrelations and by Hannan (1970) to obtain the distribution of the Durbin-Watson test statistic for first-order serial correlation under the null hypothesis of iid normal disturbances.

These results have all been obtained for the circular autoregression model which has certain symmetries involving the first and last elements which the non-circular autoregression model lacks. As a consequence very few analytic results have been obtained for the non-circular autoregression model. However, provided that the disturbances are normally distributed and that the statistic of interest is a quadratic form or ratio of quadratic forms then it is still possible to evaluate the exact cdf of the statistic numerically to an arbitrary degree of precision by using the Gil-Pelaez formula from Gil-Pelaez (1951) which forms the basis for Imhof's routine from Imhof (1961). This technique involves the numerical inversion of the characteristic function of a quadratic form in normal variables and has been used to obtain the exact distribution of the OLS estimate of the autoregressive coefficient in a first-order autoregression AR(1) model by Phillips (1977a). Similar numerical methods also based on inverting the characteristic function or a closely related function have been used by Sawa (1978) to obtain the moments of sample autocorrelation coefficients in the AR(1) model and by De Gooijer (1980) to obtain them in the first-order autoregressive moving-average (ARMA) model.

## 2.3 First Order Asymptotic Theory

Asymptotic results in econometrics are concerned with the limiting behaviour of probability statements involving the statistics of interest as some parameter value tends either to a constant or to infinity. The parameter chosen is often the sample size with the limiting behaviour being as this tends to infinity but in the SEM other parameters are sometimes chosen such as the concentration parameter with the limiting behaviour usually being as this tends to infinity. It is however possible to consider limiting behaviour as any parameter tends to some limit and thus sometimes the limiting behaviour is as the disturbance variance tends to zero in a particular way; see Section 2.5.1.

The main forms of stochastic convergence considered in econometrics, i.e. the limiting behaviour of the probability statements involving the statistics of interest, are convergence in probability and convergence in distribution. If a statistic converges in probability to some limit, which may be a random variable although it is often a constant, then the probability that the statistic lies more than an arbitrary distance away from the limit can be made arbitrarily small by requiring that the value of the limiting parameter lies sufficiently close to but not at its limit. In the case of a parameter such as the sample size tending to infinity then this is interpreted as requiring that the value of the limiting parameter be sufficiently large but not infinite. Convergence in probability is usually associated with estimators and when an estimator converges in probability to the parameter value being estimated then the estimator is referred to as being (weakly) consistent.

A statistic converges in distribution if its cdf converges (at all continuity points) to the cdf of some random variable. Estimators frequently converge in distribution to normal cdf's provided that they are appropriately standardised to have non-degenerate cdf's; if this is so then they are usually referred to as asymptotically normal statistics. Provided that the same standardisation has been applied to a set of different but still asymptotically normal estimators of the same parameter, then comparisons can be made of the estimators on the basis of the means and variances of their asymptotic cdf's even though in finite samples the estimators may have no means or variances as is the case with FIML estimates. Test statistics frequently have limiting central chi-square cdf's under the null hypothesis and thus asymptotic critical values can be constructed for them which will give asymptotically correct significance levels. Under fixed alternatives, test statistics often have explosive or occasionally degenerate limiting cdf's so that their power tends to unity or occasionally to zero. These are not very useful properties for making comparisons of different tests. However, for local alternative sequences where the parameter values for which power is evaluated tend at an appropriate rate towards the null hypothesis then the cdf's of the test statistics being considered often converge in distribution to the cdf's of non-central chi-square variates. The tests can then be compared on the basis of their differing power properties under such local alternative sequences. These stochastic convergence results of convergence in probability and distribution are very well-known, and can be found in, for example, White (1984), and we will refer to them subsequently as first-order or basic asymptotic results.

There are a number of useful relationships between these two types of stochastic convergence. Firstly, if a statistic converges in probability then it also converges in distribution to the cdf of its probability limit (plim) whether this is a random variable or a constant. Secondly, if the difference between two statistics converges in probability to zero then the two statistics are referred to as asymptotically equivalent and they have the same asymptotic cdf. Classifying statistics into asymptotic equivalence classes is very useful in making asymptotic comparisons of the properties of the different statistics. For example, as the sample size tends to infinity the 2SLS and LIML estimators in the SEM are in general asymptotically equivalent with asymptotically normal cdf's. Similarly, the Three Stage Least Squares (3SLS) and FIML estimators are also in general asymptotically equivalent with asymptotically normal cdf's. In contrast to this the 2SLS and 3SLS estimates are not in general asymptotically equivalent; the asymptotic cdf of the 3SLS estimator is generally less dispersed than that of the 2SLS estimator in a correctly specified SEM. This property is usually referred to by saying that the 3SLS estimator is asymptotically more efficient than the 2SLS estimator. However, first-order asymptotic theory provides no basis for choosing between asymptotically equivalent statistics.

The main advantage that first-order asymptotic theory has over exact finite sample theory is the ease of derivation and implementation of the latter as compared to the former. In first-order large sample asymptotics the estimators used typically behave in the limit like normal vectors and the test statistics like quadratic forms in finite dimensional normal vectors. This behaviour is the same as the exact finite sample behaviour of OLS estimates in the CLR model with known variance. This similarity between first-order asymptotics and exact CLR theory together with the ease of derivation of the first-order large sample asymptotics has led to them being the main justification of the majority of statistical techniques currently used in econometrics.

Some authors have attempted to construct a theory of statistical inference based solely on asymptotic results. For example, Pfanzagl (1982) has argued that statistical theories have two functions. The first is to analyse the performance of statistical procedures and the second is to provide methods for the construction of optimal procedures. Although non-asymptotic theory can perform these functions in certain cases its success is erratic rather than systemmatic and it cannot be regarded as being in general satisfactory. Pfanzagl also argues that asymptotic theory can treat a wider class of models than has been used to date because existing models have been devised so as to be tractable mathematically using non-asymptotic theory or so as to be simple extensions of models which are tractable mathematically. This viewpoint seems very ambitious and we would qualify this by arguing that whilst asymptotic theory does provide a framework for the analysis and construction of statistical inference procedures in a very wide class of models, it is still desirable to examine how well such procedures perform in particular models of interest.

Unfortunately, considerable evidence has accumulated that suggests that first-order large sample asymptotic results can provide highly inaccurate approximations to actual finite sample behaviour for quite plausible parameter values even with fairly large sample sizes. This evidence comes from two sources. Firstly, in some cases it has been possible to compute the exact finite sample cdf's and moments of certain estimators and test statistics (using the techniques discussed earlier). Evans and Savin (1982a) examined a finite sample relationship between the Wald, Lagrange Multiplier and Likelihood Ratio test statistics for certain models. They found that using the asymptotic equivalence of these statistics and their limiting chi-square cdf as an approximation to the cdf's of each of them can lead to considerable probabilities of conflict between the statistics and inaccurate size and power computations. Anderson and Sawa (1973) evaluated the exact cdf of the 2SLS estimator in a simple SEM and found that for low values of the concentration parameter the asymptotic distribution can be quite inaccurate. Phillips (1977a) computed the exact cdf of the ordinary least squares (OLS) estimator of the autoregressive parameter in a first-order non-circular autoregression and found that for values of the autoregressive parameter close to the unit circle the asymptotic cdf of the OLS estimator was a poor approximation to its actual cdf for moderate sample sizes.

Secondly, some evidence is also available from Monte Carlo studies. Orcutt and Winokur (1969) obtained results for the distribution of the OLS autoregression coefficient which were later confirmed by the exact results of Phillips (1977a). Mizon and Hendry (1980) found that in certain more general dynamic models test statistics such as Sargan's COMFAC test can have finite sample power properties which differ substantially from their asymptotic power properties. Finally, Summers (1965) examined the sampling distributions of LIML, 2SLS, OLS and FIML estimators in a two equation over-identified model. Summers found that if the predetermined variables in the structural equations were highly correlated then in terms of mean square error OLS could perform better than LIML and 2SLS better than FIML. It should be noted that when some estimators lack any integral moments (as is true for FIML in this case) then the use of mean square error comparisons becomes problematical.

## 2.4 Edgeworth Expansions

The underlying problem with first-order large sample asymptotic approximations is that they are only justified as the sample size gets large. Thus the first-order asymptotic approximation to the cdf of a statistic can be completely valid and yet can be completely useless for sample sizes similar to those normally encountered. Obtaining bounds on asymptotic approximation errors is still extremely difficult even for very simple statistics. The best known result in this area is the Berry-Esseen theorem which provides a bound on the error involved in approximating the cdf of a standardised mean of iid variates using a normal cdf. The bound is in terms of the variance and third absolute moment of the variates, the sample size and a universal constant; see Berry (1941) and Esseen (1945).

One approach which has been adopted to attempt to overcome the problems associated with first-order asymptotic approximations is the use of approximations which have errors of a smaller order of magnitude in the limiting parameter. The most frequently encountered of such approximations are asymptotic expansion approximations which are specifically designed to have this property. These are obtained by truncating the asymptotic expansion of the cdf of a statistic $Q_T$:

$$\Pr\{Q_T \le z\} \sim \sum_{j=0}^{\infty} T^{-j/2} a_j(z) \, , \tag{1}$$

(for large sample asymptotics where $T \to \infty$) after K+1 terms giving:

$$\Pr\{Q_T \le z\} = \sum_{j=0}^{K} T^{-j/2} a_j(z) + o(T^{-K/2}) \, . \tag{2}$$

(see Equations (21), (22) in Section 2.A). Then $a_0(z)$ is the limiting cdf of $Q_T$ which has an error of $O(T^{-\frac{1}{2}})$ or $o(T^0)$. The $\{T^{-j/2} a_j(z); j \ge 1\}$ terms are the higher-order terms in $T^{-\frac{1}{2}}$ and including them gives a higher-order asymptotic approximation with an error of $o(T^{-K/2})$. For sufficiently large sample sizes the inclusion of the higher-order terms will reduce the approximation error in absolute value. The notation used here concerning asymptotic expansions is defined in the Appendix to this chapter.

Such asymptotic expansions as Equation ( 1 ) are often called Edgeworth expansions though the terminology is sometimes restricted to cases where $Q_T$ has a limiting normal cdf and the $\{a_j; j \ge 1\}$ can be expressed as polynomials in z multiplied by a normal pdf following Edgeworth (1905). Wallace (1958) and Johnson and Kotz (1970a) who discuss the derivation and use of Edgeworth expansions in statistics. The first published paper using Edgeworth expansions in econometrics appears to have been Sargan and Mikhail (1971) but since then there have been many papers in econometrics using Edgeworth expansions. Phillips (1980b), Rothenberg (1984b) and Taylor (1983) all provide useful surveys on the derivation and use of Edgeworth expansions in econometrics.

### 2.4.1 Obtaining Edgeworth Expansions

There are two main methods used to obtain Edgeworth expansions in econometrics. The first is the calculation of probabilities method used by Anderson (1974), Sargan (1975), Sargan (1980b) and Sargan and Mikhail (1971). The second is the characteristic function method which was originally used by Edgeworth (1905) and has been used more recently by Sargan (1976).

The calculation of probabilities method at its simplest rests on transforming a probability statement about the statistic into one about a variate with a known distribution which is the same as the asymptotic distribution of the statistic of interest. Thus if $Q_T = Q_T(x)$ where x is distributed as standard normal, then:

$$\Pr\{Q_T \leq r\} = \Pr\{Q_T(x) \leq r\} = \Phi(Q_T^{-1}(r)),$$

(where $\Phi(\cdot)$ is the standard normal cdf) provided that $Q_T^{-1}(\cdot)$ exists. However, the existence of $Q_T^{-1}(\cdot)$ is not necessary since if:

$$Q_T = T^{\frac{1}{2}}H(T^{-\frac{1}{2}}x) \text{ with } H(0) = 0,\ H'(0) = 1\ ,$$

then all that is required is that $H(\cdot)$ has a unique inverse function near $x = 0$; see Wallace (1958). A more general technique can be used when the stochastic expansion of the statistic is:

$$Q_T = x + T^{-\frac{1}{2}}A(x,y) + T^{-1}B(x,y) + o_p(T^{-1})\,, \qquad (3)$$

where x is exactly standard normal and y is a vector which is distributed independent of x. Then following Rothenberg (1984b):

$$\begin{aligned}\Pr(Q_T \le r) &= E_y[\Pr(x + T^{-\frac{1}{2}}A(x,y) + T^{-1}B(x,y) \le r|y)] + o(T^{-1}) \\ &= E_y[\Phi(r - T^{-\frac{1}{2}}A(r,y) - T^{-1}B(r,y) + T^{-1}A(r,y)A'(r,y))] \\ &\quad + o(T^{-1}), \qquad (4)\end{aligned}$$

where $A'(r,y)$ is the derivative of $A(x,y)$ with respect to x at x=r. Then expanding $\Phi(\cdot)$ in a Taylor series, taking expectations and rearranging terms:

$$\begin{aligned}\Pr(Q_T \le r) &= \Phi[r - T^{-\frac{1}{2}}E_yA(r,y) - T^{-1}E_yB(r,y) + E_yA(r,y)A'(r,y)T^{-1} \\ &\quad - \tfrac{1}{2}T^{-1}\text{Var}_yA(r,y)] + o(T^{-1}). \qquad (5)\end{aligned}$$

This method can be extended to asymptotically non-normally distributed statistics such as asymptotically chi-square test criteria. Suppose x is distributed exactly $N_q(0,I)$, that y is independent of x and that $Q_T$ can be expanded:

$$Q_T = (x'x) + T^{-\frac{1}{2}}A(x,y) + T^{-1}B(x,y) + o_p(T^{-1}). \qquad (6)$$

Then x can be transformed to polar co-ordinates $(\delta,\eta)$ where $x = \delta\eta$ with $\delta \ge 0$ and $\delta^2 = x'x$ so $\eta'\eta = 1$ so that the expansion in terms of $(\delta,\eta,y)$ is:

$$Q_T = \delta^2 + T^{-\frac{1}{2}}A^*(\delta,\eta,y) + T^{-1}B^*(\delta,\eta,y) + o_p(T^{-1}). \qquad (7)$$

Then $Q_T$ can be conditioned upon $\eta$ and y and the method proceeds as before except that it uses the chi-square cdf. A detailed exposition of this approach is given by Sargan (1980b) for the asymptotically central chi-square case. Rothenberg (1977, 1984a) examined the asymptotically non-central chi-square case. Cavanagh (1981) gave a general algorithm for deriving Edgeworth expansion approximations by this method.

The characteristic function method uses the one-to-one relationship between cdf's and the corresponding characteristic functions (cf's) which are their Fourier transforms; thus this method can be regarded as an extension of the orthodox proofs of the Central Limit Theorem. The cf of $Q_T$ is defined as $\psi_T(t) = E[\exp(itQ_T)]$ and the cumulant generating function (cgf) of $Q_T$ by $K_T(t) = \log_e\psi_T(t)$. Then the application of an inverse Fourier transform to the cf of $Q_T$ gives:

$$\mathrm{pdf}(Q_T) = \int_{-\infty}^{+\infty}(2\pi)^{-1}\exp\{-itQ_T + K_T(t)\}dt. \qquad (8)$$

In the case of an asymptotically standard normal statistic, $K_T(t)$ is approximated by a low order polynomial in $(it)$ with an error of $o(T^{-K/2})$ where the leading term is $\frac{1}{2}(it)^2$. The approximate cgf is obtained firstly by approximating $Q_T$ with a stochastic expansion to $o(T^{-K/2})$ in a vector x giving $Q_T^K$, and then secondly by approximating the cumulants of $Q_T^K$ to $o(T^{-K/2})$ using the coefficients of the stochastic expansion and the cumulants of x; see McCullagh (1984) and Sargan (1976) for alternative techniques. Provided the approximated cumulants behave like the cumulants of a standardised mean of iid variates (mean zero, variance unity) then the exponential of the approximated cgf (i.e. the approximate cf) can be expanded as $\exp(\frac{1}{2}(it)^2)$, which is the cf of a N(0,1) variate, multiplied by a low order polynomial in $(it)$.

Equation ( 8 ) can then be expanded in terms of the normal pdf and its derivatives using:

$$D^n\phi(z) = \frac{\partial^n}{\partial z^n}\phi(z) = \int_{-\infty}^{+\infty} (2\pi)^{-1}(-it)^n \exp(-itz + \tfrac{1}{2}(it)^2)dt, \quad (9)$$

where $\phi(z)$ is the standard normal pdf. Then using $D^n\phi(z)=(-1)^n H_n(z)\phi(z)$, where $H_n(z)$ is the n'th Hermite polynomial, Equation ( 8 ) can be expanded in terms of the normal pdf times a polynomial in z, and integrating this gives the asymptotic expansion approximation to the cdf of $Q_T$. A more detailed discussion of Hermite polynomials can be found in Abramowitz and Stegun (1965).

The characteristic function method is not limited to asymptotically standard normal statistics; it can easily be extended to general asymptotically normal statistics and to asymptotically multivariate normal statistics. Chambers (1967) gave various algorithms which can be used in this context and McCullagh (1984) methods for computing multivariate cumulants and approximating multivariate characteristic functions. In addition, the characteristic function method can be adapted to asymptotically non-normal statistics. Thus Mauleon-Torres (1983) applied it to asymptotically chi-square test criteria which can be expanded:

$$Q_T = \delta^2 + \sum_{j=1}^{K} T^{-j/2} a_j(\delta,\eta,y) + o_p(T^{-K/2}) \quad (\delta > 0), \qquad (10)$$

where $\eta'\eta = 1$, $\delta\eta = x$ and $(x,y)$ is asymptotically $N(0,I)$ and uncorrelated in finite samples. Mauleon-Torres obtained an Edgeworth expansion for the pdf of $(x,y)$:

$$f(x,y) = [\phi(x,y)][1 + \sum_{j=1}^{K} T^{-j/2} f_j(x,y)] + o(T^{-K/2}), \qquad (11)$$

where $\phi(\cdot)$ is the multivariate standard normal distribution. From this Mauleon-Torres obtained the cf of $Q_T$ given $\eta$ and $y$ and then took expectations with respect to $\eta$ and $y$. However, this approach does not seem well suited to asymptotically non-central chi-squared statistics since then $\delta$ and $\eta$ will no longer be asymptotically independent.

In general the most efficient approach to obtaining Edgeworth expansions need not be either exclusively a calculation of probabilities approach or a characteristic function approach. Sometimes when exact distributional results are available it is more efficient to derive the expansion directly from the known distribution function as in Phillips (1983b). Different approaches may have different advantages in a given situation so that one method may be computationally more efficient but another may lead to formulae which are easier to interpret. For further discussion of characteristic functions see Greenberg and Webster (1983), Lukacs (1960) and Lukacs and Laha (1964).

### 2.4.2 The Validity of Edgeworth Expansions

Our discussion of Edgeworth expansions has so far been purely in terms of formal manipulations used to obtain such expansions with no consideration being taken of the conditions which are required for the validity of the formal manipulations and of the resulting expansions. Several theorems have been proved on the validity of Edgeworth expansions under various conditions by Feller (1971), Magdalinos (1983), Mauleon-Torres (1983), Phillips (1977b), Sargan (1976, 1980). Sargan and Mikhail (1971) and Sargan and Satchell (1986). These theorems usually prove the validity of the Edgeworth expansion for statistics which can be expressed as functions of more basic statistics. There are typically two sets of conditions: the first set are on these functions and serve to determine the asymptotic distributions of the statistics of interest. The second are on the underlying statistics and ensure that these have a valid Edgeworth expansion. The two sets of conditions combined also ensure that well behaved stochastic expansions can be made of the statistics of interest.

The conditions of these theorems usually fail in two main cases. Firstly, when discrete distributions are involved the characteristic functions of the statistics are often badly behaved in the tails. Secondly, when there are ARMA error structures the statistics of interest are often functions of a vector of statistics whose number of elements increases with the sample size. As yet there are no theorems available to cover the second case although there are some for the first case; see Bhattacharya and Rao (1976) for theorems relating to lattice distributions.

In this thesis we are mainly concerned with deriving and applying Edgeworth expansion approximations. The validity of such expansion approximations is neither a necessary nor sufficient condition for such approximations to be useful in cases where the limiting cdf is deemed an unsatisfactory approximation. Thus in this thesis we will pay relatively little attention to questions of validity. In general, if the Edgeworth expansion is not valid there is an *a priori* argument to suggest that it is not likely to be very useful because it presumably fails to pick up various features of the model. However, in econometrics it does appear that a very wide range of statistics of interest have valid Edgeworth expansions under the usual assumptions about econometric models; see Sargan and Satchell (1986) for a proof that the first and second order sample moments in the DSEM meet the required conditions for underlying statistics.

### 2.4.3. The Uses and Performance of Edgeworth Expansion Approximations

Edgeworth expansion approximations can be used in a number of ways in econometrics. Firstly, they can be used to examine the properties of particular statistical procedures; secondly, they can be used to compare alternative procedures; and thirdly, they can be used to suggest new procedures. An example of the first use is provided by Laitinen (1978) and Meisner (1979) who argued that in systems of demand equations the test statistics commonly used to test symmetry and homogeneity have true sizes which exceed their nominal sizes and that this partially explains the frequent rejection of these hypotheses in such systems. As regards the second use Edgeworth expansions have now been quite extensively used to compare alternative procedures theoretically. Thus Akahira and Takeuchi (1981) found that the Maximum Likelihood estimator with a modification to ensure that it is median unbiased to order $o(T^{-1})$ is the most efficient estimator to order $o(T^{-1})$ among the class of estimators which are median unbiased to $o(T^{-1})$ for a wide range of models. Rothenberg (1984b) reviewed these results and those on the higher order comparison of test statistics. Such modified statistics also provide an example of the use of Edgeworth expansions in generating new statistical procedures.

Pfanzagl (1980) examined the possibility of constructing a general statistical theory for parametric models based on asymptotic expansions. Certainly asymptotic expansions do seem to be closely related to certain important underlying properties of statistical models such as curvature and also to statistical concepts such as conditionality and ancillarity; these relationships are examined further in Amari (1982) and Ryall (1981).

As well as there being a number of studies which have derived, validated and used Edgeworth expansions there have also been studies which have attempted to assess the performance of Edgeworth expansion approximations. Phillips (1977a) compared Edgeworth expansion approximations (with errors $o(T^{-\frac{1}{2}})$ and $o(T^{-1})$) for the OLS estimator of the autoregressive coefficient in a pure first-order autoregression with the exact distribution computed by Imhof's routine and with the asymptotic normal distribution. Phillips found that as the sample size increased the higher-order approximations performed relatively better but that their performance worsened as the absolute value of the autoregressive parameter approached unity. This effect from increasing the strength of the dynamics in the model also appeared in Tanaka (1984) who included an intercept and used Monte Carlo simulations and also in Tse (1982) who included exogenous regressors and used both the exact distribution and Monte Carlo simulations.

Sargan and Tse (1979) considered the marginal distributions of 2SLS estimators in a simple static simultaneous equations model and in a simple dynamic simultaneous equations model. The static model is essentially the dynamic model with the lagged dependent variables replaced by non-stochastic exogenous regressors with a covariance structure such that the asymptotic cdf's of the 2SLS estimators are the same in both models. They found that the normal approximation performed worse in the dynamic model than in the static model but that the Edgeworth expansion approximation performed well in both models.

One worrying feature of Edgeworth expansion approximations in general is their performance in the tails of the distributions being approximated. This is not very surprising since the Edgeworth expansion uses information from around the origin when taking the stochastic expansion of the statistic and approximating the cumulants. For points further into the tails a larger sample size will be needed typically to achieve a given accuracy. Furthermore, there are no restrictions implicit in the Edgeworth expansion to ensure that Edgeworth expansion approximations such as Equation ( 2 ) are true cdf's, so it is quite possible for the approximate cdf to take values less than zero and greater than one, and for the approximate pdf to be negative. It is not uncommon for such phenomena to occur in the tails of the expansion in actual studies, e.g. Phillips (1977a).

One way to ensure that the approximate cdf lies in the interval [0,1] is to use the Edgeworth-B form given in Equation ( 25 ) in place of the Edgeworth-A form given in Equation ( 26 ); see the Appendix to this chapter for a discussion of how to obtain the Edgeworth-B form from the Edgeworth-A form. This does not, however, prevent the approximate pdf from being negative in certain regions so that neither form is entirely satisfactory. The use of the Edgeworth-B form can be regarded as an application of the Cornish-Fisher method whereby the statistic is transformed to have the same limiting cdf but with a smaller order of magnitude error, e.g. $o(T^{-K/2})$ rather than $o(1)$; see Cornish and Fisher (1937). Since the Cornish-Fisher statistic is a proper random variable it should completely avoid the problems of cdf's outside [0,1] or pdf's which are negative. However, it may not give very much information about what the cdf of $Q_T$ looks like.

The performance of Edgeworth expansion approximations has also been investigated for a number of test statistics. Evans and Savin (1982a) used Edgeworth based size corrections to the Wald, LM and LR test criteria in the CLR and found that these account for almost all the differences between the finite sample and asymptotic distributions both under the null hypothesis and at various alternative points. Mauleon-Torres (1983) considered asymptotically chi-square tests in two models using both the Edgeworth-A and Edgeworth-B form approximations to the cdf's of the statistics and also the Edgeworth-B form for the cdf's of the positive square roots of the statistics. The first model Mauleon-Torres used was a dynamic linear regression model with a first-order lag on the endogenous variable which is similar to that used by Tse (1982). Mauleon-Torres considered $\chi^2$ and F tests for linear restrictions on the regression coefficients including the first-order lag parameter. The Edgeworth approximations performed better than the asymptotic cdf but the performance of all the approximations worsened as the disturbance variance increased, as the first-order lag parameter approached unity and as the coefficients on the exogenous variables decreased; changing these parameters in these ways strengthened the internally generated dynamics of the model. Mauleon-Torres found that the Edgeworth-B form tended to perform better than the Edgeworth-A form but explained this by noting that all the approximations tended to over-evaluate the probabilities for the acceptance region; note that in this case the Edgeworth-B approximation was necessarily smaller than the Edgeworth-A form.

The second model Mauleon-Torres (1983) used was a dynamic simultaneous equations model with two endogenous variables, two exogenous variables, and one lag. Mauleon-Torres imposed various exclusion restrictions and then tested for over-identification on the first structural equation using both Wald and pseudo-LM tests of the implied restrictions on the reduced form. The actual LM statistic requires complicated estimation under the restrictions whereas the pseudo-LM statistic avoids this. Mauleon-Torres found that the Edgeworth expansion approximations generally performed better than the limiting cdf and that their performance did not deteriorate seriously under the influence of strong dynamics or simultaneity. In all cases Mauleon-Torres used Monte Carlo simulations to estimate true probabilities and only considered distributions under the null hypothesis.

There are a number of areas where Edgeworth expansion approximations have been neither used nor assessed; in particular they have not been used for analysing sequential testing procedures. Since such procedures are widely used in econometrics, see Sargan (1980a) for sequential Wald testing, it would be desirable to know how well their limiting properties approximate their finite sample properties. The main reason why Edgeworth expansion approximations have not been applied to such procedures appears to be the lack of Edgeworth expansion derivations for jointly distributed test statistics. This is in contrast to asymptotically normal estimators where Edgeworth expansions of their joint distributions have been considered; see for example Tanaka (1984) on Maximum Likelihood estimators in autoregressive moving average models.

### 2.5 Alternative Approximation Methods

#### 2.5.1 Asymptotic Expansions in Parameters other than the Sample Size

As noted in Section 2.A.2 in the Appendix to this chapter, asymptotic expansions can be made in parameters other than the sample size. The most commonly encountered alternative parameter sequences are small-$\sigma$ asymptotics and large-$\mu$ asymptotics. In small-$\sigma$ asymptotics the disturbance variance matrix of the reduced form of an SEM is expressed as $\sigma^2\Omega$ and then $\sigma$ is allowed to tend to zero with all the other parameters fixed.

This approach was introduced by Kadane (1970, 1974) who obtained the asymptotic distribution as $\sigma \to 0$ of a number of tests of over-identifying restrictions. Kadane (1971) then developed asymptotic expansions in $\sigma$ for the distributions and moments of k-class estimators. Large-$\mu$ asymptotics are concerned with letting the concentration parameter, $\mu^2$, tend to infinity. Anderson (1974) used large-$\mu$ asymptotics to approximate the distribution of a LIML estimator in the SEM.

The results of large-$\mu$, small-$\sigma$ and large sample size asymptotics are very similar. This is particularly true with large sample size and large $\mu$ asymptotics but this result is not surprising since, for example, with 2SLS the sample size only affects the cdf through the concentration parameter; see Richardson (1968). There are some differences, however, with small-$\sigma$ asymptotics, e.g. tests of over-identifying restrictions have asymptotic F not $\chi^2$ distributions. This again is not totally surprising since this is what would be obtained with small-$\sigma$ asymptotics in the CLR. Anderson (1977) provides a summary of some of the results for alternative asymptotics in the SEM.

Small-$\sigma$ and large-$\mu$ asymptotics have sometimes been criticised on the grounds that they involve purely artificial parameter sequences whereas large sample size asymptotics involve a naturally occurring parameter sequence. However, as Taylor (1983) noted, this objection is not really relevant. With asymptotic results the important questions are whether they are useful and whether they are easy to obtain.

### 2.5.2 Nagar Moments

Information about econometric estimators is often summarised in terms of their moments. Thus the mean indicates the location of the estimator, the variance its dispersion, the third moment its skewness, and the fourth moment its kurtosis or the thickness of its tails (this requires some comparison with the variance). Nagar (1959) attempted to develop asymptotic expansion approximations for the moments of k-class estimators and such approximations have since been generally referred to in econometrics as Nagar moments.

The most common method for evaluating Nagar moments is to form a stochastic expansion of the statistic in a vector of underlying statistics with known moments and then to obtain stochastic expansions for the powers of the statistic (for moments around zero). The expectation operator with respect to the underlying vector can then be applied and the resulting asymptotic expansions truncated to the desired order of magnitude. Thus the techniques required for the derivation of Nagar moments are very similar to those required for the derivation of Edgeworth expansions.

Like Edgeworth expansions Nagar moments have been used in a number of ways in econometrics. Firstly, they have been used to compare different estimators. Taylor (1983) summarised the results which have been obtained for the bias and mean square error (MSE) of OLS, 2SLS and LIML estimates in the SEM using both small-$\sigma$ asymptotics as in Kadane (1971), and large-$\mu$ asymptotics as in Mariano (1973). Secondly, the Nagar mean and variance have been substituted into the normal cdf to attempt to obtain a better approximation than the usual limiting normal cdf. Sargan and Mikhail (1971) considered this for IV estimates and found that the resulting approximation was only better than the limiting normal for cases where the degree of over-identification was large and did not perform as well as the Edgeworth expansion approximations. This latter result is not surprising since no adjustments are being made for skewness and kurtosis in the Nagar normal distribution whereas they are in the Edgeworth expansion approximation.

There are two main problems with Nagar moments. Firstly, the statistic of interest may not have finite moments of the desired degree. This highlights the point that Nagar moments should be interpreted as approximations to the moments of approximations to the statistics of interest. Only under certain conditions discussed by Sargan (1974) can they be interpreted as actual approximations to the moments of the statistics of interest. Secondly, the moments of the distribution of a statistic are not always the most informative guides to its features of interest. Magdalinos (1983) suggested an alternative measure of concentration to that of MSE and found that asymptotic expansions of this measure were very easy to use in making comparisons between alternative estimators in the SEM. With test statistics the moments are very rarely, if ever, used to characterise the properties of the test.

### 2.5.3 Saddlepoint Approximations

Closely related to Edgeworth expansion approximations are saddlepoint approximations for pdf's and cdf's, introduced by Daniels (1954). Like Edgeworth expansion approximations these are derived from Equation (8).

The Edgeworth expansion forms an expansion of $K_T(t)$ around t=0 and then integrates the resulting expansion of $[\exp\{-itQ_T\}\exp\{K_T(t)\}]$. However, even if $Q_T$ has asymptotically mean zero expanding around t=0 may not be very informative when $Q_T$ is far from zero in finite samples. If $K_T(t)$ is an analytic function then so also is $\exp\{-itQ_T + K_T(t)\}$ and the above integral can be regarded as a contour integral of an analytic function along the real axis. Such a contour can be distorted without changing the value of the integral providing that the contour has the same end points and lies in the same simply connected domain. The saddlepoint approximation distorts the contour so that it passes through the saddlepoint of the integrand which is the complex point t* such that:

$$\frac{d}{dt}\left[\exp\{-itQ_T + K_T(t)\}\right]\Big|_{t=t^*} = 0 \ . \qquad (12)$$

The integrand is then expanded around t=t* and integrated term-by-term. The saddlepoint t* explicitly depends on $Q_T$ so that the point around which the expansion is taken is keyed to $Q_T$; if $Q_T$ is asymptotically standard normal then $t^* = iQ_T + O(T^{-\frac{1}{2}})$ so we often expand around ( $iQ_T$) instead.

The saddlepoint approximation has two main advantages compared to the Edgeworth expansion. Firstly, it is keyed much closer to the tails when $Q_T$ lies in the tails and so it may provide a more accurate approximation to the tails of the pdf of $Q_T$. Secondly, the series expansion obtained is typically in powers of $T^{-1}$ rather than $T^{-\frac{1}{2}}$. Thus the leading term is correct to order $O(T^{-1})$. Furthermore, as was pointed out by Daniels (1956), if the leading term is renormalised by an appropriate constant (which will depend on T) so that it integrates to unity then the renormalised leading term is correct to $O(T^{-3/2})$.

However, the saddlepoint approximation also has a number of disadvantages compared to the Edgeworth expansion. Firstly, since $K_T(t)$ has to be expanded at many different points it would appear that constructing the saddlepoint approximation requires quite detailed knowledge of the cumulant generating function. This disadvantage is not as serious as it might appear since if $K_T(t)$ is analytic everywhere in the real plane, then an appropriate expansion of $K_T(t)$ to $O(T^{-K/2})$ can be constructed at any point $\tilde{t}$ using only the cumulants; see Durbin (1979) for details. Secondly, and more seriously, if $Q_T$ does not have finite moments of all orders then $K_T(t)$ is not analytic. In this case further arguments have to be put forward to support the distortion of the contour and the approximation of the integrand by its expansion around a point not on the real axis.

The saddlepoint approximation has also been obtained using the conjugate or exponential family approach by Barndorff-Nielsen and Cox (1979). Durbin (1980) gave a related method for estimators which are sufficient statistics together with a proof of the validity of this method under certain conditions. The Durbin approach is particularly simple and is based on the following argument. If t(y) is a sufficient estimator of $\theta$ where y is distributed $f(y;\theta)$ then:

$$f(y;\theta_0) = g(t;\theta_0)h(y), \tag{13}$$

where h(y) does not depend on $\theta$, and $\theta_0$ are assumed to be the true parameter values. Substituting in f(y;t) gives:

$$g(t;\theta_0) = \frac{f(y;\theta_0)}{f(y;t)}g(t,t) \qquad (14)$$

and we then substitute $\tilde{g}(t;t)$ for $g(t;t)$ to obtain $\tilde{g}(t;\theta_0)$ where $\tilde{g}(t;t)$ is the limiting pdf of the estimator t at the parameter values t. As can be seen from an Edgeworth expansion $\tilde{g}(t;t)$ is correct to $O(T^{-1})$ so the intuitive argument is that $\tilde{g}(t;\theta_0)$, constructed by multiplying the likelihood ratio by the limiting pdf, will also be correct to $O(T^{-1})$.

The theorem proved in Durbin (1980) has two parts and provides certain conditions under which the above argument is valid. The first part of the proof is a proof of the validity of an Edgeworth expansion for a vector of statistics under certain conditions including the existence of cumulants up to a certain order. These conditions are similar but not the same as the conditions on the vector of underlying statistics required for the theorem in Sargan and Satchell (1986). Thus Durbin's Edgeworth theorem may have applications elsewhere in econometrics. The second part of the proof is a validation of the saddlepoint approximation for the statistics of interest which are functions of the underlying statistics. This second stage requires the statistics of interest to be sufficient statistics and to have second moments. Since these latter conditions are frequently not met in econometrics this limits the usefulness of this saddlepoint approximation in econometrics.

Saddlepoint approximations have been used in econometrics by Holly and Phillips (1979) for k-class estimators in the SEM and by Phillips (1978) in the non-circular AR(1) model. The results in Holly and Phillips (1979) indicate that the saddlepoint approximation can be better than the Edgeworth expansion approximation but that when the concentration parameter is small then they both perform quite badly. The saddlepoint approximation for the density function does have the advantage that it is positive everywhere whereas this cannot be guaranteed for the Edgeworth expansion approximation.

### 2.5.4 Large Deviation Expansions

Both the Edgeworth expansion approximation and the saddlepoint approximation are concerned with approximating the cdf at fixed points as the sample size gets large. However, these approximations,and in particular the Edgeworth expansion, tend to be unsatisfactory in the tails of the distribution being approximated. An alternative approach referred to as large deviation expansion approximation permits the evaluation point to increase as the sample size increases. This approach has not yet been extensively investigated in econometrics. The main results refer to standardized sums of iid variates; see Petrov (1968), although Phillips (1977c) did consider more general statistics. Field and Hampel (1982) developed an interesting integration of Edgeworth, saddlepoint and large deviation methods based on the logarithmic derivative of the density which provides accurate approximations even for very small sample sizes in the models they consider.

### 2.5.5 Curve Fitting Techniques and Transformation Methods

Curve fitting techniques are the most general methods for approximating distributions in econometrics and can often use information obtained from other approximation techniques. Curve fitting consists of picking a family of functions with roughly the right properties and then finding the member of the family which fits best in some sense. The simplest approach is to pick a family of distributions and fit the low order moments of the statistic. Henshaw (1966) approximated the distribution of the Durbin-Watson statistic using a linear transformed Beta distribution with the same first four moments. The Beta distribution is an obvious choice because it is unimodal and is restricted in a specified range, and indeed the Henshaw approximation appears to work very well.

Phillips (1982) proposed a more sophisticated technique using Padé approximations. Many of the exact distributions which have been derived in econometrics consist of a leading term multiplied by an infinite series expansion where for certain parameter values the series expansion is unity. Phillips suggested retaining the leading term, or using a simpler function with similar properties, and replacing the series expansion with a finite rational function. The overall approximation can be fitted so that at various points it, and its low order derivatives, take the same values as the function of interest; Phillips suggested using the origin and the point at infinity. The approximation can then be manipulated by slight perturbation to be strictly positive everywhere. As Phillips (1982) indicated this technique can produce very accurate and easily computable approximations when other approaches such as Edgeworth expansion and saddlepoint approximations perform very badly.

Closely related to curve fitting techniques are transformation methods. These transform the statistic so that it fits some known distribution better without losing information about the statistic. Thus if x has a t distribution with K degrees of freedom the simplest approximation is to treat x as a standard normal variate. This is unsatisfactory for $K \leq 30$ but $z = x(1-2K^{-1})^{\frac{1}{2}}$, suggested by Weir (1960), is more accurate although it too deteriorates for $K \leq 20$ or when x is very large. There are a very wide range of such transformations available both for variates of known distributions, such as the t distribution above, and also for statistics of unknown distribution; e.g. the z transformation from Fisher (1915) for the sample correlation coefficient from a bivariate normal distribution.

Curve fitting techniques and transformation methods are very versatile and can give rise to extremely accurate approximations. However, this versatility can be a disadvantage in that without considerable prior information it is not clear what family of curves or transformations should be used. In econometrics such approximation methods seem likely to be of more use currently in the SEM where considerable information is available concerning the properties of estimators and test statistics than in dynamic models where much less information is typically available. This does not prevent such approximations from being of use since often the theoretical and other information is very difficult to use directly due to computational difficulties and the existence of simple and accurate approximations is clearly desirable.

### 2.5.6 Monte Carlo Methods

The remaining major technique used in econometrics is that of Monte Carlo simulation. The principle behind the Monte Carlo approach is that of generating random or pseudo-random numbers and then using these as the disturbances in a model of known structure to analyse the properties of the statistics of interest. Typically properties of the statistics will be investigated at several points in the parameter space to reduce the specificity of the results. The results of Monte Carlo studies are subject to sampling error and so the number of replications, i.e. the number of simulations at a given point in parameter space, has to be determined carefully to reduce sampling errors to acceptable levels. Various techniques for reducing sampling error without needing to increase the number of replications are available; these include the use of antithetic variates and control variates. The information produced by Monte Carlo simulations is often not very useful without further manipulation because it consists mainly of parameter values and estimated cdf's, pdf's, moments and rejection probabilities. One method for summarising such information is that of response surfaces in which the Monte Carlo estimated values are regressed on simple functions of the parameters where these functions are chosen to be reasonably compatible with the known theoretical and empirical properties of the statistics of interest; see Mizon and Hendry (1980) for further discussion.

Hendry (1984) provided a useful survey of Monte Carlo methods and their uses in econometrics, and Challen and Hagar (1983) provided a useful survey of Monte Carlo studies which have been made in econometrics.

Such studies often provide the only benchmark against which other approximation methods can be compared. Even when the exact distributions of the statistics of interest are available, it is sometimes easier to use Monte Carlo studies; see Anderson, Kunitomo and Sawa (1982) on the LIML estimator. Monte Carlo studies are also often used on their own to assess and compare econometric statistics as in Kiviet (1985) on model selection test procedures.

## 2.A Appendix to Chapter 2

### 2.A.1 Orders of Magnitude Notation

Consider two sequences of real numbers $\{g_T\}$ and $\{h_T\}$ as $T \to \infty$ where $h_T > 0$ for all T. Then:

(i) $g_T$ is defined as being of smaller order of magnitude than $h_T$ or $g_T = o[h_T]$ if and only if for all $\varepsilon > 0$ there exists $T(\varepsilon)$ such that:

$$|(g_T/h_T)| < \varepsilon \quad \text{for all } T \geq T(\varepsilon)$$

$$\text{or } \lim_{T\to\infty}(g_T/h_T) = 0 . \tag{15}$$

(ii) $g_T$ is defined as being at most of order of magnitude $h_T$ or $g_T = O[h_T]$ if and only if there exist M and T* such that:

$$|g_T| \leq Mh_T \qquad \text{for all } T \geq T^*. \tag{16}$$

Now suppose that $\{g_T\}$ is a stochastic sequence of real numbers but $\{h_T\}$ is still a sequence of fixed positive real numbers. Then:

(iii) $g_T$ is defined as being of smaller order of magnitude in probability than $h_T$ or $g_T = o_p[h_T]$ if and only if for all $\varepsilon, \delta > 0$ there exist $T(\varepsilon,\delta)$ such that:

$$\Pr\{|g_T/h_T| > \delta\} < \varepsilon \quad \text{for all } T \geq T(\varepsilon,\delta)$$

$$\text{or } \lim_{T\to\infty} \Pr\{|g_T/h_T| > \delta\} = 0 \text{ for all } \delta > 0$$

$$\text{(i.e. } \operatorname*{plim}_{T\to\infty}|g_T/h_T| = 0). \tag{17}$$

(iv) $g_T$ is defined as being at most of order of magnitude in probability $h_T$ if and only if for all $\varepsilon > 0$ there exist $M(\varepsilon)$ and $T(\varepsilon)$ such that:

$$\Pr\{|g_T/h_T| > M(\varepsilon)\} < \varepsilon \quad \text{for all } T \geq T(\varepsilon). \qquad (18)$$

If $g_T = \alpha + o_p[h_T]$ then if $g_T$ is an estimator of $\alpha$, $g_T$ is called a consistent estimator of $\alpha$. If $g_T = O_p(1)$ then $g_T$ is referred to as stochastically bounded. It should be noted that:

$$g_T = o[h_T] \rightarrow g_T = O[h_T] \text{ and}$$

$$g_T = o_p[h_T] \rightarrow g_T = O_p[h_T]$$

These definitions follow those of Greenberg and Webster (1983).

2.A.2 Asymptotic Expansions

Suppose that $\{g_n(x)\}$ is a sequence of functions of x where $g_{n+1}(x) = o[g_n(x)]$ as $x \rightarrow \infty$. Then if $f(x)$ is a function of x and:

$$f(x) - \sum_{j=0}^{K} a_j g_j(x) = o[g_K(x)], \; K = 0,1,2,\ldots, \qquad (19)$$

then $f(x)$ has the asymptotic expansion $\sum_{j=0}^{\infty} a_j g_j(x)$ in the asymptotic sequence $\{g_n(x)\}$ and this is usually written:

$$f(x) \sim \sum_{j=0}^{\infty} a_j g_j(x) \,. \qquad (20)$$

It can be noted that Equation ( 19 ) is equivalent to:

$$f(x) - \sum_{j=0}^{K} a_j g_j(x) = O[g_{K+1}(x)] \,.$$

In econometrics and statistics we are often interested in asymptotic expansions of cdf's, pdf's, cf's etc. In terms of the above definitions such asymptotic expansions are really entire families of asymptotic expansions since we usually have two variables of interest: the evaluation point; and the limiting variable (e.g. the sample size). These asymptotic expansions take the form:

$$f_T(x) \sim \sum_{j=0}^{\infty} a_j(z) g_j(T). \tag{21}$$

With T=the sample size then $g_j(T)$ is usually chosen to be $T^{-j/2}$ so:

$$f_T(x) \sim \sum_{j=0}^{\infty} T^{-j/2} a_j(z). \tag{22}$$

The sample size is not the only limiting parameter used in econometrics: sometimes small-$\sigma$ or large-$\mu$ asymptotics are considered. Small-$\sigma$ asymptotics refer to behaviour as $\sigma^2 \to 0$ where $\sigma^2$ is the disturbance variance, see Kadane (1971); large-$\mu$ asymptotics refer to $\mu^2 \to 0$ where $\mu^2$ is the concentration parameter in the SEM, see Mariano (1973). Orders of magnitude and asymptotic expansion notation can easily be extended to small-$\sigma$ (or large-$\mu$) asymptotics by considering the behaviour of $\sigma^2$-sequences as $\sigma^2 \to 0$ (or $\mu^2$-sequences as $\mu^2 \to \infty$).

Stochastic expansions of statistics are very similar to asymptotic expansions except that, firstly, they are usually made in a vector s of variates with some known distribution properties, and, secondly, the error on the truncated expansion is usually stochastic so:

$$Q_T = \sum_{j=0}^{K} T^{-j/2} f_j(s) + o_p(T^{-K/2}), \tag{23}$$

is often encountered. By suitably manipulating s it is sometimes possible to express f(s) in a very simple form. For example, if $Q_T$ is asymptotically standard normal it is often possible to transform s to (x,y) where x is exactly standard normal and y is a vector independent of x such that:

$$Q_T = x + \sum_{j=1}^{K} T^{-j/2} g_j(x,y) + o_p(T^{-K/2}). \qquad (24)$$

These definitions follow De Bruijn (1981) and Wallace (1958).

2.A.3 Transformation from the Edgeworth-B to the Edgeworth-A Form

For simplicity we will consider expansions to $O(T^{-3/2})$: these exhibit the main phenomena to be noted without being too complicated and are often those used in applications. The Edgeworth-B form approximation is:

$$\Pr\{Q_T \leq x\} = \Phi[x + T^{-\frac{1}{2}} b_1(x) + T^{-1} b_2(x)] + O(T^{-3/2}) \quad (25)$$

where $b_1(\cdot)$ and $b_2(\cdot)$ are usually low order polynomials and $\Phi(\cdot)$ is the cdf of a standard normal. Following Sargan (1976) the technique is to form an expansion of $\Phi(x^*)$ around x where $x^*-x = T^{-\frac{1}{2}}[b_1(x)+T^{-\frac{1}{2}}b_2(x)]$ $= O(T^{-\frac{1}{2}})$. Thus:

$$\Pr\{Q_T < x\} = \Phi(x) + (x^*-x)\Phi'(x)+(x^*-x)^2(\tfrac{1}{2})\Phi''(x)+O(T^{-3/2})$$

where $\Phi'(x)$ is the first derivative of the cdf $\Phi(\cdot)$ evaluated at x, i.e. the pdf $\phi(\cdot)$ at x, and $\Phi''(x)$ is the second derivative of the cdf $\Phi(\cdot)$ evaluated at x . Substituting in for $(x^* - x)$ gives:

$$\begin{aligned}
\Pr\{Q_T \le x\} &= \Phi(x) + T^{-\frac{1}{2}}b_1(x)\phi(x) + T^{-1}b_2(x)\phi(x) \\
&\quad -\tfrac{1}{2}T^{-1}(b_1(x))^2 x\phi(x) + O(T^{-3/2}) \\
&= \Phi(x) + T^{-\frac{1}{2}}b_1(x)\phi(x) \\
&\quad + \tfrac{1}{2}T^{-1}[2b_2(x) - xb_1(x)b_1(x)]\phi(x) + O(T^{-3/2}) \\
&= \Phi(x) + T^{-\frac{1}{2}}a_1(x)\phi(x) + T^{-1}a_2(x)\phi(x) + O(T^{-3/2}) \qquad (26)
\end{aligned}$$

which is the Edgeworth-A form. It is similarly possible to proceed from the Edgeworth-A form to the Edgeworth-B form. Mauleon-Torres (1983) examines the appropriate derivation for when $Q_T$ is asymptotically chi-square rather than standard normal.

# CHAPTER 3: An Approach for Obtaining Asymptotic Expansions for the CDF's of Asymptotically Chi-Square Econometric Test Criteria

## 3.1 Introduction

This chapter presents a theoretical method for deriving asymptotic expansion approximations to the cumulative distribution functions (cdf's) of asymptotically chi-square test criteria. The method is developed initially in Section 3.2 under the null hypothesis. It is then extended: firstly, in Section 3.3, for approximate inner product criteria; secondly, in Section 3.4, for multiple test criteria; and lastly, in Section 3.5, for test criteria under local alternative hypothesis sequences.

The exact inner product class of asymptotically chi-square test criteria is of interest in itself since it covers many commonly used econometric test statistics. This is because econometricians have predominantly thought of testing parameter restrictions by measuring distances in parameter space or in parameter-related spaces using inner product type metrics. The Wald test is a very direct example of this, as also is the Hausman test. The Lagrange Multiplier test also falls into this category: here the space related to the parameters is that of the first derivatives of the log-likelihood function. Exact inner product tests are often easy to compute because most (often all) of the information is generated as a side product from standard regression packages.

Not all econometric tests fall into this class: the most obvious counter-example is the Likelihood Ratio test. Except in special cases, when the log-likelihood is an exact quadratic form, this test criterion does not belong to the exact inner product class. Furthermore, the

Likelihood Ratio is computationally much harder to implement than a large sample equivalent, such as the Wald test, in most situations. However, in econometric applications the Likelihood Ratio is usually an approximate inner product statistic and so its distribution can be examined using the method presented in this chapter.

## 3.2 The Exact Inner Product Case under the Null Hypothesis

The class of test criteria covered here take the fomr:

$$s^2 = \sum_{j=1}^{h} \lambda_j^2$$

$$= \lambda_j \lambda_j \text{ in tensor summation notation}$$

$$\text{where } \lambda \xrightarrow{D} N(0, I_h),$$

that is where the vector $\lambda$ has a limiting jointly independent, standard normal distribution. The first step in the method used here is to decompose the statistic into this form thereby explicitly obtaining $\lambda$ as a vector of statistics. The second step is to obtain an Edgeworth expansion for the joint probability density function (pdf) of $\lambda$. This is discussed in Chambers (1967), McCullagh (1984) and Sargan and Satchell (1986). The method which is used here is based on McCullagh (1984) and clearly shows the relationship of the multivariate Edgewo-th expansion to the univariate Edgeworth expansion in Sargan (1976), (corrected formulae in Tse (1981)).

### 3.2.1 Multivariate Edgeworth Expansions

We assume that the decomposition vector $\lambda$ can be expressed:

$$\lambda = T^{\frac{1}{2}}(\phi(p) - \phi(\mu))$$

where $\mu = E(p)$, and

where the vector p has cumulants of all orders up to K+2 where $K \geq 2$. This condition is clearly met if p contains only quadratic and linear forms in jointly normally distributed variates. We also assume that, defining:

$$x = T^{\frac{1}{2}}(p-\mu)$$

then the first cumulant of x is zero (which follows by construction), and the cumulants of order j (j=2,... K+2) are $O(T^{-(j-2)/2})$. Equation (1) can be re-written so that $\lambda$ is a function of x:

$$\lambda = g(x)$$

and we then require some conditions of the derivatives of $g' = (g^1,\ldots,g^h)$:

$$g^r_j = \left.\frac{\partial g^r}{\partial x_j}\right|_0 = O(1)\ , \quad r=1,\ldots,h;$$

$$g^r_{jk} = \left.\frac{\partial^2 g^r}{\partial x_j \partial x_k}\right|_0 = O(T^{-\frac{1}{2}})\ , \quad r=1,\ldots,h;$$

$$g^r_{jk\ell} = \left.\frac{\partial^3 g^r}{\partial x_j \partial x_k \partial x_\ell}\right|_0 = O(T^{-1})\ , \quad r=1,\ldots,h;$$

and all higher order derivatives are $O(T^{-3/2})$. Since

$$g^r_j = \phi^r_j = \left.\frac{\partial \phi^r}{\partial p_j}\right|_\mu \ ; \ g^r_{jk} = T^{-\frac{1}{2}}\phi^r_{jk} \ ; \ g^r_{jk\ell} = T^{-1}\phi^r_{jk\ell} \text{ etc.}$$

this is easily met providing all the derivatives of $\phi$ at $p = \mu$ converge to constants as $T\to\infty$ (and the first derivative to a non-zero constant). To prove the validity of the multivariate Edgeworth expansion requires rather more complicated conditions on $\phi(.)$ and p; see Chambers (1967), Phillips (1977b), and Sargan and Satchell (1986) for theorems and more extensive discussion.

Then $\lambda_r$ can be expanded as a polynomial in x with an error of $O(T^{-3/2})$:

$$\lambda_r = (g^r_j x_j) + \tfrac{1}{2}(g^r_{jk} x_j x_k) + \tfrac{1}{6}(g^r_{jk\ell} x_j x_k x_\ell) + O(T^{-3/2})$$
$$= (G^r_o + G^r_1 + \tfrac{1}{2}G^r_2 + \tfrac{1}{6}G^r_3 + O(T^{-3/2}))X = P^r X$$

using a generalized operator notation as in McCullagh (1984). Since there is no constant term in the expansion (at $x = 0$, $\lambda_r = 0$) then we can drop the $G_o$ term.

We now need some notation for the cumulants of x . Following McCullagh (1984, p.462) we define generalized cumulants as follows:

$$\kappa^i = E(x_i);\ \kappa^{ij} = E(x_i x_j);\ \kappa^{i,j} = \text{cov}(x_i, x_j);$$

$$\kappa^{ijk} = E(x_i x_j x_k);\ \kappa^{i,jk} = \text{cov}(x_i, x_j x_k)$$
$$= \kappa^{i,j,k} + \kappa^j \kappa^{i,k} + \kappa^k \kappa^{i,j},\ \text{etc.}$$

As can be seen these generalized cumulants include both moments and cumulants as special cases. The important information about a generalized cumulant is given by its superscripts and the way in which these are divided up by commas. The commas induce a partition on the set of x variables indicated by the superscripts. Even though two superscripts may be the same they should be treated as distinct for the purposes of expanding the generalized cumulant as a sum of products of ordinary cumulants. The terms in such an expansion meet two requirements. The first is that they are products of ordinary cumulants such that each superscript in the generalized cumulant appears exactly once in the product. The second is that no terms are included that can be factorized non-trivially so as to induce a partition on the superscripts which is a weak coarsening of the partition induced by the commas.

An example is given by $\kappa^{i,j,k\ell}$, the third order joint cumulant of $x_i$ and $x_j$ and $x_k x_\ell$ :

$$\kappa^{i,j,k\ell} = \kappa^{i,j,k,\ell} + \kappa^{i,j,k}\kappa^{\ell} + \kappa^{i,j,\ell}\kappa^{k} + \kappa^{i,k}\kappa^{j,\ell} + \kappa^{i,\ell}\kappa^{j,k}$$
$$= \kappa^{i,j,k,\ell} + \kappa^{i,j,k}\kappa^{\ell}[2] + \kappa^{i,k}\kappa^{j,\ell}[2].$$

The square brackets notation is a convenient way to save space: the number in square brackets gives the number of terms of the form of the expression immediately preceding the square brackets. This has to be understood in context so in the above:

$$\kappa^{i,j,k}\kappa^{\ell}[2] = \kappa^{i,j,k}\kappa^{\ell} + \kappa^{i,j,\ell}\kappa^{k} .$$

McCullagh (1984, Section 3, pp. 463-4) gives a more detailed mathematical exposition of this rule for expanding generalized cumulants.

