

The Evaluation of Economic Forecasts

by

David I. Harvey, M.A.

Thesis submitted to the University of Nottingham for
the degree of Doctor of Philosophy, October 1997

Acknowledgements

‘To God who is able to do immeasurably more than all we can ask or imagine, according to His power, to Him be glory.’

Ephesians 3:20-21 (abridged)

I would like to express my thanks to the University of Nottingham for the research studentship funding which enabled this thesis to be written.

I am especially grateful to my supervisors Prof. Paul Newbold and Dr. Steve Leybourne for their continual help and invaluable comments throughout my time of study.

For my Mother and Father.

Abstract

The evaluation of economic forecasts is a substantial and important aspect of economic research, and a considerable part of such evaluation is performed by comparing competing forecasts. This thesis focuses on the development of statistical procedures in order that reliable comparison of contending forecasts can be made.

The study considers three issues in particular. The first two issues are closely related and concern testing the companion null hypotheses of equal forecast accuracy and forecast encompassing. The established equal accuracy and encompassing tests are found to display problematic behaviour in certain situations, and new modified tests are proposed to overcome these shortcomings. Analysis of the tests results in a recommendation for employing one of the newly proposed tests for each of the respective hypotheses. The recommended tests follow parallel formulations and have a number of attractive features, notably robustness to likely forecast error properties of contemporaneous correlation, autocorrelation, non-normality and autoregressive conditional heteroscedasticity, reliable behaviour in finite samples, and good power performance.

The third issue examines the ranking of rival forecasts according to a pre-determined evaluation criterion. A recently proposed summary criterion for multi-step-ahead forecasts, comprising a single measure for all model representations and all forecast horizons of interest, is analysed, and a more reliable alternative proposed. This summary criterion approach is compared to the more conventional method of ranking forecasts at a specific horizon for a particular model representation, and the related issue of forecast encompassing for linear combinations of forecasts is discussed.

This thesis therefore develops robust well-behaved tests for equal forecast accuracy and forecast encompassing, and advances techniques for ranking competing multi-step forecasts, providing improved, more reliable procedures for conducting economic forecast evaluation.

Contents

Chapter 1: Introduction	1
1.1 Preface	2
1.2 Tests for Equal Forecast Accuracy	4
1.3 Tests for Forecast Encompassing	9
1.4 Testing in the Presence of ARCH Errors	13
1.5 Ranking Competing Forecasts	15
Chapter 2: Tests for Equal Forecast Accuracy	19
2.1 Introduction	20
2.2 Motivation for Research	22
2.3 Diebold-Mariano Approach Tests	31
2.4 Morgan-Granger-Newbold Approach Tests	56
2.5 Conclusion	78
Chapter 3: Tests for Forecast Encompassing	82
3.1 Introduction	83
3.2 Regression Test	85
3.3 Modified Regression Tests	101
3.4 Rank Correlation Test	107
3.5 Diebold-Mariano Approach Tests	110
3.6 Power Comparisons	122
3.7 Conclusion	130
Chapter 4: Testing in the Presence of ARCH Errors	133
4.1 Introduction	134
4.2 ARCH Error Specification and Properties for Equal Accuracy Tests	135
4.3 Tests for Equal Forecast Accuracy	146
4.4 ARCH Error Specification and Properties for Encompassing Tests	175
4.5 Tests for Forecast Encompassing	180
4.6 Conclusion	196
Chapter 5: Ranking Competing Forecasts	197
5.1 Introduction	198
5.2 Literature Review	199
5.3 The GFESM Criterion	210
5.4 An Invariant Non-Reversing Criterion	243
5.5 Forecast Encompassing for Linear Combinations of Forecasts	254
5.6 Conclusion	259
Chapter 6: Conclusion	262
6.1 Summary	263
6.2 Directions for Future Research	267
References	270

Chapter 1

Introduction

1.1 Preface

The practice of forecasting comprises a fundamental and substantial part of economic study and analysis. Prediction of economic variables is of paramount importance in both the public sector and the private sector, forming an intrinsic part of decision-making processes. Given the great significance attached to such predictions, the issue of forecast performance assumes equal status, with the evaluation of economic forecasts becoming an important and valuable area of economic research.

Forecast evaluation incorporates all forms of forecast performance assessment, and Granger & Newbold (1973) provide some early comment on the general issues. These authors examine methods to determine how good a particular forecast is, and how it might be modified to achieve superior prediction performance. They argue strongly that an 'objective evaluation of forecast performance is of the greatest importance', and review a number of alternative techniques. Stekler (1991) likewise discusses general forecast evaluation questions, and provides a variety of statistical tests and criteria by which forecasts can be judged. Both papers consider measures of forecast adequacy, but the majority of their work concerns comparisons between competing forecasts. In the literature there is a plethora of such ordinal forecast comparisons, with applications being in two areas - firstly in the direct comparison of competing forecast producers or forecast-generating mechanisms, with the goal of deciding which forecast is 'best'; secondly in model building where predictive performance with respect to a hold-out sample is used as a diagnostic tool, and alternative model specifications can be compared.

The ubiquitous nature of forecast comparisons, combined with their central role in the evaluation process and consequent decision-making, generates the desire to have available formal statistical procedures by which these comparisons can be made. This thesis focuses on the development of such procedures, the aim being to establish reliable and robust statistical techniques for comparing competing forecasts. Three areas are considered - tests for equal forecast accuracy (chapters 2 and 4), tests for forecast encompassing (chapters 3 and 4) and criteria for ranking multi-step-ahead forecasts (chapter 5). The treatise is concluded in the sixth and final chapter.

1.2 Tests for Equal Forecast Accuracy

The accuracy of an economic forecast is crucial in the evaluation of competing forecasts. The question of whether one forecast is better than another in terms of accuracy needs to be addressed in a statistical framework, and much work has been done to this end. When comparing two rival forecasts, the accuracy issue can be tackled by developing tests of the null hypothesis of equal forecast accuracy of the competing forecasts against a one-sided or two-sided alternative. To do this, some measure of accuracy must be utilised and by far the most common approach is use of the mean squared forecast error, although other measures may be employed; the important point is to choose a metric which approximates the economic loss associated with the use of the forecasts, i.e. the decision at hand. Given a measure of accuracy, testing can proceed.

Tests of the null of equal forecast accuracy must be based on series of forecast errors for a given forecast horizon. These errors will most likely exhibit a number of statistical properties, and tests must obviously be robust to such properties in order to be reliable in application. Errors from economic forecasts might strongly be expected to manifest contemporaneous correlation (since forecasters have overlapping information sets and some outcomes will surprise all forecast producers) and autocorrelation (especially for multi-step-ahead forecasts on theoretical grounds), and may well be non-normally distributed. Forecast errors may possibly also be biased, although for the purpose of this analysis unbiasedness shall be assumed.

In practice, particularly with forecasts of macroeconomic variables, relatively few

observations are available, which leads to two considerations. Firstly, in general there may be insufficient data to test for the presence of the error properties described above, reiterating the need for robust tests. Secondly, tests must be well-behaved in moderately sized samples, as well as large ones, to be valuable in application.

Dhrymes *et al.* (1972) consider the evaluation of econometric models, and highlight the difficulties associated with testing, as do Howrey, Klein & McCarthy (1974) in a separate study. The earliest and simplest test of equality of prediction mean squared errors is the F-test, where the test statistic is comprised of the ratio of the two forecasts' mean squared errors. This is a variance ratio test and does not allow for any of the aforementioned properties in the forecast errors.

A more workable test is that following Morgan (1939-40) and Granger & Newbold (1986) which employs an orthogonalising transformation to achieve robustness to contemporaneous correlation. Ashley, Granger & Schmalensee (1980) use a version of this test in an applied context. The Morgan-Granger-Newbold test is uniformly most powerful unbiased when the errors are normal and are not autocorrelated, but still falls short of a reliable test if these error properties are present.

Other parametric tests have been proposed by authors such as Meese & Rogoff (1988), Diebold & Rudebusch (1991), who put forward tests which overcome the problem of autocorrelation, and Vuong (1989), whose work follows a classical hypothesis testing approach in the context of model selection on the basis of the Kullback Leibler Information Criterion. Stekler (1991) and Diebold & Mariano

(1995) also review and present a number of nonparametric approaches to testing the equal accuracy null.

Diebold & Mariano (1995) introduce a new parametric test in their paper which, by virtue of its robustness to all the above error properties, is shown to be superior to its predecessors in a simulation study. The test takes a very general specification, performing a test on the sample mean of a loss differential function which can be arbitrarily defined, e.g. as the difference between the two mean squared forecast errors. The Diebold-Mariano test is straightforward to implement and has very attractive robustness properties compared to its rivals. The caveat to its general recommendation as the best equal forecast accuracy test is its finite sample behaviour, with the test's empirical size significantly exceeding its nominal size in cases of moderate and small sample sizes.

Chapter 2 of this thesis is motivated by the above literature. In the quest to find a thoroughly robust and reliable test of the null of equal forecast accuracy, which is well behaved for all sample sizes, two issues are studied.

Firstly, the Diebold-Mariano asymptotic test is analysed, with particular reference to its undesirable property of being oversized in moderate-sized samples. The Diebold-Mariano test statistic divides the loss differential sample mean by an estimate of its standard error, and this estimate is found to be biased in finite samples. Following this, a corrected test is proposed to alleviate the original test's problems of missizing. The modified Diebold-Mariano test embodies two amendments - a finite sample correction to the variance estimator in the test statistic

to achieve approximate unbiasedness, and use of critical values from the Student's t distribution with $n - 1$ degrees of freedom (where n denotes the sample size) instead of the originally proposed standard normal critical values.

A comprehensive simulation study shows the modified Diebold-Mariano test to exhibit significant and substantial improvements to the original test. The oversizing problem is greatly reduced (although some oversizing persists in small samples for multi-step-ahead prediction) whilst all the advantages of the Diebold-Mariano approach are maintained.

The second issue regards the Morgan-Granger-Newbold test, and considers the test's behaviour when the forecast errors are not normally distributed. Simulation highlights the test's inadequacy under error non-normality with considerable oversizing present in both finite samples and in the limit. However, the known power advantages of the test under normality, also confirmed by simulation, motivate interest in discovering whether a correction exists to correct the size behaviour whilst retaining the power superiority when the error distribution departs from normality.

The source of the test's problem in a non-normal context is found to be inconsistent variance estimation in the test statistic. This in turn stems from the fact that the test is regression-based, and under forecast error non-normality the regression errors are conditionally heteroscedastic. Three corrected tests are then proposed - two parametric tests employing White (1980) type corrections to achieve heteroscedasticity-robust estimation of the regression parameter variance, and a

nonparametric approach using Spearman's rank correlation test.

A simulation study examines the size and power properties of all these tests, and finds the parametric corrected tests to display other problems - missizing and a reduction in power. The Spearman's rank correlation test features a correct test size in all samples, has lower power than the Morgan-Granger-Newbold test, but possesses significant power advantages over the modified Diebold-Mariano test when the errors are non-normal, yielding a valuable test if the consideration is purely for one-step-ahead prediction and heavy-tailed errors are suspected. Overall, however, due to its applicability to multi-step-ahead forecasts, its robustness to all the examined error properties, its general loss function specification, and its broadly reliable size and power performance, the modified Diebold-Mariano test is proposed as the recommended test for equal forecast accuracy.

1.3 Tests for Forecast Encompassing

The comparison of competing forecasts in terms of accuracy is desirable, and execution of the formal tests described above will frequently lead to the inference that one forecast is ‘significantly better’ than the other. However, as stressed by Granger & Newbold (1973), this should not result in placing complete confidence in the preferred forecast. A more stringent requirement is that the inferior forecast embodies no useful information with regard to prediction which is not already contained in the preferred forecast. If such a requirement holds, the superior forecast is said to encompass its competitor. This notion of forecast encompassing can also be defined, and subsequently tested for, by forming a combined forecast comprised of a weighted average of the individual forecasts.

The idea of combining forecasts began with Bates & Granger (1969) and Reid (1968, 1969). These authors found that a composite forecast, generated by forming a weighted average of the original forecasts, could yield a lower mean squared error than either of the competing forecasts individually.

The theory associated with combining forecasts, begun by Bates & Granger and Reid, was developed in a number of papers, notably by Dickinson (1973, 1975), Bunn (1975, 1977) and Öller (1978), and more recently by Granger & Ramanathan (1984) and Granger (1989). Much empirical work has been performed, for example the assessment of forecasting techniques by Newbold & Granger (1974), Granger & Newbold (1975), Makridakis & Hibon (1979) and Makridakis *et al.* (1982, 1983). Clemen (1989) provides a thorough review of the combination of forecasts

literature.

The link between the combination of forecasts and forecast encompassing is made clear by the Granger & Newbold (1973) definition (see also Nelson, 1972, and Cooper & Nelson, 1975) of conditional efficiency. They define a preferred forecast to be conditionally efficient with respect to (or encompass) a competitor if the optimal weight attached to the latter in a composite predictor is zero. This formal definition lends itself very naturally to a test for forecast encompassing, based on regression and a t -test on the optimal weight. This established test has been employed widely and advocated by Chong & Hendry (1986) and Clements & Hendry (1993), *inter alia*.

Additional work on the encompassing principle has been performed by Hendry & Richard (1982, 1983, 1989), Mizon (1984) and Mizon & Richard (1986), while Diebold (1989) and Wallis (1989) elaborate the relationship between combination and encompassing.

Chapter 3 studies the issue of testing for forecast encompassing. The forecast error properties discussed in the context of testing for equal forecast accuracy are equally of concern here, and the focus of the chapter is to provide a reliable, robust, well-behaved test of the null that the preferred forecast encompasses its competitor.

The regression test mentioned above bears a close resemblance to the Morgan-Granger-Newbold test for the equality of prediction mean squared errors, and consequently intuitive doubts are raised concerning its applicability to situations

where forecast error non-normality persists. Analysis of this test both theoretically and by simulation delivers the insight that the regression test for forecast encompassing is not robust to heavy-tailed error distributions, with oversizing characterising finite sample and limiting behaviour.

The problem can again be traced to inconsistent regression parameter variance estimation caused by conditional heteroscedasticity in the regression errors, and variants of the three modified tests employed to correct the Morgan-Granger-Newbold test can equally be applied in this context.

Furthermore, an additional test for forecast encompassing is proposed, using the Diebold-Mariano approach with an appropriately defined loss differential to achieve testing of the encompassing null. The aforementioned modifications to the Diebold-Mariano procedure are also utilised to improve this test's behaviour.

An extensive simulation study of the tests' empirical performances leads to similar conclusions to the equal accuracy question. For one-step-ahead evaluation, the rank correlation test behaves well and has best power under a situation of non-normal errors; also one of the parametric corrected tests displays acceptable size and relatively good power performance. However, the preferred test in general is the modified Diebold-Mariano approach test which loses little to the regression test in power whilst achieving good overall size behaviour for one-step- and multi-step-ahead prediction, irrespective of the forecast error distribution and properties.

A general recommendation is therefore proposed for the approach to the companion problems of testing for equal forecast accuracy and testing for forecast

encompassing. The modified Diebold-Mariano tests provide reliable and robust statistical techniques for these aspects of forecast evaluation.

1.4 Testing in the Presence of ARCH Errors

The tests for equal forecast accuracy and forecast encompassing discussed above are developed so as to exhibit robustness to forecast error properties which would be expected to arise in an applied context. In addition to the properties of contemporaneous correlation, autocorrelation and non-normality already considered, another property demands attention - autoregressive conditional heteroscedasticity (ARCH). ARCH implies predictable uncertainty through time and errors possessing this property would be expected whenever the volatility of the variable being forecast varies systematically through time. This is of great relevance in economics, especially with financial variables.

The notion of ARCH was introduced by Engle (1982), who developed this new class of statistical processes. In the same paper, Engle finds ARCH to be significant in UK inflation uncertainty, and finds the same to be true for the US in another study (Engle, 1983). Rich, Raymond & Butler (1992) confirm this finding of strong evidence of ARCH in inflation forecast errors. The basic ARCH model has been generalised to GARCH by Bollerslev (1986) and multivariate specifications have been summarised by Bollerslev (1990) and Engle & Kroner (1995). Bollerslev, Chou & Kroner (1992) provide a review of the wide application of GARCH models in the literature.

In chapter 4, the issues of testing for equal forecast accuracy and forecast encompassing are revisited, with the tests examined in a world where the forecast errors have ARCH as a characteristic. The example case of bivariate ARCH(1)

forecast errors is considered under three scenarios - independent errors, Engle-Kroner correlated errors (following Engle & Kroner, 1995) and Bollerslev correlated errors (following Bollerslev, 1990), and all the tests are re-examined.

The parallel results for equal accuracy and encompassing are that ARCH causes two problems - leptokurtosis in the errors, and autocorrelation in the loss differential series (or its equivalent in the regression-based tests). These effects lead to serious oversizing in finite samples and in the limit for the Morgan-Granger-Newbold test (equal accuracy) and the regression test (forecast encompassing). The tests which were previously robust to non-normality, i.e. the modified Diebold-Mariano approach tests and the corrected regression-based tests, overcome the leptokurtosis element of the ARCH effect, but still exhibit missizing due to the autocorrelation in the loss differential.

Two new tests are therefore proposed (one for testing equality of forecast accuracy, one for testing forecast encompassing), which overcome the majority of the ARCH-induced size distortions. This improvement is achieved by including additional covariance lags (the number of which being determined by a Newey & West (1994) type lag selection rule) when estimating the variance of the loss differential mean in the Diebold-Mariano test statistic. Incorporation of this information into the modified Diebold-Mariano approach tests recommended in chapters 2 and 3 is shown by Monte Carlo simulation to achieve robust reliable tests for equal forecast accuracy and forecast encompassing when ARCH is present in the forecast errors.

1.5 Ranking Competing Forecasts

In addition to the formal tests discussed above, it is also desirable to have available criteria by which competing forecasts can be ranked. These rankings will again be on the basis of some measure of accuracy, but are determined purely by the relative values of the chosen evaluation criterion for the respective forecasts, and do not involve executing statistical tests.

The literature is replete with forecast accuracy comparisons of this kind, and the predominant criterion used is the mean squared forecast error (MSFE), or variants thereof. Examples of such studies for the UK can be found in Ash & Smyth (1973), Holden & Peel (1983, 1985, 1986, 1988) and Wallis *et al.* (1986, 1987); also Engle & Yoo (1987) contains an example of a Monte Carlo forecast comparison. The minimum possible MSFE is equivalent to the conditional expectation of the quantity to be forecast given all relevant information, and this, combined with the MSFE's intuitive economic loss interpretation, generates a sound basis for its use as a ranking criterion.

However, Clements & Hendry (1993) criticise MSFE-based measures since they are not invariant to isomorphic transformations and can yield different rankings depending on the forecast horizon considered. Instead they propose a new criterion - the generalised forecast error second moment (GFESM) - which is both invariant to linear transforms and provides a unique ranking, including information on predictions at all horizons of interest. A number of discussants commented on this paper, the general response being one of scepticism centred around the concept of a

need for invariance across different transformations (e.g. levels and changes), the focus on one-step-ahead prediction which GFESM points towards, and the fact that the criterion does not correspond to a natural or intuitive economic loss function.

Chapter 5 picks up on this contentious issue of ranking competing multi-step-ahead forecasts. The literature concerning GFESM is summarised, and the criterion itself is studied.

The justification for using GFESM stems from the theory of predictive likelihood, as reviewed by Bjørnstad (1990), and this basis is explored. In a world of independent replications of the forecast-generating mechanisms, the predictive likelihood foundation is shown to provide a good footing for the use of GFESM as a ranking criterion, but when the more realistic situation of evaluating a string of forecasts in a time series framework is considered, the likelihood justification is found to be more tenuous. In this context, it is necessary to appeal to the replications being a ‘thought experiment’ for the basis to be maintained.

The second issue considered relates to the behaviour of GFESM when comparing two misspecified models. Two noteworthy results are found by way of a simple example - firstly that the ranking yielded by GFESM can be dependent on the arbitrarily chosen maximum forecast horizon, secondly that situations exist where GFESM yields a preference for one forecast whilst another forecast has a lower MSFE for all individual forecast horizons of interest. These two features of the criterion are highly undesirable, and add weight to the criticism of GFESM.

Following this, an alternative measure is proposed - the generalised mean squared forecast error matrix (GMSFEM) - which maintains many of the GFESM advantages (including invariance to isomorphic transformations) whilst ensuring that reversals when the maximum forecast horizon is changed and the counter-intuitive result mentioned above (where GFESM prefers an MSFE-dominated forecast) do not occur. The possibility of an indeterminate conclusion is introduced where neither forecast is preferred; this is in some cases a limitation of the criterion (when the indeterminacy is caused by overly stringent dominance conditions not being met), but from another perspective the detection of situations where neither forecast is completely dominant is valuable, implying that attempts to rank the forecasts by a summary measure are likely to be inappropriate, and evaluation should instead focus on the forecast horizons and representations of interest separately.

The GMSFEM criterion is based on mean squared error dominance for all linear combinations of forecasts, and this leads on to the issue of forecast encompassing for all linear combinations. A test for such forecast encompassing is developed.

Altogether, a number of methods for ranking competing multi-step-ahead forecasts are assessed. One approach is to evaluate purely using the model representation and forecast horizons (individually) of interest using a criterion which corresponds with the economic loss associated with the decision. An alternative is to make use of a single criterion which delivers a unique ranking for all representations and horizons, in which case the new GMSFEM criterion is recommended.

The tests for equal forecast accuracy and forecast encompassing developed in chapters 2-4 and the criteria recommended in chapter 5 together provide reliable and robust statistical procedures for comparing competing forecasts, making a valuable contribution to the theory of economic forecast evaluation.

Chapter 2

Tests for Equal Forecast Accuracy

2.1 Introduction

When evaluating competing economic forecasts, predictive accuracy is of vital importance, and it is therefore necessary to develop formal statistical procedures for comparing the accuracy of rival forecasts. Tests of the null hypothesis that two forecasts of the same variable have equal accuracy (against a one- or two-sided alternative) are of immense value to all involved in forecasting, and thus have a large field of application. The aim of this chapter is to investigate some of these formal forecast-comparison techniques, with a view to developing satisfactory tests of the equal accuracy null.

Tests of equal forecast accuracy must be robust to the wide variety of properties exhibited by the forecast errors (upon which such tests are based) in order to be useful in application. The forecast error characteristics which are particularly pertinent to this analysis are distribution, and correlation through space and time. Forecast errors may well be non-normal, errors from competing forecasts will almost certainly be correlated (as similarities will exist between the information sets used by the forecast producers and some aspects of the actual outcome will ‘surprise’ all forecasters), and for multi-step-ahead forecasts, autocorrelation in the forecast errors will, as a rule, be expected on theoretical grounds, even for optimal forecasts. Diebold & Mariano (1995) examine a number of tests in the light of these issues; this study extends their analysis along a couple of particular lines, with the general aim being to generate tests which are useful in practice, applicable to all sample sizes, and statistically valid for all forecast error properties.

The chapter is structured in five sections. Section 2.2 motivates the research by

considering the Diebold & Mariano (1995) paper, with specific reference to two tests of special interest. Further analysis of these two tests forms the basis of sections 2.3 and 2.4, and the study is concluded in the fifth and final section.

2.2 Motivation for Research

In an attempt to produce more formal statistical procedures for comparing the predictive accuracy of forecasts, Diebold & Mariano (1995) propose tests of the null hypothesis of no difference in the accuracy of two competing forecasts. A number of extant tests are also examined, and their properties investigated under a variety of likely economic conditions. Two main areas of interest result from this paper and shall be highlighted in turn.

2.2.1: The Diebold-Mariano Asymptotic Test

The asymptotic test proposed by Diebold & Mariano involves testing an equivalent null hypothesis to that of equal forecast accuracy. Suppose two competing forecasts are made and have forecast errors e_{1t}, e_{2t} ($t = 1, \dots, n$). Now if the economic loss functions associated with these forecasts are denoted $g(e_{1t}), g(e_{2t})$ respectively, then a loss differential series $d_t = g(e_{1t}) - g(e_{2t})$ can be constructed. The desired null can now be written as $H_0: E(d_t) = 0$, or $H_0: \mu = 0$, where μ is the population mean of the loss differential series. Given covariance stationarity and short memory with regard to d_t , Diebold & Mariano note the asymptotic distribution of the loss differential sample mean:

$$\bar{d} \xrightarrow{d} N(\mu, V) \quad (2.1)$$

which suggests the following test statistic:

$$S_1 = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \xrightarrow{d} N(0,1) \quad \text{under } H_0 \quad (2.2)$$

$$\text{where } \bar{d} = n^{-1} \sum_{t=1}^n d_t$$

$$\hat{V}(\bar{d}) = n^{-1} \left[\hat{\gamma}_0 + 2 \sum_{k=1}^{n-1} \hat{\gamma}_k \right]$$

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$$

Calculation of the variance estimate can be simplified by examination of the forecast error autocovariances. Now a general optimal h -steps-ahead forecast error will be a function of future white noise terms forming an $\text{MA}(h-1)$ type process. Given this $\text{MA}(h-1)$ structure, an optimal h -steps-ahead forecast error will have zero autocovariances for all lags greater than $h-1$. In practice this result may not hold, but would be expected for reasonably well-conceived forecasts, and serves as a useful standard for the analysis of h -steps-ahead forecast errors. Applying this result to the variance estimate of the loss differential sample mean, the following estimator for an h -steps-ahead forecast is obtained:

$$\hat{V}(\bar{d}) = n^{-1} \left[\hat{\gamma}_0 + 2 \sum_{k=1}^{h-1} \hat{\gamma}_k \right] \quad (2.3)$$

where $\hat{\gamma}_k$ is as defined below equation (2.2).

Diebold & Mariano perform Monte Carlo simulation to examine the properties of S_1 , and do so using a quadratic loss differential series to allow comparison with other extant tests. The test is analysed for 2-steps-ahead forecasts ($h = 2$), and the empirical size of S_1 is calculated for a 2-sided test at the nominal 10% level. Normal and non-normal forecast errors with varying degrees of autocorrelation and

contemporaneous correlation are examined, and the sample sizes studied range from $n = 8$ to $n = 512$. Repetition of their simulation with 10,000 replications yields the results given in table 2.1. The sample size is denoted by n , the degree of autocorrelation is given by the value of the MA(1) parameter θ , and contemporaneous correlation between the forecast errors e_{1t} and e_{2t} is given by the value ρ . More formally, the simulation procedure involves generating the following model:

draw u_{1t}, u_{2t} ($t = 1, \dots, n$) from $N(0,1)$ or t_6 distribution

transform to incorporate contemporaneous correlation:

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

transform to incorporate autocorrelation:

$$e_{1t} = (v_{1t} + \theta v_{1,t-1}) / (1 + \theta^2)^{1/2}$$

$$e_{2t} = (v_{2t} + \theta v_{2,t-1}) / (1 + \theta^2)^{1/2}$$

construct quadratic loss differential series:

$$d_t = e_{1t}^2 - e_{2t}^2 \quad t = 1, \dots, n$$

Further detail concerning this simulation procedure is given in section 2.3.

Consideration of the simulation results shows clearly the benefits of the test statistic

Table 2.1Empirical sizes for the Diebold-Mariano test at the nominal 10% level ($h = 2$)

		Normal			t_6		
		$\theta = 0$	$\theta = 0.5$	$\theta = 0.9$	$\theta = 0$	$\theta = 0.5$	$\theta = 0.9$
$n = 8$	$\rho = 0.0$	30.00	28.79	28.53	29.31	27.94	28.06
	$\rho = 0.5$	28.96	28.24	28.59	29.29	28.61	27.89
	$\rho = 0.9$	29.37	29.38	29.14	29.65	28.96	28.69
$n = 16$	$\rho = 0.0$	20.26	19.34	18.96	19.33	18.18	17.93
	$\rho = 0.5$	20.50	19.78	19.18	19.69	18.47	18.09
	$\rho = 0.9$	20.37	19.23	19.07	19.29	18.55	18.53
$n = 32$	$\rho = 0.0$	15.13	14.25	14.34	14.71	14.21	13.84
	$\rho = 0.5$	14.81	14.22	14.10	14.57	13.85	13.62
	$\rho = 0.9$	15.22	15.14	14.82	14.79	13.97	13.94
$n = 64$	$\rho = 0.0$	12.37	12.16	11.91	11.93	11.90	11.82
	$\rho = 0.5$	12.19	11.97	12.04	12.14	12.02	12.10
	$\rho = 0.9$	12.45	12.05	12.20	12.11	12.37	12.23
$n = 128$	$\rho = 0.0$	11.50	11.11	11.13	11.14	11.34	11.39
	$\rho = 0.5$	11.49	11.21	11.04	10.83	11.08	11.00
	$\rho = 0.9$	11.08	10.71	10.78	10.94	10.93	11.00
$n = 256$	$\rho = 0.0$	10.93	11.01	10.91	10.08	9.85	9.85
	$\rho = 0.5$	11.03	11.06	10.88	9.89	10.24	10.24
	$\rho = 0.9$	10.89	10.81	10.87	10.47	10.57	10.62
$n = 512$	$\rho = 0.0$	10.53	10.60	10.72	9.95	9.70	9.85
	$\rho = 0.5$	10.35	10.50	10.55	9.93	9.97	10.01
	$\rho = 0.9$	10.07	10.21	10.28	10.11	10.11	10.11

S_1 . The test is asymptotically correctly sized, with test sizes reasonably close to 10% for the larger sample sizes. These sizes are unaffected by the degree of contemporaneous correlation, and only negligibly impacted by different values of the autocorrelation parameter. The test is also robust to the particular non-normality considered, with sizes approximately equal under both normal and non-normal simulation.

The Diebold-Mariano test S_1 can therefore be used in a wide variety of economic situations, without the need for restrictive assumptions (such as non-autocorrelated, contemporaneously uncorrelated, normal forecast errors). The size is asymptotically correct for all of the examined conditions, and the test construction accommodates a large class of economic loss functions which may be quite general, especially when compared to some of the extant tests which rely on quadratic loss.

The major drawback of this test statistic lies in its small sample properties. It is most common in economic forecasting that long time series are not available, the implication being that few forecast error observations exist for predictive accuracy comparisons. Similarly, if in model estimation, observations are held back so as to perform *ex post* testing of the model's predictive capability, the number of these retained observations (which could be used to help decide between competing forecasts) will again be small in practice. The small and moderate sample properties of any test for comparing the accuracy of different forecasts are thus of great importance. Returning to the examination of table 2.1, it can be seen that the Diebold-Mariano asymptotic test statistic S_1 is seriously oversized in small samples. As shall be examined in section 2.3, this property worsens with longer forecast

horizons.

The conclusion, therefore, is that the test S_1 is highly desirable due to its very general specification and robustness to forecast error properties, but is limited in its use because of the small sample oversizing. Motivation for attempting to correct S_1 for small samples is thus established, and attention to this problem comprises section 2.3.

2.2.2: The Morgan-Granger-Newbold Test

The second area of interest generated by the Diebold & Mariano (1995) paper concerns the extant test attributed to Morgan, Granger & Newbold (Morgan, 1939-40, Granger & Newbold, 1986). This test relies on the assumption of quadratic loss, and also assumes that the forecast errors are normal and have no autocorrelation. The assumption of non-autocorrelated forecast errors implies the test will only be valid for 1-step-ahead forecasting, because for h -steps-ahead forecasts with $h > 1$, autocorrelation does appear in the form of an $MA(h-1)$ process. The test procedure transforms the forecast error vectors as follows:

$$x_t = e_{1t} - e_{2t}, \quad y_t = e_{1t} + e_{2t} \quad (2.4)$$

This orthogonalising transformation allows testing which is robust to contemporaneous correlation in the forecast errors. Given these new variables x_t, y_t , the following can be noted:

$$E(x_t y_t) = E[(e_{1t} - e_{2t})(e_{1t} + e_{2t})]$$

$$\begin{aligned}
&= E[e_{1t}^2 - e_{2t}^2] \\
&= \sigma_1^2 - \sigma_2^2
\end{aligned}$$

where σ_1^2, σ_2^2 are the variances of the forecast errors e_{1t}, e_{2t} respectively. The null hypothesis of interest, that of no difference in forecast accuracy where the economic loss function is quadratic, is now equivalent to $E(x_t y_t) = 0$. This gives rise to the Morgan-Granger-Newbold (MGN) test of zero correlation between x_t and y_t :

$$\begin{aligned}
MGN &= \frac{\hat{\rho}_{x_t y_t}}{\sqrt{(n-1)^{-1}(1 - \hat{\rho}_{x_t y_t}^2)}} \sim t_{n-1} \quad \text{under } H_0 \quad (2.5) \\
\text{where } \hat{\rho}_{x_t y_t} &= \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}}
\end{aligned}$$

The distribution result is given in Hogg & Craig (1978), with the test statistic being distributed as Student's t with $n-1$ degrees of freedom (one degree of freedom is gained by using the common population means (zero) as opposed to the sample means). It can be noted that this distribution result is exact for any sample size (for normal forecast errors), thus the problems associated with tests having an unknown finite sample distribution (as with the Diebold-Mariano test) do not arise here.

Simulation of this test is also performed by Diebold & Mariano, and comparable results for 1-step-ahead forecasting are given in table 2.2. Empirical sizes at the nominal 10% level (2-sided) are again calculated for 10,000 replications, with conditions varying to examine different sample sizes, normal and non-normal

Table 2.2

Empirical sizes for the Morgan-Granger-Newbold test
at the nominal 10% level ($h = 1$)

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	10.18	10.30	10.02	10.11	9.72	10.35	10.67
$\rho = 0.5$	10.04	9.85	10.33	10.30	10.18	10.62	10.40
$\rho = 0.9$	10.09	9.86	10.18	10.43	10.00	10.45	10.05

t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	17.92	20.45	22.59	24.83	25.99	26.07	26.76
$\rho = 0.5$	16.21	18.54	19.80	22.05	22.38	22.62	23.67
$\rho = 0.9$	11.83	12.75	12.61	13.65	13.57	13.72	13.63

forecast errors, and behaviour when the errors are contemporaneously correlated.

As would be expected, the *MGN* test is correctly sized for all sample sizes and values of ρ (the contemporaneous correlation), provided that the forecast errors are normal. The limitation of this test is its unsatisfactory behaviour under conditions of forecast error non-normality. The errors simulated are only moderately non-normal, but create alarming oversizing for all sample sizes. In practice, errors are unlikely to be normal, and samples are generally too small to permit valid testing for normality. Robustness is therefore a desirable and indeed essential property of any test of predictive accuracy.

The primary feature of the simulation results which demands explanation is that *MGN* is asymptotically oversized under non-normality; when $\rho = 0$, the simulation results suggest an asymptotic convergence of the size to a limit around 30%.

Now given the benefits of the asymptotic test S_1 , further analysis of *MGN* is only of value if the test demonstrates superiority in the cases in which it is designed to be valid - i.e. 1-step-ahead prediction with normal forecast errors. If this does occur, attention must be paid to the test's behaviour under non-normality, particularly with regard to the feature noted above. If benefits exist which are peculiar to the Morgan-Granger-Newbold test, then motivation is provided for study concerning correction to attain robustness. Examination of these issues is pursued in section 2.4.

2.3 Diebold-Mariano Approach Tests

The motivation for analysis of Diebold & Mariano's asymptotic test of predictive accuracy, S_1 , is given in section 2.2. The aim now is to attempt to correct for the small sample missizing which S_1 exhibits.

2.3.1 Theory

The key element of S_1 in this regard is the estimator of the variance of the loss differential sample mean. It is useful, therefore, to begin by examining the true variance of the sample mean:

$$\begin{aligned}
 \bar{d} &= n^{-1} \sum_{t=1}^n d_t \\
 V(\bar{d}) &= n^{-2} \left[\sum_{t=1}^n V(d_t) + 2 \sum_{t=1}^{n-1} \sum_{s=t+1}^n C(d_t, d_s) \right] \\
 &= n^{-2} \left[n\gamma_0 + 2 \sum_{k=1}^{n-1} (n-k)\gamma_k \right] \quad \text{where } \gamma_k = C(d_t, d_{t+k}) \\
 &= n^{-1} \left[\gamma_0 + 2 \sum_{k=1}^{h-1} (n-k)\gamma_k \right] \quad \text{for } h\text{-steps-ahead} \quad (2.6)
 \end{aligned}$$

From here, Diebold & Mariano assume n (the sample size) to be large relative to k (the range of which reflects the number of steps ahead forecast), and use the usual sample autocovariance estimator for γ_k to generate their variance estimate $\hat{V}(\bar{d})$, as defined in (2.3). It is convenient now to consider an alternative variance estimator which is more intuitively appealing in its construction. The approximating assumption of n large relative to k is not made, and a different autocovariance

estimator for γ_k is employed - one which is asymptotically equivalent to the above but divides through by $n - k$ rather than n (which is more appropriate when dealing with finite samples):

$$\hat{V}(\bar{d})^* = n^{-1} \left[\hat{\gamma}_0^* + 2n^{-1} \sum_{k=1}^{h-1} (n-k) \hat{\gamma}_k^* \right] \quad (2.7)$$

$$\text{where } \hat{\gamma}_k^* = (n-k)^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$$

It is trivial to show that $\hat{V}(\bar{d})$ and $\hat{V}(\bar{d})^*$ are identical.

The Diebold-Mariano test can further be examined by finding an expression for the expected value of $\hat{V}(\bar{d})$. This can be done by first finding the expectation of the sample autocovariances:

$$E(\hat{\gamma}_k^*) = (n-k)^{-1} E \left[\sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d}) \right] \quad (2.8)$$

The summation term in (2.8) can be expanded as follows:

$$\begin{aligned} & \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d}) \\ &= \sum_{t=k+1}^n d_t d_{t-k} + (n-k) \bar{d}^2 - \bar{d} \sum_{t=k+1}^n d_t - \bar{d} \sum_{t=k+1}^n d_{t-k} \\ &= \sum_{t=k+1}^n d_t d_{t-k} + (n-k) \bar{d}^2 - \bar{d} \left(n\bar{d} - \sum_{t=1}^k d_t \right) - \bar{d} \left(n\bar{d} - \sum_{t=n-k+1}^n d_t \right) \\ &= \sum_{t=k+1}^n d_t d_{t-k} - (n+k) \bar{d}^2 + \bar{d} \left(\sum_{t=1}^k d_t + \sum_{t=n-k+1}^n d_t \right) \end{aligned} \quad (2.9)$$

It is now necessary to examine the expectation of each term. The loss differential series population mean, μ , can be set to zero (as under the null) without loss of generality, giving:

$$E\left[\sum_{t=k+1}^n d_t d_{t-k}\right] = (n-k)\gamma_k \quad (2.10)$$

$$E[(n+k)\bar{d}^2] = (n+k)V(\bar{d}) \quad (2.11)$$

$$\begin{aligned} E\left[\bar{d}\left(\sum_{t=1}^k d_t + \sum_{t=n-k+1}^n d_t\right)\right] &= E\left[n^{-1}\sum_{t=1}^n d_t \sum_{t=1}^k d_t + n^{-1}\sum_{t=1}^n d_t \sum_{t=n-k+1}^n d_t\right] \\ &= 2n^{-1}\left[\sum_{j=1}^{k-1}(k-j)\gamma_j + \sum_{j=0}^{n-k} k\gamma_j \right. \\ &\quad \left. + \sum_{j=1}^{k-1}(k-j)\gamma_{n-k+j}\right] \end{aligned} \quad (2.12)$$

Substituting results (2.10)-(2.12) back into the sample autocovariance expectation (2.8) gives:

$$\begin{aligned} E(\hat{\gamma}_k^*) &= (n-k)^{-1}\{(n-k)\gamma_k - (n+k)V(\bar{d}) \\ &\quad + 2n^{-1}[\sum_{j=1}^{k-1}(k-j)\gamma_j + \sum_{j=0}^{n-k} k\gamma_j + \sum_{j=1}^{k-1}(k-j)\gamma_{n-k+j}]\} \\ &= \gamma_k - (n-k)^{-1}(n+k)V(\bar{d}) + O(n^{-2}) \end{aligned} \quad (2.13)$$

Assuming now that n is large relative to k (as Diebold & Mariano) and approximating to order n^{-1} , the following result is obtained:

$$E(\hat{\gamma}_k^*) \approx \gamma_k - V(\bar{d}) \quad (2.14)$$

The desired expectation of the loss differential sample mean variance estimate can

now be evaluated using (2.7) and (2.14):

$$\begin{aligned}
E[\hat{V}(\bar{d})^*] &\approx n^{-1} \left[\gamma_0 - V(\bar{d}) + 2n^{-1} \sum_{k=1}^{h-1} (n-k)(\gamma_k - V(\bar{d})) \right] \\
&= V(\bar{d}) - n^{-1} V(\bar{d}) \left[1 + 2n^{-1} \sum_{k=1}^{h-1} (n-k) \right] \\
&= \frac{n+1-2h+n^{-1}h(h-1)}{n} V(\bar{d}) \tag{2.15}
\end{aligned}$$

This result clearly shows that the Diebold-Mariano variance estimate is biased, and by the factor given in the final expression (2.15). This bias persists to order n^{-1} even if the final order n^{-2} term is dropped. A case for dropping this term can be made on theoretical grounds as the analysis already involves approximation to order n^{-1} . However, a case can also be made for retaining this term as shall be examined below.

The implication, therefore, is that the test statistic S_1 can be corrected for its finite sample oversizing to some extent by using the following approximately unbiased variance estimate and associated test statistic:

$$S^* = \frac{\bar{d}}{\sqrt{\hat{V}_m(\bar{d})}} \tag{2.16}$$

$$\text{where } \hat{V}_m(\bar{d}) = [n+1-2h+n^{-1}h(h-1)]^{-1} \left[\hat{\gamma}_0 + 2 \sum_{k=1}^{h-1} \hat{\gamma}_k \right]$$

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$$

$$\text{i.e. } S^* = n^{-1/2} [n+1-2h+n^{-1}h(h-1)]^{1/2} S_1$$

A new test statistic, S^* , is thus derived, correcting for the small sample bias in the original variance estimate. This correction, however, is dependent on the assumption of n being large relative to k , and is thus an approximation of the true bias. The exception to this is when the true loss differential series is white noise (it is assumed that such information is unknown and an $\text{MA}(h-1)$ process is still used in the test's construction). In this case, the following is true, using the results for white noise, $V(\bar{d}) = n^{-1}\gamma_0$ and $\gamma_k = 0$ for $k \neq 0$:

$$\begin{aligned} E(\hat{\gamma}_k^*) &= \gamma_k - (n-k)^{-1}(n+k)V(\bar{d}) \\ &\quad + 2n^{-1}(n-k)^{-1} \left[\sum_{j=1}^{k-1} (k-j)\gamma_j + \sum_{j=0}^{n-k} k\gamma_j + \sum_{j=1}^{k-1} (k-j)\gamma_{n-k+j} \right] \\ &= \gamma_k - (n-k)^{-1}(n+k)V(\bar{d}) + 2n^{-1}(n-k)^{-1}k\gamma_0 \\ &= \gamma_k - V(\bar{d}) \\ &= \gamma_0 - V(\bar{d}), k=0; -V(\bar{d}), k \neq 0 \end{aligned}$$

The result is now exact, the rest of the analysis follows through, and an expression for $E[\hat{V}(\bar{d})^*]$ identical to that in (2.15) is obtained, but this time no approximation is necessary due to the nature of the white noise forecast errors. This exact result with the order n^{-2} term included generates a case for retaining the final term, as

alluded to earlier.

Having established a corrected test statistic, it is also valuable to examine its distribution. The test statistic S_1 proposed by Diebold & Mariano has an asymptotic normal distribution, and so the correction S^* will also be normally distributed in the limit - the only difference between S_1 and S^* is the bias correction which does not affect the asymptotic distribution. Now consideration of these test statistics enables the following intuition to be made. The test constructions take a form typical of a standard test for the significance of a population mean, i.e. sample mean divided by the estimated standard deviation. Such a test has an asymptotic normal distribution, but in finite samples takes a Student's t distribution with $n - 1$ degrees of freedom - the usual t -ratio. Given the similarity of S_1 and S^* to such a test, it is intuitively appealing to compare the test statistics with critical values from a t_{n-1} distribution in finite samples. In fact when the errors are normally distributed and 1-step-ahead prediction is considered, such an approach is exactly correct.

Two modifications to the Diebold-Mariano asymptotic test are thus proposed - firstly an approximate correction for the small sample bias in the estimated variance of the loss differential sample mean, and secondly use of the Student's t distribution critical values for finite sample tests. The impact of these modifications can now be analysed by Monte Carlo simulation.

2.3.2 Simulation

Following Diebold & Mariano, Monte Carlo simulation is performed to calculate the finite sample size of the test statistics concerned. The economic loss function is assumed to be quadratic, a variety of forecast error properties are examined, and all the tests are evaluated at the nominal 10% level with a null of equal forecast accuracy and a 2-sided alternative; 10,000 replications are performed for each simulation. A range of forecast horizons are examined, with experiments conducted for 1- through 10-steps-ahead forecasts.

Normal and non-normal errors are both examined, and varying degrees of autocorrelation and contemporaneous correlation are also considered. With regard to the method of generating these errors, the respective situations of normality and non-normality are explained in turn.

2.3.2a Normal Forecast Errors

The normal forecast errors are generated by drawing realisations from a bivariate standard normal distribution:

$$u = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim N(0, I)$$

In order to incorporate contemporaneous correlation, the vector u is premultiplied by a matrix P such that:

$$v = Pu \sim N(0, \Omega)$$

$$\text{where } \Omega = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

ρ = degree of contemporaneous correlation, $0 \leq \rho \leq 1$

Given that $PP' = \Omega$, the natural choice for P is the triangular matrix:

$$P = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{bmatrix}$$

This transformation yields the vector v which can itself be transformed to include autocorrelation. For 1-step-ahead forecasts, no attention need be paid to autocorrelation as the errors are assumed to be white noise. However, for 2-steps-ahead forecasts, an MA(1) type process is expected and autocorrelation must be considered. The transformation is as follows:

$$e = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} (v_{1t} + \theta v_{1,t-1}) / (1 + \theta^2)^{1/2} \\ (v_{2t} + \theta v_{2,t-1}) / (1 + \theta^2)^{1/2} \end{bmatrix}$$

where θ = degree of autocorrelation, $0 \leq \theta \leq 1$

This transformation requires values for $v_{1,0}$ and $v_{2,0}$ - these are again drawn from a bivariate standard normal distribution and transformed as above to incorporate contemporaneous correlation (the Diebold-Mariano $v_{1,0} = v_{2,0} = 0$ assumption is unnecessary and undesirable).

For h -steps-ahead forecasts ($h > 2$), autocorrelation will appear in an MA($h - 1$) form, as theorised in the previous sub-section. Now, as shall be seen later, the test statistics S_1 and S^* are robust to forecast error autocorrelation with the sizes not significantly affected by changes in the MA parameter for 2-steps-ahead forecasting. It is reasonable, therefore, to simplify the analysis by examining only the white

noise case for 3- through 10-steps-ahead forecast simulations (note an $MA(h-1)$ process is still assumed for purposes of test statistic calculation, but white noise is simulated, ie. $\theta_1 = \theta_2 = \dots = \theta_{h-1} = 0$ so $e_{1t}, e_{2t} = v_{1t}, v_{2t}$ respectively). It can also be noted that S_1 and S^* are largely unaffected by both forecast error contemporaneous correlation and distribution, and so for longer step-ahead forecasting (5-steps through 10-steps), only normal, contemporaneously uncorrelated and non-autocorrelated errors are considered because a sufficient picture of test statistic size is gained from these conditions alone due to the robustness of the test statistics. No results are given for $n = 8$ for 8-, 9- and 10-steps-ahead forecasts because for such a small sample size, insufficient information exists to construct the test statistics.

2.3.2b Non-Normal Forecast Errors

Turning now to non-normal errors, two different methods of error generation are used in the simulations. The method used by Diebold & Mariano involves generating two independent Student's t random variables with six degrees of freedom:

$$u_{it} = \frac{z_{it}}{\sqrt{\chi_{it,6}^2/6}} \quad i = 1, 2$$

where $z_{it} \sim N(0,1)$

$\chi_{it,6}^2$ is independent of z_{it}

Contemporaneous correlation and autocorrelation are then built in following the same procedure as for normal errors, resulting again in errors e_{1t}, e_{2t} .

The second method of non-normal simulation generates two random variables which follow a bivariate Student's t_6 distribution, as formalised by Dunnett & Sobel (1954). In order to do this, realisations are drawn from a bivariate standard normal distribution:

$$z \sim N(0, I)$$

and first transformed to include contemporaneous correlation:

$$v = Pz \sim N(0, \Omega)$$

The random variables v_{1t}, v_{2t} are now transformed to follow a t_6 distribution, performed by dividing each realisation by the same chi-squared random variable at a given point in time t :

$$w_{it} = \frac{v_{it}}{\sqrt{\chi_{t,6}^2/6}} \quad i = 1, 2$$

where $\chi_{t,6}^2$ is independent of v_{it}

Autocorrelation can again be incorporated to yield the errors e_{1t}, e_{2t} .

The reasons for employing this second method in addition to the Diebold-Mariano procedure are twofold. Firstly, under the Diebold-Mariano methodology, the resulting errors (ignoring autocorrelation without loss of generality) can be

decomposed as follows:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

$$\text{i.e. } e_{1t} = u_{1t}$$

$$e_{2t} = \rho u_{1t} + (1-\rho^2)^{1/2} u_{2t}$$

Now the realisations u_{1t}, u_{2t} are independent t_6 random variables, but the forecast errors e_{1t}, e_{2t} generated in this way do not follow a t_6 distribution. The first error, e_{1t} , is t_6 , but the second, e_{2t} , is a linear combination of t_6 variates which is not then t_6 . Employment of the second method of non-normal error generation, however, does not experience this, with the errors following a bivariate t_6 distribution. The second reason for using this latter method is that the procedure involves dividing the realisations by the same chi-squared random variable at a given point in time. This implies that, even in the case where the forecast errors are not autocorrelated or contemporaneously correlated, the squared errors will be correlated:

$$e_{1t} = \frac{z_{1t}}{\sqrt{\kappa_t}}; \quad e_{2t} = \frac{z_{2t}}{\sqrt{\kappa_t}} \quad \text{where } \kappa_t = \chi_{t,6}^2/6$$

$$\begin{aligned} \text{Corr}(e_{1t}^2, e_{2t}^2) &= \frac{C(e_{1t}^2, e_{2t}^2)}{\sqrt{V(e_{1t}^2)V(e_{2t}^2)}} \\ &= \frac{E(\kappa_t^{-2} z_{1t}^2 z_{2t}^2) - E(\kappa_t^{-1} z_{1t}^2)E(\kappa_t^{-1} z_{2t}^2)}{\sqrt{[E(\kappa_t^{-2} z_{1t}^4) - [E(\kappa_t^{-1} z_{1t}^2)]^2][E(\kappa_t^{-2} z_{2t}^4) - [E(\kappa_t^{-1} z_{2t}^2)]^2]}} \end{aligned}$$

$$= \frac{A_2 \mu_2^2 - A_1^2 \mu_2^2}{A_2 \mu_4 - A_1^2 \mu_2^2}$$

$$\text{where } A_k = E(\kappa_t^{-k}) = \frac{6^k \Gamma(3-k)}{2^k \Gamma(3)}$$

$\mu_j = j$ 'th moment of N(0,1) distribution

$$= 0.2 \neq 0$$

This property of the errors is intuitively appealing, the interpretation being that the economic series concerned is harder to predict at some times than others, thus all forecasters will have a larger error variance at such times, and smaller at others. These two reasons form two distinct advantages of the latter method compared to that employed by Diebold & Mariano. The bivariate method is thus preferable, but both methods are used in the simulations for completeness. In the tables, the Diebold-Mariano non-normal errors are denoted 'DM t_6 ' and the latter method's non-normal errors denoted 'Biv. t_6 '.

The resulting forecast error series under all three distributions, e_{1t}, e_{2t} ($t = 1, \dots, n$) now contain all the properties desired for the analysis, and can be used to construct the loss differential series, d_t :

$$d_t = e_{1t}^2 - e_{2t}^2 \quad t = 1, \dots, n$$

Sample sizes of $n = 8, 16, 32, 64, 128, 256, 512$, contemporaneous correlation parameters of $\rho = 0, 0.5, 0.9$, and autocorrelation parameters of $\theta = 0, 0.5, 0.9$ are used. Simulations are performed for the original Diebold-Mariano test S_1 using standard normal critical values, the fully modified test S^* using t_{n-1} critical values,

and for the individual modifications applied separately, i.e. S_1 using t_{n-1} critical values and S^* using standard normal critical values. The results are given in tables 2.3-2.10.

2.3.3 Results

With reference to the results tables, a number of observations and inferences can be made. Firstly, table 2.3 gives the empirical test sizes for 1-step-ahead forecasts, and shows that the original Diebold-Mariano test is oversized in the smaller samples. The sizes are not significantly affected by variations in the level of contemporaneous correlation, and only marginally by the distribution of the forecast errors, confirming the observations made in section 2.2 (with 2-steps-ahead forecast simulation). As each of the two adjustments are applied to this test, the size is reduced with the fully modified test completely overcoming the problem of small sample oversizing, in certain cases to the extreme of the test being undersized.

