2475/3742

MONASH UNIVERSITY
THESIS ACCEPTED IN SATISFACTION OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
ON..... 7 September 2001 ..
.....
Sec. Research Graduate School Committee

Contributions to the Theory and Practice of Hypothesis Testing

A thesis submitted for the degree of
Doctor of Philosophy
in Econometrics

Sivagowry Srianantha kumar

B.Sc., Grad.Dip.Ec., M.Ec.

Department of Econometrics and Business Statistics
Monash University
Australia
October 2000

CONTENTS

LIST OF TABLES

LIST OF FIGURES

DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Sivagowry Srianantha kumar

ACKNOWLEDGMENTS

I am grateful to my supervisor Professor Max King for his guidance, encouragement, patience, constructive criticism and editorial advice throughout the preparation of this thesis. Without his support, I would never have completed this thesis. I am truly grateful to him.

I would like to thank Associate Professor Brett Inder for his helpful comments on various aspects of the work in its early stages.

I would also like to thank Dr. Simone Grose for her helpful comments on Chapter 5 and for painstakingly checking the code of one of my computer programs.

I am thankful to Associate Professor Paramsothy Silvapulle for giving me permission to use her results in Chapter 3.

I am also thankful to my former colleague Vince Mataassa for his useful comments about the optimization algorithm used in this thesis and for providing some invaluable references on this algorithm.

I would like to thank my former room-mates Syfun Nahar and Ivet Pitrun for their cheerful company. Their company made a stress free pleasant environment which I appreciated greatly. Particularly, I would like to thank Syfun Nahar for her friendship and encouragement throughout the period this thesis was researched and written.

The completion of this thesis within the timeframe of three and a half years would not have been possible without the highly advanced computing facilities available in the Department of Econometrics and Business Statistics. A special thanks to Inge Meldgaard for her efforts in ensuring that all my computing needs were met.

From a financial perspective, I would like to thank Monash University for providing me with an Australian Postgraduate Award throughout my graduate studies.

I would like to thank my husband, mother, brothers, family and friends who provided much needed support and encouragement at various times.

Finally I would like to dedicate this thesis to the memory of my late father, Mr K. Manikawasaker (Retd. District Secretary and Government Agent, Jaffna and Kilinochchi Districts, Sri Lanka), who always encouraged me to do this Ph.D. but unfortunately passed away two months before the submission of this thesis.

ABSTRACT

Hypothesis testing plays an important role in econometrics. It is used to test aspects of economic theory or to check the specification of an econometric model. Because econometrics is a non-experimental discipline, it is essential to have reliable testing procedures, particularly in the presence of nuisance parameters. Also, econometricians often use sample sizes of under one hundred, therefore it is preferable that their testing procedures have reliable finite-sample properties. In the absence of uniformly most powerful (UMP) tests or uniformly most powerful invariant (UMPI) tests, King (1987b) suggested the use of point optimal (PO) tests, which are most powerful at a chosen point under the alternative hypothesis. Unfortunately, these tests cannot always be constructed for a composite null hypothesis. For situations where PO tests cannot be constructed, King suggested an approximate point optimal (APO) test.

Existing studies, show that King's APO tests are not always reliable. For example, Silvapulle (1991, 1994a) applied King's APO tests to two composite non-nested testing problems, namely, testing for first-order moving average (MA(1)) errors against first-order autoregressive (AR(1)) errors and testing for AR(1) errors against first-order integrated MA (IMA(1,1)) errors with a negative MA coefficient in the linear regression model. The APO test worked extremely well for the former problem but performed poorly for the latter. This and the excellent finite-sample performance of the PO test motivated us to propose an APO test, called the g test, based on the generalised Neyman-Pearson lemma (GNPL), for testing a composite null hypothesis against a simple alternative. This test can be used for testing a composite null hypothesis versus a composite alternative.

In this thesis, we apply the g test to the above two testing problems considered by Silvapulle and to the problem of testing for a static linear regression model with AR(1) errors against a dynamic linear regression model with white noise errors. For the first two problems, we compare the small-sample sizes and powers of our g test with those of Silvapulle's test with encouraging results, while for the third problem, we compare the small-sample sizes and powers of our test with those of marginal

likelihood based one-sided classical tests, such as, the likelihood ratio (LR), Wald (W) and Lagrange multiplier (LM) tests. Because classical tests are specially designed for nested testing, they are applied to test for the significance of the dynamic regressor coefficient in a dynamic linear regression model with AR(1) errors. We consider both the Laplace approximated information based and estimated information based W (and LM) tests, in order to see which test is best in finite samples.

We selected the third problem purposely because the performance of the PO and APO tests in the presence of (unavoidable) nuisance parameters is largely unknown in practice. The presence of nuisance parameters can make tests non-similar. Because it is extremely hard to obtain exact non-similar critical values of non-similar tests, few studies exist based on (approximate) non-similar critical values. In this thesis, we propose a new approach to obtain exact non-similar critical values of general non-similar tests, using a global optimizer called simulated annealing (SA). Because this can be an extremely computer intensive procedure, we also suggest and investigate the performance of near exact non-similar critical values. Because this SA based approach involves controlling the maximum size of a non-similar test over the nuisance parameter space, this method also allows one to assess the accuracy of any approximation to the distribution of the test statistic under the null hypothesis. In this thesis, we used this approach to assess the finite-sample performance of large-sample based classical tests under the null hypothesis and to check whether the large-sample distribution based test or the approximate small disturbance asymptotic (ASDA) distribution based test is more reliable. For the latter case, we used the large-sample based Durbin's t test and the ASDA distribution based Durbin-Watson (DW) test in the context of the dynamic linear regression model.