Lastly, we need a rule for finding the cumulant generating function of $\lambda$ given the polynomial expansion of $\lambda$ in $x$ and the cumulants of $x$ . Denoting the entire structure of cumulants of $x$ by $\kappa$ we have:

$$K_\lambda(\xi) = \exp(\xi_r P^r)\kappa$$

where the exponent is formally expanded in terms of the operators $G_0, G_1, G_2, G_3$ etc. Thus

$$\begin{aligned} K_\lambda(\xi) = {} & \xi_r(G_1 + \tfrac{1}{2}G_2 + \tfrac{1}{6}G_3 + \dots)\kappa \\ & + \tfrac{1}{2}\xi_r\xi_s(\{G_1G_1\} + \tfrac{1}{2}\{G_1G_2 + G_2G_1\} + \tfrac{1}{4}\{G_2G_2\} + \dots)\kappa \\ & + \tfrac{1}{6}\xi_r\xi_s\xi_t(\{G_1G_1G_1\} + \tfrac{1}{2}G_1G_1G_2[3] + \tfrac{1}{4}G_1G_2G_2[3] + \dots)\kappa \\ & + \tfrac{1}{24}\xi_r\xi_s\xi_t\xi_u(G_1G_1G_1G_1 + \tfrac{1}{2}G_1G_1G_1G_2[4] + \dots)\kappa + \dots \qquad (2) \end{aligned}$$

McCullagh (1984, Equations (9), (10), p.465). The action of an operator on $\kappa$ is to be seen so (McCullagh (1984), Equations (7), (8), pp. 464-5)):

$$1\kappa = 0;\ G_0\kappa = g^r;\ G_1\kappa = g^r_i\kappa^i;\ G_2\kappa = g^r_{ij}\kappa^{ij};\ G_3\kappa = g^r_{ijk}\kappa^{ijk}.$$

The action of compound operators introduces commas so:

$$G_1G_1\kappa = g^r_i g^s_i \kappa^{i,j};\ G_2G_1\kappa = g^r_{ij} g^s_k \kappa^{ij,k};\ G_1G_2\kappa = g^r_i g^s_{jk}\kappa^{i,jk}.$$

With this we can expand Equation (2) as:

$$\begin{aligned} K_\lambda(\xi) &= \xi_r(g^r_j\kappa^j + \tfrac{1}{2}g^r_{jk}\kappa^{jk} + \ldots) \\ &\quad + \tfrac{1}{2}\xi_r\xi_s(g^r_j g^s_k\kappa^{j,k} + g^r_j g^s_{k\ell}\kappa^{j,k\ell} + \tfrac{1}{4}g^r_{jk}g^s_{\ell m}\kappa^{jk,\ell m} + \ldots) \\ &\quad + \tfrac{1}{6}\xi_r\xi_s\xi_t(g^r_j g^s_k g^t_\ell\kappa^{j,k,\ell} + (\tfrac{3}{2})g^r_j g^s_k g^t_{\ell m}\kappa^{j,k,\ell m} + \ldots) \\ &\quad + \tfrac{1}{24}\xi_r\xi_s\xi_t\xi_u(g^r_j g^s_k g^t_\ell g^u_m\kappa^{j,k,\ell,m} + \ldots) + \ldots \end{aligned} \tag{3}$$

McCullagh (1984, Equation (11), p.465). With the order of magnitude properties of the derivatives of g and the cumulants of x there are only 11 terms in Equation (3) which are not $O(T^{-3/2})$:

$$\begin{aligned} a^1_r &= \tfrac{1}{2}g^r_{jk}\kappa^{jk}; \\ a^2_{rs} &= \tfrac{1}{2}g^r_j g^s_k\kappa^{j,k}; \\ a^3_{rs} &= \tfrac{1}{2}g^r_j g^s_{k\ell}\kappa^{j,k\ell} = \tfrac{1}{2}g^r_j g^s_{k\ell}\kappa^{j,k,\ell}; \\ a^4_{rs} &= \tfrac{1}{8}g^r_{jk}g^s_{\ell m}\kappa^{jk,\ell m}; \\ a^5_{rs} &= \tfrac{1}{6}g^r_j g^s_{k\ell m}\kappa^{j,k\ell m}; \\ a^6_{rst} &= \tfrac{1}{6}g^r_j g^s_k g^t_\ell\kappa^{j,k,\ell}; \\ a^7_{rst} &= \tfrac{1}{4}g^r_j g^s_k g^t_{\ell m}\kappa^{j,k,\ell m}; \\ a^8_{rstu} &= \tfrac{1}{24}g^r_j g^s_k g^t_\ell g^u_m\kappa^{j,k,\ell,m}; \\ a^9_{rstu} &= \tfrac{1}{12}g^r_j g^s_k g^t_\ell g^u_{mp}\kappa^{j,k,\ell,mp}; \end{aligned}$$

$$a^{10}_{rstu} = \frac{1}{36} g^r_j g^s_k g^t_\ell g^u_{mpq} \kappa^{j,k,\ell,mpq};$$

$$a^{11}_{rstu} = \frac{1}{16} g^r_j g^s_k g^t_{\ell m} g^u_{pq} \kappa^{j,k,\ell m,pq}. \qquad (4)$$

These eleven terms correspond to Sargan's eleven Edgeworth coefficients, Sargan (1976), as can be seen by expanding the generalized cumulants used:

$$\kappa^{jk} = \kappa^{j,k};\ \kappa^{j,k\ell} = \kappa^{j,k,\ell};$$

$$\kappa^{jk,\ell m} = \kappa^{j,k,\ell,m} + \kappa^{j,\ell}\kappa^{k,m}\ [2];$$

$$\kappa^{j,k\ell m} = \kappa^{j,k,\ell,m} + \kappa^{j,k}\kappa^{\ell,m}\ [3];$$

$$\kappa^{j,k,\ell m} = \kappa^{j,k,\ell,m} + \kappa^{j,\ell}\kappa^{k,m}\ [2];$$

$$\kappa^{j,k,\ell,mp} = \kappa^{j,k,\ell,m,p} + \kappa^{j,k,m}\kappa^{\ell,p}\ [6];$$

$$\kappa^{j,k,\ell,mpq} = \kappa^{j,k,\ell,m,p,q} + \kappa^{j,k,\ell,m}\kappa^{p,q}\ [3] + \kappa^{j,k,m}\kappa^{\ell,p,q}\ [9] + \kappa^{j,m}\kappa^{k,p}\kappa^{\ell,q}\ [6];$$

$$\kappa^{j,k,\ell m,pq} = \kappa^{j,k,\ell,m,p,q} + \kappa^{j,k,\ell,p}\kappa^{m,q}\ [4] + \kappa^{j,k,\ell}\kappa^{m,p,q}\ [4] + \kappa^{j,\ell}\kappa^{k,p}\kappa^{m,q}\ [8];$$

using the result that $\kappa^j = 0$. Substituting these into (4) and taking account of orders of magnitude:

$$a^1_r = \tfrac{1}{2} g^r_{jk} \kappa^{j,k} = O(T^{-\frac{1}{2}})\ ;$$

$$a^2_{rs} = \tfrac{1}{2} g^r_j g^s_k \kappa^{j,k} = O(1)\ ;$$

$$a^3_{rs} = \tfrac{1}{2} g^r_j g^s_{k\ell} \kappa^{j,k,\ell} = O(T^{-1})\ ;$$

$$a^4_{rs} = \tfrac{1}{8} g^r_{jk} g^s_{\ell m} \kappa^{j,\ell}\kappa^{k,m}\ [2] + O(T^{-2}) = O(T^{-1});$$

$$a^5_{rs} = \tfrac{1}{6} g^r_j g^s_{k\ell m} \kappa^{j,k}\kappa^{\ell,m}\ [3] + O(T^{-2}) = O(T^{-1});$$

$$a^6_{rst} = \tfrac{1}{6} g^r_j g^s_k g^t_\ell \kappa^{j,k,\ell} = O(T^{-\frac{1}{2}});$$

$$a^7_{rst} = \tfrac{1}{4} g^r_j g^s_k g^t_{\ell m} \kappa^{j,\ell}\kappa^{k,m}\ [2] + O(T^{-3/2}) = O(T^{-\frac{1}{2}});$$

$$a^8_{rstu} = \frac{1}{24} g^r_j g^s_k g^t_\ell g^u_m \kappa^{j,k,\ell,m} = O(T^{-1})$$
$$a^9_{rstu} = \frac{1}{12} g^r_j g^s_k g^t_\ell g^u_{mp} \kappa^{j,k,m} \kappa^{\ell,p}\ [6] + O(T^{-2}) = O(T^{-1});$$
$$a^{10}_{rstu} = \frac{1}{36} g^r_j g^s_k g^t_\ell g^u_{mpq} \kappa^{j,m} \kappa^{k,p} \kappa^{\ell,q}\ [6] + O(T^{-2}) = O(T^{-1});$$
$$a^{11}_{rstu} = \frac{1}{16} g^r_j g^s_k g^t_{\ell m} g^u_{pq} \kappa^{j,\ell} \kappa^{k,p} \kappa^{m,q}\ [8] + O(T^{-2}) = O(T^{-1}). \qquad (5)$$

When r=s=t=u (e.g. the one degree of freedom case) then:

$$a^1 = (\tfrac{1}{2})\alpha_4 = O(T^{-\frac{1}{2}});\ a^2 = (\tfrac{1}{2})\sigma^2 = O(1);\ a^3 = (\tfrac{1}{2})\alpha_5 = O(T^{-1});$$
$$a^4 = (\tfrac{1}{2})\alpha_9 + O(T^{-2}) = O(T^{-1});\ a^5 = (\tfrac{1}{2})\alpha_7 + O(T^{-2}) = O(T^{-1});$$
$$a^6 = (\tfrac{1}{6})\alpha_1 = O(T^{-\frac{1}{2}});\ a^7 = (\tfrac{1}{2})\alpha_3 + O(T^{-3/2}) = O(T^{-\frac{1}{2}});$$
$$a^8 = (\tfrac{1}{24})\alpha_2 = O(T^{-1});\ a^9 = (\tfrac{1}{2})\alpha_{10} + O(T^{-2}) = O(T^{-1});$$
$$a^{10} = (\tfrac{1}{6})\alpha_6 + O(T^{-2}) = O(T^{-1});\ a^{11} = (\tfrac{1}{2})\alpha_8 + O(T^{-2}) = O(T^{-1});$$

where the $\{\alpha\}$ and $\sigma^2$ are Sargan's Edgeworth coefficients (see Sargan (1976), p.425).

We now assume that the covariance matrix assumption discussed in Appendix 3.A.2 holds so that $a^2_{rs} = \frac{1}{2}\delta_{rs} + O(T^{-1})$. Substituting the terms from (5) into (3) gives:

$$K_\lambda(\xi) = -\tfrac{1}{2}(\xi_r\xi_r) + i(b_r\xi_r) + i^2(b_{rs}\xi_r\xi_s) + i^3(b_{rst}\xi_r\xi_s\xi_t) + i^4(b_{rstu}\xi_r\xi_s\xi_t\xi_u) + O(T^{-3/2})$$

where
$$b_r = a^1_r = O(T^{-\frac{1}{2}});$$
$$b_{rs} = (a^2_{rs} - \tfrac{1}{2}\delta_{rs}) + a^3_{rs} + a^4_{rs} + a^5_{rs} = O(T^{-1});$$
$$b_{rst} = a^6_{rst} + a^7_{rst} = O(T^{-\frac{1}{2}});$$
$$b_{rstu} = a^8_{rstu} + a^9_{rstu} + a^{10}_{rstu} + a^{11}_{rstu} = O(T^{-1});\ \text{and} \qquad (6)$$

where $\delta_{rs} = \begin{cases} 1 & r = s \\ 0 & r \neq s \end{cases}$, the Kronecker delta.

From the cumulant generating function expansion we can easily obtain the characteristic function expansion by $\theta_\lambda(\xi) = \exp\{K_\lambda(\xi)\}$ so substituting in the expansion (6) for $K_\lambda(\xi)$ gives:

$$\begin{aligned}\theta_\lambda(\xi) &= \exp\{-\tfrac{1}{2}(\xi_r\xi_r) + i(b_r\xi_r) + i^2(b_{rs}\xi_r\xi_s) + i^3(b_{rst}\xi_r\xi_s\xi_t) \\ &\quad + i^4(b_{rstu}\xi_r\xi_s\xi_t\xi_u) + O(T^{-3/2})\} \\ &= [\exp(-\tfrac{1}{2}\xi_r\xi_r)].[1 + i(b_r\xi_r) + i^2(b_{rs}\xi_r\xi_s) + i^3(b_{rst}\xi_r\xi_s\xi_t) \\ &\quad + i^4(b_{rstu}\xi_r\xi_s\xi_t\xi_u) + \tfrac{1}{2}i^2(b_r b_s\xi_r\xi_s) + i^4(b_r b_{stu}\xi_r\xi_s\xi_t\xi_u) \\ &\quad + \tfrac{1}{2}i^6(b_{rst}b_{uvw}\xi_u \cdots \xi_w)] + O(T^{-3/2}).\end{aligned}$$

It is convenient to gather the terms together so that:

$$\begin{aligned}\theta_\lambda(\xi) &= [\exp(-\tfrac{1}{2}\xi_r\xi_r)][1 + i(c_r\xi_r) + i^2(c_{rs}\xi_r\xi_s) + i^3(c_{rst}\xi_r\xi_s\xi_t) \\ &\quad + i^4(c_{rstu}\xi_r\xi_s\xi_t\xi_u) + i^6(c_{rstuvw}\xi_r \cdots \xi_w)] + O(T^{-3/2})\end{aligned}$$

where:

$$\begin{aligned} c_r &= b_r = O(T^{-\frac{1}{2}}); \\ c_{rs} &= b_{rs} + \tfrac{1}{2}b_r b_s = O(T^{-1}); \\ c_{rst} &= b_{rst} = O(T^{-\frac{1}{2}}); \\ c_{rstu} &= b_{rstu} + b_r b_{stu} = O(T^{-1}); \\ c_{rstuvw} &= \tfrac{1}{2}b_{rst}b_{uvw} = O(T^{-1}). \end{aligned} \qquad (7)$$

The expansion for the joint probability density function of the $\lambda$ can now be obtained by an inverse Fourier transform, term-by-term, on the expansion for the joint characteristic function of the $\lambda$. The inverse transform is:

$$\text{p.d.f. } (\lambda\ ) = (2\pi)^{-n} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp(-i\xi_r\lambda_r)\theta_\lambda(\xi)d\xi_1 \ldots d\xi_h\ . \quad (8)$$

In the transform the integrand can be expanded as:

$$\begin{aligned}\exp(-i\xi_r\lambda_r)\theta_\lambda(\xi) = &\left[\exp(-i\xi_r\lambda_r)\right]\left[\exp(-\tfrac{1}{2}\xi_r\xi_r)\right]\left[1 + i(c_r\xi_r)\right.\\ &+ i^2(c_{rs}\xi_r\xi_s) + i^3(c_{rst}\xi_r\xi_s\xi_t) + i^4(c_{rstu}\xi_r\xi_s\xi_t\xi_u)\\ &\left.+ i^6(c_{rstuvw}\xi_r \cdots \xi_w)\right] + O(T^{-3/2}). \quad (9)\end{aligned}$$

If we now define:

$$D_r\left[\exp(-i\xi_r\lambda_r)\right] = \frac{\partial}{\partial\lambda_r}\left[\exp(-i\xi_r\lambda_r)\right] = (-i)\xi_r\left[\exp(-i\xi_r\lambda_r)\right];$$

$$D_{rs}\left[\exp(-i\xi_r\lambda_r)\right] = (-i)^2\xi_r\xi_s\left[\exp(-i\xi_r\lambda_r)\right]; \text{ etc.}$$

then we have that on re-ordering the terms of (9):

$$\begin{aligned}\exp(-i\xi_r\lambda_r)\theta_\lambda(\xi) = &\left[1 + (-1)c_rD_r + (-1)^2c_{rs}D_{rs} + (-1)^3c_{rst}D_{rst}\right.\\ &\left.+ (-1)^4c_{rstu}D_{rstu} + (-1)^6c_{rstuvw}D_{rstuvw}\right].\left[\exp(-i\xi_r\lambda_r)\right].\\ &\left[\exp(-\tfrac{1}{2}\xi_r\xi_r)\right] + O(T^{-3/2}). \quad (10)\end{aligned}$$

We now assume that the integration of (10) with respect to $\lambda$ produces an expansion for pdf($\lambda$) (8) which is still correct to $O(T^{-3/2})$. This requires conditions on the characteristic function of the $p$ as given in Sargan and Satchell (1986). Then we can interchange the operations of integration with respect to $\lambda$ and differentiation with respect to $\xi$ giving:

$$p.d.f.(\lambda) = (2\pi)^{-h} [1 + (-1)c_r D_r + (-1)^2 c_{rs} D_{rs} + (-1)^3 c_{rst} D_{rst}$$
$$+ (-1)^4 c_{rstu} D_{rstu} + (-1)^6 c_{rstuvw} D_{rstuvw}]$$
$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \exp\{-i\xi_r\lambda_r\} \exp\{-\tfrac{1}{2}\xi_r\xi_r\} d\xi_1 \dots d\xi_h + O(T^{-3/2})$$
$$= [1 + (-1)c_r D_r + (-1)^2 c_{rs} D_{rs} + (-1)^3 c_{rst} D_{rst}$$
$$+ (-1)^4 c_{rstu} D_{rstu} + (-1)^6 c_{rstuvw} D_{rstuvw}] (2\pi)^{-h/2} \exp\{-\tfrac{1}{2}\lambda_r\lambda_r\}$$
$$+ O(T^{-3/2}). \qquad (11)$$

The derivatives of the normal density function with respect to the random variable are closely related to the Hermite polynomials: see McCullagh (1984), Kendall and Stuart (1969), Olver (1974). The notation used here differs slightly from that of McCullagh (1984). If we define the joint pdf of asymptotically independent standard normal variates as $\phi(\lambda) = (2\pi)^{-h/2}\exp\{-\frac{1}{2}\lambda_r\lambda_r\}$ then differentiating this with respect to $\lambda$ gives:

$$D_r\phi(\lambda) = (-1)H_r(\lambda)\phi(\lambda);$$
$$D_{rs}\phi(\lambda) = (-1)^2 H_{rs}(\lambda)\phi(\lambda); \text{ etc.}$$

Then if we substitute these expressions for the derivatives of the jointly independent standard normal pdf into (11) we find that:

$$pdf(\lambda) = [1 + c_r H_r(\lambda) + c_{rs} H_{rs}(\lambda) + c_{rst} H_{rst}(\lambda)$$
$$+ c_{rstu} H_{rstu}(\lambda) + c_{rstuvw} H_{rstuvw}(\lambda)]\phi(\lambda) + O(T^{-3/2}) \quad (12)$$

where, following McCullagh (1984):

$$H_r(z) = z_r \; ; \; H_{rs}(z) = z_r z_s - \delta_{rs} ; \; H_{rst}(z) = z_r z_s z_t - z_r \delta_{st} \; [3];$$

$$H_{rstu}(z) = z_r z_s z_t z_u - z_r z_s \delta_{tu} \; [6] + \delta_{rs}\delta_{tu} \; [3];$$

$$H_{rstuvw}(z) = z_r \cdots z_w - z_r z_s z_t z_u \delta_{vw} \; [15]$$
$$+ z_r z_s \delta_{tu} \delta_{vw} \; [45] - \delta_{rs}\delta_{tu}\delta_{vw} \; [15] ; \qquad (13)$$

which are McCullagh's multivariate Hermite polynomials. Collecting the terms in (12) together using the formulae from (13) we find:

$$\text{pdf}(\lambda) = \left[(2\pi)^{-h/2}\exp(-(\tfrac{1}{2})\lambda_r\lambda_r)\right]\left[1 + d_o + d_r\lambda_r + d_{rs}\lambda_r\lambda_s + d_{rst}\lambda_r\lambda_s\lambda_t + d_{rstu}\lambda_r\lambda_s\lambda_t\lambda_u + d_{rstuvw}\lambda_r\lambda_s\lambda_t\lambda_u\lambda_v\lambda_w\right] + O(T^{-3/2})$$

where:
$$d_o = -c_{rs}\delta_{rs} + c_{rstu}\delta_{rs}\delta_{tu} \; [3] - c_{rstuvw}\delta_{rs}\delta_{tu}\delta_{vw} \; [15] = O(T^{-1});$$
$$d_r = c_r - c_{rst}\delta_{st} \; [3] = O(T^{-\frac{1}{2}});$$
$$d_{rs} = c_{rs} - c_{rstu}\delta_{tu} \; [6] + c_{rstuvw}\delta_{tu}\delta_{vw} \; [45] = O(T^{-1});$$
$$d_{rst} = c_{rst} = O(T^{-\frac{1}{2}});$$
$$d_{rstu} = c_{rstu} - c_{rstuvw}\delta_{vw} \; [15] = O(T^{-1});$$
$$d_{rstuvw} = c_{rstuvw} = O(T^{-1}). \qquad (14)$$

This completes the second step which was to obtain an expansion to $O(T^{-3/2})$ for the joint probability density function of $\lambda$.

### 3.2.2 The Joint Characteristic Function of the Squares

The third step in this method of expanding the density function of $s^2 = \Sigma\lambda_r^2$ is to obtain the joint characteristic function of the squares $\{\lambda_r^2\}$ to $O(T^{-3/2})$. Putting $\lambda^2 = (\lambda_1^2, \ldots, \lambda_h^2)'$ then :

$$c.f.(\lambda^2) = \Gamma_\lambda(t) = E_\lambda[\exp(i\, t_r\, \lambda_r^2)]$$

$$= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \exp(i\, t_j \lambda_j^2)\; p.d.f.(\lambda)\; d\lambda_1 \ldots d\lambda_h. \qquad (15)$$

The integrand in (15) can be expanded as:

$$\exp(i\, t_j\, \lambda_j^2) pdf(\lambda\ ) = [\exp(i\, t_j\, \lambda_j^2)][(2\pi)^{-h/2} \exp(-\tfrac{1}{2}\lambda_r\lambda_r)]$$
$$\cdot[1 + d_o + d_r\lambda_r + d_{rs}\lambda_r\lambda_s + d_{rst}\lambda_r\lambda_s\lambda_t + d_{rstu}\lambda_r \cdots \lambda_u$$
$$+ d_{rstuvw}\lambda_r \cdots \lambda_w] + O(T^{-3/2}).$$

The joint characteristic function of the $\{\lambda^2\}$, $\Gamma_\lambda(t)$, can now be expanded as a sum of integrals with error $O(T^{-3/2})$. The first integral is:

$$[1 + d_o]\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \exp(it_j\lambda_j^2)(2\pi)^{-h/2}\exp(-\tfrac{1}{2}\lambda_j\lambda_j)d\lambda_1 \ldots d\lambda_h$$
$$= (1 + d_o) \prod_{j=1}^{h}(1-2it_j)^{-\frac{1}{2}}.$$

The remaining integrals take the general form:

$$(d) \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \exp(it_j\lambda_j)f(\lambda)\; \lambda_r^{n_r}\lambda_s^{n_s}\lambda_t^{n_t}\lambda_u^{n_u}\lambda_v^{n_v}\lambda_w^{n_w} d\lambda_1 \ldots d\lambda_h \qquad (16)$$

where $n_r, \ldots, n_w \geq 0$ and $n_r + \ldots + n_w = 1,2,3,4,6$. These integrals can be expanded out by using the joint characteristic function of n independent central chi-square variates, each with one degree of freedom:

$$\int \ldots \int [\exp(it_j z_j^2)][(2\pi)^{-n/2}\exp(-\tfrac{1}{2}z_j z_j)]dz_1 \ldots dz_n$$
$$= \prod_{j=1}^{n}[(1-2it_j)^{-\frac{1}{2}}]. \tag{17}$$

Substituting this formula into the integrals (16) we obtain expressions of the form:

$$(d)\left[\prod_{\substack{a = r,s,t,\\ u,v,w}} \{\int[\exp\, it_a\lambda_a^2)][(2\pi)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}\lambda_a^2)]\lambda_a^{n_a}d\lambda_a\}\right]$$
$$\cdot\left[\prod_{\substack{b \neq r,s,t,\\ u,v,w}}(1-2it_b)^{-\frac{1}{2}}\right]$$
$$= (d)\left[\prod_{\substack{a = r,s,t,\\ u,v,w}}\{\int[\exp(it_a\lambda_a^2)]i(\lambda_a)\lambda_a^{n_a}d\lambda_a\}\right]\cdot\left[\prod_{\substack{b \neq r,s,t,\\ u,v,w}}(1-2it_b)^{-\frac{1}{2}}\right]$$

where: $i(\lambda_a) = (2\pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}\lambda_a^2)$.

To expand out the terms involving $(\lambda_a)$ we use the derivative of the $\int\exp(itz^2)i(z)dz$ with respect to t i.e. the derivative of the characteristic function of a one degree of freedom central chi-square variate with respect to the characteristic function parameter:

$$\frac{\partial}{\partial t}\{\int[\exp(it\, z^2)]i(z)dz\}$$
$$= \int\frac{\partial}{\partial t}\{[\exp(it\, z^2)]\}i(z)dz$$
$$= \int(iz)[\exp(it\, z^2)]i(z)dz. \tag{18}$$

Repeating the differentiation p times gives

$$\frac{\partial^p}{\partial t^p}\{\int[\exp(itz^2)]i(z)dz\}$$
$$= \int(iz^2)^p[\exp(itz^2)]i(z)dz.$$

However from (17) we have that:

$$\int[\exp(itz^2)]\phi(z)dz = (1-2it)^{-\frac{1}{2}}.$$

So substituting this into (18) gives

$$\frac{\partial}{\partial t}\{\int[\exp(itz^2)]\phi(z)dz\} = \frac{\partial}{\partial t}\{(1-2it)^{-\frac{1}{2}}\}$$

$$= (-2i)(-\tfrac{1}{2})(1-2it)^{-3/2}$$

$$= i(1-2it)^{-3/2}$$

$$= \int(iz^2)[\exp(itz^2)]\phi(z)dz = i(1-2it)^{-3/2}$$

$$\text{so } \int[\exp(itz^2)]\phi(z)z^2dz = (1-2it)^{-3/2}.$$

In general the even power terms in the expansion are treated by considering expressions of the form:

$$\int(i\lambda_a^2)^p\exp[it_a\lambda_a^2]\phi(\lambda_a)d\lambda_a = \frac{\partial^p}{\partial t_a^p}\ (1-2it_a)^{-\frac{1}{2}})$$

$$= \frac{i^p(2p)!}{2^p(p)!}\{(1-2it_a)^{-(1+2p)/2}\};\quad p = o,1,2\ldots \quad (19)$$

The general form for the derivatives can be verified using proof by induction. From (19) we obtain:

$$\int \exp[it_a\lambda_a^2]\lambda_a^{2p}\phi(\lambda_a)d\lambda_a = \frac{(2p)!}{2^p p!}(1-2it_a)^{-(1+2p)/2}. \quad (20)$$

This covers the even power terms in the expansion but leaves the odd power terms such as:

$$\int \exp[it_a\lambda_a^2]\lambda_a\phi(\lambda_a)d\lambda_a \ \& \int \exp[it_a\lambda_a^2]\lambda_a^3\phi(\lambda_a)d\lambda_a.$$

These terms cannot be obtained by differentiating the integral expression for the characteristic function of a squared N(0,1). However, this is not a problem since on inspection such integrals with odd powers are zero because their integrands are odd functions, and the integral of an odd function over the entire real line is zero provided the integral exists. In this case the integrals clearly exist because $|\exp(itz^2)| = 1$ and so the integrals are in absolute value less than or equal to the moments of a standard normal variable.

We can now evaluate the general term in the expansion of the joint characteristic function of the squares of the asymptotically independent standard normals:

$$J = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left[it_j\lambda_j^2\right] i(\lambda)\lambda_r^{n_r} \cdots \lambda_w^{n_w}\, d\lambda$$

where $n_r, \ldots, n_w \geq 0$ and $n_r + \ldots + n_w = n = 0,1,2,3,4,6$. We note first that $J = 0$ unless $n_r, \ldots, n_w$ are all even, i.e. $n_r = 2k_r$ etc. which implies $n = 0,2,4,6 = 2k$. Thus without loss of generality we can drop $(n_u, n_v, n_w)$ and work solely with $(n_r, n_s, n_t) = (2k_r, 2k_s, 2k_r)$ where $k_r, k_s, k_t \geq 0$ and $k_r + k_s + k_t = k = 0,1,2,3$. Thus we have for all non-zero J:

$$\begin{aligned} J &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left[it_j\lambda_j^2\right] i(\lambda)\lambda_r^{2k_r}\lambda_s^{2k_s}\lambda_t^{2k_t}\, d\lambda \\ &= \left[\prod_{b\neq r,s,t} (1-2it_b)^{-\frac{1}{2}}\right]\cdot\left[\prod_{a=r,s,t} \left\{(1-2it_a)^{-(1+2k_a)/2}\, \frac{(2k_a)!}{2^{k_a}(k_a)!}\right\}\right] \\ &= J(k_r, k_s, k_t). \end{aligned} \qquad (21)$$

We can rewrite this more generally as:

$$\prod_{j=1}^{h} \{\frac{(2k_j)!}{2^{k_j}(k_j)!} (1-2it_j)^{-(1+2k_j)/2}\}$$

where now $k_j \geq 0$ is defined for all $j=1, \ldots, h$ and $\sum_{j=1}^{h} k_j = k = 0,1,2,3.$ Before substituting the $J(k_r, k_s, k_t)$ into the integrals of the form (16) and so obtaining an expansion for the joint characteristic function of the $\{\lambda^2\}$, it is desirable to introduce some new notation.

$$g_o = d_o = O(T^{-1}); \quad g_r = d_{rr} = O(T^{-1});$$

$$g_{rs} = d_{rrss} + d_{rsrs} + d_{rssr} + d_{srrs} + d_{srsr} + d_{ssrr}$$

$$= d_{rrss} [6] = O(T^{-1}); \ (r<s)$$

$$g_{rr} = d_{rrrr} = O(T^{-1});$$

$$g_{rst} = d_{rrsstt} + d_{rrstst} + \ldots = d_{rrsstt} [90] = O(T^{-1}); \ (r<s<t)$$

$$g_{rss} = d_{rrssss} + d_{rsrsss} + \ldots = d_{rrssss} [15] = O(T^{-1}); \ (r<s)$$

$$g_{rrs} = d_{rrrrss} + d_{rrrsrs} + \ldots = d_{rrrrss} [15] = O(T^{-1}); \ (r<s)$$

$$g_{rrr} = d_{rrrrrr} = O(T^{-1}).$$

As before the square brackets indicate summation across different permutations of the subscripts. Substituting these into (15) and using (21) we obtain:

$$\begin{aligned}
\Gamma_\lambda(t) = & (1+g_o)J(k_r=0, k_s=0, k_t=0) \\
& + \sum_r g_r \ J(k_r=1, k_s=0, k_t=0) \\
& + \sum_{r<s} g_{rs} \ J(k_r=1, k_2=1, k_t=0) \\
& + \sum_r g_{rr} \ J(k_r = 2, k_s=0, k_t=0) \\
& + \sum_{r<s<t} g_{rst} \ J(k_r=1, k_s=1, k_t=1) \\
& + \sum_{r<s} g_{rrs} \ J(k_r=2, k_s=1, k_t=0) \\
& + \sum_{r<s} g_{rss} J(k_r=1, k_s=2, k_t=0) \\
& + \sum_r g_{rrr} \ J(k_r=3, k_s=0, k_t=0) + O(T^{-3/2}) . \qquad (22)
\end{aligned}$$

This concludes the third step which was to obtain an expansion to $O(T^{-3/2})$ for the joint characteristic function of the $\{\lambda^2\}$. The fourth step is to obtain the expansion for characteristic function of the sum of the squares. An inverse Fourier transform can then be applied to the latter expansion to obtain an expansion for the probability density function of the sum of the squares, $s^2 = \sum_{j=1}^{h} \lambda_j^2$.

### 3.2.3 The Characteristic Function of the Sum of Squares

The characteristic function of the sum of the squares, $\sigma(t)$, can be obtained from the characteristic function of the squares, $\Gamma_\lambda(t)$, by setting $t_j = t$ for $j=1, \ldots, h$:

$$\sigma(t) = E[\exp(it\lambda_r\lambda_r)] = \Gamma_\lambda(t\underline{1})$$

where $\underline{1}' = (1, \ldots, 1)$ an h-vector of 1's. The expansion obtained for $\Gamma_\lambda(t)$, is in terms of expressions of the general form:

$$J = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp[it_j\lambda_j^2] i(\lambda)\lambda_r^{2k_r}\lambda_s^{2k_s}\lambda_t^{2k_t} d\lambda$$

$$= \left[\prod_{b\neq r,s,t} (1-2it_b)^{-\frac{1}{2}}\right]\left[\prod_{a=r,s,t} \{(1-2it_a)^{-(1+2k_a)/2} \frac{(2k_a)!}{2^{k_a}(k_a)!}\}\right]. \quad (23)$$

Setting $t_j = t$ for all $j=1, \ldots, h$ in (23) gives

$$J^* = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp[it\lambda_j\lambda_j] i(\lambda)\lambda_r^{2k_r}\lambda_s^{2k_s}\lambda_t^{2k_t} d\lambda$$

$$= (1-2it)^{-(h+2k_r+2k_s+2k_t)\frac{1}{2}} \{ \prod_{a=r,s,t} \left[\frac{(2k_a)!}{2^{k_a}(k_a)!}\right]\} . \quad (24)$$

There are now seven distinct possible expressions of the form:

$$J^* = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp[it\lambda_j\lambda_j] i(\lambda)\lambda_r^{2k_r}\lambda_s^{2k_s}\lambda_t^{2k_t} \, d\lambda \qquad (r<s<t).$$

These are:

$$J^*(k_r=0,\ k_s=0,\ k_t=0) = (1-2it)^{-h/2};$$

$$J^*(k_r=1,\ k_s=0,\ k_t=0) = (1-2it)^{-(h+2)/2}.\{\frac{(2)!}{2^1(1)!}\} = (1-2it)^{-(h+2)/2};$$

$$J^*(k_r=2,\ k_s=0,\ k_t=0) = (1-2it)^{-(h+4)/2}\ (\frac{(4)!}{2^2(2)!}) = 3(1-2it)^{-(h+4)/2};$$

$$J^*(k_r=3,\ k_s=0,\ k_t=0) = (1-2it)^{-(h+6)/2}\ (\frac{(6)!}{2^3(3)!}) = 15(1-2it)^{-(h+6)/2};$$

$$J^*(k_r=1,\ k_s=1,\ k_t=0) = (1-2it)^{-(h+4)/2}\ (\frac{(2)!}{2\ (1)!})\ (\frac{(2)!}{2^1(1)!})$$

$$= (1-2it)^{-(h+4)/2};$$

$$J^*(k_r=2,\ k_s=1,\ k_t=0) = (1-2it)^{-(h+6)/2}\ (\frac{(4)!}{2^2(2)!})(\frac{(2)!}{2^1(1)!})$$

$$= 3(1-=it)^{-(h+6)/2};$$

$$J^*(k_r=1,\ k_s=1,\ k_t=1) = (1-2it)^{-(h+6)/2}(\frac{(2)!}{2^1(1)!})\ (\frac{(2)!}{2^1(1)!})\ (\frac{(2)!}{2^1(1)!})$$

$$= (1-2it)^{-(h+6)/2}. \qquad (25)$$

So taking (22), setting $t_j = t$ for all $j=1, \ldots, h$ and substituting in the above $J^*$ expressions (25) gives an expansion to $O(T^{-3/2})$ for $\sigma(t)$:

$$\sigma(t) = \Gamma_\lambda(t\ \underline{1}) = (1 + g_o)(1-2it)^{-h/2} + (\sum_r g_r)(1-2it)^{-(h+2)/2}$$
$$+ [(\sum_{r<s} g_{rs}) + 3(\sum_r g_{rr})](1-2it)^{-(h+4)/2} + [(\sum_{r<s<t} g_{rst})$$
$$+ 3(\sum_{r<s} g_{rss}) + 3(\sum_{r<s} g_{rrs}) + 15\ (\sum_r g_{rrr})](1-2it)^{-)h+6)/2}$$
$$+ O(T^{-3/2}).$$

This can be written more simply as:

$$\sigma(t) = (1 + a_o)(1-2it)^{-h/2} + a_1(1-2it)^{-(h+2)/2} + a_2(1-2it)^{-(h+4)/2}$$
$$+ a_3(1-2it)^{-(h+6)/2} + O(T^{-3/2})$$

where: $a_o = g_o = O(T^{-1})$;

$$a_1 = \sum_r g_r = O(T^{-1});$$

$$a_2 = (\sum_{r<s} g_{rs}) + 3(\sum_r g_{rr}) = O(T^{-1});$$

$$a_3 = (\sum_{r<s<t} g_{rst}) + 3(\sum_{r<s} g_{rss}) + 3(\sum_{r<s} g_{rrs}) + 15(\sum_r g_{rrr}) = O(T^{-1}). \quad (26)$$

3.2.4 The Density and Distribution Functions of the Sum of Squares

One of the advantages of obtaining the expansion for the characteristic function of the sum of squares in the form of (26) is that it is an expansion in terms of the characteristic functions of central chi-square variates with increasing degrees of freedom. Thus when we carry out an inverse Fourier Transform on (26) the linear combination of characteristic functions transforms to the same linear combination in the corresponding density functions. Thus:

$$\text{pdf}(s^2) = (1 + a_o)\chi^2_h(s^2) + a_1\chi^2_{h+2}(s^2) + a_2\chi^2_{h+4}(s^2) + a_3\chi^2_{h+6}(s^2) + O(T^{-3/2}), \quad (27)$$

where $\chi^2_h(.)$ is the p.d.f. of a central chi-square variate with h degrees of freedom:

$$\chi^2_h(x) = \frac{x^{(h/2)-1}e^{-x/2}}{2^{(h/2)}\Gamma(h/2)}$$

because taking an inverse Fourier Transform is a linear operation. Likewise integrating is a linear operation so the expansion for the c.d.f. of $s^2$ is:

$$\Pr[s^2 \leq r^2] = (1 + a_0)\xi_h^2(r^2) + a_1\,\xi_{h+2}^2(r^2) + a_2\,\xi_{h+4}^2(r^2)$$
$$+ a_3\,\xi_{h+6}^2(r^2) + O(T^{-3/2})$$

where $\xi_h^2(\cdot)$ is the c.d.f. of a central chi-square variate with h degrees of freedom:

$$\xi_h^2(r^2) = \int_0^{r^2} \frac{x^{(h/2)-1}e^{-x/2}}{2^{(h/2)}\Gamma(h/2)}\,dx \,. \qquad (28)$$

An alternative expression also correct to $O(T^{-3/2})$ though not numerically identical is:

$$\Pr\{s^2 \leq r^2\} = \xi_h^2[(1 + a_0^*)(r^2) + a_1^*\,(r^2)^2$$
$$+ a_2^*\,(r^2)^3] + O(T^{-3/2})$$

see Sargan (1980b). Both of these expressions can be implemented easily by computer.

A feature worth noting is that these expressions do not include any $O(T^{-\frac{1}{2}})$ terms. This contrasts with the expansions for the pdf and cdf of the $\lambda$ which do include $O(T^{-\frac{1}{2}})$ terms. In the scalar case, $s^2 = \lambda^2$, the difference between testing using $\lambda$ and using $s^2$ is that with the former tests of both one- and two-sided hypotheses (e.g. $\theta > 0$ or $\theta \neq 0$) are possible, whereas with the latter only tests of two-sided hypotheses (e.g. $\theta \neq 0$) are possible. More generally, using $s^2$ only permits spherically symmetric critical regions in $\lambda$ to be used. The information from the $O(T^{-\frac{1}{2}})$ terms in the expansion of the p.d.f. of $\lambda$ is lost because the $O(T^{-\frac{1}{2}})$ terms are associated with odd powers of $\{\lambda\}$ and these integrate to zero over spherically symmetric critical regions in $\lambda$.

## 3.3 The Approximate Inner Product Case

### 3.3.1 The Expansion of the Statistic

Although the class of exact inner product, asymptotically chi-square test statistics is very extensive, it does not include all the asymptotically chi-square test statistics used in econometrics. In particular, it omits most test statistics constructed on the Likelihood Ratio (LR) principle. Although the LR test statistic for a given hypothesis in a given model may be well approximated by a quadratic form in asymptotically independent standard normal variates, there is in general no reason why the LR statistic should exactly equal a quadratic form.

The method developed in Section 3.2.2 for the exact sum of squares case can be modified to deal with approximate inner product statistics. The first step is to make a stochastic expansion of the statistic in a vector of n asymptotically independent standard normal variates (where $n \geq h$ the number of restrictions being tested):

$$s^2 = (\sum_{j=1}^{h}\lambda_j^2) + (b_o) + (b_j\lambda_j) + (b_{jk}\lambda_j\lambda_k) + (b_{jk\ell}\lambda_j\lambda_k\lambda_\ell) + (b_{jk\ell m}\lambda_j\lambda_k\lambda_\ell\lambda_m) + 0_p(T^{-3/2})$$

where: $b_o = O(T^{-\frac{1}{2}})$; $b_j = O(T^{-\frac{1}{2}})$; $b_{jk} = O(T^{-\frac{1}{2}})$;

$b_{jk\ell} = O(T^{-\frac{1}{2}})$; $b_{jk\ell m} = O(T^{-1})$. (29)

The expansion uses tensor summation notation except for the leading term $(\sum_{j=1}^{h}\lambda_j^2)$ which it is easier to keep in explicit summation form. The error is $0_p(T^{-3/2})$ i.e. it is of order of magnitude $T^{-3/2}$ in probability. This permits the statistic to be a function of other random variates than the $\lambda$, namely w, but where these w variates do not affect the statistic except to order $T^{-3/2}$. The main advantage of this is that the joint distribution of $\lambda$ and w does not have to be approximated.

We assume that the statistic can be expanded in the form of Equation (29). The work of Barndorff-Nielsen and Cox (1984) on LR statistics, Mauleon-Torres (1983) and Sargan (1980b) seems to indicate that this is frequently the case. If (29) were not valid (e.g. there is a fifth derivative term of $O(T^{-1})$ ) then the method given would have to be extended somewhat but would remain essentially the same.

A particular problem with expanding a test statistic into the form of (29) arises in dynamic models such as:

$$y_t = \rho y_{t-1} + \beta_o x_t + \beta_1 x_{t-1} + u_t \qquad (30)$$

where test statistics often involve terms such as:

$$m_1 = \sum_{t=1}^{T} y_t^2/T \quad \text{and} \quad \sum_{t=0}^{T-1} y_t^2/T = \sum_{t=1}^{T} y_{t-1}^2/T = m_2 .$$

The difficulty here arises because although in finite samples $m_1$ and $m_2$ when suitably standardized are not perfectly correlated, asymptotically they are. Thus it is not immediately clear how the vector $\lambda$ should be defined. This does not appear to have been considered in the existing literature, e.g. Mauleon-Torres (1983). Whilst it is usually valid to expand the statistic in terms of standardized $m_1$ and $m_2$ variates, it is not possible to transform the standardized $m_1$ and $m_2$ by a linear transformation into $\lambda_1$ and $\lambda_2$, two asymptotically independent standard normal variates since even asymptotically

$$T(m_1 - m_2) = y_T^2 - y_0^2$$

is not distributed normally. The problem of asymptotic singularity cannot be resolved by a simple transformation since some of the elements of the transformation matrix must tend to infinity as T does and there is no guarantee that the transformed variables will be distributed asymptotically jointly normally.

In this case we can solve the problem by standardizing $(\Sigma y_t^2, T)$, $y_T$ and $y_0$ to have zero means and unit variances and then constructing $\lambda$ using these instead of the appropriately standardized $(\Sigma y_t^2/T)$ and $(\Sigma y_{t-1}^2/T)$. Under fairly general conditions, the standardized ( $y_t/T$), $y_T$ and $y_0$ will be asymptotically jointly normally distributed with a zero mean vector and a non-singular covariance matrix. The statistic will now be a function of T but this is allowed for in the form of Equation (29). In adition the standardization of $y_T$ depends upon $x_T$ which usually will not converge so that $(T^{1/2}b_0)$, $(T^{1/2}b_j)$ etc. will not converge as T tends to infinity. However this is not a problem provided that these terms are all of O(1) and that Equation (29) provides a sequence of approximations with an approximation error of $O_p(T^{-3/2})$ and provided that the characteristic function of $\lambda$ is appropriately well-behaved.

### 3.3.2 The Characteristic Function of the Statistic

The first step in obtaining an expansion for the characteristic function of the statistic is to expand $[\exp(its^2)]$ using (29). This is where the approximate inner product case starts to differ from the exact inner product case of Section 3.2. This expansion gives:

$$\exp(its^2) = \exp(it \sum_{j=1}^{h} \lambda_j^2) P_T(\lambda, it) + O_p(T^{-3/2})$$

where

$$\begin{aligned} P_T(\lambda, it) = {} & 1 + (it)(b_o) + (it)(b_j\lambda_j) + (it)(b_{jk}\lambda_j\lambda_k) \\ & + (it)(b_{jk\ell}\lambda_j\lambda_k\lambda_\ell) + (it)(b_{jk\ell m}\lambda_j\lambda_k\lambda_\ell\lambda_m) \\ & + (it)^2(c_o) + (it)^2(c_j\lambda_j) + (it)^2(c_{jk}\lambda_j\lambda_k) \\ & + (it)^2(c_{jk\ell}\lambda_j \dots \lambda_\ell) + (it)^2(c_{jk\ell m}\lambda_j \dots \lambda_m) \\ & + (it)^2(c_{jk\ell mp}\lambda_j \dots \lambda_p) + (it)^2(c_{jk\ell mpq}\lambda_j \dots \lambda_q) \end{aligned}$$

where

$$\begin{aligned} c_o &= \tfrac{1}{2}b_o^2 = O(T^{-1}); \\ c_j &= b_o b_j = O(T^{-1}); \\ c_{jk} &= b_o b_{jk} + \tfrac{1}{2}b_j b_k = O(T^{-1}); \\ c_{jk\ell} &= b_o b_{jk\ell} + b_j b_{k\ell} = O(T^{-1}); \\ c_{jk\ell m} &= b_j b_{k\ell m} + \tfrac{1}{2}b_{jk} b_{\ell m} = O(T^{-1}); \\ c_{jk\ell mp} &= b_{jk} b_{\ell mp} = O(T^{-1}); \\ c_{jk\ell mpq} &= \tfrac{1}{2}b_{jk\ell} b_{mpq} = O(T^{-1}). \end{aligned} \qquad (31)$$

The expansion for the characteristic function of the statistic is then obtained by taking the expectation of (31) to $O(T^{-3/2})$ using a multivariate Edgeworth expansion for the pdf of $\lambda$, as obtained in Section 3.2.1:

$$f_T(\lambda) = i(\lambda)\, Q_T(\lambda) + O(T^{-3/2})$$

where $i(\lambda) - (2\pi)^{-n/2}\exp\{-\tfrac{1}{2}\lambda_j\lambda_j\}$

$$Q_T(\lambda) = 1 + d_o + (d_j\lambda_j) + (d_{jk}\lambda_j\lambda_k) + d_{jk\ell}\lambda_j\lambda_k\lambda_\ell)$$
$$+ (d_{jk\ell m}\lambda_j \cdots \lambda_m) + (d_{jk\ell mpq}\lambda_j \cdots \lambda_q)$$

with $d_o = O(T^{-1})$;
$d_j = O(T^{-\frac{1}{2}})$;
$d_{jk} = O(T^{-1})$;
$d_{jk\ell} = O(T^{-\frac{1}{2}})$;
$d_{jk\ell m} = O(T^{-1})$;
$d_{jk\ell mpq} = O(T^{-1})$; (32)

(from Equation (14)). The characteristic function of $s^2$ is then given by:

$$\sigma(t) = E[\exp\{its_1^2\}]$$
$$= \int\exp\{its^2\}f_T(\lambda)d\lambda + O(T^{-3/2})$$
$$= \int\exp\{it\sum_{j=1}^{h}\lambda_j^2\}i(\lambda)P_T(\lambda,\ it)Q_T(\lambda)d\lambda + O(T^{-3/2}). \qquad (33)$$

For (33) to be valid we require further conditions on the joint behaviour of the $\lambda$ and $w$; see Sargan (1980). To proceed we need to expand $[P_T(\lambda),\ it)][Q_T(\lambda)]$ to $O(T^{-3/2})$ giving $[R_T(\lambda,\ it)]$:

$$R_T(\lambda,\ it) = 1 + (e_o) + (e_j\lambda_j) + (e_{jk}\lambda_j\lambda_k) + (e_{jk\ell}\lambda_j \cdots \lambda_\ell)$$
$$+ (e_{jk\ell m}\lambda_j \cdots \lambda_m) + (e_{jk\ell mp}\lambda_j \cdots \lambda_p) + (e_{jk\ell mpq}\lambda_j \cdots \lambda_q).$$

$$\text{where } e_o = (d_o) + (it)(b_o) + (it)^2(c_o) = O(T^{-\frac{1}{2}});$$
$$e_j = (d_j) + (it)(b_j + b_o d_j) + (it)^2(c_j) = O(T^{-\frac{1}{2}});$$
$$e_{jk} = (d_{jk}) + (it)(b_{jk} + d_j b_k) + (it)^2(c_{jk}) = O(T^{-\frac{1}{2}});$$
$$e_{jk\ell} = (d_{jk\ell}) + (it)(b_{jk\ell} + d_j b_{k\ell} + b_o d_{jk\ell}) + (it)^2(c_{jk\ell}) = O(T^{-\frac{1}{2}});$$
$$e_{jk\ell m} = (d_{jk\ell m}) + (it)(b_{jk\ell m} + d_j b_{k\ell m} + d_{jk\ell} b_m) + (it)^2(c_{jk\ell m}) = O(T^{-1});$$
$$e_{jk\ell mp} = (it)(d_{jk\ell} b_{mp}) + (it)^2(c_{jk\ell mp}) = O(T^{-1});$$
$$e_{jk\ell mpq} = (d_{jk\ell mpq}) + (it)(d_{jk\ell} b_{mpq}) + (it)^2(c_{jk\ell mpq}) = O(T^{-1}). \quad (34)$$

We then substitute (34) into (33) to obtain:

$$\sigma(t) = \int\left[\exp(it \sum_{j=1}^{h} \lambda_j^2)\right] i(\lambda) R_T(\lambda, it) d\lambda + O(T^{-3/2}).$$

As in Section 3.2.2 (The Joint Characteristic Function of the Squares) the terms with odd powers in the $\{\lambda_j\}$ vanish since they involve integrating odd functions :

$$\sigma(t) = \int\left[\exp(it \sum_{j=1}^{h} \lambda_j^2)\right] i(\lambda) R_T^*(\lambda, it) d\lambda + O(T^{-3/2})$$

where: $R_T^*(\lambda, it) = 1 + a_o + (a_j \lambda_j^2) + (a_{jk} \lambda_j^2 \lambda_k^2) + (a_{jk\ell} \lambda_j^2 \lambda_k^2 \lambda_\ell^2)$

$(j \leq k \leq \ell)$,

and where: $a_o = (a_o^o) + (it)(a_o^1) + (it)^2(a_o^2) = e_o = O(T^{-\frac{1}{2}});$

$$a_j = (a_j^o) + (it)(a_j^1) + (it)^2(a_j^2) = e_{jj} = O(T^{-\frac{1}{2}});$$
$$a_{jk} = (a_{jk}^o) + (it)(a_{jk}^1) + (it)^2(a_{jk}^2) = e_{jjkk}[6] = O(T^{-1}); \quad (j<k),$$
$$a_{jj} = (a_{jj}^o) + (it)(a_{jj}^1) + (it)^2(a_{jj}^2) = e_{jjjj} = O(T^{-1});$$

$$a_{jk\ell} = (a^0_{jk\ell}) + (it)(a^1_{jk\ell}) + (it)^2(a^2_{jk\ell}) = e_{jjkk\ell\ell}[90]$$
$$= O(T^{-1}); \quad (j<k<\ell)$$
$$a_{jkk} = (a^0_{jkk}) + (it)(a^1_{jkk}) + (it)^2(a^2_{jkk}) = e_{jjkkkk}[15]$$
$$= O(T^{-1}); \quad (j<k)$$
$$a_{jjk} = (a^0_{jjk}) + (it)(a^1_{jjk}) + (it)^2(a^2_{jjk}) = e_{jjjjkk}[15]$$
$$= O(T^{-1}); \quad (j<k)$$
$$a_{jjj} = (a^0_{jjj}) + (it)(a^1_{jjj}) + (it)^2(a^2_{jjj}) = e_{jjjjjj} = O(T^{-1});$$

(35)

(all terms are $O(T^{-1})$ except $(a^1_0)$ and $(a^1_j)$ which are $O(T^{-\frac{1}{2}})$).

Superscripts on the {a} refer to the power of (it) to which they are attached. The statistic is asymptotically only a function of the $\{\lambda_j\}(j=1, \ldots, h)$ so at this point we integrate out the $\{\lambda_k\}(k=h+1, \ldots, n)$ to leave the integral solely in terms of the $\{\lambda_j\}(j=1, \ldots, h)$ and (it). As in Section 3.2.2 each term in the expansion of $\sigma(t)$ is multiplicatively separable. The general term in the expansion of $\sigma(t)$ is:

$$\left[\int \exp(it \sum_{j=1}^{h} \lambda_j^2)(2\pi)^{-h/2} \exp(-\tfrac{1}{2} \sum_{j=1}^{h} \lambda_j^2)\lambda_k^{2p_k}\lambda_\ell^{2p_\ell}\lambda_m^{2p_m} d\lambda_1 \ldots d\lambda_h\right]$$

$$\cdot \left[\int (2\pi)^{-(n-h)/2} \exp(-\tfrac{1}{2} \sum_{a=h+1}^{n} \lambda_a^2)\lambda_b^{2p_b}\lambda_c^{2p_c}\lambda_d^{2p_d} d\lambda_{h+1} \ldots d\lambda_n\right]$$

multiplied by a quadratic in (it),

where: $k,\ell,m = 1,\ldots,h$

$b,c,d = h+1,\ldots,n$

$p_k, p_\ell, p_m, p_b, p_c, p_d = 0,1,2,3$

$p_k + p_\ell + p_m + p_b + p_c + p_d = 0,1,2,3.$ (36)

The second integral in Equation (36) is simply the expectation of a polynomial from a jointly independent standard normal distribution. Therefore:

$$\int (2\pi)^{-(n-h)/2} \exp(-\tfrac{1}{2} \sum_{a=h+1}^{n} \lambda_a^2)\lambda_b^{2p_b}\lambda_c^{2p_c}\lambda_d^{2p_d} d\lambda_{h+1} \ldots d\lambda_n$$

$$= \left[\frac{(2b_p)!}{2^{(p_b)}(p_b)!}\right]\left[\frac{(2p_c)!}{2^{(p_c)}(p_c)!}\right]\left[\frac{(2p_d)!}{2^{(p_d)}(p_d)!}\right]$$

from (20). Defining $\gamma_j = \lambda_j$ $(j=1,\ldots,h)$ we can then re-express (35) as:

$$\sigma(t) = \int \left[\exp(it\gamma_j\gamma_j)\right] i(\gamma) S_T(\gamma, it) d\gamma + O(T^{-3/2})$$

where: $S_T(\gamma, it) = 1 + (g_o) + (g_j\gamma_j^2) + (g_{jk}\gamma_j^2\gamma_k^2) + (g_{jk\ell}\gamma_j^2\gamma_k^2\gamma_\ell^2)$; $(j \le k \le \ell)$

and where:

$$g_o = (g_o^0) + (it)(g_o^1) + (it)^2(g_o^2) = (a_o) + \sum_{h+1 \le b}^{n} (a_b + 3a_{bb} + 15a_{bbb})$$
$$+ \sum_{h+1 \le b < c}^{n} (a_{bc} + 3a_{bcc} + 3a_{bbc}) + \sum_{h+1 \le b < c < d}^{n} (a_{bcd}) = O(T^{-\frac{1}{2}});$$

$$g_j = (g_j^0) + (it)(g_j^1) + (it)^2(g_j^2) = (a_j) + \sum_{h+1 \le b}^{n} (a_{jb} + 3a_{jbb})$$
$$+ \sum_{h+1 \le b < c}^{n} (a_{jbc}) = O(T^{-\frac{1}{2}}); \quad (1 \le j \le h)$$

$$g_{jk} = (g_{jk}^0) + (it)(g_{jk}^1) + (it)^2(g_{jk}^2) = (a_{jk}) + \sum_{h+1 \le b}^{n} (a_{jkb})$$
$$= O(T^{-1}); \quad (1 \le j \le k \le h)$$

$$g_{jk\ell} = (g_{jk\ell}^0) + (it)(g_{jk\ell}^1) + (it)^2(g_{jk\ell}^2) = (a_{jk\ell}) = O(T^{-1});$$
$$(1 \le j \le k \le \ell \le h)$$

(all terms are $O(T^{-1})$ except $(g_0^1)$ and $(g_j^1)$ which are $O(T^{-\frac{1}{2}})$).

The general term in this expansion for $\sigma(t)$ is:

$$\int[\exp(it\gamma_j\gamma_j)]l(\gamma)\gamma_j^{2p_j}\gamma_k^{2p_k}\gamma_\ell^{2p_\ell}d\gamma$$

multiplied by a quadratic in (it) (where $1\leq j\leq k\leq \ell\leq h$ and $p_j,p_k,p_\ell\leq 0$ and $p_j+p_k+p_\ell = 0,1,2,3$).

This is very similar to the expression (24) which was examined in Section 3.2.3 (The Characteristic Function of the Sum of the Squares) and using (24) the general term is then equal to:

$$(1-2it)^{-(h+2p_j+2p_k+2p_\ell)\frac{1}{2}}\{\prod_{a=j,k,\ell}\left[\frac{(2p_a)!}{2^{p_a}(p_a)!}\right]\}$$

multiplied by the quadratic in (it). Substituting these expressions into the expansion for $\sigma(t)$ gives:

$$\sigma(t) = (1+s_0)(1-2it)^{-h/2} + (s_1)(1-2it)^{-(h+2)/2}$$
$$+ (s_2)(1-2it)^{-(h+4)/2} + (s_3)(1-2it)^{-(h+6)/2} + O(T^{-3/2})$$

where: $s_0 = (s_0^0) + (it)(s_0^1) + (it)^2(s_0^2) = g_0 = O(T^{-\frac{1}{2}})$;

$s_1 = (s_1^0) + (it)(s_1^1) + (it)^2(s_1^2) = \sum_j(g_j) = O(T^{-\frac{1}{2}})$;

$s_2 = (s_2^0) + (it)(s_2^1) + (it)^2(s_2^2) = \sum_{j<k}(g_{jk}) + 3\sum_j(g_{jj}) = O(T^{-1})$;

$s_3 = (s_3^0) + (it)(s_3^1) + (it)^2(s_3^2)$

$$= \sum_{j<k<\ell}(g_{jk\ell}) + 3\sum_{j<k}(g_{jkk} + g_{jjk}) + 15\sum_j(g_{jjj}) = O(T^{-1}); \quad (37)$$

(all terms are $O(T^{-1})$ except $(s_0^1)$ and $(s_1^1)$ which are $O(T^{-\frac{1}{2}})$).

### 3.3.3 The Distribution Function of the Statistic

Equation (37) is very similar to (26) except that {s} terms are quadratics in (it) and $(s_0^1)$ and $(s_1^1)$ are $O(T^{-\frac{1}{2}})$. The former point creates a problem if h, the dimension of λ (and the degrees of freedom of the asymptotic distribution of $s^2$), is not sufficiently large. This can be seen by considering the term:

$$-(it)\,(1-2it)^{-h/2} = \left[\frac{-it}{(1-2it)}\right]\left[\frac{1}{(1-2it)}\right]^{(h-2)/2}$$

$$= \tfrac{1}{2}\{1 - \left[\frac{1}{(1-2it)}\right]\}\left[\frac{1}{(1-2it)}\right]^{(h-2)/2}$$

$$= (\tfrac{1}{2})(1-2it)^{-(h-2)/2} - (\tfrac{1}{2})(1-2it)^{-h/2} \quad (38)$$

If h=1, then $(1-2it)^{-(h-2)/2}$ is not a characteristic function and if h=2 then $(1-2it)^{-(h-2)/2}$ is the characteristic function of a degenerate distribution with all its probability mass at the origin. A similar problem occurs with:

$$(it)^2(1-2it)^{-h/2} = \{\frac{(it)}{(1-2it)}\}^2\,(1-2it)^{-(h-4)/2}$$

$$= [\tfrac{1}{2}\{\,1 - (\frac{1}{1-2it})\}]^2\,(1-2it)^{-(h-4)/2}$$

$$= \{(\tfrac{1}{4})-(\tfrac{1}{2})(1-2it)^{-1} + (\tfrac{1}{4})(1-2it)^{-2}\}(1-2it)^{-(h-4)/2}$$

$$= (\tfrac{1}{4})(1-2it)^{-(h-4)/2} - (\tfrac{1}{2})(1-2it)^{-(h-2)/2}$$

$$+ (\tfrac{1}{4})(1-2it)^{-h/2}. \quad (39)$$

The inability to express Equation (37) as a linear combination of the characteristic functions of chi-square variates with varying degrees of freedom is not in itself a problem. However it does mean that if we apply an inverse Fourier transform to Equation (37) we will not obtain a linear combination of chi-square pdf's with varying degrees of freedom as the expansion for the pdf of $s^2$. Therefore we need to find some alternative way to perform the inverse Fourier transform to that given in Section 3.2.4 where it is assumed that the expansion for the charactersitic function of $s^2$ consists of a linear combination of the characteristic functions of chi-square variates with varying degrees of freedom.