Tables 2.4-2.6 contain the 2-steps-ahead forecast results with different degrees of autocorrelation simulated in each table. These cases correspond to those examined by Diebold & Mariano and again give rise to the inference that the corrections improve the small sample test sizes. The sizes for all the simulated tests exhibit robustness to forecast error distribution and contemporaneous correlation as before; analysis of the results in these tables also now reveals that the tests are robust to varying degrees of forecast error autocorrelation. Autocorrelation and departure from normality have small effects on the test sizes, and in general the empirical sizes are closer to the nominal sizes in such cases. Taking the 2-steps-ahead forecast

Table 2.3

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($h = 1$)

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	16.67	13.49	11.58	10.94	10.29	10.62	10.80
	11.02	10.81	10.29	10.38	10.02	10.50	10.68
	13.83	12.00	10.94	10.64	10.14	10.54	10.75
	8.38	9.63	9.70	10.10	9.87	10.42	10.63
$\rho = 0.5$	16.22	13.00	11.62	11.08	10.57	10.92	10.60
	10.48	10.57	10.31	10.50	10.31	10.75	10.54
	13.33	11.68	10.99	10.80	10.44	10.80	10.54
	8.29	9.53	9.78	10.32	10.18	10.70	10.49
$\rho = 0.9$	16.66	12.90	11.31	11.18	10.35	10.93	10.42
	11.03	10.41	10.23	10.57	10.07	10.81	10.37
	13.74	11.56	10.65	10.86	10.19	10.87	10.39
	8.74	9.37	9.78	10.31	9.93	10.72	10.35

DM t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	14.98	12.10	11.04	10.62	9.67	10.00	9.62
	8.78	9.33	9.79	10.02	9.47	9.90	9.53
	11.99	10.61	10.32	10.31	9.58	9.92	9.57
	6.64	8.31	9.30	9.65	9.30	9.80	9.49
$\rho = 0.5$	15.07	12.30	11.25	10.39	9.82	9.73	9.40
	9.21	9.89	10.08	9.91	9.57	9.59	9.36
	12.12	11.01	10.60	10.11	9.66	9.66	9.37
	6.96	8.76	9.56	9.64	9.45	9.49	9.30
$\rho = 0.9$	15.01	12.85	11.11	10.85	10.28	9.61	9.63
	9.41	10.44	10.07	10.15	9.94	9.50	9.62
	12.23	11.58	10.61	10.49	10.08	9.54	9.62
	7.19	9.34	9.45	9.88	9.82	9.41	9.60

Biv. t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	15.17	12.08	10.57	9.92	10.03	9.74	9.70
	9.19	9.47	9.53	9.40	9.61	9.58	9.69
	12.24	10.67	10.04	9.65	9.83	9.62	9.69
	7.21	8.22	8.94	9.19	9.54	9.49	9.65
$\rho = 0.5$	14.76	11.81	10.92	10.33	9.82	9.72	9.44
	9.10	9.30	9.51	9.89	9.52	9.59	9.43
	11.99	10.50	10.18	10.10	9.68	9.65	9.43
	7.13	8.08	8.88	9.53	9.37	9.56	9.40
$\rho = 0.9$	14.59	11.81	10.86	10.11	10.08	10.13	10.09
	9.05	9.40	9.57	9.38	9.85	9.98	9.97
	11.60	10.55	10.19	9.77	9.96	10.06	10.03
	6.87	8.34	8.90	9.14	9.72	9.89	9.93

Note:- The first entry in each cell is for the S_1 test using $N(0,1)$ critical values, the second for the S_1 test using t_{n-1} critical values, the third for the S^* test using $N(0,1)$ critical values, and the fourth for the S^* test using t_{n-1} critical values.

Table 2.4

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($h = 2$, $\theta = 0$)

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	30.00	20.26	15.13	12.37	11.50	10.93	10.53
	23.92	17.83	13.91	12.03	11.21	10.76	10.47
	21.10	16.43	13.24	11.71	11.05	10.64	10.42
	16.42	14.18	12.19	11.22	10.75	10.49	10.34
$\rho = 0.5$	28.96	20.50	14.81	12.19	11.49	11.03	10.35
	23.25	17.91	13.85	11.63	11.25	10.93	10.24
	20.74	16.64	13.22	11.42	11.12	10.86	10.18
	15.78	14.65	12.23	10.89	10.81	10.77	10.11
$\rho = 0.9$	29.37	20.37	15.22	12.45	11.08	10.89	10.07
	23.49	17.75	14.23	12.13	10.88	10.75	10.03
	20.80	16.52	13.61	11.81	10.73	10.69	10.02
	16.46	14.10	12.54	11.26	10.48	10.53	9.99

DM t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	29.31	19.33	14.71	11.93	11.14	10.08	9.95
	23.08	16.68	13.50	11.34	10.86	9.84	9.91
	20.15	15.25	12.68	10.90	10.66	9.74	9.88
	15.39	12.94	11.51	10.37	10.24	9.60	9.75
$\rho = 0.5$	29.29	19.69	14.57	12.14	10.83	9.89	9.93
	23.02	16.94	13.36	11.47	10.59	9.75	9.84
	20.16	15.39	12.72	11.15	10.48	9.68	9.79
	15.20	13.11	11.54	10.56	10.33	9.57	9.75
$\rho = 0.9$	29.65	19.29	14.79	12.11	10.94	10.47	10.11
	23.10	16.48	13.73	11.58	10.61	10.31	10.06
	20.59	15.26	12.99	11.20	10.37	10.25	9.96
	15.73	13.05	11.58	10.58	10.07	10.17	9.91

Biv. t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	28.13	18.95	14.09	11.92	10.99	10.20	9.77
	21.88	16.20	12.75	11.45	10.66	10.05	9.72
	19.19	14.80	12.14	11.05	10.51	9.99	9.67
	14.46	12.43	11.12	10.40	10.26	9.88	9.61
$\rho = 0.5$	28.31	19.42	14.11	11.66	10.69	10.30	10.24
	22.17	16.50	12.78	11.15	10.31	10.11	10.18
	19.25	15.20	12.04	10.87	10.20	10.01	10.14
	14.29	13.02	10.80	10.31	9.83	9.87	10.05
$\rho = 0.9$	28.15	19.55	14.31	11.50	10.46	10.21	10.10
	21.98	16.66	13.02	11.01	10.20	10.04	10.04
	19.15	15.34	12.36	10.73	10.11	9.97	10.03
	14.31	12.81	11.07	10.17	9.86	9.81	9.95

Note:- The first entry in each cell is for the S_1 test using $N(0,1)$ critical values, the second for the S_1 test using t_{n-1} critical values, the third for the S^* test using $N(0,1)$ critical values, and the fourth for the S^* test using t_{n-1} critical values.

Table 2.5

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($h = 2$, $\theta = 0.5$)

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	28.79	19.34	14.25	12.16	11.11	11.01	10.60
	22.58	16.63	13.24	11.59	10.84	10.85	10.54
	19.84	15.30	12.54	11.23	10.64	10.75	10.48
	15.03	12.86	11.50	10.61	10.38	10.64	10.42
$\rho = 0.5$	28.24	19.78	14.22	11.97	11.21	11.06	10.50
	22.14	17.02	13.03	11.47	10.84	10.98	10.43
	19.32	15.51	12.42	11.18	10.75	10.87	10.40
	14.84	13.08	11.33	10.61	10.50	10.69	10.35
$\rho = 0.9$	29.38	19.23	15.14	12.05	10.71	10.81	10.21
	22.75	16.70	13.93	11.59	10.45	10.71	10.16
	19.92	15.29	13.13	11.21	10.31	10.63	10.14
	14.84	12.97	11.54	10.75	10.04	10.55	10.09

DM t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	27.94	18.18	14.21	11.90	11.34	9.85	9.70
	21.58	15.49	12.80	11.23	11.05	9.64	9.65
	18.97	14.16	12.06	10.87	10.92	9.53	9.61
	13.89	11.77	10.75	10.26	10.56	9.44	9.57
$\rho = 0.5$	28.61	18.47	13.85	12.02	11.08	10.24	9.97
	22.02	15.65	12.64	11.31	10.77	10.07	9.91
	19.29	14.21	11.88	10.99	10.59	9.98	9.89
	14.40	11.92	10.54	10.44	10.38	9.83	9.82
$\rho = 0.9$	28.96	18.55	13.97	12.37	10.93	10.57	10.11
	22.36	15.75	12.72	11.73	10.66	10.39	10.06
	19.72	14.33	11.93	11.40	10.51	10.35	10.03
	14.80	11.88	10.80	10.88	10.21	10.24	9.97

Biv. t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	27.61	18.31	13.96	11.60	11.11	10.56	9.73
	21.27	15.48	12.49	11.18	10.84	10.45	9.68
	18.67	14.03	11.88	10.80	10.63	10.40	9.62
	13.79	11.57	10.88	10.33	10.34	10.25	9.54
$\rho = 0.5$	27.71	18.41	13.64	11.81	10.65	10.61	10.04
	21.22	15.62	12.53	11.25	10.33	10.46	9.97
	18.61	14.10	11.80	10.93	10.11	10.43	9.93
	13.75	11.76	10.50	10.39	9.84	10.34	9.89
$\rho = 0.9$	28.12	18.43	13.64	11.49	10.39	10.29	10.44
	21.34	15.53	12.52	10.91	10.16	10.22	10.38
	18.52	13.91	11.72	10.51	9.96	10.12	10.34
	13.40	11.67	10.53	9.85	9.74	10.01	10.24

Note:- The first entry in each cell is for the S_1 test using $N(0,1)$ critical values, the second for the S_1 test using t_{n-1} critical values, the third for the S^* test using $N(0,1)$ critical values, and the fourth for the S^* test using t_{n-1} critical values.

Table 2.6

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($h = 2$, $\theta = 0.9$)

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	28.53	18.96	14.34	11.91	11.13	10.91	10.72
	22.06	16.22	12.98	11.42	10.85	10.79	10.64
	19.32	14.67	12.26	11.06	10.70	10.69	10.63
	14.26	12.40	11.10	10.44	10.41	10.60	10.55
$\rho = 0.5$	28.59	19.18	14.10	12.04	11.04	10.88	10.55
	22.09	16.59	12.87	11.38	10.76	10.76	10.51
	19.19	14.91	12.29	11.13	10.65	10.71	10.48
	14.61	12.70	11.33	10.60	10.43	10.65	10.38
$\rho = 0.9$	29.14	19.07	14.82	12.20	10.78	10.87	10.28
	22.58	16.15	13.60	11.66	10.41	10.80	10.18
	19.73	14.72	12.92	11.25	10.28	10.70	10.14
	14.42	12.16	11.78	10.66	10.09	10.61	10.10

DM t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	28.06	17.93	13.84	11.82	11.39	9.85	9.85
	21.24	15.07	12.49	11.23	11.08	9.74	9.82
	18.19	13.61	11.77	10.84	10.95	9.70	9.78
	13.38	11.05	10.63	10.38	10.70	9.55	9.72
$\rho = 0.5$	27.89	18.09	13.62	12.10	11.00	10.24	10.01
	21.09	15.34	12.41	11.50	10.76	10.12	9.97
	18.41	13.83	11.67	11.15	10.59	9.99	9.94
	13.94	11.30	10.49	10.53	10.38	9.84	9.87
$\rho = 0.9$	28.69	18.53	13.94	12.23	11.00	10.62	10.11
	21.95	15.54	12.77	11.64	10.72	10.46	10.04
	19.11	14.12	12.07	11.30	10.59	10.37	9.99
	14.22	11.96	10.96	10.67	10.27	10.17	9.92

Biv. t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	27.50	18.14	13.70	11.56	11.03	10.50	9.80
	20.94	15.20	12.28	10.96	10.75	10.41	9.77
	18.17	13.60	11.59	10.67	10.58	10.34	9.69
	13.35	10.95	10.52	10.11	10.22	10.24	9.61
$\rho = 0.5$	27.91	18.13	13.62	11.66	10.82	10.67	10.13
	21.10	15.13	12.31	11.04	10.57	10.54	10.10
	18.40	13.73	11.60	10.62	10.44	10.44	10.08
	13.65	11.36	10.41	10.05	10.11	10.37	10.03
$\rho = 0.9$	27.33	17.91	13.44	11.17	10.30	10.29	10.42
	20.59	14.82	11.97	10.78	10.12	10.11	10.34
	17.76	13.32	11.38	10.56	9.96	10.03	10.30
	12.90	10.99	10.35	9.98	9.79	9.87	10.26

Note:- The first entry in each cell is for the S_1 test using $N(0,1)$ critical values, the second for the S_1 test using t_{n-1} critical values, the third for the S^* test using $N(0,1)$ critical values, and the fourth for the S^* test using t_{n-1} critical values.

Table 2.7

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($h = 3$, $\theta_i = 0 \ \forall i$)

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	36.87	26.53	18.26	14.10	11.72	11.34	11.15
	30.90	24.12	17.01	13.73	11.53	11.16	11.06
	22.30	20.35	15.14	12.79	11.02	10.96	10.94
	18.14	18.47	14.27	12.16	10.73	10.84	10.89
$\rho = 0.5$	37.70	26.63	17.81	14.65	12.20	11.51	11.01
	31.60	24.02	16.70	14.03	11.95	11.39	10.96
	22.75	20.25	15.06	13.17	11.56	11.23	10.81
	18.37	17.93	14.04	12.52	11.32	11.09	10.76
$\rho = 0.9$	37.50	25.71	18.21	14.45	12.04	11.73	10.86
	32.03	23.44	16.94	13.88	11.79	11.55	10.80
	23.00	19.43	15.24	12.95	11.32	11.32	10.73
	18.16	17.32	14.17	12.37	11.10	11.19	10.65

DM t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	38.17	26.18	18.20	14.19	11.65	10.74	10.02
	31.21	23.20	16.79	13.68	11.23	10.50	9.95
	21.67	19.00	14.74	12.83	10.84	10.32	9.79
	17.16	16.91	13.67	12.28	10.53	10.11	9.71
$\rho = 0.5$	38.13	26.16	18.26	14.22	11.72	10.51	9.91
	31.76	23.33	16.92	13.62	11.39	10.39	9.84
	22.40	19.10	14.83	12.41	10.86	10.20	9.76
	17.64	16.88	13.92	11.90	10.62	10.07	9.67
$\rho = 0.9$	38.40	25.91	18.44	13.85	12.11	10.13	10.07
	31.84	23.20	17.21	13.23	11.76	10.00	10.00
	22.27	19.15	15.42	12.24	11.28	9.92	9.90
	17.30	17.10	14.17	11.64	10.96	9.79	9.88

Biv. t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	38.27	25.42	18.42	13.56	11.60	10.30	9.92
	31.80	22.51	17.04	12.97	11.27	10.20	9.88
	22.44	18.51	14.87	11.96	10.96	9.94	9.73
	17.89	16.30	13.84	11.36	10.70	9.86	9.69
$\rho = 0.5$	38.30	25.09	17.95	13.62	11.16	10.55	10.11
	31.29	22.40	16.75	12.94	10.86	10.45	10.08
	21.92	18.27	14.68	11.95	10.48	10.22	9.98
	17.06	16.25	13.74	11.41	10.25	10.11	9.89
$\rho = 0.9$	37.40	25.40	17.91	13.62	11.40	10.77	10.42
	30.69	22.45	16.75	13.00	11.17	10.69	10.34
	20.58	18.03	14.67	12.09	10.73	10.46	10.27
	15.93	15.75	13.68	11.58	10.50	10.34	10.23

Note:- The first entry in each cell is for the S_1 test using $N(0,1)$ critical values, the second for the S_1 test using t_{n-1} critical values, the third for the S^* test using $N(0,1)$ critical values, and the fourth for the S^* test using t_{n-1} critical values.

Table 2.8

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($h = 4$, $\theta_i = 0 \ \forall i$)

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	43.22	30.90	21.27	15.94	12.96	11.55	11.21
	37.40	28.29	19.95	15.40	12.69	11.48	11.14
	20.87	22.06	17.16	13.92	11.77	11.05	11.00
	16.25	19.83	16.14	13.36	11.52	10.93	10.95
$\rho = 0.5$	43.57	30.99	21.37	16.55	12.69	11.95	11.03
	37.52	28.37	20.22	15.87	12.42	11.91	10.99
	20.72	21.80	17.16	14.46	11.87	11.54	10.84
	16.63	19.71	16.05	13.93	11.53	11.36	10.79
$\rho = 0.9$	44.06	30.58	21.84	16.25	12.88	12.19	10.92
	37.73	28.06	20.61	15.56	12.63	12.04	10.85
	20.56	21.55	17.51	13.97	11.90	11.59	10.75
	16.38	19.45	16.48	13.40	11.62	11.54	10.73

DM t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	45.24	31.06	21.48	16.09	12.08	11.13	10.07
	38.44	28.20	20.32	15.54	11.82	11.03	9.99
	20.93	21.38	16.95	13.85	11.08	10.66	9.80
	16.25	19.00	16.13	13.39	10.84	10.55	9.71
$\rho = 0.5$	45.42	30.92	21.63	15.93	12.09	10.66	10.05
	38.84	28.44	20.27	15.43	11.80	10.55	9.96
	20.65	21.35	17.00	13.83	11.04	10.13	9.83
	16.24	19.35	15.92	13.29	10.81	10.01	9.78
$\rho = 0.9$	45.17	30.56	21.89	15.85	12.71	10.46	10.33
	38.81	27.70	20.76	15.19	12.46	10.36	10.29
	20.63	20.53	17.46	13.69	11.81	10.04	10.14
	16.18	18.64	16.41	13.04	11.50	9.89	10.07

Biv. t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	45.61	30.76	21.67	15.33	12.51	10.78	10.17
	39.51	27.80	20.49	14.84	12.26	10.62	10.15
	21.33	20.79	16.82	13.44	11.55	10.16	10.04
	16.65	18.66	15.82	12.88	11.33	10.00	10.00
$\rho = 0.5$	45.64	30.33	21.32	15.47	11.79	11.04	10.36
	39.08	27.89	20.07	14.85	11.48	10.92	10.33
	20.98	20.59	16.72	13.24	10.74	10.50	10.08
	16.61	18.44	15.84	12.73	10.51	10.40	9.98
$\rho = 0.9$	45.08	30.90	21.39	15.56	12.14	11.33	10.56
	38.85	28.18	20.20	15.02	11.90	11.15	10.50
	20.29	20.87	16.72	13.37	11.21	10.66	10.34
	16.07	18.65	15.60	12.83	10.90	10.52	10.28

Note:- The first entry in each cell is for the S_1 test using $N(0,1)$ critical values, the second for the S_1 test using t_{n-1} critical values, the third for the S^* test using $N(0,1)$ critical values, and the fourth for the S^* test using t_{n-1} critical values.

Table 2.9

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($\rho = 0$, $\theta_i = 0 \ \forall i$, normal errors)

	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$h = 5$	49.35	34.52	24.49	17.98	13.78	11.90	11.36
	43.52	31.80	23.38	17.41	13.51	11.66	11.29
	16.58	22.07	18.86	15.44	12.42	11.16	11.07
	12.87	19.86	17.79	14.91	12.20	11.07	10.98
$h = 6$	58.36	37.30	26.74	19.59	14.82	12.21	11.80
	52.66	34.81	25.61	19.08	14.53	12.09	11.67
	13.47	21.84	19.64	16.39	13.20	11.53	11.26
	10.55	19.78	18.77	16.02	12.92	11.42	11.21
$h = 7$	72.35	39.36	28.75	20.81	15.69	12.67	11.96
	68.28	36.94	27.51	20.38	15.44	12.50	11.91
	12.47	20.42	20.48	17.29	13.81	11.63	11.45
	9.91	18.21	19.51	16.83	13.62	11.55	11.37
$h = 8$	-	42.59	30.82	22.94	16.26	13.07	11.99
	-	39.77	29.70	22.35	16.02	12.91	11.90
	-	19.24	21.00	18.45	14.18	12.01	11.45
	-	17.43	20.20	17.99	13.82	11.91	11.39
$h = 9$	-	45.29	32.35	24.46	17.53	13.84	12.24
	-	42.66	31.31	23.88	17.18	13.75	12.20
	-	16.87	21.20	19.46	15.03	12.49	11.62
	-	15.12	20.21	19.02	14.74	12.35	11.56
$h = 10$	-	48.97	33.43	25.31	17.87	14.18	12.38
	-	46.44	32.18	24.86	17.59	14.08	12.29
	-	15.50	21.13	19.59	15.35	12.70	11.77
	-	13.97	20.24	19.07	15.14	12.60	11.75

Note:- The first entry in each cell is for the S_1 test using $N(0,1)$ critical values, the second for the S_1 test using t_{n-1} critical values, the third for the S^* test using $N(0,1)$ critical values, and the fourth for the S^* test using t_{n-1} critical values.

Table 2.10

Empirical sizes for the original and modified Diebold-Mariano tests
at the nominal 10% level ($\rho = 0$, $\theta_i = 0 \ \forall i$, normal errors)

	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$h = 1$	16.67 8.38	13.49 9.63	11.58 9.70	10.94 10.10	10.29 9.87	10.62 10.42	10.80 10.63
$h = 2$	30.00 16.42	20.26 14.18	15.13 12.19	12.37 11.22	11.50 10.75	10.93 10.49	10.53 10.34
$h = 3$	36.87 18.14	26.53 18.47	18.26 14.27	14.10 12.16	11.72 10.73	11.34 10.84	11.15 10.89
$h = 4$	43.22 16.25	30.90 19.83	21.27 16.14	15.94 13.36	12.96 11.52	11.55 10.93	11.21 10.95
$h = 5$	49.35 12.87	34.52 19.86	24.49 17.79	17.98 14.91	13.78 12.20	11.90 11.07	11.36 10.98
$h = 6$	58.36 10.55	37.30 19.78	26.74 18.77	19.59 16.02	14.82 12.92	12.21 11.42	11.80 11.21
$h = 7$	72.35 9.91	39.36 18.21	28.75 19.51	20.81 16.83	15.69 13.62	12.67 11.55	11.96 11.37
$h = 8$	- -	42.59 17.43	30.82 20.20	22.94 17.99	16.26 13.82	13.07 11.91	11.99 11.39
$h = 9$	- -	45.29 15.12	32.35 20.21	24.46 19.02	17.53 14.74	13.84 12.35	12.24 11.56
$h = 10$	- -	48.97 13.97	33.43 20.24	25.31 19.07	17.87 15.14	14.18 12.60	12.38 11.75

Note:- The first entry in each cell is for the original test (S_1 using $N(0,1)$ critical values), the second for the fully modified test (S^* using t_{n-1} critical values).

simulation results as a whole, the reduction in small sample oversizing generated by the corrections becomes clear. The original Diebold-Mariano test empirical sizes range from 27.89-30.00 for $n = 8$, 17.93-20.50 for $n = 16$, and 13.62-15.72 for $n = 32$. The comparable sizes for the modified test range from 13.38-16.42 for $n = 8$, 11.05-14.65 for $n = 16$, and 10.49-12.54 for $n = 32$.

Simulation results for 3- and 4-steps-ahead forecasts are given in tables 2.7 and 2.8 respectively (the case of no autocorrelation is now the only one examined due to the inference of test robustness to this property). As would be expected, the results again display invariance to the varying forecast error conditions of distribution and contemporaneous correlation. The tests, as with the smaller step-ahead forecasts, are asymptotically correctly sized, but oversized in small samples. The degree of this oversizing for the Diebold-Mariano test is now immense, with sizes for 3-steps-ahead prediction being approximately 38% for $n = 8$, 26% for $n = 16$, and 18% for $n = 32$, and for the 4-steps-ahead case approximately 44% for $n = 8$, 31% for $n = 16$, and 21.5% for $n = 32$. The size improvements gained by use of the new modified test are considerable, with the corresponding 3-steps-ahead approximate average sizes being 18% for $n = 8$, 17.5% for $n = 16$, and 14% for $n = 32$, and for 4-steps-ahead forecast approximately 16% for $n = 8$, 19% for $n = 16$, and 16% for $n = 32$.

This inference of test size improvement is consistent over all forecast horizons. Table 2.9 provides simulation results for 5- through 10-steps-ahead forecasts, now considering solely the representative case of normal, non-contemporaneously correlated, non-autocorrelated forecast errors. With these longer step-ahead

forecasts, the small sample oversizing of S_1 magnifies; indeed the test becomes almost unusable for small samples as the forecast horizon grows. The modified test is not immune to this feature and generally the size worsens with longer forecast horizons, but to a much smaller degree, with the test remaining significantly more workable, notably in small samples. Examination of these longer forecast horizons also highlights the result that the majority of the size correction comes through the finite sample bias correction to the estimated variance of the sample mean, with a lesser role played by the use of the Student's t critical values.

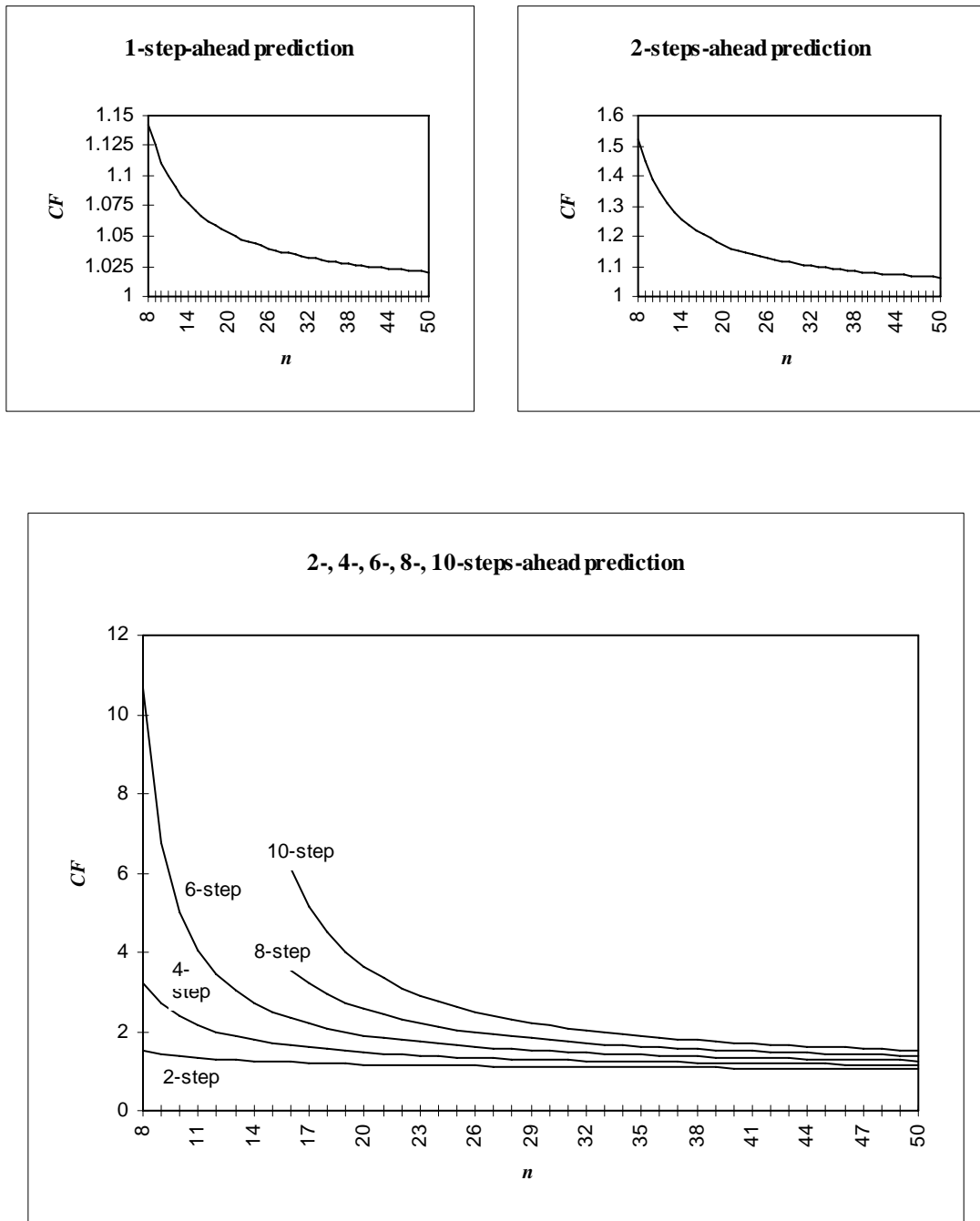
One other feature of the S^* simulation results is that of the pattern of the test size as the sample size increases. For example, the normal errors $\rho = 0$ row for 4-steps-ahead forecasting shows a test size of 16.25 for $n = 8$, 19.83 for $n = 16$, 16.14 for $n = 32$, and so on. The size initially worsens with increased n before improving again and asymptotically tending to 10. This variability is due to the nature of the bias correction which is non-linear. Plots of the correction factor:

$$CF = n[n + 1 - 2h + n^{-1}h(h - 1)]^{-1}$$

are given in figure 2.1 for a number of step-ahead forecasts, and these illustrate how the adjustment varies with n for each h -steps-ahead forecast. The correction factor is a convex function of n and so the correction when $n = 8$ is proportionally larger than when $n = 16$. As n rises, this feature becomes less marked with the curve becoming flatter, and the observed results are thus explained. One other characteristic of the correction factor can be noted from the plot containing a representative selection of

Figure 2.1

Plots of S^* correction factor, CF



different step-ahead forecasts on the same axis. It can be seen that for low forecast horizons, the factor is relatively small, with significant increases (especially in small samples) for the longer horizons.

Finally, table 2.10 summarises the simulation results for the original and fully modified tests for the representative case of normal, non-contemporaneously correlated, non-autocorrelated forecast errors, the benefits of the corrected test being clearly shown. The most apparent feature of all the test size results is that the two modifications to the Diebold-Mariano proposed test succeed in their purpose of improving the small sample sizes. Altogether, the corrected test statistic (S^* using t_{n-1} critical values) exhibits significant gains over the original Diebold-Mariano test. The new test is not completely correctly sized in the smaller sample sizes, but valuable improvements are made consistently over all the forecast horizons. The modified test is therefore preferable to Diebold & Mariano's proposed test, and has considerable value in application.

2.4 Morgan-Granger-Newbold Approach Tests

The discussion in section 2.2 highlighted issues for further study concerning the Morgan-Granger-Newbold test of predictive accuracy. Now it was noted that analysis of this test is only worthwhile if it is superior to other tests in the situations where it is correctly sized. Comparing *MGN* to the new modified Diebold-Mariano test for 1-step-ahead forecasts where the forecast errors are normal, it can be seen by examination of tables 2.2 and 2.3 that both tests are correctly sized (S^* is slightly undersized in the smallest samples). Evaluation of the value of *MGN* relative to S^* therefore relies on power considerations which are examined below. If *MGN* is found to be preferable to S^* on grounds of power, analysis of its non-normal behaviour is necessary. Given the results of the following sub-section, this area is examined and modifications to *MGN* are considered to ascertain whether a test exists which exhibits both the power advantages of *MGN* and robustness to non-normal errors.

2.4.1 Power Comparisons

In examining the power of a test, Monte Carlo simulation can again be used, now with the alternative hypothesis simulated. Given the null of equal forecast accuracy and a quadratic loss differential, H_0 is equivalent to equality of the forecast error variances, given the assumption of zero means. Construction of the alternative can hence be performed by the following transformation:

$$\text{desired alternative: } V(e_{1t}) = 1, \quad V(e_{2t}) = p \quad (p > 1)$$

$$\text{transformation: } \begin{bmatrix} e_{1t}^* \\ e_{2t}^* \end{bmatrix} = \begin{bmatrix} e_{1t} \\ p^{1/2} e_{2t} \end{bmatrix}$$

where e_{1t}, e_{2t} are the simulated forecast errors as in section 2.3, and p is a constant chosen for each sample size to provide reasonable power comparisons.

Simulation is then performed in the same way as before for 1-step-ahead forecasting (with 10,000 replications), and the powers of the two tests MGN and S^* are evaluated for normal errors, where both tests have correct size. Here and throughout this thesis non-size-adjusted powers are used, as size-adjusted powers cannot typically be realised in practice, partly because empirical sizes under the null hypothesis may depend on nuisance parameters.

In addition to the MGN and S^* tests, the simple F -test of predictive accuracy is also considered. The F -test assumes quadratic loss and thus a null of forecast error variance equality, and applies a simple variance ratio test:

$$F = \frac{\hat{V}(e_{1t}^*)}{\hat{V}(e_{2t}^*)} = \frac{\sum e_{1t}^{*2}}{\sum e_{2t}^{*2}} \sim F_{n,n} \quad (2.17)$$

This test involves the same assumptions as MGN , but also requires the forecast errors to have no contemporaneous correlation. For this reason, power is only calculated for the $\rho = 0$ case. The F -test is highly restrictive and impractical for application, but serves as a useful benchmark in the examination of other tests' power. The results are given in table 2.11 (this table also contains power results for

Table 2.11

Power comparisons for the F -test, the original and second modified Morgan-Granger-Newbold tests, the modified Diebold-Mariano test and the rank correlation test at the nominal 10% level ($h = 1$)

Normal		$n = 8$ $p = 3$	$n = 16$ $p = 2$	$n = 32$ $p = 1.5$	$n = 64$ $p = 1.375$	$n = 128$ $p = 1.25$	$n = 256$ $p = 1.1875$	$n = 512$ $p = 1.125$
$\rho = 0.0$	F	58.36	52.94	43.61	49.92	48.93	54.15	52.32
	MGN	42.84	38.70	30.65	35.47	34.91	40.85	38.06
	S^*	27.14	33.05	28.59	34.35	34.85	40.55	37.98
	r_s	28.71	33.18	27.01	32.74	32.60	37.83	35.70
	MGN_2^*	11.20	27.22	26.56	33.45	34.29	40.31	37.87
$\rho = 0.5$	MGN	50.88	46.35	35.81	43.20	42.71	48.54	46.10
	S^*	32.16	39.21	33.91	41.94	42.05	48.04	45.98
	r_s	34.29	39.15	32.57	39.29	40.34	45.37	43.24
	MGN_2^*	13.59	32.69	31.65	41.11	41.58	47.81	45.90
$\rho = 0.9$	MGN	89.63	89.19	81.19	87.87	88.87	92.76	91.93
	S^*	59.87	80.14	77.36	86.67	88.26	92.52	92.09
	r_s	71.03	81.25	74.98	84.30	86.13	90.30	89.58
	MGN_2^*	30.64	73.78	75.00	86.24	88.05	92.47	92.07

DM t_6		$n = 8$ $p = 3$	$n = 16$ $p = 2$	$n = 32$ $p = 1.5$	$n = 64$ $p = 1.375$	$n = 128$ $p = 1.25$	$n = 256$ $p = 1.1875$	$n = 512$ $p = 1.125$
$\rho = 0.0$	S^*	19.07	23.03	20.30	24.18	23.45	24.95	23.06
	r_s	27.56	30.85	26.51	30.92	31.01	34.36	32.69
	MGN_2^*	7.62	18.35	18.84	23.29	22.95	24.75	23.00
$\rho = 0.5$	S^*	25.66	31.12	27.00	31.40	29.72	32.30	30.45
	r_s	33.57	37.93	32.37	37.85	37.24	41.50	40.47
	MGN_2^*	10.19	25.29	24.92	30.45	29.17	32.04	30.41
$\rho = 0.9$	S^*	53.31	76.96	77.16	86.78	86.90	90.89	89.36
	r_s	71.57	83.46	78.96	87.23	88.17	92.66	91.64
	MGN_2^*	25.91	69.67	75.19	86.18	86.67	90.77	89.27

Biv. t_6		$n = 8$ $p = 3$	$n = 16$ $p = 2$	$n = 32$ $p = 1.5$	$n = 64$ $p = 1.375$	$n = 128$ $p = 1.25$	$n = 256$ $p = 1.1875$	$n = 512$ $p = 1.125$
$\rho = 0.0$	S^*	21.03	24.83	22.80	26.32	25.87	28.49	26.39
	r_s	28.69	32.31	28.10	31.64	32.50	36.25	34.64
	MGN_2^*	7.83	19.80	20.65	25.24	25.49	28.26	26.33
$\rho = 0.5$	S^*	24.90	29.38	26.72	31.07	31.30	33.85	31.88
	r_s	33.93	37.38	32.68	37.83	38.82	43.02	41.15
	MGN_2^*	9.52	23.84	24.65	30.14	30.91	33.58	31.73
$\rho = 0.9$	S^*	48.07	65.58	62.65	71.68	71.58	76.98	73.77
	r_s	69.69	78.37	72.45	81.96	82.86	88.32	87.15
	MGN_2^*	22.36	57.53	59.98	70.59	71.11	76.81	73.67

S^* under non-normality, and two other tests r_s and MGN_2^* ; these entries relate to later analysis conducted in sub-section 2.4.3).

Examination of the results for normal errors shows that the two tests of concern have roughly the same power for large sample sizes, but for samples of 64 observations and less, MGN is superior to S^* in terms of power, the difference being more significant as the sample becomes smaller. The F -test has consistently greater power than both these tests as expected and acts as a useful point of reference. It can also be seen that the power rises significantly with the degree of contemporaneous correlation, ρ . The reason for this is that as ρ increases, the forecast errors become more and more similar until $\rho = 1$, when they are perfectly correlated. As this limit approaches, it becomes easier to determine variance differences, as in the extreme (when $\rho = 1$) the errors are directly proportional to each other with one variance always being an exact multiple of the other.

It can hence be concluded from these simulation results that the Morgan-Granger-Newbold test has valuable power gains over the modified Diebold-Mariano test in small samples. Given the importance of small samples in the analysis of tests of predictive accuracy, the MGN test appears to be useful and have certain advantages over other tests under the circumstances considered. It is now necessary to consider the test's lack of robustness to even mild non-normality, and then examine potential corrections.

2.4.2 MGN Behaviour Under Error Non-normality

The unusual behaviour of *MGN* when the forecast errors are non-normal now demands explanation. Firstly, it is useful to note that the correlation test of which *MGN* is comprised is identical to the *t*-test on the null $\beta=0$ in the following regression:

$$y_t = \beta x_t + \varepsilon_t$$

$$\text{where } y_t = e_{1t} + e_{2t}$$

$$x_t = e_{1t} - e_{2t}$$

$$MGN = \frac{\hat{\beta}}{\sqrt{\hat{\sigma}^2 / \sum x_t^2}} \sim t_{n-1} \quad \text{under } H_0 \quad (2.18)$$

$$\text{where } \hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2}$$

$$\hat{\sigma}^2 = (n-1)^{-1} \sum (y_t - \hat{\beta} x_t)^2$$

The asymptotic distribution of $\hat{\beta}$ under the null is known and can be written as follows (e.g. White, 1984):

$$D^{-1/2} n^{1/2} (\hat{\beta} - \beta) \xrightarrow{d} N(0,1) \quad (2.19)$$

$$\text{where } D = M^{-2} Q$$

$$M = E(x_t^2)$$

$$Q = V(n^{-1/2} \sum x_t \varepsilon_t)$$

Now the distribution of the *MGN* test statistic will be:

$$\hat{D}^{-1/2}n^{1/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \text{plim}[\hat{D}^{-1}D]) \quad (2.20)$$

$$\text{where } \hat{D} = \hat{M}^{-2}\hat{Q}$$

$$\hat{M} = n^{-1}\Sigma x_t^2$$

$$\hat{Q} = \hat{\sigma}^2 n^{-1}\Sigma x_t^2$$

and will consequently only be consistent if \hat{D} consistently estimates D . Examining the constituent parts of \hat{D} , it is clear that \hat{M} is consistent for M :

$$\begin{aligned} \hat{M} = n^{-1}\Sigma x_t^2 &\xrightarrow{p} E(x_t^2) = M \\ \text{i.e. } \hat{M} &\xrightarrow{p} M \end{aligned} \quad (2.21)$$

and so (2.20) can now be written as:

$$\hat{D}^{-1/2}n^{1/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \text{plim}[\hat{Q}^{-1}Q]) \quad (2.22)$$

The analysis now proceeds to examine this estimate \hat{Q} implicit in the *MGN* test statistic, firstly for normal forecast errors and then for non-normal errors.

2.4.2a Normal Forecast Errors

The forecast errors of interest - 1-step-ahead normal errors - can be decomposed into linear functions of standard normal random variables, which allows further analysis of Q and \hat{Q} . Referring to the sub-section 2.3.2 simulation theory:

$$e_{1t} = u_{1t} \quad (2.23)$$

$$e_{2t} = \rho u_{1t} + (1 - \rho^2)^{1/2} u_{2t} \quad (2.24)$$

$$\text{where } u_{1t}, u_{2t} \sim \text{IN}(0,1)$$

An expression for Q can now be found:

$$\begin{aligned}
Q &= V(x_t \varepsilon_t) \quad \text{assuming } x_t \varepsilon_t, x_s \varepsilon_s \text{ uncorrelated for } s \neq t \\
&= E[(x_t \varepsilon_t)^2] \quad \text{as } E(x_t \varepsilon_t) = 0 \\
&= E\{[(e_{1t} - e_{2t})(e_{1t} + e_{2t})]^2\} \\
&= E(e_{1t}^4) + E(e_{2t}^4) - 2E(e_{1t}^2 e_{2t}^2)
\end{aligned}$$

$$\begin{aligned}
\text{now } E(e_{1t}^4) &= E(u_{1t}^4) \\
&= 3
\end{aligned}$$

$$\begin{aligned}
E(e_{2t}^4) &= E[\rho^4 u_{1t}^4 + (1 - \rho^2)^2 u_{2t}^4 + 4\rho^2(1 - \rho^2)u_{1t}^2 u_{2t}^2 \\
&\quad + 2\rho^2(1 - \rho^2)u_{1t}^2 u_{2t}^2 + 4\rho^3(1 - \rho^2)^{1/2} u_{1t}^3 u_{2t} \\
&\quad + 4\rho(1 - \rho^2)^{3/2} u_{1t} u_{2t}^3] \\
&= 3
\end{aligned}$$

$$\begin{aligned}
E(e_{1t}^2 e_{2t}^2) &= E[\rho^2 u_{1t}^4 + (1 - \rho^2)u_{1t}^2 u_{2t}^2 + 2\rho(1 - \rho^2)^{1/2} u_{1t}^3 u_{2t}] \\
&= 1 + 2\rho^2
\end{aligned}$$

$$\text{so } Q = 4(1 - \rho^2) \tag{2.25}$$

Similarly, the probability limit of the estimator \hat{Q} can be found under the null:

$$\begin{aligned}
\hat{Q} &\xrightarrow{p} V(\varepsilon_t)E(x_t^2) \\
&= E[(e_{1t} + e_{2t})^2]E[(e_{1t} - e_{2t})^2] \quad \text{under } H_0 \\
&= [E(e_{1t}^2) + E(e_{2t}^2) + 2E(e_{1t}e_{2t})] \\
&\quad \times [E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})]
\end{aligned}$$

$$\begin{aligned}\text{now } E(e_{1t}^2) &= E(u_{1t}^2) \\ &= 1\end{aligned}$$

$$\begin{aligned}E(e_{2t}^2) &= E[\rho^2 u_{1t}^2 + (1 - \rho^2) u_{2t}^2 + 2\rho(1 - \rho^2)^{1/2} u_{1t} u_{2t}] \\ &= 1\end{aligned}$$

$$\begin{aligned}E(e_{1t} e_{2t}) &= E[\rho u_{1t}^2 + (1 - \rho^2) u_{1t} u_{2t}] \\ &= \rho\end{aligned}$$

$$\text{so } \hat{Q} \xrightarrow{p} 4(1 - \rho^2) \quad (2.26)$$

Comparison of results (2.25) and (2.26) reveals that in this case of normal errors, \hat{Q} consistently estimates Q , thus $\hat{D} \xrightarrow{p} D$ and the following results:

$$MGN \xrightarrow{d} N(0,1) \quad (2.27)$$

This confirms the simulation observation that the *MGN* test is correctly sized for normal errors with no problems exhibited under such error properties.

2.4.2b Non-Normal Forecast Errors

The above analysis can be repeated for non-normal forecast errors, with the two methods of error generation considered separately.

Diebold-Mariano Errors

Firstly, the Diebold-Mariano approach has errors which can be decomposed as in (2.23) and (2.24):

$$e_{1t} = \kappa_{1t}^{-1/2} z_{1t} \quad (2.28)$$

$$e_{2t} = \rho \kappa_{1t}^{-1/2} z_{1t} + (1 - \rho^2)^{1/2} \kappa_{2t}^{-1/2} z_{2t} \quad (2.29)$$

$$\text{where } \kappa_{it} = \chi_{it,6}^2 / 6; \quad i = 1, 2$$

$$z_{it} \sim \text{IN}(0, 1); \quad i = 1, 2$$

$$\chi_{it,6}^2 / 6 \text{ is independent of } z_{it}$$

The quantity Q in this case can then be found using the methodology employed for the normal errors analysis:

$$Q = E(e_{1t}^4) + E(e_{2t}^4) - 2E(e_{1t}^2 e_{2t}^2)$$

$$\text{now } E(e_{1t}^4) = 3A_2 \quad \text{where } A_k = E(\kappa_{it}^{-k})$$

$$E(e_{2t}^4) = 3[\rho^4 + (1 - \rho^2)^2]A_2 + 6\rho^2(1 - \rho^2)A_1^2$$

$$E(e_{1t}^2 e_{2t}^2) = 3\rho^2 A_2 + (1 - \rho^2)A_1^2$$

$$\text{so } Q = 6(1 - \rho^2)^2 A_2 + 2(1 - \rho^2)(3\rho^2 - 1)A_1^2 \quad (2.30)$$

The limiting value of the estimator \hat{Q} under H_0 is now:

$$\begin{aligned} \hat{Q} &\xrightarrow{p} [E(e_{1t}^2) + E(e_{2t}^2) + 2E(e_{1t}e_{2t})] \\ &\quad \times [E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})] \end{aligned}$$

$$\text{now } E(e_{1t}^2) = A_1$$

$$E(e_{2t}^2) = A_1$$

$$E(e_{1t}e_{2t}) = \rho A_1$$

$$\text{so } \hat{Q} \xrightarrow{p} 4(1-\rho^2)A_1^2 \quad (2.31)$$

Under a situation of forecast error non-normality then, \hat{Q} does not consistently estimate Q and the resulting null distribution for the test statistic is:

$$MGN \xrightarrow{d} N(0, \text{plim}[\hat{Q}^{-1}Q]) \quad (2.32)$$

$$\text{where } \text{plim}[\hat{Q}^{-1}Q] = \frac{3(1-\rho^2)A_2 + (3\rho^2 - 1)A_1^2}{2A_1^2}$$

Now for t_6 errors, the A_k quantities are:

$$A_k = \frac{6^k \Gamma(3-k)}{2^k \Gamma(3)}$$

$$\text{i.e. } A_1 = 3/2, \quad A_2 = 9/2$$

which gives a precise result for the MGN test under the null:

$$MGN \xrightarrow{d} N(0, \frac{5}{2} - \frac{3}{2}\rho^2) \quad (2.33)$$

This is clearly not standard normal (unless $\rho = 1$) and so problems with the test size will be present. The degree of the problem can be found by evaluating the limiting size of the test statistic for a given value of ρ , using numerical integration:

$$\rho = 0: \quad MGN \xrightarrow{d} N(0, 2.5)$$

2-sided asymptotic test size at the nominal 10% level

$$= 2 \int_{1.64485}^{\infty} (5\pi)^{-1/2} e^{-x^2/5} .dx = 29.82\%$$

$$\rho = 0.5: MGN \xrightarrow{d} N(0, 2.125)$$

2-sided asymptotic test size at the nominal 10% level

$$= 2 \int_{1.64485}^{\infty} (4.25\pi)^{-1/2} e^{-x^2/4.25} .dx = 25.92\%$$

$$\rho = 0.9: MGN \xrightarrow{d} N(0, 1.285)$$

2-sided asymptotic test size at the nominal 10% level

$$= 2 \int_{1.64485}^{\infty} (2.57\pi)^{-1/2} e^{-x^2/2.57} .dx = 14.68\%$$

These theoretical sizes are confirmed by simulation with a sample size of $n = 10,000$. The empirical sizes are found to be 29.00% for $\rho = 0$, 25.58% for $\rho = 0.5$, and 15.08% for $\rho = 0.9$.

Bivariate Errors

This whole process can now be repeated for the second method of generating non-normal forecast errors - the bivariate Student's t approach:

$$e_{1t} = \kappa_t^{-1/2} z_{1t} \tag{2.34}$$

$$e_{2t} = \kappa_t^{-1/2} [\rho z_{1t} + (1 - \rho^2)^{1/2} z_{2t}] \tag{2.35}$$

$$\text{where } \kappa_t = \chi_{t,6}^2 / 6; \quad i = 1, 2$$

$$z_{it} \sim \text{IN}(0, 1); \quad i = 1, 2$$

$$\chi_{t,6}^2 / 6 \text{ is independent of } z_{it}$$

As before, an expression for the term Q must be derived:

$$Q = E(e_{1t}^4) + E(e_{2t}^4) - 2E(e_{1t}^2 e_{2t}^2)$$

$$\text{now } E(e_{1t}^4) = 3A_2 \quad \text{where } A_k = E(\kappa_t^{-k})$$

$$E(e_{2t}^4) = 3A_2$$

$$E(e_{1t}^2 e_{2t}^2) = (1 + 2\rho^2)A_2$$

$$\text{so } Q = 4(1 - \rho^2)A_2 \quad (2.36)$$

The limiting null value of \hat{Q} is now:

$$\begin{aligned} \hat{Q} &\xrightarrow{p} [E(e_{1t}^2) + E(e_{2t}^2) + 2E(e_{1t}e_{2t})] \\ &\quad \times [E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})] \end{aligned}$$

$$\text{now } E(e_{1t}^2) = A_1$$

$$E(e_{2t}^2) = A_1$$

$$E(e_{1t}e_{2t}) = \rho A_1$$

$$\text{so } \hat{Q} \xrightarrow{p} 4(1 - \rho^2)A_1^2 \quad (2.37)$$

and the limiting distribution for MGN when these bivariate Student's t persist becomes:

$$MGN \xrightarrow{d} N(0, A_1^{-2} A_2) \quad (2.38)$$

Substituting in values for A_1 and A_2 for the bivariate t_6 distribution gives:

$$MGN \xrightarrow{d} N(0, 2) \quad \forall \rho \quad (2.39)$$

and the test exhibits size problems again. Numerical integration shows the limiting test size (which is the same for all values of ρ) to be 24.48% (for a 2-sided test at the nominal 10% level), and simulation of a sample size of $n = 10,000$ confirms this theory with results 22.97% for $\rho = 0$, 23.12% for $\rho = 0.5$, and 22.95% for $\rho = 0.9$.

The above analysis clearly shows in detail the problems associated with the Morgan-Granger-Newbold test when the forecast errors are non-normal. The asymptotic sizes of the test under different error specifications are derived and found to be severely oversized. This behaviour clearly results from inconsistent estimation of the variance of $\hat{\beta}$ in the *MGN* regression. The reason for this failure of *MGN* when departures from normality occur is that the regression errors are then conditionally heteroscedastic. The classical regression assumption of $E(\varepsilon_t^2|x_t) = \sigma^2$ is violated, and although $(e_{1t} + e_{2t})$ and $(e_{1t} - e_{2t})$ are uncorrelated under the null, they are not in general independent, with *MGN* over-rejecting the null of zero correlation when no correlation exists due to this element of dependency.

The problems associated with the *MGN* test under conditions of forecast error non-normality have implications for the work conducted by Ashley, Granger & Schmalensee (1980). Ashley *et al.* studied the issue of causality between short run variations in aggregate advertising and the level of consumption spending. Using the Granger (1969) definition of causality between two series X_t, Y_t , the matter can be examined by constructing two forecasts of X_{T+1} (where T is the current time period). One forecast uses all the available information in existence, the other uses the same information with the exception of current and lagged values of Y_t . If the

former forecast is superior to the latter, Y_t has unique information about X_t and thus Y_t ‘causes’ X_t . Ashley *et al.* apply this theory to the aforementioned issue and employ a form of the *MGN* test to compare the two constructed forecasts. Their results provide evidence for uni-directional causality, with fluctuations in aggregate consumption causing fluctuations in aggregate advertising. However, the study does not consider the possibility of forecast error non-normality. The previous analysis has shown that with even mildly non-normal forecast errors, there is a large tendency to over-reject a valid null hypothesis, and so doubt is thrown over the reliability of the Ashley *et al.* inferences. The null of consumption fluctuations not causing advertising fluctuations is rejected at just less than the 9.2% significance level, but this could be spurious if the forecast errors are non-normal. The conclusions of the Ashley *et al.* analysis are therefore valid under a situation of normality, but due to the problems of *MGN* become more dubious when the prospect of forecast error non-normality is taken into account.

Having analysed the *MGN* non-normality problem and discovered the reason for its lack of robustness, it is now appropriate to examine alternative procedures which might capture the benefits of *MGN* (namely its power advantages) whilst being robust in application to non-normal forecast errors.

2.4.3 Robust *MGN*-Type Tests

2.4.3a Modified *MGN* Tests

The analysis in sub-section 2.4.2 concluded that the undesirable non-normal behaviour of *MGN* resulted from an inconsistent estimate of Q in (2.19) as a consequence of conditional heteroscedasticity in the regression errors. This suggests correction is possible by use of a new consistent estimator using a White (1980) approach.

Introducing the reasonable assumption of $x_t \varepsilon_t, x_s \varepsilon_s$ uncorrelated for $s \neq t$, as used in sub-section 2.4.2, allows the following to be written:

$$Q = E(x_t^2 \varepsilon_t^2) \quad (2.40)$$

Given this expression for Q , it can be seen that a consistent estimator would be $n^{-1} \sum x_t^2 \varepsilon_t^2$, but the problem is that ε_t is unobserved. The intuitive estimator to use is therefore the following:

$$\hat{Q}_{m1} = n^{-1} \sum x_t^2 \hat{\varepsilon}_t^2 \quad (2.41)$$

where $\hat{\varepsilon}_t$ are the least squares regression residuals. It is possible to show that this is consistent for Q (e.g. White, 1984):

$$\begin{aligned} \hat{Q}_{m1} - Q &= n^{-1} \sum x_t^2 \hat{\varepsilon}_t^2 - E(x_t^2 \varepsilon_t^2) \\ &= n^{-1} \sum x_t^2 \varepsilon_t^2 - E(x_t^2 \varepsilon_t^2) - 2n^{-1} \sum x_t^2 (\hat{\beta} - \beta) x_t \varepsilon_t \\ &\quad + n^{-1} \sum x_t^2 (\hat{\beta} - \beta)^2 x_t^2 \end{aligned}$$

$$\begin{aligned}
&\text{now } n^{-1}\sum x_t^2 \varepsilon_t^2 - E(x_t^2 \varepsilon_t^2) \xrightarrow{p} 0 \\
&\hat{\beta} - \beta \xrightarrow{p} 0 \\
&n^{-1}\sum x_t^4, n^{-1}\sum x_t^3 \varepsilon_t \text{ are } O_p(1) \\
&\text{so } \hat{Q}_{m1} - Q \xrightarrow{p} 0
\end{aligned} \tag{2.42}$$

This then leads to the following consistent estimator of D :

$$\hat{D}_{m1} = \frac{n \sum x_t^2 \hat{\varepsilon}_t^2}{(\sum x_t^2)^2} \xrightarrow{p} D \tag{2.43}$$

Substitution of this into the *MGN* test, replacing the implicit element \hat{D} generates a new modified *MGN* test, which is valid under both the null and alternative hypotheses and is only reliant on the one assumption of $x_t \varepsilon_t$ zero correlation through time:

$$MGN_1^* = \frac{\hat{\beta}}{\sqrt{\sum x_t^2 \hat{\varepsilon}_t^2 / (\sum x_t^2)^2}} \tag{2.44}$$

This test statistic is normally distributed asymptotically, but is of a typical t -test form and so critical values from the Student's t distribution with $n - 1$ degrees of freedom are used in finite samples.