Our study clearly shows that the marginal likelihood based one-sided asymptotic LR and W tests cannot be trusted under the null hypothesis even when the sample size is 60, because they can have maximum sizes closer to one over the nuisance parameter space. Compared to all the tests considered, the Laplace approximated information based LM test seems to be the best in this regard, however, it is not ideal either. Our study also shows that neither the large sample based t test nor the ASDA distribution based DW test is best under the null because both tests can have approximately the same (higher than nominal) maximum sizes over the nuisance parameter space.

The overall recommendation of this thesis is that, if a PO test is suitable for a particular testing problem it should be used. For the testing problems we considered, the g test performs well, which suggests that this test may be useful for other situations where the PO test cannot be constructed. For the non-similar tests we considered, near exact non-similar critical values generally seem good approximations to the exact non-similar critical values. Also our study favours near exact non-similar critical values over approximate non-similar critical values.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Hypothesis testing plays an important role in econometrics. It is used to test aspects of economic theory or to check the specification of an econometric model. The non-experimental nature of econometrics (which distinguishes econometrics from its parent discipline of statistics), means econometricians have little control over sample size and there is much less certainty about model specification than in disciplines in which experiments can be conducted. Also, econometric models often have to account for factors which may not be directly of interest. Thus, econometric models typically contain large numbers of unknown parameters, about whose true functional form, little is known. For all these reasons, econometricians need reliable testing procedures, particularly in the presence of nuisance parameters (see King (1994)). Also, these tests should be as powerful as possible.

Tests that have the correct size and good power properties can be regarded as good tests. Achieving correct sizes is a problem in the presence of nuisance parameters because their presence can make tests non-similar. In other words, the test's sizes

vary with the nuisance parameter values, thus size cannot be fixed at a desired level. The conventional approach to dealing with non-similar tests is to control the maximum probability of a Type I error (or size) over the nuisance parameter space by ones choice of critical values. These critical values are called exact non-similar critical values and if they are used, sizes for all null hypothesis data generating processes will never exceed the nominal size.

For any testing problem, we would like to use a uniformly most powerful (UMP) test or a uniformly most powerful invariant (UMPI) test[1], which always have power at least as great as that of any other test of the same size. However, such tests rarely exist. If a UMP test does not exist then selecting a test becomes complicated by the fact that no single test dominates in terms of power. For this situation, Cox and Hinkley (1974) suggested three alternative approaches to test construction, namely, using a test which maximizes power at a "somewhat arbitrary typical point" in the alternative parameter space, removing this arbitrariness by choosing the point to be close to the null hypothesis and choosing the test which maximizes some weighted average of power over the alternative parameter space. Cox and Hinkley's first suggestion is known as the point optimal (PO) solution and was investigated by King (1983, 1987b). For tests based on the second suggestion, see Wu and King (1994), and for tests based on the third suggestion, see Andrews and Ploberger (1994).

In this thesis, we propose an approximate point optimal (APO) test, based on the generalised Neyman-Pearson lemma (GNPL), for testing a composite null hypothesis against a simple alternative. We denote this test as the g test. The g test can be used for testing a composite null hypothesis against a composite alternative. As explained below, this test will be useful for situations where King's PO test cannot be constructed. The g test involves the finding of multiple critical values. These critical values can be obtained by following an iterative procedure (described in Chapter 3) or by using a global optimizer called simulated annealing (SA) (described in Chapter 5). In this thesis, we also propose a new approach to obtain exact (and near exact) non-similar critical values of general non-similar tests, using SA. To author's knowledge this is the first study that uses SA for these purposes.

[1] If a test is UMP within the class of invariant tests then it is called a UMPI test.

1.1.1 Motivation for Constructing the g Test

King (1987b) introduced a PO test (which is the most powerful test at a chosen point under the alternative hypothesis) for testing nested and non-nested hypotheses[2]. Because it is the most powerful test at a chosen point, King observed that his PO test can be used to trace out the maximum attainable power envelope for a given problem. The power envelope provides a benchmark against which test procedures can be evaluated. Also, the PO test exploits the one-sided nature of the testing problem in economic applications, which is helpful for situations where one has reasons to believe that the parameter under study is positive or negative[3].

Because, PO tests are exact, they are useful for the typical sample sizes used in econometrics. For the majority of testing situations, PO tests have excellent relative power around the point at which power is optimised. They also can have good power away from this point. For example, if a UMP test exists, then the PO test will be UMP. For some testing situations, PO tests can be approximately UMP (see Shively (1988b)) or can be UMP over a subset of the alternative parameter space (see King and Smith (1986)). However, for other testing situations, the relative power performance of the test can drop away quickly as one moves away from the point at which power is optimised (see Dufour and King (1991)).

Though the PO test has several advantages, it is not suitable for every testing situation in econometrics. It seems suitable for testing problems which involve a small number of parameters and prior knowledge of the signs of the parameters under the alternative hypothesis. As King points out, in econometric applications the knowledge of the parameter's signs can often be deduced from the underlying economic theory and the number of parameters in a testing problem, particularly if one is dealing with a linear model, can sometimes be reduced by invariance methods or by considering sufficient statistics. Even if these are realised, the PO test cannot be guaranteed always to exist as we shall discuss next.

---

[2] Nested hypotheses mean the null can be obtained as a simplified version of the alternative model. If this is not the case for all models under the null then the hypotheses are known as non-nested (see McAleer and Pesaran (1986)).

[3] In hypothesis testing, using such information should result in greater power (see Wu and King (1994)).