For h=1,2,3 then $(1-2it)^{-(h-4)/2}$ is not a true characteristic function; for h=4 it is again the characteristic function of a degenerate distribution with all its probability mass at the origin. Mauleon-Torres (1983) obtains very similar terms by a different, though related, method and attempts to carry out an inverse Fourier transform on them using the derivatives of the chi-square density function:

$$f(\delta\ ;h)\ \begin{cases} = (2\pi)^{-1}\int[\exp(-it\delta\ )](1-2it)^{-h/2}dt\ ; & \delta\ \geq 0 \\ = \ 0 & ;\ \delta\ < 0 \end{cases} \qquad (40)$$

where $\delta$ is a central chi-square with h degrees of freedom so that its density function is:

$$f(\delta\ ;h) = \begin{cases} \dfrac{(\delta\ )^{(h/2)-1}\ e^{-(\delta\ /2)}}{2^{(h/2)}\ \Gamma(h/2)} ; & \delta\ \geq 0 \\ 0 & ;\ \delta\ < 0. \end{cases}$$

Then differentiating both sides of Equation (40) gives:

$$\frac{\partial f(\delta;h)}{\partial \delta} = \begin{cases} (2\pi)^{-1}\frac{\partial}{\partial \delta}\ (\int[\exp(-it\delta\ )](1-2it)^{-h/2}dt); & \delta\ \geq 0 \\ 0 & ;\ \delta\ \quad 0 \end{cases}$$

$$= \begin{cases} -(2\pi)^{-1}\int[\exp(-it\delta\ )](it)(1-2it)^{-h/2}dt; & \delta\ \geq 0 \\ 0 & ;\ \delta\ < 0. \end{cases}$$

Thus the inverse Fourier transforms of terms such as $(it)(1-2it)^{-h/2}$ and $(it)^2(1-2it)^{-h/2}$ can be given in terms of derivatives of the density functions of central chi-squares. This still leaves a problem if h=1,2,3,4 from the $(s_0^1)$, $(s_0^2)$ and $(s_1^2)$ terms in Equation (37) at the origin. This is because when integrating the approximate density function (obtained by the inverse Fourier transform as in Equation (40)) to obtain the approximate cumulative distribution function the result includes terms such as:

$$\int \left[ \frac{\partial}{\partial\delta} \{f(\delta;h)\} \right] d\delta = f(\delta;h) + \text{constant} .$$

When h=1 then as $\delta \to 0$, $f(\delta;h) \to \infty$ and so the approximate distribution function tends to infinity. This is for the $(s_0^1)$ term when h=1; when h=2 then the $(s_0^1)$ term makes a contribution of $(\frac{1}{2})$. Similar problems arise for the $(s_1^2)$ term when h=1,2 and for the $(s_0^2)$ term when h = 1,2,3,4. Thus unless $h \geq 5$ the approximate cumulative distribution function will not take the value zero when $s^2 = 0$ and may in fact become infinite when $s^2 = 0$.

The problem lies in attempting to approximate the cdf of the statistic in an inappropriate fashion. In the exact sum of squares case the minimum of the statistic is equal to zero so that it has support on $[0,\infty)$. It is not unreasonable to approximate the cdf of a statistic with support on $[0,\infty)$ using cdf's also with support on $[0,\infty)$. However, there is nothing in Equation (29) which forces the statistic $s^2$ to have support only on $[0,\infty)$. It is convenient to expand out Equation (29) by defining $\gamma_j = \lambda_j$ $(j=1,\dots,h)$ and $\eta_a = \lambda_{h+a}(a=1,\dots,n-h)$:

$$\begin{aligned} s^2 &= (\gamma_j\gamma_j) + (c_o^o) + (c_a^o\eta_a) + (c_{ab}^o\,\eta_a\eta_b) + (c_{abc}^o\eta_a\eta_b\eta_c) \\ &\quad + (c_{abcd}^o\eta_a\eta_b\eta_c\eta_d) + (c_o^j\gamma_j) + (c_a^j\eta_a\gamma_j) + (c_{ab}^j\eta_a\eta_b\gamma_j) \\ &\quad + (c_{abc}^j\eta_a\eta_b\eta_c\gamma_j) + (c_o^{jk}\gamma_j\gamma_k) + (c_a^{jk}\eta_a\gamma_j\gamma_k) + (c_{ab}^{jk}\eta_a\eta_b\gamma_j\gamma_k) \\ &\quad + (c_o^{jk\ell}\gamma_j\gamma_k\gamma_\ell) + (c_a^{jk\ell}\eta_a\gamma_j\gamma_k\gamma_\ell) + (c_o^{jk\ell m}\gamma_j\cdots\gamma_m) \\ &\quad + O_p(T^{-3/2}). \end{aligned} \tag{41}$$

The minimum of $s^2$ given $\eta$ need neither remain constant nor occur at the same value of $\gamma$ as $\eta$ is varied. It is generally possible by transforming $\gamma$ to $\gamma^*$ using a location shift involving a polynomial in $\eta$ to ensure that a local minimum to $O(T^{-3/2})$ of $s^2$ given $\eta$ occurs at $\gamma^* = 0$ for any $\eta$. This location shift will be of $O(T^{-\frac{1}{2}})$ so that $\gamma^*$ and $\eta$ have an asymptotically jointly independent standard normal distribution.

However, this transformation does not surmount the problem that the minimum of $s^2$ given $\eta$ depends on $\eta$ and is possibly not equal to zero. If the minimum of $s^2$ given $\eta$ equals q and does not depend on $\eta$ then we can construct $\bar{s}^2 = s^2 - q$ which has a minimum value of zero irrespective of $\eta$. When this is not possible then there will always be a region of $\eta$ with probability O(1) such that the minimum value of $s^2$ is $O(T^{-\frac{1}{2}})$; see Figure 3.1 for an illustration of this.

This creates difficulties with any attempt to approximate the cdf of $s^2$ using an Edgeworth expansion. The cdf of $s^2$ at the origin is non-zero but the expansion is a linear combination of the asymptotic cdf and the derivatives of chi-square cdf's thus including terms which are non-zero at the origin when h=1,2,3,4. The response we adopt to this problem is to approximate the cdf in terms of the cdf's of central chi-square variates which have themselves been subjected to location shifts. Expanding Equation (37) gives:

$$\begin{aligned}\sigma(t) = {} & [1+r_0](1-2it)^{-h/2} + (r_1)(it)(1-2it)^{-h/2} \\ & + (r_2)(it)^2(1-2it)^{-h/2} + (r_3)(1-2it)^{-(h+2)/2} \\ & + (r_4)(1-2it)^{-(h+4)/2} + (r_5)(1-2it)^{-(h+6)/2} + O(T^{-3/2}),\end{aligned}$$

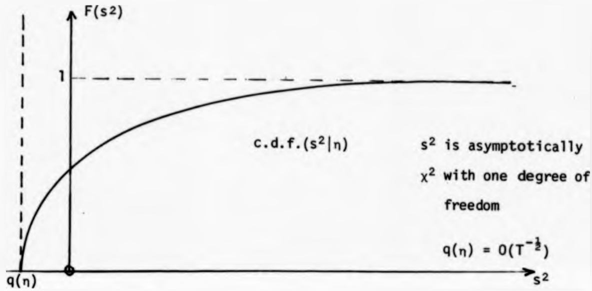


Figure 3.1 : Conditional Distribution of a Statistic with a Minimum Not Equal to Zero

$$\begin{aligned}
\text{where: } r_0 &= (s_0^0) - (\tfrac{1}{2})(s_1^1) - (\tfrac{1}{4})(s_1^2) + (\tfrac{1}{4})(s_2^2) &&= O(T^{-\frac{1}{2}}); \\
r_1 &= (s_0^1) - (\tfrac{1}{2})(s_1^2) &&= O(T^{-\frac{1}{2}}); \\
r_2 &= (s_0^2) &&= O(T^{-1}); \\
r_3 &= (s_1^0) + (\tfrac{1}{2})(s_1^1) + (\tfrac{1}{4})(s_1^2) - (\tfrac{1}{2})(s_2^1) && \\
&\quad - (\tfrac{1}{2})(s_2^2) + (\tfrac{1}{4})(s_3^2) &&= O(T^{-\frac{1}{2}}); \\
r_4 &= (s_2^0) + (\tfrac{1}{2})(s_2^1) + (\tfrac{1}{4})(s_2^2) - (\tfrac{1}{2})(s_3^1) - (\tfrac{1}{2})(s_3^2) &&= O(T^{-1}); \\
r_5 &= (s_3^0) + (\tfrac{1}{2})(s_3^1) + (\tfrac{1}{4})(s_3^2) &&= O(T^{-1}). \qquad (42)
\end{aligned}$$

Next consider a variate $\delta$ with a central chi-square distribution with h degrees of freedom. Then:

$$\exp(it(\delta + q)) = \exp(it\delta)[1 + (it)q + \tfrac{1}{2}(it)^2q^2] + O(T^{-3/2})$$

where $q = O(T^{-\frac{1}{2}})$. The characteristic function of $(\delta + q)$ is given as:

$$\begin{aligned}
E_\delta[\exp(it(\delta + q))] &= E_\delta[\exp(it\delta)\{1 + (it)q + \tfrac{1}{2}(it)^2q^2\}] + O(T^{-3/2}) \\
&= E_\delta[\exp(it\delta^2)] + q(it)E_\delta[\exp(it\delta)] + \tfrac{1}{2}q\,(it)^2E_\delta[\exp(it\delta)] \\
&\quad + O(T^{-3/2}) \\
&= [1 + q(it) + \tfrac{1}{2}q^2(it)^2](1-2it)^{-h/2} + O(T^{-3/2}). \qquad (43)
\end{aligned}$$

Using (43) it is possible to approximate the terms from (42) to $O(T^{-3/2})$:

$$\begin{aligned}
&[r_1(it) + r_2(it)^2](1-2it)^{-h/2} \\
&= E_\delta[\exp(it(\delta + \tfrac{r_1}{2})) - \exp(it(\delta - \tfrac{r_1}{2})) \\
&+ \mathrm{sign}(r_2)\exp(it(\delta + |r_2|^{\frac{1}{2}}) + \mathrm{sign}(r_2)\exp(it(\delta - |r_2|^{\frac{1}{2}})) \\
&- 2\,\mathrm{sign}(r_2)\exp(it\delta)] + O(T^{-3/2}). \qquad (44)
\end{aligned}$$

Firstly, we note that :

$$E_\delta[\exp\{it(\delta+q)\} - \exp\{it(\delta-q)\}]$$
$$= [\ 1 + q(it) + (1/2)(it)^2 - 1 + q(it) - (1/2)(it)^2\ ](1-2it)^{-h/2} + O(T^{-3/2})$$
$$= 2q(it)(1-2it)^{-h/2}.$$

Thus if we put $q = (1/2)r_1$ then we obtain the requisite parts of Equation (44). Secondly, we note that :

$$E_\delta[\exp\{it(\delta+p)\} + \exp\{it(\delta-p)\} - 2\exp\{it\delta\}]$$
$$= [\ 1+p(it)+(1/2)p^2(it)^2+1-p(it)+(1/2)p^2-2](1-2it)^{-h/2} + O(T^{-3/2})$$
$$= p^2(it)^2(1-2it)^{-h/2} + O(T^{-3/2}).$$

If we put $p = |r_2|^{1/2}$ then $p = O(T^{-1/2})$. If we now define sign(a) = 1 if a > 0, sign(a) = 0 if a = 0, and sign(a) = -1 if a < 0 then :

$$\text{sign}(r_2)E_\delta[\exp\{it(\delta+|r_2|^{1/2})\}+\exp\{it(\delta-|r_2|^{1/2})\}-2\exp\{it\delta\}]$$
$$= \ r_2(it)^2(1-2it)^{-h/2} + O(T^{-3/2}).$$

Combining this together with the expression for $r_1(it)(1-2it)^{-h/2}$ we obtain Equation (44).

After substituting (44) into (42) it is then easy to apply an inverse Fourier transform to the latter and so obtain an expansion for the probability density function of $s^2$. This can be integrated to provide an expansion for the cumulative distribution function of $s^2$:

$$\Pr\{s^2 \leq x\} = (1+r_o)\xi_h^2(x) + \xi_h^2\left(x + \frac{r_1}{2}\right) - \xi_h^2\left(x - \frac{r_1}{2}\right)$$
$$+ \operatorname{sign}(r_2)\left[\xi_h^2(x + |r_2|^{\frac{1}{2}}) - \xi_h^2(x) + \xi_h^2(x - |r_2|^{\frac{1}{2}}) - \xi_h^2(x)\right]$$
$$+ r_3\xi_{h+2}^2(x) + r_4\xi_{h+4}^2(x) + r_5\xi_{h+6}^2(x) + O(T^{-3/2}).$$

When h=3,4 then (44) can be simplified:

$$[r_1(it) + r_2(it)^2](1-2it)^{-h/2}$$
$$= -(\tfrac{1}{2})(r_1)(1-2it)^{-(h/2)/2} + (\tfrac{1}{2})(r_1)(1-2it)^{-h/2}$$
$$-(\tfrac{1}{2})(r_2)(it)(1-2it)^{-(h-2)/2} - (\tfrac{1}{4})(r_2)(1-2it)^{-(h-2)/2}$$
$$+ (\tfrac{1}{4})(r_2)(1-2it)^{-(h/2)}$$

so that the problem now only arises from $(\tfrac{1}{2})(r_2)(it)(1-2it)^{-(h-2)/2}$. In this case we can approximate the cumulative distribution function of $s^2$ by:

$$\Pr\{s^2 \leq x\} = -[(\tfrac{1}{2})(r_1) + (\tfrac{1}{4})(r_2)]\xi_{h-2}^2(x)$$
$$+ [1+(r_o)+(\tfrac{1}{2})(r_1) + (\tfrac{1}{4})(r_2)]\xi_h^2(x) - \xi_h^2\left(x + \frac{r_2}{4}\right) + \xi_h^2\left(x - \frac{r_2}{4}\right)$$
$$+ (r_3)\xi_{h+2}^2(x) + (r_4)\xi_{h+4}^2(x) + (r_5)\xi_{h+6}^2(x) + O(T^{-3/2}).$$

When $h \geq 5$ we do not need to use location shifts so the expansion for the cumulative distribution function of $s^2$ is:

$$\Pr(s^2 \leq x) = (\tfrac{1}{4})(r_2)\xi^2_{h-4}(x) - [(\tfrac{1}{2})(r_1) + (\tfrac{1}{2})(r_2)]\xi^2_{h-2}(x)$$

$$+ [1+(r_o) + (\tfrac{1}{2})(r_1) + (\tfrac{1}{4})(r_2)]\xi^2_{h}(x) + (r_3)\xi^2_{h+2}(x)$$

$$+ (r_4)\xi^2_{h+4}(x) + (r_5)\xi^2_{h+6}(x) + O(T^{-3/2}).$$

As discussed earlier it is possible to transform $\gamma$ to $\gamma^*$ (by a location shift involving polynomial in $\eta$) so that $\min(s^2|\eta)$ occurs at $\gamma^* = 0$. This is equivalent to transforming the $\{c^j_o\}$, $\{c^j_a\}$, $\{c^j_{ab}\}$ and $\{c^j_{abc}\}$ to zero in (41). Thus the problem of approximating the cumulative distribution of $s^2$ at the origin arises only if the $\{c^o_a\}$, $\{c^o_{ab}\}$, $\{c^o_{abc}\}$ or $\{c^o_{abcd}\}$ are not then equal to zero resulting in terms such as:

$$[\exp(it\gamma_j\gamma_j)][(it)(c^o_{ab}\eta_a\eta_a)],$$

in the expansion of $[\exp(its^2)]$. These then give rise to terms such as $[(it)(1-2it)^{-h/2}]$ in the expansion $\sigma(t)$ which then create problems if $h$ is not sufficiently large.

However, this problem does not usually occur with LR test criteria. This can be seen more clearly by examining the relationship between the LR test statistic and the Lagrange Multiplier (LM) test statistic. The LM statistic for the hypothesis $H_0:\phi(\theta) = 0$ in the model $y \sim f(y;\theta)$ is $(\frac{1}{T})\tilde{\lambda}'\tilde{F}\,\tilde{I}\,(\tilde{\theta})^{-1}\tilde{F}'\tilde{\lambda}$ where $\theta$ is a p-vector of parameters and $\phi(.)$ is an h-vector of restrictions ($h \leq p$). Here $\lambda$ are the Lagrange Multipliers obtained by finding a saddlepoint of:

$$\mathcal{L}(y;\theta,\lambda) = \ln[f(y;\theta)] - \lambda'\phi(\theta) = L(y;\theta) - \lambda'\phi(\theta),$$

$$\text{and } \tilde{F} = \frac{\partial\phi}{\partial\theta'}\Big|_{\tilde{\theta}}$$

where $\tilde{\theta}$ are the restricted Maximum Likelihood estimates (obtained from the saddlepoint solution of $\mathcal{L}(y;\theta,\lambda)$). $\tilde{I}(\tilde{\theta})$ is a consistent estimator of the information matrix:

$$I(\theta_o) = E(-\frac{1}{T}\frac{\partial^2 L}{\partial\theta\partial\theta'}\Big|_{\theta_o} ; \theta_o).$$

We can express the LM statistic as an exact inner product test criterion by defining $\tilde{\mu} = (T^{-\frac{1}{2}})[\tilde{F}\,\tilde{I}(\tilde{\theta})^{-1}\tilde{F}']^{\frac{1}{2}}\tilde{\lambda}$ so that $LM = \mu'\mu$. The Likelihood Ratio statistic, $LR = 2[L(y;\hat{\theta}) - L(y;\tilde{\theta})]$, (where $\hat{\theta}$ are the unrestricted Maximum Likelihood estimates) can then be expanded in the form of (41) with $\gamma = \mu$. However, when $\mu = 0$ then $\tilde{\lambda} = 0$ and under standard assumptions this implies that the restricted and unrestricted Maximum Likelihood estimates are equal and so the LR statistic takes the value zero. Since the LR statistic is always non-negative and at $\mu = 0$ it is necessarily equal to zero, therefore its minimum given any value of $\eta$ must also be equal to zero. Thus in Equation (41) we have:

$$c^o_o = c^o_a = c^o_{ab} = c^o_{abc} = c^o_{abcd} = c^j_o = c^j_a = c^j_{ab} = c^j_{abc} = 0.$$

Consequently, the Likelihood Ratio statistic can be expanded as:

$$\begin{aligned} LR = {} & (\gamma_j\gamma_j) + (c^{jk}_o\gamma_j\gamma_k) + (c^{jk}_a\eta_a\gamma_j\gamma_k) + (c^{jk}_{ab}\eta_a\eta_b\gamma_j\gamma_k) \\ & + (c^{jk\ell}_o\gamma_j\gamma_k\gamma_\ell) + (c^{jk\ell}_a\eta_a\gamma_j\gamma_k\gamma_\ell) + (c^{jk\ell m}_o\gamma_j\gamma_k\gamma_\ell\gamma_m) \\ & + O_p(T^{-3/2}). \end{aligned} \tag{45}$$

Then exp{it(LR)} can be expanded (to correspond to (31)):

$$\begin{aligned} \exp\{it(LR)\} = {} & \exp\{it\gamma_j\gamma_j\}[1 + (it)(c^{jk}_o\gamma_j\gamma_k) + (it)(c^{jk}_a\eta_a\gamma_j\gamma_k) \\ & + (it)(c^{jk}_{ab}\eta_a\eta_b\gamma_j\gamma_k) + (it)(c^{jk\ell}_o\gamma_j\gamma_k\gamma_\ell) + (it)(c^{jk\ell}_a\eta_a\gamma_j\gamma_k\gamma_\ell) \\ & + (it)(c^{jk\ell m}_o\gamma_j\gamma_k\gamma_\ell\gamma_m) + (it)^2(\tfrac{1}{2}c^{jk}_o c^{\ell m}_o\gamma_j\gamma_k\gamma_\ell\gamma_m) + \end{aligned}$$

$$+ (it)^2(c_o^{jk}c_a^{\ell m}n_a\gamma_j\gamma_k\gamma_\ell\gamma_m) + (it)^2(c_o^{jk}c_{ab}^{\ell m}n_a n_b\gamma_j\gamma_k\gamma_\ell\gamma_m)$$

$$+ (it)^2(c_o^{jk}c_o^{\ell mp}\gamma_j \cdots \gamma_p) + (it)^2(c_o^{jk}c_a^{\ell mp}n_a\gamma_j \cdots \gamma_p)$$

$$+ (it)^2(\tfrac{1}{2}c_a^{jk}c_b^{\ell m}n_a n_b\gamma_j \cdots \gamma_m) + (it)^2(c_a^{jk}c_o^{\ell mp}n_a\gamma_j \cdots \gamma_p)$$

$$+ (it)^2(\tfrac{1}{2}c_o^{jk\ell}c_o^{mpq}\gamma_j \cdots \gamma_q)] + 0_p(T^{-3/2}).$$

In the above every term involving (it) is multiplied by at least two elements of $\gamma$ and every term involving $(it)^2$ is multiplied by at least four elements of $\gamma$. All such terms give rise to expressions of the form:

$$(it)(1-2it)^{-(h+2+2k)/2} \quad \text{or}$$

$$(it)^2(1-2it)^{-(h+4+2k)} \quad \text{where } k = 0,1,2,\ldots \; .$$

All of these expressions can be expanded in terms of the characteristic functions of central chi-square variates with at least h degrees of freedom as in (26); however, in the resulting expansion of $\sigma(t)$, the coefficient on $(1-2it)^{-h/2}$ is $[1 + O(T^{-\frac{1}{2}})]$ rather than $[1 + O(T^{-1})]$ and that on $(1-2it)^{-(h+2)/2}$ is $O(T^{-\frac{1}{2}})$ rather than $O(T^{-1})$.

Barndorff-Nielsen and Cox (1984) show that for the LR statistic, the chi-square cdf is correct to an error of $O(T^{-1})$ in many cases. In such cases, they show in their Equation (2.7) that the only quadratic term in the $\gamma$ entering Equation (45) is the $(\gamma_j\gamma_j)$ term, i.e:

$$LR = (\gamma_j\gamma_j) + (c_o^{jk\ell}\gamma_j\gamma_k\gamma_\ell) + (c_a^{jk\ell}\gamma_j\gamma_k\gamma_\ell\gamma_a) + (c_o^{jk\ell m}\gamma_j\gamma_k\gamma_\ell\gamma_m)$$
$$+ O(T^{-3/2})$$

when the notation of Barndorff-Nielsen and Cox (1984) is translated into the terms used here. In the above only the $\{c_o^{jk\ell}\}$ are $O(T^{-\frac{1}{2}})$.

Expanding [exp{it(LR)}] then gives:

$$\exp\{it(LR)\} = \exp\{it\gamma_j\gamma_j\}[1 + (it)(c_o^{jk\ell}\gamma_j\gamma_k\gamma_\ell) + (it)(c_a^{jk\ell}\gamma_j\gamma_k\gamma_\ell\gamma_a) + (it)(c_o^{jk\ell m}\gamma_j \ldots \gamma_m) + \tfrac{1}{2}(it)(c_o^{jk\ell}c_o^{mpq}\gamma_j \ldots \gamma_q)] + O(T^{-3/2}).$$

When this is convoluted with the multivariate Edgeworth expansion for the joint distribution of $(\gamma,n)$ then all the $O(T^{-\frac{1}{2}})$ terms involve odd powers in $(\gamma,n)$ and so integrate to zero. Thus the resulting expansions for the probability density function and cumulative distribution function for LR consist of the asymptotic distribution term plus terms of $O(T^{-1})$.

The difficulty with the approach which Mauleon-Torres (1983) adopts can be regarded as arising at a more fundamental level from the failure of sufficient inverse moments to exist for chi-square variates with low numbers of degrees of freedom. This is because the derivative of a chi-square pdf with h degrees of freedom can be expressed as a rational function multiplied by the chi-square pdf with h degrees of freedom and therefore the inverse moments of chi-square pdf's will arise naturally from the Fourier transforms of derivarives of chi-square pdf's. The failure of such inverse moments to exist implies that the appropriate Fourier transforms of the derivatives of chi-square pdf's are not properly defined. However in the case of interest here, this does not create a problem except when the support of the statistic does not consist of $[0,\infty)$ and when h is small.

## 3.4 The Joint Distribution Case

The main case considered here is that of asymptotically independent chi-square statistics. In this case the asymptotic joint distribution is easy to obtain and manipulate since it is the product of the asymptotic marginal distributions. The case of asymptotically jointly dependent chi-square statistics is much more difficult because it is harder to manipulate the asymptotic joint distribution. Furthermore, the asymptotically independent case can be covered using the techniques developed in Sections 3.2 and 3.3 whereas the asymptotically jointly dependent case would require new techniques.

### 3.4.1 The Jointly Independent Case

As in the univariate case there are two sub-cases: firstly, where the test criteria are exact inner product statistics; and secondly, where they are approximate inner product statistics. The exact inner products sub-case follows very directly from Section 3.2 and as before provides a leading example of the approximate inner products sub-case. In the exact inner products case it is generally possible to decompose the $\{s_\ell^2\}$ as $s_\ell^2 = \gamma_j^\ell \gamma_j^\ell$ where $\ell = 1, \ldots, k$ and where summation is over $j = 1, \ldots, h_\ell$ and where

$$\gamma' = (\gamma_1^1, \ldots, \gamma_{h_k}^k)$$

is asymptotically distributed as $N(0, I_h)$ where $h = \sum_{\ell=1}^{k} h_\ell$. Then the characteristic function of the $\{s_\ell^2\}$ is:

$$\sigma(t^1, \ldots, t^k) = E[\exp(i \sum_{\ell=1}^{k} t^\ell (\gamma_j^\ell \gamma_j^\ell))].$$

However, this is just:

$$E[\exp(i \sum_{p=1}^{h} t_p (\gamma_p)^2)]$$

with appropriate subsets of the $\{t_p\}$ set to equal the various elements of $\{t^\ell\}$. Therefore the methods of Section 3.2.2 can be used to approximate the joint characteristic function of the $\{(\gamma_p)^2; p = 1,\ldots,h\}$, producing an expansion of the form in Equation (22). Rather than setting $t_p = t$ $(p = 1,\ldots,h)$ as in Section 3.2.3 and so obtaining the approximate characteristic function of the sum of squares, we set various sub-groups of the $\{t_p\}$ equal to each other. The expansion for the joint characteristic function of the sums of squares is then:

$$\begin{aligned}\phi_T(t^1, \ldots, t^k) = \{\prod_{r=1}^{k}(1-2it^r)^{-h_r/2}\}\{[1 + a_0]\\ + [\sum_{r=1}^{k} a_r(1-2it^r)^{-\frac{1}{2}}] + [\sum_{r=1}^{k}\sum_{s=1}^{k} a_{rs}(1-2it^r)^{-\frac{1}{2}}(1-2it^s)^{-\frac{1}{2}}]\\ +[\sum_{r=1}^{k}\sum_{s=1}^{k}\sum_{t=1}^{k} a_{rst}(1-2it^r)^{-\frac{1}{2}}(1-2it^s)^{-\frac{1}{2}}(1-2it^t)^{-\frac{1}{2}}]\} + O(T^{-3/2}) \quad (46)\end{aligned}$$

where the $\{a\}$ are all of $O(T^{-1})$, expressed in a similar form to that of Equation (26). The corresponding expansion for the pdf of $\{s_\ell^2\}$ is then:

$$\text{p.d.f.}(s_1^2,\dots,s_k^2) = \{\prod_{\ell=1}^{k} \chi^2_{h_\ell}(s_\ell^2)\}(1 + b_o)$$

$$+ \{\sum_{r=1}^{k} b_r\left[\prod_{\ell\neq r}^{k} \chi^2_{h_\ell}(s_\ell^2)\right]\chi^2_{h_r+2}(s_r^2)\}$$

$$+ \{\sum_{r<s}^{k}\sum^{k} b_{rs}\left[\prod_{\ell\neq r,s}^{k} \chi^2_{h_\ell}(s_\ell^2)\right]\chi^2_{h_r+2}(s_r^2)\chi^2_{h_s+2}(s_s^2)\}$$

$$+ \{\sum_{r}^{k} b_{rr}\left[\prod_{\ell\neq r}^{k} \chi^2_{h_\ell}(s_\ell^2)\right]\chi^2_{h_r+4}(s_r^2)\}$$

$$+ \{\sum_{r<s<t}^{k}\sum^{k}\sum^{k} b_{rst}\left[\prod_{\ell\neq r,s,t}^{k} \chi^2_{h_\ell}(s_\ell^2)\right]\chi^2_{h_r+2}(s_r^2)\chi^2_{h_s+2}(s_s^2)\chi^2_{h_t+2}(s_t^2)\}$$

$$+ \{\sum_{r<s}^{k}\sum^{k} b_{rrs}\left[\prod_{\ell\neq r,s}^{k} \chi^2_{h_\ell}(s_\ell^2)\right]\chi^2_{h_r+4}(s_r^2)\chi^2_{h_s+2}(s_s^2)\}$$

$$+ \{\sum_{r<s}^{k}\sum^{k} b_{rss}\left[\prod_{\ell\neq r,s}^{k} \chi^2_{h_\ell}(s_\ell^2)\right]\chi^2_{h_r+2}(s_r^2)\chi^2_{h_s+4}(s_s^2)\}$$

$$+ \{\sum_{r}^{k} b_{rrr}\left[\prod_{\ell\neq r}^{k} \chi^2_{h_\ell}(s_\ell^2)\right]\chi^2_{h_r+6}(s_r^2)\} + O(T^{-3/2}) \qquad (47)$$

where the $\{b\}$ are all of $O(T^{-1})$ and where $\chi^2_h(x) = [x^{(h/2)-1}e^{-(x/2)}]/[2^{(h/2)}\Gamma(h/2)]$, (the pdf of a central chi-square with h degrees of freedom). This Equation resembles Equation (27) which gives the expansion for the pdf of an exact sum of squares. Lastly, (47) can be integrated with respect to the $\{s_\ell^2\}$ to obtain the expansion for the cdf of the $\{s_\ell^2\}$:

$$
\begin{aligned}
&\Pr\{s_\ell^2 \le x_\ell^2 \; ; \; \ell = 1,\ldots,k\} \\
&= \Big(\prod_{\ell=1}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big)(1 + b_o) + \Big(\sum_{r=1}^{k} b_r \Big[\prod_{\ell\neq r}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big] \xi^2_{h_r+2}(x_r^2)\Big) \\
&+ \Big(\sum_{r<s}^{k}\sum^{k} b_{rs} \Big[\prod_{\ell\neq r,s}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big] \xi^2_{h_r+2}(x_r)\, \xi^2_{h_s+2}(x_s)\Big) \\
&+ \Big(\sum_{r}^{k} b_{rr} \Big[\prod_{\ell\neq r}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big] \xi^2_{h_r+4}(x_r^2)\Big) \\
&+ \Big(\sum_{r<s<t}^{k}\sum^{k}\sum^{k} b_{rst} \Big[\prod_{\ell\neq r,s,t}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big] \xi^2_{h_r+2}(x_r^2)\, \xi^2_{h_s+2}(x_s^2)\, \xi^2_{h_t+2}(x_t^2)\Big) \\
&+ \Big(\sum_{r<s}^{k}\sum^{k} b_{rrs} \Big[\prod_{\ell\neq r,s}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big] \xi^2_{h_r+4}(x_r^2)\, \xi^2_{h_s+2}(x_s^2)\Big) \\
&+ \Big(\sum_{r<s}^{k}\sum^{k} b_{rss} \Big[\prod_{\ell\neq r,s}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big] \xi^2_{h_r+2}(x_r^2)\, \xi^2_{h_s+4}(x_s^2)\Big) \\
&+ \Big(\sum_{r}^{k} b_{rrr} \Big[\prod_{\ell\neq r}^{k} \xi^2_{h_\ell}(x_\ell^2)\Big] \xi^2_{h_r+6}(x_r^2)\Big) + O(T^{-3/2}) \qquad (48)
\end{aligned}
$$

where the $\{b\}$ are the same coefficients as in Equation (47) and $\xi^2_h(x^2)$ is the cdf of a central chi-square with $h$ degrees of freedom: this equation resembles Equation (28).

By letting $x_\ell^2 \to \infty$ for $\ell = g + 1,\ldots,k$ in Equation (48) it is possible to obtain the expansion for the joint cdf of $\{s_r^2\}$ $(r = 1,\ldots,g)$. An expansion for the cdf of $\{s_j^2\}$ $(j = g+1,\ldots,k)$ conditional on $\{s_r^2 \le x_r^2\}$ $(r = 1,\ldots,g)$ can then be constructed using Equation (48) and the formula:

$$\Pr\{s_j^2 \leq x_j^2\ (j=g+1,\ldots,k) \mid s_r^2 \leq x_r^2 (r=1,\ldots,g)\}$$

$$= \frac{\Pr\{s_\ell^2 \leq x_\ell^2\ (\ell=1,\ldots,k)\}}{\Pr\{s_r^2 \leq x_r^2\ (r=1,\ldots,g)\}} .$$

Since $\Pr\{s_r^2 \leq x_r^2\ (r=1,\ldots,g)\} = \left[\prod_{r=1}^{g} \xi_{h_r}^2 (x_r^2)\right] + \text{terms } O(T^{-1})$
$+ O(T^{-3/2})$

therefore $\left[\Pr\{s_r^2 \leq x_r^2\ (r=1,\ldots,g)\}\right]^{-1} = \left[\prod_{r=1}^{g} \xi_{h_r}^2 (x_r^2)\right]^{-1}\left[1 + \text{terms } O(T^{-1})\right]$
$+ O(T^{-3/2})$.

Thus the conditional probability that $\{s_j^2 \leq x_j^2; j=g+1,\ldots,k\}$ given $\{s_r^2 \leq x_r^2;\ r = 1,\ldots g\}$ can be expanded as the asymptotic distribution term $\left[\prod_{j=g+1}^{k} \xi_{h_j}^2 (x_j^2)\right]$ which is O(1) plus a set of terms which are $O(T^{-1})$ with an error of $O(T^{-3/2})$.

The approximate inner products sub-case is rather more complicated than the exact inner product sub-case. We will assume that all the statistics can be expanded in the form of Equation (29) so that:

$$\exp\{i \sum_{r=1}^{k} t^r s_r^2\} = \exp\{i\sum_{r=1}^{k} t^r \Big[ (\lambda_j^r \gamma_j^r) + (b_o^r) + (b_j^r \lambda_j) + (b_{jk}^r \lambda_j \lambda_k)$$
$$+ (b_{jk\ell}^r \lambda_j \lambda_k \lambda_\ell) + (b_{jk\ell m}^r \lambda_j \lambda_k \lambda_\ell \lambda_m)\Big]\} + O_p(T^{-3/2})$$
$$= \exp\{i \sum_{t=1}^{k} t^r (\lambda_j^r \lambda_j^r)\}\Big[1 + \sum_{r=1}^{k} (it^r)\{(b_o^r) + (b_j^r \lambda_j)$$
$$+ (b_{jk}^r \lambda_j \lambda_k) + (b_{jk\ell}^r \lambda_j \lambda_k \lambda_\ell) + (b_{jk\ell m}^r \lambda_j \lambda_k \lambda_\ell \lambda_m)\}$$
$$+ \sum_{r=1}^{k} \sum_{s=1}^{k} (it^r)(it^s)\{\tfrac{1}{2}(b_o^r)(b_o^s) + (b_o^r b_j^s \lambda_j) + (b_o^r b_{jk}^s \lambda_j \lambda_k)$$
$$+ (b_o^r b_{jk\ell}^s \lambda_j \lambda_k \lambda_\ell) + \tfrac{1}{2}(b_j^r b_k^s \lambda_j \lambda_k) + (b_j^r b_{k\ell}^s \lambda_j \lambda_k \lambda_\ell)$$
$$+ (b_j^r b_{k\ell m}^s \lambda_j \lambda_k \lambda_\ell \lambda_m) + \tfrac{1}{2}(b_{jk}^r b_{\ell m}^s \lambda_j \lambda_k \lambda_\ell \lambda_m) + (b_{jk}^r b_{\ell mp}^s \lambda_j \lambda_k \lambda_\ell \lambda_m \lambda_p)$$
$$+ \tfrac{1}{2}(b_{jk\ell}^r\ b_{mpq}^s \lambda_j \cdots \lambda_q)\}\Big] + O_p(T^{-3/2}) \qquad (49)$$

where $\lambda$ is an n-vector ($h = \sum_{r=1}^{k} h_r \leq n$) which includes all the $\{\lambda_j^r\}$ and which is asymptotically distributed as $N(0,I_n)$. In (49) all the {b} coefficients are $O(T^{-\frac{1}{2}})$ except the $\{b_{jk\ell m}^r\}$ which are $O(T^{-1})$ (as in Equation (29)). Equation (49) resembles Equation (31) and can be rewritten:

$$\exp(i \sum_{r=1}^{k} t^r s_r^2) = \exp(i \sum_{r=1}^{k} t^r(\lambda_j^r \lambda_j^r))[1+(it^r)(c_o^r) + (it^r)(c_j^r\lambda_j)$$

$$+ (it^r)(c_{jk}^r\lambda_j\lambda_k) + (it^r)(c_{jk\ell}^r\lambda_j\lambda_k\lambda_\ell) + (it^r)(c_{jk\ell m}^r\lambda_j\lambda_k\lambda_\ell\lambda_m)$$

$$+ (it^r)(it^s)(c_o^{rs}) + (it^r)(it^s)(c_j^{rs}\lambda_j) + (it^r)(it^s)(c_{jk}^{rs}\lambda_j\lambda_k)$$

$$+ (it^r)(it^s)(c_{jk\ell}^{rs}\lambda_j\lambda_k\lambda_\ell) + (it^r)(it^s)(c_{jk\ell m}^{rs}\lambda_j\lambda_k\lambda_\ell\lambda_m)$$

$$+ (it^r)(it^s)(c_{jk\ell mp}^{rs}\lambda_j \cdots \lambda_p) + (it^r)(it^s)(c_{jk\ell mpq}^{rs}\lambda_j \cdots \lambda_q)]$$

$$+ O_p(T^{-3/2})$$

$$= [\exp(i \sum_{r=1}^{k} t^r(\lambda_j^r\lambda_j^r)] \; [P_T(\lambda, i\underline{t})] + O_p(T^{-3/2})$$

where
$$c_o^r = b_o^r = O(T^{-\frac{1}{2}})$$
$$c_j^r = b_j^r = O(T^{-\frac{1}{2}})$$
$$c_{jk}^r = b_{jk}^r = O(T^{-\frac{1}{2}})$$
$$c_{jk\ell}^r = b_{jk\ell}^r = O(T^{-\frac{1}{2}})$$
$$c_{jk\ell m}^r = b_{jk\ell m}^r = O(T^{-1})$$
$$c_o^r = \tfrac{1}{2}b_o^r b_o^s = O(T^{-1})$$
$$c_j^{rs} = b_o^r b_j^s = O(T^{-1})$$

$$c^{rs}_{jk} = b^r_o b^s_{jk} + \tfrac{1}{2} b^r_j b^s_k = O(T^{-1})$$
$$c^{rs}_{jk\ell} = b^r_o b^s_{jk\ell} + b^r_o b^s_{k\ell} = O(T^{-1})$$
$$c^{rs}_{jk\ell m} = b^r_j b^s_{k\ell m} + \tfrac{1}{2} b^r_{jk} b^s_{\ell m} = O(T^{-1})$$
$$c^{rs}_{jk\ell mp} = b^r_{jk} b^s_{\ell mp} = O(T^{-1})$$
$$c^{rs}_{jk\ell mpq} = \tfrac{1}{2} b^r_{jk\ell} b^s_{mpq} = O(T^{-1}). \qquad (50)$$

As before a multivariate Edgeworth expansion for the pdf of the $\lambda$ can be found using the methods in Section 3.2.1 and takes the same form as Equation (32). Thus the joint characteristic function of the $\{s^2_r\}$ can be obtained by combining Equations (32) and (50) in the same way that Equations (31) and (32) are combined to produce Equations (33) and (34). Thus:

$$\sigma(t^1,\ldots,t^k) = E\left[\exp\{i \sum_{r=1}^{k} t^r s^2_r\}\right]$$
$$= \int \exp\{i \sum_{t=1}^{k} t^r(\lambda^r_j \lambda^r_j)\} \mathrm{H}(\lambda) P_T(\lambda, i\underline{t}) Q_T(\lambda) d\lambda + O(T^{-3/2})$$
$$= \int \exp\{ \sum_{r=1}^{k} t^r(\lambda^r_j \lambda^r_j)\} \mathrm{H}(\lambda) R_T(\lambda, i\underline{t}) d\lambda$$

where $R_T(\lambda, i\underline{t}) = 1 + (e_o) + (e_j\lambda_j) + (e_{jk}\lambda_j\lambda_k)$
$+ (e_{jk\ell}\lambda_j\lambda_k\lambda_\ell) + (e_{jk\ell m}\lambda_j \cdots \lambda_m)$
$+ (e_{jk\ell mp}\lambda_j \cdots \lambda_p) + (e_{jk\ell mpq}\lambda_j \cdots \lambda_q)$

with $e_o = (e^o_o) + (it^r)(e^r_o) + (it^r)(it^s)(e^{rs}_o)$ and similarly for $(e_j),\ldots,(e_{jk\ell mpq})$ ;

and where $e_o^o = d_o = O(T^{-1})$; $e_j^o = d_j = O(T^{-\frac{1}{2}})$;

$$e_{jk}^o = d_{jk} = O(T^{-1});\ e_{jk\ell}^o = d_{jk\ell} = O(T^{-\frac{1}{2}});$$
$$e_{jk\ell m}^o = d_{jk\ell m} = O(T^{-1});\ e_{jk\ell mp}^o = 0;$$
$$e_{jk\ell mpq}^o = d_{jk\ell mpq} = O(T^{-1});$$

$$e_o^r = c_o^r = O(T^{-\frac{1}{2}});\ e_j^r = c_j^r + c_o^r d_j = O(T^{-\frac{1}{2}});$$
$$e_{jk}^r = c_{jk}^r + d_j c_k^r = O(T^{-\frac{1}{2}});$$
$$e_{jk\ell}^r = c_{jk\ell}^r + d_j c_{k\ell}^r + c_o^r d_{jk\ell} = O(T^{-\frac{1}{2}});$$
$$e_{jk\ell m}^r = c_{jk\ell m}^r + d_j c_{k\ell m}^r + d_{jk\ell} c_m^r = O(T^{-1});$$
$$e_{jk\ell mp}^r = d_{jk\ell} c_{mp}^r = O(T^{-1});$$
$$e_{jk\ell mpq}^r = d_{jk\ell} c_{mpq}^r = O(T^{-1});$$

and all the $\{e^{rs}\}$ coefficients are $O(T^{-1})$. (51)

As before the odd powers of $\lambda$ vanish and we can integrate out directly those elements of $\lambda$ which do not enter the $\exp\{i \sum_{r=1}^{k} t^r(\lambda_j^r \lambda_j^r)\}$ term. Defining those elements of $\lambda$ which do enter this term as $\gamma$ (an h-vector) then:

$$\sigma(t^1,\dots,t^k) = \int \exp\{i \sum_{r=1}^{k} t^r(\gamma_j^r \gamma_j^r)\} H(\gamma) S_T(\gamma, i\underline{t}) d\gamma + O(T^{-3/2})$$

where $S_T(\gamma, i\underline{t}) = 1 + (t_o) + (g_j \gamma_j^2) + (g_{jk}\gamma_j^2\gamma_k^2) + g_{jk\ell}\gamma_j^2\gamma_k^2\gamma_\ell^2)$

with
$$g_o = (g_o^o) + (it^r)(g_o^r) + (it^r)(it^s)(g_o^{rs}) = O(T^{-\frac{1}{2}})$$
$$g_j = (g_j^o) + (it^r)(g_j^r) + (it^r)(it^s)(g_j^{rs}) = O(T^{-\frac{1}{2}})$$
$$g_{jk} = (g_{jk}^o) + (it^r)(g_{jk}^r) + (it^r)(it^s)(g_{jk}^{rs}) = O(T^{-1})$$
$$g_{jk\ell} = (g_{jk\ell}^o) + (it^r)(g_{jk\ell}^r) + (it^r)(it^s)(g_{jk\ell}^{rs}) = O(T^{-1}) \quad (52)$$

where as in Equation (37) all the {g} coefficients are $O(T^{-1})$ except $\{g_o^r\}$ and $\{g_j^r\}$ which are $O(T^{-\frac{1}{2}})$. The typical term in this expansion is then:

$$\int \exp\{i \sum_{r=1}^{k} t^r(\gamma_j^r \gamma_j^{r})\} \ (\gamma)\gamma_j^{2p_j}\gamma_k^{2p_k}\gamma_\ell^{2p_\ell}d\gamma; \qquad (53)$$

$$j,k,\ell = 1,\dots,h;\ p_j, p_k, p_\ell \geq 0;\ p_j + p_k + p_\ell = 0,1,2,3$$

multiplied by a quadratic in $(i\underline{t})$. Since the quadratic in $(i\underline{t})$ does not affect the integration with respect to $\gamma$ we can temporarily ignore it and consider the term:

$$\int \exp\{it_j\gamma_j^2\} i(\gamma)\gamma_j^{2p_j}\gamma_k^{2p_k}\gamma_\ell^{2p_\ell}d\gamma$$
$$= \left[\prod_{b\neq j,k,\ell}(1-2it_b)^{-\frac{1}{2}}\right]\left[\prod_{a=j,k,\ell}(1-2it_a)^{-(\frac{1}{2}+p_a)}\left\{\frac{(2p_a)!}{2^{(p_a)}(p_a)!}\right\}\right]$$

following Equation (21). Then setting the appropriate subsets of $\{t_j\}$ equal to the corresponding $\{t^r\}$ evaluates the integral in Equation (53). Multiplying this by the appropriate quadratic in $(i\underline{t})$ and summing up all the terms in Equation (52) gives an expansion of the form:

$$\sigma(t^1,\dots,t^k) = \{\prod_{u=1}^{k}(1-2it^u)^{-h_u/2}\}\{[1+s_o]+[\sum_{u=1}^{k}s_u(1-2it^u)^{-\frac{1}{2}}]$$
$$+[\sum_{u=1}^{k}\sum_{v=1}^{k}s_{uv}(1-2it^u)^{-\frac{1}{2}}(1-2it^v)^{-\frac{1}{2}}]$$
$$+[\sum_{u=1}^{k}\sum_{v=1}^{k}\sum_{w=1}^{k}s_{uvw}(1-2it^u)^{-\frac{1}{2}}(1-2it^v)^{-\frac{1}{2}}(1-2it^w)^{-\frac{1}{2}}]\} + O(T^{-3/2})$$

where
$$s_o = (s_o^o) + (it^r)(s_o^r) + (it^r)(it^s)(s_o^{rs}) = O(T^{-\frac{1}{2}});$$
$$s_u = (s_u^o) + (it^r)(s_u^r) + (it^r)(it^s)(s_u^{rs}) = O(T^{-\frac{1}{2}});$$
$$s_{uv} = (s_{uv}^o) + (it^r)(s_{uv}^r) + (it^r)(it^s)(s_{uv}^{rs}) = O(T^{-1});$$
$$s_{uvw} = (s_{uvw}^o) + (it^r)(s_{uvw}^r) + (it^r)(it^s)(s_{uvw}^{rs}) = O(T^{-1}); \qquad (54)$$

with all the {s} coefficients of $O(T^{-1})$ except $\{s_o^r\}$ and $\{s_u^r\}$ which are of $O(T^{-\frac{1}{2}})$. This is the approximate inner products sub-case equivalent of Equation (46) in the exact inner products sub-case (NB. the {s} coefficients should not be confused with the $\{s_\ell^2\}$ which are the statistics). Equation (54) can then be inverted to give the expansions for the probability density and cumulative distribution functions of the $\{s_\ell^2\}$statistics.

As in Section 3.3 problems arise if any of the $\{h_\ell\}$ are less than or equal to four. This problem again arises from the behaviour of the statistic near zero and will not occur with LR statistics. It can be solved in a similar way to that discussed in Section 3.3 but as there are k statistics to consider, it is naturally more complicated. When all the statistics have locally minimum values of zero or of $O(T^{-3/2})$ then the cdf expansion for the $\{s_\ell^2\}$ becomes identical to Equation (48) except that the $\{b_r\}$ and $b_o$ are of $O(T^{-\frac{1}{2}})$ and all the other {b} terms are of $O(T^{-1})$. In Equation (48) for the exact sums of squares, all the {b} coefficients are of $O(T^{-1})$.

It seems plausible that in the jointly independent LR statistics case the $O(T^{-\frac{1}{2}})$ terms all vanish providing that the statistics can all be expanded in the same $\lambda$ variables with each statistic in the form of Equation (45) thus extending the single LR statistic result of Barndorff-Nielsen and Cox (1984).

### 3.4.2 The Jointly Dependent Case

In this case it is impossible to express the statistics as exact or approximate inner products of disjoint sub-vectors from a vector of asymptotically independent standard normal variates. Two alternative approaches are available: firstly, the statistics can be expressed asymptotically or exactly as idempotent quadratic forms in a common set of asymptotically independent standard normal variates $\lambda$. So in the bivariate exact case:

$$\left.\begin{array}{l} s_1^2 = b^1_{jk}\lambda_j\lambda_k \\ s_2^2 = b^2_{jk}\lambda_j\lambda_k \end{array}\right\} \{b^1_{jk}\} \text{ and } \{b^2_{jk}\} \text{ are fixed.}$$

In this case Section 3.2.1 can be used to obtain a multivariate Edgeworth expansion for the pdf of $\lambda$ in the form $[\phi(\lambda)Q_T(\lambda) + O(T^{-3/2})]$ so:

$$\sigma(t^1,t^2) = \int \exp\{(it^1)(b^1_{jk}\lambda_j\lambda_k) + (it^2)(b^2_{jk}\lambda_j\lambda_k)\}\phi(\lambda)Q_T(\lambda)d\lambda$$
$$+ O(T^{-3/2}).$$

However $\int \exp\{(it^1)(b^1_{jk}\lambda_j\lambda_k) + (it^2)(b^2_{jk}\lambda_j\lambda_k)\}\phi(\lambda)d\lambda$

$$\neq (1-2it^1)^{-h_1/2}(1-2it^2)^{-h_2/2}$$

and differentiating with respect to $(t^r)$ produces

$$\int (ib^r_{jk}\lambda_j\lambda_k)\exp\{(it^1)(b^1_{jk}\lambda_j\lambda_k) + (it^2)(b^2_{jk}\lambda_j\lambda_k)\}\phi(\lambda)d\lambda.$$

The difficulty here is that the polynomial $Q_T(\lambda)$ has to be re-expressed in terms of sums and products of the $(b^r_{jk}\lambda_j\lambda_k)$ terms and it is not clear whether this can be done, and if so how.

The second approach is to express each statistic asymptotically or exactly as the inner product of asymptotically standard normal vectors but where these vectors are asymptotically jointly dependently normally distributed. It is still possible to obtain a multivariate Edgeworth expansion for the $\lambda$ although since $\lambda$ is an asymptotically jointly dependent normal vector the methods of Section 3.2.1 need to be modified slightly. Defining $i(\lambda,\Omega)$ as the asymptotic distribution function of $\lambda$ so $i(\lambda,\Omega) = (2\pi)^{-h/2}\exp\{-\frac{1}{2}\lambda'\Omega^{-1}\lambda\}$ then in the exact bivariate case:

$$s_1^2 = \lambda_j^1\lambda_j^1$$

$$s_2^2 = \lambda_j^2\lambda_j^2$$

and $\sigma_T(t_1,t_2) = E[\exp\{(it_1)(\lambda_j^1\lambda_j^1) + (it_2)(\lambda_j^2\lambda_j^2)\}]$

$$= \int\exp\{(it_1)(\lambda_j^1\lambda_j^1) + (it_2)(\lambda_j^2\lambda_j^2)\}i(\lambda,\Omega)Q_T^*(\lambda)d\lambda.$$

In this case, however, odd power terms in $\lambda$ such as:

$$\int\exp\{(it_1)(\lambda_j^1\lambda_j^1) + (it_2)(\lambda_j^2\lambda_j^2)\}i(\lambda,\Omega)\lambda_r^1\lambda_s^2d\lambda$$

do not in general equal zero. Therefore they do not vanish and hence constructing the non-vanishing terms of $Q_T^*(\lambda)$ from combinations of derivatives of:

$$\int\exp\{i\sum_{q=1}^{n} t_q(\lambda_q)^2\}i(\lambda,\Omega)d\lambda$$

with respect to the $\{t_q\}$ is not possible since these only produce even power terms in $\lambda$. These two approaches to expressing the statistic are essentially equivalent and the problems which arise from attempting to apply the methods of Section 3.2.2 are also essentially equivalent; in both cases the real problem is that we cannot expand the joint cdf in terms of the derivatives of the asymptotic cdf with respect to the $\{t_q\}$.

## 3.5 CDF Expansions under Local Alternative Hypothesis Sequences

So far we have only considered deriving expansions for the cdf's of test statistics under the null hypothesis, i.e. examining the size properties of tests. However, it is clearly desirable to analyse the power properties of tests in addition. There is little value in considering fixed alternatives since generally power against fixed alternatives is asymptotically unity and obtaining asymptotic expansions for power at fixed alternatives is very difficult. The usual practice is to consider local alternative hypothesis sequences and this is the approach we adopt here. Thus we test $H_o:\phi(\theta) = 0$ against $H_A:\phi(\theta) \neq 0$ and evaluate the power of the test at $\theta_T = \theta_o + T^{-\frac{1}{2}}\psi$ where $\phi(\theta_o) = 0$ and $\psi$ is a constant vector such that $\phi(\theta_T) = O(T^{-\frac{1}{2}}) \neq 0$.

In the exact sum of squares case $s^2 = (\lambda_j\lambda_j) = \sum_{j=1}^{h} \lambda_j^2$ and under a local alternative sequence then: $\lambda \overset{D}{\to} N(\alpha, I_h)$, i.e. the $\lambda$ are asymptotically independent unit variance but not necessarily zero mean. In fact $\alpha$ will depend upon $\psi$, $\theta_o$, any exogenous variables and the nature of the model; it will be invariant to any reformulation of the restrictions which maintains the null hypothesis, i.e. $\phi \to \phi^*$, such that $\phi^*(\theta) = 0$ if and only if $\phi(\theta) = 0$ provided rank $(\partial\phi/\partial\theta')$ = rank $(\partial\phi^*/\partial\theta') = h$ at $\theta_o$. (See Sargan (1984) for a discussion of the consequences for the COMFAC test of the failure of this assumption). A somewhat more general formulation of a local alternative sequence is:

$$H_A^*:\theta_T^* = \theta_o + \psi_T^* \quad \text{where } \psi_T^* = O(T^{-\frac{1}{2}}).$$

In this extended case $T^{\frac{1}{2}}\psi_T^*$ can be a vector of polynomials in $T^{-\frac{1}{2}}$. However, some information must be provided about the limit of $T^{\frac{1}{2}}\psi_T^*$ as $T \to \infty$ otherwise it is impossible to obtain $\alpha$ and the asymptotic power of the test (there is little value in considering an approximation involving terms of $O(T^{-\frac{1}{2}})$ and $O(T^{-1})$ when the $O(1)$ terms are not known). Given that local alternative sequences are artificial the easiest assumption to make is that $T^{\frac{1}{2}}\psi_T^*$ is a constant.