The asymptotic test size of MGN_1^* should now be correct. Confirmation of this and analysis of the test's small sample properties are possible by way of simulation in the usual way. The results can be seen in table 2.12, which also gives the *MGN* empirical sizes under all error properties to allow comparisons to be made (results of

Table 2.12

Empirical sizes for the original and modified Morgan-Granger-Newbold tests at the nominal 10% level ($h = 1$)

Normal		$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	<i>MGN</i>	10.18	10.30	10.02	10.11	9.72	10.35	10.67
	<i>MGN</i> ₁ [*]	19.91	15.98	12.78	11.79	10.59	10.85	10.79
	<i>MGN</i> ₁ [*]	19.87	16.05	12.91	11.71	10.60	10.82	10.94
	<i>MGN</i> ₂ [*]	2.84	7.38	8.68	9.50	9.66	10.28	10.50
$\rho = 0.5$	<i>MGN</i>	10.04	9.85	10.33	10.30	10.18	10.62	10.40
	<i>MGN</i> ₁ [*]	19.80	15.39	12.96	12.03	11.02	11.16	10.71
	<i>MGN</i> ₁ [*]	20.03	15.56	13.14	11.97	10.96	11.17	10.71
	<i>MGN</i> ₂ [*]	3.06	7.08	9.73	9.89	9.92	10.57	10.41
$\rho = 0.9$	<i>MGN</i>	10.09	9.86	10.18	10.43	10.00	10.45	10.05
	<i>MGN</i> ₁ [*]	20.36	15.48	12.89	11.99	10.71	11.11	10.63
	<i>MGN</i> ₁ [*]	19.79	15.14	12.74	12.08	10.83	11.15	10.58
	<i>MGN</i> ₂ [*]	3.05	6.78	8.65	9.91	9.71	10.63	10.31

DM t_6		$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	<i>MGN</i>	17.92	20.45	22.59	24.83	25.99	26.07	26.76
	<i>MGN</i> ₁ [*]	26.26	21.16	17.98	15.68	12.86	12.01	10.69
	<i>MGN</i> ₁ [*]	25.99	21.31	17.84	15.49	13.05	11.88	10.86
	<i>MGN</i> ₂ [*]	2.02	5.96	8.07	9.26	9.11	9.69	9.45
$\rho = 0.5$	<i>MGN</i>	16.21	18.54	19.80	22.05	22.38	22.62	23.67
	<i>MGN</i> ₁ [*]	25.38	20.77	17.51	14.94	12.51	11.42	10.59
	<i>MGN</i> ₁ [*]	24.90	20.83	16.90	14.66	12.60	11.63	10.56
	<i>MGN</i> ₂ [*]	2.17	6.42	8.54	9.04	9.24	9.36	9.22
$\rho = 0.9$	<i>MGN</i>	11.83	12.75	12.61	13.65	13.57	13.72	13.63
	<i>MGN</i> ₁ [*]	22.03	18.49	15.08	13.26	12.00	10.79	10.28
	<i>MGN</i> ₁ [*]	21.85	18.34	15.16	13.35	11.88	10.98	10.31
	<i>MGN</i> ₂ [*]	2.22	6.84	8.41	9.30	9.49	9.27	9.55

Biv. t_6		$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0.0$	<i>MGN</i>	15.51	16.97	18.90	19.96	21.22	21.27	22.07
	<i>MGN</i> ₁ [*]	24.82	20.28	17.08	14.66	12.85	11.53	10.86
	<i>MGN</i> ₁ [*]	24.90	20.00	16.65	14.38	12.90	11.53	11.03
	<i>MGN</i> ₂ [*]	2.03	6.17	7.98	8.70	9.35	9.40	9.60
$\rho = 0.5$	<i>MGN</i>	15.42	17.02	19.06	20.09	20.75	21.29	22.37
	<i>MGN</i> ₁ [*]	24.84	19.95	17.21	14.89	12.53	11.40	10.74
	<i>MGN</i> ₁ [*]	24.30	20.15	17.35	14.78	12.61	11.38	10.63
	<i>MGN</i> ₂ [*]	2.30	5.69	7.70	8.95	9.16	9.47	9.37
$\rho = 0.9$	<i>MGN</i>	15.13	17.04	19.00	19.69	20.57	21.36	22.85
	<i>MGN</i> ₁ [*]	24.62	20.22	17.28	14.66	12.70	11.77	11.19
	<i>MGN</i> ₁ [*]	24.29	20.23	17.19	14.46	12.62	11.90	11.06
	<i>MGN</i> ₂ [*]	2.41	6.01	7.89	8.68	9.51	9.77	9.87

Note:- The first entry for the *MGN*₁^{*} test in each cell is for case 1 and the second for case 2.

another test, MGN_2^* , which shall be introduced shortly, are also included). One further point to note is that the modified test MGN_1^* is not symmetric and so simulation results are given for both cases (see table footnote):

$$y_t = \beta x_t + \varepsilon_t \quad \text{case 1: } y_t = e_{1t} + e_{2t}, x_t = e_{1t} - e_{2t}$$

$$\text{case 2: } y_t = e_{1t} - e_{2t}, x_t = e_{1t} + e_{2t}$$

With regard to the results, it can be seen that the modification is successful in correcting the asymptotic size of the test. However, significant oversizing in small samples is also apparent for normal and non-normal errors. This leads to the conclusion that the modifications to the Morgan-Granger-Newbold test succeed in generating a robust test asymptotically, but one that is impractical in application due to this small sample behaviour. Further to this extent, the cases in which MGN is preferable to S^* are those involving the smaller samples; the cost of correction for MGN 's erroneous non-normal behaviour leads to oversizing in precisely these cases, leaving no potential for MGN_1^* to be useful in practice.

Further analysis of this modified MGN test gives insight into its finite sample behaviour. The estimator \hat{Q}_{m1} can be written as:

$$\begin{aligned} \hat{Q}_{m1} &= n^{-1} \sum x_t^2 \varepsilon_t^2 - 2n^{-1} \sum x_t^2 (\hat{\beta} - \beta) x_t \varepsilon_t + n^{-1} \sum x_t^2 (\hat{\beta} - \beta)^2 x_t^2 \\ &= O_p(1) - O_p(n^{-1/2}) O_p(1) + O_p(n^{-1}) O_p(1) \end{aligned}$$

In the limit, $\hat{Q}_{m1} \xrightarrow{p} Q$ as noted in (2.42), but the latter two terms in the above decomposition, especially the second term, will converge to zero slowly. It is the

behaviour of these latter terms in small and moderate samples that create the poor test size performance observed in table 2.12.

A second modified *MGN* test can now be considered. This time a consistent estimator for Q is found by employing the fact that $\varepsilon_t = y_t$ under the null. The following estimator then results:

$$\hat{Q}_{m2} = n^{-1} \sum x_t^2 y_t^2 \quad (2.45)$$

This estimator is consistent for Q under the null but not under the alternative:

$$\begin{aligned} \text{under } H_0: \hat{Q}_{m2} &= n^{-1} \sum x_t^2 \varepsilon_t^2 \\ &\xrightarrow{p} Q \end{aligned} \quad (2.46)$$

$$\begin{aligned} \text{under } H_1: \hat{Q}_{m2} &= n^{-1} \sum x_t^2 (\beta x_t + \varepsilon_t)^2 \\ &= n^{-1} \sum x_t^2 \varepsilon_t^2 + 2n^{-1} \beta \sum x_t^3 \varepsilon_t + n^{-1} \beta^2 \sum x_t^4 \\ &= O_p(1) + O_p(1) + O_p(1) \\ &\xrightarrow{p} Q + \beta c_1 + \beta^2 c_2 \end{aligned} \quad (2.47)$$

where $c_i = \text{constants}$, $i = 1, 2$

Use of \hat{Q}_{m2} permits a consistent estimator of D to be derived:

$$\hat{D}_{m2} = \frac{n \sum x_t^2 y_t^2}{(\sum x_t^2)^2} \xrightarrow{p} D \quad (2.48)$$

and substitution of this as with \hat{D}_{m1} generates a second modified *MGN* test:

$$MGN_2^* = \frac{\hat{\beta}}{\sqrt{\Sigma x_t^2 y_t^2 / (\Sigma x_t^2)^2}} \quad (2.49)$$

This test would be expected to be correctly sized, but at the cost of some reduction in power due to the lack of consistent estimation under the alternative. It is important to note that the test is consistent, i.e. the power tends asymptotically to one, under both the null and the alternative, even though Q is not consistently estimated under H_1 . As with MGN_1^* , use of Student's t_{n-1} critical values is most appropriate in application.

Simulation of this test is performed for conditions identical to those considered for MGN_1^* , with the exception that MGN_2^* is symmetric and so only one case need be examined. The results, also given in table 2.12, show an improvement in size as expected, but in many cases, especially for the small and moderate samples, the test is significantly undersized. However, given that undersizing errs on the side of caution and is more tolerable than oversizing, it is also worth checking the test's power. Simulation results for the power of MGN_2^* are given in table 2.11, and are not favourable. The test has very low power for the smaller samples, in part resulting from the undersizing in these cases, and thus the cost of correcting for the MGN problems is here the removal of the small sample power advantages of MGN over S^* . Even in large samples, the MGN_2^* test does not attain the power of S^* , leaving little value in the application of this second modified test.

2.4.3b Rank Correlation Test

A second thought in the objective to find a test of the *MGN* form which is robust to the distribution of the forecast errors is to use a nonparametric test, the most obvious choice being Spearman's rank correlation test. This test of the no correlation null involves ranking the observations x_t ($t = 1, \dots, n$) and y_t ($t = 1, \dots, n$) in ascending order (generating vectors x_r, y_r), and then calculating the sample correlation of the ranks x_{rt}, y_{rt} . The resulting coefficient becomes the test statistic and is compared with specifically derived critical values:

$$r_s = \frac{\sum x_{rt} y_{rt} - n \bar{x}_r \bar{y}_r}{\sqrt{(\sum x_{rt}^2 - n \bar{x}_r^2)(\sum y_{rt}^2 - n \bar{y}_r^2)}} \quad (2.50)$$

$$= \frac{n^{-1} \sum x_{rt} y_{rt} - [(n+1)/2]^2}{(n^2 - 1)/12} \quad (2.51)$$

If there are no tied ranks, this simplifies to the following expression:

$$r_s = 1 - \frac{6 \sum d_{rt}^2}{n(n^2 - 1)} \quad \text{where } d_{rt} = x_{rt} - y_{rt} \quad (2.52)$$

This nonparametric test will overcome the problem of missizing under non-normality; the question concerning its value relates to the test's power. Simulation is thus performed in the same way as in sub-section 2.4.1 to examine the power of Spearman's rank correlation test relative to *MGN* and S^* ; the same values of p are used within each sample size to allow direct comparisons to be made. The relevant Spearman's rank correlation critical values are used for sample sizes $n = 8, 16, 32, 64$; for the larger samples, the normal approximation is used (e.g. Kendall & Gibbons, 1990):

$$\sqrt{n-1} r_s \sim_{asy.} N(0,1)$$

i.e. r_s is standardised and then compared with the relevant standard normal critical value. The exact critical values for large samples can be simulated, and execution of such experiments confirms the validity of the above approximation.

The rank correlation test simulation results are given in table 2.11 and provide the inference that the test is not as powerful as *MGN* for any sample size when the errors are normal (the non-normal case cannot be compared as *MGN* is oversized under such error properties). The nonparametric test procedure of r_s overcomes the non-normal sizing problem of *MGN*, but to the cost of a reduction in power. With normal errors, this occurs to such an extent that the small sample power advantages of *MGN* over S^* are completely removed. However, r_s does exhibit power gains over S^* under non-normality in all samples. A case exists therefore for the use of the rank correlation test with 1-step-ahead forecasts if heavy-tailed error distributions are strongly expected, due to this non-normal power gain. Generally, however, given that there is little or no advantage when the errors are normal, and that the test has no natural extension to multi-step-ahead forecasting, there remains little motivation for using r_s in more general circumstances when compared to S^* .

2.5 Conclusion

In conclusion, there is great need when examining the predictive accuracy of competing forecasts to have statistical testing procedures which are general in specification and robust to a wide range of forecast error properties. Given the nature of the situations where such tests are applied, test validity in small samples is also crucial. The paper by Diebold & Mariano (1995) examines a number of extant tests, notably one attributed to Morgan, Granger & Newbold, and proposes a new test for the null of equal forecast accuracy which possesses the desirable characteristics of a general specification and robustness to error properties.

This study has sought to analyse two areas of interest arising from the work by Diebold & Mariano. Firstly, it is observed by Monte Carlo simulation that the Diebold & Mariano proposed asymptotic test, S_1 , is heavily oversized in small samples. This problem magnifies as longer forecast horizons are examined.

Following this observation, two modifications are applied to S_1 , the aim being to capture its many advantages whilst improving the small sample sizing to an acceptable level. These corrections involve a modification to the variance estimate which accounts for the finite sample bias inherent in S_1 , and use of critical values from the Student's t distribution with $n-1$ degrees of freedom to follow the usual significance test construction.

The newly derived test, S^* , significantly reduces the small sample oversizing to a much more acceptable level, and does so consistently over all the sample sizes and

forecast horizons. The benefits of S_1 are maintained with S^* also having a very general economic loss function specification, and exhibiting robustness to forecast error distribution, autocorrelation and contemporaneous correlation. The modified test S^* is consequently preferable to S_1 and highly valuable in application.

The second issue of concern refers to the Morgan-Granger-Newbold test. Examination of the test's power relative to S^* shows that in the limited cases where *MGN*'s assumptions are satisfied (1-step-ahead forecasting, normal forecast errors), advantages exist (in terms of power) over S^* in small samples. This is useful due to the interest in small samples, but the main problem of *MGN* is its lack of robustness to even mildly non-normal forecast errors. Examination of this problem reveals the cause to be an inconsistent variance estimate (if the test is viewed as a regression t -test) resulting from conditional heteroscedasticity in the regression errors.

In response to this, three alternative tests are examined with the objective of removing the *MGN* non-normality problem whilst keeping its small sample 1-step-ahead forecasting advantages. The first two tests are parametric and use new consistent estimators of the variance which solve the *MGN* asymptotic oversizing, but again at a cost. The first modified test, MGN_1^* , generates small sample oversizing, thus where *MGN* has its power advantages, the corrections create problems with size. The second modified test, MGN_2^* , exhibits undersizing and low power in small samples, again removing the advantages of *MGN*. The third test considered is a rank correlation test, and this nonparametric approach achieves a correct test size in all samples. With regard to power, r_s exhibits significant power

advantages over S^* in the case of non-normal forecast errors, but in a world of normal errors, the *MGN* power gains are lost by use of this test. The problem of *MGN* lack of robustness can therefore be removed in a number of ways, but in doing so, either the advantages over S^* are lost or other problems are introduced, the exception to this being the superior performance of the rank correlation test to S^* for non-normal errors.

When evaluation is purely concerned with 1-step-ahead forecasts, the tests S^* and r_s are robust and correctly sized, and if non-normal errors are suspected a case exists for using r_s due to its power advantages in such circumstances. However, when analysis is extended to multi-step-ahead forecast evaluation, S^* is the only test which is robust to the autocorrelation now present in the errors, achieving approximately the correct size in most cases. Despite some small sample oversizing which remains at long horizons, the test exhibits robustness to all the examined forecast error properties and constitutes the best approach to equal accuracy testing in general.

Altogether, it is the conclusion of this analysis that the new fully modified test S^* is of great benefit in application, possessing all the desirable characteristics of the asymptotic test proposed by Diebold & Mariano but with the critical advantage of improved behaviour in small samples. The Morgan-Granger-Newbold test can and must be corrected for its characteristics under forecast error non-normality, but these corrections generally remove the possibility of circumstances under which an *MGN*-type test is preferable to S^* . The exception is r_s when non-normal errors are

suspected, but these advantages are restricted to 1-step-ahead evaluation. The proposed test for predictive accuracy, S^* , takes a very general specification, is robust to forecast error distribution, autocorrelation and contemporaneous correlation, and is reasonably sized for all samples and forecast horizons, thus making a significant contribution to the practice of testing for equal forecast accuracy.

Chapter 3

Tests for Forecast Encompassing

3.1 Introduction

Following the comparison of two competing forecasts of the same quantity, it is not necessarily optimal to choose to use just one of the forecasts in isolation. Both forecasts may contain valuable information concerning future outcomes, and it is frequently the case that a combined forecast formed from the individual ones is superior in predictive ability to either of the two available forecasts alone. This notion of the combination of forecasts has given rise to a large literature, the seminal work being done by Bates & Granger (1969); see also Newbold & Granger (1974), Granger & Newbold (1986), Clemen (1989) and Granger (1989).

Given two rival forecasts, it is useful to formulate some procedure for testing whether a preferred forecast is so superior to its competitor that combination of the two forecasts will not lead to an improved predictor. Such a procedure is developed by Nelson (1972) and formalised by Granger & Newbold (1973) as a test for conditional efficiency. One forecast is said to be conditionally efficient with respect to the other if the combined forecast has an error variance which is not smaller than that associated with the forecast in question by itself. Chong & Hendry (1986) introduce the interpretation that a conditionally efficient forecast encompasses the other forecast, i.e. the inferior forecast has no valuable information with regard to prediction to contribute to the encompassing forecast. The motivation for testing for forecast encompassing follows from these concepts.

This chapter examines tests for forecast encompassing. Section 3.2 analyses the currently applied regression test and discovers problems with its application in

situations of forecast error non-normality. Section 3.3 develops two modifications to the original test to correct this problematic behaviour, and sections 3.4 and 3.5 propose two new tests for the forecast encompassing null. Some power comparisons are given in section 3.6, and the chapter is summarised and concluded in the final section, 3.7.

3.2 Regression Test

The established test for forecast encompassing (e.g. Clements & Hendry, 1993) involves testing for conditional efficiency, as formalised by Granger & Newbold (1973). The procedure considers two 1-step-ahead forecasts f_{1t}, f_{2t} of the same actual value y_t . Now f_{1t} is said to be conditionally efficient with respect to f_{2t} if the combined forecast f_{ct} (some function of f_{1t} and f_{2t}) exhibits no significant improvement in the forecast error variance than that generated by f_{1t} alone. If a weighted average of forecasts is employed (as in equation (3.1) below) this concept of conditional efficiency can be tested directly:

$$\text{combined forecast: } f_{ct} = (1 - \lambda)f_{1t} + \lambda f_{2t} \quad 0 \leq \lambda \leq 1 \quad (3.1)$$

$$\text{forecast errors: } e_{1t} = y_t - f_{1t} \quad (3.2)$$

$$e_{2t} = y_t - f_{2t} \quad (3.3)$$

$$e_{ct} = y_t - f_{ct} \quad (3.4)$$

Rearranging (3.1)-(3.4) gives:

$$e_{1t} = \lambda(e_{1t} - e_{2t}) + e_{ct} \quad (3.5)$$

The test for conditional efficiency now amounts to running the regression formulated in equation (3.5) and conducting a t -test for the significance of λ . The null hypothesis of $\lambda = 0$ implies that combining f_{2t} with f_{1t} does not improve the error variance from the original forecast f_{1t} , and thus f_{1t} is said to be conditionally efficient with respect to f_{2t} . The test is 1-sided with an alternative hypothesis of

$\lambda > 0$. The null hypothesis amounts to saying that f_{2t} has no valuable information to add to f_{1t} , and thus the forecast f_{1t} encompasses f_{2t} , and the regression test for forecast encompassing is derived:

$$y_t = \lambda x_t + \varepsilon_t \quad (3.6)$$

$$\text{where } y_t = e_{1t}$$

$$x_t = e_{1t} - e_{2t}$$

$$R = \frac{\hat{\lambda}}{\sqrt{\hat{\sigma}^2 / \sum x_t^2}} \sim t_{n-1} \quad \text{under } H_0 \quad (3.7)$$

$$\text{where } \hat{\lambda} = \frac{\sum x_t y_t}{\sum x_t^2}$$

$$\hat{\sigma}^2 = (n-1)^{-1} \sum (y_t - \hat{\lambda} x_t)^2$$

The forecast encompassing test centres on the regression specified in equation (3.5). This has obvious similarities to the Morgan-Granger-Newbold test for equal forecast accuracy considered in chapter 2, with (3.7) bearing a very close resemblance to (2.18). Consequently, because of the problems associated with the *MGN* test when the forecast errors are non-normal, it becomes necessary to examine the behaviour of the test for forecast encompassing in these circumstances.

3.2.1 Theory

The test in question can be decomposed and analysed in a way comparable to the chapter 2 examination of the *MGN* test. Firstly, it can be noted that the parameter λ in (3.6) has the usual population interpretation:

$$\lambda = \frac{E(x_t y_t)}{E(x_t^2)}$$

which allows a useful result to be obtained under the null hypothesis:

$$H_0 : \lambda = 0$$

$$\text{i.e. } E(x_t y_t) = 0$$

$$E(e_{1t}^2) = E(e_{1t} e_{2t}) \quad (3.8)$$

From here, the analysis depends on the distribution of the forecast errors, and thus proceeds in two sub-sections corresponding to the respective situations of forecast error normality and non-normality.

3.2.1a Normal Forecast Errors

In this initial case, the errors made from the two forecasts of interest, denoted e_{1t} and e_{2t} , are assumed to follow a bivariate normal distribution:

$$e \sim N(0, \Omega) \quad (3.9)$$

$$\text{where } \Omega = \begin{bmatrix} V(e_{1t}) & C(e_{1t}, e_{2t}) \\ C(e_{1t}, e_{2t}) & V(e_{2t}) \end{bmatrix}$$

Under the null hypothesis, f_{1t} is conditionally efficient with respect to f_{2t} , thus $V(e_{2t}) > V(e_{1t})$. Furthermore, the result in (3.8) shows that $V(e_{1t}) = C(e_{1t}, e_{2t})$, again under the null. Normalising on $V(e_{1t})$, i.e. $V(e_{1t}) = C(e_{1t}, e_{2t}) = 1$, and denoting $V(e_{2t}) = \omega > 1$, gives the following variance-covariance matrix:

$$\Omega = \begin{bmatrix} 1 & 1 \\ 1 & \omega \end{bmatrix}; \quad \omega > 1 \quad (3.10)$$

Now the forecast errors can be decomposed into linear functions of independent standard normal random variables. The first step notes a pair of realisations from the standard bivariate normal distribution:

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \sim N(0, I)$$

Transformation of this vector z by premultiplication of a matrix T yields the following:

$$Tz \sim N(0, TT')$$

It is clear, then, that choosing the matrix T such that $TT' = \Omega$ allows the decomposition to be stated:

$$e = Tz \sim N(0, \Omega)$$

The obvious choice for T is the triangular matrix below (the minus coefficient in element t_{22} is to simplify the interpretation which results in equation (3.14) below):

$$T = \begin{bmatrix} 1 & 0 \\ 1 & -\sqrt{\omega-1} \end{bmatrix}$$

It is now possible to write the forecast errors as linear combinations of the independent standard normal random variables contained in z :

$$e_{1t} = z_{1t} \quad (3.11)$$

$$e_{2t} = z_{1t} - (\omega - 1)^{1/2} z_{2t} \quad (3.12)$$

Substitution of the decomposition noted in (3.11) and (3.12) into the test statistic (3.7) leads to an interesting result:

$$\begin{aligned} R &= \frac{(n-1)^{1/2} \sum e_{1t} (e_{1t} - e_{2t})}{\sqrt{\sum e_{1t}^2 \sum (e_{1t} - e_{2t})^2 - [\sum e_{1t} (e_{1t} - e_{2t})]^2}} \\ &= \frac{(n-1)^{1/2} (\omega - 1)^{1/2} \sum z_{1t} z_{2t}}{\sqrt{(\omega - 1) \sum z_{1t}^2 \sum z_{2t}^2 - [(\omega - 1)^{1/2} \sum z_{1t} z_{2t}]^2}} \\ &= \frac{(n-1)^{1/2} \sum z_{1t} z_{2t}}{\sqrt{\sum z_{1t}^2 \sum z_{2t}^2 - (\sum z_{1t} z_{2t})^2}} \end{aligned} \quad (3.13)$$

Not only is this test statistic invariant to the value of $\omega = V(e_{2t}) > 1$, it is also identical to the test statistic for a t -test of the null $\beta = 0$ in the following regression:

$$z_{1t} = \beta z_{2t} + \eta_{1t} \quad (3.14)$$

The regression test can therefore be written more simply in this case:

$$R = \frac{\hat{\beta}}{\sqrt{s_1^2 / \sum z_{2t}^2}} \sim t_{n-1} \quad \text{under } H_0 \quad (3.15)$$

$$\text{where } \hat{\beta} = \frac{\sum z_{1t} z_{2t}}{\sum z_{2t}^2}$$

$$s_1^2 = (n-1)^{-1} \sum (z_{1t} - \hat{\beta} z_{2t})^2$$

The null distribution of $\hat{\beta}$ is known asymptotically (e.g. White, 1984), and can be written down in the form of (2.19):

$$D^{-1/2}n^{1/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0,1) \quad (3.16)$$

$$\text{where } D = M^{-2}Q$$

$$M = E(z_{2t}^2)$$

$$Q = V(n^{-1/2}\Sigma z_{2t}\eta_{1t})$$

Now much of the chapter 2 analysis follows through at this point, so further analysis of the regression test can be done with reference to sub-section 2.4.2 in order to discover if any problems exist when implementing the regression test for forecast encompassing.

Estimation of D is implicitly performed in the regression test, and results in a null distribution for R identical to that in (2.20) except that now $y_t = z_{1t}$ and $x_t = z_{2t}$.

Now (2.21) will hold again, yielding the following intermediate result, c.f. (2.22):

$$R \xrightarrow{d} N(0, \text{plim}[\hat{Q}^{-1}Q]) \quad (3.17)$$

$$\text{where } \hat{Q} = s_1^2 n^{-1} \Sigma z_{2t}^2$$

In this world of forecast error normality, values for Q and $\text{plim}\hat{Q}$ can be found under the null:

$$\begin{aligned} Q &= V(z_{2t}z_{1t}) && \text{assuming } z_{2t}\eta_{1t}, z_{2s}\eta_{1s} \text{ uncorrelated for } s \neq t \\ &= E(z_{1t}^2)E(z_{2t}^2) && \text{as } z_{1t}, z_{2t} \text{ independent} \\ &= 1 \end{aligned} \quad (3.18)$$

$$\hat{Q} \xrightarrow{p} V(\eta_{1t})E(z_{2t}^2)$$

$$\begin{aligned}
&= E(z_{1t}^2)E(z_{2t}^2) \\
&= 1
\end{aligned} \tag{3.19}$$

Substituting results (3.18) and (3.19) into (3.17) gives the test statistic distribution:

$$R \xrightarrow{d} N(0,1) \tag{3.20}$$

and the regression test for forecast encompassing would therefore not be expected to exhibit any problematic size behaviour when the errors are normally distributed.

3.2.1b Non-Normal Forecast Errors

Turning now to consider non-normal forecast errors, it is assumed that the errors e_{1t} and e_{2t} follow a bivariate Student's t distribution with ν degrees of freedom as described by Dunnett & Sobel (1954). As discussed in chapter 2, use of such non-normal errors is preferable to the alternative approach of starting with two independent univariate Student's t error vectors and transforming them to assume the desired correlation properties (as employed by Diebold & Mariano, 1995). The reasons for preferring the bivariate method are, as argued in the previous chapter, that the desirable results of the errors themselves being t_ν and the squared errors being correlated both occur when using this approach. The errors can then be described as follows:

$$e_{it} = \frac{u_{it}}{\sqrt{\chi_{t,\nu}^2/\nu}} \quad i = 1,2$$

where $\chi_{t,v}^2$ is independent of u_{it}

$$u = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim N(0, \Omega)$$

$$\text{i.e. } e_{1t} = \frac{z_{1t}}{\sqrt{\kappa_{t,v}}} \quad (3.21)$$

$$e_{2t} = \frac{z_{1t} - (\omega - 1)^{1/2} z_{2t}}{\sqrt{\kappa_{t,v}}} \quad (3.22)$$

$$\text{where } \kappa_{t,v} = \chi_{t,v}^2 / \nu$$

These results can again be substituted into the test statistic (3.7):

$$\begin{aligned} R &= \frac{(n-1)^{1/2} \Sigma e_{1t} (e_{1t} - e_{2t})}{\sqrt{\Sigma e_{1t}^2 \Sigma (e_{1t} - e_{2t})^2 - [\Sigma e_{1t} (e_{1t} - e_{2t})]^2}} \\ &= \frac{(n-1)^{1/2} (\omega - 1)^{1/2} \Sigma \kappa_{t,v}^{-1} z_{1t} z_{2t}}{\sqrt{(\omega - 1) \Sigma \kappa_{t,v}^{-1} z_{1t}^2 \Sigma \kappa_{t,v}^{-1} z_{2t}^2 - [(\omega - 1)^{1/2} \Sigma \kappa_{t,v}^{-1} z_{1t} z_{2t}]^2}} \\ &= \frac{(n-1)^{1/2} \Sigma w_{1t} w_{2t}}{\sqrt{\Sigma w_{1t}^2 \Sigma w_{2t}^2 - (\Sigma w_{1t} w_{2t})^2}} \end{aligned} \quad (3.23)$$

$$\text{where } w_{it} = \kappa_{t,v}^{-1/2} z_{it} \quad i = 1, 2$$

$\kappa_{t,v}$ is independent of z_{it}

As with the normal errors, this test statistic (3.23) is invariant to $V(e_{2t})$ as it is invariant to the choice of ω (which is in this case $V(u_{2t}) > 1$); the statistic also has a useful interpretation, being identical to the test statistic for the null $\gamma = 0$ in the regression:

$$w_{1t} = \gamma w_{2t} + \eta_{2t} \quad (3.24)$$

As with the normal error case, the regression test can now be simplified:

$$R = \frac{\hat{\gamma}}{\sqrt{s_2^2 / \Sigma w_{2t}^2}} \sim t_{n-1} \quad \text{under } H_0 \quad (3.25)$$

$$\text{where } \hat{\gamma} = \frac{\Sigma w_{1t} w_{2t}}{\Sigma w_{2t}^2}$$

$$s_2^2 = (n-1)^{-1} \Sigma (w_{1t} - \hat{\gamma} w_{2t})^2$$

w_{1t}, w_{2t} are uncorrelated bivariate t_v random variables

Following the same analysis through as before, R now takes the following distribution under the null hypothesis:

$$R \xrightarrow{d} N(0, \text{plim}[\hat{Q}^{-1}Q]) \quad (3.26)$$

$$\text{where } Q = V(n^{-1/2} \Sigma w_{2t} \eta_{2t})$$

$$\hat{Q} = s_2^2 n^{-1} \Sigma w_{2t}^2$$

and the variance term $\text{plim}[\hat{Q}^{-1}Q]$ can be found as follows:

$$\begin{aligned} Q &= V(w_{2t} w_{1t}) \quad \text{assuming } w_{2t} \eta_{2t}, w_{2s} \eta_{2s} \text{ uncorrelated for } s \neq t \\ &= E(\kappa_{t,v}^{-2} z_{1t}^2 z_{2t}^2) \quad \text{using } w_{it} \text{ definition below (3.23)} \\ &= A_{2,v} \quad \text{where } A_{k,v} = E(\kappa_{t,v}^{-k}) \end{aligned} \quad (3.27)$$

$$\begin{aligned} \hat{Q} &\xrightarrow{p} V(\eta_{2t}) E(w_{2t}^2) \\ &= E(w_{1t}^2) E(w_{2t}^2) \\ &= E(\kappa_{t,v}^{-1} z_{1t}^2) E(\kappa_{t,v}^{-1} z_{2t}^2) \\ &= A_{1,v}^2 \end{aligned} \quad (3.28)$$

$$\text{so } \text{plim}[\hat{Q}^{-1}Q] = A_{1,v}^{-2}A_2 \quad (3.29)$$

The terms $A_{k,v}$ in the above equations can be expressed more explicitly:

$$\begin{aligned} A_{k,v} &= \frac{\nu^k \Gamma(\frac{\nu}{2} - k)}{2^k \Gamma(\frac{\nu}{2})} \\ A_{1,v}^{-2} A_2 &= \frac{\nu^2 \Gamma(\frac{\nu}{2} - 2) 4 [\Gamma(\frac{\nu}{2})]^2}{\nu^2 [\Gamma(\frac{\nu}{2} - 1)]^2 4 \Gamma(\frac{\nu}{2})} \\ &= \frac{\nu - 2}{\nu - 4} \end{aligned} \quad (3.30)$$

Collating results (3.26), (3.29) and (3.30) then provides a result for the regression test under the null when forecast error non-normality persists:

$$R \xrightarrow{d} N(0, (\nu - 4)^{-1}(\nu - 2)) \quad (3.31)$$

The test consequently embodies an inconsistent variance estimate in a non-normal world which will introduce significant problems in the test's application - that of asymptotic oversizing. It is important to note that this is exactly the same problem faced by *MGN* when bivariate Student's t errors are considered - the variance of (3.31) for t_6 errors is 2, as in (2.39). The regression test, like *MGN*, suffers from conditional heteroscedasticity in the regression errors, and when this is not taken into account the effect is the same for both tests and a lack of robustness is displayed.

The reason that the conditional heteroscedasticity has an identical effect on both tests, with the same asymptotic variance $(\nu - 4)^{-1}(\nu - 2)$ resulting when bivariate t_ν

errors are assumed, stems from the fact that the *MGN* and *R* regressions are special cases of a more general problem. The two regressions and tests are, in common form:

$$y_t = \delta x_t + \varepsilon_t ; \quad H_0 : \delta = 0 ; \quad \text{use } t\text{-test on } \hat{\delta}$$

$$MGN: \quad y_t = e_{1t} + e_{2t}, \quad x_t = e_{1t} - e_{2t}$$

$$R: \quad y_t = e_{1t}, \quad x_t = e_{1t} - e_{2t}$$

Now since e_{1t}, e_{2t} are distributed bivariate t_ν , the standardised linear combinations $V(y_t)^{-1/2} y_t$, $V(x_t)^{-1/2} x_t$ will also be bivariate t_ν in both cases. Denoting these bivariate t_ν variables $w_{1t} = V(y_t)^{-1/2} y_t$, $w_{2t} = V(x_t)^{-1/2} x_t$ the tests can be written in the general form:

$$V(y_t)^{1/2} w_{1t} = \delta V(x_t)^{1/2} w_{2t} + \varepsilon_t ; \quad H_0 : \delta = 0 ; \quad \text{use } t\text{-test on } \hat{\delta}$$

Given that the *MGN* and *R* tests amount to performing t -tests on $\hat{\delta}$, and that multiplying the dependent and independent variables by constants has no effect on the t -ratios, this general form can be simplified to:

$$w_{1t} = \delta^* w_{2t} + \varepsilon_t ; \quad H_0 : \delta^* = 0 ; \quad \text{use } t\text{-test on } \hat{\delta}^*$$

Furthermore, under the null w_{1t}, w_{2t} are uncorrelated. The *MGN* and *R* tests can then be seen to be special cases of a t -test on the parameter in a regression involving two uncorrelated bivariate t_ν variables. The $\hat{\delta}^*$ t -ratio is asymptotically distributed $N(0, (\nu - 4)^{-1}(\nu - 2))$, forming the general result which the Morgan-Granger-Newbold test for equal forecast accuracy and the regression test for forecast

encompassing follow. This also explains the result (3.23), and similarly (3.13); equivalent results to these also exist for *MGN*.

It is interesting to note that if the Diebold-Mariano method of non-normal error generation is employed in exactly the same way as in their paper, then repetition of the above analysis finds \hat{Q} to be consistent for Q . However, this result is peculiar to their arbitrarily chosen matrix, T , and it can be shown that \hat{Q} will not estimate Q consistently for a chosen matrix T in general, triangular or non-triangular.

The analysis can be extended from here to examine the degree of oversizing that the test exhibits under non-normality. The asymptotic distribution is normal with zero mean and variance $(\nu - 4)^{-1}(\nu - 2)$. For a given bivariate Student's t distribution, this variance can be calculated, and the resulting asymptotic test size evaluated by numerical integration. The distributions considered have six and five degrees of freedom respectively; analysis cannot proceed for lower degrees of freedom because the i 'th moment of a Student's t distribution only exists for $i < \nu$. The results are as follows:

$$t_6: (\nu - 4)^{-1}(\nu - 2) = 2$$

1-sided asymptotic test size at the nominal 10% level

$$= \int_{1.28155}^{\infty} (4\pi)^{-1/2} e^{-x^2/4} .dx = 18.24\%$$

1-sided asymptotic test size at the nominal 5% level

$$= \int_{1.64485}^{\infty} (4\pi)^{-1/2} e^{-x^2/4} .dx = 12.24\%$$

$$t_5: (\nu - 4)^{-1}(\nu - 2) = 3$$

1-sided asymptotic test size at the nominal 10% level

$$= \int_{1.28155}^{\infty} (6\pi)^{-1/2} e^{-x^2/6} .dx = 22.97\%$$

1-sided asymptotic test size at the nominal 5% level

$$= \int_{1.64485}^{\infty} (6\pi)^{-1/2} e^{-x^2/6} .dx = 17.11\%$$

Simulation of the forecast encompassing regression test under different forecast error properties allows confirmation of these theoretical results and also examination of finite sample behaviour. Execution of such experiments forms the basis of the next sub-section.

3.2.2 Simulation

Monte Carlo simulation is now performed to evaluate empirical sizes of the forecast encompassing test statistic for nominal 10% and 5% level tests against a 1-sided alternative ($H_0: \lambda = 0$, $H_1: \lambda > 0$). The forecast errors e_{1t}, e_{2t} are drawn from three bivariate distributions in turn - normal, Student's t_6 and Student's t_5 . A range of sample sizes, n , are examined, ranging from $n = 8$ to $n = 512$, plus a very large sample size of $n = 10,000$ to confirm the above theoretical results for the test's asymptotic size. All simulation experiments performed in this chapter are based on 10,000 replications.

For the simulations involving normal forecast errors, the test statistic formulation

given in (3.13) is used, with the z_{1t}, z_{2t} values generated by drawing realisations from the standard normal distribution. Similarly, the method for simulations using non-normal forecast errors involves employment of the test statistic described in (3.23), with w_{1t}, w_{2t} generated by drawing z_{1t}, z_{2t} values as before and then transforming each realisation by dividing through by the same independent chi-squared random variable.

The simulation results are given in table 3.1, and verify the theoretical analysis of sub-section 3.2.1. Under forecast error normality, the regression test is correctly sized for all sample sizes, as would be expected given the fact that the variance estimate is consistent in this case. When the forecast errors are non-normal, however, the inconsistent variance estimate impacts the test statistic and leads to oversizing in all sample sizes. This can be seen in the table with the empirical sizes ranging from 13.08% to 18.15% at the nominal 10% level, and 7.46% to 12.24% at the nominal 5% level, for the errors drawn from the bivariate Student's t_6 distribution. This problem worsens for the t_5 case, with the errors deviating further from normality. The non-normal sizes appear to converge to a limit which is approximated empirically by the $n = 10,000$ experiment test sizes. Comparison of these sizes with the theoretical asymptotic sizes given in the previous sub-section confirm the validity of the theoretical analysis. The nominal 10% level empirical sizes are 18.15% and 21.91% for the t_6 and t_5 errors respectively, compared with

Table 3.1

Empirical sizes for the regression test at the nominal 10% & 5% levels ($h = 1$)

10%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
N	10.12	9.79	9.94	10.20	10.32	10.42	10.30	9.86
t_6	13.08	14.03	15.25	16.09	16.46	17.00	17.32	18.15
t_5	15.00	16.12	16.38	17.50	18.30	18.58	19.68	21.91

5%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
N	5.18	4.97	4.95	5.29	5.27	5.18	5.32	4.75
t_6	7.46	8.45	9.74	10.29	10.49	11.13	11.54	12.24
t_5	8.93	10.00	10.61	11.47	12.30	12.75	13.93	15.81

the theoretical sizes of 18.24% and 22.97%; similarly, for the nominal 5% level the empirical sizes are 12.24% and 15.81% compared with theoretical sizes of 12.24% and 17.11%.

The simulation results therefore confirm the theoretical proposition that the regression test for forecast encompassing exhibits problematic behaviour when the forecast errors are non-normal. Corrections and alternative testing procedures must consequently be considered.

3.3 Modified Regression Tests

The presence of finite sample and asymptotic oversizing in the forecast encompassing regression test under forecast error non-normality motivates the search for improved tests which do not exhibit this undesirable feature in application.

The analysis of the previous section found the source of the problem to be conditional heteroscedasticity in the regression errors causing inconsistent estimation of the element Q in (3.26). Now given the close similarities between R and the MGN test for equal forecast accuracy, two modifications to the regression test are immediately apparent to correct for the test's inherent lack of robustness - namely equivalent tests to MGN_1^* and MGN_2^* :

$$R_{m1} = \frac{\hat{\lambda}}{\sqrt{\sum x_t^2 \hat{\varepsilon}_t^2 / (\sum x_t^2)^2}} \quad (3.32)$$

$$R_{m2} = \frac{\hat{\lambda}}{\sqrt{\sum x_t^2 y_t^2 / (\sum x_t^2)^2}} \quad (3.33)$$

where $x_t, y_t, \hat{\lambda}$ are as defined in (3.6) and (3.7), and $\hat{\varepsilon}_t$ are the least squares residuals from the regression (3.6).

The modified tests given in (3.32) and (3.33) employ consistent estimators of Q as defined in the chapter 2 analysis in equations (2.41) and (2.45), again with $x_t, y_t, \hat{\varepsilon}_t$ defined accordingly for the encompassing problem. As with all the tests of this

form, comparison of the test statistics R_{m1}, R_{m2} is made with critical values from a Student's t distribution with $n - 1$ degrees of freedom.

The two tests also have interesting and useful interpretations, comparable to that for the regression test in (3.13) and (3.23). When the forecast errors are normal, the test statistics of (3.32) and (3.33) become:

$$R_{m1} = \frac{a}{\sqrt{b + c - d}}$$

$$\text{where } a = \Sigma e_{1t}(e_{1t} - e_{2t})\Sigma(e_{1t} - e_{2t})^2$$

$$b = (\Sigma(e_{1t} - e_{2t})^2)^2 \Sigma e_{1t}^2(e_{1t} - e_{2t})^2$$

$$c = (\Sigma e_{1t}(e_{1t} - e_{2t}))^2 \Sigma(e_{1t} - e_{2t})^4$$

$$d = 2\Sigma(e_{1t} - e_{2t})^2 \Sigma e_{1t}(e_{1t} - e_{2t})\Sigma e_{1t}(e_{1t} - e_{2t})^3$$

$$\text{now } a = (\omega - 1)^{3/2} \Sigma z_{1t} z_{2t} \Sigma z_{2t}^2 \quad \text{using (3.11) and (3.12)}$$

$$b = (\omega - 1)^3 (\Sigma z_{2t}^2)^2 \Sigma z_{1t}^2 z_{2t}^2$$

$$c = (\omega - 1)^3 (\Sigma z_{1t} z_{2t})^2 \Sigma z_{2t}^4$$

$$d = 2(\omega - 1)^3 \Sigma z_{2t}^2 \Sigma z_{1t} z_{2t} \Sigma z_{1t} z_{2t}^3$$

$$\text{so } R_{m1} = \frac{\Sigma z_{1t} z_{2t} \Sigma z_{2t}^2}{\sqrt{(\Sigma z_{2t}^2)^2 \Sigma z_{1t}^2 z_{2t}^2 + (\Sigma z_{1t} z_{2t})^2 \Sigma z_{2t}^4 - 2\Sigma z_{2t}^2 \Sigma z_{1t} z_{2t} \Sigma z_{1t} z_{2t}^3}} \quad (3.34)$$

$$R_{m2} = \frac{\Sigma e_{1t}(e_{1t} - e_{2t})}{\sqrt{\Sigma(e_{1t} - e_{2t})^2 e_{1t}^2}}$$

$$= \frac{(\omega - 1)^{1/2} \Sigma z_{1t} z_{2t}}{\sqrt{(\omega - 1) \Sigma z_{1t}^2 z_{2t}^2}} \quad \text{using (3.11) and (3.12)}$$

$$= \frac{\sum z_{1t} z_{2t}}{\sqrt{\sum z_{1t}^2 z_{2t}^2}} \quad (3.35)$$

As with the regression test, these results (3.34) and (3.35) are invariant to $\omega = V(e_{2t}) > 1$. Furthermore, the test statistics are identical to those which would be obtained if the two modified regression tests were applied to $H_0: \beta = 0$ in the regression specified in (3.14). Similar results exist for non-normal errors, with the test statistics being invariant to $\omega = V(u_{2t})$ and identical to the test statistic associated with the modified regression test of the null $\gamma = 0$ in (3.24). All these results stem from the fact that the regression (3.6) is a special case of the more general (3.14) when the errors are normal and (3.24) when the errors are non-normal, as explained in sub-section 3.2.1*b*.

Monte Carlo simulation is again conducted to examine the empirical sizes of the modified regression tests R_{m1} and R_{m2} for nominal 10% and 5% level tests against a 1-sided alternative. The procedure follows the same method as that for the experiments performed in section 3.2 with the forecast errors drawn from bivariate normal, t_6 and t_5 distributions, and the same sample sizes being considered.

The results of the simulations are given in table 3.2. The first point to note is that the modification has succeeded in consistently estimating the $\hat{\lambda}$ variance, with the largest sample size ($n = 10,000$) indicating that the tests are asymptotically correctly

Table 3.2

Empirical sizes for the modified regression tests
at the nominal 10% & 5% levels ($h = 1$)

10%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
N	15.75	13.10	11.47	11.28	10.74	10.58	10.34	9.84
	8.48	9.80	9.65	10.52	10.36	10.33	10.26	9.82
t_6	18.06	15.58	14.54	13.02	12.10	11.31	10.96	10.24
	7.13	9.59	10.62	10.36	10.31	10.41	10.43	10.16
t_5	19.56	17.14	14.92	13.27	12.09	11.31	11.38	10.29
	7.56	9.64	10.09	10.18	10.23	9.85	10.46	10.17

5%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
N	10.06	7.96	6.21	6.01	5.73	5.46	5.33	4.75
	1.56	3.57	4.30	4.89	5.03	5.13	5.25	4.75
t_6	12.02	9.93	8.92	7.50	6.47	5.90	5.68	5.05
	1.07	3.16	4.28	4.51	4.69	4.86	5.00	5.00
t_5	13.47	11.42	9.20	7.67	6.83	6.09	6.02	5.16
	1.06	3.01	3.70	4.24	4.43	4.56	5.07	5.08

Note:- The first entry in each cell is for the R_{m1} test, the second for the R_{m2} test.

sized for both normal and non-normal forecast errors.

With regard to R_{m1} , however, the cost of this correction is its effect on the small and moderate sample sizes. It can be seen that serious oversizing occurs for all the forecast error distributions, and although the magnitude of this problem now decreases with more observations, it is not until $n = 128$ for normal forecast errors, and at least $n = 512$ for non-normal errors that the test size approaches the (correct) limit. In fact, for the smallest samples, the empirical sizes are actually worse than they were for the original regression test. Further to this extent, the modification makes the test size worse in moderate samples when the errors are normally, or near normally, distributed.

The first modified regression test, R_{m1} , therefore goes some way towards improving the problem of oversizing as the test statistic is now correctly sized for very large samples, but still falls short of a useful contribution to the practice of testing for forecast encompassing, with small and moderate sample oversizing still being very much apparent, and in some cases worse.

The finite sample behaviour can be more fully understood by decomposing the implicit estimator \hat{Q}_{m1} as was performed for MGN_1^* in chapter 2:

$$\begin{aligned}\hat{Q}_{m1} &= n^{-1}\sum x_t^2 \varepsilon_t^2 - 2n^{-1}(\hat{\lambda} - \lambda)\sum x_t^3 \varepsilon_t + n^{-1}(\hat{\lambda} - \lambda)^2 \sum x_t^4 \\ &= O_P(1) - O_P(n^{-1/2})O_P(1) + O_P(n^{-1})O_P(1)\end{aligned}$$

The latter two terms in the above expression, particularly the second term, converge slowly to zero, resulting in the poor size performance observed in small and moderate samples.

Turning now to the simulation results for R_{m2} , table 3.2 shows the test to compare very favourably with the empirical sizes for the regression test and its first modification. With reference to the 10% level experiments, apart from undersizing in the case of the smallest sample size, the second modified regression test is approximately correctly sized for all samples (and in the limit). This picture also applies to the nominal 5% level tests, with a slightly greater degree of undersizing in the smaller samples displayed. The test does not exhibit the problematic non-normal behaviour of the regression test, and overcomes the undesirable oversizing of the most natural modification to that test. Even when the test is missized, it is undersized; this is preferable to oversizing, with less chance of a type I error (rejection of a true null) being made.

The second modified regression test consequently embodies a valuable correction to the regression test, creating a very useful method for testing for forecast encompassing when the focus is on 1-step-ahead forecasts. Questions relating to the power of this test, following the fact that (as with MGN_2^*) the variance of the regression parameter is only consistently estimated under the null, are examined later in section 3.6.

3.4 Rank Correlation Test

The forecast encompassing regression test behaves in an undesirable manner under forecast error non-normality; the second modified regression test corrects the implicit problem of inconsistent variance estimation and provides a valuable substitute. An alternative approach, given that the problems of the regression test manifest themselves in situations of non-normality, is to devise a nonparametric test for forecast encompassing.

The regression test examines the significance of the coefficient λ in the two variable regression specified in (3.6). An equivalent correlation test exists and can be noted as follows:

$$H_0: \rho_{x_t y_t} = 0 \quad (\text{identical to } H_0: \lambda = 0)$$

$$H_1: \rho_{x_t y_t} > 0 \quad (\text{identical to } H_1: \lambda > 0)$$

$$C = \frac{\hat{\rho}_{x_t y_t}}{\sqrt{(1 - \hat{\rho}_{x_t y_t}^2)/(n-1)}} \quad (3.36)$$

$$\text{where } \hat{\rho}_{x_t y_t} = \frac{\sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2}}$$

$$x_t = e_{1t} - e_{2t}$$

$$y_t = e_{1t}$$

The obvious extension is then to employ Spearman's rank correlation test as a nonparametric approach to testing for forecast encompassing. This can be formalised as follows:

$$H_0: \rho_{x_t, y_t} = 0$$

$$H_1: \rho_{x_t, y_t} > 0$$

$$r_s = \frac{n^{-1} \sum x_{rt} y_{rt} - [(n+1)/2]^2}{(n^2 - 1)/12} \quad (3.37)$$

where x_{rt}, y_{rt} = rankings of x_t, y_t respectively

The test statistic r_s has a known distribution under the null, and the critical values are tabulated for the smaller sample sizes. For samples of size $n = 64$ and larger, an approximation is used, as described in chapter 2 where an equivalent form of the test is proposed as an alternative test for equal forecast accuracy. As in the equal accuracy case, use of this approximation can be shown to be inconsequential by simulation of the critical values for the larger samples.

For completeness, it can be shown that the test statistic is unaffected by the chosen value of ω . In order to prove this, it is necessary only to show that ω does not affect the rankings of x_t and y_t . For normal errors (the proof for non-normal errors follows directly from this):

$$x_t = e_{1t} - e_{2t} = (\omega - 1)^{1/2} z_{2t} \quad \text{using (3.11) and (3.12)} \quad (3.38)$$

$$y_t = e_{1t} = z_{1t} \quad (3.39)$$

It can be seen from these simple expansions that ω does not affect y_t at all, and only affects the scale of x_t due to its uniform effect in the decomposition given above. Clearly, therefore, the rankings x_{rt}, y_{rt} will be invariant to ω , resulting in the invariance of the test statistic.

The rank correlation test has correct size in all finite samples, and is consequently valid for small, moderate and large samples under forecast error normality and non-normality, hence becoming very valuable in application. The problematic behaviour of the regression test is not apparent, nor the finite sample oversizing of the first modified regression test. Instead, a useful nonparametric test is derived which improves upon the currently applied regression test and is attractive when testing for forecast encompassing using 1-step-ahead forecasts.

3.5 Diebold-Mariano Approach Tests

The second modified regression and rank correlation tests for forecast encompassing overcome the immediate problems associated with the regression test, and both are valuable when considering 1-step-ahead forecasts. It is also possible to derive another new test using a Diebold-Mariano (1995) type approach, which takes a more general specification and can be readily applied to h -steps-ahead prediction ($h > 1$).

The Diebold-Mariano test is a test of the null of equal forecast accuracy. The procedure supposes that the quality of a forecast is to be evaluated according to some function of the forecast errors. The null then amounts to:

$$E[g(e_{1t}, e_{2t})] = 0$$
$$\text{or } E(d_t) = 0 \quad \text{where } d_t = g(e_{1t}, e_{2t})$$

and the test is based on the sample mean of the loss differential series d_t .

It is fairly straightforward to employ this approach to testing for forecast encompassing. The regression test centres on the equation specified in (3.6) and tests the significance of λ . Now the population interpretation of λ gives the result in (3.8) under the null hypothesis. This result can be manipulated to take the form of a Diebold-Mariano test as follows:

$$\text{under } H_0 : E(e_{1t}^2 - e_{1t}e_{2t}) = 0 \quad \text{from (3.8)}$$
$$\text{or } E(d_t) = 0 \quad \text{where } d_t = e_{1t}^2 - e_{1t}e_{2t} \quad (3.40)$$

Once the loss differential series, d_t , has been specified, the Diebold-Mariano testing

procedure follows through as for the null of equal forecast accuracy, and the new test is derived.