A PO test always exists for (a) a simple null hypothesis versus a simple alternative hypothesis testing problem, and for (b) a simple null hypothesis versus a composite alternative hypothesis testing problem, but not always for (c) a composite null hypothesis versus a composite alternative hypothesis testing problem. That is, for (a), the fundamental Neyman-Pearson lemma (FNPL) provides a likelihood ratio (LR) test, which is the most powerful test, therefore, it is the PO test. Similarly, for (b), for a chosen point in the alternative parameter space, the FNPL based LR test is the most powerful test in the neighbourhood of that chosen point, therefore, it is the PO test. For (c), one might think of applying the FNPL based LR test for a chosen point under the null versus a chosen point under the alternative case, however, this test does not necessarily result in a test which is PO. The reason for this is that because the null hypothesis is composite, the sizes of the test are a function of parameters under the null hypothesis and hence vary with the values of these parameters. This means that we cannot fix the size of the test to a desired constant level. The standard approach in this situation is (as mentioned earlier) to control the maximum size to be less than or equal to some desired nominal size by ones choice of critical value. If for such a critical value, the maximum size of the FNPL based LR test occurs precisely at the chosen point under the null, then the test will be most powerful in the neighbourhood of the chosen point under the alternative, thus, it is the PO test. However, it may not be possible to obtain such critical values always, consequently this form of PO test cannot be guaranteed always to exist when testing a composite null. For situations where a PO test cannot be constructed, King (1987b) recommended an APO test.

Existing studies suggest that King's APO tests are not always reliable. For example, Silvapulle (1991, 1994a) applied King's APO tests to two composite non-nested testing problems, namely, testing for first-order moving average (MA(1)) errors against first-order autoregressive (AR(1)) errors and testing for AR(1) errors against first-order integrated MA (IMA(1,1)) errors with a negative MA coefficient in the linear regression model. The APO tests worked extremely well for the former problem but performed poorly for the latter. Studies like this indicate that it is important to have a reliable APO test for testing composite hypotheses, which econometricians frequently face. This and the remarkable finite-sample properties of the PO tests reported in previous studies motivated us to develop the g test which is introduced in Chapter 3.

1.1.2 Motivation for a New Approach to Finding Exact (and Near Exact) Non-Similar Critical Values

As noted earlier, the presence of nuisance (or unknown) parameters can make tests non-similar. To avoid the effects of nuisance parameters, researchers often search for a class of similar or invariant tests. In the former test case, the size of the test is fixed independently of the nuisance parameters and in the latter case, the nuisance parameters are avoided by making the test invariant to the class of data transformations that they represent. Invariant tests are attractive but not always available. Also the associated test statistics tend not to have standard distributions. Similarly, it may not always be possible to construct similar tests for complicated problems. Also, similar tests may be less powerful than non-similar tests (see McAleer and Pesaran (1986) and McAleer (1995) in the context of non-nested testing). Another popular approach (in the presence of nuisance parameters) is to replace the unknown parameters by consistent estimates and appeal to asymptotic theory. Recent studies show that using marginal likelihood estimates rather than classical likelihood estimates for the nuisance parameters can result in better finite-sample inferential procedures (see Ara (1995), Grose (1998) and Rahman and King (1998))[4]. However, this approach does not solve the problem of tests being non-similar.

By assuming knowledge of the nuisance parameters, exact size critical values of a non-similar test can be obtained via the Monte Carlo method. If such a critical value is used, for the nuisance parameter values we assumed the test size will be equal to the nominal size. Exact size critical values for non-similar tests are not useful in practice because they require knowledge of the unknown parameters. Exact non-similar critical values can be useful, however, they seem to be less popular because they require performing a Monte Carlo experiment each time one wants a critical value. There are few existing studies based on non-similar critical values (see Inder (1985), King and McAleer (1987), Grant (1987) and Silvapulle (1991)). However, these are typically only based on approximate non-similar critical values.

[4] Marginal likelihood is one of the likelihood based methods designed to overcome the nuisance parameters' problem and is discussed in Chapter 5.

Approximate non-similar critical values may not work well for all testing problems, design matrices and tests. Also, it is difficult to obtain them when the number of nuisance parameters is high. This motivated us to investigate a (SA based) approach to obtain exact (and near exact) non-similar critical values of general non-similar tests. This approach involves controlling the maximum size of a non-similar test over the nuisance parameter space, therefore it is also useful for assessing the accuracy of an approximation to the distribution of the test statistic under the null hypothesis.

This thesis focuses on the following issues:

(1) The theory and construction of the g test.
(2) Testing for MA(1) errors against AR(1) errors in the linear regression model, using the g test.
(3) Testing for AR(1) errors against IMA(1,1) errors in the linear regression model using the g test.
(4) Calculating exact (and near exact) non-similar critical values and exact size critical values[5] for general non-similar tests via SA.
(5) Testing a static linear regression model with AR(1) errors against a dynamic linear regression model with white noise errors, using the marginal likelihood based g test and the marginal likelihood based one-sided classical tests.

Among these, (1), (2) and (3) are discussed in Chapter 3, while (4) and (5) are discussed in Chapter 4 and Chapter 5, respectively. Unlike for the first two testing problems, nuisance parameters cannot be avoided for the last problem. We purposely selected the last problem, in order to see the performance of the g test (which is an APO test) in the presence of (unavoidable) nuisance parameters[6].

[5] Exact size critical values are obtained by assuming knowledge of the unknown parameters.
[6] Chapter 2 reveals that the performance of the PO and APO tests in the presence of unavoidable nuisance parameters is largely unknown.

1.2 Outline of the Thesis

A review of hypothesis testing is presented in Chapter 2. This chapter mainly focuses on studies involving PO tests in the context of composite hypothesis testing. Because all the applications considered in this thesis are non-nested, some popular non-nested tests are also briefly discussed. In addition, other studies that are relevant to this thesis, such as, studies based on non-similar critical values and those based on SA are also discussed.