The only way in which $\psi_T^*$ enters the analysis is through its effects on the multivariate Edgeworth expansion for $\lambda$ which can be obtained using the methods of Section 3.2.1 but with a slight modification. As in Equation (1) we assume that $\lambda$ can be expressed:

$$\lambda = T^{\frac{1}{2}}\{\phi(p) - \phi(\mu)\}$$

where $\mu = E(p;\theta_o)$. However, under the local alternative hypothesis sequence $E(p;\theta_T^*) = \mu_T^* \neq \mu(\mu_T^* - \mu = O(T^{-\frac{1}{2}}))$ so we re-express $\lambda$:

$$\lambda = T^{\frac{1}{2}}\{\phi(p) - \phi(\mu_T^*)\} + T^{\frac{1}{2}}\{\phi(\mu_T^*) - \phi(\mu)\} = g(x) + \alpha_T^*$$

where $x = T^{\frac{1}{2}}(p-\mu_T^*)$ and $\alpha_T^* = \alpha + \beta_T$ with $\beta_T = O(T^{-\frac{1}{2}})$. The methods of Section 3.2.1 can be applied to obtain an asymptotic expansion for $\varepsilon = (\lambda-\alpha_T^*) = g(x)$ to obtain:

$$\text{p.d.f.}(\varepsilon) = [(2\pi)^{-h/2}\exp\{-\tfrac{1}{2}\varepsilon'\varepsilon\}][1 + (c_o) + (c_j\varepsilon_j)$$

$$+ (c_{jk}\varepsilon_j\varepsilon_k) + (c_{jk\ell}\varepsilon_j\varepsilon_k\varepsilon_\ell) + (c_{jk\ell m}\varepsilon_j\varepsilon_k\varepsilon_\ell\varepsilon_m) + (c_{jk\ell mp}\varepsilon_j\varepsilon_k\varepsilon_\ell\varepsilon_m\varepsilon_p)$$

$$+ (c_{jk\ell mpq}\varepsilon_j \ldots \varepsilon_q)] + O(T^{-3/2})$$

with $c_o = \text{=}(T^{-\frac{1}{2}})$; $c_j = O(T^{-\frac{1}{2}})$; $c_{jk} = O(T^{-\frac{1}{2}})$; $c_{jk\ell} = O(T^{-\frac{1}{2}})$

$$c_{jk\ell m} = O(T^{-1});\ c_{jk\ell mp} = O(T^{-1});\ c_{jk\ell mpq} = O(T^{-1}). \qquad (55)$$

Since $\alpha_T^*$ is non-stochastic and does not depend on $\lambda$ then the Jacobian of the transformation from $\varepsilon$ to $\lambda$ is unity. Also we have:

$$\exp\{-\tfrac{1}{2}\varepsilon'\varepsilon\} = \exp\{-\tfrac{1}{2}[(\lambda-\alpha) - \beta_T]'[(\lambda-\alpha) - \beta_T]\}$$

$$= \exp\{-\tfrac{1}{2}(\lambda-\alpha)'(\lambda-\alpha) + \beta_T'(\lambda-\alpha) - \tfrac{1}{2}\beta_T'\beta_T\}$$

$$= \exp\{-\tfrac{1}{2}(\lambda-\alpha)'(\lambda-\alpha)\}[1 + \beta_T'(\lambda-\alpha) - \tfrac{1}{2}\beta_T'\beta_T$$

$$+ \tfrac{1}{2}(\beta_T'(\lambda-\alpha))^2] + O(T^{-3/2}). \tag{56}$$

Equation (56) can then be used to expand Equation (55) to have a leading term which is a jointly independent standard normal density in $(\lambda-\alpha)$ (i.e. the asymptotic density function of $(\lambda-\alpha)$) multiplied by a polynomial in $\lambda$. Since the Jacobian of the transformation from $\varepsilon$ to $\lambda$ is unity then this gives a multivariate Edgeworth expansion for the joint pdf of $\lambda$:

$$\text{p.d.f.}\ (\lambda) = [(2\pi)^{-h/2}\exp\{-\tfrac{1}{2}(\lambda-\alpha)'(\lambda-\alpha)\}][1 + (d_0)$$

$$+ (d_j\lambda_j) + (d_{jk}\lambda_j\lambda_k) + (d_{jk\ell}\lambda_j\lambda_k\lambda_\ell) + (d_{jk\ell m}\lambda_j\lambda_k\lambda_\ell\lambda_m)$$

$$+ (d_{jk\ell mp}\lambda_j\lambda_k\lambda_\ell\lambda_m\lambda_p) + (d_{jk\ell mpq}\lambda_j\lambda_k\lambda_\ell\lambda_m\lambda_p\lambda_q)] + O(T^{-3/2})$$

where $\{d_o\}$, $\{d_j\}$, $\{d_{jk}\}$, $\{d_{jk\ell}\}$ are $O(T^{-\frac{1}{2}})$

$$\{d_{jk\ell m}\},\ \{d_{jk\ell mp}\},\ \{d_{jk\ell mpq}\} \text{ are } O(T^{-1}). \tag{57}$$

This differs slightly from Equation (14) in the orders of magnitude and by the inclusion of a fifth order term in $\lambda$; this arises because $\lambda$ has a limiting normal distribution with non-zero mean. With this multivariate Edgeworth expansion for the pdf of $\lambda$ we can now proceed to analyse the expansion for the cdf of $s^2$. As before the simplest case to consider is that of a single exact sum of squares test statistic. The characteristic function of the statistic is given as:

$$\begin{aligned}\sigma(t) = E\left[\exp\{its^2\}\right] &= \int \exp\{it\lambda_j\lambda_j\}(2\pi)^{-h/2}\exp\{-\tfrac{1}{2}(\lambda_j\lambda_j) \\ &+ (\lambda_j\alpha_j) - \tfrac{1}{2}(\alpha_j\alpha_j)\}\left[1 + (d_o) + (d_j\lambda_j) + (d_{jk}\lambda_j\lambda_k) + (d_{jk\ell}\lambda_j\lambda_k\lambda_\ell) \right.\\ &\left. + (d_{jk\ell m}\lambda_j\lambda_k\lambda_\ell\lambda_m) + (d_{jk\ell mp}\lambda_j\lambda_k\lambda_\ell\lambda_m\lambda_p) + (d_{jk\ell mpq}\lambda_j\lambda_k\lambda_\ell\lambda_m\lambda_p\lambda_q)\right] d\lambda \\ &+ O(T^{-3/2}). \end{aligned} \tag{58}$$

The general term is equal to a coefficient multiplied by:

$$\int\left(\prod_{j=1}^{h} \exp\{it\lambda_j^2\}(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\lambda_j^2 + \lambda_j\alpha_j - \tfrac{1}{2}\alpha_j^2\}\right)$$

$$\cdot\ \lambda_a^{p_a}\lambda_b^{p_b}\lambda_c^{p_c}\lambda_d^{p_d}\lambda_e^{p_e}\lambda_f^{p_f}d\lambda$$

$$= \left[\prod_{j\neq a,b,c,d,e,f} \int \exp\{it\lambda_r^2\}(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\lambda_j^2 + \alpha_j\lambda_j - \tfrac{1}{2}\alpha_j^2\}d\lambda_j\right]$$

$$\left[\prod_{r=a,b,c,d,e,f} \int \exp\{it\lambda_r^2\}(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\lambda_r^2\alpha_r\lambda_r - \tfrac{1}{2} + \alpha_r\lambda_r - \tfrac{1}{2}\alpha_r^2\}\lambda_r^{p_r}d\lambda_r\right]$$

where $p_a, p_b, p_c, p_d, p_e, p_f \geq 0$

$$p_a + p_b + p_c + p_d + p_e + p_f \leq 6. \tag{59}$$

Since the asymptotic distribution of $\lambda$ is normal with a non-zero mean the odd powers terms in $\lambda$, i.e. $\lambda_r^p$ where $p = 2k + 1$ $(k=0,1,2)$, do not vanish on integration. This is similar to the problem in the jointly dependent case where powers of $\lambda$ such as $\lambda_r\lambda_s$ $(r\neq s)$ do not vanish on integration. The solution to the problem here is to differentiate terms of the form:

$$\int \exp\{it\lambda_r^2\}(2\pi)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\lambda_r-\alpha_r)^2\}d\lambda_r \tag{60}$$

with respect to the $\{\alpha^r\}$ rather than with respect to t. Assuming that integration with respect to $\lambda_r$ and differentiation with respect to $\alpha_r$ can be interchanged then:

$$\frac{\partial}{\partial\alpha_r}\left[\exp\{-\tfrac{1}{2}(\lambda_r-\alpha_r)^2\}\right] = (\lambda_r-\alpha_r)\exp\{-\tfrac{1}{2}(\lambda_r-\alpha_r)\}.$$

This can be used to evaluate terms of the form of Equation (60):

$$\begin{aligned}&\int \exp\{it\lambda_r^2\}(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(\lambda_r-\alpha_r)^2\}\lambda_r d\lambda_r\\
&= \int \exp\{it\lambda_r^2\}(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(\lambda_r-\alpha_r)^2\}\alpha_r d\lambda_r\\
&+ \int \exp\{it\lambda_r^2\}(2\pi)^{-\frac{1}{2}} \frac{\partial}{\partial\alpha_r}\left[\exp\{-\tfrac{1}{2}(\lambda_r-\alpha_r)^2\}\right] d\lambda_r\\
&= \left[\alpha_r + \left(\frac{\partial}{\partial\alpha_r}\right)\right]\int \exp\{it\lambda_r^2\}(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(\lambda_r-\alpha_r)^2\}d\lambda_r\end{aligned}$$

using an operator notation. The characteristic function of a non-central chi-square with k degrees of freedom and non-centrality parameter $\gamma^2$ (when k=1 this is the characteristic function of the square of a normal variate with mean $\gamma$ and variance unity) is given as:

$$\begin{aligned}\Gamma_k^2(t;\gamma^2) &= \prod_{s=1}^{k} \int (2\pi)^{-\frac{1}{2}}\exp\{itx_s^2\}\exp\{-\tfrac{1}{2}(x_s-\gamma_s)^2\}dx_s\,, \quad \gamma^2 = \sum_{s=1}^{k} \gamma_s^2;\\
&= \sum_{j=0}^{\infty} \frac{\gamma^{2j}e^{-(\gamma^2/2)}}{2^j\, j!}(1-2it)^{-(k+2j)/2}\end{aligned}$$

(from Rao (1973), p.182). Differentiating this characteristic function with respect to $\gamma$ gives :

$$\frac{\partial}{\partial\gamma}\left[\Gamma_k^2(t;\gamma^2)\right] = \frac{\partial}{\partial\gamma}\left\{\sum_{j=o}^{\infty} \frac{\gamma^{2j}\, e^{-(\gamma^2/2)}}{2^j\, j!}(1-2it)^{-(k+2j)/2}\right\}$$

$$= \left\{\sum_{j=o}^{\infty} \frac{2j\gamma^{2j-1}e^{-(\gamma^2/2)}}{2^j\, j!}(1-2it)^{-(k+2j)/2}\right\}$$

$$- \left\{\gamma\sum_{j=o}^{\infty} \frac{\gamma^{2j}e^{-(\gamma^2/2)}}{2^j\, j!}(1-2it)^{-(k+2j)/2}\right\}$$

$$= \left\{\gamma\sum_{j=1}^{\infty} \frac{\gamma^{2(j-1)}e^{-(\gamma^2/2)}}{2^{j-1}(j-1)!}(1-2it)^{-(k+2(j-1)+2)/2}\right\}$$

$$- \left\{\gamma\sum_{j=o}^{\infty} \frac{\gamma^{2j}e^{-(\gamma^2/2)}}{2^j j!}(1-2it)^{-(k+2j)/2}\right\}$$

$$= \gamma\left[\Gamma_{k+2}^2(t;\gamma^2) - \Gamma_k^2(t;\gamma)\right].$$

It is now useful to define the integral

$$I_n = \int \exp\{itx^2\}(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}(x-\gamma)^2\}x^n dx.$$

Using this expression then

$$\frac{\partial I_n}{\partial\gamma} = \int \exp\{itx^2\}(2\pi)^{-\frac{1}{2}}\,\frac{\partial}{\partial\gamma}\left[\exp\{-\tfrac{1}{2}(x-\gamma)^2\right]\,x^n dx = I_{n+1}-\gamma I_n\ ,\ n \geq 0\ .$$

This provides an iterative relationship for the $\{I_n\}$:

$$I_{n+1} = \left[\gamma + \frac{\partial}{\partial\gamma}\right]I_n.$$

Since $I_0 = \Gamma_1^2(t;\gamma^2)$ repeated application of this result gives $I_n$ for any n; in particular for n=1,...,6:

$$I_0 = \Gamma_1^2(t;\gamma^2)$$

$$I_1 = \gamma\Gamma_3^2(t;\gamma^2)$$

$$I_2 = \gamma\Gamma_5^2(t;\gamma^2) + \Gamma_3^2(t;\gamma^2)$$

$$I_3 = \gamma^3\Gamma_7^2(t;\gamma^2) + 3\gamma\Gamma_5^2(t;\gamma^2)$$

$$I_4 = \gamma^4\Gamma_9^2(t;\gamma^2) + 6\gamma^2\Gamma_7^2(t;\gamma^2) + 3\Gamma_5^2(t;\gamma^2)$$

$$I_5 = \gamma^5\Gamma_{11}^2(t;\gamma^2) + 10\gamma^3\Gamma_9^2(t;\gamma^2) + 15\gamma\Gamma_7^2(t;\gamma^2)$$

$$I_6 = \gamma^6\Gamma_{13}^2(t;\gamma^2) + 15\gamma^4\Gamma_{11}^2(t;\gamma^2) + 45\gamma^2\Gamma_9^2(t;\gamma^2) + 15\Gamma_7^2(t;\gamma^2). \qquad (61)$$

It can be noted that when $\gamma=0$ these reduce exactly to the formulae obtained by differentiating $I_n$ with respect to t $(n=0,1,\ldots,6)$. Thus this procedure is consistent with the techniques developed in Sections 3.2 and 3.3. Using these $\{I_n\}$ integrals we can now evaluate the terms from (59) of the form:

$$\int\{\prod_{j=1}^{h}\exp(it\lambda_j^2)(2\pi)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}(\lambda_j-\alpha_j)^2)\}\lambda_a^{p_a}\ldots\lambda_f^{p_f}d\lambda. \qquad (62)$$

However, this of course depends on the precise values of $(p_a,\ldots,p_f)$. At this point it is advantageous to make use of the reproductive property of non-central chi-square variates. Suppose that $\delta_1^2$ is distributed as non-central chi-square with $n_1$ degrees of freedom and non-centrality parameter $\gamma_1^2$ and that $\delta_2^2$ is independent of $\delta_1^2$ and is distributed as non-central chi-square with $n_2$ degrees of freedom and non-centrality parameter $\gamma_2^2$. Then the variate formed by summing the two $\delta^2$'s is also non-central chi-square but with $(n_1+n_2)$ degrees of freedom and non-centrality parameter $(\gamma_1^2+\gamma_2^2)$, (see Rao (1973), p.182).

Since the characteristic function of the sum of two independent variates is the product of their individual characteristic functions with the same characteristic function parameter t, therefore:

$$\Gamma^2_{n_1}(t;\gamma_1^2)\Gamma^2_{n_2}(t;\gamma_2^2) = \Gamma^2_{n_1+n_2}(t;\gamma_1^2 + \gamma_2^2). \qquad (63)$$

Equation (62) can be expanded using this result and Equation (61) so that:

$$\int\{\prod_{j=1}^{h} \exp(it\lambda_j^2)(2\pi)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}(\lambda_j-\alpha_j)^2)\}\lambda_a^{p_a}\ldots\lambda_f^{p_f}d\lambda$$

$$= \left[\prod_{j\neq a,b,c,d,e,f} \Gamma_1^2(t;\alpha_j^2)\right]\left[\prod_{r=a,b,c,d,e,f} \{\sum_{s=0}^{6} g_s^r\Gamma^2_{1+2s}(t;\alpha_r^2)\}\right]$$

where the $\{g_s^r\}$ depend on $(p_a,\ldots,p_f)$. This can be expanded as a sum of products of non-central chi-squared characteristic functions. Each product is the product of exactly h non-central chi-square characteristic functions with differing degrees of freedom but such that first multiplicand has non-centrality parameter $\alpha_1^2$ the second $\alpha_2^2$ and so on, till the h'th multiplicand has non-centrality parameter $\alpha_h^2$. Thus each product can be expressed as the characteristic function of a non-central chi-square with degrees of freedom (h+2s) (s=0,...,6) and non-centrality parameter $\sum_{j=1}^{h}\alpha_j^2 = \alpha'\alpha$:

$$\int\{\prod_{j=1}^{h} \exp(it\lambda_j^2)(2\pi)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}(\lambda_j-\alpha_j)^2)\lambda_a^{p_a}\ldots\lambda_f^{p_f}d\lambda$$

$$= \sum_{s=0}^{6} k_s\Gamma^2_{h+2s}(t;\alpha'\alpha).$$

It should be noted that if $k_s \neq 0$ for $s \geq 4$ then $p_a + p_b + \ldots p_f \geq 4$ so that the terms $\Gamma^2_{h+2s}(t;\alpha'\alpha)$ for $s \geq 4$ must be multiplied by a term of $O(T^{-1})$ since $\{d_{jk\ell m}\}$, $\{d_{jk\ell mp}\}$, $\{d_{jk\ell mpq}\}$ are all of $O(T^{-1})$. Now it is possible to expand out Equation (58) and gathering terms together gives:

$$\sigma(t) = \Gamma^2_h(t;\Delta) + \sum_{j=o}^{6} u_j \Gamma^2_{h+2j}(t;\Delta) + O(T^{-3/2})$$

where $\Delta = \alpha_j \alpha_j = \alpha'\alpha$

$$\text{with} \quad u_j = \begin{cases} O(T^{-\frac{1}{2}}) & j = 0,1,2,3 \\ O(T^{-1}) & j = 4,5,6. \end{cases} \qquad (64)$$

The expansions for the p.d.f. and c.d.f. of $s^2$ then follow immediately. Defining $\chi^2_h(x;\Delta)$ as the pdf of a non-central chi-square variate with h degrees of freedom and non-centrality parameter $\Delta$ and $\xi^2_h(x;\Delta)$ as the corresponding cdf then:

$$\text{p.d.f.}(s^2) = \chi^2_h(s^2;\Delta) + \sum_{j=o}^{6} u_j \chi^2_{h+2j}(s^2;\Delta) + O(T^{-3/2})$$

$$\text{c.d.f.}(s^2) = \Pr(s^2 \leq x) = \chi^2_h(x;\Delta) + \sum_{j=o}^{6} u_j \xi^2_{h+2j}(x;\Delta) + O(T^{-3/2}) \qquad (65)$$

with the same $\{s_j\}$ as in (64).

It is possible to extend the analysis to cover an approximate inner product statistic and also the jointly independent case both for exact and approximate sums of squares statistics. Similar problems arise with an approximate inner product statistic as arose in the analysis for the null hypothesis case in Section 3.3. when it does not behave nicely at the origin and does not have enough degrees of freedom. These problems can be solved in essentially the same way as in Section 3.3.

The difficulty with applying the techniques of this Section to the jointly dependent case discussed in Section 3.4.2 is that the joint characteristic function, pdf and cdf of dependent non-central chi-square variates are not as simple in form as the corresponding functions for independent chi-square variates. Firstly, it is no longer possible to use the reproductive property of chi-square variates under independence given in Equation (63) and, secondly, the joint pdf's and cdf's now involve considerably more complex infinite summations than in the independent case; see Johnson and Kotz (1972). The complexity of the joint pdf and cdf also raises doubts about the amenability of the pdf and cdf expansions to numerical computation. The jointly dependent case both under the null hypothesis and under local alternative sequences remains an area for further research.

## 3.A Appendix to Chapter 3

### 3.A.1 Relationship with the Approach in Mauleon-Torres (1983)

The method developed in this chapter for obtaining asymptotic expansions for the cumulative distribution functions of asymptotically chi-square test criteria has some similarities to the method developed in Mauleon-Torres (1983) which was reviewed briefly in Chapter 2. Both methods involve expanding the statistic of interest in a vector $\lambda$ of asymptotically independent standard normal variates and obtaining an Edgeworth expansion for the distribution of this vector $\lambda$. The methods differ in how they proceed from $\lambda$ to $\lambda_o'\lambda_o$ where $\lambda_o$ is a sub-vector of $\lambda$ and where $\lambda_o'\lambda_o$ is the only O(1) term in the expansion of the statistic of interest. Mauleon-Torres (1983) uses a polar co-ordinates transformation in the same way as Sargan (1980b). This, however, requires the use of the moments of variates distributed on the unit hypersphere which is somewhat cumbersome. Again as Sargan (1980b), Mauleon-Torres' method obtains an expansion for the cumulative distribution function of the positive square root of the statistic thus involving the use of chi variates.

The method developed here transforms from $\lambda$ to $\lambda_o'\lambda_o$ by equating some of the parameters $\{t_r\}$ of the joint characteristic function of the $\{\lambda_r^2\}$ to t and setting the others to zero. This method has a number of advantages over that of Mauleon-Torres (1983). Firstly, it simplifies in the case of an exact inner product test criterion and shows why there are no $O(T^{-\frac{1}{2}})$ terms in the resulting cdf expansion. Secondly, it extends very easily to the jointly independent asymptotically chi-square test criteria case whereas extending the Mauleon-Torres method to this case would be much more cumbersome. Lastly, it can be modified, as in Section 3.5, for the distribution of test criteria under local alternative sequences thus enabling power comuptations to be made as well as size computations. In this case $\lambda$ is no longer asymptotically distributed spherically symmetric about the origin. Therefore it is not clear whether a transformation to polar co-ordinates will render obtaining a cdf expansion tractable. Mauleon-Torres (1983) does not consider any of these simplifications or extensions of the problem of obtaining an asymptotic expansion for the cumulative distribution function of an asymptotically chi-square test criterion.

### 3.A.2 The Covariance Matrix Assumption

In Section 2.1 we assume that when the decomposition vector is written as a function g(x) of the underlying variables x then $g_j^r g_K^s \kappa^{j,K} = \delta_{rs} + O(T^{-1})$ where $g_j^r$ is the derivative of the r'th element of g with respect to the j'th element of x evaluated at zero (which is the expectation of x), where $\kappa^{j,K}$ is the second-order cumulant of the j'th and K'th elements of x, and where $\delta_{rs} = 1$ if r = s and 0 otherwise. This is quite a strong assumption and it provides the justification for the orders of magnitude of the terms in, for example, Equation (27) in Section 3.2 which gives the approximation to the pdf of an exact inner product statistic. The purpose of this appendix is to show that this assumption is valid for Wald tests of non-linear restrictions on regression coefficients in dynamic linear regression models with normally distributed disturbances using OLS estimators.

As discussed in Chapter 1, the Wald test of $H_0$: $\psi(\theta) = 0$ for $Y \sim f(y;\theta)$ is given by:

$$W = T\psi(\hat{\theta})'[\hat{F}\hat{\Omega}\hat{F}']^{-1}\psi(\hat{\theta}),$$

where $\hat{\theta}$ is an estimate of $\theta$ such that $T^{\frac{1}{2}}(\hat{\theta}-\theta) \xrightarrow{D} N[0,\Omega]$, $\hat{F}$ is $[\partial\psi/\partial\theta']$ evaluated at $\hat{\theta}$, and $\hat{\Omega}$ is a consistent estimator of $\Omega$. The decomposition vector $\lambda$ is given by:

$$\lambda = T^{\frac{1}{2}}[\hat{F}\hat{\Omega}\hat{F}']^{-\frac{1}{2}}\psi(\hat{\theta}), \qquad (66)$$

where $[\hat{F}\hat{\Omega}\hat{F}']^{-\frac{1}{2}}$ is an appropriate matrix square-root of $[\hat{F}\hat{\Omega}\hat{F}']^{-1}$, e.g. a Cholesky decomposition. In general we can express $\hat{\theta}$, $\hat{F}$ and $\hat{\Omega}$ as functions of some stochastic vector of underlying variables p so that $\hat{\theta} = \hat{\theta}(p)$ etc.

Frequently our estimates have the property of Fisher consistency in p so that $\hat{\theta}(\mu) = \theta$ where $E(p) = \mu$. Therefore $\lambda$ evaluated at $p = \mu$ is equal to zero under the null hypothesis and we can write:

$$\lambda = g(x) = T^{\frac{1}{2}}[\phi(p) - \phi(\mu)]$$

where $\phi(p) = [\hat{F}\hat{\Omega}\hat{F}']^{-\frac{1}{2}}\psi(\hat{\theta})$ and $x = T^{\frac{1}{2}}(p-\mu)$.

Differentiating this with respect to the j'th element of x gives:

$$\frac{\partial g}{\partial x_j} = \frac{\partial \phi}{\partial p_j} = \frac{\partial}{\partial p_j}\{[\hat{F}\hat{\Omega}\hat{F}']^{-\frac{1}{2}}\}\psi(\hat{\theta}) + [\hat{F}\hat{\Omega}\hat{F}']^{-\frac{1}{2}}\frac{\partial\psi(\hat{\theta})}{\partial p_j}.$$

Evaluating this at x = 0 gives:

$$g_j = [\hat{F}\hat{\Omega}\hat{F}']_{\mu}^{-\frac{1}{2}}\left[\frac{\partial\psi(\hat{\theta})}{\partial p_j}\bigg|_{\mu}\right]$$

$$= [\hat{F}\hat{\Omega}\hat{F}']_{\mu}^{-\frac{1}{2}}\hat{F}_{\mu}\left[\frac{\partial\hat{\theta}}{\partial p_j}\bigg|_{\mu}\right]$$

$$= [F\Omega_{(\mu)}F']^{-\frac{1}{2}}F\left[\frac{\partial\hat{\theta}}{\partial p_j}\bigg|_{\mu}\right] \quad . \qquad (67)$$

where F and $\Omega_{(\mu)}$ equal $\hat{F}$ and $\hat{\Omega}$ evaluated at $p = \mu$. Therefore we have:

$$\left[\frac{\partial g}{x_j}\right]\kappa^{j,K}\left[\frac{\partial g}{\partial x_K}\right]'$$

$$= [F\Omega_{(\mu)}F']^{-\frac{1}{2}}F\left[\frac{\partial\hat{\theta}}{\partial p_j}\bigg|_{\mu}\right]\kappa^{j,K}\left[\frac{\partial\hat{\theta}}{\partial p_K}\bigg|_{\mu}\right]'F'[F\Omega_{(\mu)}F']^{-\frac{1}{2}'}$$

$$= [F\Omega_{(\mu)}F']^{-\frac{1}{2}}F\Sigma F'[F\Omega_{(\mu)}F']^{-\frac{1}{2}'}$$

where $\Sigma = \left[\frac{\partial\hat{\theta}}{\partial p_j}\bigg|_{\mu}\right]\kappa^{j,K}\left[\frac{\partial\hat{\theta}}{\partial p_K}\bigg|_{\mu}\right]'$.

Thus provided that:

$$\hat{\theta}(\mu) = \theta, \text{ and}$$

$$\Sigma = (1+cT^{-1})\Omega_{(\mu)}, \qquad (68)$$

then $g_j^r g_K^s \kappa^{j,K} = \delta_{rs} + O(T^{-1})$. It is interesting to note that the form of the restriction $H_0$: $\psi(\theta) = 0$ does not enter into either of the conditions given by Equation (68).

We will now suppose that the model generating the data is a dynamic linear regression model:

$$y = Z\gamma + u, \tag{69}$$

where $u \sim N[0,\sigma^2 I]$, where $u_{t+s}$ is independent of $Z_t$ for all $s \geq 0$, and where $Z_t$ may contain lagged values of $y_t$.
We then estimate this model by maximum likelihood conditional upon the initial values of $y_t$ giving:

$$\hat{\gamma} = \left(\frac{Z'Z}{T}\right)^{-1} \left(\frac{Z'y}{T}\right) ,$$

$$\hat{\sigma}^2 = \left[ \left(\frac{y'y}{T}\right) - \left(\frac{y'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'y}{T}\right)\right] ,$$

$$\hat{\Omega}_{\gamma\gamma} = \hat{\sigma}^2\left(\frac{Z'Z}{T}\right)^{-1} , \hat{\Omega}_{\gamma\sigma^2} = 0, \hat{\Omega}_{\sigma^2\sigma^2} = 2(\hat{\sigma}^2)^2 . \tag{70}$$

Thus p consists of the stochastic elements of $T^{-1}W'W$ where $W = (y:Z)$. Now $E(T^{-1}Z'y) = E(T^{-1}Z'Z)\gamma$ since $E(Z'u) = 0$ from independence so that $\hat{\gamma}\rightarrow(\mu) = \gamma$. Also we know that:

$$(T^{-1}y'y) = \gamma'(T^{-1}Z'Z)\gamma + 2\gamma'(T^{-1}Z'u) + (T^{-1}u'u);\ \text{so that}$$

$$E[T^{-1}y'y] = \gamma'E(T^{-1}Z'Z)\,\gamma + \sigma .$$

Therefore $\hat{\sigma}^2(\mu) = \sigma^2$ and thus the ML estimates exhibit Fisher consistency in p, and $\Omega_{(\mu)} = \sigma^2[E(T^{-1}Z'Z)]^{-1}$. Suppose we wish to test that $\psi(\gamma) = 0$. It is convenient to change the underlying variables with which we work from p to p* where p* are the stochastic elements of $T^{-1}W^{*\prime}W^*$ and $W^* = [u:Z]$. This transformation is a linear transformation which is very convenient since:

$$\left[\frac{\partial p^*}{\partial p'}\right] = \left[\frac{\partial p'}{\partial p^*}\right]^{-1} . \qquad (71)$$

Therefore $\kappa_*^{j,k}$ the second-cumulant of the j'th and k'th elements of $x^* = T^{\frac{1}{2}}(p^*-\mu^*)$, where $\mu^* = E(p^*)$, is given by:

$$\kappa_*^{j,k} = \left(\frac{\partial p_j^*}{\partial p_i}\right)\kappa^{i,\ell}\left(\frac{\partial p_k^*}{\partial p_\ell}\right), \text{ or}$$

$\kappa_* = \left(\frac{\partial p^*}{\partial p'}\right) \kappa \left(\frac{\partial p^*}{\partial p'}\right)'$ in matrix notation, and the derivatives of $\hat{\gamma}$ with respect to $p^*$ are given by:

$$\left[\frac{\partial \hat{\gamma}}{\partial (p^*)'}\right] = \left[\frac{\partial \hat{\gamma}}{\partial p'}\right]\left[\frac{\partial p}{\partial (p^*)'}\right] = \left[\frac{\partial \hat{\gamma}}{\partial p'}\right]\left[\frac{\partial p^*}{\partial p'}\right]^{-1} .$$

Therefore $\sum$ is unchanged since:

$$\left[\frac{\partial \hat{\gamma}}{\partial p'}\right] \kappa \left[\frac{\partial \hat{\gamma}}{\partial p'}\right]' = \left[\frac{\partial \hat{\gamma}}{\partial (p^*)'}\right] \kappa_* \left[\frac{\partial \hat{\gamma}}{\partial (p^*)'}\right]' .$$

The advantage of working with $p^*$ is that:

$$\hat{\gamma} = \gamma + \left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'u}{T}\right) .$$

Therefore evaluating the derivatives with respect to $p^*$ at $\mu^*$ gives:

$$\left.\frac{\partial \hat{\gamma}}{\partial p^{*\prime}}\right|_{\mu^*} = \left[\left\{E\left(\frac{Z'Z}{T}\right)\right\}^{-1} : \; 0\right].$$

where the first block refers to the derivatives with respect to $T^{-1}Z'u$ and the second block with respect to the stochastic elements of $T^{-1}Z'Z$. Since $E(T^{-1}Z'u) = 0$ from independence then the second cumulants of $T^{\frac{1}{2}}(T^{-1}Z'u)$ are given by $\sigma^2 E(T^{-1}Z'Z)$ and therefore :

$$\Sigma = \sigma^2[E(T^{-1}Z'Z)]^{-1} = \Omega(\mu). \qquad (72)$$

Therefore both of the conditions given by Equation (68) are met and therefore the covariance matrix assumption holds.

Finally, we often choose to use the OLS based estimate $s^2 = T(T-K)^{-1}\hat{\sigma}^2$ where Z is (TxK). If we are only interested in hypotheses such as $\psi(\gamma) = 0$ then this merely re-scales $\lambda$ by a factor $[T(T-K)^{-1}]^{-\frac{1}{2}} = 1 - \frac{K}{2T} + O(T^{-2})$ and therefore the covariance matrix assumption still holds.

# CHAPTER 4: Alternative Test Statistics and the Use of the Wald Testing Principle with Reformulations of Parameter Restrictions

## 4.1 Introduction

One of the major problems facing an applied econometrician is the choice of a test statistic to test an hypothesis of interest within a model which has already been specified and estimated. The difficulty arises because there are typically several alternative test statistics which he or she can use. Ideally the choice would be made in full knowledge of the finite sample properties of these alternative test statistics both under the null hypothesis being tested and under any alternative hypotheses of interest. Then the size and power functions of the test statistics could be compared and a particular test chosen with the most desirable properties for the given situation.

In practice such full knowledge of finite sample properties is not available. Basic asymptotic theory, which is generally available, serves to partition the available test statistics into asymptotic equivalence classes. However, there are no guarantees that asymptotically equivalent test statistics behave equally well in finite samples. Therefore, it is desirable to obtain information about test statistics beyond the basic asymptotic results. In Section 4.2 we review and discuss some of the finite sample results which are available for making choices between asymptotically equivalent test statistics for a given hypothesis. The first set of results relate to the inequality relationship between the Wald (W), Likelihood Ratio (LR) and Lagrange Multiplier (LM) test criteria for linear restrictions in the linear regression model with normally distributed errors. This inequality relationship has been discussed extensively by Savin (1976), Berndt and Savin (1977), Breusch (1979) and Evans and Savin (1982a, 1982b). The other available results relate to the problem of constructing LM test statistics using alternative estimates of the information matrix and are due to Davidson and MacKinnon (1981a, 1981b, 1983).

In Section 4.3 we consider a problem of current interest, namely that of the differing properties of Wald statistics constructed using differing formulations of a given set of parameter restrictions. In Section 4.4 we examine in detail an example of this problem which has been considered by Gregory and Veall (1984, 1985 ) and derive limiting parameter asymptotics for the cdf's of the statistics used in their example. Then in Section 4.5 we compute asymptotic expansion approximations to these cdf's and compare them with the limiting parameter asymptotics of Section 4.4 and the Monte Carlo results of Gregory and Veall (1984, 1985 ). Finally, in Section 4.6 we draw some overall conclusions.

## 4.2 Comparing Alternative Test Statistics

There has been a recent move in econometrics towards justifying specific test statistics as being generated by applying particular testing principles to the given problem. Thus Aldrich (1978) and Breusch (1978) show that Durbin's h test for serial correlation in the presence of a lagged dependent variable, Durbin (1970), can be obtained via the LM test principle. This approach has been useful in assigning test statistics to asymptotic equivalence classes but usually leaves a choice between several asymptotically equivalent test statistics.

In the linear regression model with normally distributed errors a quite strong numerical result has been found relating to Wald, LR and LM tests for linear restrictions: $W \geq LR \geq LM$. This was first found by Savin (1976) and extended by Berndt and Savin (1977). The most general form of the result is to be found in Breusch (1979). Breusch postulates the following model:

$$\underset{}{y} = \underset{(T\times k)}{X} \,.\, \underset{(k\times 1)}{\beta} + \underset{(T\times 1)}{u} \quad ; \quad u \sim N[0, \Omega(\theta)]$$

where $\beta$ is a vector of k parameters and $\theta$ a vector of p parameters: no elements of $\beta$ enter into $\theta$ or vice versa. If $\beta$ and $\theta$ are estimated by maximum likelihood estimation and then Wald, LR and LM tests are constructed for the set of linear parameter restrictions:

$$H_0 : R\beta = r \; ; \quad q \text{ restrictions } q \leq k$$
$$\text{or } \beta \in B_0 = \{\beta | R\beta = r\}$$

then numerically the Wald test statistic is greater than or equal to the LR test statistic which is greater than or equal to the LM test statistic. Breusch shows this by expressing the test statistics as the differences between log-likelihoods maximised with respect to $\beta$ with differing estimates of $\Omega(\theta)$ substituted in. Thus if $(\ddot{\beta}, \ddot{\Omega})$ are the unrestricted maximum likelihood estimates (MLE) and $(\bar{\beta}, \bar{\Omega})$ are the restricted MLE then

$$W = 2[L(\hat{\beta},\hat{\Omega}) - \sup_{\beta \in B_0} L(\beta,\hat{\Omega})]$$

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

$$LM = 2[\sup_{\beta} L(\beta,\tilde{\Omega}) - L(\tilde{\beta},\tilde{\Omega})].$$

Consequently $W \geq LR$ since:

$$\sup_{\beta \in B_0} L(\beta,\hat{\Omega}) \leq \sup_{\Omega,\beta \in B_0} L(\beta,\Omega) = L(\tilde{\beta},\tilde{\Omega})$$

and also $LR \geq LM$, and hence $W \geq LR \geq LM$, since

$$\sup_{\beta} L(\beta,\tilde{\Omega}) \leq \sup_{\Omega,\beta} L(\beta,\Omega) = L(\hat{\beta},\hat{\Omega}).$$

Breusch (1979) shows that $LR \geq LM$ still holds for testing non-linear restrictions on the regression coefficients of the form $\phi(\beta) = 0$. However, it is no longer generally true that $W \geq LR$ for testing non-linear restrictions and the reasons for this will be examined in more depth in Section 4.3. The relationship between the three test statistics has also been investigated for certain models with non-normal errors. Ullah and Zinde-Walsh (1984) examine the regression model with multi-variate t-distributed disturbances and find that for the appropriately calculated W, LR and LM statistics no overall inequality relationship holds even between the LR and LM statistics.

The inequality relationship raises the possibility that the tests can produce conflicting inferences when used with the same critical values based on their common asymptotic chi-square distribution. Thus with a particular set of data it is possible for the Wald test to reject when the LR test accepts. However, only certain conflicts are possible; thus, if the LM test rejects so must the LR and Wald tests. This implies that the power functions of the tests can be ranked so, for example, the power of the Wald test is necessarily greater than or equal to the power of the LR test.

In the case where $\Omega(\theta) = \sigma^2 I$, i.e. where the covariance matrix is scalar diagonal, an even stronger result holds with the three test statistics being monotonically increasing functions of each other (see Evans and Savin (1982a)):

$$W = T(\tilde{\sigma}^2 - \hat{\sigma}^2)/\hat{\sigma}^2$$

$$LR = T \log_e[1+(W/T)] \text{ and } LM = W/[1+(W/T)] \qquad (1)$$

where $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ are the restricted and unrestricted MLE of $\sigma^2$.

Furthermore, the statistic $F = (T-K)W/(qT)$ has an $F(q,T-k)$ distribution with a zero non-centrality parameter under the null hypothesis and a non-centrality parameter under the alternative which is a known function of the parameters and the exogenous variables. Thus exact critical values for the W, LR and LM statistics can be computed and when the exact rather than asymptotic critical values are used there is no possibility of conflicting inferences. Thus in this model the problems raised by the inequality relationship can be completely surmounted.

Evans and Savin (1982a) consider making adjustments to the critical values of the three test statistics based not on the exact functional relationship and the F-distribution but on Edgeworth expansions for the three test statistics. Defining $n = (T-k)$, they initially consider the modified tests:

$$
\begin{aligned}
W_* &= n(\bar{\sigma}^2 - \bar{\sigma}^2)/\bar{\sigma}^2 \\
LR_e &= [n+(q/2)-1] \log_e [1+(W_*/n)] \\
LM_* &= [(n+q)/n]W_*/[1+(W_*/n)] \quad . \qquad (2)
\end{aligned}
$$

The modified Wald statistic is obtained by substituting n for T in W: this corrects for the bias in the estimate $\bar{\sigma}^2$ of $\sigma^2$. Likewise $LM_*$ substitutes $(n+q)$ for T in LM which corrects for the bias in the estimate of $\bar{\sigma}^2$ of $\sigma^2$. Although $W_*$ and $LM_*$ have true sizes which are closer to their nominal sizes than do W and LM, their true power functions can still differ substantially from their nominal power functions. The modified Likelihood Ratio statistic $LR_e$, obtained by replacing T with $[n+(q/2)-1]$ in LR is an Edgeworth corrected test and its true and nominal sizes differ by a term of $O(T^{-2})$, following Anderson (1958, p.208). The approximation of the true power function of $LR_e$ by its nominal power function is generally very good, even for cases where this is not true for $W_*$ and $LM_*$.

Evans and Savin (1982a) then consider making Edgeworth corrections to $W_*$ and $LM_*$ by noting:

$$
\begin{aligned}
W_* &= LR_e\{1+[(LR_e-q+2)/(2n)]\} + O(T^{-3/2}) \\
LM_* &= LR_e\{1-[(LR_e-q+2)/(2n)]\} + O(T^{-3/2}) \qquad (3)
\end{aligned}
$$

so that the appropriate critical values for $W_*$ and $LM_*$ are:

$$z_W^e = z\{1+[(z-q+2)/(2n)]\}$$
$$z_{LM}^e = z\{1-[(z-q+2)/(2n)]\}.$$

These Edgeworth size-corrected tests have true and nominal sizes which differ by a term of $O(T^{-2})$. These are essentially the same corrections as those given by Rothenberg (1977). Evans and Savin (1982a) find that these Edgeworth corrected tests have true power functions which are very well approximated by their nominal power functions. Furthermore, the probability of conflict for the Edgeworth corrected tests is very low. The only cases for which the approximation was not very good were when the nominal size, the degrees of freedom n, and the ratio of n to q were all small.

It is arguable that the success of the Edgeworth corrections in this case arises for three reasons. Firstly, the statistics are all monotonically increasing functions of each other; therefore it is possible to reduce the probability of conflict to zero by setting the critical values for the original test statistics. If this were not so then even with the correct critical values there would still be some probability of conflict of inference and this would also be true with Edgeworth corrected statistics. Secondly, the Edgeworth corrections do not depend on unknown parameters and therefore they can be computed exactly rather than having to be estimated. If they needed to be estimated, then making the corrections would introduce an extra source of sampling variation into the tests and this would be likely to provide an extra source of potential conflict of inference. Lastly, the distributions involved in the example above are all closely related to the chi-square distribution. For example, the F-distribution can be obtained as the distribution of the ratio of two independent chi-square variates, each divided by its degrees of freedom. If the distributions involved were not so closely related to chi-square distributions we might well expect that the Edgeworth corrections would not account so well for the size properties of the tests. Thus the case considered by Evans and Savin (1982a) would seem to be one of the most favourable cases for making Edgeworth corrections. In general we would not expect the circumstances to be as favourable.

Another problem involving the choice of a test statistic from a set of asymptotically equivalent test statistics arises with the LM test when it is difficult to estimate the information matrix. If the restricted MLE is $\tilde{\theta}$, then the LM test statistic takes the form:

$$LM = \left(\frac{1}{T}\right) \tilde{s}'\tilde{I}^{-1}\tilde{s} \qquad (4)$$

where $\tilde{s} = \left[ \frac{\partial L}{\partial \theta} \bigg|_{\theta = \tilde{\theta}} \right]$, the score vector evaluated at $\tilde{\theta}$; and

$\tilde{I}$ is a consistent estimate based on $\tilde{\theta}$ of the information matrix which is given by:

$$I(\theta_0) = E\left\{ -\frac{1}{T} \frac{\partial^2 L}{\partial\theta\partial\theta'}\bigg|_{\theta = \theta_0} ; \theta_0 \right\}$$

where $\theta_0$ are the true parameter values.

However the obvious estimate obtained by substituting $\tilde{\theta}$ for $\theta_o$ giving $\mathcal{I}(\tilde{\theta})$ may not be at all easy to obtain, in which case alternative estimates of $\mathcal{I}(\theta_o)$ may be used which are easier to compute and which are asymptotically equivalent to $\mathcal{I}(\tilde{\theta})$. One possibility is the Hessian at $\tilde{\theta}$ multiplied by (-1/T):

$$A(\tilde{\theta}) = \left[ -\frac{1}{T} \frac{\partial^2 L}{\partial\theta\partial\theta'} \bigg|_{\theta = \tilde{\theta}} \right] = \left[ -\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 L_t}{\partial\theta\partial\theta'} \bigg|_{\theta = \tilde{\theta}} \right]. \quad (5)$$

Another estimate is the sum of the outer products of the score vectors for the individual observations evaluated at $\tilde{\theta}$:

$$B(\tilde{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial L_t}{\partial\theta} \bigg|_{\theta = \tilde{\theta}} \right] \left[ \frac{\partial L_t}{\partial\theta} \bigg|_{\theta = \tilde{\theta}} \right]' \quad (6)$$

(advocated by Berndt, Hall, Hall and Hausman (1974)).

The justification for these is that their expectations at $\theta = \tilde{\theta}$ are both equal to $\mathcal{I}(\tilde{\theta})$. A more sophisticated method might be to conduct a Monte Carlo experiment to estimate $\mathcal{I}(\tilde{\theta})$: this is likely to be very expensive computationally. Other possibilities include any of the above but with zeros substituted for any elements of $\mathcal{I}(\tilde{\theta})$ which are known to be zero from theoretical considerations.

One problem with using $A(\tilde{\theta})$ is that the resulting test statistic is not invariant to reparameterizations of the model. This can be most easily seen in the framework of a model with a single parameter, say $y \sim f(\theta)$ and $H_o: \theta = \theta_o$. Then

$$\tilde{s} = \frac{\partial L(\theta)}{\partial \theta}\bigg|_{\theta_0} ; \; A(\tilde{\theta}) = \frac{-1}{T}\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta_0} .$$

Now consider a one-to-one reparameterization, $\theta \to \phi(\theta)$, so the hypothesis being tested becomes $H:\phi(\theta) = \phi(\theta_0)$ or $H:\phi = \phi_0$ where $\phi_0 - \phi(\theta_0)$. Then:

$$\tilde{s}^* - \frac{\partial L(\theta)}{\partial \phi}\bigg|_{\phi_0} = \frac{\partial L(\theta)}{\partial \theta}\bigg|_{\theta_0} \times \frac{\partial \theta}{\partial \phi}\bigg|_{\phi_0}$$

$$A^*(\tilde{\theta}) = -\frac{1}{T}\frac{\partial^2 L(\theta)}{\partial\phi\partial\phi'}\bigg|_{\phi_0} = -\frac{1}{T}\frac{\partial}{\partial\phi'}\left\{\frac{\partial L(\theta)}{\partial\phi}\bigg|_{\phi_0}\right\}$$

$$= -\frac{1}{T}\frac{\partial}{\partial\phi'}\left\{\frac{\partial L(\theta)}{\partial\theta}\bigg|_{\theta_0} \times \frac{\partial\theta}{\partial\phi}\bigg|_{\phi_0}\right\}$$

$$= \left[-\frac{1}{T}\frac{\partial}{\partial\phi'}\left\{\frac{\partial L(\theta)}{\partial\theta}\bigg|_{\theta_0}\right\} \times \frac{\partial\theta}{\partial\phi}\bigg|_{\phi_0}\right] + \left[-\frac{1}{T}\frac{\partial L(\theta)}{\partial\theta}\bigg|_{\theta_0} \times \frac{\partial^2\theta}{\partial\phi\partial\phi'}\bigg|_{\phi_0}\right]$$

$$= -\frac{1}{T}\left[\frac{\partial\theta}{\partial\phi}\bigg|_{\phi_0}\right]^2 \left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\bigg|_{\theta_0}\right] - \frac{1}{T}\left[\frac{\partial L(\theta)}{\partial\theta}\bigg|_{\theta_0} \times \frac{\partial^2\theta}{\partial\phi\partial\phi}\bigg|_{\phi_0}\right]$$

$$= \left[\frac{\partial\theta}{\partial\phi}\bigg|_{\phi_0}\right] A(\tilde{\theta}) \left[\frac{\partial\theta}{\partial\phi}\bigg|_{\phi_0}\right] - \frac{1}{T}\left[s(\tilde{\theta}) \times \frac{\partial^2\theta}{\partial\phi\partial\phi'}\bigg|_{\phi_0}\right] . \qquad (7)$$

Then $LM = \frac{1}{T}\tilde{s}'A(\tilde{\theta})^{-1}\tilde{s}$ and $LM^* = \frac{1}{T}\tilde{s}^{*\prime}A^*(\tilde{\theta})^{-1}\tilde{s}^*$ and it is clear that in general $LM \neq LM^*$. If expectations were taken at $\tilde{\theta}$ then no problem would arise since $E[s(\tilde{\theta})] = 0$ and so:

$$\frac{1}{T}\tilde{s}'E[A(\tilde{\theta})]^{-1}\tilde{s} = \frac{1}{T}\tilde{s}^{*\prime}E[A^*(\tilde{\theta})]^{-1}\tilde{s}^* .$$

This problem of non-invariance with respect to reparameterizations also appears with Wald tests and is discussed at greater length in Sections 4.3 and 4.4 and in Chapter 6.

Davidson and MacKinnon (1983) argue that $B(\tilde{\theta})$ does not provide a particularly well behaved LM test either and advocate an alternative estimate of $\mathcal{I}(\theta_o)$ for constructing the LM test. This alternative estimate was proposed in Davidson and MacKinnon (1981a). In somewhat altered notation their model is:

$$g_t\ (y_t;\theta) = \varepsilon_t\ ;\ \varepsilon_t \sim IN(0,1);\quad \theta \text{ is } k\times 1;$$

where $y_t$ is the t'th observation on the dependent variable, $\theta$ is a vector of unknown parameters, and $g_t(\cdot)$ is a suitably continuous function which may depend on exogenous variables and/or lagged values of y. The assumption that the $\{\varepsilon_t\}$ have unit variance is made without loss of generality.

The contribution to the log-likelihood from the t'th observation is:

$$L_t(\theta) = -\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}[g_t(y_t;\theta)]^2 + \log|g_t'(y_t;\theta)|$$

where $g_t'(y_t;\theta) = \frac{\partial}{\partial y_t}\, g_t(y_t;\theta)$.

The score vector for the t'th observation is then:

$$s_t(\theta) = \frac{\partial L_t(\theta)}{\partial\theta} = -\ g_t(y_t;\theta)\,\frac{\partial g_t}{\partial\theta} + \frac{\partial}{\partial\theta}\{\log|g_t'(\theta)|\}$$

$$= -g_t(y;\theta)G_t(\theta) + J_t(\theta). \qquad (8)$$

Davidson and MacKinnon (1981a) show that $\mathcal{I}(\theta_o)$ can be consistently estimated by:

$$C(\tilde{\theta}) = \frac{1}{T}\{\sum_{t=1}^{T}[G_t(\tilde{\theta})G_t(\tilde{\theta})' + J_t(\tilde{\theta})J_t(\tilde{\theta})']\}$$

which gives an LM test statistic:

$$[\sum_{t=1}^{T}(-\tilde{g}_t\tilde{G}_t + \tilde{J}_t)]'[\sum_{t=1}^{T}(\tilde{G}_t\tilde{G}_t' + \tilde{J}_t\tilde{J}_t')]^{-1}[\sum_{t=1}^{T}(-\tilde{g}_t\tilde{G}_t + \tilde{J}_t)] \quad (9)$$

which can be computed as the explained sum of squares from an artificial regression of the form:

$$z = Wb + u$$

where $z = [\tilde{g}_1,\ldots,\tilde{g}_T,1,\ldots,1]'$ a 2T-vector
$W = [\tilde{G}_1,\ldots\tilde{G}_T,\tilde{J}_1,\ldots,\tilde{J}_T]'$ a (2T×k) matrix.

They refer to this variant of the LM test as the DLR (Double Length Regression) variant because this artificial regression has 2T 'observations'. The LM test statistic constructed using $B(\tilde{\theta})$ they refer to as the OPG (Outer Product of the Gradient) variant. This can be constructed as the explained sum of squares from the artificial regression:

$$i = Pb + n$$

where $i = [1,\ldots,1]'$ a T-vector
$P = [\tilde{s}_1,\ldots,\tilde{s}_T]'$ a (T×k) matrix.

Davidson and MacKinnon (1983) argue that ideally an estimate of the information matrix should only depend on the dependent data through $\tilde{\theta}$, the restricted MLE. This is because the information matrix itself only depends on the exogenous variables and the parameters $\theta_0$. The less stochastic the estimate of $I(\theta_0)$ is then the more closely will the true distribution of the LM statistic match its asymptotic chi-square distribution. Davidson and MacKinnon argue that $B(\tilde{\theta})$ contains many more stochastic terms than does $C(\tilde{\theta})$. Therefore in finite samples $C(\tilde{\theta})$ should provide a more efficient estimate of $I(\theta_0)$ than does $B(\tilde{\theta})$. Hence the DLR variant of the LM statistic ought to be more nearly chi-square in finite samples than the OPG variant.

They conduct a Monte Carlo experiment to compare the properties of these two variants of the LM test for testing a logarithmic against a linear specification using the Box-Cox transformation. Since the test has only one degree of freedom, they transform the LM test statistics into asymptotic N(0,1) statistics by taking their square roots and multiplying by the sign of the 'estimated' coefficient $\tilde{b}$ in the appropriate artificial regression. They find that the DLR variant has a distribution under the null hypothesis which is much closer to N(0,1) than that of the OPG variant. Further results presented in another paper, Davidson and MacKinnon (1981b), suggest that the actual power function of the DLR variant is much closer than that of the OPG variant to their nominal power function.

### 4.3 Wald Statistics for Reformulations of a Set of Parameter Restrictions

In Section 4.2 we noted that typically no inequality relationship between W, LR and LM test statistics can be found when testing non-linear parameter restrictions. This is because the Wald test statistic is not invariant to reformulations of the parameter restrictions. This problem has been examined by Gregory and Veall (1984, 1985 , 1986 ), Breusch and Schmidt (1985) and Lafontaine and White (1985) .

Gregory and Veall (1984) consider a number of specific examples of non-invariant Wald tests. Their first example is also reported in Gregory and Veall (1985) and concerns testing a non-linear restriction in a simple CLR model; this case is examined in further detail in Sections 4.4 and 4.5. Their second example is also reported in Gregory and Veall (1986) and concerns testing for common factor (COMFAC) restrictions in a simple autoregressive distributed lag model. This example is examined further in Chapter 6. They considered four different formulations of the restriction and from a Monte Carlo study they concluded that none of the four Wald tests corresponding to the different formulations of the restriction performed uniformly the best across all the different parameter settings they used.

The third example in Gregory and Veall (1984) concerns testing for rational expectations as discussed by Hoffman and Schmidt (1981). The latter showed that the rational expectations hypothesis could be expressed in their model as a set of non-linear parameter restrictions. They derived these restrictions in terms of ratios of products of the parameters of the model. They then performed a Monte Carlo study to examine the performance of the Wald and LR tests for these restrictions; however they reformulated the restrictions to be in purely multiplicative form before computing the test statistics. Gregory and Veall (1984) performed a Monte Carlo study to compare the Wald tests from the original and multiplicative forms of the restrictions. They found that the test from the multiplicative form performed much better in terms of the accuracy of the asymptotic cdf as an approximation to the actual cdf of the test statistic under the null hypothesis.

Gregory and Veall (1984) provide an empirical example of testing for rational expectations in a model of the Canadian demand for money. They find that the Wald statistics corresponding to the different forms of the restrictions had substantially different numerical values. This is not surprising since the power functions of the tests are different and so they do sometimes conflict. However, it does raise a worrying point which is examined further by Breusch and Schmidt (1985). The latter demonstrate that for any given null hypothesis and any given data set it is possible to generate any given positive value for the Wald statistic by appropriately rewriting the null hypothesis in an algebraically equivalent form provided that the statistic does not equal zero. This phenomenon arises because both the restriction function and its derivatives need to be evaluated at the unrestricted parameter estimates. Thus suppose $y \sim f(y;\theta)$ where $\theta$ is a K-vector and that the null hypothesis under consideration is $H_0:\phi(\theta) = 0$ i.e. a set of q parameter restrictions where $q \leq K$. Then the Wald test statistic corresponding to this particular algebraic formulation of the null hypothesis is:

$$W_\phi = T(\phi(\hat{\theta}))'\{F_\phi(\hat{\theta})\hat{V}(\hat{\theta})F_\phi(\hat{\theta})'\}^{-1}\{\phi(\hat{\theta})\} \qquad (10)$$

where $\hat{\theta}$ is a consistent estimate of $\theta$ without the parameter restrictions $\phi(\theta) = 0$ imposed;

$\hat{V}(\hat{\theta})$ is a consistent estimate of the asymptotic covariance matrix of $\hat{\theta}$; and

$F_\phi(\hat{\theta}) = \frac{\partial\phi}{\partial\theta'}\Big|_{\theta = \hat{\theta}}$, a (q×K) matrix of derivatives.

However, the null hypothesis can be algebraically reformulated as $H_0^*$ : $\psi(\theta) = 0$ with $\psi(\theta) = 0$ if and only if $\phi(\theta) = 0$. For example, if the restriction is $\theta - r = 0$ (a single parameter case) then an equivalent restriction is $\exp(\theta) - \exp(r) = 0$. The Wald statistic corresponding to the reformulation $H_0^*$ of the null hypothesis is:

$$W_\psi = T\{\psi(\hat{\theta})\}'\{F_\psi(\hat{\theta})\hat{V}(\hat{\theta})F_\psi(\hat{\theta})'\}^{-1}\{\psi(\hat{\theta})\}$$

where $F_\psi(\hat{\theta}) = \left.\frac{\partial\psi}{\partial\theta'}\right|_{\theta = \hat{\theta}}$ is a (q×K) matrix.

In general for an arbitrarily chosen reformulation of the restriction, $W_\psi$ will not equal $W_\phi$. In the univariate case:

$$W = \frac{[\phi(\hat{\theta})]^2}{[F_\phi(\hat{\theta})]^2 V(\hat{\theta})} \quad \text{and} \quad W = \frac{[\psi(\hat{\theta})]^2}{[F_\psi(\hat{\theta})]^2 V(\hat{\theta})} \qquad (11)$$

and the null hypothesis does not fix $[\psi(\hat{\theta})/F_\psi(\hat{\theta})]^2$. This can be seen from Figure 4.1 (from Figure 1 in Breusch and Schmidt (1985) ):

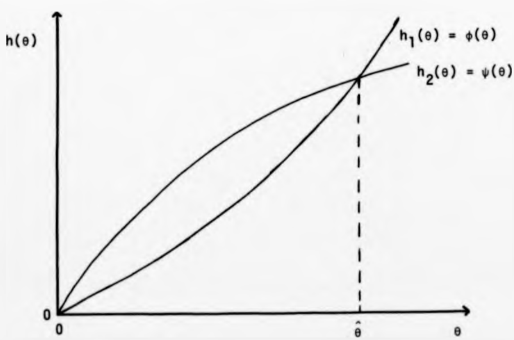


Figure 4.1: Alternative Formulations of the Restriction Function

In the diagram $\phi(\hat{\theta}) = \psi(\hat{\theta})$ but $F_\phi(\hat{\theta}) > F_\psi(\hat{\theta})$ so that $W_\phi < W_\psi$. Only if the reformulation is linear will $W_\phi = W_\psi$ but this will not generally be the case. Breusch and Schmidt (1985) make this argument and show that a $\psi(\cdot)$ can be found given $\hat{\theta}$ such that $W_\psi$ has any particular desired value provided $W_\phi \neq 0$. They then extend this to the multi-parameter case with $q \leq K$. The proviso that $W_\phi \neq 0$ is obvious since $W_\phi = 0$ if and only if $\phi(\theta) = 0$ so that $W_\phi = 0$ if and only if $\psi(\theta) = 0$ and hence if and only if $W_\psi = 0$.

Lafontaine and White (1986) consider a particular case of reformulating a null hypothesis to examine the resultant Wald statistic. They take an example from Theil (1971, p.102) concerning textile data and give the regression equation:

$$\underset{}{\hat{Y}} = \underset{(0.31)}{1.37} + \underset{(0.16)}{1.14}\,X - \underset{(0.04)}{0.83}Z\ ; \quad \hat{\sigma}^2 = 0.0001833,\ T = 17\ ; \quad \text{(standard errors)}$$

where Y is the log of per-capita consumption of textiles;

X is the log of per-capita real income; and

Z is the log of the relative price of textiles.

This is a particular case of $Y = \alpha + \beta X + \gamma Z + \varepsilon$ with $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I$. In this context it is interesting to test the linear parameter restriction $H_0 : \beta = 1$ which corresponds to the null hypothesis that income elasticity is equal to unity. However, the hypothesis could be equivalently reformulated as $H_0^K : \beta^K = 1$ for $K \neq 0$ provided $\beta > 0$ (K an integer). Then the Wald statistic would become:

$$W_K = (\hat{\beta}^K - 1)^2 / [(K\hat{\beta}^{K-1})^2 \hat{V}(\hat{\beta})]$$

where $\hat{V}(\hat{\beta})$ is the estimated variance of $\hat{\beta}$. All the $W_K$ should have the same limiting distribution which is that of a central chi-square variate with one degree of freedom under the null hypothesis. Lafontaine and White show that $W_K$ varies from 0.03 (for K=40) to 1479.65 (for K=-40). They conduct a Monte Carlo experiment giving estimated critical values ranging from $3.5 \times 10^8$ (for K=40) to 4.60 (for K=1) to $7.1 \times 10^6$ (for K=-40). Their results indicate that quite markedly different inferences can be obtained depending on the value of K.

Lafontaine and White propose approximating the power function at $\theta^*$ using a non-central $\chi^2$ distribution with a non-centrality parameter:

$$\delta_\phi^2 = T(\phi(\theta^*))'\left[F_\phi(\theta^*)E(-\frac{1}{T}\frac{\partial^2 L}{\partial\theta\partial\theta'}\Big|_{\theta^*};\ \theta^*)^{-1}F_\phi(\theta^*)'\right]^{-1}(\phi(\theta^*))$$

following Gallant and Holly (1980). Lafontaine and White use the observed Hessian at $\theta^*$ rather than its expectation at $\theta^*$, but in this model the two are identical. Although the Gallant and Holly approximation does not seem to be very successful in approximating the power function as estimated by Monte Carlo methods, it may still prove useful in explaining the behaviour observed. The general Wald statistic for $H_\phi$: $\phi(\theta) = 0$ is:

$$W_\phi = T(\phi(\hat{\theta}))'[F_\phi(\hat{\theta})\hat{V}(\hat{\theta})F_\phi(\hat{\theta})']^{-1}(\phi(\hat{\theta}))$$

and can be interpreted as being numerically equivalent to the Wald statistic for:

$$H_\phi^+ : F_\phi(\hat{\theta})\theta - [F_\phi(\hat{\theta})\hat{\theta} - \phi(\hat{\theta})] = 0 \qquad (12)$$

given the data which produced $\hat{\theta}$. This is a linearization of the restriction around the unrestricted parameter estimates. Effectively the Gallant and Holly approximation substitutes $\theta^*$ (the hypothesised alternative) for $\hat{\theta}$ in the above to obtain $H_{\phi}^{+}$ and then computes the power at $\theta^*$ for the Wald test for $H_{\phi}^{+}$ where $\theta^*$ replaces $\hat{\theta}$, assuming that the estimated variance of $\hat{\theta}$, namely $\hat{V}(\hat{\theta})$, is equal to its actual value when $\theta = \theta^*$.