Referring to chapter 2, analysis shows that the Diebold-Mariano test can be improved by making two modifications - a finite sample correction to the variance estimate $\hat{V}(\bar{d})$, and the comparison of the test against Student's t_{n-1} critical values. Application of the original and modified tests to this problem of forecast encompassing thus generate two new tests which can be summarised as below:

$$H_0: \lambda = 0$$

$$H_1: \lambda > 0$$

1. Diebold-Mariano approach:

$$DM = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \quad (3.41)$$

$$\text{where } d_t = e_{1t}^2 - e_{1t}e_{2t}$$

$$\bar{d} = n^{-1} \sum_{t=1}^n d_t$$

$$\hat{V}(\bar{d}) = n^{-1} (\hat{\gamma}_0 + 2 \sum_{k=1}^{h-1} \hat{\gamma}_k)$$

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$$

compare with $N(0,1)$ critical values

2. Modified Diebold-Mariano approach:

$$DM_m = \frac{\bar{d}}{\sqrt{\hat{V}_m(\bar{d})}} \quad (3.42)$$

$$\text{where } \hat{V}_m(\bar{d}) = [n+1-2h+n^{-1}h(h-1)]^{-1}(\hat{\gamma}_0 + 2\sum_{k=1}^{h-1}\hat{\gamma}_k)$$

$$\text{i.e. } DM_m = n^{-1/2}[n+1-2h+n^{-1}h(h-1)]^{1/2} DM$$

compare with t_{n-1} critical values

As with the other parametric tests, these Diebold-Mariano-type tests have useful interpretations as follows. For normal forecast errors, the test statistic for the 1-step-ahead modified Diebold-Mariano approach test can be written as below (equivalent results exist for h -steps-ahead forecasts and for the DM test statistic):

$$\begin{aligned} DM_m &= \frac{[n^{-1}(n-1)]^{1/2} \sum e_{1t}(e_{1t} - e_{2t})}{\sqrt{\sum e_{1t}^2(e_{1t} - e_{2t})^2 - n^{-1}[\sum e_{1t}(e_{1t} - e_{2t})]^2}} \\ &= \frac{[n^{-1}(n-1)]^{1/2} (\omega - 1)^{1/2} \sum z_{1t} z_{2t}}{\sqrt{(\omega - 1) \sum z_{1t}^2 z_{2t}^2 - n^{-1}(\omega - 1)(\sum z_{1t} z_{2t})^2}} \quad \text{using (3.11), (3.12)} \\ &= \frac{[n^{-1}(n-1)]^{1/2} \sum z_{1t} z_{2t}}{\sqrt{\sum z_{1t}^2 z_{2t}^2 - n^{-1}(\sum z_{1t} z_{2t})^2}} \quad (3.43) \end{aligned}$$

This result for DM_m is invariant to ω and is identical to the test statistic associated with the modified Diebold-Mariano approach test applied to the case where the loss differential series is $d_t = z_{1t} z_{2t}$. Similarly, when the forecast errors are non-normal, the test statistic is once again invariant to the respective ω , and is the same as the test statistic for the modified Diebold-Mariano approach test when $d_t = w_{1t} w_{2t}$, and w_{it} is as defined below (3.23). The Diebold-Mariano tests can therefore be

considered as the tests that flow naturally from consideration of the regressions (3.14) and (3.24). Again this follows from (3.6) being a special case of (3.14) and (3.24) for normal and non-normal errors respectively (see sub-section 3.2.1b).

The tests are examined by way of Monte Carlo simulation comparable to that for the other tests. For 1-step-ahead forecasts, the experiments parallel those conducted for the original and modified regression tests. More specifically, empirical sizes are evaluated for nominal 10% and 5% level tests against a 1-sided alternative. Forecast errors are drawn from the bivariate normal, t_6 and t_5 distributions, and sample sizes between $n = 8$ and $n = 512$ are considered (the asymptotic check case, $n = 10,000$, is unnecessary). The results of these simulations are given in table 3.3.

It has already been noted that the Diebold-Mariano type tests can be readily used for multi-step-ahead prediction (the other tests require further modifications to account for error autocorrelation which is present when $h > 1$), and tables 3.4-3.7 give results for the simulations of such longer forecast horizons. Table 3.4 reports the 2-steps-ahead results, with autocorrelation built in to the forecast errors according to a first order moving average process with parameter θ . Values of $\theta = 0, 0.5, 0.9$ are used and incorporated into the simulations by generating new forecast errors (e_{1t}^*, e_{2t}^*) by the following transformation:

$$e_{1t}^* = (e_{1t} + \theta e_{1,t-1}) / (1 + \theta^2)^{1/2}$$

$$e_{2t}^* = (e_{2t} + \theta e_{2,t-1}) / (1 + \theta^2)^{1/2}$$

Table 3.3

Empirical sizes for the Diebold-Mariano approach tests
at the nominal 10% & 5% levels ($h = 1$)

10%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	14.62	12.35	10.79	10.94	10.56	10.42	10.30
	10.21	10.49	9.90	10.65	10.39	10.35	10.28
t_6	13.84	12.29	11.87	10.99	10.63	10.57	10.52
	8.99	10.23	10.92	10.54	10.37	10.43	10.44
t_5	14.52	12.54	11.45	10.74	10.48	10.03	10.54
	9.50	10.40	10.35	10.25	10.28	9.87	10.49

5%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	8.37	6.49	5.38	5.46	5.37	5.31	5.29
	4.36	4.92	4.77	5.06	5.21	5.18	5.27
t_6	7.06	6.15	5.72	5.27	5.02	5.10	5.04
	3.33	4.32	4.78	4.81	4.81	4.95	5.01
t_5	7.40	6.07	5.32	4.94	4.91	4.67	5.13
	3.37	4.26	4.28	4.45	4.60	4.59	5.12

Note:- The first entry in each cell is for the DM test, the second for the DM_m test.

Table 3.4

Empirical sizes for the Diebold-Mariano approach tests
at the nominal 10% & 5% levels ($h = 2$)

10%		$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	$\theta = 0.0$	18.09	14.91	12.88	11.79	10.86	10.58	10.38
		11.28	11.79	11.31	11.04	10.46	10.34	10.27
	$\theta = 0.5$	18.86	15.34	12.68	11.56	10.89	10.48	10.96
		11.41	11.87	10.94	10.67	10.51	10.36	10.79
	$\theta = 0.9$	18.96	15.19	12.58	11.43	10.95	10.65	11.01
		11.45	11.79	10.87	10.73	10.51	10.36	10.86
t_6	$\theta = 0.0$	18.19	15.52	13.24	11.78	10.98	10.78	10.27
		10.70	11.85	11.40	10.73	10.53	10.60	10.19
	$\theta = 0.5$	19.19	15.18	13.02	11.20	10.67	10.71	10.07
		10.99	11.71	11.15	10.26	10.20	10.48	9.98
	$\theta = 0.9$	19.54	15.31	12.84	11.27	10.78	10.58	10.14
		11.39	11.52	10.98	10.37	10.38	10.33	10.05
t_5	$\theta = 0.0$	19.39	16.06	12.73	11.39	11.03	10.27	10.84
		10.91	11.88	10.97	10.66	10.55	10.11	10.77
	$\theta = 0.5$	19.34	15.58	12.83	11.36	10.64	10.54	10.23
		10.78	11.78	11.03	10.29	10.21	10.37	10.18
	$\theta = 0.9$	20.32	15.73	12.71	11.22	10.64	10.49	10.36
		11.31	11.75	10.97	10.35	10.15	10.23	10.27

5%		$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	$\theta = 0.0$	12.60	9.96	7.25	6.51	5.64	5.39	5.23
		6.38	6.89	5.81	5.69	5.16	5.16	5.19
	$\theta = 0.5$	12.81	9.57	7.05	6.15	5.79	5.61	5.36
		6.12	6.29	5.63	5.48	5.48	5.39	5.27
	$\theta = 0.9$	12.85	9.40	7.00	6.05	5.62	5.71	5.53
		5.95	6.21	5.57	5.37	5.38	5.47	5.43
t_6	$\theta = 0.0$	12.10	9.41	7.18	5.67	5.28	5.06	5.02
		5.76	6.07	5.47	4.86	4.96	4.89	4.91
	$\theta = 0.5$	12.44	8.91	6.83	5.39	5.37	5.06	4.99
		5.72	5.59	5.42	4.84	4.98	4.93	4.89
	$\theta = 0.9$	12.89	8.97	6.78	5.49	5.58	5.04	5.03
		5.89	5.45	5.15	4.75	5.06	4.89	4.94
t_5	$\theta = 0.0$	12.50	9.60	6.84	5.47	5.35	4.93	5.37
		6.23	6.07	5.16	4.75	5.01	4.82	5.24
	$\theta = 0.5$	12.06	9.13	6.72	5.19	5.04	4.94	5.54
		5.64	5.92	5.14	4.48	4.61	4.78	5.45
	$\theta = 0.9$	12.69	8.94	6.53	5.24	5.10	5.01	5.36
		5.73	5.46	5.11	4.34	4.79	4.89	5.32

Note:- The first entry in each cell is for the DM test, the second for the DM_m test.

Table 3.5

Empirical sizes for the Diebold-Mariano approach tests
at the nominal 10% & 5% levels ($h = 3, \theta_i = 0 \forall i$)

10%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	18.20	16.14	14.07	13.10	11.52	10.75	10.37
	9.68	11.86	11.73	12.01	10.92	10.52	10.26
t_6	19.18	17.40	14.63	12.63	11.52	10.83	10.56
	9.45	12.44	12.12	11.41	10.97	10.46	10.47
t_5	20.88	18.31	14.50	12.31	11.14	10.39	10.67
	10.09	12.34	11.73	11.02	10.52	10.17	10.55

5%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	13.58	11.27	8.83	7.35	6.24	5.58	5.47
	5.98	7.17	6.85	6.18	5.69	5.34	5.35
t_6	13.81	11.77	8.88	6.63	5.95	5.42	5.11
	5.63	7.19	6.76	5.46	5.33	5.25	4.98
t_5	14.64	11.57	8.30	6.73	5.73	5.08	5.31
	5.67	6.92	6.24	5.56	5.10	4.90	5.22

Note:- The first entry in each cell is for the DM test, the second for the DM_m test.

Table 3.6

Empirical sizes for the Diebold-Mariano approach tests
at the nominal 10% & 5% levels ($h = 4$, $\theta_i = 0 \ \forall i$)

10%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	18.20	16.84	15.07	13.92	12.25	11.02	10.47
	8.08	10.94	12.32	12.33	11.46	10.65	10.36
t_6	19.61	17.96	15.99	13.65	12.02	11.19	10.73
	7.57	11.56	12.73	11.80	11.13	10.75	10.55
t_5	20.24	18.68	15.46	13.06	11.59	10.54	10.81
	7.99	11.64	12.07	11.49	10.75	10.21	10.54

5%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
N	14.33	11.94	9.95	8.19	6.54	5.87	5.55
	4.84	6.96	7.67	6.85	5.96	5.43	5.37
t_6	14.44	12.51	10.23	7.64	6.30	5.67	5.25
	4.45	7.21	7.35	6.13	5.61	5.40	5.08
t_5	15.31	12.59	9.85	7.36	6.04	5.34	5.35
	4.73	6.78	7.05	5.88	5.38	4.94	5.24

Note:- The first entry in each cell is for the DM test, the second for the DM_m test.

Table 3.7

Empirical sizes for the Diebold-Mariano approach tests
at the nominal 10% & 5% levels ($\theta_i = 0 \ \forall i$, bivariate normal errors)

10%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$h = 5$	17.91	16.62	15.67	14.63	12.63	11.07	10.43
	5.74	10.13	12.29	12.59	11.52	10.63	10.21
$h = 6$	18.32	16.51	15.77	14.98	12.86	11.36	10.59
	4.46	9.10	11.81	12.97	11.78	10.73	10.33
$h = 7$	18.85	16.86	15.78	15.48	13.59	11.58	10.78
	3.46	8.48	11.37	12.90	12.10	10.82	10.47
$h = 8$	-	16.45	15.74	15.70	13.97	11.81	10.98
	-	7.57	11.02	12.78	12.35	10.91	10.68
$h = 9$	-	16.65	16.17	16.00	14.32	12.05	11.05
	-	6.23	10.73	12.74	12.73	11.04	10.70
$h = 10$	-	16.80	15.92	16.06	14.56	12.36	11.25
	-	5.48	9.93	12.44	12.60	11.46	10.84

5%	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$h = 5$	14.46	12.30	10.85	9.10	6.98	5.93	5.71
	3.47	6.43	7.86	7.60	6.10	5.65	5.43
$h = 6$	15.51	12.52	11.02	9.68	7.41	6.17	5.77
	2.57	5.76	7.53	7.83	6.59	5.84	5.53
$h = 7$	17.14	12.75	11.35	10.10	7.93	6.41	5.81
	2.21	5.28	7.50	7.88	6.73	5.94	5.57
$h = 8$	-	12.65	11.57	10.46	8.47	6.54	5.99
	-	4.80	7.22	8.00	7.03	6.07	5.68
$h = 9$	-	13.08	12.02	10.77	9.06	6.79	6.09
	-	3.87	7.17	8.07	7.53	6.06	5.78
$h = 10$	-	13.69	11.80	10.82	9.16	7.14	6.27
	-	3.52	6.48	7.97	7.45	6.31	5.87

Note:- The first entry in each cell is for the DM test, the second for the DM_m test.

Tables 3.5 and 3.6 contain results from simulations of 3-steps- and 4-steps-ahead respectively. In these cases, the forecast errors are generated without autocorrelation; the reason for this is that from table 3.4 the inference can be drawn that the empirical sizes of the test statistics are robust to autocorrelated errors. The final table - table 3.7 - also considers only the non-autocorrelated forecast error case, and reports test sizes for 5- through 10-steps-ahead forecasts. For these longer horizons, forecast errors are drawn solely from the bivariate normal distribution because, as with autocorrelated errors, the tests are robust to non-normality.

With regard to the results, a number of points can be drawn. First and foremost, it can be seen from table 3.3 that for 1-step-ahead forecasts, the modified Diebold-Mariano forecast encompassing test, DM_m , is correctly sized over all forecast error distributions and for all sample sizes (with the exception of slight undersizing for the $n = 8$ sample under non-normality). This test is consequently valuable in the practice of testing for forecast encompassing, overcoming the undesirable presence of oversizing in the standard regression test.

The second point to note is that although the unmodified test, DM , is correctly sized in the limit, small sample oversizing is again apparent. This serves as another example of the conclusions drawn in chapter 2 that the original Diebold-Mariano test exhibits significant problems of oversizing in small samples and can be improved upon by use of the modified test. This feature of DM_m superiority to DM in terms of size is evident over all forecast horizons, as can be seen in tables 3.4-3.7.

Not only is the test robust to the distribution of the forecast errors, it is also robust to

autocorrelation in the errors of h -steps-ahead forecasts ($h > 1$). This can be seen in table 3.4 where for the 2-steps-ahead case, with forecast errors generated with first order autocorrelation, the tests' empirical sizes are not impacted by changes in the moving average parameter, θ . The Diebold-Mariano approach to forecast encompassing testing therefore brings with it the advantages of being immediately applicable to multi-step-ahead forecasts, and being robust to any forecast error non-normality, autocorrelation or contemporaneous correlation.

Fourthly, with reference to tables 3.5-3.7, it can be seen that the DM_m test remains a valid and attractive test for longer forecast horizons. As was found with the chapter 2 analysis of this approach to testing, the empirical sizes worsen somewhat as the forecast horizon grows. However, table 3.7 shows that the small and moderate sample explosion in size which occurred with the quadratic loss differential series ($d_t = e_{1t}^2 - e_{2t}^2$) does not occur here. Instead, the empirical sizes for DM_m never go above 13% at the nominal 10% level or much above 8% at the nominal 5% level, even when using forecasts as long as 8-, 9- and 10-steps-ahead. Further research into the question of why the size explodes in the case of quadratic loss, but not here, might be interesting.

Altogether, then, the modified Diebold-Mariano approach test DM_m forms an important contribution to the practice of testing for forecast encompassing. The test is robust to forecast error properties of non-normality, autocorrelation and contemporaneous correlation, and is readily applicable to testing with h -steps-ahead forecasts ($h > 1$), with the empirical sizes correct in the 1-step-ahead case and only

marginally oversized in moderate samples for longer forecast horizons.

3.6 Power Comparisons

Following the analysis of the problems associated with the regression test for forecast encompassing, three new tests have been proposed which are approximately correctly sized in large and moderate samples - the second modified regression test, the rank correlation test and the modified Diebold-Mariano approach test. It is now important to compare the power of these tests both with each other and with the original regression test.

Monte Carlo simulation is used to establish these relative powers, and this involves simulating the alternative hypothesis which can be written as follows:

$$\begin{aligned} H_1: \lambda > 0 \\ \text{or } \frac{E(x_t y_t)}{E(x_t^2)} > 0 \end{aligned} \quad (3.44)$$

Now the denominator in (3.44) is always positive so it is possible to condense the alternative to:

$$\begin{aligned} H_1: E(x_t y_t) > 0 \\ \text{i.e. } E(e_{1t}^2) > E(e_{1t} e_{2t}) \end{aligned} \quad (3.45)$$

The procedure for simulation is the same as that for the empirical size calculations, except now the variance-covariance matrix Ω differs. Recalling the elements of this matrix, as given in (3.9), it is necessary to normalise on one of the constituent elements and then choose values for the remaining terms such that the alternative is generated. The choices of these values will be restricted to ensure that the

alternative holds. The first stage is to normalise by letting $V(e_{1t}) = 1$; the first of the aforementioned restrictions on the choices of the remaining terms can then be derived:

$$V(e_{1t}) = 1: \quad C(e_{1t}, e_{2t}) < 1 \quad \text{from (3.45)} \quad (3.46)$$

The second restriction can be found by using the denominator of (3.44):

$$\begin{aligned} E(x_t^2) &> 0 \\ \text{i.e. } V(e_{1t}) + V(e_{2t}) - 2C(e_{1t}, e_{2t}) &> 0 \\ V(e_{2t}) &> 2C(e_{1t}, e_{2t}) - 1 \end{aligned} \quad (3.47)$$

Thirdly, it is necessary to ensure that the correlation between the forecast errors is less than one in absolute value:

$$\begin{aligned} |Corr(e_{1t}, e_{2t})| &= \left| \frac{C(e_{1t}, e_{2t})}{\sqrt{V(e_{1t})V(e_{2t})}} \right| < 1 \\ \text{i.e. } [C(e_{1t}, e_{2t})]^2 &< V(e_{2t}) \end{aligned} \quad (3.48)$$

Now this third restriction (3.48) subsumes the second restriction (3.47) given that (3.46) is imposed. Collating this information the matrix Ω for the alternative hypothesis becomes:

$$\Omega = \begin{bmatrix} 1 & \zeta \\ \zeta & \omega \end{bmatrix} \quad (3.49)$$

where $\zeta = C(e_{1t}, e_{2t}) < 1$

$$\omega = V(e_{2t}) > [C(e_{1t}, e_{2t})]^2$$

Referring to the method of simulation employed for the tests' size, the next step is to choose a matrix T such that $TT' = \Omega$. The most natural choice is again the triangular matrix:

$$T = \begin{bmatrix} 1 & 0 \\ \zeta & \sqrt{\omega - \zeta^2} \end{bmatrix}$$

The procedure then follows through with appropriately chosen parameters substituted into the matrix T , then premultiplying a vector of standard normal random realisations by T yields the normal forecast errors, and transformation of these generates the Student's t_v forecast errors.

Further analysis shows that each test's power varies according to one quantity alone, p , which is a combination of the chosen parameters $V(e_{2t})$ and $C(e_{1t}, e_{2t})$. This can be shown for each test in turn for (1-step-ahead) normal errors; as with the other proofs of this style, extension to the non-normal case follows directly:

$$e = Tz: \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \zeta & \sqrt{\omega - \zeta^2} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

$$\text{i.e. } e_{1t} = z_{1t} \tag{3.50}$$

$$e_{2t} = \zeta z_{1t} + (\omega - \zeta^2)^{1/2} z_{2t} \tag{3.51}$$

1. Regression test:

$$\begin{aligned}
 R &= \frac{(n-1)^{1/2} \Sigma e_{1t} (e_{1t} - e_{2t})}{\sqrt{\Sigma e_{1t}^2 \Sigma (e_{1t} - e_{2t})^2 - [\Sigma e_{1t} (e_{1t} - e_{2t})]^2}} \\
 &= \frac{(n-1)^{1/2} \Sigma [(1-\zeta) z_{1t}^2 - (\omega - \zeta^2)^{1/2} z_{1t} z_{2t}]}{\sqrt{\Sigma z_{2t}^2 \Sigma [(1-\zeta) z_{1t} - (\omega - \zeta^2)^{1/2} z_{2t}]^2 - \{\Sigma [(1-\zeta) z_{1t}^2 - (\omega - \zeta^2)^{1/2} z_{1t} z_{2t}]\}^2}} \\
 &\quad \text{using (3.50) and (3.51)} \\
 &= \frac{(n-1)^{1/2} (\Sigma z_{1t}^2 - p \Sigma z_{1t} z_{2t})}{p \sqrt{\Sigma z_{1t}^2 \Sigma z_{2t}^2 - (\Sigma z_{1t} z_{2t})^2}} \\
 &\quad \text{where } p = \frac{\sqrt{\omega - \zeta^2}}{1 - \zeta}
 \end{aligned}$$

2. Second modified regression test:

$$\begin{aligned}
 R_{m2} &= \frac{\Sigma e_{1t} (e_{1t} - e_{2t})}{\sqrt{(e_{1t} - e_{2t})^2 e_{1t}^2}} \\
 &= \frac{\Sigma z_{1t} [(1-\zeta) z_{1t} - (\omega - \zeta^2)^{1/2} z_{2t}]}{\sqrt{\Sigma [(1-\zeta) z_{1t} - (\omega - \zeta^2)^{1/2} z_{2t}]^2 z_{1t}^2}} \\
 &= \frac{\Sigma z_{1t}^2 - p \Sigma z_{1t} z_{2t}}{\sqrt{\Sigma (z_{1t} - p z_{2t})^2 z_{1t}^2}}
 \end{aligned}$$

3. Rank correlation test:

$$\begin{aligned}
 x_t &= e_{1t} - e_{2t} & y_t &= e_{1t} \\
 &= (1-\zeta) z_{1t} - (\omega - \zeta^2)^{1/2} z_{2t} & &= z_{2t} \\
 (1-\zeta)^{-1} x_t &= z_{1t} - p z_{2t}
 \end{aligned}$$

i.e. rankings derived from $(1-\zeta)^{-1} x_t$ are identical to rankings derived from x_t , and $y_t = z_{2t}$; therefore x_{rt}, y_{rt} and the test statistic and power vary according to p alone

4. Modified Diebold-Mariano approach test:

$$\begin{aligned}
 DM_m &= \frac{[n^{-1}(n-1)]^{1/2} \sum e_{1t}(e_{1t} - e_{2t})}{\sqrt{\sum e_{1t}^2(e_{1t} - e_{2t})^2 - n^{-1}[\sum e_{1t}(e_{1t} - e_{2t})]^2}} \\
 &= \frac{[n^{-1}(n-1)]^{1/2} \sum [(1-\zeta)z_{1t}^2 - (\omega - \zeta^2)^{1/2} z_{1t}z_{2t}]}{\sqrt{\sum [(1-\zeta)z_{1t}^2 - (\omega - \zeta^2)^{1/2} z_{1t}z_{2t}]^2 - n^{-1}\{\sum [(1-\zeta)z_{1t}^2 - (\omega - \zeta^2)^{1/2} z_{1t}z_{2t}]\}^2}} \\
 &= \frac{[n^{-1}(n-1)]^{1/2} [\sum (z_{1t}^2 - pz_{1t}z_{2t})]}{\sqrt{\sum (z_{1t}^2 - pz_{1t}z_{2t})^2 - n^{-1}[\sum (z_{1t}^2 - pz_{1t}z_{2t})]^2}}
 \end{aligned}$$

For each test, then, the power will be affected purely by changes in the value of p ; a decrease in p represents a movement further away from the null hypothesis (λ is larger) and so the power rises with decreased p .

Simulation experiments are conducted for all four tests at the nominal 10% level for the usual sample sizes between $n = 8$ and $n = 512$. In each case, values for $V(e_{2t})$ and $C(e_{1t}, e_{2t})$ are chosen to comply with the necessary restrictions highlighted above, and to give a value for p which allows meaningful comparisons between the tests' powers. Situations of forecast error normality and non-normality (using t_6 errors) are considered, but powers of the regression test are not calculated for non-normal errors as the test is not correctly sized in such a case. Non-size-adjusted powers are used for the reasons described in chapter 2; see Harvey, Leybourne & Newbold (1998) for size-adjusted powers corresponding to the results of this section. Two sets of results are quoted, reflecting relatively high and low powers respectively.

The simulation results are given in table 3.8. With regard to the interpretation of these results, observations can first be made to compare the new tests with the original regression test. It can be seen that the modified Diebold-Mariano approach test, DM_m , compares favourably with the regression test - for all sample sizes except the very small $n = 8$ case the difference in power between the tests is negligible. The implication is thus that nothing substantial in terms of power is lost by moving from the problematic regression test to DM_m . Regarding the second modified regression test, laying aside the $n = 8$ case where R_{m2} is undersized, the test is only marginally less powerful than DM_m - the potential problem of inconsistent estimation under the alternative not having a significant effect. Under forecast error normality, the rank correlation test is not far behind DM_m in power and is again not considerably inferior to the regression test. Furthermore, when the errors are non-normal, the rank correlation test, as might be expected for a nonparametric test, has more power than DM_m . Power comparisons with the regression test under non-normality are not valid as the test statistic is oversized.

All things considered, the three new tests compare very positively with the regression test - none are significantly inferior in terms of power, but all achieve good size properties in finite samples and in the limit. For 1-step-ahead prediction, any of these tests could be used in place of the regression test. The tests r_s and DM_m are generally more reliable than R_{m2} , and if heavy-tailed errors are expected, then a case exists for employing r_s due to its power advantages. For longer forecast horizons, the rank correlation test has no natural extension to the situation of

Table 3.8

Power comparisons for the original and second modified regression tests, the rank correlation test and the modified Diebold-Mariano approach test at the nominal 10% level ($h = 1$)

High power		$n = 8$ $p = 1$	$n = 16$ $p = 1.75$	$n = 32$ $p = 2.75$	$n = 64$ $p = 4$	$n = 128$ $p = 6$	$n = 256$ $p = 8$	$n = 512$ $p = 12$
N	R	85.77	79.68	76.59	75.50	72.52	76.91	73.26
	R_{m2}	69.91	75.66	74.17	74.79	72.23	76.85	72.92
	r_s	71.20	72.53	71.06	71.87	69.27	73.67	70.17
	DM_m	76.33	77.07	74.69	75.01	72.35	76.86	72.94
t_6	R_{m2}	61.57	67.56	64.95	63.43	59.06	61.00	55.67
	r_s	70.20	70.97	69.21	70.28	66.87	70.57	66.97
	DM_m	68.58	69.22	65.55	63.76	59.20	61.04	55.70

Moderate power		$n = 8$ $p = 3$	$n = 16$ $p = 4$	$n = 32$ $p = 6$	$n = 64$ $p = 8$	$n = 128$ $p = 12$	$n = 256$ $p = 16$	$n = 512$ $p = 20$
N	R	34.16	37.55	36.17	39.32	37.64	39.70	44.43
	R_{m2}	26.50	34.20	35.22	38.75	37.46	39.57	44.31
	r_s	25.39	33.19	33.88	37.17	35.67	37.73	42.25
	DM_m	30.87	35.61	35.67	38.98	37.72	39.66	44.34
t_6	R_{m2}	22.39	31.23	31.47	32.72	30.83	31.58	33.68
	r_s	26.34	33.12	34.25	36.91	35.38	37.27	41.21
	DM_m	27.37	32.94	32.15	32.98	30.94	31.63	33.70

multi-step-ahead prediction, and R_{m2} needs further modifications. The modified Diebold-Mariano approach test, DM_m , comes into its own in such scenarios with direct application to the h -steps-ahead case ($h > 1$) making it a very valuable test for practitioners testing the forecast encompassing null.

3.7 Conclusion

In summary, this study has analysed issues associated with the practice of testing for forecast encompassing when evaluating two different forecasts of the same quantity. The current procedure applied to perform such testing - the regression test - has been shown to be valid only when the forecast errors concerned are normal. If the errors deviate into even mild non-normality, the test becomes oversized both in finite samples and asymptotically as a result of a now inconsistent variance estimate implicit in the test.

Correction of this inconsistent estimator is performed by two modifications to the regression test. Although the first of these modified tests is correctly sized in the limit, the undesirable feature of small and moderate sample oversizing remains and is actually worse in some cases, pertaining to normal forecast errors as well. The second modification is more successful, being approximately correctly sized for all sample sizes, with the exception of the very smallest sample cases.

Moving away from the regression test approach, two other new tests for forecast encompassing are proposed. Firstly, the rank correlation test is derived which forms a nonparametric approach to the problem. This test yields correct sizes for all samples. Secondly, the Diebold-Mariano approach tests are examined; the modified test is again superior to the original Diebold-Mariano approach test in terms of size, and again achieves the correct test sizes in all samples. Both of these tests are thus robust to the distribution of the forecast errors. The analysis is then extended to h -steps-ahead forecasts ($h > 1$), and here the modified Diebold-Mariano approach test

demonstrates its particular value. The modified Diebold-Mariano approach test can be immediately applied to longer forecast horizons unlike the second modified regression and rank correlation tests, and under such conditions is only marginally oversized in moderate sample sizes and is robust to autocorrelation in the forecast errors.

Examination of the tests' powers reveals very positive properties for the three new correctly sized tests - second modified regression test, rank correlation test and modified Diebold-Mariano approach test. None of the tests lose much in power to the regression test under normality, and no single test is materially superior to the others overall. Distinctions can be made, however, with r_s being the best under forecast error non-normality, whilst DM_m exhibits good overall power, reliable for different distributions and sample sizes.

It is the conclusion of this analysis that the regression test lacks robustness to forecast error non-normality, and that the second modified regression, rank correlation and modified Diebold-Mariano approach tests provide robust superior alternatives without a significant loss in power. Furthermore, the modified Diebold-Mariano approach test has the added advantage of direct application to multi-step-ahead forecasts, being robust to the presence of autocorrelation. If the interest is solely 1-step-ahead prediction and non-normality is strongly suspected, the rank correlation test may be the best procedure to employ. Generally, though, it is recommended that the modified Diebold-Mariano approach test should be favoured in application, with all the indications showing that it comprises the best all-round method available for testing for forecast encompassing.

This test, combined with the modified Diebold-Mariano test of chapter 2, yields a unified approach to the companion hypotheses of equal forecast accuracy and forecast encompassing, providing robust evaluation tests which can be reliably applied to different forecast horizons and sample sizes in the presence of a variety of forecast error properties.

Chapter 4

Testing in the Presence of ARCH Errors

4.1 Introduction

In chapters 2 and 3, tests for equal forecast accuracy and tests for forecast encompassing are examined respectively. Following this analysis, tests are proposed which display robustness to non-normality, contemporaneous correlation and autocorrelation in the forecast errors, and behave well in finite samples. One further property that might be expected to be present in economic forecast errors is autoregressive conditional heteroscedasticity (ARCH), as introduced by Engle (1982). Intuitively, ARCH errors imply predictable uncertainty through time; more specifically, that the conditional variance of the errors depends on past squared errors. Such behaviour is common in a number of economic situations, particularly in financial markets; ARCH is to be expected wherever the volatility of a variable appears to vary systematically over time. It is important, then, to consider this property and its effects on the evaluation tests, again focusing on the development of robust and reliable procedures for testing the respective null hypotheses.

This chapter examines the behaviour of the tests for equal forecast accuracy and forecast encompassing when the errors concerned exhibit ARCH, and considers modifications to improve finite sample and asymptotic properties of the tests. Section 4.2 contains the specification of the ARCH errors used in the analysis of tests for equal forecast accuracy, section 4.4 likewise specifies ARCH errors for the analysis of tests for forecast encompassing. Sections 4.3 and 4.5 study the tests for equal forecast accuracy and tests for forecast encompassing respectively, and lastly section 4.6 concludes the chapter.

4.2 ARCH Error Specification and Properties for Equal Accuracy Tests

Before analysis of tests for equal forecast accuracy where the errors follow ARCH processes can commence, the specification of the forecast errors must first be established and a number of properties made explicit. This chapter focuses on errors which follow ARCH(1) processes and the introductory theory of this section follows Hamilton (1994). If two forecasts f_{1t}, f_{2t} are considered with errors e_{1t}, e_{2t} respectively, then ARCH(1) implies that the conditional variance of the errors (conditional on information on past squared errors up to time $t-1$) follows the process:

$$E(e_{it}^2 | e_{i,t-1}^2, e_{i,t-2}^2, \dots) = \alpha_0 + \alpha_1 e_{i,t-1}^2 \quad i = 1, 2 \quad (4.1)$$

Implicit here is the restriction that the errors follow the same ARCH process which is assumed for this section's analysis. Neither this assumption nor the ARCH(1) specification is necessary, they are made purely to illustrate the impact of conditional heteroscedasticity on the tests.

Now for (4.1) to be a sensible representation, constraints of $\alpha_0 > 0$ and $\alpha_1 \geq 0$ must be imposed as e_{it}^2 cannot be negative. Furthermore, stationarity shall be assumed which in this case results in the need for $\alpha_1 < 1$.

It is also possible to think of the squared forecast error series e_{1t}^2, e_{2t}^2 as following AR(1) processes themselves:

$$e_{it}^2 = \alpha_0 + \alpha_1 e_{i,t-1}^2 + w_{it} \quad i = 1, 2 \quad (4.2)$$

where $w_{it} \sim \text{white noise}$

$$E(w_{it}) = 0$$

$$E(w_{it}^2) = \tau^2$$

$$E(w_{it}w_{is}) = 0 \quad s \neq t$$

which is an AR(1) specification for e_{1t}^2, e_{2t}^2 consistent with the conditional variance property in (4.1).

There are two specific cases of ARCH(1) errors which will be studied in the next section (considering tests for equal forecast accuracy), and more detail on the error specifications under each scenario follows. The two cases relate to independent errors and contemporaneously correlated errors, and shall be examined in turn.

4.2.1 Independent ARCH Forecast Errors

In the case where the errors e_{1t}, e_{2t} are independently distributed, an alternative representation for the above ARCH(1) processes can be employed:

$$e_{it} = v_{it} \sqrt{h_{it}} \quad i = 1, 2 \quad (4.3)$$

$$\text{where } h_{it} = \alpha_0 + \alpha_1 e_{i,t-1}^2$$

$$v_{it} \sim \text{IID}(0,1)$$

It shall further be assumed for the purpose of this study that $v_{it} \sim \text{IN}(0,1)$. This representation is useful as it permits discovery of some of the properties of the forecast errors which are needed for the analysis of the tests. The four properties

required are as follows:

$$\begin{aligned}
 1. \quad E(e_{it}^2) &= \alpha_0 + \alpha_1 E(e_{i,t-1}^2) + E(w_{it}) \quad i = 1,2 \quad \text{using (4.2)} \\
 &= \frac{\alpha_0}{1 - \alpha_1} \quad \text{as } e_{it}^2 \text{ stationary} \quad (4.4)
 \end{aligned}$$

$$\begin{aligned}
 2. \quad E(e_{1t}e_{2t}) &= E(e_{1t})E(e_{2t}) \\
 &= 0 \quad (4.5)
 \end{aligned}$$

$$\begin{aligned}
 3. \quad E(e_{it}^4) &= V(e_{it}^2) + [E(e_{it}^2)]^2 \quad i = 1,2 \\
 &= \frac{\tau^2}{1 - \alpha_1^2} + \frac{\alpha_0^2}{(1 - \alpha_1)^2} \quad \text{using (4.2)} \quad (4.6)
 \end{aligned}$$

$$\begin{aligned}
 4. \quad E(e_{1t}^2 e_{2t}^2) &= E(e_{1t}^2)E(e_{2t}^2) \\
 &= \frac{\alpha_0^2}{(1 - \alpha_1)^2} \quad (4.7)
 \end{aligned}$$

It is also possible to derive an expression for τ^2 :

$$e_{it}^2 = h_{it} + w_{it}, \quad e_{it}^2 = v_{it}^2 h_{it} \quad \text{using (4.2) and (4.3)}$$

$$\text{i.e. } h_{it} + w_{it} = v_{it}^2 h_{it}$$

$$w_{it} = h_{it}(v_{it}^2 - 1)$$

$$E(w_{it}^2) = E(h_{it}^2)E(v_{it}^4 - 2v_{it}^2 + 1)$$

$$\text{so } \tau^2 = 2E(h_{it}^2) \quad \text{since } E(v_{it}^4) = 3$$

$$= 2E(\alpha_0^2 + \alpha_1^2 e_{i,t-1}^4 + 2\alpha_0\alpha_1 e_{i,t-1}^2)$$

$$\begin{aligned}
&= \frac{2\tau^2\alpha_1^2}{1-\alpha_1^2} + \frac{2\alpha_0^2}{(1-\alpha_1)^2} \quad \text{using (4.4) and (4.6)} \\
&= \frac{2\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)} \quad (4.8)
\end{aligned}$$

This equation (4.8) only has a real solution if $\alpha_1^2 < 1/3$, and thus the second moment of w_{it} and the fourth moment of e_{it} do not exist if $\alpha_1^2 \geq 1/3$. The results above will be used in the examination of the equal forecast accuracy tests in section 4.3.

4.2.2 Contemporaneously Correlated ARCH Forecast Errors

In addition to examining the case of independent errors, it is important to consider the behaviour of tests when the forecast errors are contemporaneously correlated as such a property would be expected in the real world. There are a number of ways in which correlated ARCH errors could be specified, and two common methods are now studied - one following Engle & Kroner (1995) and one following Bollerslev (1990).

4.2.2a Engle-Kroner Approach

Engle & Kroner (1995) propose the following specification for a multivariate GARCH(p, q) model when there are no exogenous influences on the system:

$$E(e_t e_t' | e_{t-1} e_{t-1}', e_{t-2} e_{t-2}', \dots) = H_t \quad (4.9)$$

$$\text{where } H_t = C_0' C_0 + \sum_{i=1}^q A_i' e_{t-i} e_{t-i}' A_i + \sum_{i=1}^p G_i' H_{t-i} G_i$$

where C_0 is a triangular matrix. Simplifying this to bivariate ARCH(1), the

specification becomes:

$$H_t = C + A_1' e_{t-1} e_{t-1}' A_1 \quad \text{where } e_t = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \quad (4.10)$$

Further simplification is possible by letting A_1 be diagonal. This corresponds to Engle & Kroner's 'diagonal representation', first used by Engle, Granger & Kraft (1984) and Bollerslev, Engle & Wooldridge (1988) for ARCH and GARCH models respectively. If the particular simplification is that of $A_1 = aI$, then (4.10) reduces to:

$$H_t = C + a^2 e_{t-1} e_{t-1}' \quad (4.11)$$

This simplification is both useful and intuitively appealing since the errors e_{1t}, e_{2t} are individually ARCH(1). This follows from the fact that (4.11) implies:

$$E(e_{it}^2 | e_{i,t-1}^2, e_{i,t-1}, \dots) = c_{ii} + a^2 e_{i,t-1}^2 \quad i = 1, 2$$

where $c_{ii} = (i, i)$ 'th element of C

An additional step is to make this correlated specification consistent with that for independent errors by letting $c_{11} = c_{22} = \alpha_0$ and $a = \sqrt{\alpha_1}$. In full, the Engle-Kroner specification for contemporaneously correlated ARCH(1) forecast errors, simplified for this section's analysis can be described as:

$$H_t = \begin{bmatrix} \alpha_0 + \alpha_1 e_{1,t-1}^2 & c_{12} + \alpha_1 e_{1,t-1} e_{2,t-1} \\ c_{12} + \alpha_1 e_{1,t-1} e_{2,t-1} & \alpha_0 + \alpha_1 e_{2,t-1}^2 \end{bmatrix} \quad (4.12)$$

It is also possible to find a result for the element c_{12} . Taking expectations

(unconditionally) in (4.12) gives:

$$E(e_t e_t') = \begin{bmatrix} E(e_{1t}^2) & E(e_{1t} e_{2t}) \\ E(e_{1t} e_{2t}) & E(e_{2t}^2) \end{bmatrix} = \begin{bmatrix} \frac{\alpha_0}{1-\alpha_1} & \frac{c_{12}}{1-\alpha_1} \\ \frac{c_{12}}{1-\alpha_1} & \frac{\alpha_0}{1-\alpha_1} \end{bmatrix}$$

Given that the unconditional correlation between e_{1t} and e_{2t} is ρ :

$$\rho = \frac{E(e_{1t} e_{2t})}{\sqrt{E(e_{1t}^2) E(e_{2t}^2)}}$$

it follows that $c_{12} = \rho\alpha_0$, yielding:

$$H_t = \begin{bmatrix} \alpha_0 + \alpha_1 e_{1,t-1}^2 & \rho\alpha_0 + \alpha_1 e_{1,t-1} e_{2,t-1} \\ \rho\alpha_0 + \alpha_1 e_{1,t-1} e_{2,t-1} & \alpha_0 + \alpha_1 e_{2,t-1}^2 \end{bmatrix} \quad (4.13)$$

$$E(e_t e_t') = \begin{bmatrix} E(e_{1t}^2) & E(e_{1t} e_{2t}) \\ E(e_{1t} e_{2t}) & E(e_{2t}^2) \end{bmatrix} = \begin{bmatrix} \frac{\alpha_0}{1-\alpha_1} & \frac{\rho\alpha_0}{1-\alpha_1} \\ \frac{\rho\alpha_0}{1-\alpha_1} & \frac{\alpha_0}{1-\alpha_1} \end{bmatrix} \quad (4.14)$$

Now, as with the independent errors case, a representation can be derived where the errors are a function of past errors and normal disturbances, $v_{it} \sim \text{IN}(0,1)$, $i = 1,2$:

$$\text{if } e_t = T_{t-1} v_t, \text{ then } H_t = T_{t-1} T_{t-1}'$$

For this to be an alternative representation of the Engle-Kroner correlated ARCH(1) specification, $T_{t-1} T_{t-1}'$ must equal H_t as defined in (4.13). A natural choice for T_{t-1} is then the triangular matrix:

$$T_{t-1} = \begin{bmatrix} t_{11,t-1} & 0 \\ t_{21,t-1} & t_{22,t-1} \end{bmatrix}$$

$$\text{where } t_{11,t-1} = \sqrt{\alpha_0 + \alpha_1 e_{1,t-1}^2}$$

$$t_{21,t-1} = \frac{\rho\alpha_0 + \alpha_1 e_{1,t-1} e_{2,t-1}}{\sqrt{\alpha_0 + \alpha_1 e_{1,t-1}^2}}$$

$$t_{22,t-1} = \sqrt{\alpha_0 + \alpha_1 e_{2,t-1}^2 - \frac{(\rho\alpha_0 + \alpha_1 e_{1,t-1} e_{2,t-1})^2}{\alpha_0 + \alpha_1 e_{1,t-1}^2}}$$

giving the required representation for e_{1t}, e_{2t} :

$$e_{1t} = v_{1t} \sqrt{\alpha_0 + \alpha_1 e_{1,t-1}^2} \quad (4.15)$$

$$e_{2t} = v_{1t} \left(\frac{\rho\alpha_0 + \alpha_1 e_{1,t-1} e_{2,t-1}}{\sqrt{\alpha_0 + \alpha_1 e_{1,t-1}^2}} \right) + v_{2t} \sqrt{\alpha_0 + \alpha_1 e_{2,t-1}^2 - \frac{(\rho\alpha_0 + \alpha_1 e_{1,t-1} e_{2,t-1})^2}{\alpha_0 + \alpha_1 e_{1,t-1}^2}} \quad (4.16)$$

It is useful to note that the $\rho = 0$ Engle-Kroner correlated errors case does not correspond to the case of independent ARCH errors. This follows from the difference in error specifications and is most clearly highlighted by comparing equation (4.3) with (4.15) and (4.16).

Again the four properties required in the following analysis can be found:

$$1. E(e_{it}^2) = \frac{\alpha_0}{1 - \alpha_1} \quad i = 1, 2 \quad \text{from (4.14)} \quad (4.17)$$

$$2. E(e_{1t}e_{2t}) = \frac{\rho\alpha_0}{1-\alpha_1} \quad \text{from (4.14)} \quad (4.18)$$

$$3. E(e_{it}^4) = \frac{\tau^2}{1-\alpha_1^2} + \frac{\alpha_0^2}{(1-\alpha_1)^2} \quad i = 1,2 \quad \text{using (4.2)} \quad (4.19)$$

where τ^2 is as defined in (4.8)

$$4. E(e_{1t}^2 e_{2t}^2) = E\{(v_{1t}h_{11t}^{1/2})^2[v_{1t}h_{11t}^{1/2}h_{12t} + v_{2t}(h_{22t} - h_{11t}^{-1}h_{12t}^2)^{1/2}]\}^2\}$$

using (4.15) and (4.16)

$$= 2E(h_{12t}^2) + E(h_{11t}h_{22t})$$

$$\text{as } E(v_{it}^4) = 3, E(v_{it}^2) = 1, i = 1,2$$

$$v_{1t}, v_{2t}, h_{ijt} \text{ independent}$$

$$= 2E(\rho^2\alpha_0^2 + \alpha_1^2 e_{1,t-1}^2 e_{2,t-1}^2 + 2\rho\alpha_0\alpha_1 e_{1,t-1}e_{2,t-1})$$

$$+ E(\alpha_0^2 + \alpha_1^2 e_{1,t-1}^2 e_{2,t-1}^2 + \alpha_0\alpha_1 e_{1,t-1}^2 + \alpha_0\alpha_1 e_{2,t-1}^2)$$

$$\text{using (4.13)}$$

$$= 3\alpha_1^2 E(e_{1t}^2 e_{2t}^2) + \frac{\alpha_0^2(1+2\rho^2)(1+\alpha_1)}{1-\alpha_1}$$

$$= \frac{\alpha_0^2(1+2\rho^2)(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)} \quad (4.20)$$

4.2.2b Bollerslev Approach

Bollerslev (1990) introduces a second multivariate conditional heteroscedastic specification in which the conditional correlation between e_{1t} and e_{2t} is fixed over time. This Bollerslev approach takes the following form:

$$E(e_t e_t' | e_{t-1} e_{t-1}', e_{t-2} e_{t-2}', \dots) = H_t \quad (4.21)$$

$$\text{where } h_{ijt} = \rho_{ij} \sqrt{h_{iit} h_{jtt}} \quad i = 1, 2; \quad j = 1, 2; \quad i \neq j$$

Denoting the conditional correlation $\rho_{12} = \rho_{21} = \rho_c$, this yields:

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{12t} & h_{22t} \end{bmatrix} \quad (4.22)$$

$$\text{where } h_{11t} = \alpha_0 + \alpha_1 e_{1,t-1}^2$$

$$h_{12t} = \rho_c \sqrt{(\alpha_0 + \alpha_1 e_{1,t-1}^2)(\alpha_0 + \alpha_1 e_{2,t-1}^2)}$$

$$h_{22t} = \alpha_0 + \alpha_1 e_{2,t-1}^2$$

Once again, an alternative representation involving $v_{it} \sim \text{IN}(0,1)$, $i = 1, 2$, can be found, using a method parallel to that for the Engle-Kroner approach:

$$e_t = T_{t-1} v_t, \quad H_t = T_{t-1} T_{t-1}'$$

$$T_{t-1} = \begin{bmatrix} t_{11,t-1} & 0 \\ t_{21,t-1} & t_{22,t-1} \end{bmatrix}$$

$$\text{where } t_{11,t-1} = \sqrt{\alpha_0 + \alpha_1 e_{1,t-1}^2}$$

$$t_{21,t-1} = \rho_c \sqrt{\alpha_0 + \alpha_1 e_{2,t-1}^2}$$

$$t_{22,t-1} = \sqrt{(1 - \rho_c^2)(\alpha_0 + \alpha_1 e_{2,t-1}^2)}$$

This yields the following representation for e_{1t}, e_{2t} :

$$e_{1t} = v_{1t} \sqrt{\alpha_0 + \alpha_1 e_{1,t-1}^2} \quad (4.23)$$

$$e_{2t} = v_{1t} \rho_c \sqrt{\alpha_0 + \alpha_1 e_{2,t-1}^2} + v_{2t} \sqrt{(1 - \rho_c^2)(\alpha_0 + \alpha_1 e_{2,t-1}^2)} \quad (4.24)$$

In this case, the $\rho_c = 0$ Bollerslev correlated errors do correspond to the independent errors case, as is clear from (4.3) and (4.22). The independent errors case is therefore a special case of Bollerslev correlated ARCH errors, but it is still useful for clarity to examine both cases separately.

Finally, the four properties necessary for the analysis of the tests performed in the next section must be derived:

$$1. E(e_{it}^2) = \frac{\alpha_0}{1-\alpha_1} \quad i = 1,2 \quad \text{using (4.2)} \quad (4.25)$$

$$2. E(e_{1t}e_{2t}) = \rho\sqrt{E(e_{1t}^2)E(e_{2t}^2)}$$

where ρ = unconditional correlation

$$= \frac{\rho\alpha_0}{1-\alpha_1} \quad (4.26)$$

$$3. E(e_{1t}^4) = \frac{\tau^2}{1-\alpha_1^2} + \frac{\alpha_0^2}{(1-\alpha_1)^2} \quad \text{using (4.2)} \quad (4.27)$$

$$4. E(e_{1t}^2 e_{2t}^2) = E\{(v_{1t} h_{11t}^{1/2})^2 [\rho_c v_{1t} h_{22t}^{1/2} + v_{2t} (1-\rho_c^2)^{1/2} h_{22t}^{1/2}]^2\}$$

using (4.23) and (4.24)

$$= (1+2\rho_c^2)E(h_{11t}h_{22t})$$

as $E(v_{it}^4) = 3$, $E(v_{it}^2) = 1$, $i = 1,2$

v_{1t}, v_{2t}, h_{ijt} independent

$$= (1+2\rho_c^2)E(\alpha_0^2 + \alpha_1^2 e_{1,t-1}^2 e_{2,t-1}^2$$

$$+ \alpha_0 \alpha_1 e_{1,t-1}^2 + \alpha_0 \alpha_1 e_{2,t-1}^2) \quad \text{using (4.22)}$$

$$\begin{aligned}
&= (1 + 2\rho_c^2)\alpha_1^2 E(e_{1t}^2 e_{2t}^2) + \frac{\alpha_0^2 (1 + 2\rho_c^2)(1 + \alpha_1)}{1 - \alpha_1} \\
&= \frac{\alpha_0^2 (1 + 2\rho_c^2)(1 + \alpha_1)}{[1 - \alpha_1^2 (1 + 2\rho_c^2)](1 - \alpha_1)} \tag{4.28}
\end{aligned}$$

4.3 Tests for Equal Forecast Accuracy

Having established the specification of the forecast errors, it is now possible to evaluate the behaviour of the tests for equal forecast accuracy (considered in chapter 2) when ARCH is present in those errors. The first test examined is the Morgan-Granger-Newbold test, then the parametric non-normality corrections MGN_1^* , MGN_2^* and the rank correlation correction test r_s , plus the modified Diebold-Mariano test S^* . For each test, theoretical analysis of its null distribution under ARCH errors is presented, followed by simulation of its size behaviour in finite samples; the analysis here is restricted to examine 1-step-ahead prediction. Lastly, a new modified test is proposed which exhibits robustness to ARCH forecast errors, and longer forecast horizons are considered.

4.3.1 Morgan-Granger-Newbold Test

4.3.1a Theory

The MGN test of the null of mean squared forecast error equality (a measure of forecast accuracy) takes the formulation given in equation (2.18) of chapter 2. Now as in the examination of MGN 's behaviour under error non-normality, analysis centres around the distribution of the test statistic. Under the null, MGN follows a distribution as given in (2.20), and given the result (2.21) which clearly holds in this case, the starting point for the ARCH analysis is equation (2.22):

$$MGN \xrightarrow{d} N(0, \text{plim}[\hat{Q}^{-1}Q]) \quad (4.29)$$

where $Q = V(n^{-1/2} \sum x_t \varepsilon_t)$

$$\hat{Q} = \hat{\sigma}^2 n^{-1} \sum x_t^2$$

$$x_t = e_{1t} - e_{2t}, \varepsilon_t = y_t = e_{1t} + e_{2t}$$

The analysis then focuses on the estimate \hat{Q} implicit in the *MGN* test statistic under the cases of independent and correlated ARCH(1) errors.