This chapter reveals the importance of developing tests which have excellent finite-sample properties, such as, PO tests, rather than relying on large-sample based tests. For example, in the context of non-nested testing, almost all the existing tests are large-sample based and many are observed to perform poorly in finite samples. Studies involving APO tests of a composite null indicate the need for a reliable APO test. We also note that PO and APO tests have been applied mainly for situations where the nuisance parameters can be avoided via invariance arguments, which confirms the fact that their performance in the presence of unavoidable nuisance parameters is largely unknown. Studies based on approximate non-similar critical values indicate that obtaining such critical values can be quite difficult when the null distribution of the test statistic depends on more than one nuisance parameter. Finally, studies on SA confirm that the SA algorithm is much more robust than conventional algorithms.

Chapter 3 discusses the theory behind the g test and how it can be applied for testing composite non-nested disturbance covariance matrices in the linear regression model. The g test is then applied for the first two (above mentioned) testing problems and the small-sample size and power properties are compared with those from Silvapulle (1991, 1994a) and Silvapulle and King (1991). Also, the use of the g test is illustrated by its application to two real world data sets.

Chapter 4, proposes a new approach to calculating exact size critical values (t / assuming knowledge of the unknown parameters) and exact non-similar critical values of general non-similar tests. Because obtaining the exact non-similar critical

values can be time consuming, we also suggest and investigate the use of near exact non-similar critical values. For this study we consider two non-similar tests, namely, the Durbin-Watson (DW) test and Durbin's t test in the context of the dynamic linear regression model. In order to obtain near exact non-similar critical values of these tests, we use approximate small disturbance asymptotic (ASDA) and large-sample based critical values, respectively, and SA. Therefore, in this case, our SA based approach can be expected to indicate which asymptotic approach is more reliable under the null hypothesis. In this chapter, we calculate the sizes of the tests for a variety of nuisance parameter values and design matrices, in order to check whether the SA based near exact non-similar critical values are indeed working well in terms of controlling the sizes over the nuisance parameter space. We also compare near exact non-similar critical values with approximate non-similar critical values used in previous studies.

Chapter 5 explores the problem of testing for a static linear model with AR(1) errors against a dynamic linear model with white noise errors. As mentioned earlier, for this problem, the nuisance parameters cannot be avoided. Therefore, to lessen the nuisance parameters' effect, here we deal with marginal likelihood based tests. The tests considered for this problem are the marginal likelihood based g test and the marginal likelihood based one-sided LR, Lagrange multiplier (LM), and Wald (W) tests. Both Laplace approximated information based and estimated information based W (and LM) tests are considered in order to see which test is best in finite samples. The size and power comparisons of this chapter are based on near exact non-similar critical values obtained via SA. Here, the size and power calculations are made for a variety of nuisance parameter values in order to determine the best test over the nuisance parameter space.

A summary and some concluding remarks are given in the final chapter.

CHAPTER 2

HYPOTHESIS TESTING AND RESEARCH FINDINGS RELATED TO OUR STUDY: A REVIEW

2.1 Introduction

The non-experimental nature of econometrics leads to much less certainty about model specification than in disciplines in which experiments can be conducted. That is, for a particular economic data set there may be more than one plausible model. Any model seems plausible should it be able to stand up to rigorous testing using empirical data. For this purpose there exist many diagnostic tests such as the DW test for serial correlation, LM tests for various forms of heteroscedasticity, tests for functional form, structural stability, exogeneity of regressors and so on (see Beggs (1988) for an excellent survey on diagnostic tests). It is possible for these diagnostic tests to accept one or more models within the same (nested) or across different (non-nested) paradigms. It is then important to test the models which are not rejected by diagnostic tests against each other using powerful testing procedures because correct model specification is essential for inference, forecasting and policy making.

Econometricians typically use small to moderate sized samples. Because it is extremely hard to derive the finite-sample distributions of test statistics in the presence of nuisance parameters, econometricians often use large-sample based tests. However, large-sample based tests generally perform poorly in terms of size properties. For example, the majority of the tests proposed for non-nested problems are large-sample based and many are known to perform poorly in finite samples (see McAleer and Pesaran (1986) and McAleer (1987, 1995) and Godfrey (1998)). The reason for the failure of most of the asymptotic tests in finite samples may be due to their dependence on the likelihood principle. That is, the most popular basis for estimation and inference in econometrics is the maximum likelihood or likelihood principle. It is well known that maximum likelihood estimates are biased in finite samples, consequently, maximum likelihood based tests can be expected to perform relatively poorly in finite samples. Recent studies show that marginal likelihood based tests behave better than conventional likelihood based tests in small samples (see Ara (1995), Rahman and King (1998), and Grose (1998)). This suggests that better handling of nuisance parameters may improve the small-sample properties of large-sample based approaches. Because marginal likelihood based tests have been applied mainly to linear regression models and dynamic linear regression models with white noise errors, their performance for more complicated models is largely unknown.

King's (1987b) point optimal (PO) testing approach is particularly aimed at small-sample testing problems. Since its introduction, PO tests and approximate point optimal (APO) tests have been applied successfully to many testing situations. Strangely, many text books fail to mention the PO testing approach and its progress. We also observe that recent studies on non-nested testing procedures also mostly ignore this approach (see McAleer (1995) and Pesaran and Weeks (2000)). This chapter's aim is to survey contributions on PO testing in the context of composite hypothesis testing. Because all the applications considered in this thesis are non-nested, we briefly discuss some popular non-nested tests as well. Other studies relevant to this thesis are also reviewed.

The plan of this chapter is as follows. Section 2.2 introduces PO and APO tests. Section 2.3 briefly discusses some non-nested tests. Section 2.4 compares the PO

(and APO) tests of a composite null hypothesis with other existing tests. Section 2.5 reviews some other relevant studies to this thesis. Finally, concluding remarks are given in section 2.6.