Unlike the Evans and Savin (1982a) example, the Lafontaine and White statistics are not monotonic functions of each other. Therefore the power functions of the various tests are not identical when the correct critical values are used. Consequently these test statistics even when correct critical values are used are intrinsically able to produce conflicting inferences. The Wald statistic for $H_0:\beta=1$ gives a uniformly most powerful unbiased test, see Silvey (1975), but not a uniformly most powerful test. When $K > 1$ the resultant correct-size test is relatively more powerful against alternatives where $\beta < 1$ than is the test resulting from $K = 1$ and it is relatively less powerful against alternatives where $\beta > 1$. This behaviour is reversed when $K < 0$. These results are obtained from Monte Carlo studies but are not too surprising because when $K \neq 1$ the parameter restriction function does not treat the parameter space symmetrically around $\beta = 1$.

The Breusch and Schmidt (1985) and Lafontaine and White (1986) results raise the worrying point that an unscrupulous investigator can achieve a desired inference simply by appropriately formulating the null hypothesis. This is clearly an undesirable feature for a testing principle and is not shared by the Likelihood Ratio testing principle. The Lagrange Multiplier testing principle can suffer from this problem to a lesser extent because of the problems which estimation of the information matrix may raise. This is much less likely to be a serious problem if several hypotheses are tested separately since then the investigator is unlikely to have much justification for switching between different estimation methods for the information matrix or between different parameterizations of the model.

The Wald testing principle does have the advantage that the different possible Wald tests will typically have different power functions even when corrected critical values are used. Therefore a particular formulation of the restrictions could be obtained to have high power for various areas of the parameter space. This kind of flexibility is not available when the Likelihood Ratio testing principle is used and is only likely to be available to a small extent when the Lagrange Multiplier testing principle used.

## 4.4 Testing for a Non-Linear Restriction in a Simple Model using the Wald Principle

Gregory and Veall (1984) present another example of testing using the Wald principle. They propose the model:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t \; ; \; \epsilon_t \sim IN(0,\sigma^2); \qquad (13)$$

$x_1$ and $x_2$ are exogenous

and then consider testing the null hypothesis that the two slope coefficients are reciprocals i.e. $\beta_1\beta_2 = 1$. This can be written in various forms of which Gregory and Veall (1984) consider four:

(i) $H_M : h_M(\beta) = \beta_1\beta_2 - 1 = 0$;

(ii) $H_{R1} : h_{R1}(\beta) = \beta_1 - (1/\beta_2) = 0$;

(iii) $H_{R2} : h_{R2}(\beta) = \beta_2 - (1/\beta_1) = 0$; and

(iv) $H_{R3} : h_{R3}(\beta) = 1 - (1/\beta_1\beta_2) = 0$. (14)

In Gregory and Veall (1985) they present some of the results for cases (i) and (ii). Unlike the rational expectations restrictions and COMFAC restrictions of Gregory and Veall (1984) it is difficult to find any particular justification for this kind of non-linear restriction. However, this case does provide an extremely simple example of non-linearity in a well behaved model. There are no lagged variables or other sources of dynamics; similarly there are no other endogenous variables to provide simultaneity problems. The error structure is assumed to be normal and homoscedastic. In this model OLS provides a Best Linear Unbiased estimator of the regression coefficients and an unbiased variance estimate: therefore computing the Wald statistics for the various restriction formulations is very simple. The model seems slightly more complicated than is absolutely necessary since there is an intercept term in the equation. However, as is shown in the Appendix to this chapter, the intercept term plays no fundamental role in the fixed regressors case. Without the intercept term the model becomes the simplest one which allows us to consider non-linear restrictions involving more than one regression coefficient.

Returning to the comparison of Wald tests for the null hypothesis it is now valuable to examine the actual statistics in some detail. Gregory and Veall (1984) used OLS estimates of $\beta$ and $\sigma^2$ to construct the Wald tests. In this example it is well known that OLS and ML estimates for $\beta$ are identical and only differ for $\sigma^2$ in the degrees of freedom used as a divisor :

$$\hat{\beta} = (X'X)^{-1}X'y;$$
$$\hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/(T-K); \quad \text{(here } K = 3)$$
$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1};$$

where $X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & x_{1T} & x_{2T} \end{bmatrix}$ is a $(T\times3)$ matrix;

$y = (y_1, \ldots, y_T)'$ is a T-vector;

$\beta = (\beta_0, \beta_1, \beta_2)'$ is a 3-vector.

The four Wald test statistics then are given as:

$$W_M = \frac{T(\hat{\beta}_1\hat{\beta}_2 - 1)^2}{\hat{\sigma}^2\left\{[0, \hat{\beta}_2, \hat{\beta}_1]\left(\frac{X'X}{T}\right)^{-1}\begin{bmatrix} 0 \\ \hat{\beta}_2 \\ \hat{\beta}_1 \end{bmatrix}\right\}}$$
$$= T(\hat{\beta}_1\hat{\beta}_2 - 1)^2/\{[\hat{\beta}_2^2 c_{11} + 2\hat{\beta}_1\hat{\beta}_2 c_{12} + \hat{\beta}_1^2 c_{22}]\cdot\hat{\sigma}^2\}$$
$$W_{R1} = T(\hat{\beta}_1\hat{\beta}_2 - 1)^2/\{\hat{\sigma}^2[\hat{\beta}_2^2 c_{11} + 2c_{12} + (c_{22}/\hat{\beta}_2^2)]\}$$
$$W_{R2} = T(\hat{\beta}_1\hat{\beta}_2 - 1)^2/\{\hat{\sigma}^2[(c_{11}/\hat{\beta}_1^2) + 2c_{12} + \hat{\beta}_1^2 c_{22}]\}$$
$$W_{R3} = T(\hat{\beta}_1\hat{\beta}_2 - 1)^2/\{\hat{\sigma}^2[(c_{11}/\hat{\beta}_1^2) + (2c_{12}/\hat{\beta}_1\hat{\beta}_2)) + (c_{22}/\hat{\beta}_2^2)]\} \quad (15)$$

where $C = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix}$ is the lower right-hand corner (2×2) block of $(\frac{X'X}{T})^{-1}$.

Gregory and Veall (1984) examined, using Monte Carlo simulations, the properties of the test statistics at various parameter settings both under the null and for alternative hypotheses. However, it is possible to explain the qualitative behaviour which they observed by the use of various limiting arguments as $\beta_1 \to \infty$ or $\beta_2 \to \infty$.

Assuming that the X matrix is taken as fixed then OLS estimation in this particular example has a number of features which reduce the problem somewhat. Firstly, $\sqrt{T}(\hat{\beta} - \beta)$ is distributed as a normal vector with mean zero and covariance matrix $\sigma^2 C$. Secondly, $(T-K)(\frac{1}{\sigma^2})\hat{\sigma}^2$ is distributed as a central chi-square with (T-K) degrees of freedom. Thirdly, $\hat{\beta}$ and $\hat{\sigma}^2$ are independent; see Silvey (1975) for details. It is therefore possible to re-write the estimates as :

$$\hat{\beta}_1 = \beta_1 + u_1$$
$$\hat{\beta}_2 = \beta_2 + u_2$$
with $u \sim N[0,(\frac{1}{T})\sigma^2 C]$ where $u = (u_1, u_2)'$.

so that the statistics can then be examined in terms of u and $\hat{\sigma}^2$ and the actual parameters.

The statistic $W_M$ corresponding to the purely multiplicative formulation of the hypothesis, $H_M$, can be rewritten as:

$$W_M = \eta_M^2 \quad \text{where } \eta_M = \frac{\sqrt{T}[\hat{\beta}_2 - (1/\hat{\beta}_1)]}{\left\{\hat{\sigma}^2\left[c_{11}\left(\frac{\hat{\beta}_2}{\hat{\beta}_1}\right)^2 + 2c_{12}\left(\frac{\hat{\beta}_2}{\hat{\beta}_1}\right) + c_{22}\right]\right\}^{\frac{1}{2}}} \qquad (16)$$

which can be decomposed into three parts:

(a) $(\hat{\sigma}^2)^{\frac{1}{2}}$ which is independent of the rest of the statistic;

(b) $\sqrt{T}[\hat{\beta}_2 - (1/\hat{\beta}_1)]$; and

(c) $[c_{11}(\hat{\beta}_2/\hat{\beta}_1)^2 + 2c_{12}(\hat{\beta}_2/\hat{\beta}_1) + c_{22}]^{\frac{1}{2}}$.

The behaviour of (a) does not depend on $\beta$ at all. In contrast, (b) can be rewritten:

$$\sqrt{T}[\hat{\beta}_2 - (1/\hat{\beta}_1)] = \sqrt{T}[(u_2 + \beta_2) - (\frac{1}{u_1 + \beta_1})]$$

$$= \sqrt{T}\left[\frac{(u_1 + \beta_1)(u_2 + \beta_2) - 1}{(u_1 + \beta_1)}\right]$$

$$= \sqrt{T}\left[\frac{(u_1 u_2 + \beta_1 u_2 + \beta_2 u_1 + \beta_1 \beta_2 - 1)}{u_1 + \beta_1}\right]$$

$$= \sqrt{T}\left[\frac{u_2 + (1/\beta_1)u_1 u_2 + (\beta_2/\beta_1)u_1 + \beta_2 - (1/\beta_1)}{1 + (1/\beta_1)u_1}\right].$$

Lastly, (c) can be rewritten as:

$$[c_{11}(\hat{\beta}_2/\hat{\beta}_1)^2 + 2c_{12}(\hat{\beta}_2/\hat{\beta}_1) + c_{22}]^{\frac{1}{2}}$$

$$= \left\{c_{11}\left[\frac{u_2(1/\beta_1) + (\beta_2/\beta_1)}{1 + (1/\beta_1)u_1}\right]^2 + 2c_{12}\left[\frac{u_2(1/\beta_1) + (\beta_2/\beta_1)}{1 + (1/\beta_1)u_1}\right] + c_{22}\right\}^{\frac{1}{2}}.$$

If we define $\gamma^* = (1/\beta_1)$ then we can rewrite $\eta_M$ as:

$$\eta_M = \frac{\sqrt{T}\left[\dfrac{u_2 + \gamma^* u_1 u_2 + \beta_2 \gamma^* u_1 + \beta_2 - \gamma^*}{1 + \gamma^* u_1}\right]}{(\hat{\sigma}^2)^{\frac{1}{2}}\left\{c_{11}\left[\dfrac{\gamma^* u_2 + \beta_2 \gamma^*}{1 + \gamma^* u_1}\right]^2 + 2c_{12}\left[\dfrac{\gamma^* u_2 + \beta_2 \gamma^*}{1 + \gamma^* u_1}\right] + c_{22}\right\}^{\frac{1}{2}}} = \eta_M(u_1, u_2, \gamma^*, \beta_2, \hat{\sigma}^2). \quad (17)$$

Consider firstly its behaviour as $\beta_1 \to \infty$ while maintaining $\beta_1\beta_2 = 1$ so that $\gamma^* \to 0$ and $\beta_2 \to 0$. The distribution of $(u_1,u_2)$ is clearly unaffected by changing $\beta_1$ or $\beta_2$. Applying Cramer's theorem in its generalised form (see Rao (1973, p.122)) gives:

$$\lim_{\beta_1\to\infty} \Pr\{\eta_M \leq x\} = \Pr\{\eta_M(u_1,u_2, \lim_{\beta_1\to\infty} \gamma^*, \lim_{\beta_1\to\infty} \beta_2, \hat{\sigma}^2) \leq x\} \quad (18)$$

with $\beta_1\beta_2 = 1$ maintained. Therefore:

$$\lim_{\beta_1\to\infty} \Pr\{\eta_M \leq x\} = \lim \Pr\{\eta_M^* \leq x\}$$

$$\text{where} \quad \eta_M^* = \frac{(\sqrt{T})(u_2)}{[\hat{\sigma}^2 c_{22}]^{\frac{1}{2}}} \quad . \quad (19)$$

But $\sqrt{T}(u_2) \sim N(0,\sigma^2 c_{22})$ so that $\eta_M^*$ has a central t-distribution with (T-K) = (T-3) degrees of freedom. Therefore so does $\eta_M$ in the limit as $\beta_1 \to \infty$ with $\beta_1\beta_2 = 1$ and thus $W_M$ converges in distribution to an F(1,T-3) as $\beta_1 \to \infty$ with $\beta_1\beta_2 = 1$. This result holds true as $\beta_1 \to \infty$ with $\beta_1\beta_2 = a$ for any fixed $a \neq 0$. Intuitively, the test becomes one of $\beta_2 = 0$, as $\beta_1 \to \infty$ when $\beta_1\beta_2 = a$ (fixed $a \neq 0$), and $\eta_M^*$ is the t-statistic for testing that $\beta_2 = 0$.

A similar argument can be made for the limit as $\beta_2 \to \infty$ with $\beta_1\beta_2 = a$ (fixed $a \neq 0$). Then $\eta_M^* = (\sqrt{T}(u_1)/[\hat{\sigma}^2 c_{11}]^{\frac{1}{2}}$ which is the t-statistic for testing that $\beta_1 = 0$. Now the limiting behaviour can also be investigated for $\beta_1 \to \infty$ with fixed $\beta_2$ in which case $\eta_M^* = [(\sqrt{T})(u_2 + \beta_2)]/[\hat{\sigma}^2 c_{22}]^{\frac{1}{2}}$ so that $\eta_M$ has a limiting non-central t-distribution with (T-3) degrees of freedom and non-centrality parameter $\delta = (\sqrt{T}\,\beta_2/[\sigma^2 c_{22}]^{\frac{1}{2}}$ . Therefore $W_M$ converges in distribution to $F(1,T-3,[T\,\beta_2^2/(\sigma^2 c_{22})])$ as $\beta_1 \to \infty$ with fixed $\beta_2$. Again $\eta_M^*$ is the t-statistic for testing that $\beta_2 = 0$ so that this limiting result is intuitively reasonable. Also as before a similar result is obtained when $\beta_2 \to \infty$ for fixed $\beta_1$.

This agrees reasonably well with the Monte Carlo results obtained by Gregory and Veall (1984, 1985 ) who found that $W_M$ behaves more like an F(1, T-K) than a $\chi^2(1)$ for all the parameter settings under the null hypothesis which they considered. Gregory and Veall (1984) considered experiments both for fixed X and for stochastic X matrices. The limiting arguments given above for $\beta_1 \to \infty$ with $\beta_1\beta_2 = 1$ can be extended to the stochastic X case as follows. Since the limiting distribution of $W_M$ given X as $\beta_1 \to \infty$, with $\beta_1\beta_2 = a$ (fixed $a \neq 0$), does not depend on X then when X is stochastic $W_M$ retains the same limiting distribution.

The behaviour of $W_{R1}$ as $\beta_1 \to \infty$ with $\beta_1\beta_2 = 1$ is somewhat different. The statistic for this first ratio formulation of the hypothesis can be written:

$$W_{R1} = \eta_{R1}^2 \quad \text{where}$$

$$\eta_{R1} = \frac{\sqrt{T}(\hat{\beta}_1\hat{\beta}_2 - 1)\hat{\beta}_2}{\{\hat{\sigma}^2[\beta_2^4 c_{11} + 2\beta_2^2 c_{12} + c_{22}]\}^{\frac{1}{2}}} . \tag{20}$$

Again this can be split into three parts:

(a) $(\hat{\sigma}^2)^{\frac{1}{2}}$ which is independent of the rest of the statistic;

(b) $\sqrt{T}(\hat{\beta}_1\hat{\beta}_2 - 1)\hat{\beta}_2$

$$= \sqrt{T}[u_2 + \beta_2^2 u_1 + 2\beta_2 u_1 u_2 + \beta_1 u_2^2 + u_1 u_2^2] \text{ ; and}$$

(c) $[\beta_2^4 c_{11} + 2\beta_2^2 c_{12} + c_{22}]^{\frac{1}{2}}$

$$= [(\beta_2 + u_2)^4 c_{11} + 2(\beta_2 + u_2)^2 c_{12} + c_{22}]^{\frac{1}{2}}$$

$$= [u_2^4 c_{11} + 4u_2^3\beta_2 c_{11} + u_2^2(6\beta_2^2 c_{11} + 2c_{12}) + u_2(4\beta_2^3 c_{11} + 4\beta_2 c_{12}) + (\beta_2^4 c_{11} + 2\beta_2^2 c_{12} + c_{22})]^{\frac{1}{2}}.$$

It is now helpful to consider $(\eta_{R1}/\beta_1)$ rather than $\eta_{R1}$ itself. As $\beta_1 \to \infty$ then $\gamma^* = (1/\beta_1) \to 0$ and $\beta_2 \to 0$ for $\beta_1\beta_2 = 1$ so that applying Cramer's theorem gives:

$$\lim_{\beta_1 \to \infty} \Pr\{\eta_{R1}/\beta_1) \le x\} = \Pr\{(\eta_{R1}/\beta_1)^* \le x\} \qquad (21)$$

where $$(\eta_{R1}/\beta_1)^* = \frac{\sqrt{T}\, u_2^2}{\{(\sigma^2)[u_2^4 c_{11} + 2u_2^2 c_{12} + c_{22}]\}^{\frac{1}{2}}}$$

which has a well-behaved distribution with no probability mass at zero. Since $(\eta_{R1}/\beta_1)$ has a well behaved limiting distribution as $\beta_1 \to \infty$ with $\beta_1\beta_2 = 1$ then $\eta_{R1}$ has an exploding distribution i.e. $\Pr\{\eta_{R1} \le t\} \to 0$ as $\beta_1 \to \infty$. As before this argument (with slight modifications) also holds when $\beta_1 \to \infty$ with $\beta_1\beta_2 = a$ for fixed $a \neq 0$ and when $\beta_1 \to \infty$ with $\beta_2$ fixed. However, for fixed $\beta_1$ and $\beta_2$ with $\beta_1\beta_2 = 1$ then as $T \to \infty$ the statistic $W_{R1}$ tends to a central chi-squared distribution with one degree of freedom. This agrees very well with the qualitative aspects of the Monte Carlo results for $W_{R1}$ given in Gregory and Veall (1984, 1985 ).

The behaviour of $W_{R1}$ as $\beta_2 \to \infty$ with $\beta_1\beta_2 = 1$ is like that of $W_M$ as $\beta_2 \to \infty$ with $\beta_1\beta_2 = 1$. Thus, rewriting $W_{R1}$ gives:

$W_{R1} = \eta_{R1}^2$ where

$$\eta_{R1} = \frac{\sqrt{T}[(\hat{\beta}_1\hat{\beta}_2 - 1)\hat{\beta}_2/(\beta_2^2)]}{\{\hat{\sigma}^2[c_{11}(\hat{\beta}_2/\beta_2)^4 + 2(\hat{\beta}_2^2 c_{12}/\beta_2^4) + (c_{22}/\beta_2^4)]\}^{\frac{1}{2}}} . \qquad (22)$$

Again the $\hat{\sigma}^2$ term can be temporarily set aside and the statistic split into:

(a) $\sqrt{T}[(\hat{\beta}_1\hat{\beta}_2 - 1)\hat{\beta}_2/\beta_2^2]$

$= \sqrt{T}[(u_2/\beta_2^2) + u_1 + 2(u_1u_2/\beta_2) + (u_2^2\beta_1/\beta_2^2) + (u_1u_2^2/\beta_2^2)]$

(b) $[c_{11}(\hat{\beta}_2/\beta_2)^4 + 2c_{12}(\hat{\beta}_2^2/\beta_2^4) + (c_{22}/\beta_2^4)]^{\frac{1}{2}}$

$= [u_2^4(c_{11}/\beta_2^4) + 4u_2^3(c_{11}/\beta_2^3) + u_2^2(6c_{11}/\beta_2^2) + 2u_2^2(c_{12}/\beta_2^4)$

$+ 4u_2(c_{11}/\beta_2) + 4u_2(c_{12}/\beta_2^3) + c_{11} + 2(c_{12}/\beta_2^2) + (c_{22}/\beta_2^4)]^{\frac{1}{2}}$.

Defining $\gamma^* = (1/\beta_2)$ so that $\gamma^* \to 0$ as $\beta_2 \to \infty$ with $\beta_1\beta_2 = 1$ then:

$$\eta_{R1} = \frac{\sqrt{T}[\gamma^{*2}u_2 + u_1 + 2\gamma^* u_1u_2 + u_2^2\beta_1\gamma^{*2} + u_1u_2^2\gamma^{*2}]}{\{\hat{\sigma}^2[c_{11}\gamma^{*4}u_2^4 + 4c_{11}\gamma^{*3}u_2^3 + 6c_{11}\gamma^{*2}u_2^2 + 2c_{12}\gamma^{*4}u_2^2 + 4c_{11}\gamma^* u_2 + 4c_{12}\gamma^{*3}u_2 + c_{11} + 2c_{12}\gamma^{*2} + c_{22}\gamma^{*4}]\}^{\frac{1}{2}}} . \qquad (23)$$

Applying Cramer's theorem to $\eta_{R1}$ then gives:

$$\lim_{\beta_2 \to \infty} \Pr\{\eta_{R1} \leq x\} = \Pr\{\eta_{R1}^* \leq x\} \qquad (24)$$

where $\eta_{R1}^* = \dfrac{\sqrt{T}\,\hat{u}_1}{\{\hat{\sigma}^2 c_{11}\}^{\frac{1}{2}}}$ which is the t-statistic for testing $\beta_1 = 0$.

Likewise for $\beta_2 \to \infty$ with $\beta_1\beta_2 = a$ for any fixed $a \neq 0$, and as when $\beta_2 \to \infty$ with fixed $\beta_1 \neq 0$ then $W_{R1}$ has a limiting power function which is the same as that of the t-test statistic for testing $\beta_1 \neq 0$.

Symmetry arguments imply that $W_{R2}$ as $\beta_1 \to \infty$ behaves like $W_{R1}$ as $\beta_2 \to \infty$, (i.e. like $W_M$ as $\beta_2 \to \infty$) and that $W_{R2}$ as $\beta_2 \to \infty$ behaves like $W_{R1}$ as $\beta_1 \to \infty$. Lastly, there is the statistic $W_{R3}$ which tends to explode if either $\beta_1$ or $\beta_2 \to \infty$. This can be seen from re-writing the statistic so:

$$W_{R3} = \eta_{R3}^2 \text{ where}$$

$$\eta_{R3} = \frac{\sqrt{T}[(\hat{\beta}_1\hat{\beta}_2 - 1)\hat{\beta}_2]}{\{\hat{\sigma}^2[c_{11}(\hat{\beta}_2/\hat{\beta}_1)^2 + 2c_{12}(\hat{\beta}_2/\hat{\beta}_1) + c_{22}]\}^{\frac{1}{2}}} . \qquad (25)$$

Then $(\eta_{R3}/\beta_1)$ has a non-degenerate limiting distribution as $\beta_1 \to \infty$ with $\beta_1\beta_2 = a \neq 0$ or with fixed $\beta_2$ and so $\eta_{R3}$ and $W_{R3}$ have exploding distributions as $\beta_1 \to \infty$. By symmetry they also have exploding distributions as $\beta_2 \to \infty$.

The only one of these four test statistics which is satisfactory is the first since the other three all have significance levels which tend to unity as the parameters move in certain directions within the parameter space. This property is clearly undesirable in any test statistic since it implies that for any critical value the probability of rejecting the null hypothesis when it is true can be made arbitrarily close to one by suitably choosing the parameter values.

All of the analysis above is for fixed sample sizes with varying parameter values. The usual asymptotic properties as $T \to \infty$ for fixed parameter values under the null (or local alternative sequences) also hold true. However, they provide no quantitative information as to what happens with fixed parameter values and fixed sample size T. Thus they give no indication as to how large will be the deviations of the actual distributions of $W_M$, $W_{R1}$, $W_{R2}$ and $W_{R3}$ from their common asymptotic distribution for any given parameter values, exogenous data and sample size. This suggests the possibility of using asymptotic expansions to approximate the distributions of the statistics for various circumstances.

### 4.5 Asymptotic Expansion Approximations to the Distributions of the Alternative Wald Statistics

There are various ways in which asymptotic expansion approximations to the distributions of $W_M$, $W_{R1}$, $W_{R2}$ and $W_{R3}$ can be calculated. To implement the expansion the statistics each have to be defined in an appropriate fashion as functions of appropriate random variates and constants. The first decision to be taken is that of whether X is to be stochastic or fixed. For this example we have chosen to use fixed exogenous variables since this reduces the number of random variates of which the statistics are functions and considerably simplifies the computations. If X is stochastic then each statistic is a function of the nine random variates:

$$s_0 = \frac{\Sigma y_t}{T}; \quad s_1 = \frac{\Sigma x_{1t}}{T}; \quad s_2 = \frac{\Sigma x_{2t}}{T}$$

$$s_{00} = \frac{\Sigma y_t^2}{T}; \quad s_{01} = \frac{\Sigma y_t x_{1t}}{T}; \quad s_{02} = \frac{\Sigma y_t x_{2t}}{T}$$

$$s_{11} = \frac{\Sigma x_{1t}^2}{T}; \quad s_{12} = \frac{\Sigma x_{1t} x_{2t}}{T}; \quad s_{22} = \frac{\Sigma x_{2t}^2}{T}$$

since $\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 1 & s_1 & s_2 \\ s_1 & s_{11} & s_{12} \\ s_2 & s_{12} & s_{22} \end{bmatrix}^{-1} \begin{bmatrix} s_0 \\ s_{01} \\ s_{02} \end{bmatrix}$

and $\tilde{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/T = [(y'y) - (y'X)(X'X)^{-1}(X'y)]/T$

$$= \left\{ s_{00} - (s_0, s_{01}, s_{02}) \begin{bmatrix} 1 & s_1 & s_2 \\ s_1 & s_{11} & s_{12} \\ s_2 & s_{12} & s_{22} \end{bmatrix}^{-1} \begin{pmatrix} s_0 \\ s_{01} \\ s_{02} \end{pmatrix} \right\} . \qquad (26)$$

Each statistic must be differentiated 213 times (although many of the derivatives will be zero at the evaluation point) whereas in the fixed X case only 34 differentiations of each statistic are needed (again many derivatives are zero). Computing the cumulants of the variates is therefore much more complex in the stochastic X case. Firstly, there are considerably more cumulants to be computed; and secondly, under stochastic X six of the nine random variates are data second order sample moments ($s_{00}$, $s_{01}$, $s_{02}$, $s_{11}$, $s_{12}$, $s_{22}$) whereas under fixed X there is only one ($s_{00}$). The program used to compute asymptotic expansion approximations, discussed in the Appendix on Computation, requires that the number of observations plus lags all multiplied by the number of endogenous variables does not exceed eighty. This would impose an upper sample size of twenty-five assuming that the {x} variables are generated by a first-order vector auto-regressive process as in Gregory and Veall (1984, 1985 ). Of their results only those for sample size 20 would be comparable; those for sample sizes 30, 50, 100 and 500 would clearly not be comparable.

In view of the restricted comparability of the stochastic regressors case and computational cost that would be involved, we will only consider the fixed regressors case. For this case each statistic is a function of only four data sample moments of which three are data first-order sample moments and one is a data second-order sample moment. This considerably reduces the number of derivatives and cumulants which are required as noted above. It also permits sample sizes of 20, 30 and 50 to be considered. However, some compatability is lost because the particular fixed regressor matrices used become important in determining the distributions of the test statistics.

In order to compute asymptotic expansion approximations (AEA's) for the distributions of $W_M$, $W_{R1}$, $W_{R2}$ and $W_{R3}$ we have used the program ESSACS which was developed to implement the theoretical approach of Chapter 3. The exogenous variables were generated using the same data generation process as that in Gregory and Veall (1984, 1985)) although the algorithm used was not the same :

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.3 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$$

where $\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \sim IN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$ .

This does not provide exact comparability with Gregory and Veall's results since in the fixed regressors case the actual values of the exogenous variables are important. We considered sample sizes $T = 20, 30, 50$, and parameter values $(\beta_1,\beta_2) = (1.0, 1.0), (2.0, 0.5), (5.0, 0.2), (10.0, 0.1)$, $\sigma^2 = 1$ and $\beta_o = 1.0$ to retain some comparability with Gregory and Veall's results. The results most comparable to those of Gregory and Veall are presented in Table 4.1. It is clear that the AEA's are reasonably well behaved for the multiplicative form. However, they are much less well behaved for the ratio form except for $(\beta_1,\beta_2) = (1.0, 1.0)$. It is clear when the ratio form has true rejection probabilities substantially different from the nominal 5% level that the AEA rejection probabilities also differ substantially from the nominal level. However, the deviations are in the opposite direction to that indicated by Gregory and Veall's Monte Carlo results and by the arguments presented in Section 4.4 concerned with asymptotics in the parameters $\beta_1$ and $\beta_2$.

| Parameter Values | Form of Restriction | T = 20 | T = 30 | T = 50 |
|---|---|---|---|---|
| $\beta_1 = 10.0, \beta_2 = 0.1$ | A | -6.77 | -4.17 | -1.25 |
| | | (29.3) | (25.3) | (20.6) |
| | B | 6.38 | 5.92 | 5.55 |
| | | (6.5) | (6.2) | (6.3) |
| $\beta_1 = 5.0, \beta_2 = 0.2$ | A | 0.38 | 3.82 | 4.18 |
| | | (20.1) | (15.2) | (11.9) |
| | B | 6.36 | 5.91 | 5.54 |
| | | (6.4) | (5.8) | (5.7) |
| $\beta_1 = 2.0, \beta_2 = 0.5$ | A | 4.00 | 4.57 | 4.63 |
| | | (8.6) | (8.9) | (7.8) |
| | B | 6.00 | 5.72 | 5.40 |
| | | (6.1) | (5.3) | (7.1) |
| $\beta_1 = 1.0, \beta_2 = 1.0$ | A | 5.60 | 5.00 | 4.96 |
| | | (5.3) | (4.5) | (5.3) |
| | B | 6.84 | 6.33 | 5.82 |
| | | (6.9) | (4.7) | (6.5) |

Table 4.1: Test Statistic Rejection Probabilities at the 5% Critical Value for a $\chi^2(1)$ Distribution

Here A refers to the First Ratio Form of the restriction : $h_{R1}(\cdot)$.
B refers to the Multiplicative Form of the restriction : $h_M(\cdot)$.

The rejection probabilities are presented as %'s and the figures given in brackets refer to the comparable % rejections given by Gregory and Veall (1985).

| Parameter Values | Form of Restriction | T = 20 | T = 30 | T = 50 |
|---|---|---|---|---|
| $\beta_1 = 10.0,\ \beta_2 = 0.1$ | A | 44.41 | 19.52 | 7.67 |
| | | (28.3) | (24.4) | (20.1) |
| | B | 4.73 | 4.88 | 4.96 |
| | | (4.8) | (5.3) | (5.9) |
| $\beta_1 = 5.0,\ \beta_2 = 0.2$ | A | 15.60 | 9.03 | 6.00 |
| | | (18.9) | (14.6) | (11.6) |
| | B | 4.70 | 4.87 | 4.95 |
| | | (4.3) | (4.8) | (5.1) |
| $\beta_1 = 2.0,\ \beta_2 = 0.5$ | A | 4.47 | 4.53 | 4.42 |
| | | (6.7) | (8.2) | (7.3) |
| | B | 4.37 | 4.69 | 4.82 |
| | | (4.2) | (4.3) | (5.8) |
| $\beta_1 = 1.0,\ \beta_2 = 1.0$ | A | 4.03 | 4.03 | 4.40 |
| | | (4.0) | (3.2) | (4.8) |
| | B | 5.20 | 5.30 | 5.23 |
| | | (5.1) | (3.9) | (5.5) |

Table 4.2: Test Statistic Rejection Probabilities at the 5% Critical Values for $F(1,T-3)$ Distributions

The results in brackets refer to the comparable probabilities estimated by Monte Carlo methods in Gregory and Veall (1984).

For $T = 20$ : $c_{11} = 0.5903$, $c_{12} = -0.1441$, $c_{22} = 0.3951$.

For $T = 30$ : $c_{11} = 0.5046$, $c_{12} = -0.1598$, $c_{22} = 0.4729$.

For $T = 40$ : $c_{11} = 0.5986$, $c_{12} = -0.1992$, $c_{22} = 0.4464$.

| | Nominal Significance Level | | T=20 | T=30 | T=50 | T=20 | T=30 | T=50 | T=20 | T=30 | T=50 | T=20 | T=30 | T=50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $W_M$ | 5% | $X^2$ | 6.84 | 6.33 | 5.82 | 6.00 | 5.72 | 5.40 | 6.36 | 5.91 | 5.54 | 6.38 | 5.92 | 5.55 |
| | | F | 5.20 | 5.30 | 5.23 | 4.37 | 4.69 | 4.82 | 4.70 | 4.87 | 4.95 | 4.73 | 4.88 | 4.96 |
| | 1% | $X^2$ | 2.11 | 1.81 | 1.51 | 1.50 | 1.36 | 1.20 | 1.70 | 1.47 | 1.28 | 1.71 | 1.47 | 1.28 |
| | | F | 1.06 | 1.17 | 1.17 | 0.66 | 0.82 | 0.90 | 0.78 | 0.90 | 0.97 | 0.78 | 0.90 | 0.97 |
| $W_{R1}$ | 5% | $X^2$ | 5.60 | 5.00 | 4.96 | 4.00 | 4.57 | 4.63 | 0.38 | 3.82 | 4.18 | -6.77 | -4.17 | -1.25 |
| | | F | 4.03 | 4.03 | 4.40 | 4.47 | 4.53 | 4.42 | 15.60 | 9.03 | 6.00 | 44.41 | 19.52 | 7.67 |
| | 1% | $X^2$ | 1.32 | 1.05 | 1.01 | 5.82 | 4.89 | 3.63 | 37.32 | 27.75 | 19.26 | 139.39 | 103.91 | 71.11 |
| | | F | 0.57 | 0.61 | 0.74 | 5.23 | 4.58 | 3.46 | 35.56 | 27.89 | 19.52 | 135.94 | 106.41 | 73.16 |
| $W_{R2}$ | 5% | $X^2$ | 3.77 | 4.82 | 4.71 | 6.38 | 5.92 | 5.55 | 6.38 | 5.93 | 5.56 | 6.39 | 5.93 | 5.56 |
| | | F | 2.49 | 3.86 | 4.16 | 4.72 | 4.88 | 4.96 | 4.73 | 4.88 | 4.97 | 4.73 | 4.88 | 4.97 |
| | 1% | $X^2$ | 0.69 | 0.98 | 0.91 | 1.71 | 1.47 | 1.28 | 1.71 | 1.47 | 1.28 | 1.71 | 1.47 | 1.28 |
| | | F | 0.34 | 0.57 | 0.67 | 0.78 | 0.90 | 0.97 | 0.78 | 0.90 | 0.97 | 0.78 | 0.90 | 0.97 |
| $W_{R3}$ | 5% | $X^2$ | 5.23 | 5.31 | 5.10 | 4.79 | 4.95 | 4.87 | 2.62 | 3.14 | 3.66 | -8.22 | -4.97 | -1.86 |
| | | F | 4.17 | 4.49 | 4.60 | 5.12 | 4.86 | 4.64 | 14.08 | 8.24 | 5.43 | 42.64 | 18.58 | 7.00 |
| | 1% | $X^2$ | 2.46 | 1.99 | 1.61 | 6.03 | 4.94 | 3.62 | 35.50 | 26.72 | 18.47 | 137.32 | 102.75 | 70.22 |
| | | F | 1.74 | 1.51 | 1.34 | 5.27 | 4.57 | 3.43 | 34.06 | 26.96 | 18.76 | 134.25 | 105.36 | 72.32 |

Table 4.3 : Complete Approximate Rejection Probability Results Expressed as %'s

Table 4.2 presents results for the F(1,T-3) 5% critical value. Again the results are qualitatively useful except for the case $(\beta_1,\beta_2) = (2.0, 0.5)$, and for the cases T = 50, $(\beta_1,\beta_2) = (10.0, 0.1)$, (5.0, 0.2). One immediately apparent feature is that for the ratio form the AEA to the cdf of the statistic slopes the wrong way in the region of the $\chi^2(1)$ and F(1,T-3) 5% critical values. The F(1,T-3) critical value is higher than the $\chi^2(1)$ critical value so the rejection probability associated with the former should be smaller than that associated with the latter. Table 4.3 presents the complete results comparable to those of Gregory and Veall (1984).

It is important not to rely too heavily on the results for any given $(\beta_1, \beta_2, T)$ combination in drawing conclusions. Gregory and Veall (1984) used the same replications for each of the different statistics given the $(\beta_1, \beta_2, T)$ values and for each $(\beta_1, \beta_2, T)$ combination they generated a particular set of exogenous variables .

## 4.6 Conclusions

Non-linear parameter restrictions typically give rise to test statistics whose exact finite sample distributions can differ quite substantially from their asymptotic distributions. Furthermore, the use of these non-linear restrictions can give rise to various different Wald test statistics which are numerically different and have differing finite sample properties. Applying asymptotic expansion methods to obtain approximations to the cdf's of such alternative Wald statistics seems to provide qualitatively useful information in the particular example considered by Gregory and Veall (1985). However, it does not seem to provide very useful quantitative information in this example. In this particular case, the asymptotic expansion approximations support the conclusion from Gregory and Veall's Monte Carlo study that the Wald test from the multiplicative form of the hypothesis behaves better as regards its size properties, i.e. its rejection probabilities under the null hypothesis, than do the Wald tests for the three suggested ratio forms of the hypothesis. However, this conclusion can be obtained from the limiting behaviour of the various test statistics under certain parameter sequences with the samole size fixed. The only additional information given by the asymptotic expansion approximations is a very rough indication of the extent of the approximation error from using the asymptotic cdf as an approximation to the actual cdf.

## 4.A Appendix to Chapter 4

The results presented in Section 4.5 appear at first sight to be dependent on the parameters and exogenous data. However, it is possible to make various invariance arguments which reduce the dimensionality of the parameters determining the distributions of the various test statistics considered. It might also appear at first sight that the model is more complicated than is necessary to consider this problem in that the model contains an intercept. In fact, this is not a real problem since the families of test distributions are identical whether there is an intercept or not.

The first invariance argument is that multiplying $\beta_o, x_{1t}, x_{2t}, \epsilon_t$ and $y_t$ by the same constant k does not change the distributions of the test statistics:

$$
\begin{aligned}
y_t^* = ky_t &= k\beta_o + k\beta_1 x_{1t} + k\beta_2 x_{2t} + k\epsilon_t \\
&= \beta_o^* + \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \epsilon_t^*
\end{aligned}
$$

where $\epsilon_t^* \sim IN(0, k^2\sigma^2)$.

In this transformed model $\hat{\beta}_1^* = \hat{\beta}_1$, $\hat{\beta}_2^* = \hat{\beta}_2$, $\hat{\sigma}^2 = k^2\hat{\sigma}^2$, $c_{11}^* = k^{-2}c_{11}$, $c_{12}^* = k^{-2}c_{12}$, $c_{22}^* = k^{-2}c_{22}$ so that $W_M^* = W_M$ (and likewise for the other Wald test statistics) so that the distributions of the test statistics are unaltered. The second invariance argument is that transforming the model to:

$$
y_t = \beta_o + k\beta_1 k^{-1} x_{1t} + k^{-1}\beta_2 k x_{2t} + \epsilon_t
$$

or $y_t = \beta_0 + \beta_1^+ x_{1t}^+ + \beta_2^+ x_{2t}^+ + \varepsilon_t$

(where $\beta_1^+ = k\beta_1$, $x_{1t}^+ = k^{-1}x_{1t}$, $\beta_2^+ = k^{-1}\beta_2$ and $x_{2t}^+ = kx_{2t}$) does not change y and maintains the null hypothesis since $\beta_1\beta_2 = \beta_1\beta_2$. Furthermore:

$$c_{11}^+ = k^{-2}c_{11} \;;\; c_{12}^+ = c_{12} \;;\; c_{22}^+ = k^2 c_{22} \;;$$
$$\hat{\beta}_1^+ = k\hat{\beta}_1 \;;\; \hat{\beta}_2^+ = k^{-1}\hat{\beta}_2 ;\; \hat{\sigma}^2 = \hat{\sigma}^2 \;;$$

so
$$(\hat{\beta}_1^+)^2 c_{22}^+ = \hat{\beta}_1^2 c_{22} \;;\; \hat{\beta}_1^+\hat{\beta}_2^+ c_{12}^+ = \hat{\beta}_1\hat{\beta}_2 c_{12} \;;$$
$$(\hat{\beta}_2^+)^2 c_{11}^+ = \hat{\beta}_2^2 c_{11} \;;\; (\hat{\beta}_1^+)^{-2} c_{11}^+ = (\hat{\beta}_1)^{-2} c_{11} \;;$$
$$(\hat{\beta}_1^+\hat{\beta}_2^+)^{-1} c_{12}^+ = (\hat{\beta}_1\hat{\beta}_2)^{-1} c_{12} ;\; (\hat{\beta}_2^+)^{-2} c_{22}^+ = (\hat{\beta}_2)^{-2} c_{22} ;$$

and therefore $W_M^+ = W_M$ (and likewise for the other test statistics). These two invariance arguments mean that of the seven parameters ($\beta_1, \beta_2, \sigma^2, c_{11}, c_{12}, c_{22}, T$) two are effectively redundant, e.g. we could work with ($\beta_1^2 c_{22}, \beta_1\beta_2 c_{12}, \beta_2^2 c_{11}, \sigma^2, T$) instead. Also the AEA's are only being obtained for cases where the null hypothesis holds, i.e. $\beta_1\beta_2 = 1$. Therefore another possibility would be to pre-set $\beta_1 = \beta_2 = \sigma^2 = 1$ and work with ($c_{11}, c_{12}, c_{22}, T$). Considering things in this form is useful because $c_{11}c_{22} - c_{12}^2 > 0$ must hold for ($c_{11}, c_{12}, c_{22}$) to come from a covariance matrix (and also $c_{11} > 0$, $c_{22} > 0$). In addition with $\beta_1 = \beta_2 = 1$ $W_M$ is symmetric in ($c_{11}, c_{22}$) since the ordering of the (x) variables is unimportant, $W_{R1}$ with ($c_{11} = a$, $c_{22} = b$) is the same as $W_{R2}$($c_{11} = b$, $c_{22} = a$) and $W_{R3}$ is like $W_M$ symmetric in ($c_{11}, c_{22}$). So attention can be restricted to having $c_{11} \geq c_{22}$ and $c_{11}c_{22} - c_{12}^2 > 0$.

As regards the intercept term this only really enters through the degrees of freedom of the variance estimator. So for example the first Wald statistic is:

$$W_M = \frac{[\hat{\beta}_1\hat{\beta}_2 - 1]^2}{\hat{\sigma}^2(\hat{\beta}^2 d_{22} + 2\hat{\beta}_1\hat{\beta}_2 d_{12} + \hat{\beta}_2 d_{11})}$$

$$\text{with } \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \sigma^2 \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}\right].$$

In both the model with the intercept and the model without the intercept $(d_{11}, d_{12}, d_{22})$ can be set to any desired values such that $d_{11}, d_{22} > 0$, $(d_{11}d_{22} - d_{12}^2) > 0$. The only difference is that $\hat{\sigma}^2$ has (T-3) degrees of freedom (i.e. $(T-3)\hat{\sigma}^2 \sim \chi^2(T-3)$) in the intercept model and (T-2) degrees of freedom in the model with no intercept. However, all that is necessary is to increase T to (T+1) in the intercept model whilst using a new set of exogenous data such that $(d_{11}, d_{12}, d_{22})$ remain unchanged. Clearly then, since $\hat{\sigma}^2$ is independent of $(\hat{\beta}_1, \hat{\beta}_2)$ the distribution of $W_M$ in the intercept model, with sample size (T+1) and the new exogenous data, is the same as the distribution of $W_M$ in the non-intercept model, with sample size T and the old exogenous data. Hence the family of distributions for the intercept and the non-intercept model are identical (and this is also true for $W_{R1}$, $W_{R2}$, $W_{R3}$).

# CHAPTER 5: Predictive Failure Tests in the AR(1) Model

## 5.1 Introduction

Predictive failure and structural change tests are now frequently used in applied econometric studies. The properties of such tests have been extensively studied in static models by Anderson and Mizon (1983) and Rugaimukamu (1982) but to date little analysis has been made of their properties in dynamic models. In this chapter we investigate the size properties of two of these tests in the AR(1) model using the asymptotic expansion techniques developed in Chapter 3. We have chosen the AR(1) model since it is the simplest dynamic model in which it is possible to investigate the properties of these tests.

In Section 5.2 we discuss the general background to predictive failure tests. In Section 5.3 we review the results which are available on forecast errors and predictive failure tests within the AR(1) model and more general dynamic models. We then consider what these results may imply about predictive failure tests in the AR(1) model. In Section 5.4 we examine the size properties of two particular types of predictive failure test using asymptotic expansion approximations and also the extent to which these results can be explained using the arguments put forward in Section 5.3. Finally, in Section 5.5 we consider what conclusions can be drawn from this study and also some possible directions for future research.

## 5.2 The Use of Predictive Failure Tests

### 5.2.1 Parameter Constancy and Predictive Failure Tests

One of the main objectives of econometric model building is to provide accurate forecasts of economic variables of interest. It is not surprising, therefore, that the accuracy of forecasts from econometric models has been used to compare and assess these models. Thus, Cooper (1972) and Nelson (1972) compared the accuracy of forecasts from large macro-econometric models with that of forecasts from simple time series models and found that the forecasts from the time series models often outperformed those from the large macro-econometric models which supposedly incorporated much more information.

As well as comparing models, it is also desirable to be able to assess how well a model performs within its own frame of reference, and this is generally formalised by the use of predictive failure tests, as discussed in Hendry (1980). These tests have been developed from tests of parameter constancy and structural change. Suppose that we consider the following linear regression model:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} , \qquad (1)$$

(or $y = X\beta + u$) where $y_j$ is a $T_j$-vector of observations on the endogenous variable for the j'th sub-period (j=1,2), $X_j$ is a $(T_j{\times}K)$ matrix of non-stochastic regressors for the j'th sub-period, $\beta_j$ is a K-vector of regression coefficients for the j'th sub-period and $E(u) = 0, E(uu')=\sigma^2 I_T$

where $I_T$ is a (T×T) identity matrix with $T = T_1 + T_2$, i.e. the total number of observations, and u is normally distributed.

Then for $T_1 > K$ and $T_2 > K$, the hypothesis that the regression coefficients are the same in both sub-periods, i.e. $H_0 : \beta_1 = \beta_2$, can be tested by using a standard F-test:

$$F = \frac{[(\tilde{e}'\tilde{e}) - (\hat{e}'\hat{e})]/K}{(\hat{e}'\hat{e}) \,/\, (T-2K)} ,$$

where $\hat{e}$ are the unrestricted OLS residuals and $\tilde{e}$ are the residuals from the regression subject to $H_0$. This test statistic is distributed as F(K, T-2K) because there are K restrictions and 2K regression coefficients. However, if $T_2 \leq K$ then the residuals for the second sub-period from the unrestricted regression will all be equal to zero so that $(\hat{e}'\hat{e})/\sigma^2$ will be distributed as a central chi-square variate with $(T_1 - K)$ degrees of freedom rather than (T-2K) degrees of freedom. Furthermore, it can be shown that $[(\tilde{e}'\tilde{e})-(\hat{e}'\hat{e})]/\sigma^2$ is distributed as a chi-square variate with $T_2$ degrees of freedom and is independent of $(\hat{e}'\hat{e})/\sigma^2$; see Johnston (1984, Chapter 10). Therefore, when $T_2 \leq K$, the F statistic:

$$W^c = \frac{[(\tilde{e}'\tilde{e}) - (\hat{e}'\hat{e})] \,/\, T_2}{(\hat{e}'\hat{e}) \,/\, (T_1-K)}$$

is distributed as an $F(T_2, T_1 - K)$ variate. Chow (1960) showed that this statistic could be alternatively expressed as:

$$W^c = \frac{(y_2 - X_2\hat{\beta}_1)'[I_2 + X_2(X_1'X_1)^{-1}X_2']^{-1}(y_2 - X_2\hat{\beta}_1)}{T_2\hat{\sigma}_1^2}$$

where $(\hat{\beta}_1, \hat{\sigma}_1^2)$ are the OLS estimates of $(\beta_1, \sigma^2)$ from the regression for the first sub-period. In simpler notation this can be written as:

$$W^c = \frac{(\hat{f}_2' V^{-1} \hat{f}_2)/T_2}{\hat{\sigma}_1^2} \qquad (2)$$

where $\hat{f}_2 = (y_2 - X_2\hat{\beta}_1)$, the prediction errors for the second sub-period using the estimates for the first sub-period, and $V = [I_2 + X_2(X_1'X_1)^{-1}X_2']$ with $I_2$ being a $(T_2 \times T_2)$ identity matrix. Thus $W^c$ which is the Chow (1960) test for structural change can be expressed as a predictive failure test when $T_2 \leq K$ and this form of the test can easily be extended to $T_2 > K$.

The analysis above is for the case of non-stochastic regressors. If the regressors are stochastic but independent of the disturbances then under the null hypothesis of no structural change the test statistic has the same central F distribution as in the non-stochastic regressors case. Under the alternative hypothesis the marginal distribution of the statistic is in general no longer an F-distribution. However, the distribution of the test statistic conditional upon the regressors is still an F-distribution. In models with lagged dependent regressors, such as the AR(1) model considered later in this chapter, the test statistic does not even have an F-distribution under the null hypothesis; however, the test statistic when multiplied by $T_2$ does have an asymptotically chi-square distribution with $T_2$ degrees of freedom for fixed $T_2$ as $T_1 \to \infty$. This provides the justification for a simplified predictive failure test:

$$W^h = \frac{(\hat{f}_2'\hat{f}_2)}{\hat{\sigma}_1^2} \qquad (3)$$

which is asymptotically equivalent to $T_2W^c$ if $\text{plim}(X_1'X_1)^{-1} = 0$ as $T_1 \to \infty$ so that $\text{plim}\, V^{-1} = I_2$. This simplified predictive failure test was suggested in Hendry (1974). As is noted in Kiviet (1986), the Hendry test statistic is always greater than or equal to $T_2$ x the Chow test statistic. Since the critical values for an F-distributed variate are always greater than the corresponding critical values for the appropriately scaled chi-square variate (with the same number of numerator degrees of freedom), it is clear that the rejection probability of the Hendry test using chi-square critical values is necessarily greater than or equal to that of the Chow test using F distribution critical values. However, it is not clear on inspection for a dynamic model whether the actual significance level of the Hendry test is closer to or further from its nominal significance level than is that of the Chow test.

These tests for structural change have explicit null and alternative hypotheses but the predictive failure test interpretation suggests that they may be useful for testing against a wider range of alternatives than that specified above. Using the latter interpretation we can see that the two tests will tend to reject more frequently the further the expectation of the forecast errors, $E(\tilde{f}_2)$, deviates from zero. They will also tend to reject more frequently the larger the variance of the disturbances in the second sub-period is compared to that of the disturbances in the first sub-period. Thus, the two tests will tend to pick up model mis-specification provided that the mis-specification causes either or both of the above effects.

However, as Hendry (1980) points out, in a stationary environment model mis-specification will not typically be picked up by these predictive failure tests. This can be easily explained by noting that even when the model is mis-specified we are still estimating a conditional distribution, and given the stationarity assumption this conditional distribution will not change when we move from the first to the second sub-period. Essentially, in a stationary environment the model mis-specification appears through an increased residual variance and thus the model will typically forecast within the accuracy limits provided by its estimated variance, and so the test will not be consistent.

Thus for a predictive failure test to have a reasonable probability of showing model mis-specification, the conditional distribution being estimated must not be time invariant (so that the true data generation process cannot be stationary). This can occur, as suggested by Hendry (1980), either because there has been a genuine structural change or because the model has been mis-specified and there has been a change in the behaviour of the variables which are relevant to the mis-specification.

It can be noted that if the true data generation process is dynamic then certain kinds of mis-specification, such as omitted variables which are serially correlated, may well give rise to residual autocorrelation in the estimated model; see Hendry and Mizon (1978). If this is also reflected in the forecasts, then it will affect the distributions of predictive failure tests. In particular, multi-period predictive failure tests will not typically have asymptotically chi-square distributions even when appropriately scaled. Thus, although they may not be consistent tests, they may still have some power against certain alternatives.

In summary then, predictive failure tests formalise an important criterion for assessing econometric models, namely their forecasting performance. However, predictive failure tests can only test the performance of a model within its own accuracy limits and thus they are not very useful for comparing different models. The non-rejection of a model on the basis of a predictive failure test is quite compatible with substantial mis-specification of that model. Nevertheless, it seems worthwhile to investigate the size performance of predictive failure tests since only their asymptotic sizes are usually available and it is desirable to know how the tests behave when the model is correctly specified.

### 5.2.2 Predictive Failure Tests in the AR(1) Model

The main model with which we will be concerned is the pure AR(1) model which takes the form:

$$y_i = \beta y_{i-1} + \varepsilon_i , \quad |\beta| < 1 ; \quad i = 1,2,\ldots,n, \qquad (4)$$

where $\{\varepsilon_i\}$ are distributed as independent $N[0,\sigma^2]$ variates. The model is stationary provided that the distribution of $y_0$ is taken to be $N[0,\sigma^2(1-\beta^2)^{-1}]$. In this chapter we use OLS estimates, based on observations for $i = 0, \ldots, n$ together with the observed post-estimation period data to construct the predictive failure tests. In the AR(1) model the OLS estimates of $(\beta,\sigma^2)$ are:

$$\hat{\beta}_n = \frac{\sum_{i=1}^{n} y_i y_{i-1}}{\sum_{i=1}^{n} y_{i-1}^2} ,$$

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{\beta}_n y_{i-1})^2}{(n - 1)} . \quad (5)$$

The forecast errors used in the predictive failure tests for the AR(1) model are defined as:

$$\begin{aligned} \hat{e}_{j|n} &= y_{n+j} - y_{j|n}, \quad (j = 1, 2, \ldots) \\ &= \epsilon_{n+j} + \beta y_{n+j-1} - \hat{\beta}_n y_{n+j-1} \\ &= \epsilon_{n+j} + (\beta - \hat{\beta}_n) y_{n+j-1} . \end{aligned} \quad (6)$$

where $y_{j|n} = \hat{\beta}_n y_{n+j-1}$ , $j = 1, 2, \ldots$ , are often referred to as unchained forecasts as opposed to chained forecasts where each forecast is based on the previous forecast. In the AR(1) model the chained forecasts are:

$$\hat{y}_{j|n} = \begin{cases} \hat{\beta}_n y_n , & j = 1 \\ \hat{\beta}_n y_{j-1|n} = (\hat{\beta}_n)^j y_n , & j = 2, 3, \ldots . \end{cases}$$

In the remaining part of this chapter we re-scale the Chow test by its numerator degrees of freedom to give it the same asymptotic distribution as the corresponding Hendry test. The Hendry and Chow predictive failure tests at time n for the one-step ahead forecasts in the AR(1) model are then defined as:

$$F^h_{1|n} = (\hat{e}_{1|n})^2/\hat{\sigma}^2_n$$

$$F^c_{1|n} = (\hat{e}_{1|n})^2/ \{[1 + (y^2_n) (\sum_{i=1}^{n} y^2_{i-1})^{-1}] \; \hat{\sigma}^2_n\} . \qquad (7)$$

In this notation the multi-period Hendry predictive failure test is easy to express:

$$F^h_{k|n} = [\sum_{j=1}^{k} \hat{e}^2_{j|n}] / \hat{\sigma}^2_n . \qquad (8)$$

However, the multi-period Chow predictive failure test is not quite so easily expressed in this notation:

$$F^c_{k|n} = \frac{(\hat{f}_2)' V^{-1} (\hat{f}_2)}{\hat{\sigma}^2_n} \qquad (9)$$

where $\hat{f}_{2,i} = \hat{e}_{i|n}$ and $V = [v_{ij}]$ where

$$v_{ij} = \delta_{ij} + (y_{n+i-1}y_{n+j-1})(\sum_{s=1}^{n} y^2_{s-1})^{-1} ,$$

$$\delta_{ij} = 1 \text{ if } i = j \text{ and } 0 \text{ otherwise.}$$

## 5.3 Properties of Forecast Errors and Predictive Failure Tests

Although there is no general theory of the properties of predictive failure tests in the AR(1) model, there is a considerable amount of work which has been done regarding the properties of forecast errors within this model. In addition, some Monte Carlo studies have been made which include examinations of the properties of predictive failure tests in certain more general dynamic models; see Kiviet (1985, 1986). These results are of value in forming conjectures about the properties of predictive failure tests in the AR(1) model which can then be examined using asymptotic expansion approximations.

### 5.3.1 Properties Relevant to One-Step Ahead Tests

A number of results have been obtained which are particularly relevant to analysing the distributions of one-step ahead predictive failure tests and therefore we will start by reviewing these.

(i) The most useful single result comes from Phillips (1979) and concerns the distribution of the one-step ahead chained forecast error conditional upon $y_n$. The result also holds for the one-step ahead unchained forecast error since the two are identical in the one-step ahead case. Phillips (1979) in his Equation (12) gives (in our notation):

$$
\begin{aligned}
&\Pr\{\dot{y}_{1|n} - y_{n+1} \leq x \mid y_n)) \\
&\quad = I(x/\sigma) + i(x/\sigma)\left[n^{-1}\{\beta/(1-\beta^2)^{\frac{1}{2}}\}\{y_n/\sigma_y\}\{3-(y_n/\sigma_y)^2\}\right. \\
&\quad\quad \left. - (2n)^{-1}(y_n/\sigma_y)^2(x/\sigma)\right] + O(n^{-2}) \,. \qquad (10)
\end{aligned}
$$

where $\sigma_y = \sigma/(1-\beta^2)^{\frac{1}{2}}$ and where $I(\cdot)$ and $i(\cdot)$ are the cdf and pdf of a standard normal variate. Following Chapter 3, Section 3.2 then we can obtain an approximation to the conditional distribution of the square of the one-step ahead forecast error:

$$\Pr\{(\hat{y}_{1|n} - y_{n+1})^2 \leq x^2 \mid y_n\}$$
$$= [1-(2n)^{-1}(y_n^2/\sigma_y^2)]\Gamma_1^2(x^2/\sigma^2)$$
$$+ (2n)^{-1}(y_n^2/\sigma_y^2)\ \Gamma_3^2(x^2/\sigma^2) + O(n^{-2}) ,$$

where $\Gamma_k^2(\cdot)$ is the cdf of a central chi-square variate with k degrees of freedom. From Section 5A.3 we know this expression is the same to $O(n^{-2})$ as the cdf of a central chi-square variate which has been multiplied by $[1 + n^{-1}(y_n^2/\sigma_y^2)]$:

$$\Pr\{(\hat{y}_{1|n} - y_{n+1})^2 \leq x^2 \mid y_n\}$$
$$= \Gamma_1^2([1 + n^{-1}(y_n^2/\sigma_y^2)]^{-1}[x^2/\sigma^2]) + O(n^{-2}) . \qquad (11)$$

In the denominator term of the Chow test an adjustment is made to take into account the sampling variation in $\hat{\beta}_n$. For the one-period test this adjustment is:

$$V = 1 + (y_n^2)\ (\sum_{i=1}^{n} y_{i-1}^2)^{-1}$$
$$= 1 + n^{-1}(y_n^2)\ (\sum_{i=1}^{n} y_{i-1}^2/n)^{-1} . \qquad (12)$$

Conditioning upon $y_n$, then plim $(\sum_{i=1}^{n} y_{i-1}^2/n) = \sigma_y^2$ so that:

$$V = 1 + n^{-1}(y_n^2/\sigma_y^2) + o_p(n^{-1})$$

and thus $V^{-1}$ appears to be an appropriate adjustment to standardise the square of the forecast error, given $y_n$.