Independent Errors

As noted in section 4.2, independent ARCH errors are a special case of Bollerslev correlated errors (where $\rho_c = 0$), but for exposition this case is considered separately here. When the forecast errors are independent, the true Q can be expressed as follows under the null (note that the chapter 2 assumption of $x_t \varepsilon_t, x_s \varepsilon_s$ uncorrelated for $s \neq t$ cannot be assumed in this case):

$$\begin{aligned} Q &= n^{-1} V(\sum (e_{1t}^2 - e_{2t}^2)) \\ &= n^{-1} V(\sum d_t) \\ &\quad \text{where } d_t = e_{1t}^2 - e_{2t}^2 \\ &= \gamma_0 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \gamma_k \end{aligned} \tag{4.30}$$

where $\gamma_k = C(d_t, d_{t+k}) = E(d_t d_{t+k})$

The term γ_k can be decomposed as follows, beginning with (4.2):

$$\begin{aligned} e_{1t}^2 &= \alpha_0 + \alpha_1 e_{1,t-1}^2 + w_{1t} \\ e_{2t}^2 &= \alpha_0 + \alpha_1 e_{2,t-1}^2 + w_{2t} \\ d_t &= \alpha_1 d_{t-1} + w_{3t} \end{aligned} \tag{4.31}$$

where $w_{3t} \sim \text{white noise}$

$$E(w_{3t}) = 0$$

$$E(w_{3t}^2) = \delta_1^2 \quad (\delta_1^2 = \text{constant})$$

$$E(w_{3t}w_{3s}) = 0 \quad s \neq t$$

$$\gamma_k = E(d_t d_{t+k})$$

$$= \frac{\delta_1^2 \alpha_1^k}{1 - \alpha_1^2}$$

$$= \gamma_0 \alpha_1^k \tag{4.32}$$

Substituting (4.32) back into (4.30) yields:

$$Q = \gamma_0 \left[1 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \alpha_1^k \right] \tag{4.33}$$

Next, γ_0 can be evaluated:

$$\gamma_0 = E(d_t^2)$$

$$= E(e_{1t}^4) + E(e_{2t}^4) - 2E(e_{1t}^2 e_{2t}^2)$$

$$= 2 \left[\frac{\tau^2}{1 - \alpha_1^2} + \frac{\alpha_0^2}{(1 - \alpha_1)^2} \right] - 2 \left[\frac{\alpha_0^2}{(1 - \alpha_1)^2} \right] \quad \text{using (4.6) and (4.7)}$$

$$= \frac{4\alpha_0^2}{(1 - 3\alpha_1^2)(1 - \alpha_1)^2} \quad \text{using (4.8)} \tag{4.34}$$

Also, the bracketed term in (4.33) can be found in the limit:

$$2n^{-1} \sum_{k=1}^{n-1} (n-k) \alpha_1^k = 2 \sum_{k=1}^{n-1} \alpha_1^k - 2n^{-1} \sum_{k=1}^{n-1} k \alpha_1^k$$

$$\text{now } 2 \sum_{k=1}^{n-1} \alpha_1^k = \frac{2\alpha_1(1-\alpha_1^{n-1})}{1-\alpha_1} \longrightarrow \frac{2\alpha_1}{1-\alpha_1} \quad \text{given } 0 \leq \alpha_1 < 1/3$$

$$\text{and } 2n^{-1} \sum_{k=1}^{n-1} k \alpha_1^k \longrightarrow 0 \quad \text{given } 0 \leq \alpha_1 < 1/3$$

$$\text{so } \left[1 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \alpha_1^k \right] \longrightarrow \frac{1+\alpha_1}{1-\alpha_1} \quad (4.35)$$

Collating results (4.33)-(4.35), the limiting value of Q can be found:

$$Q \longrightarrow \frac{4\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)^3} \quad (4.36)$$

The second part of the theory relates to finding the probability limit of the *MGN* estimate of Q under H_0 :

$$\begin{aligned} \hat{Q} &\xrightarrow{p} V(\varepsilon_t)E(x_t^2) \\ &= [E(e_{1t}^2) + E(e_{2t}^2) + 2E(e_{1t}e_{2t})] \\ &\quad \times [E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})] \\ &= \frac{4\alpha_0^2}{(1-\alpha_1)^2} \quad \text{using (4.4) and (4.5)} \end{aligned} \quad (4.37)$$

The expressions given in (4.29), (4.36) and (4.37) can now be used to determine the asymptotic distribution of the *MGN* test statistic under independent errors:

$$MGN \xrightarrow{d} N\left(0, \frac{1+\alpha_1}{(1-3\alpha_1^2)(1-\alpha_1)}\right) \quad (4.38)$$

It is clear, therefore, that the *MGN* test is not robust to ARCH in the forecast errors, and will be incorrectly sized in the limit. Numerical integration can be performed to

find the exact degree of the problem for different ARCH parameters α_1 , and table 4.1a shows a number of *MGN* theoretical asymptotic test sizes for a 2-sided test at the nominal 10% level for these independent errors, and also for correlated errors studied below. The variance term in (4.38) does not exist for $\alpha_1^2 \geq 1/3$ and thus α_1 parameter values are only considered up to this boundary. It can be seen from the entries in the table that the problem is severe. Even with moderate amounts of ARCH, e.g. $\alpha_1 = 0.3$, *MGN* has an asymptotic test size of 30.24% which is totally unsatisfactory and requires correction.

Engle-Kroner Correlated Errors

Considering now the case where the errors are correlated as specified by the simplified Engle-Kroner approach, parallel theory to that conducted above can be written down, beginning by analysing the true Q under the null, which again takes the form given in (4.33). Now in this case, γ_0 becomes:

$$\begin{aligned}
 \gamma_0 &= E(e_{1t}^4) + E(e_{2t}^4) - 2E(e_{1t}^2 e_{2t}^2) \\
 &= 2 \left[\frac{\tau^2}{1 - \alpha_1^2} + \frac{\alpha_0^2}{(1 - \alpha_1)^2} \right] - 2 \left[\frac{\alpha_0^2 (1 + 2\rho^2)(1 + \alpha_1)}{(1 - 3\alpha_1^2)(1 - \alpha_1)} \right] \\
 &\quad \text{using (4.19) and (4.20)} \\
 &= \frac{4\alpha_0^2 (1 - \rho^2)(1 + \alpha_1)}{(1 - 3\alpha_1^2)(1 - \alpha_1)} \quad \text{using (4.8)} \tag{4.39}
 \end{aligned}$$

which, combined with (4.35), leads to the following limiting expression for Q :

Table 4.1a

Theoretical asymptotic sizes for the *MGN* test at the nominal 10% level

Error specification	α_1	Distribution variance	Asymptotic test size
Independent	0.1	1.2600	14.28
	0.2	1.7045	20.77
	0.3	2.5440	30.24
	0.4	4.4872	43.75
	0.5	12	63.49
Engle-Kroner correlated ($\forall \rho$)	0.1	1.2474	14.08
	0.2	1.6364	19.85
	0.3	2.3151	27.97
	0.4	3.7692	39.69
	0.5	9	58.35
Bollerslev correlated ($\rho_c = 0.5$)	0.1	1.2636	14.34
	0.2	1.7262	21.06
	0.3	2.6260	31.01
	0.4	4.7911	45.24
	0.5	13.6071	65.57
Bollerslev correlated ($\rho_c = 0.9$)	0.1	1.2714	14.46
	0.2	1.7759	21.71
	0.3	2.8342	32.86
	0.4	5.7303	49.20
	0.5	21.0692	72.01

Note:- Calculations of *MGN* theoretical sizes with Bollerslev correlated ARCH(1) errors use simulated values of ρ , as given in table 4.1b.

Table 4.1b

Simulated values of ρ under Bollerslev correlated ARCH(1) errors

α_1	$\rho_c = 0.5$	$\rho_c = 0.9$
0.1	0.4983	0.8992
0.2	0.4936	0.8969
0.3	0.4854	0.8927
0.4	0.4729	0.8859
0.5	0.4542	0.8745

$$Q \longrightarrow \frac{4\alpha_0^2(1-\rho^2)(1+\alpha_1)^2}{(1-3\alpha_1^2)(1-\alpha_1)^2} \quad (4.40)$$

As with the independent error case, the probability limit of \hat{Q} must also be found:

$$\begin{aligned} \hat{Q} &\xrightarrow{p} [E(e_{1t}^2) + E(e_{2t}^2) + 2E(e_{1t}e_{2t})] \\ &\quad \times [E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})] \\ &= \frac{4\alpha_0^2(1-\rho^2)}{(1-\alpha_1)^2} \quad \text{using (4.17) and (4.18)} \end{aligned} \quad (4.41)$$

Combining (4.29), (4.40) and (4.41) yields the asymptotic distribution of *MGN* when the errors are Engle-Kroner correlated ARCH(1):

$$MGN \xrightarrow{d} N\left(0, \frac{(1+\alpha_1)^2}{1-3\alpha_1^2}\right) \quad (4.42)$$

As might be expected, the *MGN* test is found to be incorrectly sized, with the limiting test distribution being normal with a variance greater than one; in this simplified Engle-Kroner specification the variance is invariant to the amount of contemporaneous correlation. The degree of the problem can be seen in table 4.1a, and as before a very serious lack of robustness to ARCH errors is displayed, reinforcing the motivation for finding a new correctly sized test.

Bollerslev Correlated Errors

The final scenario requiring examination is that of contemporaneously correlated ARCH(1) errors as specified by Bollerslev. Once again the first stage is to find an expression for Q :

$$\gamma_0 = E(e_{1t}^4) + E(e_{2t}^4) - 2E(e_{1t}^2 e_{2t}^2)$$

$$\begin{aligned}
&= 2 \left[\frac{\tau^2}{1-\alpha_1^2} + \frac{\alpha_0^2}{(1-\alpha_1)^2} \right] - 2 \left[\frac{\alpha_0^2(1+2\rho_c^2)(1+\alpha_1)}{[1-\alpha_1^2(1+2\rho_c^2)](1-\alpha_1)} \right] \\
&\quad \text{using (4.27) and (4.28)} \\
&= \frac{4\alpha_0^2(1-\rho_c^2)(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)[1-\alpha_1^2(1+2\rho_c^2)]} \tag{4.43}
\end{aligned}$$

$$\begin{aligned}
\text{so } Q &\longrightarrow \frac{4\alpha_0^2(1-\rho_c^2)(1+\alpha_1)^2}{(1-3\alpha_1^2)(1-\alpha_1)^2[1-\alpha_1^2(1+2\rho_c^2)]} \tag{4.44} \\
&\quad \text{using (4.33)}
\end{aligned}$$

Turning now to the estimator of \hat{Q} :

$$\begin{aligned}
\hat{Q} &\xrightarrow{p} [E(e_{1t}^2) + E(e_{2t}^2) + 2E(e_{1t}e_{2t})] \\
&\quad \times [E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})] \\
&= \frac{4\alpha_0^2(1-\rho^2)}{(1-\alpha_1)^2} \quad \text{using (4.25) and (4.26)} \tag{4.45}
\end{aligned}$$

The asymptotic distribution of the *MGN* test statistic can now be found for the Bollerslev approach to specifying contemporaneously correlated ARCH(1) errors, using (4.29), (4.44) and (4.45):

$$MGN \xrightarrow{d} N \left(0, \frac{(1+\alpha_1)^2(1-\rho_c^2)}{(1-3\alpha_1^2)[1-\alpha_1^2(1+2\rho_c^2)](1-\rho^2)} \right) \tag{4.46}$$

This more complex result again shows the lack of *MGN* robustness to ARCH errors. The asymptotic test size now depends on the conditional correlation, ρ_c , and also on the unconditional correlation, ρ . It can again be noted that independent ARCH errors are a special case of this result, with (4.46) reducing to (4.38) when $\rho_c = 0$,

since this implies $\rho = 0$ (found by taking expectations in (4.21)).

Now for a given conditional correlation, the unconditional correlation has no analytic expression. This follows from:

$$\begin{aligned}\rho &= \frac{E(e_{1t}e_{2t})}{\sqrt{E(e_{1t}^2)E(e_{2t}^2)}} \\ &= \rho_c \left(\frac{1-\alpha_1}{\alpha_0} \right) E \left(\sqrt{(\alpha_0 + \alpha_1 e_{1,t-1}^2)(\alpha_0 + \alpha_1 e_{2,t-1}^2)} \right) \\ &\quad \text{using (4.22) and (4.25)}\end{aligned}$$

where the expectation of the square root term cannot be expressed analytically. The unconditional correlation, ρ , must therefore be simulated.

Table 4.1a contains the theoretical test sizes for *MGN* for these Bollerslev correlated errors, with ρ being simulated using 10,000 replications of a sample size of 10,000, the results of these simulations being given in table 4.1b. The test again exhibits considerable oversizing in the limit, as with the previous cases of error specification.

4.3.1b Simulation

It is now important to simulate the behaviour of *MGN* to establish its finite sample size properties and also to check the theoretical asymptotic sizes. Simulation can be performed by using equations (4.3) for independent errors, (4.15) and (4.16) for Engle-Kroner correlated errors, and (4.23) and (4.24) for Bollerslev correlated errors. A value of $\alpha_0 = 0.2$ is used - α_0 does not affect the limiting test size, and does not distort the general picture in finite samples. In each case of ARCH(1)

errors, simulations are performed for sample sizes between $n = 8$ and $n = 512$, and in some representative cases $n = 10,000$ to check the asymptotics. To minimise the effects of the starting value, $n + 100$ observations are simulated and then the first 100 are discarded before the test is applied. All the experiments are evaluated at the nominal 10% level with a null of equal mean squared forecast errors against a 2-sided alternative. One-step-ahead forecasts ($h = 1$) are solely considered, and 10,000 replications are performed for each simulation.

Table 4.2 gives simulation results for independent errors for a range of ARCH parameters. These results clearly confirm the theoretical conclusion of lack of test robustness to ARCH errors. In all cases where α_1 is positive (i.e. ARCH is present), *MGN* is oversized, to greater degrees as the amount of ARCH increases. Similar results are found for the Engle-Kroner and Bollerslev correlated cases - simulations for these are performed for $\alpha_1 = 0.2$ and $\alpha_1 = 0.4$, and are given in tables 4.3 and 4.4 (along with other tests which are considered below). The asymptotics are confirmed (tables 4.2 and 4.3), with the $n = 10,000$ results coming very close to the theoretical limiting sizes given in table 4.1a - less so as the ARCH parameter approaches the $\alpha_1^2 = 1/3$ boundary.

The foremost inference from all these results is that *MGN* is not a valid test under any conditions where the forecast errors follow ARCH processes, and correction is imperative.

Table 4.2

Empirical sizes for the *MGN* test at the nominal 10% level
 ($h = 1$, independent ARCH(1) errors)

α_1	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
0.0	9.90	10.02	10.01	10.35	10.13	10.09	9.82	10.13
0.1	12.09	12.92	13.75	14.03	14.37	14.36	14.23	14.31
0.2	14.56	16.87	18.35	19.20	19.99	20.04	20.26	20.63
0.3	17.90	21.41	24.15	26.35	27.38	28.50	28.34	29.76
0.4	21.40	26.32	29.87	33.32	35.59	37.44	38.64	42.02
0.5	24.82	31.55	35.92	41.19	43.85	46.87	49.58	56.52

Table 4.3

Empirical sizes for the MGN , MGN_1^* , MGN_2^* , r_s , S^* and LS^* tests
at the nominal 10% level ($h = 1$, ARCH(1) errors, $\alpha_1 = 0.2$)

Ind.	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
MGN	14.56	16.87	18.35	19.20	19.99	20.04	20.26	20.63
MGN_1^*	25.35	22.86	21.17	19.87	19.29	18.53	18.25	17.96
MGN_2^*	4.62	11.65	15.02	16.35	17.44	17.51	17.79	17.93
r_s	11.56	15.41	16.28	16.76	16.87	16.94	16.52	17.39
S^*	11.89	14.85	16.24	16.98	17.67	17.68	17.86	17.93
LS^*	14.79	12.96	11.92	11.79	10.81	11.13	10.15	10.44

EK 0.5	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
MGN	14.79	17.04	17.63	18.50	19.48	19.13	19.44	19.94
MGN_1^*	25.45	23.60	20.95	19.52	18.99	18.32	17.95	17.66
MGN_2^*	4.50	11.70	14.71	16.36	17.12	17.43	17.47	17.64
r_s	11.08	15.67	15.78	16.56	16.51	16.77	16.80	17.25
S^*	11.64	14.87	16.29	16.94	17.41	17.57	17.54	17.64
LS^*	14.45	13.17	11.69	11.69	10.46	11.04	9.97	9.97

EK 0.9	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
MGN	14.92	16.90	17.31	18.87	19.18	18.99	19.76	19.74
MGN_1^*	25.33	23.26	20.59	19.87	18.99	18.20	18.18	17.85
MGN_2^*	4.74	11.89	14.58	16.56	17.07	17.06	17.70	17.82
r_s	11.53	15.45	15.56	16.40	16.77	16.70	16.95	16.90
S^*	11.85	15.09	15.98	17.26	17.30	17.24	17.72	17.84
LS^*	14.88	13.56	11.63	11.83	10.80	10.40	10.57	10.07

BV 0.5	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
MGN	14.64	17.20	18.16	19.50	20.03	20.22	20.47	20.95
MGN_1^*	25.64	23.39	21.33	19.64	19.04	18.26	17.98	17.79
MGN_2^*	4.44	11.60	14.80	16.21	17.08	17.26	17.42	17.79
r_s	11.12	15.33	16.13	16.57	16.73	17.17	16.77	17.20
S^*	11.60	14.73	16.29	16.76	17.38	17.40	17.50	17.79
LS^*	13.87	13.02	11.65	11.40	10.52	10.71	10.15	10.09

BV 0.9	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
MGN	14.51	16.76	17.67	20.39	20.04	20.33	21.42	21.52
MGN_1^*	25.56	23.43	20.92	20.30	18.79	17.98	18.54	18.02
MGN_2^*	4.64	11.54	14.29	16.60	16.93	17.14	17.87	18.00
r_s	11.43	15.31	15.87	16.88	16.87	16.76	17.20	17.02
S^*	11.80	15.02	15.62	17.11	17.07	17.26	17.94	18.00
LS^*	14.67	13.39	11.33	11.89	10.41	10.58	10.50	10.03

Note:- ‘Ind.’ denotes independent errors, ‘EK 0.5’ denotes Engle-Kroner correlated errors where $\rho = 0.5$, ‘EK 0.9’ denotes Engle-Kroner correlated errors where $\rho = 0.9$, ‘BV 0.5’ denotes Bollerslev correlated errors where $\rho_c = 0.5$, and ‘BV 0.9’ denotes Bollerslev correlated errors where $\rho_c = 0.9$.

Table 4.4

Empirical sizes for the MGN , MGN_1^* , MGN_2^* , r_s , S^* and LS^* tests
at the nominal 10% level ($h = 1$, ARCH(1) errors, $\alpha_1 = 0.4$)

Ind.	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
MGN	21.40	26.32	29.87	33.32	35.59	37.44	38.64
MGN_1^*	32.39	31.76	30.71	29.90	29.18	28.73	27.84
MGN_2^*	6.57	16.24	21.08	23.71	25.59	26.47	26.31
r_s	15.52	22.01	23.84	24.48	25.53	25.35	24.77
S^*	15.43	19.92	22.99	24.58	25.94	26.70	26.44
LS^*	14.42	12.77	12.43	11.25	10.63	10.41	10.23

EK 0.5	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
MGN	21.62	25.86	27.86	30.71	32.30	33.70	35.36
MGN_1^*	32.34	31.79	30.25	29.56	28.96	27.89	28.02
MGN_2^*	6.15	16.44	21.16	23.96	25.62	25.87	26.79
r_s	15.64	21.77	23.20	24.00	24.31	24.55	24.41
S^*	15.32	20.40	23.02	24.83	25.94	26.03	26.85
LS^*	13.84	12.93	12.43	11.20	10.56	10.65	10.13

EK 0.9	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
MGN	21.22	25.44	27.38	30.50	33.06	33.70	35.92
MGN_1^*	32.00	31.55	29.08	29.76	28.75	28.03	28.01
MGN_2^*	6.27	16.77	21.19	24.29	25.61	26.04	26.74
r_s	15.41	21.71	22.68	23.92	24.18	24.69	24.74
S^*	15.43	20.70	22.72	25.06	26.01	26.22	26.80
LS^*	14.59	13.21	12.36	11.25	11.13	10.18	10.00

BV 0.5	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
MGN	21.00	26.52	30.17	33.34	35.93	38.28	39.67
MGN_1^*	32.65	31.90	30.44	30.64	29.33	28.42	28.10
MGN_2^*	5.94	15.91	20.87	23.70	25.05	25.84	26.32
r_s	15.59	21.89	23.94	24.52	25.07	25.66	25.21
S^*	14.96	20.21	22.70	24.54	25.39	25.99	26.39
LS^*	13.38	12.88	12.28	11.08	10.49	10.11	9.91

BV 0.9	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
MGN	20.99	26.35	29.91	34.23	37.85	40.14	41.34
MGN_1^*	32.37	32.32	30.22	30.53	29.87	29.37	28.37
MGN_2^*	5.93	15.65	19.77	23.47	25.15	25.75	26.22
r_s	15.46	21.72	22.91	24.81	25.26	25.33	25.50
S^*	15.27	20.01	21.53	24.31	25.41	26.04	26.28
LS^*	13.70	12.47	11.64	10.99	10.19	10.34	9.53

Note:- ‘Ind.’ denotes independent errors, ‘EK 0.5’ denotes Engle-Kroner correlated errors where $\rho = 0.5$, ‘EK 0.9’ denotes Engle-Kroner correlated errors where $\rho = 0.9$, ‘BV 0.5’ denotes Bollerslev correlated errors where $\rho_c = 0.5$, and ‘BV 0.9’ denotes Bollerslev correlated errors where $\rho_c = 0.9$.

4.3.2 Non-Normality Robust Tests

4.3.2a Theory

The next tests to consider are the chapter 2 corrections to MGN , providing alternative testing procedures which were robust to non-normality. Firstly, it is possible to establish the asymptotic distributions of the two variance-correction tests MGN_1^*, MGN_2^* (given in equations (2.44) and (2.49) respectively) when the errors are ARCH(1) in the same way as for MGN :

$$MGN_1^* \xrightarrow{d} N(0, \text{plim}[\hat{Q}_{m1}^{-1}Q]) \quad (4.47)$$

$$MGN_2^* \xrightarrow{d} N(0, \text{plim}[\hat{Q}_{m2}^{-1}Q]) \quad (4.48)$$

$$\begin{aligned} \text{where } \hat{Q}_{m1} &= n^{-1} \sum x_t^2 \hat{\varepsilon}_t^2 \\ \hat{Q}_{m2} &= n^{-1} \sum x_t^2 y_t^2 \end{aligned}$$

Now the true element Q can be written under the null as in (4.30) or (4.33) for all the considered ARCH specifications, and these two equations lead to two useful results below. Firstly, the probability limits of $\hat{Q}_{m1}, \hat{Q}_{m2}$ can be found under H_0 as follows:

$$\begin{aligned} \hat{Q}_{m1}, \hat{Q}_{m2} &\xrightarrow{p} E(x_t^2 \varepsilon_t^2) \\ &= E[(e_{1t} - e_{2t})^2 (e_{1t} + e_{2t})^2] \\ &= E(e_{1t}^4) + E(e_{2t}^4) - 2E(e_{1t}^2 e_{2t}^2) \\ &= E(d_t^2) \\ &= \gamma_0 \end{aligned} \quad (4.49)$$

Now substituting results (4.30) and (4.49) into (4.47) and (4.48) gives one version of

the distribution results for MGN_1^* and MGN_2^* :

$$MGN_1^*, MGN_2^* \xrightarrow{d} N\left(0, \lim_{n \rightarrow \infty} \left[1 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \rho_k\right]\right) \quad (4.50)$$

$$\text{where } \rho_k = \text{Corr}(d_t, d_{t+k})$$

The interesting point to be made here is that the tests will only be correctly sized if all the autocorrelations of d_t are zero. Problems will therefore occur if d_t exhibits any autocorrelation, with ARCH being an example of this more general conclusion.

If results (4.33) and (4.49) are now substituted into (4.47) and (4.48), an explicit result for the MGN_1^*, MGN_2^* null distribution is found:

$$MGN_1^*, MGN_2^* \xrightarrow{d} N\left(0, \lim_{n \rightarrow \infty} \left[1 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \alpha_1^k\right]\right) \quad (4.51)$$

$$\text{i.e. } MGN_1^*, MGN_2^* \xrightarrow{d} N\left(0, \frac{1+\alpha_1}{1-\alpha_1}\right) \quad \text{using (4.35)} \quad (4.52)$$

The asymptotic distribution variance of the test statistics is the same under independent, Engle-Kroner correlated and Bollerslev correlated errors, and a lack of robustness to ARCH forecast errors is displayed due to the autocorrelation present in the loss differential d_t .

The rank correlation variant of the MGN test, as defined in (2.51), cannot be examined in such a clear way theoretically, and inference concerning this test is left to simulation.

The remaining alternative test to *MGN* when testing for equal forecast accuracy is the preferred test of chapter 2 - the modified Diebold-Mariano test, S^* . For 1-step-ahead errors, this test's formulation is:

$$S^* = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \quad (4.53)$$

$$\text{where } \bar{d} = n^{-1} \sum d_t$$

$$\hat{V}(\bar{d}) = n^{-1}(n-1)^{-1} \sum (d_t - \bar{d})^2$$

and the asymptotic distribution can be written as:

$$S^* \xrightarrow{d} N(0, \text{plim}[\hat{V}(\bar{d})^{-1} V(\bar{d})]) \quad (4.54)$$

Once again, the issue of the test's behaviour centres around variance estimation, and it is necessary to find expressions for $V(\bar{d})$ and its estimate in the limit:

$$\begin{aligned} nV(\bar{d}) &= nV(n^{-1} \sum d_t) \\ &= \gamma_0 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \gamma_k \end{aligned} \quad (4.55)$$

$$= \gamma_0 \left[1 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \alpha_1^k \right] \quad (4.56)$$

$$\begin{aligned} n\hat{V}(\bar{d}) &= (n-1)^{-1} \sum (d_t - \bar{d})^2 \\ &\xrightarrow{p} E[(d_t - \bar{d})^2] \end{aligned}$$

$$\begin{aligned}
&= E(d_t^2) + \bar{d}^2 \\
&\xrightarrow{p} \gamma_0
\end{aligned} \tag{4.57}$$

Collating (4.54)-(4.57) gives two results for the distribution of S^* under H_0 , parallel to those of (4.50) and (4.52):

$$S^* \xrightarrow{d} N\left(0, \lim_{n \rightarrow \infty} \left[1 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \rho_k\right]\right) \tag{4.58}$$

$$S^* \xrightarrow{d} N\left(0, \frac{1+\alpha_1}{1-\alpha_1}\right) \tag{4.59}$$

S^* therefore has the same asymptotic distribution as MGN_1^* and MGN_2^* in this equal mean squared forecast error null case, for independent, Engle-Kroner correlated and Bollerslev correlated ARCH(1) errors.

Interpreting results (4.52) and (4.59), it can be seen that problems caused by ARCH errors persist in these tests. Some of the effects are removed and the distortion is less severe than for MGN , but these previously robust tests are found to be oversized in the limit when ARCH is present. The factor remaining which drives the problem relates not to the leptokurtosis inherent in the ARCH specification, as these tests are robust to such effects (this also explains why MGN_1^* , MGN_2^* and S^* improve on MGN which is not robust to leptokurtosis), but relates to the autocorrelation in d_t (or $x_t \varepsilon_t$ depending on the view point: $d_t = x_t \varepsilon_t$ under H_0).

Table 4.5 reports theoretical asymptotic test sizes for MGN_1^* , MGN_2^* and S^* for

Table 4.5

Theoretical asymptotic sizes for the MGN_1^* , MGN_2^* and S^* tests at the nominal 10% level with independent, Engle-Kroner correlated and Bollerslev correlated ARCH(1) errors

α_1	Distribution variance	Asymptotic test size
0.1	1.2222	13.68
0.2	1.5	17.93
0.3	1.8571	22.74
0.4	2.3333	28.16
0.5	3	34.23

different α_1 , and highlights the conclusion that the problem of ARCH remains, albeit to a lesser degree. The two points of particular importance are that under ARCH errors, the autocorrelation in d_t causes problems for the tests beyond merely the introduction of non-normality, and that the limiting distributions of all the non-normality robust tests are both the same and oversized. It is clear that a new test must be derived to take account of the autocorrelation in d_t and provide a robust alternative to all the tests so far considered.

4.3.2b Simulation

Simulation of the four tests MGN_1^* , MGN_2^* , r_s and S^* allows examination of their finite sample properties under an ARCH error specification. Tables 4.3 and 4.4 contain results of simulation experiments for independent and correlated (by both Engle-Kroner and Bollerslev models) errors for ARCH(1) parameters $\alpha_1 = 0.2, 0.4$.

A number of points are clear. Firstly, the $n = 10,000$ asymptotic check simulation confirms the theory of the previous sub-section, with MGN_1^* , MGN_2^* and S^* test statistics all converging to the theoretical limiting size. A second observation is that the rank correlation variant of MGN , the r_s test, also appears to have the same asymptotic distribution as the other three tests, the simulations giving every indication of convergence to the same limit. Lastly, the theoretical conclusion that the problem of ARCH remains in these ‘corrected’ tests is very much true empirically, verifying the analysis. None of the tests are correctly sized for any sample size under any error specification. All four tests converge to an incorrect size, MGN_1^* from above, MGN_2^* , r_s and S^* from below (as might be expected

from their properties discovered in chapter 2). Overall, oversizing is the predominant and most prevalent feature of all the simulations, confirming the conclusion to the theory that a new test is required.

4.3.3 A Modified Test

The theoretical and simulation results of the previous sub-sections give rise to the conclusion that none of the tests for equal forecast accuracy so far considered are robust to forecast errors which follow an ARCH specification.

In order to develop such a robust test, the source of the problem must be found. It was noted in sub-section 4.3.2a that the problem of incorrect asymptotic test size for the MGN_1^* , MGN_2^* and S^* tests was the presence of autocorrelation in d_t . Consistent variance estimation in the test statistics and hence a correct limiting test size can therefore only be achieved by taking account of this autocorrelation.

Examining S^* in detail shows that the true variance can be written as in (4.55) with $n-1$ autocovariance lags. The 1-step-ahead S^* test does not include any of these lags in its estimation of the true variance, and as a result has probability limit as given in (4.58). The most obvious possibility for correcting S^* for ARCH errors is thus to add autocovariance lags to the variance estimate even in the 1-step-ahead case. This is then effectively equivalent to running the S^* h -steps-ahead test ($h > 1$) for 1-step-ahead errors, and this way the autocorrelation in d_t might be picked up sufficiently to generate an approximately correctly sized test.

With regard to MGN_1^* and MGN_2^* , the autocorrelation present in $x_t \varepsilon_t$ ($= d_t$) can

be accounted for by including estimation of some of the autocovariances, i.e. the new tests would have new estimators of Q as follows:

modified MGN_1^* :

$$\hat{Q}_{m3} = n^{-1} \sum_{t=1}^n x_t^2 \hat{\varepsilon}_t^2 + 2n^{-1} \sum_{k=1}^q \sum_{t=1}^{n-k} x_t^2 \hat{\varepsilon}_t^2 x_{t+k}^2 \hat{\varepsilon}_{t+k}^2 \quad (4.60)$$

modified MGN_2^* :

$$\hat{Q}_{m4} = n^{-1} \sum_{t=1}^n x_t^2 y_t^2 + 2n^{-1} \sum_{k=1}^q \sum_{t=1}^{n-k} x_t^2 y_t^2 x_{t+k}^2 y_{t+k}^2 \quad (4.61)$$

where q is the number of autocovariance lags included. However, given that MGN_1^* and MGN_2^* exhibit size problems even when no ARCH is present, and that S^* is the preferred test for equal forecast accuracy in such cases, it makes most sense to consider modifying S^* .

Simulation can be performed as before to determine the success or otherwise of this S^* modification. Table 4.6 gives simulation results for S^* with additional lags added to study the effects of the new test. For a given number of lags, the finite sample correction part of S^* expounded in chapter 2 is altered to maintain approximate unbiasedness in the estimation of the variance of \bar{d} , e.g. for 3 additional lags, this is equivalent to a 4-steps-ahead test and the correction factor is $n(n-7+12n^{-1})^{-1}$. Lags from zero (unadjusted S^*) to six are considered, and it can be seen that for each sample size, adding lags to the original S^* test initially

Table 4.6

Empirical sizes for the S^* test with additional lags
at the nominal 10% level ($h = 1$, independent ARCH(1) errors, $\alpha_1 = 0.2$)

	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
0 lags	11.89	14.85	16.24	16.98	17.67	17.68	17.86
1 lags	14.79	12.96	11.92	11.41	11.27	11.56	11.04
2 lags	17.84	16.66	13.73	11.79	10.81	10.95	10.22
3 lags	17.22	18.88	16.12	12.59	11.19	11.13	10.01
4 lags	14.14	20.64	17.75	13.93	11.91	11.11	10.15
5 lags	10.91	20.46	19.08	15.32	12.33	11.27	10.26
6 lags	10.29	19.90	20.25	16.27	12.92	11.49	10.44

improves the test size, and then the cost of including greater numbers of lags outweighs their contribution to consistent variance estimation and the test size becomes worse again.

It can be concluded, then, that this modification of adding covariance lags to S^* is successful in attaining a correct test size and consequently robustness to ARCH errors. However, the question still remains of how many lags to include - table 4.6 shows that the optimal number of lags would increase with the sample size, starting with 1 lag for $n = 8$ and rising non-linearly to 3 or 4 lags for $n = 512$. For a given application, though, the degree of ARCH will be unknown, and some general purpose lag selection rule is required.

One of the most established lag selection criteria for covariance estimation is the Newey-West procedure, described in Newey & West (1994). In this paper they propose a rule which chooses the number of lags, or the bandwidth, m according to:

$$m = [g \times n^{1/3}] \quad \text{where '[]' denotes 'integer part of'} \quad (4.62)$$

Newey & West, following Priestley (1981), *inter alia*, give a data dependent method for choosing the optimal value of the parameter g in (4.62). Furthermore, the Newey-West procedure involves use of the Bartlett kernel to weight the included autocovariance lags. Application of this procedure to the problem of selecting additional lags to S^* does not prove favourable. The numbers of lags suggested by the criterion are somewhat unpredictable due to the data dependency of the parameter, and are not monotonically increasing with the sample size.

A more appropriate selection criterion can be found by taking the Newey-West rule

as given in (4.62) and specifying the parameter arbitrarily. Such an approach avoids data dependency and can yield sensible bandwidths which fit with the intuition of optimal lag selection derived from observing table 4.6.

As for the kernel, Newey & West conclude from their simulations that the ‘choice of kernel is of secondary importance’. An alternative kernel which is useful in this case (to maintain variance estimation approximate unbiasedness and equally weight included covariance lags) is what Christiano & den Haan (1996) refer to as the unweighted, truncated kernel (following Hansen & Hodrick, 1980 and White, 1984), i.e. the kernel is equal to one for all included lags.

One point to note is that in the case of ARCH errors, the squared error process is AR(1). Now in chapters 2 and 3, d_t follows an MA process of known order which gives strong justification for using the unweighted truncated kernel as employed in the multi-step-ahead S^* test. When ARCH errors are present this justification is lost; however, it is valuable to maintain the Diebold-Mariano-type structure and simply add lags to S^* as is done when using multi-step-ahead forecasts. This is then equivalent to approximating the AR(1) process with an MA process of low order, and the ARCH effects should be picked up without need for altering the whole test structure.

With regard to choosing this arbitrary bandwidth, the minimum value for the parameter g which gives $n \geq 1$ for all sample sizes considered is 0.5. This and a parameter value of 0.75 give sensible but different bandwidths and the new tests derived by including lags according to these rules are simulated, the results being

shown in table 4.7b. Table 4.7a gives the bandwidths suggested by the two criteria for the different sample sizes, plus entries for MGN and S^* for the purpose of comparison. The simulation results are very encouraging, with good size properties being found for both criteria across different ARCH parameters and different sample sizes. Wherever ARCH is present, the modification under both criteria improves substantially on the MGN and S^* tests (with the trivial exception of when there is very little ARCH and the sample size is very small). Choosing between the two criteria, on the whole where they differ in their bandwidth selection, the $[0.5n^{1/3}]$ criterion appears to outperform the $[0.75n^{1/3}]$ criterion, leading to the conclusion that it should be preferred and used as the general purpose lag selection rule.

The analysis thus far has concentrated on 1-step-ahead errors alone. It is already known from chapter 2 that S^* is robust to autocorrelation in the forecast errors, and achieves this property by adding $h-1$ lags for an h -steps-ahead forecast. It makes sense, therefore, to use $m+h-1$ lags for the new test, where h is the forecast horizon and m the bandwidth included to capture the effects of ARCH.

Collating these results and conclusions, a new modified test for equal forecast accuracy can be formulated:

$$LS^* = \frac{\bar{d}}{\sqrt{\hat{V}_L(\bar{d})}} \quad (4.63)$$

Table 4.7a

Bandwidths suggested by alternative selection criteria

	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$[0.5n^{1/3}]$	1	1	1	2	2	3	4
$[0.75n^{1/3}]$	1	1	2	3	3	4	6

Table 4.7b

Empirical sizes for MGN , S^* , and the S^* test with additional lags
 (bandwidth chosen by alternative selection criteria)
 at the nominal 10% level ($h = 1$, independent ARCH(1) errors)

α_1	Test	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
0.0	MGN	9.90	10.02	10.01	10.35	10.13	10.09	9.82
	S^*	8.31	9.57	10.12	10.12	10.06	10.32	9.80
	$S^* 0.5$	15.68	14.16	12.11	12.63	11.15	10.93	10.34
	$S^* 0.75$	15.68	14.16	15.12	13.61	11.47	11.16	10.91
0.1	MGN	12.09	12.92	13.75	14.03	14.37	14.36	14.23
	S^*	10.18	11.96	13.41	13.30	13.95	13.83	13.72
	$S^* 0.5$	15.19	13.43	12.04	12.14	10.87	11.11	10.43
	$S^* 0.75$	15.19	13.43	14.62	13.14	11.39	11.28	10.67
0.2	MGN	14.56	16.87	18.35	19.20	19.99	20.04	20.26
	S^*	11.89	14.85	16.24	16.98	17.67	17.68	17.86
	$S^* 0.5$	14.79	12.96	11.92	11.79	10.81	11.13	10.15
	$S^* 0.75$	14.79	12.96	13.73	12.59	11.19	11.11	10.44
0.3	MGN	17.90	21.41	24.15	26.35	27.38	28.50	28.34
	S^*	13.61	17.55	19.57	20.72	21.88	22.19	21.96
	$S^* 0.5$	14.51	12.85	12.05	11.54	10.64	10.71	10.26
	$S^* 0.75$	14.51	12.85	13.05	12.07	10.88	10.74	10.59
0.4	MGN	21.40	26.32	29.87	33.32	35.59	37.44	38.64
	S^*	15.43	19.92	22.99	24.58	25.94	26.70	26.44
	$S^* 0.5$	14.42	12.77	12.43	11.25	10.63	10.41	10.23
	$S^* 0.75$	14.42	12.77	12.77	11.59	10.60	10.39	10.30
0.5	MGN	24.82	31.55	35.92	41.19	43.85	46.87	49.58
	S^*	17.04	22.22	26.15	28.65	29.84	30.78	30.95
	$S^* 0.5$	14.12	12.79	12.72	10.92	10.99	10.47	9.97
	$S^* 0.75$	14.12	12.79	12.16	11.10	10.49	10.07	9.79

Note:- ' $S^* 0.5$ ' and ' $S^* 0.75$ ' denote the S^* test with additional lags chosen by the selection criteria $[0.5n^{1/3}]$ and $[0.75n^{1/3}]$ respectively.

where $\bar{d} = n^{-1} \sum d_t$

$$d_t = e_{1t}^2 - e_{2t}^2$$

$$\hat{V}_L(\bar{d}) = [n+1-2(q+1)+n^{-1}q(q+1)]^{-1}(\hat{\gamma}_0 + 2\sum_{k=1}^q \hat{\gamma}_k)$$

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$$

$$q = m + h - 1$$

$$m = [0.5n^{1/3}]$$

compare with t_{n-1} critical values

Table 4.7b shows the comparison between *MGN* and the new LS^* (S^* with $[0.5n^{1/3}]$) test under different degrees of ARCH, and the earlier tables 4.3 and 4.4 also contain results for LS^* compared to all the tests under different ARCH specifications. The value of LS^* is clearly illustrated, with dramatic reductions in oversizing achieved by the new test whenever ARCH is present. LS^* remains slightly oversized in the smaller samples, but is generally robust, and provides a workable alternative to all other tests for equal forecast accuracy. Finally, table 4.8 shows results for the 2-steps-ahead case, with LS^* again performing well, being robust to forecast error autocorrelation and, although not as well sized as for 1-step-ahead errors, still forming a useful multi-step-ahead test.

Altogether, this new test LS^* maintains all the advantages of S^* over its competitors, such as a general loss function specification, robustness to contemporaneous correlation and autocorrelation in the forecast errors, is readily

Table 4.8

Empirical sizes for the LS^* test at the nominal 10% level
 ($h = 2$, independent ARCH(1) errors)

α_1	θ	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
0.0	0.0	17.94	17.84	15.12	13.61	11.47	11.16	10.66
	0.5	18.92	17.23	13.76	13.16	11.74	11.14	10.73
	0.9	19.18	17.06	13.32	13.07	11.69	10.81	10.58
0.1	0.0	17.98	17.36	14.62	13.14	11.39	11.28	10.46
	0.5	18.37	16.71	13.45	13.12	11.73	11.19	10.49
	0.9	18.78	16.25	12.97	12.94	11.79	11.17	10.51
0.2	0.0	17.84	16.66	13.73	12.59	11.19	11.11	10.26
	0.5	18.14	16.02	13.00	12.82	11.47	11.05	10.41
	0.9	18.47	15.84	12.64	12.63	11.71	10.96	10.47
0.3	0.0	17.77	16.16	13.05	12.07	10.88	10.74	10.39
	0.5	17.72	15.07	12.88	12.34	11.13	10.81	10.42
	0.9	18.09	15.16	12.42	12.21	11.40	10.76	10.39
0.4	0.0	17.37	15.53	12.77	11.59	10.60	10.39	10.40
	0.5	17.60	14.48	12.35	11.75	11.07	10.52	10.19
	0.9	17.93	14.73	12.10	11.76	11.18	10.60	10.05
0.5	0.0	17.11	14.90	12.16	11.10	10.49	10.07	9.88
	0.5	17.72	14.02	12.00	11.35	10.64	10.18	9.64
	0.9	17.73	14.21	11.72	11.28	10.89	10.32	9.83

Note:- For these 2-steps-ahead errors, autocorrelation is incorporated by generating ARCH(1) errors u_{it} , $i = 1, 2$ and then transforming them to follow an MA(1) process with parameter θ , i.e. $e_{it} = (u_{it} + \theta u_{i,t-1}) / (1 + \theta^2)^{1/2}$, $i = 1, 2$.

applicable to multi-step-ahead forecasting, and most importantly has the added advantage of providing a much needed test which exhibits robustness to ARCH forecast errors. The example case considered in the analysis is that of ARCH(1) errors, but the root problems (leptokurtosis in the errors and autocorrelation in d_t) and the associated solution (including additional covariance lags in the estimation of $V(d_t)$), are common to general specifications of ARCH forecast errors.

4.4 ARCH Error Specification and Properties for Encompassing Tests

In the same way that section 4.2 outlined the specification of ARCH forecast errors which had equal forecast accuracy in preparation for the section 4.3 analysis of the tests, so it is now necessary to determine the specification for ARCH forecast errors where one forecast encompasses the other, in order to enable the next section to study tests for forecast encompassing. Again, ARCH(1) processes are assumed for the sake of clarity.

Given two forecasts f_{1t}, f_{2t} with corresponding errors e_{1t}, e_{2t} it can be said that f_{1t} encompasses f_{2t} if the error variance of some combined forecast f_{ct} does not significantly improve upon the error variance of f_{1t} alone (see chapter 3). This can be expressed as $\lambda = 0$ in the following regression:

$$e_{1t} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t \quad (4.64)$$

and gives rise to two conditions which characterise forecast encompassing:

$$V(e_{2t}) > V(e_{1t}) \quad (4.65)$$

$$V(e_{1t}) = C(e_{1t}, e_{2t}) \quad (4.66)$$

It is clear, then, that a model of contemporaneously correlated ARCH forecast errors must be employed so as to impose the forecast encompassing conditions on the error specification. The Engle-Kroner approach allows this to be done easily; the Bollerslev model, however, has complications since the unconditional correlation, ρ , has no analytic expression. In a Bollerslev world, the forecast encompassing errors

cannot be specified analytically, and so, given that this analysis is simply illustrating the effects of ARCH on forecast evaluation tests, attention is restricted to the Engle-Kroner error model.

The Engle-Kroner bivariate ARCH(1) specification is given in (4.10), and simplified further in (4.11). In section 4.2, the next assumption was to let $c_{11} = c_{22} = \alpha_0$ which gave errors with equal variance. In this case it is now necessary to let c_{11} and c_{22} differ, i.e. $c_{11} = \alpha_{01}$, $c_{22} = \alpha_{02}$; if it is further assumed that $\alpha_{02} > \alpha_{01}$, the forecast encompassing condition (4.65) is imposed upon the errors. The second condition, (4.66), is added by letting $c_{12} = c_{21} = \alpha_{01}$. Putting all this together gives the following error specification:

$$E(e_t e'_t | e_{t-1} e'_{t-1}, e_{t-2} e'_{t-2}, \dots) = H_t \quad (4.67)$$

$$\text{where } H_t = \begin{bmatrix} \alpha_{01} + \alpha_1 e_{1,t-1}^2 & \alpha_{01} + \alpha_1 e_{1,t-1} e_{2,t-1} \\ \alpha_{01} + \alpha_1 e_{1,t-1} e_{2,t-1} & \alpha_{02} + \alpha_1 e_{2,t-1}^2 \end{bmatrix}$$

$$\begin{aligned} E(e_t e'_t) &= \begin{bmatrix} E(e_{1t}^2) & E(e_{1t} e_{2t}) \\ E(e_{1t} e_{2t}) & E(e_{2t}^2) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\alpha_{01}}{1-\alpha_1} & \frac{\alpha_{01}}{1-\alpha_1} \\ \frac{\alpha_{01}}{1-\alpha_1} & \frac{\alpha_{02}}{1-\alpha_1} \end{bmatrix}; \alpha_{02} > \alpha_{01} \end{aligned} \quad (4.68)$$

It is also useful to normalise on $V(e_{1t})$, i.e. $V(e_{1t}) = C(e_{1t}, e_{2t}) = 1$, which amounts to setting $\alpha_{01} = 1 - \alpha_1$, and can be done without loss of generality:

$$H_t = \begin{bmatrix} 1 - \alpha_1 + \alpha_1 e_{1,t-1}^2 & 1 - \alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1} \\ 1 - \alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1} & \alpha_{02} + \alpha_1 e_{2,t-1}^2 \end{bmatrix} \quad (4.69)$$

$$E(e_t e_t') = \begin{bmatrix} 1 & 1 \\ 1 & \frac{\alpha_{02}}{1-\alpha_1} \end{bmatrix}; \quad \alpha_{02} > 1-\alpha_1 \quad (4.70)$$

Having established the basic error specification, as in previous sections a representation can be derived yielding errors as a function of past errors and normal disturbances, $v_{it} \sim \text{IN}(0,1)$, $i = 1,2$:

$$e_t = T_{t-1} v_t, \quad H_t = T_{t-1} T_{t-1}'$$

$$T_{t-1} = \begin{bmatrix} t_{11,t-1} & 0 \\ t_{21,t-1} & t_{22,t-1} \end{bmatrix}$$

$$\text{where } t_{11,t-1} = \sqrt{\alpha_0 + \alpha_1 e_{1,t-1}^2}$$

$$t_{21,t-1} = \frac{1-\alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1}}{\sqrt{1-\alpha_1 + \alpha_1 e_{1,t-1}^2}}$$

$$t_{22,t-1} = \sqrt{\alpha_{02} + \alpha_1 e_{2,t-1}^2 - \frac{(1-\alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1})^2}{1-\alpha_1 + \alpha_1 e_{1,t-1}^2}}$$

yielding the following representation for e_{1t}, e_{2t} :

$$e_{1t} = v_{1t} \sqrt{1-\alpha_1 + \alpha_1 e_{1,t-1}^2} \quad (4.71)$$

$$\begin{aligned} e_{2t} = v_{1t} & \left(\frac{1-\alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1}}{\sqrt{1-\alpha_1 + \alpha_1 e_{1,t-1}^2}} \right) \\ & + v_{2t} \sqrt{\alpha_{02} + \alpha_1 e_{2,t-1}^2 - \frac{(1-\alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1})^2}{1-\alpha_1 + \alpha_1 e_{1,t-1}^2}} \end{aligned} \quad (4.72)$$

This representation given in (4.71) and (4.72) can be used to determine a number of properties which will be needed in the section 4.5 analysis of tests for forecast encompassing:

$$1. E(e_{1t}^2) = 1 \quad \text{from (4.70)} \quad (4.73)$$

$$2. E(e_{2t}^2) = \frac{\alpha_{02}}{1-\alpha_1} \quad \text{from (4.70)} \quad (4.74)$$

$$3. E(e_{1t}e_{2t}) = 1 \quad \text{from (4.70)} \quad (4.75)$$

$$\begin{aligned} 4. E(e_{1t}^4) &= E(v_{1t}^4 h_{11t}^2) && \text{using (4.71)} \\ &= 3E(h_{11t}^2) && \text{as } E(v_{1t}^4) = 3 \\ &&& v_{1t}, h_{ijt} \text{ independent} \\ &= 3E[(1-\alpha_1)^2 + \alpha_1^2 e_{1,t-1}^4 + 2(1-\alpha_1)\alpha_1 e_{1,t-1}^2] \\ &&& \text{using (4.69)} \\ &= 3\alpha_1^2 E(e_{1t}^4) + 3(1-\alpha_1)(1+\alpha_1) \\ &= \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} \end{aligned} \quad (4.76)$$

$$\begin{aligned} 5. E(e_{1t}^2 e_{2t}^2) &= E\{v_{1t}^2 h_{11t} [v_{1t}^2 h_{11t}^{-1} h_{12t}^2 + v_{2t}^2 (h_{22t} - h_{11t}^{-1} h_{12t}^2) \\ &\quad + 2v_{1t} v_{2t} h_{11t}^{-1/2} h_{12t} (h_{22t} - h_{11t}^{-1} h_{12t}^2)^{1/2}] \} \\ &= 2E(h_{12t}^2) + E(h_{11t} h_{22t}) \\ &= 2E[(1-\alpha_1)^2 + \alpha_1^2 e_{1,t-1}^2 e_{2,t-1}^2 + 2(1-\alpha_1)\alpha_1 e_{1,t-1} e_{2,t-1}] \\ &\quad + E[(1-\alpha_1)\alpha_{02} + \alpha_1^2 e_{1,t-1}^2 e_{2,t-1}^2 + (1-\alpha_1)\alpha_1 e_{2,t-1}^2 \\ &\quad + \alpha_{02}\alpha_1 e_{1,t-1}^2] \\ &= 3\alpha_1^2 E(e_{1t}^2 e_{2t}^2) + 4(1-\alpha_1)\alpha_1 + 2(1-\alpha_1)^2 \\ &\quad + (1-\alpha_1)\alpha_{02} + 2\alpha_1\alpha_{02} \end{aligned}$$

$$= \frac{\alpha_{02}(1+\alpha_1)+2(1-\alpha_1^2)}{1-3\alpha_1^2} \quad (4.77)$$

$$\begin{aligned}
6. \quad E(e_{1t}^3 e_{2t}) &= E\{v_{1t}^3 h_{11t}^{3/2} [v_{1t} h_{11t}^{-1/2} h_{12t} + v_{2t} (h_{22t} - h_{11t}^{-1} h_{12t}^2)^{1/2}]\} \\
&= 3E(h_{11t} h_{12t}) \\
&= 3E[(1-\alpha_1)^2 + \alpha_1^2 e_{1,t-1}^3 e_{2,t-1} + (1-\alpha_1)\alpha_1 e_{1,t-1}^2 \\
&\quad + (1-\alpha_1)\alpha_1 e_{1,t-1} e_{2,t-1}] \\
&= 3\alpha_1^2 E(e_{1t}^3 e_{2t}) + 3(1-\alpha_1)^2 + 6(1-\alpha_1)\alpha_1 \\
&= \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} \quad (4.78)
\end{aligned}$$

4.5 Tests for Forecast Encompassing

Turning now to the examination of tests for forecast encompassing, the section 4.4 specification of the forecast errors can be used to evaluate the different tests contained in chapter 3 when ARCH is a feature of the errors. The analysis mirrors that performed in section 4.3 for tests for equal forecast accuracy, with both theoretical and simulation considerations given for each test. The tests studied are firstly the regression test, R , followed by the non-normality robust tests R_{m1}, R_{m2}, r_s and DM_m . Finally, a modified test equivalent to the LS^* test proposed in section 4.3 is presented.

4.5.1 Regression Test

4.5.1a Theory

The regression test for forecast encompassing has the specification given in (3.7), and takes the same form as MGN , the difference being the specification of y_t . The distribution analysis of MGN contained in section 4.3 holds for R as well, and it is possible to write an equivalent result to (4.29) for the regression test under the null hypothesis:

$$R \xrightarrow{d} N(0, \text{plim}[\hat{Q}^{-1}Q]) \quad (4.79)$$

$$\text{where } Q = V(n^{-1/2} \sum x_t \varepsilon_t)$$

$$\hat{Q} = \hat{\sigma}^2 n^{-1} \sum x_t^2$$

$$x_t = e_{1t} - e_{2t}, \varepsilon_t = y_t = e_{1t}$$

As with MGN , the analysis now centres around the estimate \hat{Q} implicit in the

regression test under ARCH(1) forecast errors as specified in section 4.4, i.e. Engle-Kroner correlated errors where one forecast encompasses the other.