2.2 Introduction to PO Tests and APO Tests

In the absence of UMP or UMPI tests, King (1987b) suggested a PO test, which is the most powerful test at a chosen point in the alternative parameter space. The idea behind his PO test is as follows. If we focus on one point in the alternative hypothesis parameter space, then for a given significance level, all tests have a single power value at this point. According to theory, for a given class of tests, the maximum (or supremum) of these power values exists and a test whose power attains this maximum is a most powerful test in the neighbourhood of the predetermined point; it is the PO test.

King proposed the PO test for a very general framework and applied it to the problem of testing for AR(1) errors against MA(1) errors in the linear regression model. He proposed the PO test for the general problem of testing

$$H_0: x \text{ has density } f(x|\omega) \tag{2.2.1}$$

against

$$H_a: x \text{ has density } g(x|\theta) \tag{2.2.2}$$

where x is the observed sample, ω is a $p\times 1$ vector of parameters restricted to the set Λ and θ is a $q\times 1$ vector of parameters restricted to the set Θ. Observe that this general testing problem incorporates both nested and non-nested problems as special cases. We assume that any knowledge about the possible range of parameter values has been used to keep the parameter sets, Λ and Θ, as small as possible.

For the simpler problem of testing

$$H_0^*: x \text{ has density } f(x,\omega_1) \tag{2.2.3}$$

against

$$H_a^*: x \text{ has density } g(x,\theta_1), \tag{2.2.4}$$

where $\omega_1 \in \Lambda$ and $\theta_1 \in \Theta$ are fixed and known, the fundamental Neyman-Pearson lemma (FNPL) implies that rejecting H_0^* for large values of $s(\omega_1,\theta_1) = g(x,\theta_1)/f(x,\omega_1)$ is the most powerful test. Suppose we wish to test a simple hypothesis H_0^* against a composite alternative H_a, the same test can serve as the most powerful test in the neighbourhood of θ_1. However, as discussed in the previous chapter, the same test does not necessarily result in a most powerful test in the neighbourhood of θ_1, when testing H_0 against H_a. King observed that this could happen only if the critical value c and the point ω_1 can be chosen such that

$$\Pr[s(\omega_1,\theta_1) > c | x \sim f(x,\omega_1)] = \alpha \tag{2.2.5}$$

and

$$\Pr[s(\omega_1,\theta_1) > c | x \sim f(x,\omega)] \leq \alpha, \text{ for all } \omega \in \Lambda, \tag{2.2.6}$$

where α is the desired level of significance. That is, we have to choose ω_1 such that the maximum size of the test occurs at this point. For situations where appropriate values of ω_1 and c cannot be found, King suggested an APO test. The APO test requires ω_1 to be chosen such that (2.2.6) holds and

$$\alpha - \Pr[s(\omega_1,\theta_1) > c | x \sim f(x,\omega_1)] \tag{2.2.7}$$

is minimised. Obviously, if (2.2.7) equals zero, then, the APO test is the PO test. Thus, the closer (2.2.7) is to zero the better.

The majority of successful applications of the PO approach have been in the context of the linear regression model. Some examples are: testing for AR(1) disturbances (Berenblut and Webb (1973), King (1985a) and Dufour and King (1991)); testing for MA(1) disturbances (King (1985b)); testing for heteroscedasticity (Evans and King (1985, 1988)), testing for random walk disturbances (Sargan and Bhargava (1983)); testing for random coefficients (Franzini and Harvey (1983), Shively (1986, 1988a, 1988b), Brooks and King (1994), Brooks (1993, 1995) and Rahman and King (1994)); testing for moving average unit roots in ARIMA models (Saikkonen and Luukkonen (1993)); testing for autoregressive disturbances in a time series regression with missing observations (Shively (1993)), testing for block effects in regression disturbances (Bhatti (1992) and Bhatti and King (1994)); and tests of non-nested error processes (King (1983, 1987b), Silvapulle (1991, 1994a, 1994b) and Silvapulle and King (1991, 1993)).

Despite its excellent performance on many occasions, King's PO test has had its critics (see Dastoor and Fisher (1988), Bierens (1988) and Potscher (1988)). The main criticisms King's PO test have received are as follows.

(1) The finite-sample distributions of the PO tests are complex, therefore, it is computationaly difficult to obtain critical values for the PO tests.

(2) If the null hypothesis is simple (possibly after reduction by invariance), PO tests can be easily constructed using FNPL. But in practice, there exist testing problems which cannot be simplified by using the invariance principle or by any other means. Will PO tests be able to be found and will they continue to perform well in such situations?

For the first reason, some researchers have ignored the PO test (even for situations where it can be easily constructed) and have developed or used tests which have known asymptotic distributions (see Burke et al. (1990) and Smith and Tremayne (1990)). Asymptotic tests are easier to apply because of their known asymptotic null distributions. However, computational simplicity is not the only criterion for making a choice of testing procedure. A good test should have right size and good power properties.

2.3 A Brief Introduction to Non-Nested Tests

The classical tests are not appropriate for non-nested hypothesis testing, because the distributions of statistics such as LR or W test statistics are not centred at zero under the null hypothesis when the hypotheses under consideration are non-nested. However, if conventional test statistics are appropriately centred (at least asymptotically), the same tests can be applied for non-nested hypothesis testing. Observing that the standard LR test is not applicable for non-nested hypothesis testing and needs to be properly centred, Cox (1961, 1962) suggested a modification of the LR test statistic (hereafter, the Cox test) which does not have the usual central chi-squared distribution under the null hypothesis and can be quite demanding in terms of computation. The Cox test has been mainly applied to linear and simple non-linear regression models (for example, see Pesaran and Deaton (1978) and Evans and Deaton (1980)). This may be due to the complex and often intractable derivations that are involved in the computation of the numerator of the Cox statistic in non-regression situations. To overcome this problem, Pesaran and Pesaran (1993) proposed a simulation based approach for computing the Cox statistic. A drawback of this approach is that it depends on a reference distribution which is valid asymptotically.