(ii) The final term in the predictive failure tests is the error variance estimate $\hat{\sigma}_n^2$. This has not been the subject of much investigation. Orcutt and Winokur (1969) from a Monte Carlo study of the AR(1) model with an intercept term:

$$y_i = \alpha + \beta y_{i-1} + u_i \; ; \; i = 0, \pm 1, \pm 2, \ldots , \qquad (13)$$

where u is distributed as $N(0, \sigma^2 I)$, found that $E(\hat{\sigma}_n^2) \doteq \sigma^2$ and that $var(\hat{\sigma}_n^2) \doteq 2\sigma^2/(n-2)$. As these are the appropriate values for the classical linear regression model they concluded that the rescaled error variance estimate, $(n-2)\hat{\sigma}_n^2/\sigma^2$, had a distribution which was adequately approximated by that of a central chi-square variate with (n-2) degrees of freedom.

(iii) Lastly, Fuller and Hasza (1980) investigated the performance of a regression t-statistic for the one-step ahead forecast error in the Orcutt and Winokur (1969) model. They found that the two-tailed test based upon this t-statistic (which is the signed square root of the one-step ahead Chow predictive failure test for that model) had a tendency to under-reject when $|\beta|$ was close to zero and to over-reject when $|\beta|$ was close to unity, in their Monte Carlo study.

This last Monte Carlo result does not seem entirely compatible with the previous results which tend to suggest that the marginal distributions of the one-step ahead predictive failure tests should not depend, to an error of $o(n^{-1})$, on the value of $\beta$. However, it can be argued that the Fuller and Hasza (1980) and Orcutt and Winokur (1969) results are not really applicable to the pure AR(1) model since they were obtained in an AR(1) model with an intercept and it is well-known that the OLS estimators of $\beta$ have different properties in these two models, e.g. see Sawa (1978) for the exact moments of OLS estimators in these models. Also, there is nothing in the theoretical results to suggest that the true distributions of the predictive failure tests do not actually depend on $\beta$; they only seem to suggest that to an error of $o(n^{-1})$ the distributions do not depend on $\beta$.

### 5.3.2. Properties Relevant to Multi-Period Tests

There are no generalizations of the Phillips (1979) result on the one-step ahead forecast error available for the joint distribution of multiple unchained forecast errors or even for the individual r-step ahead unchained forecast error. Nonetheless, there are a number of results, some of which are related to the Phillips (1979) result, which are relevant to multi-period predictive failure tests.

(i) Phillips (1979) takes the expectation of Equation (10) with respect to $y_n$ which gives:

$$\Pr\{\hat{y}_{1|n} - y_{n+1} \leq x\} = I(x/\sigma) - (2n)^{-1}(x/\sigma)\, i(x/\sigma) ,$$

so that the unconditional distribution of $\hat{e}_{1|n}$ is approximately symmetric about zero, i.e.:

$$\Pr\{\hat{e}_{1|n} \leq x\} = \Pr\{-\hat{e}_{1|n} \leq x\} .$$

In fact the unconditional distribution is exactly symmetric about zero, as shown by Fuller and Hasza (1980), and we can extend their argument to show that the unconditional joint distribution of the forecast errors is symmetric about zero:

$$\begin{aligned}\Pr \{\hat{e}_{1|n} \leq x_1, \ldots, \hat{e}_{k|n} \leq x_k\} \\ = \Pr \{-\hat{e}_{1|n} \leq x_1, \ldots, -\hat{e}_{k|n} \leq x_k\} \text{ for } k \geq 1 . \quad (14)\end{aligned}$$

This is easily shown in the pure AR(1) model. If we define $z_i = -y_i$ and $\omega_i = -\varepsilon_i$ then:

$$z_i = \beta z_{i-1} + \omega_i \; ; \; i = 1, 2, \ldots ; \; Z_0 \sim N[0, \sigma^2/(1-\beta^2)],$$

where $\omega_i$ is distributed $IN(0,\sigma^2)$ and $(z_{n+j} - \hat{z}_{j|n}) = -(y_{n+j} - \hat{y}_{j|n})$ where $j = 1,\ldots,k$. Both sets of forecast errors, $\{(z_{n+j} - \hat{z}_{j|n})\}$ and $\{(y_{n+j} - \hat{y}_{j|n})\}$ for $j = 1,\ldots,k$, have the same distribution since the processes generating the $\{y_i\}$ and $\{z_i\}$ are identical. Therefore the forecast errors, $\{(y_{n+j} - \hat{y}_{j|n}); j=1,\ldots,k\}$, must have a symmetric distribution. This symmetry result also holds for the standardised forecast errors since transforming from $\{y_i\}$ to $\{z_i\}$ will not change the observed value of $\hat{\sigma}^2_n$.

This symmetry in the distribution of the standardised forecast errors in the AR(1) model matches the symmetry of the standardised forecast errors in the static model. The main implication of this result for asymptotic expansion approximations to the distributions of predictive failure tests is that in the expansion of the joint cdf of the single-period Hendry predictive failure tests (following Chapter 3, Equation (48)), all the $\{b_{rst}; 1 \leq r,s,t \leq k\}$ are necessarily equal to zero so that:

$$\Pr\{\tau_i^2 \le x_i^2\ ;\ i=1,\dots,k\}$$
$$= (1+b_0)\{\prod_{i=1}^{k} \Gamma_1^2(x_i^2)\}$$
$$+ \{\sum_{r=1}^{k} b_r\, \Gamma_3^2(x_r^2)\, [\prod_{i\neq r}^{k} \Gamma_1^2(x_i^2)]\}$$
$$+ \{\sum_{r<s}^{k} \sum^{k} b_{rs}\, \Gamma_3^2(x_r^2)\, \Gamma_3^2(x_s^2)\, [\prod_{i\neq r,s}^{k} \Gamma_1^2(x_i^2)]\}$$
$$+ \{\sum_{r=1}^{k} b_{rr}\, \Gamma_5^2(x_r^2)[\prod_{i\neq r}^{k} \Gamma_1^2(x_i^2)]\} + O(n^{-3/2}), \qquad (15)$$

where $\{\tau_i^2\}$ are the squares of the standardised forecast errors, i.e. the single-period Hendry tests; see Section 5.A.1 for further details.

Furthermore, if we express the model in vector form as $y = X\beta + u$ then the sampling variation adjustment term for the Chow test, $V^{-1} = [I+X_2(X_1'X_1)^{-1}X_2']^{-1}$, is invariant to changes in the sign of the $\{y_i\}$. We can decompose the Chow Statistics as $(\hat{f}_2^o)'(\hat{f}_2^o)$ where $\hat{f}_2^o = (\hat{\sigma}_n)^{-1}V^{-\frac{1}{2}}\hat{f}_2$. Since $\hat{f}_2^o$ has a symmetric distribution the asymptotic expansion approximation for the joint distribution of the squares of the elements of $\hat{f}_2^o$ can also be written in the form of Equation (15).

(ii) A rather different symmetry occurs with the joint distribution of the squares of the forecast errors, namely that their distribution is symmetric in $\beta$ so that the sign of $\beta$ does not matter. This can be seen from the following invariance argument. If we define a new series by $y_i^* = \lambda(-1)^i y_i$ where $\lambda \neq 0$ then:

$$y_i^* = \lambda(-1)^i\,[\beta y_{i-1} + \varepsilon_i]$$
$$= \beta^* y_{i-1} + \varepsilon_i^*,$$
$$\text{where } y_o^* \sim N[0,\ \lambda^2\sigma^2/(1-\beta^{*2})], \qquad (16)$$

where $\beta^* = -\beta$ and the $\{\epsilon_i^*\}$ are $IN[0, \lambda^2\sigma^2]$. The OLS estimate of $\beta^*$ is:

$$\hat{\beta}_n^* = \frac{\sum_{i=1}^{n} y_i^* y_{i-1}^*}{\sum_{i=1}^{n} (y_{i-1}^*)^2} = -\hat{\beta}_n ,$$

and so the forecast error for $y_{n+s}$ is:

$$\hat{e}_{s|n}^* = y_{n+s}^* - \hat{\beta}_n^* y_{n+s-1}^* = \lambda(-1)^{n+s}\hat{e}_{s|n}$$

giving a squared forecast error of:

$$(\hat{e}_{s|n}^*)^2 = \lambda^2(\hat{e}_{s|n})^2 . \qquad (17)$$

Thus transforming from $\{y_i\}$ to $\{y_i^*\}$ simply multiplies the values of the squared forecast errors by $\lambda^2$. If we put $\lambda = 1$ then the density of the $\{\epsilon_i^*\}$ is the same as that of the $\{\epsilon_i\}$ and so the distribution of the forecast errors in Equation (16) with $\lambda = 1$ is the same as the distribution of the forecast errors in Equation (4). However, provided that $\lambda = 1$ so that $\lambda^2\sigma^2 = \sigma^2$, the only difference between these two models is that in one model $\beta$ has a plus sign and in the other it has a minus sign. Therefore the distribution of the squared forecast errors is the same for both $\beta$ and $-\beta$ given $\sigma^2$.

The transformation from $\{y_i\}$ to $\{y_i^*\}$ simply multiplies the value of the OLS estimate of $\sigma^2$ by $\lambda^2$:

$$(\hat{\sigma}_n^*)^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (y_i^* - \hat{\beta}_n^* y_{i-1}^*)^2$$
$$= \lambda^2\hat{\sigma}_n^2 .$$

Consequently the values of the Hendry tests are unaltered since the terms involving $\lambda^2$ cancel out. Therefore the transformation from $\{y_i\}$ to $\{y_i^*\}$ will not affect the values of the Hendry tests either and so their joint distribution is the same for both $\beta$ and $-\beta$ and does not depend on $\sigma^2$. The r-period Chow test for Equation (16) is:

$$F^{C*}_{r|n} = \frac{(\hat{f}_2^*)' (V^*)^{-1} (\hat{f}_2^*)}{(\hat{\sigma}_n^*)^2} \qquad (18)$$

where $\hat{f}_2^*$ and $V^*$ are defined like $\hat{f}_2$ and V in Equation (9) but using $\{y_i^*\}$ in place of $\{y_i\}$ . However, this is equal to $F^C_{r|n}$ but where $f_2$ has been pre-multiplied by a diagonal matrix with the i'th diagonal element equal to $\lambda(-1)^{n+i}$, and where V has been pre- and post-multiplied by the same matrix. Therefore $F^{C*}_{r|n} = F^C_{r|n}$ since the diagonal matrices cancel out and consequently the distribution of the Chow tests is the same for both $\beta$ and $-\beta$ and does not depend on $\sigma^2$.

(iii) A second implication of the Phillips (1979) result is that:

$$E(\hat{e}^2_{1|n}) = \sigma^2(1 + n^{-1})$$

for $|\beta| < 1$ ; see Phillips (1979). This is not strictly justifiable from the Phillips (1979) distribution result but Fuller and Hasza (1981) prove that it is correct to an error of $O(n^{-3/2})$. We can extend this result to the general single-period s-step ahead unchained forecast error for $s \geq 2$:

$$E(\hat{e}^2_{s|n}) = \sigma^2(1 + n^{-1}) + O(n^{-3/2}) \qquad (19)$$

using the method and results of Fuller and Hasza (1981); see Section 3.A.2 for further details.

(iv) As well as considering the variances of the forecast errors in the multi-period case we also need to consider the covariances between the forecast errors. In the general case of $\hat{e}_{r|n}$ and $\hat{e}_{s|n}$ where r > s then:

$$\hat{e}_{r|n}\,\hat{e}_{s|n} = (\epsilon_{n+r}\,\epsilon_{n+s}) - (\hat{\beta}_n - \beta)\epsilon_{n+r}y_{n+s-1} \\ -(\hat{\beta}_n-\beta)y_{n+r-1}\epsilon_{n+s} + (\hat{\beta}_n - \beta)^2 y_{n+r-1}y_{n+s-1}. \qquad (20)$$

Taking expectations in Equation (20) then $E(\epsilon_{n+r}\epsilon_{n+s}) = 0$ and $E[(\hat{\beta}_n-\beta)\epsilon_{n+r}y_{n+s-1}] = 0$ since $\epsilon_{n+r}$ has mean zero and is independent of the rest of the second term. The expectation of the last term is the normal covariance term arising from the sampling variation in $\hat{\beta}_n$ which we would expect by analogy with the non-stochastic CLR model. However, unlike the CLR case, the expectation of the third term is not equal to zero because $\hat{\beta}_n$ is typically a biased estimate of $\beta$ (due to the dynamics), and $y_{n+r-1}$ and $\epsilon_{n+s}$ are now dependent (again due to the dynamics). We can expand the expectation of the third term:

$$\begin{aligned} &E[(\hat{\beta}_n-\beta)y_{n+r-1}\epsilon_{n+s}] \\ &= E[(\hat{\beta}_n-\beta)(\beta^{r-s}y_{n+s-1} + \sum_{j=0}^{r-s-1} \beta^j\epsilon_{n+r-1-j})\epsilon_{n+s}] \\ &= E[(\hat{\beta}_n-\beta)\ \beta^{r-s-1}\epsilon^2_{n+s}] \\ &= \beta^{r-s-1}\sigma^2 E(\hat{\beta}_n-\beta) \quad . \end{aligned}$$

From White (1961) we know that $E(\hat{\beta}_n-\beta) = -2\beta n^{-1} + O(n^{-2})$ so that:

$$E[(\hat{\beta}_n-\beta)y_{n+r-1}\epsilon_{n+s}] = -2\sigma^2\beta^{r-s}n^{-1} + O(n^{-2}). \qquad (21)$$

We can also expand the expectation of the final term:

$$E[(\hat{\beta}_n-\beta)^2 y_{n+r-1}y_{n+s-1}]$$
$$= E[(\hat{\beta}_n-\beta)^2(\beta^{r-s}y_{n+s-1} + \sum_{j=0}^{r-s-1} \beta^j \varepsilon_{n+r-1-j})y_{n+s-1}]$$
$$= E[(\hat{\beta}_n-\beta)^2 \beta^{r-s} y^2_{n+s-1}]$$
$$= \beta^{r-s}\beta^{2(s-1)}E[(\hat{\beta}_n-\beta)^2 y_n^2]$$
$$+ \beta^{r-s} \frac{[1-\beta^{2(s-1)}]}{(1-\beta^2)} E[(\hat{\beta}_n-\beta)^2]\sigma^2 .$$

From White (1961) we also know that:

$$E[(\hat{\beta}_n-\beta)^2] = (1-\beta^2)n^{-1} + O(n^{-2}) \tag{22}$$

and from Fuller and Hasza (1980) that:

$$E[(\hat{\beta}_n-\beta)^2 y_n^2] = \sigma^2 n^{-1} + O(n^{-3/2}). \tag{23}$$

Combining these results we obtain:

$$E[(\hat{\beta}_n-\beta)^2 y_{n+r-1}y_{n+s-1}] \tag{24}$$
$$= \beta^{r-s}[\beta^{2(s-1)} + 1 - \beta^{2(s-1)}]\sigma^2 n^{-1} + O(n^{-3/2})$$
$$= \beta^{r-s}\sigma^2 n^{-1} + O(n^{-3/2}) .$$

Therefore the expectation of the third term is approximately twice as large as the expectation of the final term. The Chow sampling variation adjustment does not take this third term into consideration since in the CLR model its expectation is necessarily zero. The relevant element of the Chow adjustment matrix is:

$$(y_{n+r-1}y_{n+s-1})(\sum_{i=1}^{n} y_{i-1}^{2})^{-1}$$

and since $\hat{\sigma}_n^2(\sum_{i=1}^{n} y_{i-1}^2/n)^{-1}$ is a consistent estimate of the variance of $n^{\frac{1}{2}}(\hat{\beta}_n-\beta)$ it is clear that the adjustment only attempts to take the final term into account. This failure of the Chow sampling variation adjustment to properly account for the covariance between forecast errors in dynamic models does not appear to have been noted before. In the AR(1) model with OLS estimates, making an appropriate modification should not be difficult since the expectation of $\hat{\beta}_n$ is known in terms of $\beta$ to an error of $O(n^{-2})$; see White (1961). One approach would be to estimate the bias and then to make an appropriate adjustment which takes the bias into consideration. A more indirect approach would be to use a bias adjusted estimator in place of the OLS estimator of $\beta$. The bias could be estimated or a bias-adjusted estimator constructed either using the formula from White (1961) or by using a variant of the jacknife technique adopted by Quenouille (1949) in the AR(1) model with an intercept. In more complicated dynamic models with more lags and with exogenous variables it may be considerably more difficult to make an appropriate adjustment.

The additional term in the covariance matrix of the r'th and s'th forecast errors is of order of magnitude $O(n^{-1})$ so that it is clearly relevant in considering the joint behaviour of the forecast errors. However as noted in Section 3.2 in Chapter 3, the off-diagonal terms of $O(n^{-1})$ in the covariance matrix of asymptotically normal variates will not affect the asymptotic expansion to $o(n^{-1})$ of the joint distribution of the squares of these variates. Thus these terms will not affect the joint distribution to $o(n^{-1})$ of the squares of the forecast errors. To examine the relationship between the squares of the forecast errors we should look at the covariance of the squares:

$$\bar{e}^2_{r|n} = \varepsilon^2_{n+r} - 2\varepsilon_{n+r}(\bar{\beta}_n-\beta)y_{n+r-1} + (\bar{\beta}_n-\beta)^2y^2_{n+r-1}\,;\ r \geq 1.$$

Therefore the covariance between the squared forecast errors is:

$$\begin{aligned}
\mathrm{cov}[\bar{e}_{r|n}, \bar{e}_{s|n}] &= \mathrm{cov}[\varepsilon^2_{n+r}, \varepsilon^2_{n+s}] \\
&- 2\mathrm{cov}[\varepsilon^2_{n+r}, \varepsilon_{n+s}(\hat{\beta}_n-\beta)y_{n+s-1}] \\
&+ \mathrm{cov}[\varepsilon^2_{n+r},(\hat{\beta}_n-\beta)^2y^2_{n+s-1}] \\
&- 2\mathrm{cov}[\varepsilon_{n+r}(\hat{\beta}_n-\beta)y_{n+r-1}, \varepsilon^2_{n+s}] \\
&+ 4\mathrm{cov}[\varepsilon_{n+r}(\hat{\beta}_n-\beta)y_{n+r-1}, \varepsilon_{n+s}(\hat{\beta}_n-\beta)y_{n+s-1}] \\
&- 2\mathrm{cov}[\varepsilon_{n+r}(\hat{\beta}_n-\beta)y_{n+r-1}, (\hat{\beta}_n-\beta)^2y^2_{n+s-1}] \\
&+ \mathrm{cov}[(\hat{\beta}_n-\beta)^2y^2_{n+r-1}, \varepsilon^2_{n+s}] \\
&-2\,\mathrm{cov}[(\hat{\beta}_n-\beta)^2y^2_{n+r-1},\varepsilon_{n+s}(\hat{\beta}_n-\beta)y_{n+s-1}] \\
&+ \mathrm{cov}[(\hat{\beta}_n-\beta)^2y^2_{n+r-1}, (\hat{\beta}_n-\beta)^2y^2_{n+s-1}] \,. \qquad (25)
\end{aligned}$$

Since $\varepsilon_{n+r}$ is independant of $\varepsilon_{n+s}$, $y_{n+r-1}$ and $y_{n+s-1}$ then all the covariance terms which involve $\mathrm{cov}[\varepsilon^2_{n+r}, \ldots]$ are equal to zero. The $\varepsilon_{n+r}$ can be extracted from the other covariances using the multiplication property of expectations under independence. Thus for example:

$$\text{cov}[\varepsilon_{n+r}(\hat{\beta}_n-\beta)y_{n+r-1}, \varepsilon^2_{n+s}]$$
$$= E(\varepsilon_{n+r})\ \text{cov}[(\hat{\beta}_n-\beta)y_{n+r-1}, \varepsilon^2_{n+s}] = 0.$$

Therefore, eliminating these terms from Equation (25) gives:

$$\text{cov}[\hat{e}^2_{r|n}, \hat{e}^2_{s|n}] = \text{cov}[(\hat{\beta}_n-\beta)^2y^2_{n+r-1}, \varepsilon^2_{n+s}]$$
$$- 2\text{cov}[(\hat{\beta}_n-\beta)^2y^2_{n+r-1}, \varepsilon_{n+s}(\hat{\beta}_n-\beta)y_{n+s-1}]$$
$$+ \text{cov}[(\hat{\beta}_n-\beta)^2y^2_{n+r-1}, (\hat{\beta}_n-\beta)^2y^2_{n+s-1}]. \qquad (26)$$

The first two terms here arise from the correlation of current y with lagged $\varepsilon$. Thus, it is only the final term in this expression which is taken into account by the sampling variation adjustment in the Chow test. We would expect the first two terms to depend on $\beta$ which suggests that both the multi-period Chow and Hendry tests have distributions which depend on $\beta$ even when terms of $o(n^{-1})$ are dropped.

(v) Lastly, there have been some Monte Carlo studies of the properties of multi-period predictive failure tests for certain rather more general dynamic models with exogenous variables; see Kiviet (1985, 1986). Kiviet (1986) considered an autoregressive distributed lag model with one lag on both the endogenous variable and the exogenous variable:

$$y_i = \gamma y_{i-1} + \beta_0 x_i + \beta_1 x_{i-1} + \omega_i, \quad |\gamma| < 1,$$

where $\omega_i$ is distributed $IN(0,\sigma^2_\omega)$. Two different processes were used to generate the exogenous variable. The first was a pure AR(1) process, and the second was a random walk with drift; various different parameter values were used for both types of processes. A number of different predictive failure tests are considered in Kiviet (1986). These include the Chow test, two versions of the Hendry test using slightly different error variance estimates, the Likelihood Ratio (LR) test statistic, and the LR test statistic with the Anderson (1958) small sample correction. Kiviet (1986) presented averaged results for m-period tests for m = 4, 8 and 20 for a number of differently specified models where the averaging is over 18 different coefficient and exogenous variable process parameter combinations.

The main conclusions were that for the stationary exogenous variable processes the Chow test performed very well under the null hypothesis whereas the Hendry test had a tendency to over-reject, particularly when the sample size was small or when a large number of future periods were used to construct the tests. Kiviet (1986) also found that the Chow test has very low power even when the exogenous variables are non-stationary and are omitted in the particular type of mis-specification considered. Kiviet (1985) considers a generalised Chow test based on instrumental variables estimation for a single equation taken from a simultaneous equation system but here the results are much less clear.

## 5.4 Asymptotic Expansion Results

The data generation processes and estimation models were both the pure AR(1) model as discussed earlier. Five different sample sizes were considered, n = (10,20,30,40,50), the variance of the disturbances was set to unity, and five different values of the autoregressive parameter were used, $\beta = (0.0, 0.2, 0.5, 0.8, 0.9)$. Since the sign of $\beta$ is unimportant these results are equally valid for $\beta = (0.0, -0.2, -0.5, -0.8, -0.9)$. Similarly, since the value of $\sigma^2$ does not matter, these results are valid for all $\sigma^2$; see Section 5.3.2.

Asymptotic expansion approximations for the one-period and two-period predictive failure tests were computed using the ESSACS program described in the Appendix on Computing. In order to do this it was necessary to decompose the test statistics as the sums of squares of asymptotically independent standard normal variates. For the Hendry test this was very simple since we could use the standardised forecast errors as discussed in Section 5.3 so that the two-period Hendry test statistic was just the sum of the Hendry test statistics for the individual periods, $F^h_{2|n} = F^h_{1|n} + F^{h+}_{2|n}$. However, for the Chow tests this was more difficult. Following Section 5.3 we defined $\hat{f}^o_2 = (\hat{\sigma}_n)^{-1} V^{-\frac{1}{2}} \hat{f}_2$ where $V^{-\frac{1}{2}}$ was the lower triangular Cholesky decomposition of $V^{-1}$ so that $F^c_{k|n} = (\hat{f}^o_2)'(\hat{f}^o_2)$, and thus $(\hat{f}^o_2)$ were the asymptotically independent standard normal variates. This decomposition has the advantage that the sub-vector $g_2$ formed by taking the first r elements of $\hat{f}^o_2$ gives the decomposition vector for the r-period Chow test statistic so $F^c_{r|n} = (g_2)'(g_2)$. Thus we can form an incremental statistic by $F^{c+}_{2|n} = F^c_{2|n} - F^c_{1|n}$.

Two sorts of results are reported. Firstly we give the numerically computed asymptotic expansion approximation coefficients, and secondly we give numerically evaluated approximate rejection probabilities at various chi-square and F-distribution critical values. The first type of result enables us to consider some of the conjectures based on the theoretical arguments from Section 5.3. The second type of result enables us to assess the relative performance of the tests and some of the conjectures based on Monte Carlo evidence.

5.4.1 Results for One-Step Ahead Tests

The asymptotic expansion approximations for the marginal cdf's of $F^{h}_{1|n}$, $F^{h+}_{2|n}$ and $F^{c}_{1|n}$ were computed to be:

$$\Pr\{F^{h}_{1|n} \leq r_1^2\} = \Pr\{F^{h+}_{2|n} \leq r_1^2\} + o(n^{-1})$$

$$= [1 - (3/4)n^{-1}]\Gamma_1^2(r_1^2) + (3/4)n^{-1}\Gamma_5^2(r_1^2) + o(n^{-1});$$

$$\Pr\{F^{c}_{1|n} \leq r_2^2\} = [1 - (1/4)n^{-1}]\Gamma_1^2(r_2^2)$$

$$- (1/2)n^{-1}\Gamma_3^2(r_2^2) + (3/4)n^{-1}\Gamma_5^2(r_2^2) + o(n^{-1}). \quad (27)$$

Various points arise immediately from examining these approximate cdf's.

(i) There are no $\Gamma_7^2(r^2)$ terms in the approximations. This is not surprising since we would expect this from the symmetry of the distributions of the vector decompositions of the Hendry and Chow predictive failure tests as demonstrated in Section 5.3.

(ii) The approximations do not depend on $\beta$. This confirms the conjecture made in Section 5.3 for $F^{h}_{1|n}$ and $F^{c}_{1|n}$.

(iii) The approximate cdf of $F^{c}_{1|n}$ is the same, to an error of $o(n^{-1})$, as the cdf of an F-variate with (1,n) degrees of freedom; see Section 5.A3 for further details. Furthermore, the cdf of $F^{h}_{1|n}$ is the same, to an error of $o(n^{-1})$, as the cdf of an F-variate with $(1,n)$ degrees of freedom multiplied by $(1+n^{-1})$; again see Section 5.A.3.

From these theoretical results we can see that the one-period predictive failure tests behave the same in the AR(1) model as they would, to an error of $o(n^{-1})$, in the CLR model were $X_2(X_1'X_1)^{-1}X_2$ equal to $n^{-1}$. This is supported by the numerical evaluations given in Table 5.1.

In Table 5.1 we report the evaluated approximate rejection probabilities for $F^h_{1|n}$ and $F^c_{1|n}$, when n=10, using chi-square and F-distribution critical values for 10%, 5% and 1% significance levels. Two sets of F-distribution critical values were used; firstly, those for an F-variate with 1 and 9 degrees of freedom since these incorporate the degrees of freedom adjustment used in practice; and secondly, those for an F-variate with 1 and 10 degrees of freedom because $F^c_{1|n}$ should be closely approximated in distribution by such a variate. The shift from 1 and (n-1) degrees of freedom to 1 and n degrees of freedom is only an order $o(n^{-1})$ alteration and thus is of smaller order of magnitude than the terms retained in the asymptotic expansion approximation. Thus the choice of degrees of freedom is somewhat arbitrary although it is clearly important for small n.

The numerical evaluations given in Table 5.1 show that the one-period Chow test for n=10 appears to behave like an F-variate with 1 and 10 degrees of freedom at the 5% and 10% significance level critical values but rather less so at the 1% significance level critical value. It appears to have a tendency to under-reject when used with the critical values for an F-variate with 1 and 9 degrees of freedom. This is noticeably weaker than its apparent tendency to

| Nominal Significance Level | | Critical Values | Hendry Test ($F^h_{1\|n}$) | Chow Test ($F^c_{1\|n}$) |
|---|---|---|---|---|
| | F(1,10) | 3.2850 | 11.39 | 9.99 |
| 10% | F(1, 9) | 3.3603 | 11.01 | 9.65 |
| | $\chi^2$ | 2.7055 | 14.84 | 13.14 |
| | F(1,10) | 4.9646 | 5.54 | 4.80 |
| 5% | F(1, 9) | 5.1174 | 5.20 | 4.51 |
| | $\chi^2$ | 3.8415 | 8.92 | 7.77 |
| | F(1,10) | 10.0440 | 0.69 | 0.61 |
| 1% | F(1, 9) | 10.5610 | 0.56 | 0.50 |
| | $\chi^2$ | 6.6350 | 2.79 | 2.42 |

Table 5.1: Approximate Rejection Probabilities Expressed as %'s for $F^h_{1|n}$ and $F^c_{1|n}$ when n = 10

over-reject when used with chi-square critical values which suggests that the choice of degrees of freedom for the F-distribution is less important than the choice between a chi-square and an F-distribution.

The Hendry test appears, when using the F-distribution critical values, to over-reject moderately at the 10% significance level, to over-reject slightly at the 5% significance level and to under-reject at the 1% significance level; also, it appears to over-reject more strongly than the Chow test when using the chi-square critical values. This accords with the result of Section 5.2 that $\Pr\{F^{c}_{1|n} > c_\alpha\} \leq \Pr\{F^{h}_{1|n} > c_\alpha\}$ for all $c_\alpha$. The overall conclusion of the theoretical and evaluation results is that the Chow test appears to behave well when used with the F-distribution critical values and that the Hendry test appears to strongly over-reject when used with chi-square critical values.

Table 5.1 only reports evaluated approximate rejection probabilities for n=10. There are two reasons for this. Firstly, the differences between the nominal significance levels and the evaluated approximate rejection probabilities for the tests at any given point appear from Equation (27) to be constants multiplied by $n^{-1}$. Thus there is no value in reporting the evaluated approximate rejection probabilities for n=20,...,50 at the chi-square critical values since these can be inferred from Table 5.1; doubling n simply halves the deviation of the evaluated approximate rejection probability from the nominal significance level. Secondly, the F-distribution critical values are essentially only presented to examine the behaviour of the Chow test. Since the results seem to indicate that for n=10 the Chow test behaves very well when used with 1 and 10 degrees of freedom, there is little value in reporting results for n=20,...,50.

### 5.4.2 Results for Multi-Period Tests

The coefficients of the asymptotic expansion approximations to the joint cdf's of $(F^{c}_{1|n}, F^{c+}_{2|n})$ and of $(F^{h}_{1|n}, F^{h+}_{2|n})$ are given in Table 5.2 where the approximations take the form:

$$\Pr(z_1 \leq r_1^2 \;\&\; z_2 \leq r_2^2) = (1 + b_0^* n^{-1})\Gamma_1^2(r_1^2)\Gamma_1^2(r_2^2)$$

$$+ b_1^* n^{-1}\Gamma_3^2(r_1^2)\Gamma_1^2(r_2^2) + b_2^* n^{-1}\Gamma_1^2(r_1^2)\Gamma_3^2(r_2^2)$$

$$+ b_{11}^* n^{-1}\Gamma_5^2(r_1)\Gamma_1^2(r_2^2) + b_{12}^* n^{-1}\Gamma_3^2(r_1^2)\Gamma_3^2(r_2^2)$$

$$+ b_{22}^* n^{-1}\Gamma_1^2(r_1^2)\Gamma_5^2(r_2^2) + o(n^{-1}) \,. \qquad (28)$$

following Equation (15) which again confirms the symmetry argument presented in Section 5.3. The coefficients of the asymptotic expansion approximations to the cdf's of $F^{c}_{2|n}$ and $F^{h}_{2|n}$ are given in Table 5.3 where the approximations take the form:

$$\Pr(z \leq r_3^2) = (1 + c_0^* n^{-1})\Gamma_2^2(r_3^2) + c_1^* n^{-1}\Gamma_4^2(r_3^2)$$

$$+ c_2^* n^{-1}\Gamma_6^2(r_3^2) + o(n^{-1}). \qquad (29)$$

The coefficients in Table 5.3 are obtained from those in Table 5.2 by:

$$c_0^* = b_0^*, \quad c_1^* = b_1^* + b_2^*, \quad c_2^* = b_{11}^* + b_{12}^* + b_{22}^*,$$

following Chapter 3, Sections 3.2.3 and 3.2.4. As with the results for the individual single-period predictive failure tests various points are immediately apparent from examining the coefficients.

| Coefficient | Hendry Tests ($F^h_{1\|n}$ and $F^{h+}_{2\|n}$) Parameter Values: $\beta = 0.0$ | $\beta = 0.2$ | $\beta = 0.5$ | $\beta = 0.8$ | $\beta = 0.9$ | Chow Tests ($F^c_{1\|n}$ and $F^{c+}_{2\|n}$) |
|---|---|---|---|---|---|---|
| $b^*_o$ | $-\frac{1}{2}$ | -26/50 | $-\frac{5}{8}$ | -41/50 | -181/200 | 0 |
| $b^*_1$ | -1 | -49/50 | $-\frac{7}{8}$ | -34/50 | -119/200 | -1 |
| $b^*_2$ | -1 | -49/50 | $-\frac{7}{8}$ | -34/50 | -119/200 | -1 |
| $b^*_{11}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ |
| $b^*_{12}$ | 1 | 49/50 | $\frac{7}{8}$ | 34/50 | 119/200 | $\frac{1}{2}$ |
| $b^*_{22}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{3}{4}$ |

Table 5.2: Asymptotic Expansion Coefficients for the Joint CDF's of ($F^h_{1|n}$, $F^{h+}_{2|n}$) and of ($F^c_{1|n}$, $F^{c+}_{2|n}$)

| Hendry Test ($F^h_{2\|n}$) | Coefficients $c_0^*$ | $c_1^*$ | $c_2^*$ |
|---|---|---|---|
| Parameter Value | | | |
| $\beta = 0.0$ | $-\frac{1}{2}$ | $-2$ | $2\frac{1}{2}$ |
| $\beta = 0.2$ | $-26/50$ | $-98/50$ | $124/50$ |
| $\beta = 0.5$ | $-5/8$ | $-7/4$ | $19/8$ |
| $\beta = 0.8$ | $-41/50$ | $-68/50$ | $109/50$ |
| $\beta = 0.9$ | $-181/200$ | $-238/200$ | $419/200$ |
| Chow Test ($F^c_{2\|n}$) | 0 | $-2$ | 2 |

Table 5.3: Asymptotic Expansion Coefficients for the CDF's of $F^h_{2|n}$ and $F^c_{2|n}$

(i) The coefficients of the expansion for the joint cdf of $F^{h}_{1|n}$ and $F^{h+}_{2|n}$ do depend on $\beta$ as suggested in Section 5.3 and an exact fit can be found for these coefficients by putting:

$$b^*_0 = \tfrac{1}{2}(1-\beta^2)-1, \quad b^*_{11} = b^*_{22} = 3/4,$$
$$b^*_1 = b^*_2 = -\tfrac{1}{2}(1-\beta^2)-\tfrac{1}{2}, \quad b^*_{12} = \tfrac{1}{2}(1-\beta^2) + \tfrac{1}{2}. \qquad (30)$$

Consequently, the coefficients of the expansion for the cdf of the two-period Hendry test are given by:

$$c^*_0 = \tfrac{1}{2}(1-\beta^2)-1, \quad c^*_1 = -(1-\beta^2)-1,$$

$$c^*_2 = \tfrac{1}{2}(1-\beta^2) + 2.$$

The dependence on $\beta$ through $\beta^2$ reflects the lack of dependence of the joint cdf on the sign of $\beta$ as argued in Section 5.3.

(ii) The coefficients of the expansions for the joint cdf of $F^{c}_{1|n}$ and $F^{c+}_{2|n}$ and for the cdf of $F^{c}_{2|n}$ do not depend on $\beta$. Thus the critical values of the tests could be set to achieve rejection probabilities of $\alpha$ to $o(n^{-1})$ under the null hypothesis irrespective of $\beta$. This is somewhat surprising because the arguments presented in Section 5.3 suggested that the Chow tests failed to make a full adjustment for the sampling variation in $\hat{\beta}_n$ so that their cdf's should depend on $\beta$.

(iii) The expansion coefficients for the joint cdf of the single-period Hendry tests $F^{h}_{1|n}$ and $F^{h+}_{2|n}$ are symmetric so that, to $o(n^{-1})$, in the joint distribution of the Hendry tests they are interchangeable. Thus the asymptotic expansion approximation for the cdf of the second-period test is the same as that of the first-period test as would be suggested by the results of Sections 5.3 and 5.4.1. This is also true for $F^{C}_{1|n}$ and $F^{C+}_{2|n}$.

From the approximate rejection probabilities presented in Table 5.4 we can see that the two-period Chow test seems to perform very well at the 10% significance level using F-distribution critical values with 1 and 10 degrees of freedom, less well at the 5% level and moderately under-rejects at the 1% level. We can also see that the use of chi-square critical values appears to lead to much greater deviations of actual rejection probabilities to $o(n^{-1})$ from nominal significance levels than does the use of F-distribution critical values. This effect is more marked for the two-period test than for the one-period test which accords with the results of Kiviet (1986), reported in Section 5.3 that lengthening the post-estimation period used causes the size performance of the test to deteriorate. As with the one-period tests the results presented in Table 5.4 meet the restriction that the rejection probability of the Chow test must be less than that of the Hendry test.

As we would expect from the expansion results, the rejection probabilities of the two-period Hendry test depend on $\beta$ and we can see that raising $\beta$ lowers the rejection probabilities. However, the variation in the rejection probabilities for either chi-square or F-distribution critical values arising from varying $\beta$ is much smaller than the difference in rejection probabilities for any $\beta$ arising from switching from F-distribution to chi-square distribution critical values. Also the difference between the rejection probabilities for

| Nominal Significance Level | | Critical Value | Hendry Test ($F^h_{2\|n}$) | | | | | Chow Test ($F^c_{2\|n}$) |
|---|---|---|---|---|---|---|---|---|
| | | | β = 0.0 | β = 0.2 | β = 0.5 | β = 0.8 | β = 0.9 | |
| | F(1,10) | 5.8490 | 11.89 | 11.87 | 11.80 | 11.66 | 11.60 | 9.96 |
| 10% | F(1, 9) | 6.0130 | 11.28 | 11.26 | 11.19 | 11.04 | 10.96 | 9.42 |
| | $\chi^2$ | 4.6052 | 17.78 | 17.77 | 17.74 | 17.67 | 17.64 | 15.30 |
| | F(1,10) | 8.2056 | 5.47 | 5.45 | 5.38 | 5.24 | 5.18 | 4.43 |
| 5% | F(1, 9) | 8.5130 | 4.93 | 4.91 | 4.84 | 4.71 | 4.65 | 3.98 |
| | $\chi^2$ | 5.9915 | 11.36 | 11.34 | 11.26 | 11.12 | 11.06 | 9.49 |
| | F(1,10) | 15.1188 | 0.44 | 0.44 | 0.43 | 0.41 | 0.40 | 0.35 |
| 1% | F(1, 9) | 16.0430 | 0.31 | 0.31 | 0.30 | 0.28 | 0.28 | 0.24 |
| | $\chi^2$ | 9.2103 | 3.88 | 3.87 | 3.81 | 3.69 | 3.64 | 3.12 |

Table 5.4: Approximate Rejection Probabilities Expressed as %'s for $F^h_{2|n}$ and $F^c_{2|n}$ when n=10.

*Note:* *The critical values used for the F(1,10) and F(1,9) nominal significance levels are equal to two times those normally used. This is because we have not divided the statistic by the numerator degrees of freedom, in this case equal to two, so as to maintain comparability with the critical values used for the chi-square nominal significance levels.*

the two-period Hendry and Chow tests is much less than the difference in rejection probabilities for either the Hendry or the Chow two-period tests arising from switching critical values.

As with the one-period tests we only present results in Table 5.4 for n=10; the reasons for doing this for the two-period tests are the same as for the one-period tests.

## 5.5 Conclusions

The clearest conclusion which we can draw from the theoretical results of Section 5.3 and the numerical results of Section 5.4 is that the one-period tests seem to behave substantially the same in the AR(1) model as they would in the CLR model. If we turn to the two-period tests then the results of Section 5.3 seem to indicate that neither the Hendry nor the Chow test properly take into account the interdependence between the forecast errors arising from the sampling variation. The results of Section 5.4 indicate that this is a more serious problem with the Hendry test; its distribution does depend on the unknown autoregressive parameter $\beta$, and the test seems to substantially over-reject when used with chi-square critical values. However, the results of Section 5.4 seem to indicate that the two-period Chow test performs well when used with F-distribution critical values in spite of the results in Section 5.3. Whether the good performance is due to the particular model being considered or whether it carries over to more general dynamic models is unclear. The major implication of this study for actual practice is that researchers should be somewhat cautious when interpreting the results of Chow predictive failure tests.

The numerical results of Section 5.4 indicate that the coefficients of the asymptotic expansion approximations to the cdf's of the Hendry and Chow tests are relatively simple functions of $\rho^2$ which suggests that it may be possible to obtain the coefficients analytically. If this is so, then it may be possible to examine in more detail why the two-period Chow test appears to perform well in the AR(1) model in spite of the reservations raised in Section 5.3.

Another direction for future research is to consider sequential predictive failure tests. There are two ways in which this might be approached. Firstly, we could examine the properties of a k-period predictive failure test conditional upon the result of an r-period predictive failure test where $k>r$. This would require the use of asymptotic expansion approximations for the distributions of asymptotically jointly dependent chi-square variates which were briefly discussed in Chapter 3, Section 3.4. Secondly, we could examine the properties of individual period predictive failure tests conditional upon the results of previous individual period predictive failure tests. This would be a less satisfactory approach for the Chow tests than for the Hendry tests since in constructing the Chow tests we wish explicitly to take account of sampling variation in the parameter estimates and its effect on the forecast errors, and by only considering individual period tests we would fail to take account of the effects of sampling variation on the covariances of the forecast errors.

We have not considered the power properties of these tests in this study although, in principle, this should be possible using the techniques of Chapter 3. However, as argued in Section 5.2, predictive failure tests are unlikely to have particularly good power properties against stationary alternatives since they will not in general be consistent against such alternatives. Thus the alternatives of interest would primarily be non-stationary; this though would create difficulties in the application of asymptotic expansion methods.

## 5.A Appendix to Chapter 5

### 5.A.1 Implications of Distributional Symmetry for Unconditional Forecast Errors

Consider the univariate case and let $\omega$ be the standardised forecast error. Then the pdf of $\omega$ can be approximated to $O(n^{-3/2})$ by:

$$\begin{aligned} \text{pdf}(\omega) = {} & [(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\omega^2\}][1 + d_0 + d_1\omega \\ & + d_2\omega^2 + d_3\omega^3 + d_4\omega^4 + d_6\omega^6] + O(n^{-3/2}), \end{aligned} \qquad (31)$$

(where $d_6 = d_3^2/2$) using the methods of Chapter 3, Section 3.2. Since $\omega$ has a symmetric distribution about 0, the pdf of $\omega$ is the same as that of $(-\omega)$:

$$\begin{aligned} \text{pdf}(-\omega) = {} & [(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\omega^2\}][1 + d_0 - d_1\omega \\ & + d_2\omega^2 - d_3\omega^3 + d_4\omega^4 + d_6\omega^6] + O(n^{-3/2}). \end{aligned}$$

Equating pdf($\omega$) with pdf($-\omega$) implies that $d_1 = d_3 = 0$ and thus $d_6 = 0$. Therefore Equation (31) can be simplified to:

$$\text{pdf}(\omega) = [(2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\omega^2\}][1 + d_0 + d_2\omega^2 + d_4\omega^4] + O(n^{-3/2}),$$

and so the pdf and cdf of $\omega^2$ can be approximated to $o(n^{-1})$ by:

$$\begin{aligned} \text{pdf}(\omega^2) &= [1+d_0]\chi_1^2(\omega^2) + d_2\chi_3^2(\omega^2) + 3d_4\chi_5^2(\omega^2) + o(n^{-1}) \\ \text{cdf}(\omega^2) &= [1+d_0]\Gamma_1^2(\omega^2) + d_2\Gamma_3^2(\omega^2) + 3d_4\Gamma_5^2(\omega^2) + o(n^{-1}). \end{aligned} \qquad (32)$$

A similar argument can be used for the multivariate case using Equations (15), (22), (26), (46), (47) and (48) of Chapter 3 to give Equation (15).

5.A.2 Variances of Unconditional Forecast Errors

To show that Equation (19) holds we expand $\hat{e}_{s|n}$ for $s \geq 2$ as:

$$\begin{aligned}\hat{e}_{s|n} &= y_{n+s} - \hat{y}_{s|n} = \epsilon_{n+s} - (\hat{\beta}_n-\beta)y_{n+s-1} \\ &= \epsilon_{n+s} - (\hat{\beta}_n-\beta)(\beta^{s-1}y_n + \sum_{j=0}^{s-2} \beta^j \epsilon_{n+s-1-j}) .\end{aligned} \tag{33}$$

Squaring this gives:

$$\begin{aligned}\hat{e}^2_{s|n} &= \epsilon^2_{n+s} - 2(\hat{\beta}_n-\beta)\beta^{s-1}y_n\epsilon_{n+s} \\ &\quad - 2(\hat{\beta}_n-\beta)\Big[\sum_{j=0}^{s-2} \beta^j\epsilon_{n+s-1-j}\Big]\epsilon_{n+s} \\ &\quad + (\hat{\beta}_n-\beta)^2\beta^{2(s-1)}y_n^2 \\ &\quad + 2(\hat{\beta}_n-\beta)^2\beta^{(s-1)}y_n\Big[\sum_{j=0}^{s-2} \beta^j\epsilon_{n+s-1-j}\Big] \\ &\quad + (\hat{\beta}_n-\beta)^2\Big[\sum_{j=0}^{s-2} \beta^j\epsilon_{n+s-1-j}\Big]^2,\end{aligned}$$

and then taking expectations gives:

$$\begin{aligned}E\big[\hat{e}^2_{s|n}\big] &= \sigma^2 + \beta^{2(s-1)}E\big[(\hat{\beta}_n-\beta)^2y_n^2\big] \\ &\quad + \sigma^2E\big[(\hat{\beta}_n-\beta)^2\big]\Big[\sum_{j=0}^{s-2} \beta^{2j}\Big].\end{aligned} \tag{34}$$

From White (1961) and Fuller and Hasza (1980) we have Equations (22) and (23):

$$\begin{aligned}E\big[(\hat{\beta}_n-\beta)^2\big] &= (1-\beta^2)n^{-1} + O(n^{-2}), \\ E\big[(\hat{\beta}_n-\beta)^2y_n^2\big] &= \sigma^2n^{-1} + O(n^{-3/2}).\end{aligned}$$

Substituting these into Equation (34) gives Equation (19):

$$\begin{aligned}E(\hat{e}_{s|n}) &= \sigma^2 + \sigma^2\beta^{2(s-1)}n^{-1} + \sigma^2(1-\beta^2)(\sum_{j=0}^{s-2} \beta^{2j})n^{-1} + O(n^{-3/2}) \\ &= \sigma^2\big[1+\beta^{2(s-1)}n^{-1} + n^{-1} - \beta^{2(s-1)}n^{-1}\big] + O(n^{-3/2}) \\ &= \sigma^2\big[1+n^{-1}\big] + O(n^{-3/2}).\end{aligned}$$

### 5.A.3 Asymptotic Expansion of the CDF of a F(1,n) Variate in Terms of the CDF's of $\chi_r^2$ Variates

Fisher (1925) gave an Edgeworth expansion for the pdf of a t-variate with n degrees of freedom where $i(t) = [(2\pi)^{-\frac{1}{2}}\exp\{-\frac{1}{2}t^2\}]$ as:

$$p_n(t) = i(t)[1-n^{-1}(\tfrac{1}{4})(1 + 2t^2 - t^4)] + o(n^{-1}). \qquad (35)$$

Integrating this gives:

$$Pr\{|t| \leq r\} = 2\int_0^r i(t)[1-n^{-1}(\tfrac{1}{4})(1 + 2t^2 - t^4)]dt + o(n^{-1}),$$

and transforming from t to $z = t^2$ gives:

$$Pr\{z \leq r^2\} = Pr\{|t| \leq r\}$$

$$= \int_0^{r^2} \frac{e^{-z/2}}{(2\pi)^{-\frac{1}{2}}z^{\frac{1}{2}}}[1-n^{-1}(\tfrac{1}{4}) - n^{-1}(\tfrac{1}{2})z + n^{-1}(\tfrac{1}{4})z^2]dz + o(n^{-1})$$

$$= [1-n^{-1}(\tfrac{1}{4})]\Gamma_1^2(r^2) - n^{-1}(\tfrac{1}{2})\Gamma_3^2(r^2) + n^{-1}(\tfrac{3}{4})\Gamma_5^2(r^2) + o(n^{-1}). \qquad (36)$$

Since t is distributed as a t-variate with n degrees of freedom then z is distributed as an F-variate with (1,n) degrees of freedom so that Equation (36) gives an expansion for the cdf of an F(1,n) variate in terms of the cdf's of $\chi_r^2$ variates. Also since:

$$(n-1)^{-1} = (\tfrac{1}{n})[1-(\tfrac{1}{n})]^{-1} = n^{-1} + o(n^{-1}),$$

then the cdf of an F(1,n-1) variate is the same to $o(n^{-1})$ as that of an F(1,n) variate.

Now suppose that x is distributed as a chi-square variate with q degrees of freedom so that the pdf of x is:

$$f_q(x) = \frac{x^{(q/2)-1} e^{-(x/2)}}{\Gamma(q/2) 2^{(q/2)}} = \chi^2_q(x) ,$$

where $\Gamma(\cdot)$ here is the gamma function, not the cdf of a $\chi^2$ variate.

If we define $y = [1+(a/n)]x$ then the pdf of y is:

$$g_q(y) = \frac{\{y/[1+(a/n)]\}^{(q/2)-1} e^{-\{y/[1+(a/n)]\}/2}}{\Gamma(q/2)\, 2^{(q/2)}\, [1+(a/n)]}$$

$$= \chi^2_q(y)[1+(q/n)]^{-(q/2)} e^{(ay)/(2n)} + o(n^{-1})$$

$$= [\, 1 - \{(aq)/(2n)\}\,]\, \chi^2_q(y) + \{(aq)/(2n)\}\, \chi^2_{q+2}(y) + o(n^{-1})$$

Therefore, if z is distributed as an F(1,n) variate then the cdf of $(1+n^{-1})z$ can be approximated by:

$$\Pr\{(1+n^{-1})z \le r^2\} = [1 - n^{-1}(\tfrac{1}{2})]\Gamma^2_1(r^2) \qquad (37)$$
$$+ n^{-1}(\tfrac{1}{2})\Gamma^2_5(r^2) + o(n^{-1}),$$

where $\Gamma^2_k(\cdot)$ here is the cdf of a chi-square variate with k degrees of freedom. Equation (37) also holds if z is an F(1,n-1) variate.

# CHAPTER 6: An Asymptotic Expansion Analysis of a Simple Dynamic Specification Testing Sequence in the AD(1,1) Model

## 6.1 Introduction

This chapter is concerned with analysing a simple testing sequence in the Autoregressive Distributed lag model with one lag on the endogenous variable and one lag on a single exogenous variable (the AD(1,1) model). The first test in the sequence is a Wald test of the hypothesis that the lag polynomials have a common factor so that the model can be re-written as a simple regression model with autocorrelated disturbances; see Hendry and Mizon (1978) and Sargan (1980a). The second test in the sequence is an incremental Wald test of the additional hypothesis that there are no endogenous dynamics. The two hypotheses together imply that the model can be re-written as a simple regression model with white noise disturbances.

In Section 6.2 we consider the background to sequential testing procedures. Then in Section 6.3 we examine in more detail the case mentioned above. In Section 6.4 we perform a numerical evaluation study using asymptotic expansion approximations to the joint cdf's of the test statistics for the case mentioned above. Lastly, in Section 6.5 we attempt to draw conclusions from this study.

## 6.2 Sequential Test Procedures

Sequential testing procedures are frequently adopted in applied econometric studies, both when deciding whether to accept or reject particular models and when testing various restrictions on the parameters of a model. Often there are situations where one set of restrictions would not be considered unless some other set of restrictions had been accepted, so leading to a series of monotonically nested hypotheses. Such nested hypotheses can be tested sequentially using Wald tests. These have been shown in Sargan (1980a) and Kiviet and Phillips (1986) to have some useful asymptotic properties for testing such hypotheses sequences.

Following Chapter 1, suppose that Y is a random variable distributed with density function $f(y;\theta)$ where $\theta$ is a k-vector of parameters, $\theta \in \Theta$, and where the density function meets the standard regularity conditions for Maximum Likelihood estimation to have the usual asymptotic properties; see Wilks (1962). If $\hat{\theta}$ is a consistent estimator of $\theta$ and $\hat{\Omega}$ is a consistent estimator of the asymptotic covariance matrix of $\hat{\theta}$, i.e. plim $\hat{\Omega} = \Omega$ where $T^{\frac{1}{2}}(\hat{\theta} - \theta)$ is asymptotically $N[0,\Omega]$, then we can attempt to construct Wald tests for the sequence of nested hypotheses:

$$H_1: \ \phi_1(\theta) = 0;$$

$$H_2: \ \phi_1(\theta) = 0, \ \phi_2(\theta) = 0;$$

$$\vdots$$

$$H_r: \ \phi_1(\theta) = 0, \ \ldots, \ \phi_r(\theta) = 0; \text{ where } r \leq k. \qquad (1)$$

If we define $[\phi^j(\theta)] = [\phi_1(\theta)', \ldots, \phi_j(\theta)']'$ then the Wald statistic for $H_j$ is:

$$W_j = T[\phi^j(\hat{\theta})]' \, [\hat{F}^j \hat{\Omega} \hat{F}^{j\prime}]^{-1} \, [\phi^j(\hat{\theta})]$$

$$\text{where } \hat{F}^j = \left. \frac{\partial \phi^j}{\partial \theta} \right|_{\hat{\theta}} . \qquad (2)$$

From the sequence of test statistics $\{W_j\}$ we can construct a sequence of incremental Wald tests:

$$W_{(j)} = \begin{cases} W_j & ; \quad j = 1 \\ W_j - W_{j-1} & ; \quad j > 1 , \end{cases} \qquad (3)$$

so that $W_{(j)}$ is a test of $H_j$ given that $H_{j-1}$ is assumed to hold.

Provided that the matrices $\{\bar{F}^j\}$ are of full rank in some neighbourhood of the true parameter value, such incremental Wald tests have the useful property that if the most restrictive hypothesis is true then the incremental test statistics are asymptotically distributed as independent, central Chi-square variates with degrees of freedom given by the dimensions of the $\{\phi_j(\theta)\}$ vectors. Thus the overall asymptotic size of a sequence of m incremental Wald tests can be controlled given the value of m. Furthermore, under a local alternative hypothesis sequence converging to the most restrictive hypothesis the incremental test statistics are asymptotically distributed as independent, non-central Chi-square variates with the same degrees of freedom as before and with non-centrality parameters depending on the particular local alternative hypothesis sequence chosen; see Hogg (1961), Sargan (1980a) and Kiviet and Phillips (1986). The power properties of a sequence of incremental Wald tests can then be approximated using this result.

Wald tests are particularly appropriate for testing sequences of nested hypotheses within a maintained model because they require estimation only of the general model. The Likelihood Ratio and Lagrange Multiplier tests, although asymptotically equivalent to the Wald tests under local alternative hypothesis sequences, require estimation of the model subject to the restrictions and therefore they generally require more computation than the Wald tests. However the Wald tests suffer from the problem of a lack of invariance to algebraic reformulations of the restrictions. This is discussed in Section 6.3 and also in Chapter 4.

Nested hypothesis sequences can arise either from purely economic considerations or from mathematical and statistical considerations. An example of the former would be where a researcher wishing to test for rational expectations might first test that expectations were constructed using a particular information set and then that the expectations were rational with respect to that information set. An example of the latter is provided by the COMFAC (COMmon FACtor) restrictions in the AD model:

$$y_t = \sum_{j=1}^{r} \alpha_j y_{t-j} + \sum_{i=0}^{r} \beta_i' x_{t-i} + u_t, \qquad (4)$$

or $\alpha(L)y_t = \beta(L)'x_t + u_t,$

where the $\{u_t\}$ are distributed as $IN[0,\sigma^2]$ and are independent of the $\{x_t\}$. The common factor restriction of order m, where $m \leq r$, can be expressed as:

$$\alpha(L) = \rho(L)\alpha^*(L); \; \beta(L) = \rho(L)\beta^*(L)$$

where $\rho(L)$ is an m'th order lag polynomial, $\alpha^*(L)$ is an (r-m)'th order lag polynomial and $\beta^*(L)$ is a vector of (r-m)'th order lag polynomials. The model can be generalised to have different numbers of lags on the endogenous and exogenous variables simply by requiring some of the elements of $\{\alpha_j\}$ and $\{\beta_i\}$ to be identically equal to zero. Sargan (1980a) shows that these restrictions (and their extensions to models with unequal lag lengths) can be expressed in terms of the determinants of certain matrices and provides an algorithm for their practical implementation. Hendry and Mizon (1978) illustrate how tests for these restrictions, based on the Sargan approach, can be used in dynamic specification searches with an example looking at the demand for money.