Firstly, the true element Q can be found as follows, beginning with a version of (4.30):

$$Q = \gamma_0 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \gamma_k \quad (4.80)$$

$$\text{where } \gamma_k = E(d_t d_{t+k})$$

$$d_t = e_{1t}^2 - e_{1t} e_{2t}$$

The distinction between (4.80) and (4.30) is the definition of d_t which distinguishes the encompassing case. Now the γ_k term can be expressed more explicitly by the following, beginning with two results which follow from the conditional expectation specification given in (4.69):

$$e_{1t}^2 = 1 - \alpha_1 + \alpha_1 e_{1,t-1}^2 + \eta_{1t}$$

$$e_{1t} e_{2t} = 1 - \alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1} + \eta_{2t}$$

$$d_t = \alpha_1 d_{t-1} + \eta_{3t} \quad (4.81)$$

where $\eta_{3t} \sim \text{white noise}$

$$E(\eta_{3t}) = 0$$

$$E(\eta_{3t}^2) = \delta_2^2 \quad (\delta_2^2 = \text{constant})$$

$$E(\eta_{3t} \eta_{3s}) = 0 \quad s \neq t$$

$$\gamma_k = E(d_t d_{t+k})$$

$$= \frac{\delta_2^2 \alpha_1^k}{1 - \alpha_1^2}$$

$$= \gamma_0 \alpha_1^k \quad (4.82)$$

Substituting into (4.80) then gives the following for Q :

$$Q = \gamma_0 \left[1 + 2n^{-1} \sum_{k=1}^{n-1} (n-k) \alpha_1^k \right] \quad (4.83)$$

Once again γ_0 must be found:

$$\begin{aligned} \gamma_0 &= E(d_t^2) \\ &= E[(e_{1t}^2 - e_{1t}e_{2t})^2] \\ &= E(e_{1t}^4) + E(e_{1t}^2 e_{2t}^2) - 2E(e_{1t}^3 e_{2t}) \\ &= \frac{[\alpha_{02} - (1 - \alpha_1)](1 + \alpha_1)}{1 - 3\alpha_1^2} \quad \text{using (4.76)-(4.78)} \end{aligned} \quad (4.84)$$

The bracketed term in (4.83) has already been found in the limit, as in (4.35), and substitution of both of these results into (4.83) gives the limiting expression for Q :

$$Q \longrightarrow \frac{[\alpha_{02} - (1 - \alpha_1)](1 + \alpha_1)^2}{(1 - 3\alpha_1^2)(1 - \alpha_1)} \quad (4.85)$$

Turning to the estimate of Q , the probability limit of \hat{Q} must again be derived under H_0 :

$$\begin{aligned} \hat{Q} &\xrightarrow{p} V(\varepsilon_t) E(x_t^2) \\ &= E(e_{1t}^2) [E(e_{1t}^2) + E(e_{2t}^2) - 2E(e_{1t}e_{2t})] \\ &= \frac{\alpha_{02} - (1 - \alpha_1)}{1 - \alpha_1} \quad \text{using (4.73)-(4.75)} \end{aligned} \quad (4.86)$$

The desired asymptotic distribution of the regression test when the errors are ARCH(1) can now be found using (4.79), (4.85) and (4.86):

$$R \xrightarrow{d} N\left(0, \frac{(1 + \alpha_1)^2}{1 - 3\alpha_1^2}\right) \quad (4.87)$$

This result shows that the regression test for forecast encompassing does not have an asymptotic standard normal distribution, but has a limiting variance which exceeds one and depends on the degree of ARCH in the errors. It is also noteworthy that the asymptotic variance for R is the same as that for MGN under the Engle-Kroner correlated error case given in (4.42). This follows from the autocorrelation structure of d_t being the same in both cases plus the chapter 3 result of regression error conditional heteroscedasticity impacting both tests in the same way. Table 4.9 (along with entries concerning tests analysed below) shows the extent of the lack of robustness, and, as with the MGN analysis, motivates the need for a new correctly sized test. The theoretical sizes given in table 4.9 differ from those for MGN given in table 4.1a purely because the encompassing alternative is 1-sided as opposed to the equal accuracy 2-sided alternative hypothesis.

4.5.1b Simulation

Monte Carlo simulation can now be performed to check the theoretical sizes of R given in table 4.9, and to examine the test's finite sample properties. Paralleling the experiments for MGN , sample sizes between $n = 8$ and $n = 512$ are considered, plus

Table 4.9

Theoretical asymptotic sizes for the R, R_{m1}, R_{m2} and DM_m tests
at the nominal 10% level with Engle-Kroner ARCH(1) errors

α_1	Test	Distribution variance	Asymptotic test size
0.1	R	1.2474	12.56
	R_{m1}, R_{m2}, DM_m	1.2222	12.32
0.2	R	1.6364	15.82
	R_{m1}, R_{m2}, DM_m	1.5	14.77
0.3	R	2.3151	19.98
	R_{m1}, R_{m2}, DM_m	1.8571	17.35
0.4	R	3.7692	25.46
	R_{m1}, R_{m2}, DM_m	2.3333	20.07
0.5	R	9	33.46
	R_{m1}, R_{m2}, DM_m	3	22.97

in one case $n = 10,000$, with 1-step-ahead forecasts evaluated at the nominal 10% level. The null is one of encompassing against a 1-sided alternative; 10,000 replications are performed for each simulation. Equations (4.71) and (4.72) are used to generate the errors, and the starting value has its impact minimised by simulating $n + 100$ observations each time and discarding the first 100 observations.

One further point to note is that the test statistic R is invariant to the choice of α_{02} , provided it lies within the specified range $\alpha_{02} > 1 - \alpha_1$. This can be shown as follows, beginning with the following version of the test statistic:

$$R = \frac{(n-1)^{1/2} \sum x_t y_t}{\sqrt{\sum x_t^2 \sum y_t^2 - (\sum x_t y_t)^2}}$$

Now the error representations given in (4.71) and (4.72) can be employed in this proof, plus the fact that in the simulations, starting values for e_{1t}, e_{2t} are arbitrarily assumed to be zero (the effect of this becomes negligible since the first 100 observations are dropped in each experiment). Collating this information gives:

$$y_t = v_{1t} \sqrt{1 - \alpha_1 + \alpha_1 e_{1,t-1}^2} \quad \text{from (4.71)}$$

$$\text{i.e. } y_1 = v_{11} \sqrt{1 - \alpha_1} \quad \text{since } e_{1,0} = 0$$

$$y_2 = v_{12} \sqrt{1 - \alpha_1 + \alpha_1 y_1^2}$$

$$y_3 = v_{13} \sqrt{1 - \alpha_1 + \alpha_1 y_2^2}$$

etc.

$$\text{so } y_t = f_t(\alpha_1; v_{11}, v_{12}, \dots, v_{1t}) \quad (4.88)$$

$$\begin{aligned}
x_t = & v_{1t} \sqrt{1 - \alpha_1 + \alpha_1 e_{1,t-1}^2} - v_{1t} \left(\frac{1 - \alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1}}{\sqrt{1 - \alpha_1 + \alpha_1 e_{1,t-1}^2}} \right) \\
& - v_{2t} \sqrt{\alpha_{02} + \alpha_1 e_{2,t-1}^2 - \frac{(1 - \alpha_1 + \alpha_1 e_{1,t-1} e_{2,t-1})^2}{1 - \alpha_1 + \alpha_1 e_{1,t-1}^2}} \quad \text{from (4.72)}
\end{aligned}$$

now $\alpha_{02} > 1 - \alpha_1$, or $\alpha_{02} = 1 - \alpha_1 + c$, $c > 0$

$$\begin{aligned}
\text{i.e. } x_1 = & \frac{-v_{21} \sqrt{c(1 - \alpha_1)}}{\sqrt{1 - \alpha_1}} \quad \text{since } e_{1,0}, e_{2,0} = 0 \\
= & \sqrt{c}(-v_{21})
\end{aligned}$$

$$x_2 = \sqrt{c} \left(\frac{-v_{11} v_{12} v_{21} \alpha_1 \sqrt{1 - \alpha_1} - v_{22} \sqrt{(1 - \alpha_1)[1 + \alpha_1(v_{11}^2 + v_{21}^2)]}}{\sqrt{(1 - \alpha_1)(1 + \alpha_1 v_{11}^2)}} \right)$$

etc.

$$\text{so } x_t = \sqrt{c} \ g_t(\alpha_1; v_{11}, v_{12}, \dots, v_{1t}; v_{21}, v_{22}, \dots, v_{2t}) \quad (4.89)$$

Thus:

$$\begin{aligned}
R = & \frac{(n-1)^{1/2} c^{1/2} \Sigma g_t(.) f_t(.)}{\sqrt{c \Sigma g_t(.)^2 \Sigma f_t(.)^2 - [c^{1/2} \Sigma g_t(.) f_t(.)]^2}} \\
= & \frac{(n-1)^{1/2} \Sigma f_t(.) g_t(.)}{\sqrt{\Sigma f_t(.)^2 \Sigma g_t(.)^2 - [\Sigma f_t(.) g_t(.)]^2}} \quad (4.90)
\end{aligned}$$

Now (4.90) shows that R is invariant to c , and thus provided $\alpha_{02} > 1 - \alpha_1$, this proves that R is not affected by the parameter α_{02} , given that forecast encompassing holds.

Table 4.10 provides the results of the simulation experiments for a number of ARCH

parameters α_1 , and includes results for another test, LDM_m , to enable comparison later in sub-section 4.5.3. As would be expected, the picture is one of serious oversizing. A distinct lack of robustness to ARCH in the forecast errors is revealed at all sample sizes, which worsens as the degree of ARCH rises. Tables 4.11 and 4.12 contain comparisons with other tests which shall be discussed shortly, and results in table 4.11 provide confirmation of the theoretical limiting size for $\alpha_1 = 0.2$.

The general inference is that, like MGN , the regression test for forecast encompassing is not statistically valid when ARCH is present in the forecast errors, and once again the need for correction is highlighted.

4.5.2 Non-Normality Robust Tests

4.5.2a Theory

Using the same analysis as in the previous sub-section, the chapter 3 non-normality correction tests can now be examined. The two parametric corrections R_{m1}, R_{m2} have the same asymptotic distribution as MGN_1^*, MGN_2^* respectively, as given in (4.47) and (4.48), but with the difference that y_t is now defined as $y_t = e_{1t}$. Repeating the sub-section 4.3.2a theory for the forecast encompassing case yields:

$$\begin{aligned}\hat{Q}_{m1}, \hat{Q}_{m2} &\xrightarrow{P} E(x_t^2 \varepsilon_t^2) \\ &= E[(e_{1t} - e_{2t})^2 e_{1t}^2] \\ &= E(e_{1t}^4) + E(e_{1t}^2 e_{2t}^2) - 2E(e_{1t}^3 e_{2t})\end{aligned}$$

Table 4.10

Empirical sizes for the R and LDM_m tests
at the nominal 10% level ($h = 1$, Engle-Kroner ARCH(1) errors)

α_1	Test	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
0.1	R	11.51	11.65	11.65	12.66	12.33	12.70	12.76
	LDM_m	11.30	11.63	11.01	11.55	10.90	10.51	10.41
0.2	R	13.53	14.01	14.33	15.67	15.61	15.77	15.71
	LDM_m	11.65	11.78	11.20	11.50	11.12	10.56	10.56
0.3	R	15.37	16.43	17.05	18.66	18.94	19.45	19.15
	LDM_m	12.09	12.19	11.84	11.58	11.38	10.35	10.90
0.4	R	16.96	18.70	19.82	21.69	22.03	23.27	23.45
	LDM_m	12.50	12.71	12.60	11.86	11.87	10.62	10.85
0.5	R	18.52	20.57	22.41	24.35	25.16	27.11	27.70
	LDM_m	12.94	13.32	13.41	12.20	12.15	11.03	10.80

Table 4.11

Empirical sizes for the $R, R_{m1}, R_{m2}, r_s, DM_m$ and LDM_m tests
at the nominal 10% level ($h = 1$, Engle-Kroner ARCH(1) errors, $\alpha_1 = 0.2$)

	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$	$n=10000$
R	13.53	14.01	14.33	15.67	15.61	15.77	15.71	15.83
R_{m1}	19.32	17.33	15.93	15.76	15.46	15.39	14.94	14.79
R_{m2}	11.37	13.41	13.77	14.75	14.85	15.01	14.75	14.79
r_s	11.06	12.94	13.83	14.41	14.04	14.31	14.19	14.80
DM_m	13.37	14.14	14.19	14.85	14.91	15.06	14.77	14.79
LDM_m	11.65	11.78	11.20	11.50	11.12	10.56	10.56	10.03

Table 4.12

Empirical sizes for the $R, R_{m1}, R_{m2}, r_s, DM_m$ and LDM_m tests
 at the nominal 10% level ($h = 1$, Engle-Kroner ARCH(1) errors, $\alpha_1 = 0.4$)

	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
R	16.96	18.70	19.82	21.69	22.03	23.27	23.45
R_{m1}	22.12	21.31	20.83	21.07	20.56	20.36	20.15
R_{m2}	13.86	16.92	18.33	19.49	19.48	19.80	19.70
r_s	13.58	16.47	17.76	18.20	18.36	18.80	18.65
DM_m	16.09	17.74	18.66	19.65	19.53	19.81	19.73
LDM_m	12.50	12.71	12.60	11.86	11.87	10.62	10.85

$$\begin{aligned}
&= E(d_t^2) \quad \text{where } d_t = e_{1t}^2 - e_{1t}e_{2t} \\
&= \gamma_0
\end{aligned} \tag{4.91}$$

The rest of the theory in sub-section 4.3.2a then follows through yielding an asymptotic distribution for the parametric corrections which is the same as that for MGN_1^* and MGN_2^* :

$$R_{m1}, R_{m2} \xrightarrow{d} N\left(0, \frac{1+\alpha_1}{1-\alpha_1}\right) \tag{4.92}$$

Investigation into the behaviour of the rank correlation variant of R , as given in (3.36) is studied later through simulation.

The modified Diebold-Mariano approach test, DM_m , mirrors the earlier S^* test, the distinction being the definition of d_t . Specifying $d_t = e_{1t}^2 - e_{1t}e_{2t}$ for the encompassing hypothesis, the result of (4.81) again holds, and this allows the deduction that the chapter 3 preferred test for forecast encompassing has the following limiting distribution:

$$DM_m \xrightarrow{d} N\left(0, \frac{1+\alpha_1}{1-\alpha_1}\right) \tag{4.93}$$

As with tests for equal forecast accuracy, the non-normality robust tests for forecast encompassing are affected by ARCH in the forecast errors, causing oversizing as a result of the autocorrelation now present in d_t . Table 4.9 gives the limiting sizes for these ‘corrected’ tests for different ARCH parameters, and the degree of the problem

is exposed. A robust test for forecast encompassing is clearly needed.

4.5.2b Simulation

Simulation of these tests provides confirmation of the above theoretical results (see tables 4.11 and 4.12), showing that all the non-normality robust tests R_{m1}, R_{m2}, r_s and DM_m are significantly oversized in finite samples and in the limit. The tests are all invariant to α_{02} given that $\alpha_{02} > 1 - \alpha_1$ is satisfied, and proofs equivalent to that performed for R exist to demonstrate this property. All four tests tend to the same limiting, and incorrect, size. Once again, a lack of robustness is displayed in these tests when the forecast errors follow ARCH processes.

4.5.3 A Modified Test

The above analysis and simulation show that a test for forecast encompassing which exhibits robustness to ARCH forecast errors is required. The problem in the tests considered in the previous sub-section is that of unaccounted-for autocorrelation in d_t . The tests R_{m1}, R_{m2} and DM_m can be modified to overcome this problem by adding additional autocovariance lags in the variance estimation contained in the test statistics. The theory is exactly the same as for the modified tests for equal forecast accuracy in section 4.3.

Focusing on the preferred test, DM_m , the modified test becomes:

$$LDM_m = \frac{\bar{d}}{\sqrt{\hat{V}_L(\bar{d})}} \quad (4.94)$$

where $\bar{d} = n^{-1} \sum d_t$

$$d_t = e_{1t}^2 - e_{1t}e_{2t}$$

$$\hat{V}_L(\bar{d}) = [n+1-2(q+1)+n^{-1}q(q+1)]^{-1}(\hat{\gamma}_0 + 2\sum_{k=1}^q \hat{\gamma}_k)$$

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^n (d_t - \bar{d})(d_{t-k} - \bar{d})$$

$$q = m + h - 1$$

$$m = [0.5n^{1/3}]$$

compare with t_{n-1} critical values

The choice of bandwidth, m , can again be determined by a lag selection criterion as given in (4.62); table 4.13 gives empirical sizes for DM_m with additional lags running from zero to six, and inspection confirms that the best choice for the parameter g in (4.62) is again 0.5, as noted in (4.94) above. The result is a test for forecast encompassing, LDM_m , which is robust and has good size properties, as given in tables 4.10-4.12. Table 4.14 provides results for 2-steps-ahead prediction, with LDM_m remaining robust for this longer horizon.

The new test, LDM_m , therefore provides an alternative to DM_m which maintains all the DM_m advantages, whilst having the highly desirable additional property of being robust to ARCH forecast errors. This result applies broadly beyond the case of bivariate ARCH(1) errors, and demonstrates how a test can be made robust to ARCH errors in general.

Table 4.13

Empirical sizes for the DM_m test with additional lags
at the nominal 10% level ($h = 1$, Engle-Kroner ARCH(1) errors, $\alpha_1 = 0.2$)

	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
0 lags	13.37	14.14	14.19	14.85	14.91	15.06	14.77
1 lags	11.65	11.78	11.20	11.56	11.50	11.00	11.22
2 lags	10.13	12.62	11.42	11.50	11.12	10.59	10.74
3 lags	8.52	12.29	12.16	12.01	11.32	10.56	10.55
4 lags	6.37	11.15	12.03	12.48	11.48	10.60	10.56
5 lags	4.49	10.48	11.87	12.75	11.84	10.87	10.68
6 lags	3.89	9.50	11.71	13.01	11.83	10.95	10.54

Table 4.14

Empirical sizes for the LDM_m test at the nominal 10% level
 ($h = 2$, Engle-Kroner ARCH(1) errors)

α_1	θ	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
0.1	0.0	9.65	12.44	11.46	12.22	11.37	10.65	10.34
	0.5	10.32	12.28	11.40	12.23	11.06	10.95	10.19
	0.9	10.75	12.77	11.42	12.03	10.88	11.02	10.30
0.2	0.0	10.13	12.62	11.42	12.01	11.32	10.60	10.68
	0.5	10.51	12.47	11.50	11.88	11.28	10.61	10.20
	0.9	10.61	12.81	11.42	11.96	11.12	10.88	10.25
0.3	0.0	10.52	12.65	11.54	12.05	11.34	10.56	10.87
	0.5	10.77	12.59	11.58	12.09	11.40	10.72	10.35
	0.9	10.92	12.77	11.64	11.92	11.33	10.67	10.34
0.4	0.0	10.97	12.55	11.70	12.03	11.57	10.48	10.76
	0.5	11.06	12.59	11.84	12.07	11.31	10.89	10.47
	0.9	10.96	13.03	11.93	12.04	11.37	11.08	10.51
0.5	0.0	11.14	12.68	11.91	12.01	11.56	10.70	10.74
	0.5	11.01	12.92	12.08	12.26	11.30	10.97	10.69
	0.9	11.10	12.11	12.11	12.23	11.48	10.93	10.60

Note:- For these 2-steps-ahead errors, autocorrelation is incorporated by generating ARCH(1) errors u_{it} , $i = 1, 2$ and then transforming them to follow an MA(1) process with parameter θ , i.e. $e_{it} = (u_{it} + \theta u_{i,t-1}) / (1 + \theta^2)^{1/2}$, $i = 1, 2$.

4.6 Conclusion

In conclusion, the behaviour of tests for equal forecast accuracy and tests for forecast encompassing have been analysed in the important case of ARCH forecast errors. The established tests (MGN and R respectively) are found to be seriously oversized in finite samples and in the limit. The alternative tests explored in chapter 2 (MGN_1^* , MGN_2^* , r_s and S^*) and in chapter 3 (R_{m1} , R_{m2} , r_s and DM_m) correct for the leptokurtosis element of the ARCH problem, but the problem of autocorrelation in the respective loss differentials d_t remains, with these ‘corrected’ tests also having an incorrect asymptotic test size.

Two new modified tests are proposed: LS^* for testing for equal forecast accuracy, and LDM_m for testing for forecast encompassing. These tests take account of the effects of ARCH in the errors by adding additional covariance lags to the variance estimates in S^* and DM_m respectively according to a given rule for lag selection, and using an unweighted truncated kernel. The result is then two tests which exhibit robustness to ARCH errors, as well as to non-normal, contemporaneously correlated and autocorrelated errors. They are applicable to multi-step-ahead forecast evaluation (albeit with some moderate size distortions for LS^* in the smallest samples) and also take very general loss function specifications. These newly proposed tests consequently make a valuable contribution to the practice of testing for equal forecast accuracy and forecast encompassing.

Chapter 5

Ranking Competing Forecasts

5.1 Introduction

In the evaluation of competing forecasts, it is desirable to be able to rank rival forecasts in order of their predictive ability. The forecasts of the variable of interest can be used in this way to examine the past performance of the forecast-generating models or methods, and thereby give an indication of the likely value of each in future application.

A number of criteria for ranking competing forecasts exist. Some of the most common criteria are mean squared forecast error (MSFE) based measures which evaluate a simple form of economic loss to the user. Clements & Hendry (1993) criticise these measures due to their lack of invariance across alternative isomorphic representations of the system concerned, and instead propose a new invariant summary criterion based on the determinant of the generalised forecast error second moment matrix - denoted GFESM. This chapter analyses some aspects of the GFESM criterion and explores alternative measures and related issues.

The study is comprised of six sections. Section 5.2 reviews the literature on the GFESM criterion; section 5.3 then provides a critique of the use of this measure based on two premises - a questioning of the criterion's underlying justification, and analysis of its behaviour when comparing misspecified models. Section 5.4 investigates the possibility of another, more well-behaved, invariant criterion, and section 5.5 examines the related issue of testing for forecast encompassing with linear combinations of forecasts. The chapter is concluded in section 5.6.

5.2 Literature Review

In the practice of ranking competing forecasts, one of the most established and frequently used measures is the mean squared forecast error (MSFE):

$$\text{MSFE} = E[e_{T+h}^2] \quad (5.1)$$

where e = forecast error

h = forecast horizon

The forecast which has the minimum possible MSFE is known as the optimal forecast, and equates to the conditional expectation of the quantity to be predicted. Given a number of competing forecasts, then, rankings can be established by this mean squared error measure of economic loss - the lower the MSFE, the better the model or method.

In a multivariate situation, a number of MSFE-based criteria are used. The MSFE matrix becomes the variance-covariance matrix between the forecast errors of the variables concerned:

$$V_h = E[e_{T+h} e_{T+h}'] \quad (5.2)$$

where $e_{T+h} = (k \times 1)$ vector of h -steps-ahead forecast errors

and criteria are derived using this matrix. One approach is to use the trace MSFE (TMSFE):

$$\text{TMSFE} = \text{trace}(V_h) \quad (5.3)$$

Another is to use the MSFE determinant (DMSFE):

$$\text{DMSFE} = |V_h| \quad (5.4)$$

A third criterion proposed by Granger & Newbold (1986) chooses the forecast which has the lowest value of MSFEM (mean squared forecast error matrix):

$$\text{MSFEM} = d'V_h d \quad (5.5)$$

for every non-zero vector d . MSFEM optimality then implies optimality for all linear functions of the series. Such an optimum will not necessarily exist unless the conditional mean of the future given the past is included in the set of forecasts compared, i.e. the criterion may not provide a ranking of the rival forecasts.

MSFE-based measures of forecast accuracy therefore exist in univariate and multivariate settings, and are frequently applied. The criteria are simple, intuitive, correspond to the conditional expectation of the future quantity, and have a clear economic loss interpretation. Such is the basis for the widespread use of MSFE measures when making comparisons between contending forecasts.

In criticism of MSFE criteria, Clements & Hendry (1993) argue that these measures ‘constitute an inadequate and potentially misleading basis for model selection’. Their motivation for this statement stems from an analysis of how MSFE type criteria behave under transformations to the models concerned. Clements & Hendry consider a class of transformations which are linear, non-singular and scale-preserving. If a linear system is denoted by:

$$\Phi s_t = u_t \quad (5.6)$$

where $u_t \sim \text{ID}(0, \Omega)$

$$s'_t = [x'_t : z'_t]$$

$x_t = k$ variables in system to be forecast

$z_t = N$ predetermined variables

$$\Phi = [I : -B]$$

$\Omega, B = \text{model parameters}$

then the desired class of transformations can be written as:

$$M\Phi P^{-1}P s_t = M u_t \quad (5.7)$$

where M is $(k \times k)$, $|M| = 1$

P is $(k+N \times k+N)$ upper block triangular

The model specified in (5.6) is invariant to these transformations, thus (5.7) summarises the class of isomorphic representations of the system associated with the linear non-singular scale-preserving transformations. Clements & Hendry then examine the effects of applying these M and P transforms on the rankings suggested by the MSFE-based forecast comparison criteria (5.1), (5.3), (5.4) and (5.5).

Firstly, the simple MSFE used for univariate models is found to be invariant to isomorphic transformations of the type considered above when 1-step-ahead forecasts are employed. However, when longer forecast horizons are used, the MSFE is no longer invariant under M and P transformations, and so comparisons for the purpose of forecast evaluation between one representation of a variable and a

different representation of that same variable are not valid. Furthermore, when parameter uncertainty is included, the rankings of rival forecasts of a given variable are also not invariant to these transformations, with the rankings possibly switching when different representations of the system are examined.

Each of the other MSFE-based criteria exhibit similar problems. The multivariate measure TMSFE varies under both M and P transformations for all forecast horizons, with the solitary exception of invariance 1-step-ahead when a P transform is applied. DMSFE and MSFEM are slightly better, being invariant for all steps ahead forecasts for M transformations, plus 1-step-ahead for P transformations, but again both fall down under P transforms where multi-step-ahead forecasts are concerned.

The isomorphic transformations considered above are widely applied. Different representations of a system such as levels, differences and cointegrating combinations are all contained in the P class of transforms, and thus the Clements & Hendry findings have serious implications for the validity of forecast comparisons using MSFE-type evaluation criteria. Put simply, multi-step MSFE-based measures cannot in general be used to give valid comparisons between different isomorphic representations of the same model, or to provide rankings which will be consistent over alternative representations of the models concerned, even when a common basis for comparison is used across models. There are a few exceptions to this generalisation, e.g. DMSFE and MSFEM under M transformations, but the broad picture is of MSFE-based criteria unreliability when different model representations are considered. No MSFE-type measure is invariant to multi-step P transformations,

thus forecast performance comparisons between a model's level and its change are not valid for h -steps-ahead forecasts ($h > 1$). In addition to this, rankings between models using levels representations, achieved using multi-step MSFE-based criteria, may switch if the models are transformed to changes representations. This fundamental lack of invariance leads Clements & Hendry to their conclusion mentioned above, that MSFE-based criteria are inadequate and possibly misleading in the evaluation of competing forecasts.

Out of this background, Clements & Hendry motivate the desire for a measure which yields unique forecast accuracy rankings. Such a measure must clearly be invariant to transformations which generate different but isomorphic representations of a system.

In order to obtain an invariant criterion, Clements & Hendry take into account the covariance terms between the forecast errors from different horizons from the multi-step-ahead forecasts considered. Their proposed criterion is the determinant of the generalised forecast error second moment matrix, denoted GFESM:

$$\text{GFESM} = |\phi_h| = |E[E_v E_v']| \quad (5.8)$$

$$\text{where } E_v' = [e'_{T+1}, e'_{T+2}, \dots, e'_{T+h}]$$

The forecast errors from all horizons up to and including h are thus stacked in a vector E_v , then used to form the generalised forecast error second moment matrix $E[E_v E_v']$; the determinant of this matrix is the GFESM criterion which is invariant to both M and P type transformations, i.e. invariant to the desired class of linear

scale-preserving non-singular transformations to the system. GFESM consequently provides a unique ranking of competing models for a given maximum forecast horizon, consistent over all isomorphic representations of the models of interest.

Clements & Hendry show that when the true model is known (i.e. errors are white noise), and no parameter uncertainty exists, the GFESM h -steps-ahead criterion is $|\phi_1|^h$, thus a recursion exists with the complete ranking determined by the 1-step-ahead forecast errors. This result is used to argue that ‘model evaluation should focus on one-step performance’. However, when deviation away from the particular conditions necessary for this result occurs, the recursion breaks down and analysis of the 1-step-ahead GFESM is insufficient to determine multi-step-ahead forecast rankings.

The GFESM criterion is not the only measure of forecast evaluation which has the property of invariance to linear scale-preserving non-singular transformations. Clements & Hendry cite two other criteria - the log-likelihood, and forecast encompassing. The log-likelihood criterion is derived from a predictive likelihood approach to the evaluation of competing forecasts. The log-likelihood of the variables to be predicted by a model reduces to a concentrated log-likelihood function which varies with $|\hat{\phi}_h|$ alone, where $\hat{\phi}_h$ is the maximum likelihood estimate of ϕ_h . This result forms the justification for the Clements & Hendry use of the GFESM criterion, and is analysed more fully in section 3. Forecast encompassing is concerned with whether one forecast-generating mechanism is conditionally efficient with respect to a rival generating mechanism (see Nelson, 1972, Granger &

Newbold, 1973, and Chong & Hendry, 1986), i.e. one forecast encompasses the other if it can explain the errors made by the inferior forecast; the inferior forecast then has no information to contribute to the encompassing forecast with regard to prediction (see chapter 3).

A number of discussants commented on the Clements & Hendry paper, and several aspects relating to the GFESM measure are examined. Diebold argues that the lack of invariance exhibited by MSFE rankings across alternative model representations is actually desirable. Each user of forecasts has individual preferences, and these preferences should dictate their choice of measure and model representation. McNees adds to this, pointing out that all applied forecasting operates in the realm of misspecified models, and it is far from intolerable to expect one model to be preferred for, say, forecasts of a variable's level, and another model for prediction of that variable's change. The practitioner should decide which model is most appropriate for their particular situation, given their preferences. Rather than using GFESM as a summary invariant measure with which to choose a preferred model for all purposes, McNees argues that several models should be considered as each may be useful and preferable in a specified context. Again, the emphasis is on the user - once a particular use of a model is established, one of the available forecasts can be chosen in accordance with the practitioner's interests. Wallis supports this view: forecasters and forecast evaluators cannot take the loss functions of all users into consideration; it is best, therefore, to present information which can be used in a wide spectrum of individual cases (e.g. forecast comparisons in levels and changes, for different forecast horizons etc.), rather than produce one single summary evaluation conclusion based on GFESM. West notes that if a natural measure of

forecast quality exists in a given context, then it is relatively unimportant that different measures, which may be appropriate in other contexts, yield conflicting rankings of the competing forecasts. Baillie and Engle are two further exponents of this view that different models perform better in terms of prediction under different representations, with the question being which representation is of interest, rather than the seeking of an invariant criterion which, Engle claims, actually disguises the issue.

The second common criticism of the Clements & Hendry GFESM proposal relates to the focus on 1-step-ahead forecasts. Baillie, Meese and Newbold all give examples of how two misspecified models will be ranked differently by the MSFE criterion depending on whether the forecast horizon is one or two (or more). The danger of a purely 1-step-ahead analysis is thus highlighted, and economic situations frequently arise where one model may forecast the short term best, and another the long term. Baillie also adds that examples such as the prediction of stock returns, which are more predictable over long horizons than short ones, illustrate the need for long forecast horizon consideration. The GFESM analysis implies concentration on 1-step-ahead errors which loses this information. Furthermore, even when a longer horizon GFESM criterion is used, the different step-ahead forecast errors are implicitly weighted and a conclusive ranking is obtained; the implication is then that one model is preferable absolutely, and the fact that other models may perform better at a specified forecast horizon is masked.

Diebold and Engle note in criticism of GFESM that the criterion does not correspond easily to the established decision theory of minimising expected loss.

The conventional approach to forecast evaluation specifies some notion of loss, or cost of error, and selects the model which incurs the minimum expected loss. While the loss functions of users can rarely be accurately defined, this methodology points to a certain model for a certain use. This relates to the criticisms outlined above, where one forecast may well involve less loss for a given use compared to another forecast, which itself may be preferable if that use were to change. Diebold argues that rather than analysing the loss function associated with each particular problem, the GFESM measure is a generalisation whose primary advantage is that of convenience. GFESM is comprised of a determinant of an expected value, not an expected value of a determinant, hence it is difficult to see how it conforms to the concept of cost of error, and therefore how the measure fits with the intuition of decision via minimised loss.

Further criticisms of GFESM are also made. Wallis states that applied empirical work shows that there is commonly 'no unambiguous ranking of competing forecasts across variables, subperiods, and forecast horizons'. The Clements & Hendry GFESM criterion seeks to establish an unambiguous ranking; Wallis claims this only 'sweeps the problem under the carpet'. Howrey observes that cases frequently occur where the difference between the GFESMs of two competing models is indefinite, i.e. neither of the models dominates the other in all dimensions. In such cases it is possible that GFESM will choose the model which actually performs worst in terms of forecasting the individual variables of interest. Armstrong & Fildes (1995) argue that 'invariance of rankings to transformations is only one of the many criteria that are helpful for examining forecast accuracy'.

In response to these criticisms, Clements & Hendry reply to the discussants, reiterating the problems which may arise when using MSFE-based criteria for forecast evaluation. Rankings of forecasts obtained by such criteria may be unique to the representation used, and no generality to other isomorphic representations can be claimed. This creates obvious dangers when alternative representations of a model are of interest to the user. They also add further justification to the use of GFESM by highlighting its connections with predictive likelihood and arguing that it is a natural criterion since it ‘measures the volume of space around a forecast error, centred on zero, with smaller volumes being preferable’. This conviction of the need for invariance in forecast evaluation criteria is illustrated by the use of GFESM in Clements & Hendry (1995) where the imposition of unit roots and cointegrating restrictions in linear systems of $I(1)$ variables in levels, differences, and cointegrating combinations is assessed in terms of the impact on forecast accuracy, and also in Clements & Hendry (1996) where the issue of testing for seasonal unit roots in the context of forecasting is considered. However, in a more recent paper, Hendry (1996) tempers this conviction, stating that ‘although invariance is useful to determine a unique measure for a fixed model independently of its representation, it is not compelling, and often several forecast-accuracy indices are reported’.

Altogether, Clements & Hendry criticise the use of MSFE-based criteria in forecast evaluation due to their lack of invariance to linear non-singular scale-preserving transformations to the models concerned. The problems which may result are especially pertinent to cases where there is no unique data transformation of interest to the user. Instead, Clements & Hendry propose an invariant criterion - GFESM -

which has faced considerable criticism for its approach, underlying principles, and nature of reducing the forecast evaluation problem to a single number. The GFESM criterion has its drawbacks, therefore, but remains a useful contribution to the literature on the comparison of competing forecasts, overcoming the basic lack of invariance of the established measures.

5.3 The GFESM Criterion

5.3.1 Justification for GFESM

The justification for using GFESM as a criterion for ranking competing forecasts stems from an analysis of predictive likelihood. The system that Clements & Hendry (1993) consider follows a study by Engle & Yoo (1987) where forecasts are compared over horizons $1, \dots, h$. The study takes the form of a Monte Carlo simulation where the system is independently replicated n times. In this framework, the predictive likelihood of a given forecast can be examined. The previous section discussed the prediction of several variables, but for simplicity attention is now restricted to univariate series. The results and conclusions of the analysis follow through in a multivariate setting.

For a given replication, it is now assumed that there are two forecasts for each horizon, i.e. forecast 1 is f_{1k} , forecast 2 is f_{2k} , where $k = 1, \dots, h$. It can further be assumed that the actual values y_k are specified as follows:

$$y_k = g_k(f_{1k}, f_{2k}; \beta_k) + e_k; \quad k = 1, \dots, h \quad (5.9)$$

where e_k is an error term and g_k is a function of the two forecasts with parameter β_k , continuous such that y_k can be purely associated with one forecast and the error. The following assumption is also made about the errors:

$$(e_1, \dots, e_h) \sim N(0, \Omega) \quad (5.10)$$

i.e. the errors are assumed to follow a multivariate normal distribution with mean

zero and variance-covariance matrix Ω . In vector form the system is:

$$y = g(.) + e \quad (5.11)$$

where each element is an $(h \times 1)$ vector for a given replication. The joint density of y can then be written as follows:

$$y \sim N(g(.), \Omega)$$

$$f(y|\beta, \Omega) = (2\pi)^{-h/2} |\Omega|^{-1/2} \exp\{-\frac{1}{2}[y - g(.)]'\Omega^{-1}[y - g(.)]\} \quad (5.12)$$

Now the analysis so far has considered one replication. It is necessary then to find the joint density for all n replications (indexed $i = 1, \dots, n$):

$$f(y|\beta, \Omega) =$$

$$\prod_{i=1}^n (2\pi)^{-h/2} |\Omega|^{-1/2} \exp\{-\frac{1}{2}[y_i - g_i(.)]'\Omega^{-1}[y_i - g_i(.)]\} \quad (5.13)$$

The result is then the likelihood function for the system:

$$L(\beta, \Omega|y) =$$

$$(2\pi)^{-nh/2} |\Omega|^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n [y_i - g_i(.)]'\Omega^{-1}[y_i - g_i(.)]\right\} \quad (5.14)$$

Taking natural logs yields the log-likelihood function:

$$\ln L(\beta, \Omega|y) =$$

$$-\frac{nh}{2} \ln(2\pi) - \frac{n}{2} \ln(|\Omega|) - \frac{1}{2} \sum_{i=1}^n [y_i - g_i(.)]'\Omega^{-1}[y_i - g_i(.)] \quad (5.15)$$

which can be concentrated for Ω :

$$\ln L_c(\beta|y) = C - \frac{n}{2} \ln |\hat{\Omega}(\beta)| \quad (5.16)$$

$$\text{where } C = -\frac{nh}{2} [\ln(2\pi) - 1]$$

$$\hat{\Omega}(\beta) = n^{-1} \sum_{i=1}^n [y_i - g_i(\cdot)][y_i - g_i(\cdot)]'$$

The concentrated log-likelihood function therefore depends solely on $|\hat{\Omega}(\beta)|$. Now vectors (β^1, β^2) exist such that $g(f_1, f_2; \beta^j) = f_j$; $j = 1, 2$ which allows predictive likelihood to be used as a ranking criterion. Substitution of β^1 into the concentrated log-likelihood (5.16) gives a value for the predictive likelihood of forecast f_1 , and substitution of β^2 yields the likelihood of f_2 ; the forecast with the higher likelihood will thus be preferred (ranked first).

Now ranking on the basis of predictive likelihood in this way is the same as ranking on the basis of $|\hat{\Omega}(\beta)|$, the preferred forecast having the smallest value of $|\hat{\Omega}(\beta)|$. With enough replications, $\hat{\Omega} \xrightarrow{p} \Omega$, so predictive likelihood amounts to a forecast comparison based on an estimate of the population quantity $|\Omega|$. Given that

$|\Omega| = |\phi_h|$, the link with the GFESM criterion becomes clear, and justification for employing GFESM in the Clements & Hendry framework is derived with a firm and valid base in predictive likelihood.

However, when the framework is changed to the more realistic setup of constructing the GFESM criterion in the context of applied time series analysis, the justification for using the GFESM measure becomes less clear. In practice, estimation of Ω

comes about by finding the component sample variances and covariances. Thus the estimate $\hat{\Omega}$ is constructed as in the Clements & Hendry framework, with the crucial difference that the $i = 1, \dots, n$ replications are now n different starting points in the time series. The property of independence between replications which yielded the earlier justification no longer applies since overlapping occurs, e.g. $\hat{y}_T(3)$ is forecasting the same value as $\hat{y}_{T-1}(4)$, and many additional covariance terms are introduced. The joint density does not condense to a function of $|\hat{\Omega}(\beta)|$ as before, with more forecasts in existence than observations, and it is far from clear that the link between predictive likelihood and the GFESM criterion still holds. Clements & Hendry argue for the use of the GFESM measure, but when the estimate is constructed in the context of an applied time series, the justification for using an estimate of the population quantity $|\phi_h|$ as a forecast ranking criterion breaks down.

Having said this, the predictive likelihood justification can be maintained if the possibility of replication is viewed as a ‘thought experiment’. Under such a philosophy, the replication itself need not be executed and the population quantity $|\Omega|$ of the thought experiment could be consistently estimated from a series of forecasts. Then if the population quantity $|\Omega|$, i.e. GFESM, is a useful quantity to estimate in the thought experiment, it is equally useful and sensible to estimate it in the real world.

In closing, one result which holds under both of the frameworks mentioned above is that the GFESM criterion will always rank the true model as the best if it is available, provided the errors are normal. If the true model is evaluated in

competition with one or more misspecified models, then the GFESM for the true model must be the lowest asymptotically. This follows in the normality framework because if the true β is such that y is purely associated with one model's forecast (the true model's), then the estimate $\hat{\Omega}(\beta)$ for that β will, in the limit, be the minimum $|\Omega|$ possible, thus the likelihood and the GFESM criterion will give this true model the highest ranking.

5.3.2 GFESM Behaviour Under Model Misspecification

The GFESM criterion for the evaluation of competing forecasts is now studied to examine its behaviour when misspecified models are employed. In applied forecasting, the true model is rarely, if ever, known. Contending forecasts are consequently generated by a number of approximations to this unknown truth. Analysis of misspecifications is therefore vital. This sub-section considers the behaviour of GFESM for the simplest non-trivial example of two rival forecasts generated by two non-nested misspecified models. More specifically, the true model is assumed to be an ARMA(1,1) process, with the two misspecifications being an AR(1) model and an MA(1) model:

$$\text{true model A: } y_t = \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \quad t = 1, \dots, T \quad (5.17)$$

$$\text{model B: } y_t = \alpha y_{t-1} + \eta_t \quad t = 1, \dots, T \quad (5.18)$$

$$\text{model C: } y_t = u_t - \beta u_{t-1} \quad t = 1, \dots, T \quad (5.19)$$

where $\varepsilon_t \sim \text{IID}(0, \sigma^2)$. The criterion's behaviour is studied by examining the rankings generated when comparing the three models given above in a pairwise

manner. 1-step- and 2-steps-ahead forecasts are considered, and comparison is made with the rankings suggested by the MSFE evaluation criterion.

5.3.2a Specification of Criteria

True Model A

For the true model, the optimal forecasts and associated errors are as follows:

$$\text{1-step-ahead: } \hat{y}_T(1) = \phi y_T - \theta \varepsilon_T$$

$$e_T(1) = \varepsilon_{T+1}$$

$$\text{2-steps-ahead: } \hat{y}_T(2) = \phi^2 y_T - \phi \theta \varepsilon_T$$

$$e_T(2) = \varepsilon_{T+2} + (\phi - \theta) \varepsilon_{T+1}$$

From this, the MSFE and GFESM criteria can be found (see section appendix):

$$\text{MSFE}_{a1} = \text{GFESM}_{a1} = \sigma^2 \quad (5.20)$$

$$\text{MSFE}_{a2} = \sigma^2 [1 + (\phi - \theta)^2] \quad (5.21)$$

$$\text{GFESM}_{a2} = \sigma^4 \quad (5.22)$$

Model B

The AR(1) misspecified model has the representation given in (5.18). Now in practice the autoregressive parameter α must be estimated. The least squares approach minimises $\sum \eta_t^2$ and yields the following estimator:

$$\hat{\alpha} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$$

It is then possible to note that:

$$\hat{\alpha} = \frac{n^{-1} \sum y_t y_{t-1}}{n^{-1} \sum y_{t-1}^2} \xrightarrow{p} \frac{C(y_t, y_{t-1})}{V(y_t)}$$

provided that y_t is a stationary process. The estimator $\hat{\alpha}$ therefore tends in probability to the first autocorrelation of the true ARMA(1,1) model, which is known for given values of the parameters ϕ and θ :

$$\hat{\alpha} \xrightarrow{p} \rho(1) = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta} \quad (5.23)$$

The optimal forecasts and forecast errors can now be found:

$$\text{1-step-ahead: } \hat{y}_T(1) = \hat{\alpha} y_T$$

$$e_T(1) = (\phi - \hat{\alpha}) y_T + \varepsilon_{T+1} - \theta \varepsilon_T$$

$$\text{2-steps-ahead: } \hat{y}_T(2) = \hat{\alpha}^2 y_T$$

$$e_T(2) = (\phi^2 - \hat{\alpha}^2) y_T + \varepsilon_{T+2} + (\phi - \theta) \varepsilon_{T+1} - \phi \theta \varepsilon_T$$

As with the true model, the MSFE and GFESM criteria can be derived from this information (see section appendix):

$$\text{MSFE}_{b1} = \text{GFESM}_{b1} \xrightarrow{p} \sigma^2 (1 + p_2^{-1} \theta^2 p_1^2) \quad (5.24)$$

$$\text{where } p_1 = \phi - \theta$$

$$p_2 = 1 + \theta^2 - 2\phi\theta$$

$$\begin{aligned} \text{MSFE}_{b2} \xrightarrow{p} \sigma^2 [& (\phi^2 - p_2^{-2} p_1^2 p_3^2)^2 p_4^{-1} p_2 + 1 + p_1^2 + \phi^2 \theta^2 \\ & - 2\phi\theta(\phi^2 - p_2^{-2} p_1^2 p_3^2)] \end{aligned} \quad (5.25)$$

where $p_3 = 1 - \phi\theta$

$$p_4 = 1 - \phi^2$$

$$\begin{aligned} \text{GFESM}_{b_2} \xrightarrow{p} \sigma^4 \{ & (1 + p_2^{-1} \theta^2 p_1^2 [(\phi^2 - p_2^{-2} p_1^2 p_3^2)^2 p_4^{-1} p_2 \\ & + 1 + p_1^2 + \phi^2 \theta^2 - 2\phi\theta(\phi^2 - p_2^{-2} p_1^2 p_3^2)] \\ & - (p_1 + p_2^{-1} \phi \theta^2 p_1^2)^2 \} \end{aligned} \quad (5.26)$$

Model C

This second model misspecification takes an MA(1) formulation as given in (5.19).

As with model B, the parameter β must be estimated, and the least squares method of doing this amounts to minimising Σu_t^2 . Now it is possible to note that:

$$u_t = (1 - \beta L)^{-1} y_t$$

which leads to the following result:

$$\min_{\beta} \Sigma u_t^2 \xrightarrow{p} \min_{\beta} V[(1 - \beta L)^{-1} y_t] \quad (5.27)$$

The true model y_t is ARMA(1,1) which then gives, by substitution into (5.27):

$$\min_{\beta} \Sigma u_t^2 \xrightarrow{p} \min_{\beta} V[(1 - \beta L)^{-1} (1 - \phi L)^{-1} (1 - \theta L) \varepsilon_t] \quad (5.28)$$

The least squares estimator $\hat{\beta}$ can now be found by minimising the right hand side of (5.28), i.e. minimising the variance of an ARMA(2,1) process. For given values of ϕ , θ and β , the variance of such a process can be written as (see section appendix):

$$\gamma_0^* = \sigma^2 \left[\frac{1 + \theta^2 - 2\phi\theta + \beta(\phi - 2\theta + \phi\theta^2)}{(1 - \phi^2)(1 - \beta\phi - \beta^2 + \beta^3\phi)} \right]; -1 < \beta < 1 \quad (5.29)$$

Minimisation of (5.29) with respect to β involves solving the following function for β (see section appendix):

$$(1 - \beta\phi - \beta^2 + \beta^3\phi)(\phi - 2\theta + \phi\theta^2) + (1 + \theta^2 - 2\phi\theta + \beta(\phi - 2\theta + \phi\theta^2))(\phi + 2\beta - 3\beta^2\phi) = 0 \quad (5.30)$$

and the result is the estimator $\hat{\beta}$, found numerically for given values of ϕ and θ .

Once again, the optimal forecasts and associated errors can now be derived:

$$\text{1-step-ahead: } \hat{y}_T(1) = -\hat{\beta}u_T = -\hat{\beta}(1 - \hat{\beta}L)^{-1}y_T$$

$$e_T(1) = y_{T+1} + \hat{\beta}(1 - \hat{\beta}L)^{-1}y_T$$

$$\text{2-steps-ahead: } \hat{y}_T(2) = 0$$

$$e_T(2) = \phi^2 y_T + \varepsilon_{T+2} + (\phi - \theta)\varepsilon_{T+1} - \phi\theta\varepsilon_T$$

The MSFE and GFESM criteria for the MA(1) model can be specified using these results (see section appendix):

$$\begin{aligned} \text{MSFE}_{c1} = \text{GFESM}_{c1} &= \sigma^2 [p_4(1 - \beta\phi - \beta^2 + \beta^3\phi)]^{-1} \\ &\times [p_2 + \beta(\phi - 2\theta + \phi\theta^2)] \end{aligned} \quad (5.31)$$

$$\text{MSFE}_{c2} = \sigma^2 p_4^{-1} p_2 \quad (5.32)$$

$$\begin{aligned} \text{GFESM}_{c_2} = \sigma^4 \{ & p_4^{-2} (1 - \beta\phi - \beta^2 + \beta^3\phi)^{-1} p_2 [p_2 \\ & + \beta(\phi - 2\theta + \phi\theta^2)] - [p_4^{-1} (1 - \beta\phi)^{-1} p_1 p_3]^2 \} \quad (5.33) \end{aligned}$$

where the p_j are as defined above.

5.3.2b Comparison of Criteria

Behaviour of the GFESM criterion can now be examined by comparison with the MSFE criterion. Such comparison is performed by considering the rankings that each measure suggests for the evaluation of a given pair of models. Noting that the 1-step-ahead GFESM is simply the 1-step-ahead MSFE, the following ranking comparisons are analysed:

- (i) MSFE_{b1} vs. MSFE_{a1}
- (ii) MSFE_{b2} vs. MSFE_{a2}
- (iii) MSFE_{c1} vs. MSFE_{a1}
- (iv) MSFE_{c2} vs. MSFE_{a2}
- (v) MSFE_{b1} vs. MSFE_{c1}
- (vi) MSFE_{b2} vs. MSFE_{c2}
- (vii) GFESM_{b2} vs. GFESM_{a2}
- (viii) GFESM_{c2} vs. GFESM_{a2}
- (ix) GFESM_{b2} vs. GFESM_{c2}

In each case, the ranking can be established by taking a ratio of the two criterion values, e.g. to compare MSFE_{b1} vs. MSFE_{a1} , the ratio $\text{MSFE}_{b1} / \text{MSFE}_{a1}$ is used.

Use of such ratios cancels out the σ^2 and σ^4 terms in the 1-step- and 2-steps-ahead

criteria respectively. It is then possible to evaluate the ratios for given values of ϕ and θ . In this study, a large range of possible (ϕ, θ) values are considered, initially in steps of 0.1 (i.e. $\phi = -0.9, -0.8, \dots, 0.9$; $\theta = 0, 0.1, \dots, 0.9$). The extreme values are also considered by including $\phi = -0.99, 0.99$; $\theta = 0.99$. Note that negative values of θ are unnecessary as their inclusion would merely generate duplicates of other values, since ratio values for $(\phi, -\theta)$ are identical to those for $(-\phi, \theta)$.

Taking the $MSFE_{b1}$ vs. $MSFE_{a1}$ comparison as an example, the interpretation of the ratio values can be made clear: a value greater than 1 implies $MSFE_{b1} > MSFE_{a1}$, and therefore the true model is preferred (ranked above model B) for 1-step-ahead forecasts using the MSFE criterion; a value less than 1 leads to the reverse conclusion; and a value equal to 1 implies the models have equal MSFEs 1-step-ahead and thus the conclusion is indeterminate. The ratio values for the nine ranking comparisons listed above are given in tables 5.1-5.9.