Atkinson's (1970) test is closely related to the Cox test and is based on a comprehensive model constructed by artificially nesting the non-nested models of interest. Because his testing framework is nested, standard tests designed for nested hypothesis testing can be used. However, Atkinson's test suffers some limitations (see Pesaran and Weeks (2000)).

Many other tests that are asymptotically equivalent to the Cox test have been proposed and investigated by various authors in the context of non-nested linear regression models. Among these, the following tests seem popular: two artificial nesting procedures known as the J and P tests proposed by Davidson and MacKinnon (1981, 1982), the JA test (which is an exact test under normally and independently distributed (NID) errors) proposed by Fisher and McAleer (1981), the Cox-type N test proposed by Pesaran (1974) and adjusted Cox-type tests derived by

Godfrey and Pesaran (1983). Though easy to implement, Deaton's (1982) F test (which is exact under NID errors) has been used less frequently than either the J or JA tests. This may be due to its poor power properties (see McAleer and Pesaran (1986)).

According to the literature, the Cox test and Cox-type tests can have significantly higher than nominal sizes, particularly in finite samples. Though, the JA and F tests have accurate sizes, they can be less powerful than the Cox and J tests. In order to overcome the poor performance of the Cox test under the null, Godfrey and Pesaran (1983) proposed some modifications to the test based upon mean and variance adjustments which resulted in a substantial improvement in its finite-sample performance. In particular, Godfrey and Pesaran showed that their adjusted Cox-type tests have empirical sizes that agree very closely with the nominal size, while having power that is typically much higher than those of the F and JA tests. However, their adjusted Cox tests do not have known null distributions.

In the context of non-nested linear regression models, Godfrey (1998) noted that the J test has several useful features: it is easily generalised to allow for several non-nested alternative regression models; and it has considerable intuitive appeal. He also noted that after adjustment of critical values (because the J test is known to perform poorly in small samples), the J test might be more powerful than other procedures. These observations led him to apply the bootstrap method to the J test to reduce the problem of over rejection of true models[7]. He also applied this technique to the Cox-type N test, JA test, F test and adjusted Cox-type tests of Godfrey and Pesaran (1983) and noticed a substantial improvement in finite-sample sizes, when the errors are normal and non-normal (when errors are non-normal, the JA and F tests are no longer exact and are asymptotically valid only). Godfrey's power results indicate that the J and Godfrey and Pesaran's tests are equally powerful, whereas, the F and JA tests are less powerful. He applied the bootstrap method to the multiple non-nested alternatives case as well. Based on his results, he recommended the bootstrap based joint J test over a joint F test for testing multiple alternatives. The remarkable performance of the bootstrap approach led Godfrey to comment that the bootstrap

[7] Bootstrap is a data based simulation method which is useful to approximate an unknown sampling distribution (see Horowitz (1997)).

samples could be used to make more reliable inferences from diagnostic checks as well. Fan and Li (1995) and Davidson and MacKinnon (1996) also successfully applied the bootstrap method to the J test.

Walker (1967) was the first to introduce a test based on Cox's principle for testing non-nested time series models, especially AR(p) against MA(q) processes. However, his test turns out to be unattractive in terms of computation, particularly when both p and q exceed one. Godfrey and Tremayne (1988) proposed pure significance (PS) tests for testing AR(1) against MA(1) models. These tests are asymptotically valid and are easy to calculate but can have higher than nominal sizes in finite samples. Observing this, Smith and Tremayne (1990) introduced various correction factors to the PS tests to ensure that sizes of the corrected PS tests are reasonably close to the nominal size, in the context of testing AR(1) against MA(1) models. Burke et al. (1990) introduced an easy to implement PS test to the problem of testing for AR(1) errors against MA(1) errors in the linear regression model. Some other studies that considered the problem of testing between AR and MA models are McAleer et al. (1988), Hall and McAleer (1989), Godfrey and Tremayne (1992), and Franses (1992).

Baltagi and Li (1995) derived some modifications to the PS test of Burke et al. (1990) and applied their modified test to the problem of testing for AR(1) errors against MA(1) errors (and vice-versa) in an error components model. Their Monte Carlo results suggest that the PS test for testing AR(1) against MA(1) errors is trustworthy only when the sample size, $n \geq 60$ and n is greater than the number of individuals. However, the PS test for MA(1) errors against AR(1) errors performs well when the number of individuals is large and does not rely on n to achieve its asymptotic distribution.

McKenzie et al. (1999) developed some simple (large-sample based) separate (or non-nested) tests for testing AR(p) against MA(q) errors (and vice-versa) in the linear regression model. After estimating both the null and alternative non-nested models by maximum likelihood methods, their testing procedure involves testing the significance of variables added to a linearised version of the null model, the added variables being the predictions, or the residuals from the specified alternative model,

or the difference of the predictions of the two models[8]. For the simpler case of testing for AR(1) against MA(1) errors (and vice-versa), they compared the finite-sample properties of their tests with those of the PS tests of Burke et al. (1990) and LM tests of AR(1) against ARMA(1,1) errors. Their Monte Carlo results, supported the prediction tests which use information from the alternative hypothesis. The LM tests (which use information about the null model only) are the second best and PS tests perform the worst.