It is natural to require that we accept the existence of (m+1) common factors only if we have already accepted the existence of m common factors. Therefore common factor restrictions would appear to provide a very natural example of a nested hypothesis sequence. However there are two sources of difficulties with common factor restrictions. Firstly, it is possible for there to be complex common factors, which necessarily occur in complex conjugate pairs; therefore it has been suggested that it would be more sensible to test for the existence of a pair of additional common factors and if this is rejected then test for an additional single common factor. Secondly, there may be problems associated with testing for $m_1$ common factors when there are $m_2$ common factors present with $m_1 < m_2$. Sargan (1984) provides some theoretical justification for this. The difficulty with the Wald tests here is that the matrix of first derivatives of the restrictions will not be of full rank asymptotically so that the test statistics will not typically have the usual chi-square distributions asymptotically.

## 6.3 The Testing Sequence

The model which we are considering in more detail has only one lag on one endogenous and one exogenous variable:

$$y_t = \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \tag{5}$$

where the $\{x_t\}$ are non-stochastic and the $\{\varepsilon_t\}$ are distributed $IN[0,\sigma^2]$. The hypotheses which we are considering are:

$$H_1: \quad \alpha\beta_0 + \beta_1 = 0; \tag{6}$$

$$H_2: \quad \alpha\beta_0 + \beta_1 = 0, \quad \alpha = 0.$$

Under $H_1$ the model can be written as:

$$y_t = \beta_0 x_t + \omega_t \; ; \quad \omega_t = \alpha\omega_{t-1} + \varepsilon_t \tag{7}$$

and under $H_2$ the model can be written as:

$$y_t = \beta_0 x_t + \varepsilon_t . \tag{8}$$

The Wald tests $W_1$ and $W_2$ for $H_1$ and $H_2$ respectively do not suffer from either of the problems discussed at the end of Section 6.2. Firstly, there is only one lag in the model so that there is no possibility of complex common factors. Secondly, the matrix of first derivatives of the restriction functions is:

$$F(\alpha,\beta_0,\beta_1) = \begin{bmatrix} \beta_0 & \alpha & 1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad (9)$$

which is clearly of full rank for all $(\alpha,\beta_0,\beta_1)$.

Using the invariance theorem proved in Appendix 6.A, we can prove that the joint distribution of $W_1$ and $W_2$ depends only on $\alpha$, $(\alpha\beta_0+\beta_1)^2/\sigma^2$, T and the exogenous data $\{x_t\}$. Therefore under the null hypothesis $H_2$ we can set $\alpha = \beta_0 = \beta_1 = 0$ and $\sigma^2 = 1$ without any loss of generality so that in the evaluation study in Section 6.4 the only factors affecting the joint distribution of $W_1$ and $W_2$ are T and the exogenous data $\{x_t\}$.

We can invoke the invariance theorem of Appendix 6.A by defining $y_t^*$ as:

$$y_t^* = \lambda(y_t + \gamma x_t), \ \lambda \neq 0. \qquad (10)$$

Then $\{y_t^*\}$ is generated by the AD(1,1) model:

$$y_t^* = \alpha^* y_{t-1}^* + \beta_0^* x_t + \beta_1^* x_{t-1} + \varepsilon_t^*$$

where $\alpha^* = \alpha$, $\beta_0^* = \lambda(\beta_0+\gamma)$, $\beta_1^* = \lambda(\beta_1-\alpha\gamma)$,

$$\varepsilon_t^* = \lambda\varepsilon_t, \text{ and var } (\varepsilon_t^*) = (\sigma^*)^2 = \lambda^2\sigma^2. \qquad (11)$$

Therefore we can define two groups of transformations, G and G*, operating on $\{y_t\}$ and $(\alpha, \beta_0,\beta_1,\sigma^2)$ respectively, by the families of transformations given in Equations (10) and (11).

Clearly, the transformation on $\{y_t\}$ given in Equation (10) for different values of $(\lambda,\gamma)$ form a group where the operation is to follow one transformation belonging to G, $y_t \rightarrow y_t^*$, by another transformation belonging to G, $y_t^* \rightarrow y_t^{**}$. Firstly, the result of any transformation belonging to G followed by another transformation belonging to G is also a transformation belonging to G, so that this operation of following one transformation by another is closed. Secondly, this operation is associative. Thirdly, there is a unique identity transformation given by putting $\lambda = 1$ and $\gamma = 0$. Fourthly, each transformation belonging to G has a unique inverse given by $y_t = \lambda^{-1}(y_t^* - \lambda\gamma x_t)$. Similarly, the induced transformations on $(\alpha, \beta_0, \beta_1, \sigma^2)$ given in Equation (11) also form a group G*.

The null hypotheses $H_1$: $\alpha\beta_0+\beta_1 = 0$ and $H_2$: $\alpha\beta_0+\beta_1 = 0$, $\alpha = 0$ are invariant under the induced group of transformations G* on $(\alpha,\beta_0,\beta_1,\sigma^2)$ given by Equation (11) since:

$$\alpha^*\beta_0^* + \beta_1^* = \lambda(\alpha\beta_0+\beta_1), \quad \alpha^* = \alpha. \tag{12}$$

Thus $H_1$ holds for $(\alpha, \beta_0, \beta_1, \sigma^2)$ if and only if it holds for $(\alpha^*, \beta_0^*, \beta_1^*, \sigma^{*2})$, and similarly for $H_2$. It is then reasonable to ask whether we can construct Wald test statistics for $H_1$ and $H_2$ which are numerically invariant under the group of transformations G acting on $\{y_t\}$ and therefore have joint distributions which are invariant under the induced group of transformations G* acting on $(\alpha, \beta_0, \beta_1, \sigma^2)$. The theorem in Section 6.A enables us to construct such test statistics providing that four conditions are met.

Firstly, we need to be able to partition $\theta = (\theta_1', \theta_2')'$ so that the restrictions only involve $\theta_1$, and $g^*\theta = (g_1^*\theta_1', g_2^*\theta')'$ where $g^* \varepsilon G^*$. Here $\theta_1 = (\alpha, \beta_0, \beta_1)'$ and $\theta_2 = \sigma^2$ and this condition is then clearly met by the hypotheses and the group of transformations $G^*$. Secondly, we need to find an estimator $\bar{\theta}_1(y)$ of $\theta_1$ such that for any $g \varepsilon G$ then $\bar{\theta}_1(gy) = g_1^*\bar{\theta}_1(y)$ for the corresponding $g^* \varepsilon G^*$. The estimator which we choose is the OLS estimator given by:

$$\min S(\theta_1^*; y) = \sum_{t=1}^{T} e_t^2 \tag{13}$$

where $e_t = y_t - ay_{t-1} - b_0x_t - b_1x_{t-1}$,

with respect to $\theta_1^* = (a, b_0, b_1)'$ giving $\bar{\theta}_1 = (\bar{\alpha}, \bar{\beta}_0, \bar{\beta}_1)'$ and then putting $\bar{\sigma}^2 = (T-3)^{-1}S(\bar{\theta}_1; y)$. But then:

$$S(g_1^*\theta_1^*; gy) = \lambda^2 S(\theta_1; y), \tag{14}$$

so that if $S(.;y)$ is minimised at $\bar{\theta}_1$ then $S(.;gy)$ is minimised at $g_1^*\bar{\theta}_1$. Therefore $\hat{\theta}_1(gy) = g_1^*\hat{\theta}_1(y)$ as required. We can also note that $\sigma^2(gy) = \lambda^2\sigma^2(y)$ so that $\bar{\theta}(gy) = g^*\bar{\theta}(y)$.

Thirdly, we need to find an estimator $\bar{\Omega}_{11}(y)$ of $\Omega_{11}$, the asymptotic covariance matrix of $T^{\frac{1}{2}}[\bar{\theta}_1(y) - \theta_1]$ so that:

$$\bar{\Omega}_{11}(gy) = \left[ \left. \frac{\partial g_1^*}{\partial \theta_1'} \right|_{\bar{\theta}_1(y)} \right] \hat{\Omega}_{11}(y) \left[ \left. \frac{\partial g_1^*}{\partial \theta_1'} \right|_{\bar{\theta}_1(y)} \right]' . \tag{15}$$

The covariance matrix estimate which we choose is the standard OLS estimate:

$$\hat{\Omega}_{11}(y) = \hat{\sigma}^2(y) \sum_{t=1}^{T} z_t z_t')^{-1} \qquad (16)$$

where $z_t = (y_t, x_t, x_{t-1})'$.

The covariance matrix estimate using the transformed data is:

$$\hat{\Omega}_{11}(gy) = \hat{\sigma}^2(gy) \left(\sum_{t=1}^{T} z_t^* z_t^{*\prime}\right)^{-1}$$

where $z_t^* = (y_{t-1}^*, x_t, x_{t-1})'$

$$= \begin{bmatrix} \lambda & 0 & \lambda\gamma \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} z_t = Cz_t.$$

Then $\hat{\Omega}_{11}(gy) = [\lambda^{-1}C']^{-1} \hat{\Omega}_{11}(y)[\lambda^{-1}C]^{-1}$ since $\hat{\sigma}^2(gy) = \lambda^2\hat{\sigma}^2(y)$ and:

$$(\lambda^{-1}C')^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ -\lambda\gamma & 0 & \lambda \end{bmatrix} = \left[ \left. \frac{\partial g_1^* \theta_1}{\partial \theta_1^*} \right| \hat{\theta}_1(y) \right].$$

Fourthly, we need to find a representation $\phi(\theta_1) = 0$ of the restrictions being tested and a group of matrices $G^+$ such that for any $g^* \in G$ there exists $J^+ \in G^+$ so that:

$$\phi(g_1^*\theta_1) = J^+\phi(\theta_1) \text{ for all } \theta_1. \qquad (17)$$

This is met by putting $\alpha\beta_0+\beta_1 = 0$ for $H_1$ and $\alpha\beta_0+\beta_1 = 0, \alpha = 0$ for $H_2$ since:

$$\begin{bmatrix} \alpha^*\beta_0^* + \beta_1^* \\ \alpha^* \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha\beta_0+\beta_1 \\ \alpha \end{bmatrix}.$$

Thus all the conditions of the theorem are met and so the Wald test statistics $W_1$ and $W_2$ for $H_1$: $\alpha\beta_0+\beta_1 = 0$ and $H_2$: $\alpha\beta_0+\beta_1 = 0$, $\alpha = 0$ are numerically invariant under the group of transformations G given by Equation (10); therefore their joint distribution is invariant under the group of transformations G* given by Equation (11).

If we now put $\gamma = -\beta_0$ and $\lambda = (\sigma^2)^{-\frac{1}{2}}$ then $\alpha^* = \alpha$, $\beta_0^* = 0$, $\beta_1^* = (\alpha\beta_0+\beta_1)/(\sigma^2)^{\frac{1}{2}}$ and $(\sigma^*)^2 = 1$. Furthermore, we can choose $\lambda$ to be either the positive or the negative square root of $\sigma^2$ and therefore the sign of $\beta_1^*$ does not matter. Since none of the transformations affect T or $\{x_t\}$ we can therefore parameterize the joint distribution of $W_1$ and $W_2$ by $\alpha$, $(\alpha\beta_0+\beta_1)^2/\sigma^2$, T and $\{x_t\}$.

Under $H_2$, $\alpha\beta_0+\beta_1 = 0$ and $\alpha = 0$ so that the joint distribution of $W_1$ and $W_2$ only depends on T and $\{x_t\}$. The exact functional form of this distribution in terms of T and $\{x_t\}$ is not known.

However $W_1$ and $W_2$ are functions of $\{x_t\}$ through various terms: $\Sigma x_t^2$, $\Sigma x_t x_{t-1}$, $\Sigma x_{t-1}^2$, $\Sigma y_t^2$, $\Sigma y_t y_{t-1}$, $\Sigma y_{t-1}^2$, $\Sigma y_t x_t$, $\Sigma y_t x_{t-1}$, $\Sigma y_{t-1} x_t$, and $\Sigma y_{t-1} x_{t-1}$. Since both $H_1$ and $H_2$ are concerned with the dynamic properties of the $\{y_t\}$ it seems plausible that the distribution of $W_1$ and $W_2$ will depend on the dynamic properties of the $\{x_t\}$. Rescaling the $\{x_t\}$ by a constant is equivalent to rescaling $(\alpha\beta_0+\beta_1)/(\sigma^2)^{\frac{1}{2}}$ by the reciprocal of the constant. The crudest summary statistics for the dynamic structure of the $\{x_t\}$ which are invariant to rescaling the $\{x_t\}$ are the first and second order autocorrelation coefficients and therefore in the numerical evaluation study in Section 6.4 we characterise the dynamic structure of the exogenous data sets in terms of these coefficients.

Another useful property of $W_1$ and $W_{(2)}$ is that if we put $\lambda = -1$ and $\gamma = 0$ in Equation (10) then the signed square-root of $W_1$ changes sign whilst that of $W_{(2)}$ does not. The decomposition vector of $W_2$ which consists of these signed square-roots is given by:

$$\hat{\eta}(y) = [\hat{F}_1(y)\hat{\Omega}_{11}(y)\hat{F}_1(y)']^{-\frac{1}{2}}\phi[\hat{\theta}_1(y)], \qquad (19)$$

where $\hat{F}_1(y)$ denotes the derivative of $\phi(\theta_1)$ with respect to $\theta_1$ evaluated at $\hat{\theta}_1(y)$, where $[\hat{F}_1(y)\hat{\Omega}_{11}(y)\hat{F}_1(y)']^{-\frac{1}{2}}$ is the lower triangular Cholesky decomposition of $[\hat{F}_1(y)\hat{\Omega}_{11}(y)\hat{F}_1(y)']^{-1}$ and where $\phi(\theta_1) = [\alpha\beta_0+\beta_1, \alpha]'$. If we put $\lambda = -1$ and $\gamma = 0$ then:

$$\tilde{\eta}(gy) = (J^+)^{-1}[\tilde{F}_1(y)\hat{\Omega}_{11}\tilde{F}_1'(y)]^{-\frac{1}{2}}(J^+)^{-1}(J^+)\phi[\tilde{\theta}_1(y)]$$

$$= (J^+)^{-1}\tilde{\eta}(y) = J^+\tilde{\eta}(y),$$

where $J^+ = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. (20)

Firstly, since $J^+$ is a diagonal orthogonal matrix, so that $(J^+)^{-1} = (J^+)'$ $= J^+$, the lower triangular Cholesky decomposition of:

$$[\tilde{F}_1(gy)\hat{\Omega}_{11}(gy)\tilde{F}_1(gy)']^{-1} = (J^+)^{-1}[\tilde{F}(y)\hat{\Omega}_{11}(y)\tilde{F}_1(y)']^{-1}(J^+)^{-1},$$

is given by :

$$(J^+)[\tilde{F}_1(y)\hat{\Omega}_{11}(y)F_1(y)']^{-\frac{1}{2}}(J^+)^{-1}. \qquad (21)$$

Secondly, we know from earlier that $\phi[\tilde{\theta}_1(gy)] = J^+[\tilde{\theta}_1(y)]$.

Thus we have:

$$\begin{bmatrix} \tilde{\eta}_1(gy) \\ \tilde{\eta}_2(gy) \end{bmatrix} = \begin{bmatrix} -\tilde{\eta}_1(y) \\ \tilde{\eta}_2(y) \end{bmatrix},$$

and $T\tilde{\eta}_1(y)^2 = W_1$, $T\tilde{\eta}_2(y)^2 = W_{(2)}$.

This symmetry result has several consequences following Appendix 5.A.1 of Chapter 5. Firstly, the distribution of $\tilde{\eta}_1(y)$ is symmetric under the null hypothesis $H_2$ and therefore the first, third and sixth order terms in the Edgeworth expansion of its distribution must all be equal to zero. Thus $d_{111} = 0$ in Equation (14) of Chapter 3 and therefore $b_{111} = 0$ in Equation (48) of Chapter 3 (which is also Equation (27) in this Chapter). Secondly, the symmetry result implies that $d_{122} = 0$ and therefore:

$$b_{122} = (d_{122}d_{122}[9] + d_{112}d_{222}[6])3$$

$$= (d_{112}d_{222}[6])3;$$

$$b_{112} = (d_{112}d_{112}[9] + d_{111}d_{122}[6])3$$

$$= (d_{112}d_{112}[9])3;$$

$$b_{222} = (d_{222}^{2})15; \qquad (22)$$

where the square brackets indicate summation across different permuations of the subscripts. Thus we would anticipate that $b_{112}$ would be positive and that $b_{112}b_{222} = {}^{5}/_{4}\ b_{122}^{2}$; in particular if $d_{222} = 0$ (or equivalently $b_{222} = 0$) then $b_{122} = 0$ also.

Wald tests for $H_1$ have already been investigated by Gregory and Veall (1986) using Monte Carlo methods. As discussed in Chapter 4, they were concerned with the differing behaviour of Wald tests constructed from four alternative formulations of the hypothesis $H_1$:

$$H_{(i)} : \alpha\beta_0 + \beta_1 = 0;$$
$$H_{(ii)} : \beta_0 + (\beta_1/\alpha) = 0;$$
$$H_{(iii)} : \alpha + (\beta_1/\beta_0) = 0;$$
$$H_{(iv)} : (\alpha\beta_0/\beta_1) + 1 = 0. \qquad (23)$$

Gregory and Veall found that none of the tests performed uniformly best over all the different parameter combinations which they considered because the power functions for all the tests varied considerably over the parameter space. However, the statistics $W^3$ & $W^4$ corresponding to $H_{(iii)}$ and $H_{(iv)}$ respectively seemed to perform worse in general than the statistics $W^1$ and $W^2$ corresponding to $H_{(i)}$ and $H_{(ii)}$ in that their rejection probabilities varied more over the parameter space even under the null hypothesis.

This can partly be explained using the invariance result developed in Appendix 6.A. The hypothesis formulation $H_{(i)}$ is the same as that which we have used to construct $W_1$ which gives us an invariant Wald test. If we now transform the parameters as in Equation (11) and substitute the results into the restriction functions for $H_{(ii)}$, $H_{(iii)}$ and $H_{(iv)}$ we obtain:

$$\beta_0^* + (\beta_1^*/\alpha^*) = \lambda[\beta_0 + (\beta_1/\alpha)];$$

$$\alpha^* + (\beta_1^*/\beta_0^*) = [\alpha\beta_0/(\beta_0+\gamma)] + [\beta_1/(\beta_0 + \gamma)];$$

$$(\alpha^*\beta_0^*/\beta_1^*) + 1 = [\alpha\beta_0/(\beta_1-\alpha\gamma)] + [\beta_1/(\beta_1-\alpha\gamma)]. \qquad (24)$$

The restriction function for $H_{(ii)}$ undergoes a linear transformation and therefore meets the fourth condition of the theorem given in Appendix 6.A; however, this is not so for the restriction functions for $H_{(iii)}$ and $H_{(iv)}$. Thus the Wald test from $H_{(ii)}$ is invariant under the group of transformations G* whereas the Wald tests from $H_{(iii)}$ and $H_{(iv)}$ are not.

Gregory and Veall generated the exogenous variable data in their model separately for each replication using an AR(1) process with an autoregressive coefficient of 0.75 together with a standard normal pseudo-random number generator. They set $\sigma^2 = 1$ and then considered ten different parameter combinations for $(\alpha,\beta_0,\beta_1)$:

$$
\begin{array}{llcll}
(I): & (0.5,0.1,-0.05) & ; & (VI): & (0.9,0.1,-0.09); \\
(II): & (0.5,0.5,-0.25) & ; & (VII): & (0.9,0.5,-0.45); \\
(III): & (0.5,1.0,-0.5) & ; & (VIII): & (0.9,1.0,-0.9)\ ; \\
(IV): & (0.5,0.5,-0.1) & ; & (IX): & (0.9,0.5,-0.2)\ ; \\
(V): & (0.5,0.5,-0.4) & ; & (X): & (0.9,0.5,-0.7).
\end{array} \qquad (25)
$$

Combinations (I), (II) and (III) set $\alpha = 0.5$, $(\alpha\beta_0 + \beta_1)^2 = 0$. Therefore we would expect that $W^1$, which is equal to $W_1$ from earlier, would have the same distribution for combinations (I), (II) and (III). This is indeed supported by visual examination of Table 1 in Gregory and Veall (1986) which gives numbers of rejections out of 1000 replications. Similarly we would expect $W^2$ to have the same distribution for combinations (I), (II) and (III), and this is also supported by visual examination of Table 1 in Gregory and Veall (1986). However, the invariance argument does not hold for $W^3$ and $W^4$ and visual examination of Table 1 in Gregory and Veall (1986) does indeed indicate that their distributions do vary considerably across combinations (I), (II) and (III). Combinations (IV) and (V) set $\alpha = 0.5$, $(\alpha\beta_0 + \beta_1)^2 = 0.0225$; combinations (VI), (VII) and (VIII) set $\alpha = 0.9$, $(\alpha\beta_0 + \beta_1)^2 = 0$; and combinations (IX) and (X) set $\alpha = 0.9$, $(\alpha\beta_0 + \beta_1)^2 = 0.0625$. Visual examination of Tables 1 and 2 in

Gregory and Veall (1986) confirms the presence of the invariance properties of the distributions of $W^1$ and $W^2$ for all these parameter combinations and the absence of such invariance properties for the distributions of $W^3$ and $W^4$.

When $\alpha = 0.5$ and $(\alpha\beta_0 + \beta_1) = 0$ the results in Table 1 in Gregory and Veall (1986) suggest that $W_1$ tends to over-reject slightly using the chi-square distribution crictical values. When $\alpha = 0.9$ and $(\alpha\beta_0 + \beta_1) = 0$ then $W_1$ tends to over-reject more strongly using chi-square critical values. However, Gregory and Veall did not consider $\alpha = 0$ and they generated the $\{x_t\}$ separately for each replication rather than conditioning on fixed $\{x_t\}$; therefore, their results are not directly comparable with those of the evaluation study in Section 6.4 of this chapter.

Mizon and Hendry (1980) conducted a Monte Carlo study in the AD(1,1) model with an intercept (using our notation):

$$y_t = \theta + \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t \qquad (26)$$

where the $\epsilon_t$ are distributed $IN(0,\sigma^2)$, and used the Sargan algorithm to test for the COMFAC restriction. In this model the Sargan algorithm is equivalent to constructing a Wald test using the formulation $H_{(1)}$: $\alpha\beta_0 + \beta_1 = 0$ but with numerical rather than analytical derivatives. Mizon and Hendry generated their exogenous variables separately for each replication using an AR(1) process with parameter $\rho$ and with a standard normal pseudo-random number generator. Defining $\gamma = (\alpha\beta_0 + \beta_1)T^{\frac{1}{2}}$, they generated nine separate experiments for each value of

$\gamma$ = (0.0, 0.5, 1.0, 1.5, 2.0) with $\beta_1 = 1$, $\theta = 0$ using a Graeco-Latin 3×3 design generated from T = (25,55,75), $\alpha$ = (0.0, 0.4, 0.8), $\rho$ = (0.3, 0.6, 0.9) and $\sigma^2$ = (0.1, 1.0, 10.0). They found that the Wald test had rejection levels close to its nominal rejection levels using chi-square critical values. However, for T = 75, $\alpha = 0.0$, $\sigma^2 = 10.0$, $\rho = 0.9$ and $\gamma = 0$, the test under-rejected. Mizon and Hendry used only 100 replications for each experiment so that the results for any individual experiment are not very precise. Also their model does contain an intercept term and had exogenous data generated separately for each replication. Therefore this result is not very informative on its own and is not directly comparable with the results of the evaluation study in Section 6.4.

## 6.4 Expansion Results

In order to examine further the properties of the Wald tests for $H_1$: $\alpha\beta_0 + \beta_1 = 0$ and $H_2$: $\alpha\beta_0 + \beta_1 = 0$, $\alpha = 0$ from Section 6.3 we carried out a numerical study using the ESSACS program. We considered three different sets of data generated from the AR(1) process $x_t = \rho x_{t-1} + v_t$ where $v_t = (1-\rho^2)^{\frac{1}{2}} \eta_t$ and $\eta_t \sim IN(0,1)$ with $\rho = 0.2$, $0.5$ and $0.8$. The first and second sample autocorrelations for this series are presented in Table 6.1. We used five different sample sizes, T = 10, 20, 30, 40 and 50, by taking the first 11, the first 21 etc. data points from each series. As can be seen from Table 6.1 the series generated with $\rho = 0.2$ has very weak inherent dynamics, that for $\rho = 0.5$ has moderately strong inherent dynamics, and that for $\rho = 0.8$ has very strong inherent dynamics. These measures of dynamic structure are quite crude as discussed in Section 6.3 but do indicate that the dynamic properties of the three sets of exogenous data vary quite markedly.

In Table 6.2 we report the expansion approximation coefficients multiplied by the sample size for the cdf approximation:

$$\begin{aligned}
\Pr\{W_1 \le s_1^2 \ \& \ W_{(2)} \le s_2^2\} &= (1+b_0)\Gamma_1^2(s_1^2)\Gamma_1^2(s_2^2) \\
&+ b_1\Gamma_3^2(s_1^2)\Gamma_1^2(s_2^2) + b_2\Gamma_1^2(s_1^2)\Gamma_3^2(s_2^2) \\
&+ b_{11}\Gamma_5^2(s_1^2)\Gamma_1^2(s_2^2) + b_{12}\Gamma_3^2(s_1^2)\Gamma_3^2(s_2^2) \\
&+ b_{22}\Gamma_1^2(s_1^2)\Gamma_5^2(s_2^2) + b_{111}\Gamma_7^2(s_1^2)\Gamma_1^2(s_2^2) \\
&+ b_{112}\Gamma_5^2(s_1^2)\Gamma_3^2(s_s^2) + b_{122}\Gamma_3^2(s_1^2)\Gamma_5^2(s_2^2) \\
&+ b_{222}\Gamma_1^2(s_1^2)\Gamma_7^2(s_2^2) + o(T^{-1}), \qquad (27)
\end{aligned}$$

| | | $r_1$ | $r_2$ |
|---|---|---|---|
| | T = 10 | 0.0995 | -0.0476 |
| | 20 | 0.0454 | 0.0501 |
| ρ = 0.2 | 30 | 0.0305 | 0.0188 |
| | 40 | 0.0732 | 0.1330 |
| | 50 | 0.0293 | 0.0821 |
| | T = 10 | 0.3172 | -0.0018 |
| | 20 | 0.3612 | 0.1364 |
| ρ = 0.5 | 30 | 0.3411 | 0.0970 |
| | 40 | 0.4675 | 0.3028 |
| | 50 | 0.3822 | 0.2568 |
| | T = 10 | 0.8065 | 0.0386 |
| | 20 | 0.9470 | 0.4584 |
| ρ = 0.8 | 30 | 0.9975 | 0.6606 |
| | 40 | 1.0280 | 0.7980 |
| | 50 | 0.9421 | 0.7461 |

Table 6.1: First and Second Order Order Sample Autocorrelation Coefficients for the 3 Exogenous Data Sets

n.b. $r_1 = [\sum_{t=1}^{T}(x_t-\bar{x})(x_{t-1}-\bar{x})]/[\sum_{t=0}^{T}(x_t-\bar{x})^2]$

$\bar{x} = (T+1)^{-1}\sum_{t=0}^{T} x_t$

| | | $b_0T$ | $b_1T$ | $b_2T$ | $b_{11}T$ | $b_{12}T$ | $b_{22}T$ | $b_{111}T$ | $b_{112}T$ | $b_{122}T$ | $b_{222}T$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | T = 10 | 0.8427 | -0.8730 | -1.3651 | 0.6844 | -0.0670 | 0.75 | 0 | 0.0280 | 0 | 0 |
| | 20 | 1.0207 | -1.1643 | -1.4333 | 0.7336 | 0.0795 | 0.75 | 0 | 0.0139 | 0 | 0 |
| $\rho$=0.2 | 30 | 1.0554 | -1.1776 | -1.4604 | 0.7294 | 0.0896 | 0.75 | 0 | 0.0136 | 0 | 0 |
| | 40 | 1.1064 | -1.2684 | -1.4639 | 0.7323 | 0.1334 | 0.75 | 0 | 0.0102 | 0 | 0 |
| | 50 | 1.0489 | -1.1761 | -1.5267 | 0.7406 | 0.1589 | 0.75 | 0 | 0.0046 | 0 | 0 |
| | T = 10 | 0.8242 | -0.6924 | -1.4373 | 0.3202 | 0.0179 | 0.75 | 0 | 0.2173 | 0 | 0 |
| | 20 | 1.0504 | -1.0634 | -1.5249 | 0.3889 | 0.1715 | 0.75 | 0 | 0.2276 | 0 | 0 |
| $\rho$=0.5 | 30 | 1.1095 | -1.0799 | -1.5422 | 0.2843 | 0.2253 | 0.75 | 0 | 0.2530 | 0 | 0 |
| | 40 | 1.1500 | -1.1670 | -1.5510 | 0.2842 | 0.2671 | 0.75 | 0 | 0.2667 | 0 | 0 |
| | 50 | 1.0799 | -1.1251 | -1.5946 | 0.3755 | 0.3102 | 0.75 | 0 | 0.2041 | 0 | 0 |
| | T = 10 | 0.8312 | -0.5048 | -1.5541 | -0.0152 | 0.1123 | 0.75 | 0 | 0.3806 | 0 | 0 |
| | 20 | 1.0723 | -0.8830 | -1.6856 | -0.0885 | 0.3225 | 0.75 | 0 | 0.5123 | 0 | 0 |
| $\rho$=0.8 | 30 | 1.1316 | -0.7648 | -1.7461 | -0.7598 | 0.5460 | 0.75 | 0 | 0.8430 | 0 | 0 |
| | 40 | 1.1925 | -0.8572 | -1.7509 | -0.7098 | 0.5420 | 0.75 | 0 | 0.8334 | 0 | 0 |
| | 50 | 1.1349 | -0.8847 | -1.7623 | -0.4925 | 0.5441 | 0.75 | 0 | 0.6905 | 0 | 0 |

Table 6.2: Asymptotic Expansion Approximation Coefficients multiplied by the Sample Size

where $W_1$ is the test for $\alpha\beta_0 + \beta_1 = 0$, $W_{(2)}$ is the incremental test for $\alpha = 0$, and $\Gamma_k^2(s^2)$ is the cdf of a chi-square variate with k degrees of freedom evaluated as $s^2$. From Table 6.2 we can note a number of points. Firstly, $b_{111} = 0$, $b_{122} = 0$, $b_{222} = 0$ and $b_{22} = (3/4)T^{-1}$ in all cases. As noted in Section 6.3, $b_{111} = 0$ follows from the symmetry of the distribution of signed square-root of $W_1$. Also since numerically $b_{222} = 0$ then $b_{122} = 0$ also follows from the symmetry arguments of Section 6.3; however it is not clear why $b_{222} = 0$ or why $b_{22} = (3/4)T^{-1}$. Secondly, $b_0$ is approximately equal to $T^{-1}$, $b_1$ is approximately equal to $-T^{-1}$, and $b_2$ is approximately equal to $-(3/2)T^{-1}$. This is in contrast to $b_{11}$, $b_{12}$ and $b_{112}$ which all vary markedly across the different data sets and across the different sample sizes for any given data set. In Table 6.3 we report the main quantitative features of the expansion coefficients.

The evaluated approximate rejection probabilities are given in Tables 6.4, 6.5, 6.6 and 6.7, and from these we can also make a number of qualitative observations. In Table 6.4 we report the approximate rejection probabilities, expressed as percentages for $W_1$, the test statistic for $H_1$: $\alpha\beta_0 + \beta_1 = 0$. These are evaluated at the 10%, 5% and 1% nominal significance levels using two sets of critical values: firstly, those for the chi-square distribution with one degree of freedom; and, secondly, those for the F-distribution with (1,T-3) degrees of freedom. For T = 30, 40 and 50, the F-distribution critical

| | $\rho = 0.2$ | $\rho = 0.5$ | $\rho = 0.8$ |
|---|---|---|---|
| $b_0$ | $\approx T^{-1}$ | $\approx T^{-1}$ | $\approx T^{-1}$ |
| $b_1$ | $\approx -T^{-1}$ | $\approx -T^{-1}$ | $\approx -T^{-1}$ |
| $b_2$ | $\approx -(3/2)T^{-1}$ | $\approx -(3/2)T^{-1}$ | $\approx -(3/2)T^{-1}$ |
| $b_{11}$ | $\approx 0.7T^{-1}$ | $\approx 0.3T^{-1}$ | negative and varies |
| $b_{12}$ | $\approx 0$ | $\approx 0.2T^{-1}$ | $\approx (1/2)T^{-1}$ |
| $b_{22}$ | $(3/4)T^{-1}$ | $(3/4)T^{-1}$ | $(3/4)T^{-1}$ |
| $b_{111}$ | 0 | 0 | 0 |
| $b_{112}$ | $\approx 0$ | $\approx 0.2T^{-1}$ | $\approx 0.7T^{-1}$ but varies |
| $b_{122}$ | 0 | 0 | 0 |
| $b_{222}$ | 0 | 0 | 0 |

Table 6.3: Qualitative Behaviour of Expansion Coefficients for Different Data Sets

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 11.41 | 6.57 | 2.00 |
| | F | 7.38 | 3.07 | 0.21 |
| T = 20, | $\chi^2$ | 10.57 | 5.71 | 1.49 |
| | F | 8.83 | 4.19 | 0.68 |
| T = 30, | $\chi^2$ | 10.37 | 5.46 | 1.32 |
| | F | 9.25 | 4.48 | 0.81 |
| T = 40, | $\chi^2$ | 10.24 | 5.32 | 1.23 |
| | F | 9.41 | 4.60 | 0.86 |
| T = 50, | $\chi^2$ | 10.27 | 5.31 | 1.20 |
| | F | 9.63 | 4.74 | 0.91 |

Case I: Data Set with $\rho = 0.2$

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 11.18 | 6.26 | 1.78 |
| | F | 7.09 | 2.80 | 0.17 |
| T = 20, | $\chi^2$ | 10.48 | 5.59 | 1.41 |
| | F | 8.72 | 4.07 | 0.63 |
| T = 30, | $\chi^2$ | 10.19 | 5.28 | 1.22 |
| | F | 9.07 | 4.31 | 0.73 |
| T = 40, | $\chi^2$ | 10.13 | 5.20 | 1.16 |
| | F | 9.30 | 4.48 | 0.80 |
| T = 50, | $\chi^2$ | 10.20 | 5.23 | 1.16 |
| | F | 9.55 | 4.66 | 0.86 |

Case II: Data Set with $\rho = 0.5$

Table 6.4: Evaluated Approximate Rejection Probabilities expressed as %'s for $W_1$

| | | Nominal Significiance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 11.03 | 6.01 | 1.58 |
| | F | 6.85 | 2.56 | 0.14 |
| T = 20, | $\chi^2$ | 10.42 | 5.46 | 1.30 |
| | F | 8.64 | 3.93 | 0.56 |
| T = 30, | $\chi^2$ | 9.93 | 4.98 | 1.01 |
| | F | 8.79 | 4.01 | 0.57 |
| T = 40, | $\chi^2$ | 9.93 | 4.98 | 1.02 |
| | F | 9.10 | 4.26 | 0.68 |
| T = 50, | $\chi^2$ | 10.04 | 5.06 | 1.05 |
| | F | 9.39 | 4.49 | 0.77 |

Case III: Data Set with $\rho$ = 0.8

Table 6.4 (contd.)

values were interpolated from the F(1,T-2) and F(1,T-4) critical values. However, following the argument of Appendix 5.A.3 of Chapter 5, the differences between the approximate rejection probabilities using F(1, T-2) and F(1, T-4) critical values should only be an $o(T^{-1})$ term and therefore the effects of the interpolation should not be too serious. It is clear that in all cases the approximate rejection probabilities are closer to the nominal significance levels for the chi-square distribution critical values than for the F-distribution critical values. The test based on $W_1$ appears to over-reject when used with chi-square critical values and to under-reject when used with the F-distribution critical values. The deviations of approximate rejection probabilities from nominal significance levels decrease as T is increased, except when T increases from 40 to 50 using the chi-square distribution critical values. The probability of rejection appears to decrease as the inherent dynamics of the exogenous data are strengthened. i.e. by shifting from the data set with $\rho = 0.2$ to that with $\rho = 0.5$ and then to that with $\rho = 0.8$.

In Table 6.5 we report the evaluated approximate rejection probabilities for $W_{(2)}$, the incremental test of $\alpha = 0$ after testing $\alpha\beta_0 + \beta_1 = 0$. As with $W_1$, the test using chi-square distribution critical values appears to over-reject whereas that with F-distribution critical values appears to under-reject; also the approximate rejection probabilities are closer to the nominal significance levels for the chi-square distribution critical values than for the F-distribution critical values. The deviations of approximate rejection probabilities from nominal significance levels decrease as T is increased, except

Case I: Data Set with $\rho = 0.2$

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 10.08 | 5.70 | 1.75 |
| | F | 6.43 | 2.66 | 0.19 |
| T = 20, | $\chi^2$ | 10.15 | 5.42 | 1.40 |
| | F | 8.45 | 3.96 | 0.64 |
| T = 30, | $\chi^2$ | 10.08 | 5.27 | 1.26 |
| | F | 8.98 | 4.31 | 0.77 |
| T = 40, | $\chi^2$ | 10.09 | 5.22 | 1.20 |
| | F | 9.27 | 4.51 | 0.83 |
| T = 50, | $\chi^2$ | 10.04 | 5.16 | 1.16 |
| | F | 9.41 | 4.60 | 0.87 |

Case II: Data Set with $\rho = 0.5$

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 10.76 | 6.16 | 1.90 |
| | F | 6.93 | 2.89 | 0.21 |
| T = 20, | $\chi^2$ | 10.51 | 5.67 | 1.48 |
| | F | 8.77 | 4.16 | 0.67 |
| T = 30, | $\chi^2$ | 10.41 | 5.49 | 1.33 |
| | F | 9.29 | 4.51 | 0.81 |
| T = 40, | $\chi^2$ | 10.35 | 5.40 | 1.26 |
| | F | 9.52 | 4.67 | 0.88 |
| T = 50, | $\chi^2$ | 10.24 | 5.29 | 1.20 |
| | F | 9.59 | 4.72 | 0.90 |

Table 6.5: Evaluated Approximate Rejection Probabilities expressed as %'s for $W_{(2)}$

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| $T = 10,$ | $\chi^2$ | 11.24 | 6.49 | 2.00 |
| | F | 7.29 | 3.06 | 0.21 |
| $T = 20,$ | $\chi^2$ | 10.98 | 5.98 | 1.58 |
| | F | 9.20 | 4.41 | 0.72 |
| $T = 30,$ | $\chi^2$ | 11.21 | 6.03 | 1.51 |
| | F | 10.04 | 4.98 | 0.93 |
| $T = 40,$ | $\chi^2$ | 10.89 | 5.76 | 1.38 |
| | F | 10.04 | 5.00 | 0.96 |
| $T = 50,$ | $\chi^2$ | 10.62 | 5.55 | 1.28 |
| | F | 9.97 | 4.96 | 0.97 |

Case III: Data Set with $\rho = 0.8$

Table 6.5: (contd.)

when T increases from 10 to 20 and from 30 to 40 using the nominal 10% chi-square critical value and the $\rho = 0.2$ data set and except when T increases from 20 to 30 using the 10% and 5% chi-square critical values and the $\rho = 0.8$ data set. However with the $\rho = 0.2$ data set the 10% chi-square critical values work extremely well. Unlike the results for $W_1$, the approximate rejection probabilities for $W_{(2)}$ increase as the inherent dynamics of the exogenous data are strengthened and the effect of varying the inherent dynamics appears more marked for $W_{(2)}$ than for $W_1$.

In Table 6.6 we report the evaluated approximate rejection probabilities for $W_2$, the test statistic for $H_2$: $\alpha\beta_0+\beta_1 = 0$, $\alpha = 0$. These are evaluated at the 10%, 5% and 1% nominal significance levels using two sets of critical values: firstly, those for the chi-square distribution with two degrees of freedom; and, secondly, two times those for the F-distribution with (2, T-3) degrees of freedom (again interpolated for T = 30, 40, 50). As with $W_1$ and $W_{(2)}$, the test appears to over-reject with the chi-square distribution critical values and to under-reject with the F-distribution based critical values; also the chi-square distribution critical values appear to work better in general. The deviations of approximate rejection probabilities from nominal significance levels decrease as T increases, and strengthening the inherent dynamics of the exogenous data appears to increase the probability of rejection. However this increase is much weaker than the decrease observed for $W_1$ or the increase observed for $W_{(2)}$.

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 11.82 | 6.93 | 2.14 |
| | F | 5.70 | 1.94 | 0.05 |
| T = 20, | $\chi^2$ | 10.93 | 6.02 | 1.61 |
| | F | 8.12 | 3.67 | 0.48 |
| T = 30, | $\chi^2$ | 10.60 | 5.67 | 1.40 |
| | F | 8.76 | 4.13 | 0.67 |
| T = 40, | $\chi^2$ | 10.45 | 5.50 | 1.31 |
| | F | 9.09 | 4.36 | 0.76 |
| T = 50, | $\chi^2$ | 10.40 | 5.43 | 1.26 |
| | F | 9.10 | 4.52 | 0.82 |

Case I: Data Set with $\rho = 0.2$

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 12.00 | 7.18 | 2.36 |
| | F | 5.96 | 2.15 | 0.07 |
| T = 20, | $\chi^2$ | 11.06 | 6.19 | 1.76 |
| | F | 8.28 | 3.84 | 0.56 |
| T = 30, | $\chi^2$ | 10.66 | 5.77 | 1.50 |
| | F | 8.84 | 4.24 | 0.74 |
| T = 40, | $\chi^2$ | 10.51 | 5.60 | 1.39 |
| | F | 9.16 | 4.46 | 0.83 |
| T = 50, | $\chi^2$ | 10.45 | 5.50 | 1.31 |
| | F | 9.16 | 4.59 | 0.86 |

Case II: Data Set with $\rho = 0.5$

Table 6.6: Evaluated Approximate Rejection Probabilities expressed as %'s for $W_2$

| | | Nominal Significance Level | | |
|---|---|---|---|---|
| | | 10% | 5% | 1% |
| T = 10, | $\chi^2$ | 12.11 | 7.36 | 2.54 |
| | F | 6.16 | 2.33 | 0.08 |
| T = 20, | $\chi^2$ | 11.27 | 6.45 | 1.96 |
| | F | 8.52 | 4.10 | 0.67 |
| T = 30, | $\chi^2$ | 10.92 | 6.10 | 1.77 |
| | F | 9.13 | 4.57 | 0.93 |
| T = 40, | $\chi^2$ | 10.68 | 5,81 | 1.58 |
| | F | 9.34 | 4.68 | 0.98 |
| T = 50, | $\chi^2$ | 10.56 | 5.65 | 1.44 |
| | F | 9.28 | 4.74 | 0.97 |

**Case III: Data Set with $\rho = 0.8$**

Table 6.6 (contd.)

| Nominal Significance Levels | | Sample Size | | | | | Sample Size | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\delta$ | 10 | 20 | 30 | 40 | 50 | 10 | 20 | 30 | 40 | 50 |
| | 10% | 10.09 | 10.08 | 10.03 | 10.04 | 10.00 | 6.43 | 8.39 | 8.93 | 9.23 | 9.37 |
| 10% | 5% | 10.08 | 10.10 | 10.05 | 10.06 | 10.02 | 6.42 | 8.41 | 8.95 | 9.25 | 9.38 |
| | 1% | 10.08 | 10.13 | 10.07 | 10.08 | 10.03 | 6.42 | 8.44 | 8.97 | 9.27 | 9.40 |
| | 10% | 5.71 | 5.38 | 5.24 | 5.19 | 5.13 | 2.66 | 3.93 | 4.29 | 4.48 | 4.57 |
| 5% | 5% | 5.70 | 5.39 | 5.25 | 5.20 | 5.14 | 2.66 | 3.94 | 4.30 | 4.49 | 4.58 |
| | 1% | 5.70 | 5.41 | 5.26 | 5.22 | 5.15 | 2.66 | 3.94 | 4.31 | 4.51 | 4.59 |
| | 10% | 1.75 | 1.38 | 1.25 | 1.19 | 1.15 | 0.19 | 0.63 | 0.76 | 0.83 | 0.86 |
| 1% | 5% | 1.75 | 1.39 | 1.25 | 1.20 | 1.15 | 0.19 | 0.63 | 0.76 | 0.83 | 0.86 |
| | 1% | 1.75 | 1.39 | 1.26 | 1.20 | 1.15 | 0.19 | 0.64 | 0.77 | 0.83 | 0.87 |
| | | Using $x^2$ - Distribution Critical Values | | | | | Using F - Distribution Critical Values | | | | |

Table 6.7: Evaluated Approximate Rejection Probabilities expressed as %'s for Sequential Tests, Case I: Data Set with $\rho = 0.2$

| Nominal Significance Levels | | Sample Size | | | | | Sample Size | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 | 20 | 30 | 40 | 50 | 10 | 20 | 30 | 40 | 50 |
| α | δ | | | | | | | | | | |
| | 10% | 10.21 | 10.12 | 10.11 | 10.10 | 10.06 | 6.60 | 8.45 | 9.02 | 9.29 | 9.43 |
| 10% | 5% | 10.34 | 10.23 | 10.19 | 10.17 | 10.11 | 6.74 | 8.56 | 9.10 | 9.36 | 9.48 |
| | 1% | 10.58 | 10.39 | 10.32 | 10.28 | 10.19 | 6.92 | 8.72 | 9.23 | 9.47 | 9.56 |
| | 10% | 5.79 | 5.41 | 5.29 | 5.23 | 5.17 | 2.74 | 3.97 | 4.35 | 4.52 | 4.61 |
| 5% | 5% | 5.88 | 5.48 | 5.35 | 5.28 | 5.20 | 2.81 | 4.03 | 4.39 | 4.57 | 4.64 |
| | 1% | 6.04 | 5.59 | 5.43 | 5.35 | 5.26 | 2.89 | 4.12 | 4.47 | 4.63 | 4.69 |
| | 10% | 1.78 | 1.39 | 1.27 | 1.20 | 1.16 | 0.20 | 0.64 | 0.77 | 0.84 | 0.87 |
| 1% | 5% | 1.81 | 1.42 | 1.29 | 1.22 | 1.17 | 0.20 | 0.65 | 0.79 | 0.85 | 0.88 |
| | 1% | 1.86 | 1.45 | 1.31 | 1.24 | 1.19 | 0.21 | 0.67 | 0.81 | 0.87 | 0.89 |
| | | Using $\chi^2$ - distribution Critical Values | | | | | Using F-Distribution Critical Values | | | | |

Table 6.7 (contd.), Case II: Data Set with $\rho = 0.5$

| Nominal Significance Levels | | Sample Size | | | | | Sample Size | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 | 20 | 30 | 40 | 50 | 10 | 20 | 30 | 40 | 50 |
| α | δ | | | | | | | | | | |
| | 10% | 10.17 | 10.15 | 10.29 | 10.21 | 10.14 | 6.65 | 8.51 | 9.22 | 9.41 | 9.52 |
| 10% | 5% | 10.44 | 10.37 | 10.54 | 10.39 | 10.27 | 6.93 | 9.74 | 9.47 | 9.60 | 9.65 |
| | 1% | 10.90 | 10.73 | 10.93 | 10.69 | 10.48 | 7.26 | 9.07 | 9.86 | 9.89 | 9.86 |
| | 10% | 5.76 | 5.42 | 5.41 | 5.31 | 5.23 | 2.77 | 4.00 | 4.47 | 4.60 | 4.67 |
| 5% | 5% | 5.95 | 5.57 | 5.58 | 5.43 | 5.31 | 2.89 | 4.13 | 4.62 | 4.82 | 4.86 |
| | 1% | 6.26 | 5.81 | 5.85 | 5.63 | 5.46 | 3.04 | 4.34 | 4.87 | 4.90 | 4.89 |
| | 10% | 1.77 | 1.40 | 1.31 | 1.23 | 1.18 | 0.20 | 0.64 | 0.80 | 0.86 | 0.89 |
| 1% | 5% | 1.83 | 1.45 | 1.36 | 1.27 | 1.21 | 0.21 | 0.67 | 0.84 | 0.89 | 0.91 |
| | 1% | 1.93 | 1.53 | 1.45 | 1.33 | 1.25 | 0.21 | 0.71 | 0.90 | 0.94 | 0.95 |
| | | Using $\chi^2$ - Distribution Critical Values | | | | | Using F-Distribution Critical Values | | | | |

Table 6.7 (contd.), Case III: Data Set with $\rho = 0.8$

Note on Table 6.7

This table reports evaluated approximate rejection probabilities for a test based on $W_{(2)}$ with nominal significance level $\alpha$ conditional upon $W_1$ lying in the acceptance region for a test with nominal significance level $\delta$. Nine different combinations for $\alpha$ and $\delta$ are given with $\alpha, \delta$ = 10%, 5% and 1%. Two sets of results are given for each data set: firstly, where chi-square distribution critical values are used throughout the sequence; and, secondly, where F-distribution critical values are used throughout the sequence.

In Table 6.7 we report the evaluated approximate rejection probabilities for the sequential test based on $W_{(2)}$ conditional upon $W_1$ lying in the acceptance region for an earlier test in the sequence. These exhibit the same qualitative results as do the evaluated approximate rejection probabilities for $W_{(2)}$. However an additional feature appears when we vary the nominal significance level of the intitial test in the sequence. For the $\rho = 0.2$ data set, changing the nominal significance level of the initial test has very little effect on the approximate rejection probability of the subsequent test. However for the $\rho = 0.5$ data set, lowering the nominal significance level of the intial test tends to decrease the approximate rejection probability of the subsequent test and for the $\rho = 0.8$ data set this effect becomes more marked. This effect can be explained by noting that for the $\rho = 0.2$ data set both $T\ b_{12}$ and $T\ b_{112}$ are close to zero whereas for the $\rho = 0.5$ and the $\rho = 0.8$ data sets they are noticeably greater than zero. If $T\ b_{12} = 0$ and $T\ b_{112} = 0$ then $W_1$ and $W_{(2)}$ would be independent to $o(T^{-1})$ and so changing the nominal significance level of the initial test would have no effect on the approximate rejection probability of the subsequent test. Increasing $T\ b_{12}$ and $T\ b_{112}$ increases the dependence to $o(T^{-1})$ of the two tests so that we would expect the effect noted above to be more pronounced.

We also observe that strengthening the inherent dynamics of the exogenous data has a weaker effect on the approximate rejection probability the higher is the nominal significance level of the initial test. This seems to be a consequence of the effect noted above:

strengthening the inherent dynamics of the exogenous data increases the approximate rejection probability of the unconditional test based on $W_{(2)}$ but this is offset for the conditional test based on $W_{(2)}$ by the increased joint dependence of $W_1$ and $W_{(2)}$ and the decreased approximate rejection probability of the test based on $W_1$. The higher the nominal significance level of the initial test, the stronger is this offsetting effect. This also helps to explain why strengthening the inherent dynamics of the exogenous data appears to have a weaker effect on the approximate rejection probabilities of tests based on $W_2$ than on those of tests based on $W_1$ and $W_{(2)}$.

## 6.5 Conclusions

In this chapter we have studied the properties of testing for $H_1$: $\alpha\beta_0 + \beta_1 = 0$, the COMFAC restriction, and $H_2$: $\alpha\beta_0 + \beta_1 = 0$, $\alpha = 0$, the purely static model restriction, in the AD(1,1) model. We have shown in Section 6.3 that the joint distribution of the Wald statistics $W_1$ and $W_2$ for $H_1$ and $H_2$ using OLS estimates is determined by $T, \{x_t\}, \alpha$ and $(\alpha\beta_0 + \beta_1)^2/\sigma^2$; thus under the null hypothesis $H_2$ the joint distribution is determined solely by T and $\{x_t\}$. In Section 6.4 we have performed a numerical study based on asymptotic expansion approximations to the joint cdf of $W_1$ and $W_{(2)} = W_2 - W_1$. This study seems to indicate that the inherent dynamic structure of the exogenous data is very important as a determinant of the joint distribution of $W_1$ and $W_{(2)}$ and has systematic effects on the rejection probabilities of the tests based on $W_1$ and $W_{(2)}$. In particular, strengthening the inherent dynamics of the exogenous data series appears to strengthen the dependence between $W_1$ and $W_{(2)}$ which makes using sequential testing procedures more difficult.

There are a number of possible directions for future research in this area. Firstly, it would be very useful to known to which functions of the exogenous data are important in determining the properties of the tests and how these affect the properties of the tests. This would help to elucidate the difficulties which are likely to occur when applying such tests with various data sets. Secondly, it would be helpful to examine the power properties of such tests; these could be analysed using the techniques developed in Chapter 3. Thirdly, it would be helpful to examine the consequences of using alternative Wald statistics, e.g. those based on $H_1^*$: $\beta_0 + (\beta_1/\alpha) = 0$ and $H_2^*$: $\beta_0 + (\beta_1/\alpha) = 0$, $\alpha = 0$, as is briefly discussed in Section 6.3.

Lastly, it would be helpful to examine the properties of such tests in more general AD models. However, as noted in Section 6.3, increasing the number of lags would create serious problems since it may alter the asymptotic distributions of the test statistics involved; thus it would require the development of asymptotic expansions for test statistics with asymptotic distributions which are not chi-square distributions.

It is important to note that the tables given in this chapter are only tables of approximate probabilities. At best they have approximation errors of $o(T^{-1})$ although we have not demonstrated the validity of the asymptotic expansions in this case. It is therefore possible that the changes in behaviour we observe in the approximate probabilities arising from varying the exogenous variables do not adequately reflect the changes in the true probabilities from varying the exogenous variables. Subject to this caveat, the main conclusion of this study is that the dynamic properties of the exogenous variables seem very important in determining the properties of sequential testing procedures in dynamic models.

## 6.A Appendix: An Invariance Theorem for Wald Tests

Suppose that, as in Chapter 1, Y is a (vector) random variable with distribution function $F(y;\theta)$ where $\theta \in \Theta$. In certain cases it may be possible to find a group of transformations G, operating on Y, and an induced group of transformation G*, operating on θ, such that for any $g \in G$ there exists a unique $g^* \in G^*$ with the following property: that if $Y \sim F(y;\theta)$ then $Z = gY \sim F(z;g^*\theta)$ where $g^*\theta \in \Theta$ for any $\theta \in \Theta$.

If we wish to test $H_0$: $\theta \in \Theta_0$ against $H_A$: $\theta \in \Theta_A$ where $\Theta_0$ and $\Theta_A$ are disjoint subsets of $\Theta$ then providing $\theta \in \Theta_0$ iff $g^*\theta \in \Theta_0$, and $\theta \in \Theta_A$ iff $g^*\theta \in \Theta_A$ we can say that this hypothesis testing problem is invariant under the groups of transformations G and G*; see Cox and Hinckley (1974), Lehmann (1959).

In econometrics we are frequently interested in testing functic restrictions upon a subset of the parameters and therefore we will assume that we can partition $\theta = (\theta_1', \theta_2')'$ and that we can represent $\theta \in \Theta_0$ by $\psi(\theta_1) = 0$ and $\theta \in \Theta_A$ by $\psi(\theta_1) \neq 0$. Provided that $\psi$ (.) is a differentiable function we can consider formulating a Wald test of $H_0$ against $H_A$. It is well known that the likelihood ratio test invariant under the groups of transformations G and G*; see Cox anc Hinkley (1974). It is interesting to ask whether it is also possibl to construct an invariant Wald test. The following theorem gives sufficient conditions for this to be possible.

**Theorem 1**

Under the following conditions there exists an invariant Wald test of $H_0$ under the groups of transformations G and G*:

A1) For any $g^* \varepsilon G^*$ there exist $g_1^*$ and $g_2^*$ such that $g^*\theta = [(g_1^*\theta_1)', (g_2^*\theta\ )']'$ which is partitioned conformably with $(\theta_1', \theta_2')'$.

A2) There exists an estimator $\hat{\theta}_1(y)$ of $\theta_1$ such that:

$$\hat{\theta}_1(gy) = g_1^*\hat{\theta}_1(y).$$

A3) There exists an estimator $\hat{\Omega}_{11}(y)$ of $\Omega_{11}$ the asymptotic covariance matrix of $T^{\frac{1}{2}}[\hat{\theta}_1(y) - \theta_1]$ such that:

$$\hat{\Omega}_{11}(gy) = \left[\frac{\partial g_1^*}{\partial \theta_1'}\bigg|\hat{\theta}_1(y)\right] \hat{\Omega}_{11}(y) \left[\frac{\partial g_1^*}{\partial \theta_1'}\bigg|\hat{\theta}_1(y)\right].$$

A4) There exists a representation $\phi(\theta_1) = 0$ of $\theta \varepsilon \Theta_0$ such that $\phi(\phi_1) = 0$ if and only if $\theta \varepsilon\ \Theta_0$;

and there exists a group of matrix transformations $G^+$ such that for all $g^* \varepsilon G^*$ there exists a matrix $J^+ \varepsilon G^+$ so that $\phi(g_1^*\ \theta_1) = J^+\phi(\theta_1)$ for all $(\theta_1', \theta_2')' \varepsilon \Theta$.

The invariant Wald test takes the form:

$$\hat{W}\ (y) = T\phi[\hat{\theta}_1(y)]'[\hat{F}_1(y)\hat{\Omega}_{11}(y)\hat{F}_1(y)']^{-1}\phi[\hat{\theta}_1(y)]$$

where $\hat{F}_1(y) = \left[\frac{\partial \phi}{\partial \theta_1'}\bigg|\hat{\theta}_1(y)\right]$.

Proof:

Consider firstly $\phi[\hat{\theta}_1(gy)] - \phi[g_1^*\hat{\theta}_1(y)] = J^+\phi[\hat{\theta}_1(y)]$.

Next consider $\hat{F}_1(gy)$:

$$\hat{F}_1(gy) = \left[\frac{\partial\phi}{\partial\theta_1'}\bigg|_{\hat{\theta}_1(gy)}\right] = \left[\frac{\partial\phi}{\partial\theta_1'}\bigg|_{\theta_1=g_1^*\hat{\theta}_1(y)}\right]$$

$$= \left[\frac{\partial\phi(g_1^*\theta_1)}{\partial(g_1^*\theta_1)'}\bigg|_{g_1^*\theta_1=g_1^*\hat{\theta}_1(y)}\right]$$

$$= \left[\frac{\partial\phi(g_1^*\theta_1)}{\partial\theta_1'}\bigg|_{\theta_1=\hat{\theta}_1(y)}\right]\left[\frac{\partial g_1^*}{\partial\theta_1'}\bigg|_{\theta_1=\hat{\theta}_1(y)}\right]^{-1}$$

$$= \left[\frac{\partial(J^+\phi[\theta_1])}{\partial\theta_1'}\bigg|_{\theta_1=\hat{\theta}_1(y)}\right]\left[\frac{\partial g_1^*}{\partial\theta_1'}\bigg|_{\theta_1=\hat{\theta}_1(y)}\right]^{-1}$$

$$= J^+\left[\frac{\partial\phi}{\partial\theta_1'}\bigg|_{\theta_1=\hat{\theta}_1(y)}\right]\left[\frac{\partial g_1^*}{\partial\theta_1'}\bigg|_{\theta_1=\hat{\theta}_1(y)}\right]^{-1}$$

$$= J^+\hat{F}_1(y)\left[\frac{\partial g_1^*}{\partial\theta_1'}\bigg|_{\theta_1=\hat{\theta}_1(y)}\right]^{-1}. \qquad (28)$$

Substituting this into $[\hat{F}_1(gy)\ \hat{\Omega}_{11}(gy)\hat{F}_1(gy)']^{-1}$ gives:

$$[\hat{F}_1(gy)\ \hat{\Omega}_{11}(gy)\hat{F}_1(gy)']^{-1} = [J^+\hat{F}_1(y)\ \hat{\Omega}_{11}(y)\hat{F}_1(y)'J^{+\prime}]^{-1}$$

$$= (J^{+\prime})^{-1}[\hat{F}_1(y)\ \hat{\Omega}_{11}(y)\hat{F}_1(y)']^{-1}(J^+)^{-1}$$

Substituting this and Equation (28) into $\hat{W}$ (gy) gives:

$$\hat{W}(gy) = T\phi[\hat{\theta}_1(y)]'J^{+\prime}(J^{+\prime})^{-1}[\hat{F}_1(y)\hat{\Omega}_{11}(y)\hat{F}_1(y)']^{-1}(J^{+})^{-1}J^{+}\phi[\hat{\theta}_1(y)]$$

$$= T\phi[\hat{\theta}_1(y)]'[\hat{F}_1(y)\hat{\Omega}_{11}(y)\hat{F}_1(y)']^{-1}\phi[\hat{\theta}_1(y)]$$

$$= \hat{W}(y). \qquad (29)$$

Thus $\hat{W}$ is invariant under all transformations $g \in G$ operating on Y and its distribution is invariant under all transformations $g^* \in G^*$ operating on $\theta$. Q.E.D.