The first point to note regarding the results occurs where the models are ranked equally, i.e. a ratio value of 1 is obtained. This happens in three cases. Firstly, where $\theta = 0$, the true model and the AR(1) model (model B) are identical, and thus their forecasts will also be identical in the limit. This same result occurs for the true model and the MA(1) model (model C) when $\phi = 0$. Thirdly, when $\phi = \theta$, a common factor exists and the true model reduces to a white noise process, i.e. $y_t = \varepsilon_t$. The estimates $\hat{\alpha}$ and $\hat{\beta}$ in models B and C respectively will therefore tend in probability

Table 5.1

AR(1) vs. ARMA(1,1) comparison: $MSFE_{b1} / MSFE_{a1} = GFESM_{b1} / GFESM_{a1}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	1.0000	1.0098	1.0394	1.0889	1.1584	1.2478	1.3572	1.4866	1.6360	1.8055	1.9752
$\phi = -0.9$	1.0000	1.0084	1.0346	1.0795	1.1438	1.2279	1.3320	1.4561	1.6005	1.7651	1.9306
$\phi = -0.8$	1.0000	1.0069	1.0294	1.0694	1.1280	1.2061	1.3041	1.4224	1.5611	1.7203	1.8811
$\phi = -0.7$	1.0000	1.0056	1.0245	1.0596	1.1126	1.1846	1.2765	1.3888	1.5217	1.6754	1.8316
$\phi = -0.6$	1.0000	1.0043	1.0200	1.0503	1.0976	1.1635	1.2492	1.3554	1.4825	1.6306	1.7821
$\phi = -0.5$	1.0000	1.0032	1.0158	1.0414	1.0831	1.1429	1.2222	1.3222	1.4433	1.5858	1.7326
$\phi = -0.4$	1.0000	1.0023	1.0120	1.0332	1.0692	1.1227	1.1957	1.2892	1.4042	1.5411	1.6831
$\phi = -0.3$	1.0000	1.0015	1.0086	1.0255	1.0560	1.1032	1.1695	1.2565	1.3653	1.4963	1.6336
$\phi = -0.2$	1.0000	1.0009	1.0057	1.0186	1.0436	1.0845	1.1440	1.2242	1.3265	1.4517	1.5841
$\phi = -0.1$	1.0000	1.0004	1.0033	1.0125	1.0323	1.0667	1.1192	1.1924	1.2880	1.4070	1.5346
$\phi = 0$	1.0000	1.0001	1.0015	1.0074	1.0221	1.0500	1.0953	1.1611	1.2498	1.3625	1.4851
$\phi = 0.1$	1.0000	1.0000	1.0004	1.0035	1.0133	1.0348	1.0726	1.1307	1.2119	1.3180	1.4356
$\phi = 0.2$	1.0000	1.0001	1.0000	1.0009	1.0064	1.0214	1.0514	1.1012	1.1745	1.2737	1.3861
$\phi = 0.3$	1.0000	1.0004	1.0004	1.0000	1.0017	1.0105	1.0324	1.0733	1.1379	1.2296	1.3366
$\phi = 0.4$	1.0000	1.0010	1.0018	1.0011	1.0000	1.0029	1.0164	1.0474	1.1024	1.1858	1.2872
$\phi = 0.5$	1.0000	1.0018	1.0043	1.0046	1.0021	1.0000	1.0047	1.0248	1.0686	1.1424	1.2377
$\phi = 0.6$	1.0000	1.0028	1.0080	1.0111	1.0094	1.0038	1.0000	1.0075	1.0376	1.0999	1.1882
$\phi = 0.7$	1.0000	1.0041	1.0132	1.0215	1.0240	1.0182	1.0069	1.0000	1.0123	1.0589	1.1387
$\phi = 0.8$	1.0000	1.0058	1.0200	1.0369	1.0492	1.0500	1.0360	1.0132	1.0000	1.0219	1.0893
$\phi = 0.9$	1.0000	1.0077	1.0288	1.0589	1.0909	1.1143	1.1157	1.0852	1.0320	1.0000	1.0401
$\phi = 0.99$	1.0000	1.0098	1.0388	1.0864	1.1513	1.2309	1.3183	1.3962	1.4126	1.2343	1.0000

Note:- 1 denotes indeterminate, >1 denotes ARMA(1,1) preferred, and <1 denotes AR(1) preferred

Table 5.2

AR(1) vs. ARMA(1,1) comparison: $MSFE_{b2} / MSFE_{a2}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	1.0000	1.0045	1.0162	1.0331	1.0535	1.0761	1.1000	1.1246	1.1492	1.1736	1.1952
$\phi = -0.9$	1.0000	1.0041	1.0148	1.0302	1.0486	1.0690	1.0905	1.1123	1.1342	1.1557	1.1746
$\phi = -0.8$	1.0000	1.0036	1.0133	1.0271	1.0437	1.0619	1.0808	1.0999	1.1188	1.1372	1.1532
$\phi = -0.7$	1.0000	1.0032	1.0118	1.0244	1.0393	1.0556	1.0722	1.0887	1.1048	1.1202	1.1335
$\phi = -0.6$	1.0000	1.0027	1.0104	1.0218	1.0354	1.0500	1.0647	1.0790	1.0925	1.1051	1.1157
$\phi = -0.5$	1.0000	1.0022	1.0089	1.0193	1.0318	1.0452	1.0583	1.0706	1.0819	1.0920	1.1001
$\phi = -0.4$	1.0000	1.0017	1.0074	1.0168	1.0285	1.0410	1.0530	1.0638	1.0731	1.0810	1.0869
$\phi = -0.3$	1.0000	1.0012	1.0059	1.0143	1.0253	1.0372	1.0485	1.0584	1.0663	1.0724	1.0764
$\phi = -0.2$	1.0000	1.0007	1.0043	1.0116	1.0219	1.0336	1.0448	1.0543	1.0614	1.0662	1.0687
$\phi = -0.1$	1.0000	1.0004	1.0027	1.0087	1.0182	1.0298	1.0414	1.0512	1.0583	1.0625	1.0640
$\phi = 0$	1.0000	1.0001	1.0014	1.0057	1.0141	1.0256	1.0379	1.0487	1.0566	1.0611	1.0625
$\phi = 0.1$	1.0000	1.0000	1.0004	1.0030	1.0097	1.0206	1.0336	1.0461	1.0558	1.0618	1.0639
$\phi = 0.2$	1.0000	1.0001	1.0000	1.0009	1.0052	1.0147	1.0280	1.0425	1.0550	1.0637	1.0680
$\phi = 0.3$	1.0000	1.0004	1.0004	1.0000	1.0016	1.0083	1.0209	1.0370	1.0530	1.0660	1.0739
$\phi = 0.4$	1.0000	1.0009	1.0019	1.0011	1.0000	1.0026	1.0124	1.0289	1.0484	1.0668	1.0803
$\phi = 0.5$	1.0000	1.0016	1.0044	1.0049	1.0023	1.0000	1.0042	1.0181	1.0398	1.0643	1.0853
$\phi = 0.6$	1.0000	1.0024	1.0078	1.0119	1.0105	1.0042	1.0000	1.0065	1.0266	1.0563	1.0863
$\phi = 0.7$	1.0000	1.0033	1.0120	1.0222	1.0270	1.0210	1.0077	1.0000	1.0104	1.0408	1.0801
$\phi = 0.8$	1.0000	1.0041	1.0166	1.0354	1.0533	1.0587	1.0432	1.0151	1.0000	1.0183	1.0637
$\phi = 0.9$	1.0000	1.0049	1.0210	1.0496	1.0881	1.1260	1.1404	1.1067	1.0375	1.0000	1.0345
$\phi = 0.99$	1.0000	1.0055	1.0241	1.0597	1.1161	1.1964	1.3015	1.4216	1.4996	1.2979	1.0000

Note:- 1 denotes indeterminate, >1 denotes ARMA(1,1) preferred, and <1 denotes AR(1) preferred

Table 5.3

MA(1) vs. ARMA(1,1) comparison: $MSFE_{c1} / MSFE_{a1} = GFESM_{c1} / GFESM_{a1}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	15.238	17.821	20.607	23.599	26.793	30.193	33.796	37.604	41.616	45.831	49.802
$\phi = -0.9$	2.3322	2.5660	2.8134	3.0740	3.3478	3.6347	3.9346	4.2474	4.5732	4.9117	5.2274
$\phi = -0.8$	1.5185	1.6210	1.7291	1.8425	1.9613	2.0851	2.2139	2.3476	2.4862	2.6296	2.7627
$\phi = -0.7$	1.2476	1.3053	1.3664	1.4307	1.4982	1.5686	1.6417	1.7176	1.7961	1.8772	1.9523
$\phi = -0.6$	1.1218	1.1563	1.1934	1.2329	1.2745	1.3181	1.3636	1.4109	1.4598	1.5104	1.5572
$\phi = -0.5$	1.0569	1.0773	1.0998	1.1241	1.1501	1.1775	1.2063	1.2363	1.2675	1.2999	1.3299
$\phi = -0.4$	1.0234	1.0347	1.0475	1.0617	1.0772	1.0937	1.1113	1.1299	1.1493	1.1695	1.1883
$\phi = -0.3$	1.0076	1.0129	1.0193	1.0266	1.0349	1.0439	1.0537	1.0641	1.0751	1.0868	1.0977
$\phi = -0.2$	1.0015	1.0034	1.0058	1.0087	1.0122	1.0161	1.0205	1.0253	1.0304	1.0359	1.0411
$\phi = -0.1$	1.0001	1.0004	1.0009	1.0015	1.0023	1.0033	1.0044	1.0056	1.0070	1.0085	1.0099
$\phi = 0$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\phi = 0.1$	1.0001	1.0000	1.0001	1.0004	1.0010	1.0018	1.0028	1.0042	1.0058	1.0078	1.0099
$\phi = 0.2$	1.0015	1.0004	1.0000	1.0004	1.0018	1.0043	1.0080	1.0132	1.0202	1.0295	1.0403
$\phi = 0.3$	1.0076	1.0035	1.0009	1.0000	1.0011	1.0045	1.0110	1.0213	1.0369	1.0604	1.0938
$\phi = 0.4$	1.0234	1.0139	1.0066	1.0018	1.0000	1.0021	1.0091	1.0230	1.0469	1.0875	1.1550
$\phi = 0.5$	1.0569	1.0387	1.0233	1.0112	1.0030	1.0000	1.0037	1.0168	1.0440	1.0942	1.1736
$\phi = 0.6$	1.1218	1.0901	1.0619	1.0376	1.0182	1.0050	1.0000	1.0064	1.0298	1.0800	1.1604
$\phi = 0.7$	1.2476	1.1939	1.1444	1.1000	1.0614	1.0301	1.0084	1.0000	1.0112	1.0528	1.1287
$\phi = 0.8$	1.5185	1.4220	1.3319	1.2487	1.1735	1.1077	1.0536	1.0153	1.0000	1.0210	1.0868
$\phi = 0.9$	2.3322	2.1120	1.9060	1.7145	1.5385	1.3791	1.2387	1.1214	1.0356	1.0000	1.0398
$\phi = 0.99$	15.238	12.860	10.687	8.7180	6.9544	5.3964	4.0452	2.9032	1.9767	1.2901	1.0000

Note:- 1 denotes indeterminate, >1 denotes ARMA(1,1) preferred, and <1 denotes MA(1) preferred

Table 5.4

MA(1) vs. ARMA(1,1) comparison: $MSFE_{c2} / MSFE_{a2}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	25.378	27.743	29.867	31.764	33.454	34.956	36.292	37.479	38.536	39.479	40.242
$\phi = -0.9$	2.9078	3.1316	3.3341	3.5160	3.6783	3.8229	3.9514	4.0656	4.1672	4.2577	4.3307
$\phi = -0.8$	1.6938	1.7956	1.8889	1.9734	2.0492	2.1169	2.1772	2.2308	2.2784	2.3208	2.3549
$\phi = -0.7$	1.3160	1.3749	1.4300	1.4804	1.5260	1.5670	1.6036	1.6362	1.6652	1.6909	1.7116
$\phi = -0.6$	1.1489	1.1850	1.2195	1.2517	1.2813	1.3080	1.3320	1.3534	1.3725	1.3894	1.4031
$\phi = -0.5$	1.0667	1.0882	1.1096	1.1301	1.1492	1.1667	1.1825	1.1967	1.2094	1.2207	1.2298
$\phi = -0.4$	1.0263	1.0381	1.0504	1.0626	1.0743	1.0852	1.0952	1.1043	1.1124	1.1197	1.1255
$\phi = -0.3$	1.0082	1.0136	1.0198	1.0262	1.0325	1.0386	1.0443	1.0495	1.0541	1.0584	1.0618
$\phi = -0.2$	1.0016	1.0034	1.0057	1.0083	1.0110	1.0137	1.0163	1.0186	1.0208	1.0228	1.0244
$\phi = -0.1$	1.0001	1.0004	1.0008	1.0014	1.0020	1.0027	1.0033	1.0039	1.0045	1.0051	1.0055
$\phi = 0$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\phi = 0.1$	1.0001	1.0000	1.0001	1.0004	1.0008	1.0014	1.0020	1.0027	1.0033	1.0039	1.0045
$\phi = 0.2$	1.0016	1.0004	1.0000	1.0004	1.0016	1.0034	1.0057	1.0083	1.0110	1.0137	1.0160
$\phi = 0.3$	1.0082	1.0038	1.0010	1.0000	1.0010	1.0038	1.0082	1.0136	1.0198	1.0262	1.0319
$\phi = 0.4$	1.0263	1.0157	1.0073	1.0019	1.0000	1.0019	1.0073	1.0157	1.0263	1.0381	1.0492
$\phi = 0.5$	1.0667	1.0460	1.0275	1.0128	1.0033	1.0000	1.0033	1.0128	1.0275	1.0460	1.0645
$\phi = 0.6$	1.1489	1.1125	1.0776	1.0464	1.0216	1.0056	1.0000	1.0056	1.0216	1.0464	1.0743
$\phi = 0.7$	1.3160	1.2543	1.1922	1.1325	1.0793	1.0370	1.0095	1.0000	1.0095	1.0370	1.0745
$\phi = 0.8$	1.6938	1.5846	1.4706	1.3556	1.2452	1.1468	1.0684	1.0176	1.0000	1.0176	1.0619
$\phi = 0.9$	2.9078	2.6637	2.4020	2.1285	1.8526	1.5880	1.3520	1.1640	1.0422	1.0000	1.0343
$\phi = 0.99$	25.378	22.769	19.926	16.886	13.717	10.536	7.5021	4.8207	2.7160	1.3957	1.0000

Note:- 1 denotes indeterminate, >1 denotes ARMA(1,1) preferred, and <1 denotes MA(1) preferred

Table 5.5

AR(1) vs. MA(1) comparison: $MSFE_{b1} / MSFE_{c1} = GFESM_{b1} / GFESM_{c1}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	0.0656	0.0567	0.0504	0.0461	0.0432	0.0413	0.0402	0.0395	0.0393	0.0394	0.0397
$\phi = -0.9$	0.4288	0.3930	0.3677	0.3512	0.3417	0.3378	0.3385	0.3428	0.3500	0.3594	0.3693
$\phi = -0.8$	0.6585	0.6212	0.5954	0.5804	0.5751	0.5784	0.5891	0.6059	0.6279	0.6542	0.6809
$\phi = -0.7$	0.8015	0.7704	0.7498	0.7406	0.7426	0.7552	0.7776	0.8086	0.8472	0.8925	0.9382
$\phi = -0.6$	0.8915	0.8686	0.8547	0.8519	0.8611	0.8827	0.9161	0.9607	1.0155	1.0796	1.1444
$\phi = -0.5$	0.9462	0.9312	0.9236	0.9264	0.9417	0.9706	1.0133	1.0695	1.1386	1.2200	1.3028
$\phi = -0.4$	0.9771	0.9687	0.9661	0.9731	0.9926	1.0265	1.0759	1.1410	1.2218	1.3177	1.4163
$\phi = -0.3$	0.9925	0.9888	0.9896	0.9989	1.0204	1.0568	1.1100	1.1809	1.2699	1.3769	1.4883
$\phi = -0.2$	0.9985	0.9975	0.9999	1.0098	1.0310	1.0673	1.1210	1.1941	1.2874	1.4014	1.5216
$\phi = -0.1$	0.9999	1.0000	1.0025	1.0110	1.0299	1.0632	1.1143	1.1858	1.2791	1.3952	1.5195
$\phi = 0$	1.0000	1.0001	1.0015	1.0074	1.0221	1.0500	1.0953	1.1611	1.2498	1.3625	1.4851
$\phi = 0.1$	0.9999	1.0000	1.0003	1.0031	1.0124	1.0330	1.0696	1.1260	1.2049	1.3079	1.4216
$\phi = 0.2$	0.9985	0.9997	1.0000	1.0005	1.0046	1.0171	1.0431	1.0869	1.1512	1.2372	1.3325
$\phi = 0.3$	0.9925	0.9969	0.9995	1.0000	1.0007	1.0060	1.0212	1.0509	1.0974	1.1595	1.2220
$\phi = 0.4$	0.9771	0.9872	0.9953	0.9993	1.0000	1.0009	1.0072	1.0239	1.0530	1.0904	1.1144
$\phi = 0.5$	0.9462	0.9644	0.9814	0.9935	0.9991	1.0000	1.0010	1.0079	1.0235	1.0440	1.0546
$\phi = 0.6$	0.8915	0.9199	0.9493	0.9744	0.9913	0.9988	1.0000	1.0011	1.0076	1.0184	1.0239
$\phi = 0.7$	0.8015	0.8411	0.8853	0.9287	0.9648	0.9884	0.9985	1.0000	1.0011	1.0058	1.0089
$\phi = 0.8$	0.6585	0.7073	0.7658	0.8303	0.8941	0.9479	0.9833	0.9980	1.0000	1.0009	1.0024
$\phi = 0.9$	0.4288	0.4771	0.5398	0.6176	0.7091	0.8080	0.9007	0.9678	0.9965	1.0000	1.0003
$\phi = 0.99$	0.0656	0.0785	0.0972	0.1246	0.1656	0.2281	0.3259	0.4809	0.7146	0.9568	1.0000

Note:- 1 denotes indeterminate, >1 denotes MA(1) preferred, and <1 denotes AR(1) preferred

Table 5.6

AR(1) vs. MA(1) comparison: $MSFE_{b2} / MSFE_{c2}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	0.0394	0.0362	0.0340	0.0325	0.0315	0.0308	0.0303	0.0300	0.0298	0.0297	0.0297
$\phi = -0.9$	0.3439	0.3206	0.3044	0.2930	0.2851	0.2796	0.2760	0.2736	0.2722	0.2714	0.2712
$\phi = -0.8$	0.5904	0.5589	0.5364	0.5205	0.5093	0.5016	0.4964	0.4931	0.4910	0.4900	0.4897
$\phi = -0.7$	0.7599	0.7296	0.7076	0.6920	0.6811	0.6736	0.6686	0.6654	0.6635	0.6625	0.6622
$\phi = -0.6$	0.8704	0.8462	0.8285	0.8163	0.8081	0.8028	0.7994	0.7972	0.7960	0.7954	0.7952
$\phi = -0.5$	0.9375	0.9209	0.9092	0.9020	0.8979	0.8959	0.8950	0.8946	0.8946	0.8945	0.8945
$\phi = -0.4$	0.9744	0.9649	0.9591	0.9569	0.9574	0.9592	0.9614	0.9633	0.9647	0.9655	0.9657
$\phi = -0.3$	0.9919	0.9877	0.9864	0.9884	0.9930	0.9987	1.0041	1.0085	1.0116	1.0132	1.0137
$\phi = -0.2$	0.9984	0.9973	0.9985	1.0032	1.0108	1.0196	1.0281	1.0350	1.0398	1.0424	1.0432
$\phi = -0.1$	0.9999	1.0000	1.0019	1.0073	1.0162	1.0271	1.0380	1.0471	1.0535	1.0571	1.0582
$\phi = 0$	1.0000	1.0001	1.0014	1.0057	1.0141	1.0256	1.0379	1.0487	1.0566	1.0611	1.0625
$\phi = 0.1$	0.9999	1.0000	1.0003	1.0026	1.0088	1.0191	1.0315	1.0433	1.0523	1.0576	1.0592
$\phi = 0.2$	0.9984	0.9997	1.0000	1.0005	1.0036	1.0112	1.0222	1.0339	1.0435	1.0494	1.0512
$\phi = 0.3$	0.9919	0.9966	0.9995	1.0000	1.0006	1.0045	1.0126	1.0231	1.0326	1.0388	1.0407
$\phi = 0.4$	0.9744	0.9854	0.9946	0.9992	1.0000	1.0007	1.0050	1.0129	1.0216	1.0277	1.0297
$\phi = 0.5$	0.9375	0.9576	0.9775	0.9922	0.9990	1.0000	1.0009	1.0052	1.0119	1.0175	1.0195
$\phi = 0.6$	0.8704	0.9010	0.9353	0.9670	0.9891	0.9986	1.0000	1.0009	1.0048	1.0094	1.0112
$\phi = 0.7$	0.7599	0.7998	0.8489	0.9026	0.9515	0.9846	0.9982	1.0000	1.0009	1.0037	1.0052
$\phi = 0.8$	0.5904	0.6337	0.6913	0.7638	0.8459	0.9232	0.9764	0.9975	1.0000	1.0007	1.0017
$\phi = 0.9$	0.3439	0.3773	0.4251	0.4931	0.5874	0.7091	0.8435	0.9508	0.9954	1.0000	1.0002
$\phi = 0.99$	0.0394	0.0442	0.0514	0.0628	0.0814	0.1136	0.1735	0.2949	0.5521	0.9299	1.0000

Note:- 1 denotes indeterminate, >1 denotes MA(1) preferred, and <1 denotes AR(1) preferred

Table 5.7

AR(1) vs. ARMA(1,1) comparison: $\text{GFESM}_{b2} / \text{GFESM}_{a2}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	1.0000	1.0101	1.0416	1.0981	1.1856	1.3125	1.4896	1.7300	2.0495	2.4660	2.9406
$\phi = -0.9$	1.0000	1.0098	1.0408	1.0966	1.1832	1.3091	1.4849	1.7240	2.0418	2.4563	2.9289
$\phi = -0.8$	1.0000	1.0091	1.0385	1.0923	1.1764	1.2994	1.4717	1.7067	2.0197	2.4286	2.8955
$\phi = -0.7$	1.0000	1.0081	1.0351	1.0856	1.1659	1.2842	1.4511	1.6796	1.9850	2.3852	2.8432
$\phi = -0.6$	1.0000	1.0068	1.0308	1.0772	1.1523	1.2645	1.4241	1.6441	1.9396	2.3283	2.7745
$\phi = -0.5$	1.0000	1.0055	1.0259	1.0673	1.1363	1.2410	1.3919	1.6015	1.8849	2.2598	2.6918
$\phi = -0.4$	1.0000	1.0041	1.0207	1.0565	1.1184	1.2146	1.3553	1.5530	1.8227	2.1817	2.5976
$\phi = -0.3$	1.0000	1.0028	1.0156	1.0453	1.0994	1.1861	1.3155	1.5000	1.7543	2.0958	2.4938
$\phi = -0.2$	1.0000	1.0016	1.0107	1.0342	1.0799	1.1562	1.2733	1.4434	1.6812	2.0036	2.3824
$\phi = -0.1$	1.0000	1.0008	1.0064	1.0237	1.0606	1.1259	1.2298	1.3845	1.6045	1.9068	2.2654
$\phi = 0$	1.0000	1.0002	1.0030	1.0144	1.0424	1.0961	1.1860	1.3244	1.5257	1.8069	2.1444
$\phi = 0.1$	1.0000	1.0000	1.0008	1.0069	1.0260	1.0678	1.1430	1.2641	1.4457	1.7050	2.0209
$\phi = 0.2$	1.0000	1.0002	1.0000	1.0018	1.0127	1.0422	1.1020	1.2049	1.3660	1.6025	1.8963
$\phi = 0.3$	1.0000	1.0008	1.0009	1.0000	1.0035	1.0209	1.0645	1.1482	1.2875	1.5005	1.7719
$\phi = 0.4$	1.0000	1.0018	1.0036	1.0021	1.0000	1.0059	1.0327	1.0957	1.2118	1.4001	1.6488
$\phi = 0.5$	1.0000	1.0032	1.0082	1.0089	1.0042	1.0000	1.0095	1.0499	1.1405	1.3024	1.5280
$\phi = 0.6$	1.0000	1.0048	1.0146	1.0212	1.0185	1.0077	1.0000	1.0151	1.0764	1.2087	1.4103
$\phi = 0.7$	1.0000	1.0066	1.0224	1.0390	1.0459	1.0359	1.0138	1.0000	1.0247	1.1210	1.2962
$\phi = 0.8$	1.0000	1.0083	1.0308	1.0616	1.0889	1.0961	1.0717	1.0266	1.0000	1.0442	1.1865
$\phi = 0.9$	1.0000	1.0096	1.0382	1.0851	1.1453	1.2026	1.2236	1.1724	1.0648	1.0000	1.0817
$\phi = 0.99$	1.0000	1.0101	1.0416	1.0979	1.1849	1.3099	1.4795	1.6863	1.8366	1.5113	1.0000

Note:- 1 denotes indeterminate, >1 denotes ARMA(1,1) preferred, and <1 denotes AR(1) preferred

Table 5.8

MA(1) vs. ARMA(1,1) comparison: $\text{GFESM}_{c2} / \text{GFESM}_{a2}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	78.416	92.020	104.82	115.95	124.44	129.20	129.01	122.54	108.32	84.786	53.128
$\phi = -0.9$	4.1503	4.6556	5.1383	5.5779	5.9519	6.2353	6.4009	6.4194	6.2590	5.8857	5.3350
$\phi = -0.8$	2.0582	2.2444	2.4242	2.5914	2.7393	2.8607	2.9477	2.9921	2.9849	2.9168	2.7950
$\phi = -0.7$	1.4789	1.5777	1.6743	1.7656	1.8484	1.9193	1.9746	2.0105	2.0230	2.0079	1.9670
$\phi = -0.6$	1.2309	1.2893	1.3475	1.4036	1.4555	1.5012	1.5387	1.5659	1.5804	1.5801	1.5651
$\phi = -0.5$	1.1076	1.1426	1.1784	1.2137	1.2471	1.2773	1.3030	1.3228	1.3354	1.3394	1.3344
$\phi = -0.4$	1.0446	1.0644	1.0855	1.1069	1.1277	1.1470	1.1640	1.1777	1.1873	1.1919	1.1909
$\phi = -0.3$	1.0147	1.0243	1.0352	1.0469	1.0586	1.0699	1.0801	1.0886	1.0950	1.0986	1.0990
$\phi = -0.2$	1.0030	1.0065	1.0108	1.0157	1.0210	1.0263	1.0313	1.0356	1.0390	1.0411	1.0417
$\phi = -0.1$	1.0002	1.0008	1.0016	1.0028	1.0041	1.0055	1.0069	1.0081	1.0092	1.0099	1.0101
$\phi = 0$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\phi = 0.1$	1.0002	1.0000	1.0002	1.0008	1.0018	1.0032	1.0049	1.0066	1.0083	1.0096	1.0101
$\phi = 0.2$	1.0030	1.0008	1.0000	1.0009	1.0036	1.0082	1.0146	1.0226	1.0312	1.0389	1.0418
$\phi = 0.3$	1.0147	1.0070	1.0019	1.0000	1.0021	1.0089	1.0210	1.0390	1.0628	1.0900	1.1023
$\phi = 0.4$	1.0446	1.0271	1.0130	1.0035	1.0000	1.0041	1.0180	1.0444	1.0875	1.1549	1.2675
$\phi = 0.5$	1.1076	1.0748	1.0457	1.0221	1.0061	1.0000	1.0074	1.0334	1.0867	1.1860	1.3560
$\phi = 0.6$	1.2309	1.1742	1.1213	1.0745	1.0363	1.0101	1.0000	1.0128	1.0598	1.1633	1.3406
$\phi = 0.7$	1.4789	1.3811	1.2871	1.2000	1.1232	1.0604	1.0169	1.0000	1.0224	1.1078	1.2727
$\phi = 0.8$	2.0582	1.8713	1.6892	1.5168	1.3590	1.2211	1.1089	1.0307	1.0000	1.0423	1.1809
$\phi = 0.9$	4.1503	3.6406	3.1424	2.6697	2.2346	1.8473	1.5172	1.2543	1.0723	1.0000	1.0812
$\phi = 0.99$	78.416	64.740	51.610	39.522	28.852	19.858	12.678	7.3310	3.7226	1.6582	1.0000

Note:- 1 denotes indeterminate, >1 denotes ARMA(1,1) preferred, and <1 denotes MA(1) preferred

Table 5.9

AR(1) vs. MA(1) comparison: $\text{GFESM}_{b_2} / \text{GFESM}_{c_2}$

	$\theta=0$	$\theta=0.1$	$\theta=0.2$	$\theta=0.3$	$\theta=0.4$	$\theta=0.5$	$\theta=0.6$	$\theta=0.7$	$\theta=0.8$	$\theta=0.9$	$\theta=0.99$
$\phi = -0.99$	0.0128	0.0110	0.0099	0.0095	0.0095	0.0102	0.0115	0.0141	0.0189	0.0291	0.0553
$\phi = -0.9$	0.2409	0.2169	0.2026	0.1966	0.1988	0.2099	0.2320	0.2686	0.3262	0.4173	0.5490
$\phi = -0.8$	0.4859	0.4496	0.4284	0.4215	0.4295	0.4542	0.4993	0.5704	0.6766	0.8326	1.0360
$\phi = -0.7$	0.6762	0.6390	0.6182	0.6149	0.6308	0.6691	0.7349	0.8354	0.9812	1.1879	1.4455
$\phi = -0.6$	0.8124	0.7809	0.7649	0.7675	0.7917	0.8423	0.9255	1.0500	1.2272	1.4735	1.7727
$\phi = -0.5$	0.9029	0.8800	0.8706	0.8794	0.9112	0.9716	1.0682	1.2107	1.4115	1.6872	2.0172
$\phi = -0.4$	0.9573	0.9433	0.9404	0.9545	0.9918	1.0589	1.1644	1.3187	1.5352	1.8305	2.1812
$\phi = -0.3$	0.9855	0.9790	0.9810	0.9985	1.0385	1.1086	1.2180	1.3779	1.6021	1.9076	2.2690
$\phi = -0.2$	0.9970	0.9952	0.9999	1.0182	1.0577	1.1266	1.2347	1.3938	1.6181	1.9245	2.2871
$\phi = -0.1$	0.9998	1.0000	1.0048	1.0209	1.0563	1.1198	1.2215	1.3734	1.5900	1.8882	2.2427
$\phi = 0$	1.0000	1.0002	1.0030	1.0144	1.0424	1.0961	1.1860	1.3244	1.5257	1.8069	2.1444
$\phi = 0.1$	0.9998	1.0000	1.0006	1.0061	1.0241	1.0644	1.1375	1.2558	1.4338	1.6888	2.0006
$\phi = 0.2$	0.9970	0.9994	1.0000	1.0010	1.0091	1.0338	1.0861	1.1783	1.3246	1.5425	1.8202
$\phi = 0.3$	0.9855	0.9939	0.9990	1.0000	1.0014	1.0119	1.0426	1.1051	1.2114	1.3765	1.6075
$\phi = 0.4$	0.9573	0.9754	0.9907	0.9986	1.0000	1.0017	1.0144	1.0491	1.1143	1.2122	1.3008
$\phi = 0.5$	0.9029	0.9334	0.9641	0.9871	0.9981	1.0000	1.0020	1.0160	1.0495	1.0981	1.1269
$\phi = 0.6$	0.8124	0.8558	0.9048	0.9503	0.9828	0.9976	1.0000	1.0023	1.0156	1.0390	1.0519
$\phi = 0.7$	0.6762	0.7289	0.7944	0.8658	0.9312	0.9769	0.9970	1.0000	1.0023	1.0120	1.0185
$\phi = 0.8$	0.4859	0.5388	0.6102	0.6999	0.8012	0.8976	0.9664	0.9960	1.0000	1.0018	1.0048
$\phi = 0.9$	0.2409	0.2773	0.3304	0.4064	0.5126	0.6510	0.8064	0.9348	0.9930	1.0000	1.0005
$\phi = 0.99$	0.0128	0.0156	0.0202	0.0278	0.0411	0.0660	0.1167	0.2300	0.4934	0.9114	1.0000

Note:- 1 denotes indeterminate, >1 denotes MA(1) preferred, and <1 denotes AR(1) preferred

to zero and the theoretical result is then that all three models are white noise and have the same forecasts asymptotically. In each of these three cases, forecast performance will be identical across the models by both criteria, and the ratio results will clearly be 1.

Examination of tables 5.1-5.4 confirms that when the true model is compared with a misspecified alternative, the true model always has the lower MSFE. This is illustrated by the ratio values all exceeding 1 (with the exception of the special cases noted above where the models are the same and the ratio equals 1). The same observation can be made for the GFESM criterion using tables 5.7 and 5.8 (tables 5.1 and 5.3 are the GFESM_1 comparisons as $\text{MSFE}_1 = \text{GFESM}_1$) with the true model ranked first in all but the special cases, illustrating the proof of this result contained in sub-section 5.3.1.

The most interesting inference is found by studying tables 5.5 and 5.9. Table 5.5 contains the results of the comparison between the two misspecified models by the GFESM_1 criterion ($= \text{MSFE}_1$). Table 5.9 contains results of the GFESM_2 comparison of the same models. The expected result occurs that the AR(1) model is preferred for some values of ϕ and θ , and the MA(1) model preferred for others. In general, this is consistent across the forecast horizon. However, there are a few cases where this does not hold. These cases (boxed and highlighted in the tables) represent values of ϕ and θ for which GFESM_1 ranks the AR(1) model as best, but GFESM_2 concludes that the MA(1) model should be ranked first. Tables 5.10 and 5.12 magnify the (ϕ, θ) ranges around which this phenomenon occurs, and a clear

Table 5.10

AR(1) vs. MA(1) comparison: $MSFE_{b1} / MSFE_{c1} = GFESM_{b1} / GFESM_{c1}$ (region magnified)

ϕ	θ	0.5	0.52	0.54	0.56	0.58	0.6	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98
-0.84		0.491	0.492	0.493	0.494	0.496	0.497	0.499	0.501	0.504	0.506	0.509	0.511	0.514	0.518	0.521	0.524	0.528	0.532	0.535	0.539	0.543	0.548	0.552	0.556	0.561
-0.82		0.536	0.537	0.539	0.540	0.542	0.544	0.547	0.549	0.552	0.555	0.559	0.562	0.566	0.569	0.573	0.577	0.582	0.586	0.591	0.595	0.600	0.605	0.610	0.615	0.620
-0.8		0.578	0.580	0.582	0.584	0.586	0.589	0.592	0.595	0.599	0.602	0.606	0.610	0.614	0.619	0.623	0.628	0.633	0.638	0.643	0.649	0.654	0.660	0.666	0.672	0.678
-0.78		0.618	0.620	0.623	0.625	0.628	0.631	0.635	0.638	0.642	0.647	0.651	0.656	0.660	0.665	0.671	0.676	0.682	0.688	0.694	0.700	0.706	0.713	0.719	0.726	0.733
-0.76		0.656	0.658	0.661	0.664	0.668	0.671	0.675	0.679	0.684	0.689	0.694	0.699	0.704	0.710	0.716	0.722	0.729	0.735	0.742	0.749	0.756	0.763	0.771	0.778	0.786
-0.74		0.691	0.694	0.697	0.701	0.705	0.709	0.713	0.718	0.723	0.728	0.734	0.740	0.746	0.753	0.759	0.766	0.773	0.780	0.788	0.796	0.804	0.812	0.820	0.829	0.837
-0.72		0.724	0.728	0.731	0.735	0.740	0.744	0.749	0.755	0.760	0.766	0.772	0.779	0.786	0.793	0.800	0.808	0.816	0.824	0.832	0.840	0.849	0.858	0.867	0.877	0.886
-0.7		0.755	0.759	0.763	0.768	0.772	0.778	0.783	0.789	0.795	0.802	0.809	0.816	0.823	0.831	0.839	0.847	0.856	0.865	0.874	0.883	0.893	0.902	0.912	0.923	0.933
-0.68		0.784	0.789	0.793	0.798	0.803	0.809	0.815	0.822	0.828	0.835	0.843	0.851	0.859	0.867	0.876	0.885	0.894	0.904	0.914	0.924	0.934	0.945	0.955	0.966	0.978
-0.66		0.812	0.816	0.821	0.827	0.832	0.838	0.845	0.852	0.859	0.867	0.875	0.884	0.892	0.901	0.911	0.920	0.930	0.941	0.951	0.962	0.973	0.985	0.996	1.008	1.020
-0.64		0.837	0.842	0.847	0.853	0.859	0.866	0.873	0.881	0.889	0.897	0.905	0.914	0.924	0.934	0.944	0.954	0.965	0.976	0.987	0.999	1.011	1.023	1.035	1.048	1.061
-0.62		0.861	0.866	0.872	0.878	0.885	0.892	0.900	0.908	0.916	0.925	0.934	0.944	0.954	0.964	0.975	0.986	0.997	1.009	1.021	1.033	1.046	1.059	1.073	1.086	1.100
-0.6		0.883	0.888	0.895	0.901	0.909	0.916	0.924	0.933	0.942	0.951	0.961	0.971	0.981	0.992	1.004	1.016	1.028	1.040	1.053	1.066	1.080	1.094	1.108	1.122	1.137
-0.58		0.903	0.909	0.916	0.923	0.931	0.939	0.947	0.956	0.966	0.975	0.986	0.996	1.008	1.019	1.031	1.044	1.056	1.070	1.083	1.097	1.111	1.126	1.141	1.156	1.172
-0.56		0.922	0.929	0.936	0.943	0.951	0.960	0.969	0.978	0.988	0.998	1.009	1.020	1.032	1.044	1.057	1.070	1.083	1.097	1.111	1.126	1.141	1.157	1.172	1.189	1.205
-0.54		0.940	0.947	0.954	0.962	0.970	0.979	0.988	0.998	1.009	1.019	1.031	1.043	1.055	1.068	1.081	1.095	1.109	1.123	1.138	1.153	1.169	1.185	1.202	1.219	1.236
-0.52		0.956	0.963	0.971	0.979	0.988	0.997	1.007	1.017	1.028	1.039	1.051	1.063	1.076	1.089	1.103	1.117	1.132	1.147	1.163	1.179	1.195	1.212	1.230	1.247	1.266
-0.5		0.971	0.978	0.986	0.995	1.004	1.013	1.023	1.034	1.045	1.057	1.070	1.082	1.096	1.110	1.124	1.139	1.154	1.170	1.186	1.203	1.220	1.238	1.256	1.274	1.293
-0.48		0.984	0.992	1.000	1.009	1.018	1.028	1.039	1.050	1.062	1.074	1.087	1.100	1.114	1.128	1.143	1.158	1.174	1.191	1.208	1.225	1.243	1.261	1.280	1.299	1.319
-0.46		0.996	1.004	1.013	1.022	1.032	1.042	1.053	1.064	1.077	1.089	1.102	1.116	1.130	1.145	1.161	1.176	1.193	1.210	1.227	1.245	1.264	1.283	1.303	1.323	1.343

Note:- 1 denotes indeterminate, >1 denotes MA(1) preferred, and <1 denotes AR(1) preferred

Table 5.11

AR(1) vs. MA(1) comparison: $MSFE_{b2} / MSFE_{c2}$ (region magnified)

ϕ	θ	0.5	0.52	0.54	0.56	0.58	0.6	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98
-0.84		0.419	0.418	0.417	0.416	0.415	0.414	0.414	0.413	0.412	0.412	0.411	0.411	0.410	0.410	0.410	0.409	0.409	0.409	0.409	0.409	0.409	0.408	0.408	0.408	0.408
-0.82		0.462	0.460	0.459	0.458	0.457	0.456	0.456	0.455	0.454	0.454	0.453	0.453	0.452	0.452	0.452	0.451	0.451	0.451	0.451	0.450	0.450	0.450	0.450	0.450	0.450
-0.8		0.502	0.500	0.499	0.498	0.497	0.496	0.496	0.495	0.494	0.494	0.493	0.493	0.492	0.492	0.491	0.491	0.491	0.491	0.490	0.490	0.490	0.490	0.490	0.490	0.490
-0.78		0.540	0.539	0.537	0.536	0.535	0.535	0.534	0.533	0.532	0.532	0.531	0.531	0.530	0.530	0.529	0.529	0.529	0.529	0.528	0.528	0.528	0.528	0.528	0.528	0.528
-0.76		0.576	0.575	0.574	0.573	0.572	0.571	0.570	0.569	0.569	0.568	0.567	0.567	0.566	0.566	0.566	0.565	0.565	0.565	0.565	0.564	0.564	0.564	0.564	0.564	0.564
-0.74		0.610	0.609	0.608	0.607	0.606	0.605	0.604	0.604	0.603	0.602	0.602	0.601	0.601	0.600	0.600	0.600	0.599	0.599	0.599	0.599	0.599	0.599	0.598	0.598	0.598
-0.72		0.643	0.642	0.641	0.640	0.639	0.638	0.637	0.636	0.636	0.635	0.634	0.634	0.634	0.633	0.633	0.632	0.632	0.632	0.632	0.632	0.631	0.631	0.631	0.631	0.631
-0.7		0.674	0.672	0.671	0.670	0.670	0.669	0.668	0.667	0.667	0.666	0.665	0.665	0.665	0.664	0.664	0.664	0.663	0.663	0.663	0.663	0.663	0.662	0.662	0.662	0.662
-0.68		0.703	0.702	0.701	0.700	0.699	0.698	0.697	0.697	0.696	0.695	0.695	0.694	0.694	0.694	0.693	0.693	0.693	0.693	0.692	0.692	0.692	0.692	0.692	0.692	0.692
-0.66		0.730	0.729	0.728	0.727	0.726	0.726	0.725	0.724	0.724	0.723	0.723	0.722	0.722	0.722	0.721	0.721	0.721	0.721	0.720	0.720	0.720	0.720	0.720	0.720	0.720
-0.64		0.756	0.755	0.754	0.753	0.752	0.752	0.751	0.750	0.750	0.749	0.749	0.749	0.748	0.748	0.748	0.747	0.747	0.747	0.747	0.747	0.747	0.747	0.746	0.746	0.746
-0.62		0.780	0.779	0.778	0.778	0.777	0.776	0.776	0.775	0.775	0.774	0.774	0.774	0.773	0.773	0.773	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.772	0.772
-0.6		0.803	0.802	0.801	0.801	0.800	0.799	0.799	0.798	0.798	0.798	0.797	0.797	0.797	0.796	0.796	0.796	0.796	0.796	0.796	0.796	0.795	0.795	0.795	0.795	0.795
-0.58		0.824	0.823	0.823	0.822	0.822	0.821	0.821	0.820	0.820	0.820	0.819	0.819	0.819	0.819	0.818	0.818	0.818	0.818	0.818	0.818	0.818	0.818	0.818	0.818	0.818
-0.56		0.844	0.843	0.843	0.842	0.842	0.842	0.841	0.841	0.841	0.840	0.840	0.840	0.840	0.839	0.839	0.839	0.839	0.839	0.839	0.839	0.839	0.839	0.839	0.839	0.839
-0.54		0.863	0.862	0.862	0.861	0.861	0.861	0.860	0.860	0.860	0.860	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859	0.859
-0.52		0.880	0.880	0.879	0.879	0.879	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.878	0.877	0.877	0.877	0.877	0.877	0.877	0.877	0.877	0.877	0.877	0.877	0.877
-0.5		0.896	0.896	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895	0.895
-0.48		0.911	0.911	0.911	0.911	0.910	0.910	0.910	0.910	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.911
-0.46		0.924	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.926	0.926	0.926	0.926	0.926	0.926	0.926	0.926	0.926	0.926	0.926	0.926	0.926

Note:- 1 denotes indeterminate, >1 denotes MA(1) preferred, and <1 denotes AR(1) preferred

Table 5.12

AR(1) vs. MA(1) comparison: GFESM_{b2} / GFESM_{c2} (region magnified)

ϕ	θ	0.5	0.52	0.54	0.56	0.58	0.6	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98
-0.84		0.359	0.364	0.371	0.378	0.386	0.395	0.404	0.415	0.427	0.439	0.453	0.468	0.484	0.501	0.520	0.541	0.563	0.587	0.613	0.642	0.673	0.706	0.743	0.783	0.827
-0.82		0.407	0.414	0.421	0.429	0.438	0.448	0.458	0.470	0.483	0.497	0.512	0.529	0.547	0.566	0.587	0.610	0.634	0.661	0.689	0.720	0.754	0.790	0.830	0.873	0.920
-0.8		0.454	0.462	0.470	0.479	0.488	0.499	0.511	0.524	0.538	0.554	0.570	0.589	0.608	0.629	0.652	0.677	0.703	0.732	0.763	0.796	0.833	0.872	0.914	0.960	1.010
-0.78		0.500	0.508	0.517	0.527	0.538	0.550	0.563	0.577	0.592	0.609	0.627	0.647	0.668	0.691	0.715	0.742	0.771	0.801	0.835	0.870	0.909	0.951	0.996	1.044	1.096
-0.76		0.545	0.553	0.563	0.574	0.586	0.598	0.613	0.628	0.644	0.662	0.682	0.703	0.726	0.750	0.777	0.805	0.836	0.868	0.904	0.942	0.983	1.027	1.074	1.125	1.180
-0.74		0.588	0.597	0.608	0.619	0.632	0.646	0.661	0.677	0.695	0.714	0.735	0.758	0.782	0.808	0.836	0.866	0.898	0.933	0.971	1.011	1.054	1.100	1.150	1.203	1.261
-0.72		0.629	0.639	0.651	0.663	0.676	0.691	0.707	0.725	0.744	0.764	0.786	0.810	0.836	0.863	0.893	0.925	0.959	0.996	1.035	1.077	1.122	1.171	1.223	1.279	1.338
-0.7		0.669	0.680	0.692	0.705	0.719	0.735	0.752	0.770	0.790	0.812	0.835	0.861	0.888	0.917	0.948	0.981	1.017	1.056	1.097	1.141	1.188	1.238	1.293	1.351	1.413
-0.68		0.707	0.719	0.731	0.745	0.760	0.777	0.795	0.814	0.835	0.858	0.883	0.909	0.937	0.968	1.000	1.035	1.073	1.113	1.156	1.202	1.251	1.303	1.359	1.419	1.484
-0.66		0.744	0.756	0.769	0.784	0.800	0.817	0.836	0.856	0.878	0.902	0.928	0.955	0.985	1.017	1.051	1.087	1.126	1.168	1.212	1.260	1.311	1.365	1.423	1.485	1.552
-0.64		0.778	0.791	0.805	0.820	0.837	0.855	0.875	0.896	0.919	0.944	0.971	0.999	1.030	1.063	1.098	1.136	1.177	1.220	1.266	1.315	1.368	1.424	1.484	1.548	1.616
-0.62		0.811	0.825	0.839	0.855	0.872	0.891	0.912	0.934	0.958	0.984	1.011	1.041	1.073	1.107	1.144	1.183	1.225	1.269	1.317	1.368	1.422	1.480	1.542	1.607	1.677
-0.6		0.842	0.856	0.871	0.888	0.906	0.926	0.947	0.970	0.995	1.021	1.050	1.081	1.114	1.149	1.187	1.227	1.270	1.316	1.366	1.418	1.474	1.533	1.596	1.664	1.735
-0.58		0.872	0.886	0.902	0.919	0.938	0.958	0.980	1.004	1.029	1.057	1.086	1.118	1.152	1.189	1.228	1.269	1.313	1.361	1.411	1.465	1.522	1.583	1.648	1.717	1.790
-0.56		0.899	0.914	0.930	0.948	0.967	0.988	1.011	1.036	1.062	1.090	1.121	1.153	1.188	1.226	1.266	1.308	1.354	1.402	1.454	1.509	1.568	1.630	1.696	1.766	1.841
-0.54		0.925	0.940	0.957	0.976	0.995	1.017	1.040	1.065	1.092	1.122	1.153	1.186	1.222	1.261	1.302	1.345	1.392	1.441	1.494	1.551	1.610	1.674	1.741	1.813	1.889
-0.52		0.949	0.965	0.982	1.001	1.021	1.044	1.067	1.093	1.121	1.151	1.183	1.217	1.254	1.293	1.335	1.380	1.427	1.478	1.532	1.589	1.650	1.715	1.784	1.857	1.934
-0.5		0.972	0.988	1.005	1.025	1.046	1.068	1.093	1.119	1.147	1.178	1.211	1.246	1.283	1.323	1.366	1.412	1.460	1.512	1.567	1.625	1.687	1.753	1.823	1.897	1.976
-0.48		0.992	1.009	1.027	1.047	1.068	1.091	1.116	1.143	1.172	1.203	1.236	1.272	1.310	1.351	1.395	1.441	1.490	1.543	1.599	1.658	1.721	1.788	1.859	1.935	2.015
-0.46		1.011	1.028	1.047	1.067	1.089	1.112	1.138	1.165	1.194	1.226	1.260	1.296	1.335	1.377	1.421	1.468	1.518	1.572	1.628	1.689	1.753	1.821	1.893	1.969	2.050

Note:- 1 denotes indeterminate, >1 denotes MA(1) preferred, and <1 denotes AR(1) preferred

zone is visible where this switching between model preferences by the GFESM criterion occurs. The zone is where ϕ and θ are of opposite sign and, approximately, $0.46|\theta| + 0.23 < |\phi| < 0.71|\theta| + 0.11$. The occurrence of this GFESM ranking reversal between misspecified models when different maximum forecast horizons are employed adds weight to the criticism of GFESM. The argument against the Clements & Hendry emphasis on a mainly 1-step-ahead focus is reinforced (given that the true model is normally unknown in practice and misspecification is to be expected), with GFESM failing to give an unambiguous invariant ranking across all forecast horizons.

Furthermore, examination of these findings alongside table 5.6 - the results of the AR(1) vs. MA(1) comparison using the 2-steps-ahead MSFE criterion - shows that in each of the cases where GFESM has a rank reversal, the MSFE criterion does not switch between horizons. The same is true in table 5.11 where the region is magnified consistent with tables 5.10 and 5.12. This peculiar behaviour can then be summarised in the following way. For a range of (ϕ, θ) values, MSFE concludes that the AR(1) model is preferable in terms of forecasting performance to the MA(1) model over both 1-step- and 2-steps-ahead forecast horizons. In contrast, whilst GFESM_1 obviously gives this inference (since it is equal to MSFE_1), GFESM_2 (which takes both the 1-step- and 2-steps-ahead forecasts into account) concludes that the MA(1) model should be preferred to the AR(1) model. This unusual characteristic of GFESM would be expected to be even more widespread in a more complex example - this study purely considers the simplest example of two non-trivial non-nested misspecified models, and only examines 1-step- and 2-steps-ahead

forecasts.

The GFESM criterion for ranking competing forecasts was proposed as an invariant summary measure for all isomorphic representations of a system and to give a unique unambiguous ranking of the models of interest for a given maximum forecast horizon, h (incorporating information from all 1-,2-,..., h -steps-ahead forecasts). These are valuable properties *per se*, but the criterion has a number of disadvantages. Many of these are expounded by the discussants of the Clements & Hendry paper, the justification for using the criterion in an applied time series context is questioned in sub-section 5.3.1, and criticism of the GFESM measure is strengthened by the above analysis. There are a number of cases in the simple example studied where the GFESM ranking changes according to the maximum step-ahead forecast employed. Moreover, in these cases the undesirable result is obtained that whilst the MSFE (which corresponds directly to the notion of economic loss) ranks one model consistently above the other 1-step-ahead and 2-steps-ahead, the GFESM criterion associated with both forecast horizons (GFESM_2) prefers the MSFE-dominated, or ‘inferior’, model absolutely. It is interesting to note, however, that these problems occur predominantly when ϕ and θ are relatively large in absolute value and of opposite sign. In such cases, the misspecifications are particularly severe, and given sufficient data an analyst should be able to detect the model misspecifications, forming something of a caveat to the conclusions of this analysis.

Altogether, the GFESM criterion has invariance as its primary advantage, but the question remains of whether invariance is necessary, and more particularly, whether

the costs of employing a criterion such as GFESM, with its undesirable features and somewhat tenuous justification, outweigh the first-best desire for the property of invariance.