Mizon and Richard (1986) proposed two encompassing testing procedures, namely, the Wald encompassing test (WET) and the score encompassing test (SET), for testing encompassing hypotheses (see Gourieroux and Monfort (1995) and references therein). The encompassing approach involves asking whether the null model can explain one or more features of the rival alternative model. When all the features of the alternative model can be explained by the null model, it is said that the latter model encompasses the former model. Similarly, the alternative model can also encompass the null model[9]. McAleer and Pesaran (1986) noted that encompassing tests are also straightforward applications of Cox's principle. Both encompassing tests are harder to implement because the derivations involved in the test statistics are not always easy to evaluate.

2.4 Comparing PO (APO) Tests of a Composite Null Hypothesis with Other Tests

PO and APO tests have been applied to both nested and non-nested testing problems, which are discussed separately below. Also, for each category (nested and non-nested), we explain how the PO test and APO test can be constructed for a selected testing problem.

[8] Thus, separate tests are also known as prediction tests.
[9] For a formal definition of encompassing, see Gourieroux and Monfort (1995).

2.4.1 Comparing Tests for Composite Non-Nested Problems

Because PO tests have been applied more to the problem of testing for AR(1) errors against MA(1) errors in the linear regression model, we explain how to construct the PO test for this popular problem. Consider the linear regression model,

$$y = X\beta + u, \tag{2.4.1}$$

where y is an $n\times 1$ vector, X is an $n\times k$ nonstochastic matrix of rank $k<n$, β is a $k\times 1$ parameter vector and u is the $n\times 1$ vector of disturbances.

Suppose we wish to test the null hypothesis that the elements of u are generated by the stationary AR(1) process,

$$H_0: u_t = \rho u_{t-1} + e_t, \quad 0 \le \rho < 1, \tag{2.4.2}$$

against the alternative hypothesis that the elements are generated by the MA(1) process,

$$H_a: u_t = e_t + \gamma e_{t-1}, \quad 0 < \gamma \le 1, \tag{2.4.3}$$

where, $e = (e_0, e_1, \ldots, e_n)' \sim N(0,\sigma^2 I_{n+1})$. These hypotheses can be rewritten in matrix form as

$$H_0: u \sim N(0,\sigma^2\Sigma(\rho)), \quad 0 \le \rho < 1, \tag{2.4.4}$$

against

$$H_a: u \sim N(0,\sigma^2\Omega(\gamma)), \quad 0 < \gamma \le 1, \tag{2.4.5}$$

where $\Sigma(\rho)$ is an $n \times n$ matrix whose $(i,j)th$ element is $\rho^{|i-j|}/(1-\rho^2)$ and $\Omega(\gamma)$ is the $n \times n$ tridiagonal matrix with $1+\gamma^2$ as the main diagonal elements and γ as the nonzero off-diagonal elements. For this problem, the nuisance parameters β and σ^2 can be avoided by using the invariance method (explained in the next chapter).

The construction of King's point optimal invariant (POI) test[10] for testing (2.4.4) against (2.4.5) involves first pre-specifying values for the autocorrelation parameters ρ and γ. Let these values be ρ_0 and γ_0, respectively, and the corresponding covariance matrices under the null and alternative hypotheses be $\sigma^2\Sigma(\rho_0)$ and $\sigma^2\Omega(\gamma_0)$, respectively. Here γ_0 is the point under the alternative hypothesis at which we wish to optimize power while ρ_0 is an arbitrary value of ρ under H_0. Now the FNPL provides a POI test which possesses optimal power at γ_0. The critical regions of such a test can be written as

$$s(\gamma_0, \rho_0) = \hat{u}'\Omega(\gamma_0)^{-1}\hat{u} / \tilde{u}'\Sigma(\rho_0)^{-1}\tilde{u} < c, \tag{2.4.6}$$

where $\tilde{u}$ and $\hat{u}$ are the generalised least squares (GLS) residual vectors from (2.4.1) for covariance matrices $\Sigma(\rho_0)$ and $\Omega(\gamma_0)$, respectively.

King explored whether the $s(\gamma_0, \rho_0)$ test still optimizes power at γ_0 when $\Sigma(\rho_0)$ is broadened to $\Sigma(\rho)$. He observed that this could happen if the critical value c and the parameter ρ_0 can be chosen such that

$$\Pr[s(\gamma_0, \rho_0) < c | u \sim N(0, \sigma^2\Sigma(\rho_0))] = \alpha \tag{2.4.7}$$

and

$$\Pr[s(\gamma_0, \rho_0) < c | u \sim N(0, \Sigma(\rho)), 0 \le \rho \le 0.999] \le \alpha, \tag{2.4.8}$$

[10] If a test achieves optimum power at a particular point within the class of invariant tests then it is called a POI test.

hold simultaneously, where α is the desired level of significance. Here, we approximated $0 \leq \rho < 1$ with $0 \leq \rho \leq 0.999$ for practical reasons. The LHS of (2.4.8) (of which (2.4.7) is a special case) can be evaluated using standard numerical methods such as Koerts and Abrahamse's (1969) or Davies' (1980) versions of Imhof's (1961) algorithm. Alternatively, Palm and Sneek's (1984) and Shively et al.'s (1990) procedures can be used (for more detail see King (1987b)).

If c and ρ_0 cannot be chosen to solve (2.4.7) and (2.4.8) together, as mentioned earlier, King's APOI test can be constructed. Such tests have critical regions of the form of (2.4.6) with c chosen such that (2.4.8) holds and ρ_0 chosen to make the left side of (2.4.7) as close to α as possible.

King suggested an iterative procedure to obtain appropriate ρ_0 and c values. He also provided solved ρ_0 and c values for an X matrix which can be regarded as a representative matrix for a number of matrices used in practice. Such values might be used to apply an approximate test, or they could be used as good starting values for the iterative process. In order to make the POI test (or APOI test) operational, the point at which power is to be optimized should be decided (see King (1987b) and Wu and King (1994) on this issue).