If we specify $\theta_1 \equiv \theta$ then this theorem is applicable to testing functional restrictions on the full set of parameters.

## CHAPTER 7: Conclusion

Asymptotic expansion approximations to the distributions of test statistics have not been very widely used in econometrics for studying the properties of testing procedures in spite of the known inadequacies of first-order asymptotic theory in many cases. This appears to be the result mainly of the perceived difficulty in deriving and computing such approximations. In this thesis we have attempted to demonstrate that these approximations can be used to help solve actual problems of interest in econometrics.

In order to do this we have developed a method for deriving and computing asymptotic expansion approximations to the joint distributions of asymptotically independent chi-square statistics which can be exactly or approximately decomposed as the inner products of asymptotically standard normal vectors under the null hypothesis. We have also extended this method to approximate the joint distributions of such statistics under local alternative hypothesis sequences. In principle this enables us to analyze the size and power properties of tests used in testing nested hypothesis sequences to a smaller order of magnitude error than is possible with the asymptotic distributions. However, this method is still very costly to use in practice except for very simple cases with moderate sample sizes and we have implemented it only for distributions under the null hypothesis. Furthermore, this method does not enable us to analyze the properties of tests used in testing non-nested hypothesis sequences.

The main conclusion of the numerical evaluation studies involving asymptotic expansion approximations to the distributions of test statistics in actual situations is that such approximations are best used in conjunction with other approaches to small sample theory, e.g. exact invariance theory, parameter sequence asymptotics with a fixed sample size and Monte Carlo technqiues.

Asymptotic expansion approximations often seem to provide more useful qualitative than quantitative information. Sometimes they are useful in confirming and explaining previously observed phenomena for existing test statistics. For example, in Chapter 4 we find that they confirm the observed differences in behaviour, across the parameter space, of alternative Wald tests of a given non-linear hypothesis. However these results are qualitatively explainable using parameter sequence asymptotics with a fixed sample size; the only additional information which asymptotic expansion approximations provide is a rough indication of the size of the approximation error from using the asymptotic distribution.

Asymptotic expansion approximations can also suggest the presence of problems with existing testing procedures. For example, in Chapter 5 we find from moment expansions that the Chow predictive failure test contains an inappropriate adjustment for the sampling variation in the parameter estimates in dynamic models. However the numerically evaluated asymptotic expansion approximations to the distributions of the Chow statistics indicate that this is not a serious problem. In Chapter 6 we find that asymptotic expansion approximations confirm the presumption that the dynamic properties of the exogenous variables in an autoregressive distributed lag model strongly influence the properties of tests of hypotheses concerning the endogenous variable dynamics in such a model.

A number of problems still remain when considering the application of asymptotic expansion approximations in actual practice. Firstly, they are not easy to compute numerically. Secondly, they are not as yet applicable to testing procedures for non-nested hypothesis sequences. Thirdly, it is still unclear as to what features of a given situation determine the usefulness of asymptotic expansion approximations in that situation. Finally, in most situations encountered in practice the model used will be mis-specified. Very little research has been done on the derivation and use of asymptotic expansion approximations to the distributions of test statistics in mis-specified models. All of these problems suggest directions for future research to improve our usage of asymptotic expansion approximations in econometrics.

## APPENDIX ON COMPUTING

### A.1 The ESSACS Program

ESSACS is an acronym for Exact Sum of Squares Asymptotically Chi-square Statistics, and the ESSACS program was written to implement the techniques developed in Sections 3.2 and 3.4 of Chapter 3 for obtaining asymptotic expansion approximations to the joint cdf's of such statistics. The algorithm used draws extensively in places on the algorithm given in Tse (1983) and implemented in the program EDGE written by Y. K. Tse.

The EDGE algorithm computed the Edgeworth expansion, to an error of $O(T^{-3/2})$, of the cdf of a single asymptotically normal statistic. It has two modes of input: firstly, for a statistic based on data first and second order sample moments in a DESM; and, secondly, for a statistic based on quadratic and linear forms in a jointly normal vector y:

$$p = (y'D_1y, \dots, y'D_{N_1}y, y'D_{N_1+1}, \dots, y'D_N). \quad (1)$$

The first mode of input is a subcase of the second.

In the first mode of input the vector y is assumed to be generated by a DSEM which in reduced form is:

$$y_t = \sum_{s=1}^{r} A_s y_{t-s} + BZ_t + U_t, \; t=1, \dots, T, \quad (2)$$

where $y_t$ is an m-vector of endogenous variables so that $y = (y'_{1-r}, \dots, y'_t)'$, $Z_t$ is a q-vector of fixed exogenous variables, and the $U_t$ are serially independently distributed as $N[0,\Omega]$ vectors. The input data for the model consists of the values of m, q and T, the elements of the matrices $\{A_s\}$, B and $\Omega$, the exogenous variables data $\{Z_t\}$, and, when $r \geq 1$, starting means, $E(y_0), \ldots, E(y_{1-r})$, for the endogenous variables. The data first and second order sample moments take the forms:

$$M^k_{yab} = \sum_{t=1-b}^{(T-k)-b} y_{t+a}y_t'/(T-k), \quad 0 \leq a \leq k + b \leq k + r \quad ;$$

$$M^k_{zyab} = \sum_{t=1-b}^{(T-k)-b} Z_{t+a}y_t'/(T-k), \quad -k \leq b \leq a \leq k + b \leq k + r \; ;$$

$$M_{zz} = \sum_{t=1}^{T} Z_t Z_t'/T. \qquad (3)$$

Given (k,b,a,i,j) to indicate the (i,j)'th element of $M^k_{yab}$ the program will compute the elements of the appropriate quadratic form matrix; similarly given (k,b,a) to indicate the vector of linear forms $M^k_{zyab}$ the program will compute the elements of the appropriate linear form vectors. The program automatically computes $M_{zz}$, the mean of $y, \mu_y$, and the covariance matrix of $y, \Omega_y$. In the second mode of input the elements of $\mu_y$, $\Omega_y$ and $D_1, \ldots, D_N$ must be entered explicitly.

The EDGE program requires next the functional form of the statistic $\phi(p)$ in terms of the elements of p, and the derivatives of $\phi(.)$ up to the third order, evaluated at $\mu = E(p)$. The derivatives can either be expressed as functions of $\mu$ or computed numerically using central differences; see Tse (1981). The EDGE program then computes the second-order cumulants and the third-order and fourth-order directional cumulants of p where the direction is given by the first derivative of the statistic evaluated at $\mu$.

The general strategy of the ESSACS algorithm is to write the statistics of interest as the inner products of subvectors of some asymptotically standard normal vector. The algorithm proceeds by obtaining certain coefficients of the Edgeworth expansion of the pdf of this asymptotically N[0,I] decomposition vector and then using those coefficients to obtain the expansion, to an error of $O(T^{-3/2})$, of the cdf of the asymptotically chi-square statistics. The decomposition vector, $\lambda$, is written as T multiplied by a vector function of some vector p of linear and quadratic forms in a jointly normal vector y. Thus the input stage of the ESSACS program parallels directly that of the EDGE program. The provision or numerical computation of the derivatives and the computation of the second-order cumulants and the third-order directional cumulants of p also parallels directly that in the EDGE program. However the two programs then start to differ substantially.

The asymptotic expansion approximation of the pdf of the asymptotically chi-square statistics, given by Equation (47) of Chapter 3, can be obtained from the asymptotic expansion approximation of the cf of $\{\lambda^2\}$, the squares of the elements of $\lambda$, given by Equation (22) of Chapter 3. This can be obtained from the Edgeworth expansion approximation of the pdf of $\lambda$, given by Equation (14) of Chapter 3. However, only certain coefficients of this approximation are needed to compute the approximation to the cf of the $\{\lambda^2\}$. Also, it is unnecessary to store separately all the coefficients with distinct permutations of a given set of subscripts; it is only necessary to store the sum of such coefficients across all the distinct permutations of the given set of subscripts.

Given any intial set of coefficients, e.g. $\{a_{rs}\}$, we define:

$$\tilde{a}_{rs} = \begin{cases} a_{rr} \quad , & r = s \\ a_{rs} + a_{sr}, & r < s \end{cases} \tag{4}$$

so that in this notation the subscripts are given in ascending order and we have summed over all coefficients with distinct permutations of a given set of subscripts. We refer to the $\{\tilde{a}_{rs}\}$ as the summed-coefficients corresponding to the original coefficients $\{a_{rs}\}$.

The {g} coefficients of the approximation to the cf of the $\{\lambda^2\}$ can then be written in terms of the $\{\tilde{d}\}$ summed-coefficients of the approximation to the pdf of $\lambda$:

$$g_{rst} = \tilde{d}_{rrsstt}, \qquad g_{rs} = \tilde{d}_{rrss},$$
$$g_r = \tilde{d}_r, \quad g_0 = \tilde{d}_0. \tag{5}$$

where $r \leq s \leq t$. The $\{\tilde{d}\}$ summed-coefficients are given in terms of the summed-coefficients $\{\tilde{c}\}$ of the approximation to the cf of $\lambda$, given by Equation (7) of Chapter 3:

$$\tilde{d}_{rrsstt} = \tilde{c}_{rrsstt} \quad , \quad (r \leq s \leq t);$$

$$\tilde{d}_{rrss} = - \sum (\tilde{c}_{jjrrss} + \tilde{c}_{rrkkss} + \tilde{c}_{rrssll})$$
$$- 6(\tilde{c}_{rrrrss} + \tilde{c}_{rrssss}) + \tilde{c}_{rrss}, \quad (r < s) \ ;$$

$$\tilde{d}_{rrrr} = - 6\sum (\tilde{c}_{jjrrrr} + \tilde{c}_{rrrrkk}) - 15\tilde{c}_{rrrrrr} + \tilde{c}_{rrrr} \ ;$$

$$\tilde{d}_{rr} = \sum(\tilde{c}_{jjkkrr} + \tilde{c}_{jjrrll} + \tilde{c}_{rrllmm}) + 45\tilde{c}_{rrrrrr}$$
$$+ 3\sum (\tilde{c}_{jjjjrr} + 2\tilde{c}_{jjrrrr} + \tilde{c}_{rrllll} + 2\tilde{c}_{rrrrll})$$

$$- \sum(\bar{c}_{jjrr} + \bar{c}_{rrkk}) - 6\bar{c}_{rrrr} + \bar{c}_{rr};$$

$$\bar{d}_0 = - \sum(\bar{d}_{jj} + 3\bar{d}_{jjjj} + 15\bar{d}_{jjjjjj}) - \bar{d}_{jjkkll}$$

$$- \sum(\bar{d}_{jjkk} + 3\bar{d}_{jjjjkk} + 3\bar{d}_{jjkkkk}); \qquad (6)$$

where in each summation the subscripts are in strictly ascending order. The $\{\bar{c}\}$ summed-coefficients are given in terms of the summed-coefficients $\{\bar{b}\}$ of the approximation to the cgf of $\lambda$, given by Equation (6) of Chapter 3:

$$\bar{c}_{rrsstt} = \bar{b}^2_{rst} + 2\bar{b}_{rrs}\bar{b}_{stt} + 2\bar{b}_{rrt}\bar{b}_{sst} + 2\bar{b}_{rss}\bar{b}_{rtt};$$

$$\bar{c}_{rrrrss} = \bar{b}^2_{rrs} + 2\bar{b}_{rrr}\bar{b}_{rss} \; ; \; \bar{c}_{rrssss} = \bar{b}^2_{rss} + 2\bar{b}_{rrs}\bar{b}_{sss};$$

$$\bar{c}_{rrrrrr} = \bar{b}^2_{rrr} \; ; \; \bar{c}_{rrss} = \bar{b}_{rrss} + \bar{b}_r\bar{b}_{rss} + \bar{b}_{rrs}\bar{b}_s \; ;$$

$$\bar{c}_{rrrr} = \bar{b}_{rrrr} + \bar{b}_r\bar{b}_{rrr} \; ; \; \bar{c}_{rr} = \bar{b}_{rr} + (1/2)\bar{b}^2_r \; ; \qquad (7)$$

where $r < s < t$. The $\{\bar{b}\}$ summed-coefficients are finally given in terms of the summed multivariate Edgeworth coefficients, given by Equation (5) of Chapter 3 :

$$\bar{b}_r = \bar{a}^1_r \; ; \; \bar{b}_{rr} = [\bar{a}^2_{rr} - (1/2)] + \bar{a}^3_{rr} + \bar{a}^4_{rr} + \bar{a}^5_{rr} \; ;$$

$$\bar{b}_{rst} = \bar{a}^6_{rst} + \bar{a}^7_{rst} \; ;$$

$$\bar{b}_{rrss} = \bar{a}^8_{rrss} + \bar{a}^9_{rrss} + \bar{a}^{10}_{rrss} + \bar{a}^{11}_{rrss} \; ; \qquad (8)$$

where $r \leq s \leq t$ .

If we write the decomposition vector $\lambda$ as $g(x)$ where $x = T^{\frac{1}{2}}(p-\mu)$ then we can compute the summed multivariate Edgeworth coefficients from the cumulants of $x$ and the derivatives of $g$ with respect to $x$ at $x = 0$. We denote the second-order cumulant of $x_j$ and $x_K$ by $\kappa^{j,K}$ etc., and the first derivative of $g^r$ with respect to $x_j$ at $x = 0$ by $g_j^r$ etc. It is convenient to define a number of intermediate expressions for use in computing the required summed multivariate Edgeworth coefficients:

$$\beta_K^r = g_j^r \kappa^{j,K} \quad ; \quad \gamma_{K\ell}^r = g_j^r \kappa^{j,K,\ell} \quad ;$$

$$\lambda_{K\ell}^r = g_{jK}^r \kappa^{j,\ell} \quad ; \quad \delta_\ell^{r,s} = \beta_K^r g_{K\ell}^s \quad ;$$

$$\eta_{\ell m}^{r,s} = \beta_K^r g_{K\ell m}^s \quad ; \quad \chi_\ell^{r,s} = \gamma_{K\ell}^r g_K^s \quad ; \qquad (9)$$

where $\gamma_{K\ell}^r$, the third-order directional cumulant, is symmetric in $(K,\ell)$, $\eta_{\ell m}^{r,s}$ is symmetric in $(\ell,m)$, and $\chi_\ell^{r,s}$ is symmetric in $(r,s)$. The required summed multivariate Edgeworth coefficients are given by:

$$\bar{a}_r^1 = (\tfrac{1}{2})(g_{jK}^r \kappa^{j,K}); \quad \bar{a}_{rr}^2 = (\tfrac{1}{2})(\beta_K^r g_K^r);$$

$$\bar{a}_{rr}^3 = (\tfrac{1}{2})(\gamma_{K\ell}^r g_{K\ell}^r); \quad \bar{a}_{rr}^4 = (\tfrac{1}{4})(\lambda_{K\ell}^r \lambda_{\ell K}); \quad \bar{a}_{rr}^5 = (\tfrac{1}{2})(\eta_m^{r,s} \kappa^{\ell,m});$$

$$\bar{a}_{rst}^6 = (\chi_\ell^{r,s} g_\ell^t); \quad \bar{a}_{rrs}^6 = (\tfrac{1}{2})(\chi_\ell^{r,r} g_\ell^s);$$

$$\bar{a}_{rss}^6 = (\tfrac{1}{2})(\chi_\ell^{s,s} g_\ell^r); \quad \bar{a}_{rrr}^6 = (\tfrac{1}{6})(\chi_\ell^{r,r} g_\ell^r);$$

$$\bar{a}_{rst}^7 = (\delta_m^{r,t} \beta_m^s) + (\delta_m^{s,r} \beta_m^t) + (\delta_m^{t,s} \beta_m^r);$$

$$\bar{a}_{rrs}^7 = (\tfrac{1}{2})[(\delta_m^{r,s} \beta_m^r) + 2(\delta_m^{r,r} \beta_m^s)];$$

$$\tilde{a}^{7}_{rss} = (\tfrac{1}{2})[2(\delta^{s,s}_{m}\beta^{r}_{m}) + (\delta^{s,r}_{m}\beta^{s}_{m})];$$

$$\tilde{a}^{7}_{rrr} = (\tfrac{1}{2})(\delta^{r,r}_{m}\beta^{r}_{m});$$

$$\tilde{a}^{8}_{rrss} = (\tfrac{1}{4})(g^{r}_{j}g^{r}_{K}g^{s}_{\ell}g^{s}_{m}\kappa^{j,K,\ell,m});$$

$$\tilde{a}^{8}_{rrrr} = (\tfrac{1}{24})(g^{r}_{j}g^{r}_{K}g^{r}_{\ell}g^{r}_{m}\kappa^{j,K,\ell,m});$$

$$\tilde{a}^{9}_{rrss} = (\tfrac{1}{2})(\chi^{r,r}_{m}\delta^{s,s}_{m}) + (\tfrac{1}{2})(\chi^{s,s}_{m}\delta^{r,r}_{m}) + (\chi^{r,s}_{m})(\delta^{r,s}_{m} + \delta^{s,r}_{m});$$

$$\tilde{a}^{9}_{rrrr} = (\tfrac{1}{2})(\chi^{r,r}_{m}\delta^{r,r}_{m});$$

$$\tilde{a}^{10}_{rrss} = (\tfrac{1}{2})[(\beta^{r}_{m}\beta^{r}_{p}\eta^{s,s}_{mp}) + (\beta^{s}_{m}\beta^{s}_{p}\eta^{r,r}_{mp})];$$

$$\tilde{a}^{10}_{rrrr} = (\tfrac{1}{6})(\beta^{r}_{m}\beta^{r}_{p}\eta^{r,r}_{mp});$$

$$\tilde{a}^{11}_{rrss} = (\tfrac{1}{2})[2(\delta^{r,r}_{m}\delta^{s,s}_{q}) + (\delta^{r,s}_{m} + \delta^{s,r}_{m})(\delta^{r,s}_{q} + \delta^{s,r}_{q})]\kappa^{m,q};$$

$$\tilde{a}^{11}_{rrrr} = (\tfrac{1}{2})(\delta^{r,r}_{m}\delta^{r,r}_{q}\kappa^{m,q}); \qquad (10)$$

where $r < s < t$.

The method for computing $\kappa^{j,K}$, $\gamma^{r}_{K\ell}$ etc. follows to a large extent that adopted in the EDGE program. From Tse (1983), the joint cgf of p is given by:

$$\Psi(z) = \alpha_0 + (\tfrac{1}{2})\log_e\{\det(\Omega_y^{-1} - 2\theta)\}$$
$$+ (\tfrac{1}{2})(\mu_y^* + \psi)'(\Omega_y^{-1} - 2\theta)(\mu_y^* + \psi),$$

$$\text{where: } \theta = \sum_{j=1}^{N_1} iz_jD_j \quad ; \quad \psi = \sum_{j=N_1+1}^{N} iz_jD_j \quad ;$$

$$y \sim N[\mu_y, \Omega_y] \quad ; \quad \mu_y^* = \Omega_y^{-1}\mu_y \quad ;$$

$$\alpha_0 = (\tfrac{1}{2})[\log_e\{\det(\Omega_y^{-1})\} - \mu_y'\Omega_y^{-1}\mu_y]. \qquad (11)$$

If we now define:

$$\Sigma^r = \sum_{j=1}^{N_1} g^r D_j \quad ; \quad F^r = \Omega_y \Sigma^r \quad ; \quad q^r = \sum_{j=N_1+1}^{N} g_j^r D_j ;$$

$$F_j = \Omega_y D_j \quad , \quad (j=1,\dots,N_1) \quad ; \quad m^r = \Omega_y q^r \quad ;$$

$$H_{ij} = D_i\Omega_y D_j + D_j\Omega_y D_j \quad , \quad (i,j=1,\dots,N_1) \quad ;$$

$$G^r = \Omega_y \Sigma^r \Omega_y \quad ; \quad \omega^r = m^r + 2F^r\mu_y \quad ;$$

then we can compute $\kappa^{j,K}$, $\gamma_{K\ell}^r$ and $a_{rrss}^8$:

$$\kappa^{j,K} \begin{cases} = 2T[\mathrm{tr}(F_jF_k) + \mu_y'H_{jk}\mu_y] , & (j,k=1,\dots,N_1) \quad ; \\ = 2T\mu_y'D_j\Omega_y D_k , & (j=1, \dots , N_1 \ \& \ k=N_1+1, \dots, N) \quad ; \\ = TD_j'\Omega_y D_k , & (j,k=N_1+1, \dots , N) \quad ; \end{cases}$$

$$\gamma_{K\ell}^r \begin{cases} = T^{3/2}4[\ \mathrm{tr}\ (F^r(F_kF_\ell + F_\ell F_k)\ ) + \mu_y'H_{k\ell}m^r \\ \quad + 2\mu_y'\ \{\ (F^r)\ H_{k\ell} + D_kG^rD_\ell\ \}\mu_y\ ]\ , \ (k,\ell=1, \dots , N_1\ ) \quad ; \\ = T^{3/2}\ 2[\ (m^r)'D_k\Omega_y D_\ell + 2\mu_y'\ \{(F^r)'D_k + D_kF^r\ \}\Omega_y D_\ell] \\ \qquad (k=1, \dots, N_1 \ \& \ \ell = N_1 + 1, \dots , N\ ) \\ = T^{3/2}2D_k'G^rD_\ell\ , \ (k,\ell=N_1+1, \dots, N\ ) \quad ; \end{cases}$$

$$a^*_{rrss} = T^2 8[2(\omega^r)'\Sigma^r\Sigma^r F^s(\omega^s) + (\omega^r)'\Sigma^s F^s(\omega^r) + (\omega^s)'\Sigma^r F^r(\omega^s)$$
$$+ 2(\omega^r)'\Sigma^s F^r(\omega^s) + 4tr\{F^r F^r F^s F^s\} + 2tr\{F^r F^s F^r F^s\}];$$

(where $r \leq s$). Then $\tilde{a}^*_{rrss} = 6a^*_{rrss}$ $(r < s)$ and $\tilde{a}^*_{rrrr} = a^*_{rrrr}$ since $a^*_{rstu}$ is symmetric in $(r,s,t,u)$. The calculation of $\kappa^{j,k}$ and $\gamma^r_{k,\ell}$ is exactly the same as in Tse (1983) whereas that of $a^*_{rrss}$ is slightly modified.

## A.2 The Evaluation Study in Chapter 4

This study is concerned with the behaviour of Wald tests for different algebraic formulations of a non-linear hypothesis in the CLR with an intercept term and two other regressors:

$$y_t = \beta_o + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t, \; t = 1, \ldots, T, \qquad (13)$$

where the $\{x_{it}\}$ are known constants and the $\{\varepsilon_t\}$ are distributed as $IN(0,\sigma^2)$. This model can be written in the form of Equation (2) above with $m = 1$, $r = 0$, $q = 3$, $Z_t' = (1, x_{1t}, x_{2t})$, $B = \beta_o, \beta_1, \beta_2)$ and $\Omega = \sigma^2$.

Four Wald test statistics are considered: $W_M, W_{R1}, W_{R2}$ and $W_{R3}$. Each can be written as $W = T\mu^2$ where:

$$\mu_M = (\hat\beta_1\hat\beta_2 - 1)/\{\hat\sigma^2[\hat\beta_2^2 C_{11} + 2\hat\beta_1\hat\beta_2 C_{12} + \hat\beta_1^2 C_{22}]\}^{\frac{1}{2}};$$

$$\mu_{R1} = (\hat\beta_1\hat\beta_2 - 1)/\{\hat\sigma^2[\hat\beta_2^2 C_{11} + 2C_{12} + (C_{22}/\hat\beta_2^2)]\}^{\frac{1}{2}};$$

$$\mu_{R2} = (\hat\beta_1\hat\beta_2 - 1)/\{\hat\sigma^2[(C_{11}/\hat\beta_1^2) + 2C_{12} + \hat\beta_1^2 C_{22}]\}^{\frac{1}{2}};$$

$$\mu_{R3} = \hat\beta_1\hat\beta_2\eta_M;$$

from Equation (15) of Chapter 4, with:

$$\hat\beta = [\hat\beta_o, \hat\beta_1, \hat\beta_2]' = (\sum_{t=1}^{T} Z_t Z_t')^{-1}(\sum_{t=1}^{T} Z_t y_t) \quad ;$$

$$\hat\sigma^2 = (T-3)^{-1}[(\sum_{t=1}^{T} y_t^2) - \hat\beta'(\sum_{t=1}^{T} Z_t y_t)] \quad ;$$

and $\begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix}$ is the lower right-hand corner $(2\times2)$ block of $(\sum_{t=1}^{T} Z_t Z_t'/T)^{-1}$.

Instead of computing the asymptotic expansion approximations to the cdf's of the {W} directly it is more convenient to compute them for the cdf's of $\tilde{W}_M$, $\tilde{W}_{R1}$, $\tilde{W}_{R2}$ and $\tilde{W}_{R3}$ which are obtained by substituting $\tilde{\sigma}^2$ for $\hat{\sigma}^2$ in the {W}, where:

$$\tilde{\sigma}^2 = (\sum_{t=1}^{T} y_t^2/T) - \hat{\beta}'(\sum_{t=1}^{T} Z_t y_t/T),$$

is the ML estimate of $\sigma^2$. Thus we have:

$$\tilde{W}_M = (T-3)T^{-1} W_M = T\tilde{\mu}_M^2, \text{ etc.}$$

where the $\{\tilde{\mu}\}$ are obtained by substituting $\tilde{\sigma}^2$ for $\hat{\sigma}^2$ in the $\{\mu\}$. The $\{\tilde{\mu}\}$ can then be expressed as functions of:

$$M^o_{y00}(1,1) = \sum_{t=1}^{T} y_t^2/T, \; M^o_{Zy00} = \sum_{t=1}^{T} Z_t y_t/T, \text{ and}$$

$$M_{ZZ} = \sum_{t=1}^{T} Z_t Z_t'/T. \qquad (14)$$

so that given the values of $\beta_0$, $\beta_1$, $\beta_2$ and $\sigma^2$, the DSEM mode of input of the ESSACS program can be used to compute the approximations to the cdf's of the $\{\tilde{W}\}$. Finally, the approximations to the cdf's of the {W} can be computed from those of the $\{\tilde{W}\}$ by transforming the coefficients of the latter approximations following Appendix 5.A.3 of Chapter 5.

## A.3 The Evaluation Study in Chapter 5

This study is concerned with predictive failure tests in the AR(1) model:

$$y_i = \beta y_{i-1} + \varepsilon_i, \ |\beta| < 1, \ i = 1, 2, \ldots, n, n+1, n+2, \quad (15)$$

where the $\{\varepsilon_i\}$ are distributed as $IN(0,\sigma^2)$ and $y_0$ is distributed $N[0, \sigma^2/(1-\beta^2)]$. This model can be written in the form of Equation (2) above with $m = r = 1$, $A_1 = \beta$, $\Omega = \sigma^2$, $E(y_0) = 0$ and $T = n+2$. To construct the tests we also define:

$$Z_i = \begin{cases} (0,0,0,0,0)', & i = 1, \ldots, n-1, \\ (0,0,(n+2)^{\frac{1}{2}}, 0,(n+2)n^{-\frac{1}{2}})', & i = n, \\ (0,(n+2)^{\frac{1}{2}}, 0,(n+2)n^{-\frac{1}{2}},0), & i = n+1, \\ ((n+2)^{\frac{1}{2}}, 0, 0, 0, 0), & i = n+2, \end{cases} \quad (16)$$

so that $q = 5$, and $B = (0,0,0,0,0)$.

Six predictive failure tests are considered: $F^h_{1|n}$, $F^h_{2|n}$ and $F^{h+}_{2|n} = F^h_{2|n} - F^h_{1|n}$; and $F^c_{1|n}$, $F^c_{2|n}$ and $F^{c+}_{2|n} = F^c_{2|n} - F^c_{1|n}$. These are given by Equations (7), (8) and (9) of Chapter 5 and are expressed as functions of $y_n$, $y_{n+1}$, $y_{n+2}$, $\hat\beta_n$, $\hat\sigma^2_n$ and $\sum_{i=1}^{n} y^2_{i-1}$ where:

$$\hat\beta_n = (\sum_{i=1}^{n} y^2_{i-1})/(\sum_{i=1}^{n} y_i y_{i-1}), \text{ and}$$

$$\hat\sigma^2_n = (n-1)^{-1}[(\sum_{i=1}^{n} y_i^2) - \hat\beta_n(\sum_{i=1}^{n} y_i y_{i-1})].$$

are the OLS estimates of $\beta$ and $\sigma^2$. As in the evaluation study in Chapter 4 it is more convenient to obtain asymptotic expansion approximations to the cdf's of the modified statistics $\{\bar{F}\}$ where $\hat{\sigma}^2_n$ is replaced by the conditional ML estimate $\tilde{\sigma}^2_n$ than to obtain such approximations to the cdf's of the original statistics $\{F\}$. Here $\tilde{\sigma}^2_n$ is defined by:

$$\tilde{\sigma}^2_n = \left(\sum_{i=1}^{n} y_i^2/n\right) - \tilde{\beta}_n\left(\sum_{i=1}^{n} y_i y_{i-1}/n\right).$$

Thus $\{\bar{F}\}$ are given by $\bar{F}^h_{1|n} = (n-1)n^{-1}F^h_{1|n}$ etc., and the approximations to the cdf's of the $\{F\}$ can be obtained from those of the $\{\bar{F}\}$ following Appendix 5.A.3 of Chapter 5. The modified statistics can then be expressed by $\bar{F}^h_{1|n} = T(\bar{\mu}^h_1)^2$ and $\bar{F}^h_{2|n} = T(\bar{\mu}^{h+}_2)^2$ etc. where the $\{\bar{\mu}\}$ are functions of:

$$M^2_{y00}(1,1) = \sum_{i=1}^{n} y_i^2/n \qquad = p_1;$$

$$M^2_{y11}(1,1) = \sum_{i=1}^{n} y_i y_{i-1}/n \qquad = p_2;$$

$$M^2_{y01}(1,1) = \sum_{i=1}^{n} y_{i-1}^2/n \qquad = p_3;$$

$$
M^{0}_{Zy00} = \begin{cases} y_{n+2}(n+2)^{-\frac{1}{2}} & = p_4; \\ y_{n+1}(n+2)^{-\frac{1}{2}} & = p_5; \\ y_n(n+2)^{-\frac{1}{2}} & = p_6; \\ y_{n+1}n^{-\frac{1}{2}} & = p_7; \\ y_n n^{-\frac{1}{2}} & = p_8. \end{cases}
$$

Then $\bar{\mu}_1^{h}$ and $\bar{\mu}_2^{h+}$ are given by:

$$\bar{\mu}_1^{h} = (p_3p_5 - p_2p_6)\,(p_1p_3^2 - p_2^2p_3)^{-\frac{1}{2}} ;$$

$$\bar{\mu}_2^{h+} = (p_3p_4 - p_2p_5)(p_1p_3^2 - p_2^2p_3)^{-\frac{1}{2}}. \tag{17}$$

If we now define:

$$v_1 = 1 + (p_8^2/p_3) ; \quad v_2 = p_7p_8/p_3 ;$$

$$v_3 = 1 + (p_7^2/p_3) ; \quad s_1 = v_1^{-\frac{1}{2}} ;$$

$$s_2 = v_2(v_1^2v_3 - v_1v_2^2)^{-\frac{1}{2}} ; \quad s_3 = v_1^{\frac{1}{2}}(v_1v_3 - v_2^2)^{-\frac{1}{2}};$$

then $\bar{\mu}_1^{c}$ and $\bar{\mu}_2^{c+}$ are given by:

$$\bar{\mu}_1^{c} = s_1\bar{\mu}_1^{h}; \quad \bar{\mu}_2^{c+} = s_2\bar{\mu}_1^{h} + s_3\bar{\mu}_2^{h+}. \tag{18}$$

## A.4 The Evaluation Study in Chapter 6

This study is concerned with sequential testing in the AD(1,1) model:

$$y_t = \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T, \qquad (19)$$

where the $\{x_t\}$ are non-stochastic and the $\{\varepsilon_t\}$ are distributed $IN(0,\sigma^2)$. This model can be written in the form of Equation (2) above with $m = r = 1$, $q = 2$, $A = \alpha$, $Z_t' = (x_t, x_{t-1})$, $B = (\beta_0, \beta_1)$ and $\Omega = \sigma^2$. The value of $E(y_0)$ is set to zero since in this study we set $\alpha = \beta_0 = \beta_1 = 0$; see Section 6.3 of Chapter 6. If we write the model as

$$y_t = \gamma' W_t + \varepsilon_t, \quad t = 1, \ldots, T,$$

where $\gamma' = (\alpha, \beta_0, \beta_1)$ and $w_t' = (y_{t-1}, x_t, x_{t-1})$ then the test statistics considered can be expressed as:

$$W_1 = T\hat{\phi}_1^2[\hat{\sigma}^2\hat{f}'\hat{C}\hat{f}]^{-1};$$

$$W_2 = T[\hat{\phi}_1, \hat{\phi}_2]'[\hat{\sigma}^2\hat{F}_1\hat{C}\hat{F}_1']^{-1}[\hat{\phi}_1, \hat{\phi}_2]; \qquad (20)$$

where: $\hat{\phi}_1 = (\hat{\alpha}\hat{\beta}_0 + \hat{\beta}_1)$; $\hat{\phi}_2 = \hat{\alpha}$;

$$\hat{f}' = (\hat{\beta}_0, \hat{\alpha}, 1) \quad ; \quad \hat{F}_1 = \begin{bmatrix} \hat{\beta}_0 & \hat{\alpha} & 1 \\ 1 & 0 & 0 \end{bmatrix};$$

$$\hat{C} = (\sum_{t=1}^{T} w_t w_t'/T) \quad ;$$

$$\hat{\gamma} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\sum_{t=1}^{T} w_t w_t')^{-1} (\sum_{t=1}^{T} w_t y_t) \quad ;$$

$$\hat{\sigma}^2 = (T-3)^{-1}[(\sum_{t=1}^{T} y_t^2) - \hat{\gamma}'(\sum_{t=1}^{T} w_t y_t)].$$

As in the evaluation study in Chapter 4 it is more convenient to obtain asymptotic expansion approximations to the cdf's of the modified statistics $\{\tilde{W}\}$ , where $\hat{\sigma}^2$ is replaced by the conditional ML estimate $\tilde{\sigma}^2$, than to obtain such approximations for the original statistics $\{W\}$ . Here $\tilde{\sigma}^2$ is defined by:

$$\tilde{\sigma}^2 = (\sum_{t=1}^{T} y_t^2/T) - \hat{\gamma}'(\sum_{t=1}^{T} w_t y_t/T).$$

Thus $\tilde{W}_1 = (T-3)T^{-1}W_1$ etc. and the approximations to the cdf's of the $\{W\}$ can be obtained from those of $\{\tilde{W}\}$ following Appendix 5.A.3 of Chapter 5. If we now define:

$$M^0_{y00}(1,1) = \sum_{t=1}^{T} y_t^2/T = p_1 \quad ;$$

$$M^0_{y011}(1,1) = \sum_{t=1}^{T} y_t y_{t-1}/T = p_2 \quad ;$$

$$M^0_{y01}(1,1) = \sum_{t=1}^{T} y_{t-1}{}^2/T = p_3 \quad ;$$

$$M^0_{Zy00} = \begin{cases} \sum_{t=1}^{T} y_t x_t / T = p_4 \; ; \\ \sum_{t=1}^{T} y_t x_{t-1} / T = p_5 \; ; \end{cases}$$

$$M^0_{Zy11} = \begin{cases} \sum_{t=1}^{T} y_{t-1} x_t / T = p_6 \; ; \\ \sum_{t=1}^{T} y_{t-1} x_{t-1} / T = p_7 \; ; \end{cases}$$

$$M_{ZZ} = \sum_{t=1}^{T} Z_t Z_t' / T \; ;$$

then $\hat{\alpha}$, $\hat{\beta}_0$, $\hat{\beta}_1$, $\bar{\sigma}^2$ and $\hat{C}$ can be expressed as functions of $p = (p_1, \ldots, p_7)'$ and $M_{ZZ}$.

The $\{\bar{W}\}$ can be decomposed as $\bar{W}_1 = T\bar{\mu}_1^2$ and $\bar{W}_2^+ = \bar{W}_2 - \bar{W}_1 = T(\bar{\mu}_2^+)^2 = \tilde{W}_{(2)}$ where $\bar{\mu}_1$ and $\bar{\mu}_2^+$ are given by:

$$\bar{\mu}_1 = s_1\hat{\phi}_1 \; ; \; \bar{\mu}_2^+ = s_2\hat{\phi}_1 + s_3\hat{\phi}_2 \qquad (22)$$

where: $s_1 = v_1^{-\frac{1}{2}}$ ; $s_2 = v_2(v_1^2 v_3 - v_1 v_2^2)^{-\frac{1}{2}}$ ;

$s_3 = v_1^{\frac{1}{2}}(v_1 v_3 - v_2^2)^{-\frac{1}{2}}$ ; and

$$V = \begin{bmatrix} v_1 & v_2 \\ v_2 & v_3 \end{bmatrix} = \bar{\sigma}^{-2} FCF' .$$

Thus $\bar{\mu}_1$ and $\bar{\mu}_2^+$ can be expressed as functions of the elements of p and $M_{ZZ}$.

# BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. (1965) Handbook of Mathematical Functions. New York: Dover.

Akahira, M. and Takeuchi, K. (1981) Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency. New York: Springer-Verlag.

Aldrich, J. (1978) An alternative derivation of Durbin's h statistic. Econometrica, 46, 1493-1494.

Amari, S.-I. (1982) Geometrical theory of asymptotic ancillarity and conditional inference. Biometrika, 69, 1-17.

Amemiya, T. (1983) Advanced Econometrics. Cambridge: Harvard University Press.

Anderson, G.J. and Mizon, G.E. (1983) Parameter constancy tests: old and new. Discussion Paper 8325, Department of Economics, University of Southampton.

Anderson, R.L. (1942) Distribution of the serial correlation coefficient. Annals of Mathematical Statistics, 13, 1-13.

Anderson, T.W. (1958) An Introduction to Multivariate Statistical Analysis. New York: Wiley.

------- (1974) An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system. Journal of the American Statistical Association, 69, 565-573.

------- (1977) Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences. Econometrica, 45, 509-518.

Anderson, T.W., Kunitomo, N. and Sawa, T. (1982) Evaluation of the distribution function of the limited information maximum likelihood estimator. Econometrica, 50, 1009-1027.

Anderson, T.W. and Rubin, H. (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics, 20 , 570-582.

Anderson, T.W. and Sawa, T. (1973) Distributions of estimators of coefficients of a single equation in a simultaneous system and their asymptotic expansions. Econometrica, 41, 279-312.

------- (1979) Evaluation of the distribution function of the two stage least squares estimate. Econometrica, 47, 163-182.

Barndorff-Nielsen, O. and Cox, D.R. (1979) Edgeworth and saddlepoint approximations with statistical applications, (with Discussion). Journal of the Royal Statistical Society, Ser. B, 41, 279-312.

Basmann, R.L. (1965) On the application of the identifiability test statistic in predictive testing of explanatory economic models. Econometric Annual of the Indian Economic Journal, 13,387-423.

------- (1974) Exact finite sample distributions for some econometric estimators and test statistics: a survey and appraisal. In Frontiers of Quantitative Economics, Vol. II (M.D. Intriligator and D.A. Kendrick, eds.). Amsterdam: North Holland.

Basmann, R.L. and Richardson, D.H. (1973) The exact finite sample distribution of a non-consistent structural variance estimator. Econometrica, 41, 41-58.

Bergstrom, A.R. (1962) The exact sampling distribution of least squares and maximum likelihood estimators of the marginal propensity to consume. Econometrica, 30, 480-490.

Berry, A.C. (1941) The accuracy of the Gaussian approximations to the sum of independent variates. Transactions of the American Mathematical Society, 48, 122-136.

Berndt, E.R., Hall, B., Hall, R.E. and Hausman, J.A. (1974) Estimation and inference in non-linear structural models. Annals of Economic and Social Measurement, 3, 653-665.

Berndt, E.R. and Savin, N.E. (1977) Conflict among criteria for testing hypotheses in the multivariate linear regression model. Econometrica, 45, 1263-1278.

Bhattacharya, R.N. and Rao, R.R. (1976) Normal Approximations and Asymptotic Expansions. New York: Wiley.

Breusch, T.S. (1978) Testing for autocorrelation in dynamic linear models. Australian Economic Papers, 17, 334-355.

------- (1979) Conflict among criteria for testing hypotheses: extensions and comment. Econometrica, 47, 203-208.

Breusch, T.S. and Schmidt, P. (1985) Alternative forms of the Wald test: how long is a piece of string? Mimeo, Department of Economics, University of Southampton.

Cavanagh, C. (1981) Hypothesis testing in non-linear models. Working Paper, Institute of Economic and Business Research, University of California-Berkeley.

Challen, D.W. and Hagar, A.J. (1983) Macroeconometric Systems: Construction, Validation and Applications. London: Macmillan.

Chambers, J.M. (1967) On methods of asymptotic approximation for multivariate distributions. Biometrika, 54, 367-383.

Chow, G.C. (1960) Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28, 591-605.

Cooper, R.L. (1972) The predictive performance of quarterly econometric models of the United States. In Econometric Models of Cyclical Behaviour (B.G. Hickman, ed.). New York: Columbia University Press.

Cornish, E.A. and Fisher, R.A. (1937) Moments and cumulants in the specification of distributions. Review of the International Statistical Institute, 5, 307-320.

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

Daniels, H.E. (1954) Saddlepoint approximations in statistics. Annals of Mathematical Statistics, 25, 631-650.

------- (1956) The approximate distribution of the serial correlation coefficient. Biometrika, 43, 169-185.

Davidson, R. and MacKinnon, J.G. (1981a) Model specification tests based on artificial linear regressions. Discussion Paper 426, Institute for Economic Research, University of Queen's.

------- (1981b) Small sample properties of alternative forms of the Lagrange Multiplier test. Discussion Paper 439, Institute for Economic Research, University of Queen's.

------- (1983) Small sample properties of alternative forms of the Lagrange Multiplier test. Economics Letters, 12, 269-275.

Davies, R.B. (1973) Numerical inversion of a characteristic function. Biometrika, 60, 415-417.

De Bruijn, N.G. (1981) Asymptotic Methods in Analysis. New York: Dover.

De Grooijer, J.G. (1980) Exact moments of the sample autocorrelations from series generated by general ARIMA processes of order (p,d,q), d = 0 or 1. Journal of Econometrics, 14, 365-379.

Durbin, J. (1970) Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. Econometrica, 38, 410-421.

------- (1979) Discussion of paper by O. Barndorff-Nielsen and D.R. Cox. Journal of the Royal Statistical Society, Ser. B, 41, 279-312.

------- (1980) Approximations for densities of sufficient estimators. Biometrika, 67, 311-333.

Edgeworth, F.Y. (1905) The law of error. Transactions of the Cambridge Philosophical Society, 20, 36-65 & 113-141.

Esseen, G.G. (1945) Fourier analysis of distribution functions. Acta Mathematica, 77, 1-25.

Evans, G.B.A. and Savin, N.E. (1982a) Conflict among the criteria revisited: the W, LR and LM tests. Econometrica, 50, 737-748.

------- (1982b) Testing procedures in a model with lagged dependent variables. In Advances in Econometrics (W. Hildenbrand, ed.). Cambridge: Cambridge University Press.

Feller, W. (1971) An Introduction to Probability, Vol. II. New York: Wiley.

Field, C.A. and Hampel, F.R. (1982) Small-sample asymptotic distributions of M-estimators of location. Biometrika, 69, 29-46.

Fisher, R.A. (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika, 10, 507-521.

------- (1925) Expansion of "Student's" integral in powers of $n^{-1}$. Metron, 5, 109-112.

------- (1929) Tests of significance in harmonic analysis. Proceedings of the Royal Society, Ser. A, 125, 54-59.

Fuller, W.A. (1977) Some properties of a modification of the limited information estimator. Econometrica, 45, 939-954.

Fuller, W.A. and Hasza, D.P. (1980) Predictors for the first-order autoregressive process. Journal of Econometrics, 13, 139-157.

------- (1981) Properties of predictors for autoregressive time series. Journal of the American Statistical Association, 76, 155-161.

Gallant, A.R. and Holly, A. (1980) Statistical inference in an implicit, non-linear simultaneous equation model in the context of maximum likelihood estimation. Econometrica, 49, 697-720.

Gil-Pelaez, J. (1951) Note on the inversion theorem. Biometrika, 38, 481-482.

Greenberg, E. and Webster, C.E. (1983) Advanced Econometrics: A Bridge to the Literature. New York: Wiley.

Gregory, A.W. and Veall, M.R. (1984) On formulating Wald tests of non-linear restrictions. Working Paper 8401, University of Western Ontario.

------- (1985) Formulating Wald tests of non-linear restrictions. Econometrica, 53, 1465-1468.

------- (1986) Wald tests of common factor restrictions. Economics Letters, 22, 203-208.

Hannan, E.J. (1970) Multiple Time Series. New York: Wiley.

Hendry, D.F. (1974) Stochastic specification in an aggregate demand model of the United Kingdom. Econometrica, 42, 559-578.

------- (1980) Predictive failure and econometric modelling in macroeconomics: the transactions demand for money. In Econometric Modelling (P. Ormerod, ed.). London: Heinemann Educational Books.

------- (1984) Monte Carlo experimentation in econometrics. In Handbook of Econometrics, Vol. II (Z. Griliches and M.D. Intriligator, eds.). Amsterdam: North Holland.

Hendry, D.F. and Mizon, G.E. (1978) Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England. Economic Journal, 88, 549-563.

Hendry, D.F. and Wallis, K.F. (eds.) (1984) Econometrics and Quantitative Economics. Oxford: Basil Blackwell.

Henshaw, R.C. (1966) Testing single equation least squares models for autocorrelated disturbances. Econometrica, 44, 646-660.

Hoffman, D.L. and Schmidt, P. (1981) Testing the restrictions implied by the rational expectations hypothesis. Journal of Econometrics, 15, 265-287.

Hogg, R.B. (1961) On the resolution of statistical hypotheses. Journal of the American Statistical Association, 56, 978-989.

Holly, A. and Phillips, P.C.B. (1979) A saddlepoint approximation to the distribution of the k-class estimator of a coefficient in a simultaneous system. Econometrica, 47, 1527-1548.

Imhof, J.P. (1961) Computing the distribution of quadratic forms in normal variates. Biometrika, 48, 419-426.

Johnson, N.L. and Kotz, S. (1969) Distributions in Statistics: Discrete Distributions. New York: Wiley.

------- (1970a) Distributions in Statistics: Continuous Univariate Distributions I. New York: Wiley.

------- (1970b) Distributions in Statistics: Continuous Univariate Distributions II. New York: Wiley.

------- (1972) Distributions in Statistics: Continuous Multivariate Distributions. New York: Wiley.

Johnston, J. (1984) Econometric Methods. Tokyo: McGraw-Hill.

Kadane, J.B. (1970) Testing over-identifying restrictions when the disturbances are small. Journal of the American Statistical Association, 65, 182-185.

------- (1971) Comparison of k-class estimators when the disturbances are small. Econometrica, 39, 723-737.

------- (1974) Testing a subset of the over-identifying restrictions. Econometrica, 42, 853-867.

Kendall, M.G. and Stuart, A. (1967) The Advanced Theory of Statistics, Vol. I, 3rd edition. London: Griffin.

Kiviet, J.F. (1985) Model selection test procedures in a single linear equation of a dynamic simultaneous system and their defects in small samples. Journal of Econometrics, 28, 327-362.

------- (1986) On the rigour of some misspecification tests for modelling dynamic relationships. Review of Economic Studies, 53, 241-261.

Kiviet, J.F. and Phillips, G.D.A. (1986) Testing strategies for model specification. Mimeo, University of Amsterdam.

Lafontaine, F. and White, K.J. (1986) Obtaining any Wald statistic you want. Economics Letters, 21, 35-40.

Laitinen, K. (1978) Why is demand homogeneity so often rejected? Economics Letters, 1, 187-191.

Lehmann, E.L. (1959) Testing Statistical Hypotheses. New York: Wiley.

Lukacs, E. (1960) Characteristic Functions. London: Griffin.

Lukacs, E. and Laha, R.G. (1964) Applications of Characteristic Functions. London: Griffin.

McCullagh, P. (1984) Tensor notation and the cumulants of polynomials. Biometrika, 71, 461-476.

McDonald, J.B. (1972) The exact finite sample distribution of the limited information maximum likelihood identifiability test statistic. Econometrica, 40, 1109-1119.

Magdalinos, M.A. (1983) Applications of the refined asymptotic theory in econometrics. Unpublished Ph.D. thesis, University of Southampton.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) Multivariate Analysis. London: Academic Press.

Mariano, R.S. (1973) Approximations to the distribution of the ordinary least-squares and two stage least-squares estimators in the case of two included endogenous variables. Econometrica, 41, 67-77.

------- (1982) Analytical small sample distribution theory in econometrics: the simultaneous equations case. International Economic Review, 23, 503-533.

Mariano, R.S. and Sawa, T. (1972) The exact finite sample distribution of the limited information maximum likelihood estimator in the case of two included endogenous variables. Journal of the American Statistical Association, 67, 159-163.

Mauleon-Torres, I. (1983) Approximations to the finite sample distribution of econometric chi-square type of criteria. Unpublished Ph.D. thesis, University of London.

Meisner, J.F. (1979) The sad fate of the asymptotic Slutsky symmetry test for large systems. Economics Letters, 2, 231-233.

Mizon, G.E. and Hendry, D.F. (1980) An empirical application and Monte Carlo analysis of tests of dynamic specification. Review of Economic Studies, 47, 21-45.

Nagar, A.L. (1959) The bias and moment matrix of the general k-class estimator of the parameters in simultaneous equations. Econometrica, 27, 573-595.

Nelson, C.R. (1972) The prediction performance of the FRB-MIT-PENN model of the US economy. American Economic Review, 62, 902-917.

Olver, F.W.R. (1974) Asymptotics and Special Functions. New York: Wiley.

Orcutt, G.H. and Winokur, H.S. (1969) First order autoregression: inference, estimation and prediction. Econometrica, 37, 1-14.

Petrov, V.V. (1968) Asymptotic behaviour of probabilities of large deviations. Theory of Probability and its Applications, 9, 312-320.

Pfanzagl, J. (1980) Asymptotic expansions in parametrical statistical theory. In Developments in Statistics, Vol. 3 (P.R. Krishnaiah, ed.). New York: Academic Press.

------- (1982) Contributions to a General Asymptotic Statistical Theory. With assistance of W. Wefelmeyer. New York: Springer-Verlag.

Phillips, P.C.B. (1977a) Approximations to some finite sample distributions associated with a first-order stochastic difference equation. Econometrica, 45, 463-485.

------- (1977b) A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators. Econometrica, 45, 1517-1534.

------- (1977c) A large deviation limit theorem for multivariate distributions. Journal of Multivariate Analysis, 7, 50-62.

------- (1978) Edgeworth and saddlepoint approximations in the first-order non-circular autoregression. Biometrika, 65, 91-98.

------- (1979) The sampling distribution of forecasts from a first-order autoregression. Journal of Econometrics, 9, 241-261.

------- (1980a) The exact finite sample density function of instumental variables estimators in an equation with n+1 endogenous variables. Econometrica, 48, 861-878.

------- (1980b) Finite sample theory and the distribution of alternative estimators of the marginal propensity to consume. Review of Economic Studies, 47, 183-224.

------- (1982) Best uniform and modified Padé approximants to probability densities in econometrics. In Advances in Econometrics (W. Hildenbrand, ed.). Cambridge: Cambridge University Press.

------- (1983a) Exact small sample theory in the simultaneous equations model. In Handbook of Econometrics, Vol. I (Z. Griliches and M.D. Intriligator, eds.). Amsterdam: North Holland.

------- (1983b) Marginal densities of instrumental variables estimators in the general single equation case. Advances in Econometrics, 2, 1-24.

------- (1984a) The exact distribution of LIML : I. International Economic Review, 25, 249-261.

------- (1984b) The exact distribution of the Stein-rule estimator. Journal of Econometrics, 25, 123-131.

------- (1985a) The exact distribution of the SUR estimator. Econometrica, 53, 745-756.

------- (1985b) The exact distribution of LIML : II. International Economic Review, 26, 1-36.

Quenouille, M.L. (1949) Approximate tests of correlation in time series. Journal of the Royal Statistical Society, Ser. B, 11, 68-83.

Rao, C.R. (1973) Linear Statistical Inference, 2nd edition.New York: Wiley.

Rhodes, G.F. (1981) Exact density function and approximate critical regions for likelihood ratio identifiability test statistics. Econometrica, 49, 1035-1056.

Richardson, D.H. (1968) The exact distribution of a structural coefficient estimate. Journal of the American Statistical Association, 63, 1214-1226.

Richardson, D.H. and Rohr, R.J. (1971) The distribution of a structural t-statistic for the case of two included endogenous variables. Journal of the American Statistical Association, 66, 375-382.

Rothenberg, T.J. (1977) Edgeworth expansions for multivariate test statistics. Working Paper IP-225, Centre for Research in Management Science, University of California-Berkeley.

------- (1984a) Hypothesis testing in linear models when the error covariance matrix is non-scalar. Econometrica, 52, 827-842.

------- (1984b) Approximating the distributions of econometric estimators and test statistics. In Handbook of Econometrics, Vol. II (Z. Griliches and M.D. Intriligator, eds.). Amsterdam: North Holland.

Rugaimukamu, D.M.B. (1982) Post sample parameter stability tests under linear aggregation of economic relationships. Unpublished Ph.D. thesis, University of Southampton.

Ryall, T.A. (1981) Extensions of the concept of local ancillarity. Biometrika, 68, 677-683.

Sargan, J.D. (1970) The finite sample distribution of FIML estimators. Paper presented to the World Congress of the Econometric Society, Cambridge.

------- (1974) The validity of Nagar's expansion for the moments of econometric estimators. Econometrica, 42, 169-176.

------- (1975) Gram-Charlier approximations applied to t-ratios of k-class estimators. Econometrica, 43, 327-346.

------- (1976) Econometric estimators and the Edgeworth approximation. Econometrica, 44, 421-448.

------- (1980a) Some tests of dynamic specification for a single equation. Econometrica, 48, 879-897.

------- (1980b) Some approximations to the distribution of econometric criteria which are asymptotically distributed as chi-square. Econometrica, 48, 1107-1138.

------- (1984) A study of alternative Wald tests for dynamic specification. Discussion Paper A.46, London School of Economics.

Sargan, J.D. and Mikhail, W.M. (1971) A genera approximation to the distribution of instrumental variable estimates. Econometrica, 39, 131-169.

Sargan, J.D. and Satchell, S.E. (1986) A theorem of validity for Edgeworth expansions. Econometrica, 54, 184-213.

Sargan, J.D. and Tse, Y.K. (1979) Edgeworth approximations for 2SLS estimates of a dynamic model. Discussion Paper A.25, London School of Economics.

Savin, N.E. (1976) Conflict among testing procedures in a linear regression model with autocorrelated disturbances. Econometrica, 44, 1303-1315.

Sawa, T. (1969) The exact finite sampling distribution of ordinary least squares and two stage least squares estimators. Journal of the American Statistical Association, 64, 923-926.

------- (1978) The exact moments of the least squares estimator for the autoregressive model. Journal of Econometrics, 8, 159-172.

Silvey, S.D. (1975) Statistical Inference. London: Chapman Hall.

Summers, R. (1965) A capital intensive approach to the small sample properties of various simultaneous equations estimators. Econometrica, 33, 1-41.

Tanaka, K. (1984) An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. Journal of the Royal Statistical Society, Ser. B, 46, 58-67.

Tanaka, K. and Maekawa, K. (1984) The sampling distributions of the predictor for an autoregressive model under misspecification. Journal of Econometrics, 25, 327-351.

Taylor, W.E. (1983) On the relevance of finite sample theory. Econometric Reviews, 2, 1-39.

Theil, H. (1971) Principles of Econometrics. New York: Wiley.

Tse, Y.K. (1981) Edgeworth approximations to the finite sample distribution of econometric estimators and test statistics. Unpublished Ph.D. thesis, University of London.

------- (1982) Edgeworth approximations in first-order stochastic difference equations with exogenous variables. Journal of Econometrics, 20, 175-195.

------- (1983) On calculating the Edgeworth approximate distibution of an econometric estimator or test statistic. Economic Letters, 12,37-41.

Ullah, A. and Zinde-Walsh, V. (1984) On the robustness of LM, LR and W tests in regression models. Econometrica, 52, 1055-1066.

Von Neumann, J. (1941) Distribution of the ratio of the mean square sucessive difference to the variance. Annals of Mathematical Statistics, 12, 367-395.

Wald, A. (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society, 54, 426-482.

Wallace, D.L. (1958) Asymptotic approximations to distributions. Annals of Mathematical Statistics, 29, 635-654.

Watson, G.S. (1956) On the distribution of the circular serial correlation coefficient. Biometrika, 43, 161-168.

Weir, J.B. de V. (1960) Standardized t. Nature, 185, page 558.

White, H. (1984) Asymptotic Theory for Econometricians. New York: Academic Press.

White, J.S. (1961) Asymptotic expansions for the mean and variance of the serial correlation coefficient. Biometrika, 48, 85-94.

Wilks, S.S. (1962) Mathematical Statistics. New York: Wiley.

TITLE ASYMPTOTIC EXPANSION APPROXIMATIONS AND THE DISTRIBUTIONS OF VARIOUS TEST STATISTICS IN DYNAMIC ECONOMETRIC MODELS

AUTHOR Gordon Christopher Russell Kemp

INSTITUTION and DATE THE UNIVERSITY OF WARWICK, 1987

D87433