Section 5.3 Appendix

True Model A

$$\begin{aligned}\text{MSFE}_{a1} &= \text{GFESM}_{a1} = E[e_T^2(1)] \\ &= E[\varepsilon_{T+1}^2] \\ &= \sigma^2\end{aligned}$$

$$\begin{aligned}\text{MSFE}_{a2} &= E[e_T^2(2)] \\ &= E[\varepsilon_{T+2}^2 + (\phi - \theta)^2 \varepsilon_{T+1}^2 + 2(\phi - \theta)\varepsilon_{T+1}\varepsilon_{T+2}] \\ &= \sigma^2[1 + (\phi - \theta)^2]\end{aligned}$$

$$\begin{aligned}\text{GFESM}_{a2} &= \begin{vmatrix} E[e_T^2(1)] & E[e_T(1)e_T(2)] \\ E[e_T(1)e_T(2)] & E[e_T^2(2)] \end{vmatrix} \\ &= \begin{vmatrix} \sigma^2 & \sigma^2(\phi - \theta) \\ \sigma^2(\phi - \theta) & \sigma^2[1 + (\phi - \theta)^2] \end{vmatrix} \\ &= \sigma^4\end{aligned}$$

Model B

$$\text{MSFE}_{b1} = \text{GFESM}_{b1} = E[e_T^2(1)]$$

$$\xrightarrow{p} E\{[(\phi - p_2^{-1} p_1 p_3)y_T + \varepsilon_{T+1} - \theta\varepsilon_T]^2\}$$

$$\text{where } p_1 = \phi - \theta$$

$$p_2 = 1 + \theta^2 - 2\phi\theta$$

$$p_3 = 1 - \phi\theta$$

$$\begin{aligned}
&= (\phi - p_2^{-1} p_1 p_3)^2 E(y_T^2) + E(\varepsilon_{T+1}^2) + \theta^2 E(\varepsilon_T^2) \\
&\quad - 2\theta(\phi - p_2^{-1} p_1 p_3)E(y_T \varepsilon_T) \\
&= \sigma^2(1 + p_2^{-1} \theta^2 p_1^2)
\end{aligned}$$

$$\text{MSFE}_{b2} = E[e_T^2(2)]$$

$$\begin{aligned}
&\xrightarrow{p} E\{[(\phi^2 - (p_2^{-1} p_1 p_3)^2)y_T + \varepsilon_{T+2} + p_1 \varepsilon_{T+1} - \phi \theta \varepsilon_T]^2\} \\
&= [\phi^2 - (p_2^{-1} p_1 p_3)^2]^2 E(y_T^2) + E(\varepsilon_{T+2}^2) + p_1^2 E(\varepsilon_{T+1}^2) \\
&\quad + \phi^2 \theta^2 E(\varepsilon_T^2) - 2\phi \theta (\phi^2 - (p_2^{-1} p_1 p_3)^2) E(y_T \varepsilon_T) \\
&= \sigma^2[(\phi^2 - p_2^{-2} p_1^2 p_3^2)^2 p_4^{-1} p_2 + 1 + p_1^2 + \phi^2 \theta^2 \\
&\quad - 2\phi \theta (\phi^2 - p_2^{-2} p_1^2 p_3^2)]
\end{aligned}$$

$$\text{where } p_4 = 1 - \phi^2$$

$$\text{GFESM}_{b2} = \begin{vmatrix} E[e_T^2(1)] & E[e_T(1)e_T(2)] \\ E[e_T(1)e_T(2)] & E[e_T^2(2)] \end{vmatrix}$$

$$\begin{aligned}
\text{now } E[e_T(1)e_T(2)] &\xrightarrow{p} E\{[(\phi - p_2^{-1} p_1 p_3)y_T + \varepsilon_{T+1} - \theta \varepsilon_T] \\
&\quad [(\phi^2 - (p_2^{-1} p_1 p_3)^2)y_T + \varepsilon_{T+2} + p_1 \varepsilon_{T+1} - \phi \theta \varepsilon_T]\} \\
&= (\phi - p_2^{-1} p_1 p_3)(\phi^2 - (p_2^{-1} p_1 p_3)^2)E(y_T^2) \\
&\quad - [\phi \theta (\phi - p_2^{-1} p_1 p_3) + \theta(\phi^2 - (p_2^{-1} p_1 p_3)^2)] \\
&\quad \times E(y_T \varepsilon_T) + p_1 E(\varepsilon_{T+1}^2) + \phi \theta^2 E(\varepsilon_T^2) \\
&= \sigma^2(p_1 + p_2^{-1} \phi \theta^2 p_1^2)
\end{aligned}$$

$$\text{so } \text{GFESM}_{b2} = E[e_T^2(1)]E[e_T^2(2)] - E[e_T(1)e_T(2)]^2$$

$$\begin{aligned}
&\xrightarrow{p} \sigma^4\{(1 + p_2^{-1} \theta^2 p_1^2)[(\phi^2 - p_2^{-2} p_1^2 p_3^2)^2 p_4^{-1} p_2 + 1 + p_1^2 + \phi^2 \theta^2 \\
&\quad - 2\phi \theta (\phi^2 - p_2^{-2} p_1^2 p_3^2)] - (p_1 + p_2^{-1} \phi \theta^2 p_1^2)^2\}
\end{aligned}$$

Model C

Estimate of β :

$$\text{let: } \gamma_0^* = V[(1 - \beta L)^{-1}(1 - \phi L)^{-1}(1 - \theta L)\varepsilon_t]$$

$$\text{now let: } z_t = (1 - \beta L)^{-1}(1 - \phi L)^{-1}(1 - \theta L)\varepsilon_t$$

$$(1 - \beta L)(1 - \phi L)z_t = (1 - \theta L)\varepsilon_t$$

$$z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2} = \varepsilon_t - \theta \varepsilon_{t-1} \quad (i)$$

$$\text{where } \phi_1 = \beta + \phi$$

$$\phi_2 = -\beta\phi$$

multiply each term in (i) by ε_t and take expectations:

$$E(z_t \varepsilon_t) = \sigma^2 \quad (ii)$$

multiply each term in (i) by ε_{t-1} and take expectations:

$$E(z_t \varepsilon_{t-1}) = \sigma^2(\phi_1 - \theta) \quad (iii)$$

multiply each term in (i) by z_t, z_{t-1}, z_{t-2} in turn:

$$z_t^2 - \phi_1 z_t z_{t-1} - \phi_2 z_t z_{t-2} = z_t \varepsilon_t - \theta z_t \varepsilon_{t-1} \quad (iv)$$

$$z_t z_{t-1} - \phi_1 z_{t-1}^2 - \phi_2 z_{t-1} z_{t-2} = z_{t-1} \varepsilon_t - \theta z_{t-1} \varepsilon_{t-1} \quad (v)$$

$$z_t z_{t-2} - \phi_1 z_{t-1} z_{t-2} - \phi_2 z_{t-2}^2 = z_{t-2} \varepsilon_t - \theta z_{t-2} \varepsilon_{t-1} \quad (vi)$$

take expectations of (iv), (v), (vi) and use information in (ii), (iii):

$$\gamma_0^* - \phi_1 \gamma_1^* - \phi_2 \gamma_2^* = \sigma^2 - \sigma^2 \theta(\phi_1 - \theta) \quad (vii)$$

$$\gamma_1^* - \phi_1 \gamma_0^* - \phi_2 \gamma_1^* = -\sigma^2 \theta \quad (viii)$$

$$\gamma_2^* - \phi_1 \gamma_1^* - \phi_2 \gamma_0^* = 0 \quad (ix)$$

where $\gamma_0^* = V(z_t)$; $\gamma_k^* = C(z_t z_{t-k})$

solve (vii), (viii), (ix) for γ_0^* :

$$(ix) \text{ gives } \gamma_2^* = \phi_1 \gamma_1^* + \phi_2 \gamma_0^*$$

$$(viii) \text{ gives } \gamma_1^* = (1 - \phi_2)^{-1}(\phi_1 \gamma_0^* - \sigma^2 \theta)$$

substitution into (vii) gives:

$$\begin{aligned} \gamma_0^* &= \phi_1 (1 - \phi_2)^{-1} (\phi_1 \gamma_0^* - \sigma^2 \theta) + \phi_2 [\phi_1 (1 - \phi_2)^{-1} (\phi_1 \gamma_0^* - \sigma^2 \theta) \\ &\quad + \phi_2 \gamma_0^*] + \sigma^2 - \sigma^2 \theta (\phi_1 - \theta) \\ &= \sigma^2 \left[\frac{(1 - \phi_2)[1 - \theta(\phi_1 - \theta)] - \theta \phi_1 (1 + \phi_2)}{(1 - \phi_2)(1 - \phi_2^2) - \phi_1^2 (1 + \phi_2)} \right] \\ &= \sigma^2 \left[\frac{1 + \theta^2 - 2\phi\theta + \beta(\phi - 2\theta + \phi\theta^2)}{(1 - \phi^2)(1 - \beta\phi - \beta^2 + \beta^3\phi)} \right] \end{aligned}$$

now minimise γ_0^* with respect to β :

$$\begin{aligned} \min_{\beta} [\gamma_0^*] &\equiv \min_{\beta} \left[\frac{1 + \theta^2 - 2\phi\theta + \beta(\phi - 2\theta + \phi\theta^2)}{(1 - \phi^2)(1 - \beta\phi - \beta^2 + \beta^3\phi)} \right] \\ &\equiv \min_{\beta} [W] \end{aligned}$$

$$\frac{dW}{d\beta} = b^{-2} \left(b \frac{da}{d\beta} - a \frac{db}{d\beta} \right)$$

$$\text{where } a = 1 + \theta^2 - 2\phi\theta + \beta(\phi - 2\theta + \phi\theta^2)$$

$$b = (1 - \phi^2)(1 - \beta\phi - \beta^2 + \beta^3\phi)$$

$$\begin{aligned} &= [(1 - \phi^2)(1 - \beta\phi - \beta^2 + \beta^3\phi)]^{-2} \\ &\quad [(1 - \phi^2)(1 - \beta\phi - \beta^2 + \beta^3\phi)(\phi - 2\theta + \phi\theta^2) \\ &\quad - (1 + \theta^2 - 2\phi\theta + \beta(\phi - 2\theta + \phi\theta^2))(1 - \phi^2)(-\phi - 2\beta + 3\beta^2\phi)] \end{aligned}$$

set $\frac{dW}{d\beta} = 0$ and solve following for β :

$$(1 - \beta\phi - \beta^2 + \beta^3\phi)(\phi - 2\theta + \phi\theta^2) + (1 + \theta^2 - 2\phi\theta + \beta(\phi - 2\theta + \phi\theta^2))(\phi + 2\beta - 3\beta^2\phi) = 0$$

$$\text{MSFE}_{c1} = \text{GFESM}_{c1} = E[e_T^2(1)]$$

$$\text{now } e_T(1) = y_{T+1} + \hat{\beta}(1 - \hat{\beta}L)^{-1}y_T$$

$$\text{i.e. } (1 - \hat{\beta}L)e_T(1) = (1 - \hat{\beta}L)y_{T+1} + \hat{\beta}y_T$$

$$\begin{aligned} e_T(1) &= (1 - \hat{\beta}L)y_{T+1} \\ &= (1 - \hat{\beta}L)^{-1}(1 - \phi L)^{-1}(1 - \theta L)\varepsilon_{T+1} \end{aligned}$$

$$\text{so } \text{MSFE}_{c1} = \text{GFESM}_{c1} = V[e_T(1)]$$

$$\begin{aligned} &= V[(1 - \hat{\beta}L)^{-1}(1 - \phi L)^{-1}(1 - \theta L)\varepsilon_{T+1}] \\ &= \sigma^2[p_4(1 - \beta\phi - \beta^2 + \beta^3\phi)]^{-1} \\ &\quad \times [p_2 + \beta(\phi - 2\theta + \phi\theta^2)] \end{aligned}$$

$$\text{MSFE}_{c2} = E[e_T^2(2)]$$

$$\begin{aligned} &= E\{[\phi^2 y_T + \varepsilon_{T+2} + (\phi - \theta)\varepsilon_{T+1} - \phi\theta\varepsilon_T]^2\} \\ &= \phi^4 E(y_T^2) + E(\varepsilon_{T+2}^2) + (\phi - \theta)^2 E(\varepsilon_{T+1}^2) \\ &\quad + \phi^2 \theta^2 E(\varepsilon_T^2) - 2\phi^3 \theta E(y_T \varepsilon_T) \\ &= \sigma^2 p_4^{-1} p_2 \end{aligned}$$

$$\text{GFESM}_{c2} = \begin{vmatrix} E[e_T^2(1)] & E[e_T(1)e_T(2)] \\ E[e_T(1)e_T(2)] & E[e_T^2(2)] \end{vmatrix}$$

$$\text{now } (1 - \hat{\beta}L)e_T(1) = y_{T+1}$$

$$\text{i.e. } e_{T+1} = \hat{\beta}e_T + y_{T+1}$$

$$= \sum_{i=0}^{\infty} \hat{\beta}^i y_{T+1-i} \quad \text{by recursion, provided } |\hat{\beta}| < 1$$

$$\text{then } e_T(1)e_T(2) = \sum_{i=0}^{\infty} \hat{\beta}^i y_{T+1-i} y_{T+2}$$

$$E[e_T(1)e_T(2)] = \sum_{i=0}^{\infty} \hat{\beta}^i \gamma_{i+1} \quad \text{where } \gamma_k = C(y_t y_{t-k})$$

$$= \gamma_1(1 + \hat{\beta}\phi + \hat{\beta}^2\phi^2 + \dots) \quad \text{as } \gamma_k = \phi\gamma_{k-1}; k = 1, 2, \dots$$

$$= \gamma_1(1 - \hat{\beta}\phi)^{-1} \quad \text{provided } |\hat{\beta}\phi| < 1$$

$$= \sigma^2 p_4^{-1} (1 - \hat{\beta}\phi)^{-1} p_1 p_3$$

$$\text{so } \text{GFESM}_{c_2} = E[e_T^2(1)]E[e_T^2(2)] - E[e_T(1)e_T(2)]^2$$

$$= \sigma^4 \{ p_4^{-2} (1 - \beta\phi - \beta^2 + \beta^3\phi)^{-1} p_2 [p_2 + \beta(\phi - 2\theta + \phi\theta^2)] \\ - [p_4^{-1} (1 - \beta\phi)^{-1} p_1 p_3]^2 \}$$

5.4 An Invariant Non-Reversing Criterion

The desire for a forecast ranking criterion which is invariant for all isomorphic representations of a system, combined with the highly undesirable behaviour of GFESM when ranking competing misspecified models motivates analysis concerning alternative methods of obtaining an unambiguous ranking of competing forecasts where all 1-step- through h -steps-ahead forecasts are incorporated in a summary criterion.

A stricter ranking criterion which has many of the GFESM advantages without the unwelcome reversals revealed in the previous section can be derived from the MSFEM measure proposed by Granger & Newbold (1986). This new criterion is mentioned by Clements & Hendry (1993), and ranks forecasts on the basis of dominance.

The MSFEM criterion selects the forecast which has the lowest value of $d'V_h d$ for all non-zero vectors d , where V_h is as defined in (5.2). This criterion can be generalised to apply to the stacked forecast error second moment matrix ϕ_h , as defined in (5.8), i.e. selecting the forecast associated with the lowest $d'\phi_h d$ for all non-zero d . Note that $V_h = \text{MSFE}_h$ in a univariate setting and thus purely considers the h -steps-ahead forecast, whereas ϕ_h includes information on all step-ahead forecasts up to and including h -steps-ahead. Denoting this new criterion GMSFEM (generalised MSFEM), analysis proceeds by considering the choice between two competing forecasts. Forecast 1 is chosen in preference to forecast 2 by the

GMSFEM criterion if it is dominant in terms of forecast performance. More formally, forecast 1 is preferred when:

$$d'\phi_{1h}d \leq d'\phi_{2h}d \quad \forall d \neq 0 \quad (5.34)$$

where the inequality is strict for at least one vector d . This can be rearranged to give:

$$\begin{aligned} d'\phi_{1h}d - d'\phi_{2h}d &\leq 0 \quad \forall d \neq 0 \\ d'(\phi_{1h} - \phi_{2h})d &\leq 0 \quad \forall d \neq 0 \end{aligned} \quad (5.35)$$

again where for at least one vector d , the inequality is strict. Given that the left hand side of (5.35) is a quadratic form, it follows that, by the GMSFEM criterion, forecast 1 dominates and is preferred to forecast 2 if $(\phi_{1h} - \phi_{2h})$ is negative semi-definite. From here it is clear that all the criterion requires is that the eigenvalues of $(\phi_{1h} - \phi_{2h})$ be less than or equal to zero, with at least one eigenvalue being non-zero. Three possibilities can then be conceived:

1. ALL EIGENVALUES OF $(\phi_{1h} - \phi_{2h}) \leq 0$, AT LEAST ONE EIGENVALUE $\neq 0$:
implies forecast 1 dominant and preferred to forecast 2
2. ALL EIGENVALUES OF $(\phi_{1h} - \phi_{2h}) \geq 0$, AT LEAST ONE EIGENVALUE $\neq 0$:
implies forecast 2 dominant and preferred to forecast 1
3. SOME EIGENVALUES OF $(\phi_{1h} - \phi_{2h}) \leq 0$, SOME ≥ 0 ; OR BOTH $= 0$:
implies neither forecast dominant and conclusion indeterminate
i.e. the GMSFEM criterion does not return a ranking

The GMSFEM criterion therefore allows simple evaluation of the two competing forecasts, resulting in one of the above inferences.

The major advantage of the GMSFEM criterion over the GFESM measure is that it cannot reverse when different maximum forecast horizons, h , are considered. Forecast preference according to GMSFEM for a given maximum forecast horizon, h , involves dominance for all d . Now use of the criterion with a maximum horizon $k < h$ must return the same ranking since all the linear combinations of forecast errors included in GMSFEM_k are a subset of those for GMSFEM_h as dominance for all vectors d must be satisfied. It is not therefore possible theoretically for GMSFEM to find forecast 1 dominant at one horizon and forecast 2 dominant at a different horizon. This abolition of possible switching in forecast rankings instantly overcomes the GFESM criticisms in this area discussed in section 5.3, and appears very attractive.

It is also useful to note that GMSFEM dominance is sufficient, but not necessary, for GFESM preference of a forecast. This follows because for a positive definite matrix X , and a positive semi-definite matrix Y , $|X + Y| \geq |X|$. Letting $X = \phi_{1h}$ (positive definite) and $Y = \phi_{2h} - \phi_{1h}$ (positive semi-definite if GMSFEM dominance of forecast 1 holds), the rule yields $|\phi_{2h}| \geq |\phi_{1h}|$ which implies GFESM preference of forecast 1.

Another useful relation is that GMSFEM_h dominance is sufficient, but not necessary, for MSFE dominance at any horizon $k \leq h$, since the MSFE measure is one of the linear combinations considered by GMSFEM ($\text{MSFE}_k = \text{GMSFEM}_h$ when d has k 'th element one and all other elements zero).

The GMSFEM criterion can now be examined by application to the three model

study of sub-section 5.3.2. The matrices ϕ_h for each model and forecast horizon are then:

$$\text{true model A: } \phi_{a1} = \sigma^2 \quad (5.36)$$

$$\phi_{a2} = \sigma^2 \begin{bmatrix} 1 & (\phi - \theta) \\ (\phi - \theta) & 1 + (\phi - \theta)^2 \end{bmatrix} \quad (5.37)$$

$$\text{model B: } \phi_{b1} \xrightarrow{p} \sigma^2 [1 + p_2^{-1} \theta^2 p_1^2] \quad (5.38)$$

$$\phi_{b2} \xrightarrow{p} \sigma^2 \begin{bmatrix} 1 + p_2^{-1} \theta^2 p_1^2 & p_1 + p_2^{-1} \phi \theta^2 p_1^2 \\ p_1 + p_2^{-1} \phi \theta^2 p_1^2 & p_5 \end{bmatrix} \quad (5.39)$$

$$\text{where } p_5 = (\phi^2 - p_2^{-2} p_1^2 p_3^2)^2 p_4^{-1} p_2 + 1 + p_1^2 \\ + \phi^2 \theta^2 - 2\phi\theta(\phi^2 - p_2^{-2} p_1^2 p_3^2)]$$

$$\text{model C: } \phi_{c1} = \sigma^2 \left[\frac{p_2 + \hat{\beta}(\phi - 2\theta + \phi\theta^2)}{p_4(1 - \hat{\beta}\phi - \hat{\beta}^2 + \hat{\beta}^3\phi)} \right] \quad (5.40)$$

$$\phi_{c2} = \sigma^2 \begin{bmatrix} \frac{p_2 + \hat{\beta}(\phi - 2\theta + \phi\theta^2)}{p_4(1 - \hat{\beta}\phi - \hat{\beta}^2 + \hat{\beta}^3\phi)} & \frac{p_1 p_3}{p_4(1 - \hat{\beta}\phi)} \\ \frac{p_1 p_3}{p_4(1 - \hat{\beta}\phi)} & \frac{p_2}{p_4} \end{bmatrix} \quad (5.41)$$

where p_i ($i = 1, \dots, 4$) are as defined in section 5.3. The six comparisons and criteria now become:

$$(i) \text{ AR}(1) \text{ vs. ARMA}(1,1): \text{GMSFEM}_1 = d'(\phi_{a1} - \phi_{b1})d$$

$$(ii) \text{ AR}(1) \text{ vs. ARMA}(1,1): \text{GMSFEM}_2 = d'(\phi_{a2} - \phi_{b2})d$$

$$(iii) \text{ MA}(1) \text{ vs. ARMA}(1,1): \text{GMSFEM}_1 = d'(\phi_{a1} - \phi_{c1})d$$

$$(iv) \text{ MA}(1) \text{ vs. ARMA}(1,1): \text{GMSFEM}_2 = d'(\phi_{a2} - \phi_{c2})d$$

$$(v) \text{ AR}(1) \text{ vs. MA}(1): \quad \text{GMSFEM}_1 = d'(\phi_{c1} - \phi_{b1})d$$

$$(vi) \text{ AR}(1) \text{ vs. MA}(1): \quad \text{GMSFEM}_2 = d'(\phi_{c2} - \phi_{b2})d$$

In each of these six cases, the eigenvalues of the respective matrix are studied and one of the three conclusions described above is drawn. It is clear that when comparing a misspecified model with the true model, the true model will be dominant, so results are not given for the cases (i)-(iv). Tables 5.13 and 5.14 contain results of the comparisons of the misspecified models noted under (v) and (vi) above, with a zero value indicating indeterminacy, and the values 1 and 2 denoting model preference accordingly. It is important to note that this analysis concerns asymptotic results, and account is not taken of errors which will arise from estimation of the population quantities comprising GMSFEM.

For 1-step-ahead comparisons (table 5.13), the GMSFEM is merely comparing the models' MSFE₁; consequently the conclusions are identical to those gleaned from table 5.5 (i.e. model AR(1) preferred in some cases, model MA(1) preferred in others). The new contribution of GMSFEM comes in when 2-steps-ahead forecasts are included (table 5.14). The criterion cannot, by definition, reverse, and comparison of GMSFEM₁ with GMSFEM₂ shows that the rankings do not change when the maximum forecast horizon is increased. The peculiar cases associated with the GFESM criterion (again highlighted and boxed in the table) are now found to give an indeterminate conclusion rather than a switch in model preferences. Tables 5.15 and 5.16 magnify the areas where the GFESM measure experienced

Table 5.13

AR(1) vs. MA(1) comparison: GMSFEM₁

	$\theta = 0$	$\theta = 0.1$	$\theta = 0.2$	$\theta = 0.3$	$\theta = 0.4$	$\theta = 0.5$	$\theta = 0.6$	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.9$	$\theta = 0.99$
$\phi = -0.99$	2	2	2	2	2	2	2	2	2	2	2
$\phi = -0.9$	2	2	2	2	2	2	2	2	2	2	2
$\phi = -0.8$	2	2	2	2	2	2	2	2	2	2	2
$\phi = -0.7$	2	2	2	2	2	2	2	2	2	2	2
$\phi = -0.6$	2	2	2	2	2	2	2	2	1	1	1
$\phi = -0.5$	2	2	2	2	2	2	1	1	1	1	1
$\phi = -0.4$	2	2	2	2	2	1	1	1	1	1	1
$\phi = -0.3$	2	2	2	2	1	1	1	1	1	1	1
$\phi = -0.2$	2	2	2	1	1	1	1	1	1	1	1
$\phi = -0.1$	2	2	1	1	1	1	1	1	1	1	1
$\phi = 0$	0	1	1	1	1	1	1	1	1	1	1
$\phi = 0.1$	2	0	1	1	1	1	1	1	1	1	1
$\phi = 0.2$	2	2	0	1	1	1	1	1	1	1	1
$\phi = 0.3$	2	2	2	0	1	1	1	1	1	1	1
$\phi = 0.4$	2	2	2	2	0	1	1	1	1	1	1
$\phi = 0.5$	2	2	2	2	2	0	1	1	1	1	1
$\phi = 0.6$	2	2	2	2	2	2	0	1	1	1	1
$\phi = 0.7$	2	2	2	2	2	2	2	0	1	1	1
$\phi = 0.8$	2	2	2	2	2	2	2	2	0	1	1
$\phi = 0.9$	2	2	2	2	2	2	2	2	2	0	1
$\phi = 0.99$	2	2	2	2	2	2	2	2	2	2	0

Note:- 0 denotes indeterminate, 1 denotes MA(1) preferred, and 2 denotes AR(1) preferred

Table 5.14

AR(1) vs. MA(1) comparison: GMSFEM₂

	$\theta = 0$	$\theta = 0.1$	$\theta = 0.2$	$\theta = 0.3$	$\theta = 0.4$	$\theta = 0.5$	$\theta = 0.6$	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.9$	$\theta = 0.99$
$\phi = -0.99$	2	2	2	2	2	2	2	2	2	2	0
$\phi = -0.9$	2	2	2	2	2	2	2	2	0	0	0
$\phi = -0.8$	2	2	2	2	2	2	0	0	0	0	0
$\phi = -0.7$	2	2	2	2	2	2	0	0	0	0	0
$\phi = -0.6$	2	2	2	2	2	0	0	0	0	0	0
$\phi = -0.5$	2	2	2	2	0	0	0	0	0	0	0
$\phi = -0.4$	2	2	2	0	0	0	0	0	0	0	0
$\phi = -0.3$	2	2	2	0	0	0	1	1	1	1	1
$\phi = -0.2$	2	2	0	1	1	1	1	1	1	1	1
$\phi = -0.1$	2	0	1	1	1	1	1	1	1	1	1
$\phi = 0$	0	1	1	1	1	1	1	1	1	1	1
$\phi = 0.1$	2	0	1	1	1	1	1	1	1	1	1
$\phi = 0.2$	2	2	0	1	1	1	1	1	1	1	1
$\phi = 0.3$	2	2	2	0	1	1	1	1	1	1	0
$\phi = 0.4$	2	2	2	2	0	1	1	1	0	0	0
$\phi = 0.5$	2	2	2	2	2	0	1	1	0	0	0
$\phi = 0.6$	2	2	2	2	2	2	0	0	0	0	0
$\phi = 0.7$	2	2	2	2	2	2	2	0	0	0	0
$\phi = 0.8$	2	2	2	2	2	2	2	0	0	0	0
$\phi = 0.9$	2	2	2	2	2	2	2	2	0	0	0
$\phi = 0.99$	2	2	2	2	2	2	2	2	2	0	0

Note:- 0 denotes indeterminate, 1 denotes MA(1) preferred, and 2 denotes AR(1) preferred

Table 5.15

AR(1) vs. MA(1) comparison: GMSFEM₁ (region magnified)

ϕ	θ	0.5	0.52	0.54	0.56	0.58	0.6	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98
-0.84		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.82		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.8		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.78		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.76		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.74		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.72		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.7		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.68		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.66		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.64		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.62		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.6		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.58		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.56		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.54		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.52		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.5		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.48		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
-0.46		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Note:- 0 denotes indeterminate, 1 denotes MA(1) preferred, and 2 denotes AR(1) preferred

Table 5.16

AR(1) vs. MA(1) comparison: GMSFEM₂ (region magnified)

ϕ	θ	0.5	0.52	0.54	0.56	0.58	0.6	0.62	0.64	0.66	0.68	0.7	0.72	0.74	0.76	0.78	0.8	0.82	0.84	0.86	0.88	0.9	0.92	0.94	0.96	0.98
-0.84		2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.82		2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.8		2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.78		2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.76		2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.74		2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.72		2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.7		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.68		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.66		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.64		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.62		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.6		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.58		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.56		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.54		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.52		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.5		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.48		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-0.46		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Note:- 0 denotes indeterminate, 1 denotes MA(1) preferred, and 2 denotes AR(1) preferred

problems, and this pattern persists with an inference of indeterminacy in these unusual cases.

A clear feature of the GMSFEM criterion is that a significant range of values for ϕ and θ exists for which no conclusive model ranking is returned, with neither model dominant over the other. Much of this indeterminacy results from the potentially overly stringent requirement that dominance must hold for all vectors d , although in practice many such linear combinations of forecast errors are likely to be important. In one sense this indeterminacy can be viewed as a disadvantage since a ranking is not obtained. This is especially true if one forecast is dominant for all vectors d of interest to the practitioner, with indeterminacy arising from lack of dominance for some other irrelevant d . In these cases GMSFEM indeterminacy is a drawback and the cost of employing a reliable summary criterion. In other circumstances however, remembering that many d are likely to be relevant, it can be argued that indeterminacy implies neither forecast is preferable overall and use of a summary measure to try and ‘force’ a ranking is inappropriate. This feature of GMSFEM indeterminacy can also therefore be seen as a benefit, highlighting cases where an alternative to the summary approach, involving evaluation for the horizons, representations and loss functions of interest individually, would be better.

The MSFE measure, then, does not include information from both forecast horizons simultaneously and is not invariant for all isomorphic representations of the system, and the GFESM criterion possesses disturbing characteristics (rank reversals and counter-intuitive rankings when compared with MSFE rankings at each horizon) when comparing two misspecified models’ forecasts. The GMSFEM criterion,

however, exhibits the advantages of the GFESM measure, without the possibility of reversals or rankings which conflict with MSFE. Where GFESM behaved problematically, GMSFEM gives an indeterminate ranking. This new criterion consequently has significant value in the practice of ranking competing forecasts.

5.5 Forecast Encompassing for Linear Combinations of Forecasts

The new criterion examined in section 5.4, the GMSFEM criterion, chooses the forecast which has the smallest $d'\phi_h d$ for all non-zero vectors d (if such an optimum exists). This is identical to choosing the smallest MSFE for all non-zero d for the following linear combination of forecasts:

$$f_{cT}(h) = d_1 \hat{y}_T(1) + d_2 \hat{y}_T(2) + \dots + d_h \hat{y}_T(h) \quad (5.42)$$

which has the forecast error:

$$e_{cT}(h) = d_1 e_T(1) + d_2 e_T(2) + \dots + d_h e_T(h) \quad (5.43)$$

It is then trivial to show that:

$$E[e_{cT}^2(h)] = d'\phi_h d \quad (5.44)$$

Given this notion of a linear combination of forecasts, the approach of which is implicitly suggested by the GMSFEM criterion, an interesting issue to analyse is that of forecast encompassing. One forecast is said to encompass another if the inferior forecast has no valuable information with regard to prediction to contribute to the encompassing forecast. Clements & Hendry note the value of forecast encompassing in the evaluation of single forecasts, and so the natural question arises as to whether it is possible to characterise the conditions for and test for forecast encompassing when two linear combinations of forecasts are being considered. The issue then is whether one linear combination of forecasts encompasses another for all possible weightings of the composite forecasts. More formally, the combined

forecasts can be written as:

$$\text{actual value: } y_{c,T+h} = d_1 y_{T+1} + d_2 y_{T+2} + \dots + d_h y_{T+h} \quad (5.45)$$

$$\text{forecast 1: } f_{1cT}(h) = d_1 \hat{y}_{1T}(1) + d_2 \hat{y}_{1T}(2) + \dots + d_h \hat{y}_{1T}(h) \quad (5.46)$$

$$\text{forecast 2: } f_{2cT}(h) = d_1 \hat{y}_{2T}(1) + d_2 \hat{y}_{2T}(2) + \dots + d_h \hat{y}_{2T}(h) \quad (5.47)$$

and the consideration is whether $f_{1cT}(h)$ encompasses $f_{2cT}(h)$ for all non-zero vectors d (or vice-versa).

Forecast encompassing is examined thoroughly in chapter 3; in this case the condition for forecast 1 encompassing forecast 2 is that the optimal value of λ in the following combined forecast is zero:

$$f_{cT}^c(h) = (1 - \lambda) f_{1cT}(h) + \lambda f_{2cT}(h) \quad (5.48)$$

Manipulation of expression (5.48) yields the forecast encompassing regression:

$$e_{1ct}(h) = \lambda [e_{1ct}(h) - e_{2ct}(h)] + \varepsilon_t \quad (5.49)$$

and the condition for forecast 1 to encompass forecast 2 then revolves around λ .

Now the population interpretation of λ is:

$$\lambda = \frac{E\{e_{1ct}(h)[e_{1ct}(h) - e_{2ct}(h)]\}}{E\{[e_{1ct}(h) - e_{2ct}(h)]^2\}}$$

and thus with a null of $H_0: \lambda = 0$ (i.e. a null of forecast 1 encompassing forecast 2),

the following forecast encompassing conditions result:

$$H_0: E\{e_{1ct}(h)[e_{1ct}(h) - e_{2ct}(h)]\} = 0$$

$$\text{now } e_{ict}(h) = d_1 e_{it}(1) + d_2 e_{it}(2) + \dots + d_h e_{it}(h) \quad i = 1, 2$$

$$= d' e_{ith}$$

$$\text{where } d' = [d_1 \quad d_2 \quad \dots \quad d_h]$$

$$e'_{ith} = [e_{it}(1) \quad e_{it}(2) \quad \dots \quad e_{it}(h)]$$

$$\text{giving } H_0: E[d' e_{1th} e'_{1th} d - d' e_{1th} e'_{2th} d] = 0 \quad \forall d \neq 0$$

$$H_0: d' E[e_{1th} e'_{1th} - e_{1th} e'_{2th}] d = 0 \quad \forall d \neq 0 \quad (5.50)$$

By supposing appropriate hypothetical vectors d , it is simple to deduce that the diagonal terms of $E[e_{1th} e'_{1th} - e_{1th} e'_{2th}]$ must be zero, and that symmetric pairs of off-diagonal terms of $E[e_{1th} e'_{1th} - e_{1th} e'_{2th}]$ must sum to zero. Under these conditions, forecast 1 will encompass forecast 2. More formally, the encompassing conditions can be described as follows. Let:

$$C = E[e_{1th} e'_{1th} - e_{1th} e'_{2th}]$$

$$\text{conditions: } c_{ii} = 0 \quad i = 1, \dots, h \quad (5.51)$$

$$c_{ij} + c_{ji} = 0 \quad i, j = 1, \dots, h; i \neq j \quad (5.52)$$

This is proven by letting d be a vector with k 'th element one, all other elements zero, for $k = 1, \dots, h$, giving conditions (5.51), then letting d be all possible vectors with two elements one, all other elements zero, giving conditions (5.52).

This can also be illustrated by an example, e.g. $h = 2$:

$$d'Cd = 0 \quad \forall d \neq 0$$

$$\begin{bmatrix} d_1 & d_2 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 0 \quad \forall d \neq 0$$

$$\text{if } d' = [1 \ 0], \text{ then } c_{11} = 0$$

$$\text{if } d' = [0 \ 1], \text{ then } c_{22} = 0$$

$$\text{if } d' = [1 \ 1], \text{ then } c_{12} + c_{21} = 0$$

The conditions for forecast encompassing when two linear combinations of multi-step-ahead forecasts are considered can therefore be characterised in this way.

The next issue regards testing for forecast encompassing in such cases. Returning to the example, the conditions for forecast 1 encompassing forecast 2 are:

$$(i) \quad E[e_{1t}^2(1) - e_{1t}(1)e_{2t}(1)] = 0 \quad (5.53)$$

$$(ii) \quad E[e_{1t}^2(2) - e_{1t}(2)e_{2t}(2)] = 0 \quad (5.54)$$

$$(iii) \quad E[e_{1t}(1)e_{1t}(2) - e_{1t}(1)e_{2t}(2)] \\ + E[e_{1t}(1)e_{1t}(2) - e_{1t}(2)e_{2t}(1)] = 0 \quad (5.55)$$

The first two of these conditions, (5.53) and (5.54), amount to testing that the individual forecasts from forecast 1, i.e. $\hat{y}_{1T}(1), \hat{y}_{1T}(2)$, encompass the corresponding individual forecasts from forecast 2, i.e. $\hat{y}_{2T}(1), \hat{y}_{2T}(2)$, respectively. This is necessary, but not sufficient, for encompassing, and the third condition (5.55) is also required to ensure forecast encompassing for all linear combinations of the next two predictions.

Testing of these conditions individually can easily be performed using the modified

Diebold-Mariano approach test (following Diebold & Mariano, 1995). Application of Bonferroni bounds (following Dunn, 1961; see Campbell & Ghysels, 1995 for an example) allows the combined tests to have the desired level of significance. The problem is that for moderate and large values of h , the number of conditions to be tested will be $h + h(h - 1) / 2$, thus the large number of individual tests required to test for encompassing will most likely lead to a test with low power, unless the number of available observations is very large compared with h .

Altogether, it is useful, following forecast comparisons by a summary criterion such as GMSFEM, to characterise the conditions needed for the related issue of forecast encompassing for linear combinations of forecasts. The conditions can be easily written down, and a formal test devised using modified Diebold-Mariano approach tests with Bonferroni bounds. Problems may also arise, however, with regard to the power of the test in practical applications.

5.6 Conclusion

In conclusion, this chapter has sought to examine a number of criteria which may be used to rank competing forecasts. The common MSFE-based approach has many advantages, employing simple measures with intuitive interpretations in terms of economic loss. Clements & Hendry (1993) criticise such an approach due to the lack of MSFE invariance across all isomorphic representations of the system concerned. Several discussants to the Clements & Hendry paper counter this criticism with compelling arguments for why MSFE criteria are appropriate, with the lack of invariance expected (because misspecified models are the norm) and even desirable, with different representations being utilised by different users according to their purposes and preferences.

Clements & Hendry insist on the need for an invariant criterion and propose the GFESM measure. This criterion overcomes the basic lack of invariance exhibited by many of the currently applied measures and provides a single ranking of competing multi-step-ahead forecasts, but exhibits significant drawbacks. Its justification stems from a predictive likelihood approach to forecast ranking in a world where estimation of the criterion is performed using repeated independent simulations of the system. This justification can only be maintained by appealing to a thought experiment when the criterion is estimated using consecutive starting points in a time series, and only applies in certain circumstances. The value and reliability of the criterion are more dramatically called into question by the analysis showing that when comparing misspecified models, GFESM can yield rank reversals when the maximum forecast horizon considered is changed. The rankings

obtained can also totally conflict with the more intuitive MSFE conclusions for the individual forecast horizons evaluated separately. Further criticisms are added by other discussants to the Clements & Hendry paper, such as the implicit focus on 1-step-ahead forecasts, and the lack of GFESM correspondence to established intuition as regards minimising expected economic loss.

A second invariant criterion, GMSFEM, is considered, which cannot by definition reverse when different maximum forecast horizons are used. The measure is based on forecast dominance and retains many of the advantages of the GFESM criterion - in particular providing a summary criterion incorporating all steps-ahead forecasts which is invariant to isomorphic transformations of the system - whilst overcoming the undesirable behaviour of GFESM when misspecified models are compared. Cases do exist where GMSFEM yields an indeterminate conclusion and fails to rank the forecasts; this can be seen as an indication of when an approach other than this summary technique might be more suitable since neither forecast is dominant overall, although in some situations indeterminacy will arise from overly stringent dominance conditions not being met, in which case this failure to rank is disadvantageous, and the cost of using a reliable summary criterion. The predictive likelihood justification which can be argued in favour of GFESM also supports this new criterion since a forecast preferred by GFESM cannot have its competitor preferred by GMSFEM.

Following analysis of the GMSFEM criterion, the related issue of forecast encompassing for linear combinations of forecasts is examined. The necessary and sufficient conditions for such encompassing can be characterised for a given

maximum forecast horizon, and tested for using an approach following Diebold & Mariano (1995), albeit with some practical problems.

Altogether, then, MSFE-based measures have shortcomings in the area of invariance which is overcome by use of the GFESM criterion. However, GFESM faces immense criticism on a number of grounds. Invariance is a desirable property for ranking criteria, but the price of employing the GFESM measure is too high; where invariance is essential, the GMSFEM criterion provides a more reliable alternative. Two sensible approaches for ranking contending forecasts are recommended - one is to evaluate according to a criterion which best represents the economic loss of the decision being made, using the forecast horizon(s) and model representation of interest alone (e.g. MSFE); the other is to employ a summary criterion which has the property of invariance to isomorphic transformations and includes information on all horizons up to a chosen maximum (e.g. GMSFEM). The best overall approach in the practice of ranking competing forecasts therefore seems to be one which centres around the user's purpose, taking account of a variety of available criteria, and maintaining careful awareness of the limitations of each.

Chapter 6

Conclusion

6.1 Summary

This thesis has studied the issue of economic forecast evaluation with the aim of developing statistical procedures to enable reliable comparison of competing forecasts. Three key subjects are analysed which have wide application and comprise some of the most important forecast evaluation techniques - testing for equal forecast accuracy and testing for forecast encompassing, and ranking forecasts according to a pre-determined criterion.

With regard to testing the companion hypotheses of equal accuracy and encompassing, the qualities of a ‘good’ test are explained. These features are an approximately correct test size in all sample sizes (not just in the limit), robustness to likely forecast error properties (particularly contemporaneous correlation, autocorrelation, non-normality and ARCH), and relatively good power performance (relative to rival procedures).

The recommended tests for the two hypotheses which best meet these criteria are the modified Diebold-Mariano approach tests developed in this thesis. The original Diebold-Mariano test for equal forecast accuracy is shown to display robustness to the aforementioned error properties (with the exception of ARCH), and takes a very general loss function specification. The limitation of the test in terms of finite sample oversizing is greatly reduced by the modifications of chapter 2, and a variant of this modified Diebold-Mariano test is developed in chapter 3 to provide a similarly attractive test for forecast encompassing.

Other extant tests (regression-based) for equal accuracy and encompassing are

shown to lack robustness to non-normal forecast errors due to the resulting conditional heteroscedasticity in the regression errors, and corrections attaining robustness have mixed success, in general being eclipsed by the superior size and power performance of the modified Diebold-Mariano approach tests. One exception is the newly proposed rank correlation approach for each hypothesis, where power gains can be achieved when one-step-ahead forecasts are the sole concern and there is a strong suspicion of error non-normality. However, the rank correlation tests' advantages are restricted to this special case, and the lack of extension to multi-step predictions leads to the inference that the modified Diebold-Mariano approach tests are the best available tests for the respective evaluation hypotheses in general.

When ARCH is a suspected characteristic of the forecast errors, further extensions to the tests are required to maintain robustness. Two new tests proposed in chapter 4 adapt the respective modified Diebold-Mariano approach tests to again provide reliable, robust, well-behaved procedures for testing the equal forecast accuracy and forecast encompassing hypotheses.

The remaining issue studied - criteria for ranking competing forecasts - is more contentious. When evaluating multi-step forecasts, two approaches are possible and debate surrounds which should be employed in practice. One approach makes use of the MSFE-type criteria which represent economic loss and are specific to the forecast horizon(s) and model representation of interest. The other approach utilises a summary criterion which attempts to provide a unique ranking of the contending forecasts for all horizons and representations using information at every horizon of interest simultaneously.

The Clements & Hendry GFESM criterion performs the task of a summary criterion, but has faced much criticism by advocates of the former approach. Analysis in chapter 5 finds GFESM to have weaknesses in its theoretical justification and to possess disturbing characteristics when ranking two misspecified models - namely rank reversals when the maximum horizon considered is changed, and results that directly conflict with the more intuitive MSFE rankings for each horizon evaluated separately. A more reliable criterion, GMSFEM, is proposed for this 'summary' approach to ranking which maintains the advantages of GFESM and overcomes its undesirable characteristics, allowing for the possibility of an indeterminate ranking when neither forecast is dominant overall.

The related issue of forecast encompassing for linear combinations of forecasts is also discussed and a method for testing such a hypothesis is proposed.

Altogether, the main contributions of this thesis are the development of robust well-behaved tests for equal forecast accuracy and forecast encompassing, and the advancement of techniques for ranking competing multi-step forecasts. With regard to the former, the modified Diebold-Mariano approach tests proposed in this study are recommended as the best available tests for the companion evaluation hypotheses of equal accuracy and encompassing, with the respective variants of the tests to be employed when ARCH is suspected in the forecast errors. Turning to the latter issue, the most sensible recommendation appears to be for an approach which is user-centred and combines the two methods of ranking discussed above, using both MSFE-type criteria and GMSFEM as a summary criterion, remaining alert to

the limitations of the different criteria employed.

6.2 Directions for Future Research

There are a number of areas following the analysis conducted in this thesis which provide potential ideas for future research. Firstly, this study has focused on theoretical developments of forecast evaluation procedures, and it would be most interesting to perform a variety of applied studies. The tests for equal forecast accuracy and forecast encompassing could be applied to data on the past performance of a number of competing forecasters, for example predictions of UK macroeconomic time series by leading forecasters such as Her Majesty's Treasury, the London Business School, the National Institute of Economic and Social Research, *etc.* Such a study would be useful not only to compare the forecasts from the respective forecast producers using the recommended tests of this thesis, but also to evaluate the relative performances of all the evaluation tests considered in chapters 2-4. Furthermore, the various ranking criteria analysed in chapter 5 could be applied, yielding greater understanding of the behaviour of each criterion, their relative advantages and disadvantages, and the extent to which the GFESM problems of reversals and the GMSFEM drawback of indeterminacy occur in practice. Tests for forecast encompassing for a linear combination of forecasts could also be executed, bearing in mind the limitations of the procedure discussed in chapter 5.

A second important issue for further research concerns the properties of economic forecast errors. The likely properties examined in the context of the forecast evaluation tests are contemporaneous correlation, autocorrelation, non-normality and ARCH. Given sufficiently long time series data on past runs of forecasts and

actuals, the presence of these error properties could be tested for and the extent of each determined for practical applications. Information gleaned from such research would be invaluable, providing detail on the expected behaviour of tests for both equal accuracy and encompassing.

Another application of the forecast evaluation tests and ranking criteria is in model building. If a number of observations are kept back as a ‘hold-out’ sample, then different model specifications can be evaluated according to how well they forecast the hold-out data. In line with this, Clark (1996a, 1996b) provides simulation studies evaluating the behaviour of equal accuracy and encompassing tests when applied to out-of-sample forecasts from a bivariate vector autoregression. The tests and criteria can therefore be used as a diagnostic tool, as mentioned in chapter 1. This area generates two avenues for possible research. One is to employ the newly proposed tests and criteria in applied model building. The other is to examine the question of how large the hold-out sample should be, trading off information for model specification with data for forecast evaluation.

Returning to more theoretical concerns, one observation from the simulation results of this thesis is that the asymptotic test sizes are approached extremely slowly empirically, with very large samples required to come close to the theoretical limiting sizes. Research into why this occurs would be interesting.

Finally, useful research could be conducted into some generalisations of the problems considered in this study. Firstly, the situation where several forecasts are being considered is of interest, where for example the equal accuracy tests would be

for a null hypothesis that all forecasts have equal accuracy, with the t -tests employed in chapter 2 requiring extension to F -tests. Secondly, analysis of the extant evaluation tests highlighted the problems associated with testing hypotheses based on parameters in a regression where the variables follow a bivariate Student's t distribution (causing conditional heteroscedasticity in the regression errors) and when the variables follow ARCH processes (causing autocorrelation in the loss differential). It would consequently be interesting to study other situations where these problems arise - both from the perspective of other forecast error features generating similar effects (e.g. error properties other than non-normal which cause regression error conditional heteroscedasticity) and the angle of examining whether other economic problems exhibit these same difficulties, outside the sphere of forecast evaluation studied in this thesis.

References

- Armstrong, J. S. and Fildes, R. (1995), 'Correspondence on the selection of error measures for comparisons among forecasting methods (with reply)', *Journal of Forecasting*, 14, 67-75.
- Ash, J. C. K. and Smyth, D. J. (1973), *Forecasting the United Kingdom Economy*, Saxon House, Farnborough.
- Ashley, R., Granger, C. W. J. and Schmalensee, R. (1980), 'Advertising and Aggregate Consumption: An Analysis of Causality', *Econometrica*, 48, 1149-1167.
- Bates, J. M. and Granger, C. W. J. (1969), 'The Combination of Forecasts', *Operational Research Quarterly*, 20, 451-468.
- Bjørnstad, J. F. (1990), 'Predictive Likelihood: A Review', *Statistical Science*, 5, 242-265.
- Bollerslev, T. (1986), 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics*, 31, 307-328.
- Bollerslev, T. (1990), 'Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model', *Review of Economics and Statistics*, 72, 498-505.
- Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992), 'ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence', *Journal of Econometrics*, 52, 5-59.
- Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988), 'A Capital Asset Pricing Model with Time Varying Covariances', *Journal of Political Economy*, 96, 116-131.
- Bunn, D. W. (1975), 'A Bayesian Approach to the Linear Combination of Forecasts', *Operational Research Quarterly*, 26, 325-329.

- Bunn, D. W. (1977), 'A Comparative Evaluation of the Outperformance and Minimum Variance Procedures for Linear Syntheses of Forecasts', *Operational Research Quarterly*, 28, 653-660.
- Campbell, B. and Ghysels, E. (1995), 'Is the Outcome of the Federal Budget Process Unbiased and Efficient? A Nonparametric Assessment', *Review of Economics and Statistics*, 77, 17-31.
- Chong, Y. Y. and Hendry, D. F. (1986), 'Econometric Evaluation of Linear Macroeconomic Models', *Review of Economic Studies*, 53, 671-690.
- Christiano, L. J. and den Haan, W. J. (1996), 'Small-Sample Properties of GMM for Business-Cycle Analysis', *Journal of Business and Economic Statistics*, 14, 309-327.
- Clark, T. E. (1996a), 'Finite-Sample Properties of Tests for Forecast Equivalence', Discussion Paper, Research Division, Federal Reserve Bank of Kansas City.
- Clark, T. E. (1996b), 'Finite-Sample Properties of Forecast Encompassing Tests', Discussion Paper, Research Division, Federal Reserve Bank of Kansas City.
- Clemen, R. T. (1989), 'Combining Forecasts: A Review and Annotated Bibliography', *International Journal of Forecasting*, 5, 559-583.
- Clements, M. P. and Hendry, D. F. (1993), 'On the Limitations of Comparing Mean Square Forecast Errors (with discussion)', *Journal of Forecasting*, 12, 617-676.
- Clements, M. P. and Hendry, D. F. (1995), 'Forecasting in Cointegrated Systems', *Journal of Applied Econometrics*, 10, 127-146.
- Clements, M. P. and Hendry, D. F. (1996), 'An Empirical Study of Seasonal Unit Roots in Forecasting', Discussion Paper, Department of Economics, University of Warwick.
- Cooper, J. P. and Nelson, C. R. (1975), 'The Ex-Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models, and Some Results on Composite Predictions', *Journal of Money, Credit and Banking*, 7, 1-32.

- Dhrymes, P. J., Howrey, E. P., Hymans, S. H., Kmenta, J., Learner, E. E., Quandt, R. E., Ramsey, J. B., Shapiro, H. T. and Zarnowitz, V. (1972), 'Criteria for Evaluation of Econometric Models', *Annals of Economic and Social Measurement*, 1, 291-324.
- Dickinson, J. P. (1973), 'Some Statistical Results on the Combination of Forecasts', *Operational Research Quarterly*, 24, 253-260.
- Dickinson, J. P. (1975), 'Some Comments on the Combination of Forecasts', *Operational Research Quarterly*, 26, 205-210.
- Diebold, F. X. (1989), 'Forecast Combination and Encompassing: Reconciling Two Divergent Literatures', *International Journal of Forecasting*, 5, 589-592.
- Diebold, F. X. and Mariano, R. S. (1995), 'Comparing Predictive Accuracy', *Journal of Business and Economic Statistics*, 13, 253-263.
- Diebold, F. X. and Rudebusch, G. D. (1991), 'Forecasting Output with the Composite Leading Index: An Ex Ante Analysis', *Journal of the American Statistical Association*, 86, 603-610.
- Dunn, O. J. (1961), 'Multiple Comparisons Among Means', *Journal of the American Statistical Association*, 56, 52-64.
- Dunnett, C. W., and Sobel, M. (1954), 'A Bivariate Generalisation of Student's t-distribution, with Tables for Certain Special Cases', *Biometrika*, 41, 153-169.
- Engle, R. F. (1982), 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica*, 50, 987-1007.
- Engle, R. F. (1983), 'Estimates of the Variance of United-States Inflation Based Upon the ARCH Model', *Journal of Money, Credit and Banking*, 15, 286-301.
- Engle, R. F., Granger, C. W. J. and Kraft, D. F. (1984), 'Combining Competing Forecasts of Inflation Using a Bivariate ARCH Model', *Journal of Economic Dynamics and Control*, 8, 151-165.

- Engle, R. F. and Kroner, K. F. (1995), 'Multivariate Simultaneous Generalized ARCH', *Econometric Theory*, 11, 122-150.
- Engle, R. F. and Yoo, B. S. (1987), 'Forecasting and Testing in Cointegrated Systems', *Journal of Econometrics*, 35, 143-159.
- Granger, C. W. J. (1969), 'Investigating Causal Relations by Econometric Models and Cross-Spectral Methods', *Econometrica*, 37, 424-438.
- Granger, C. W. J. (1989), 'Combining Forecasts - Twenty Years Later', *Journal of Forecasting*, 8, 167-173.
- Granger, C. W. J. and Newbold, P. (1973), 'Some Comments on the Evaluation of Economic Forecasts', *Applied Economics*, 5, 35-47.
- Granger, C. W. J. and Newbold, P. (1975), 'Economic Forecasting - The Atheist's Viewpoint', in *Modelling the Economy*, ed. G. A. Renton, Heinemann Educational Books, London.
- Granger, C. W. J. and Newbold, P. (1986), *Forecasting Economic Time Series*, 2nd edn., Academic Press, Orlando, Florida.
- Granger, C. W. J. and Ramanathan, R. (1984), 'Improved Methods of Forecasting', *Journal of Forecasting*, 3, 197-204.
- Hamilton, J. D. (1984), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.
- Hansen, L. P. and Hodrick, R. J. (1980), 'Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis', *Journal of Political Economy*, 88, 829-853.
- Harvey, D. I., Leybourne, S. J. and Newbold, P. (1998), 'Tests for Forecast Encompassing', *Journal of Business and Economic Statistics*, forthcoming.
- Hendry, D. F. (1996), 'The Econometrics of Macroeconomic Forecasting', Discussion Paper, Nuffield College, University of Oxford.

- Hendry, D. F. and Richard, J.-F. (1982), 'On the Formulation of Empirical Models in Dynamic Econometrics', *Journal of Econometrics*, 20, 3-33.
- Hendry, D. F. and Richard, J.-F. (1983), 'The Econometric Analysis of Economic Time Series (with discussion)', *International Statistical Review*, 51, 111-163.
- Hendry, D. F. and Richard, J.-F. (1989), 'Recent Developments in the Theory of Encompassing', in *Contributions to Operations Research and Econometrics. The Twentieth Anniversary of CORE*, eds. B. Cornet and H. Tulkens, MIT Press, Cambridge, Massachusetts.
- Hogg, R. V. and Craig, A. T. (1978), *Introduction to Mathematical Statistics*, 4th edn., MacMillan, New York.
- Holden, K. and Peel, D. A. (1983), 'Forecasts and Expectations: Some Evidence for the UK', *Journal of Forecasting*, 2, 51-58.
- Holden, K. and Peel, D. A. (1985), 'An Evaluation of Quarterly National Institute Forecasts', *Journal of Forecasting*, 4, 227-234.
- Holden, K. and Peel, D. A. (1986), 'An Empirical Investigation of Combinations of Economic Forecasts', *Journal of Forecasting*, 5, 229-242.
- Holden, K. and Peel, D. A. (1988), 'A Comparison of Some Inflation, Growth and Unemployment Forecasts', *Journal of Economic Studies*, 15, no.5, 45-52.
- Howrey, E. P., Klein, L. R. and McCarthy, M. D. (1974), 'Notes on Testing the Predictive Performance of Econometric Models', *International Economic Review*, 15, 366-383.
- Kendall, M. G. and Gibbons, J. D. (1990), *Rank Correlation Methods*, 5th edn., Edward Arnold, London.
- Makridakis, S., Anderson, A. Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982), 'The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition', *Journal of Forecasting*, 1, 111-153.

- Makridakis, S., Anderson, A. Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1983), *The Forecasting Accuracy of Major Time Series Methods*, Wiley, London.
- Makridakis, S. and Hibon, M. (1979), 'Accuracy of Forecasting: An Empirical Investigation (with discussion)', *Journal of the Royal Statistical Society, Series A*, 142, 97-145.
- Meese, R. A. and Rogoff, K. (1988), 'Was it Real? The Exchange Rate-Interest Differential Relation Over the Modern Floating-Rate Period', *Journal of Finance*, 43, 933-948.
- Mizon, G. E. (1984), 'The Encompassing Approach in Econometrics', in *Analyzing Modern Business Cycles*, ed. P. A. Klein, M. E. Sharpe Inc., London.
- Mizon, G. E. and Richard, J.-F. (1986), 'The Encompassing Principle and its Application to Non-Nested Hypothesis Tests', *Econometrica*, 54, 657-678.
- Morgan, W. A. (1939-40), 'A Test for Significance of the Difference Between the Two Variances in a Sample From a Normal Bivariate Population', *Biometrika*, 31, 13-19.
- Nelson, C. R. (1972), 'The Prediction Performance of the F.R.B.-M.I.T.-PENN Model of the U.S. Economy', *American Economic Review*, 62, 902-917.
- Newbold, P. and Granger, C. W. J. (1974), 'Experience with Forecasting Univariate Time Series and the Combination of Forecasts (with discussion)', *Journal of the Royal Statistical Society, Series A*, 137, 131-149.
- Newey, W. K. and West, K. D. (1994), 'Automatic Lag Selection in Covariance Matrix Estimation', *Review of Economic Studies*, 61, 631-653.
- Öller, L.-E. (1978), 'A Method for Pooling Forecasts', *Journal of the Operational Research Society*, 29, 55-63.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic Press, New York.

- Reid, D. J. (1968), 'Combining Three Estimates of Gross Domestic Product', *Economica*, 35, 431-444.
- Reid, D. J. (1969), 'A Comparative Study of Time Series Prediction Techniques on Economic Data', Ph.D. thesis, University of Nottingham.
- Rich, R. W., Raymond, J. E. and Butler, J. S. (1992), 'The Relationship Between Forecast Dispersion and Forecast Uncertainty - Evidence from a Survey Data ARCH Model', *Journal of Applied Econometrics*, 7, 131-148.
- Stekler, H. O. (1991), 'Macroeconomic Forecast Evaluation Techniques', *International Journal of Forecasting*, 7, 375-384.
- Vuong, Q. H. (1989), 'Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses', *Econometrica*, 57, 307-334.
- Wallis, K. F. (1989), 'Macroeconomic Forecasting: A Survey', *Economic Journal*, 99, 28-61.
- Wallis, K. F. (ed.), Andrews, M. J., Fisher, P. G., Longbottom, J. A. and Whitley, J. D. (1986), *Models of the UK Economy: A Third Review by the ESRC Macroeconomic Modelling Bureau*, Oxford University Press, Oxford.
- Wallis, K. F. (ed.), Fisher, P. G., Longbottom, J. A., Turner, D. S. and Whitley, J. D. (1987), *Models of the UK Economy: A Fourth Review by the ESRC Macroeconomic Modelling Bureau*, Oxford University Press, Oxford.
- White, H. (1980), 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica*, 48, 817-838.
- White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, Orlando, Florida.