King (1983) originally considered an approximate version of the $s(\gamma_0, \rho_0)$ test which we call the pseudo POI (PPOI) test. The PPOI test is obtained by controlling the maximum probability of a type I error by ones choice of critical value alone. That is, c is determined by solving

$$\Pr[s(\rho_0, \gamma_0) < c | u \sim N(0, \sigma^2 \Sigma(\rho)] \leq \alpha . \tag{2.4.9}$$

King (1987b) compared his $s(\gamma_0, \rho_0)$ tests with PPOI tests and found that the former tests are preferable over the latter.

King and McAleer (1987) considered testing for AR(1) errors against MA(1) errors in the linear regression model using the Cox test, P test, prediction test, LM test of AR(1) disturbances against ARMA(1,1) disturbances, and PPOI tests. Their Monte

Carlo study revealed that the Cox test and its versions can have actual sizes much higher than the nominal size, while the LM test sizes are acceptable. On the other hand, (by construction) the PPOI test has sizes less than or equal to the nominal size. Therefore, in order to compare the powers at roughly the same significance levels, they followed two steps. The first step involves controlling the maximum size of an asymptotic test by ones choice of critical value (explained more in section 2.5.1) and the second step involves calculating the powers using that critical value (hereafter, tests based on this approach are called size corrected tests). Based on their empirical results, King and McAleer recommended the PPOI test and they also commended the LM test as having reasonable power properties even when it is applied for an inappropriate alternative.

When testing for AR(1) errors against MA(1) errors, King (1983) commented that any good test should have a size which tends to zero, asymptotically, for autocorrelation coefficients greater than 0.5. Burke et al. (1990) introduced an easy to implement PS test which satisfies King's suggestion, based on ordinary least squares (OLS) residuals. In particular, they compared their (size corrected) PS test results with those of King and McAleer (1987). Their study also supported the PPOI test while their new test seems to be slightly more powerful than the LM test. This is in contrast to McKenzie et al.'s (1999) finding that the LM test is more powerful than the PS test.

Silvapulle (1991, 1994b) considered testing for AR(1) errors against MA(1) errors in the first-order dynamic linear regression model, using POI, PS (denoted the τ test) and LM tests. Consider the first-order dynamic linear regression model

$$y_t = \mu y_{t-1} + x_t'\beta + u_t, \quad t = 1,\ldots,n, \tag{2.4.10}$$

where y_t is the dependent variable at time t, x_t is a $k \times 1$ vector of non-stochastic exogenous variables at time t, μ is an unknown scalar assumed to be such that $|\mu| < 1$[11], β is an unknown $k \times 1$ parameter vector and u_t is the disturbance term.

[11] μ is believed to be typically non-negative in economic applications.

Obviously, if μy_{t-1} is not present in (2.4.10), the testing problem can be simplified using invariance arguments. This influenced Silvapulle to estimate μ using the instrumental variable (IV) estimator and subtract $\hat{\mu} y_{-1}$ from y_t[12]. Thereafter, she applied a modified version of King's POI test, denoted the MPOI test, to the problem. She showed that the approximate critical value for the MPOI test can be obtained using an approximate small disturbance asymptotic (ASDA) distribution of the test statistic. Her ASDA distribution based critical value for the MPOI test statistic turns out to be the true critical value of the statistic for the corresponding regression with the lagged dependent variable omitted. This approach is analogous to the one adopted by Inder (1985) when testing for AR(1) disturbances in the dynamic linear regression model.

Silvapulle found that MPOI tests generally perform better than asymptotic tests in terms of size and power properties and the LM test is the second best[13]. However, her study is rather limited in the sense that she carried it out almost always in the positive nuisance parameter space. That is, she nearly always used a positive β vector and varied it by varying σ, which can take only positive values. Therefore, Silvapulle's ((1991), p.129) conclusion that, "ASDA distribution based MPOI tests perform equally well with choppy, random and economic data series as exogenous regressors. Thus, we are tempted to say that the ASDA distribution based critical values of the MPOI tests are good approximations to the non-similar critical values" is questionable. Also, her results indicate that the ASDA distribution based MPOI tests can have estimated sizes as high as 0.120 when the nominal size is 0.050.

Silvapulle and King (1993) considered testing for higher-order autocorrelations such as joint AR(1)-AR(4) disturbances against joint MA(1)-MA(4) disturbances in the linear regression model, using POI tests, the LM test of restricted AR(5) errors against restricted ARMA(5,5) errors and the prediction test. Because the asymptotic tests are observed to have significantly higher than nominal sizes, they used size corrected tests. Their empirical power results favoured the POI tests when positive first-order and fourth-order autocorrelations are considered while they favoured the asymptotic

[12] Because the OLS estimator is inconsistent under the null, the IV estimator is used.
[13] She did not size correct the tests in this case.

tests when wrongly assuming positive autocorrelation in the presence of mild negative autocorrelation. On many occasions, the POI tests are found to possess more than double the power of their competitors. In addition, the POI tests are the only ones to have the desirable property that the test's size tends to zero as the sample size increases whenever the AR(1)-AR(4) disturbance process has either a first-order or fourth-order autocorrelation coefficient greater than 0.5 (see King (1983) for more detail). Based on their results, Silvapulle and King commented that the extra computation required to perform a POI is worthwhile. Their study encourages the use of POI tests for testing higher-order autocorrelations.

Silvapulle and King (1991) and Silvapulle (1994a) encountered two situations where the POI tests cannot be constructed, when testing for MA(1) errors against AR(1) errors and testing for AR(1) errors against IMA(1,1) errors in the linear regression model. For the latter problem, Silvapulle was able to construct a PO test (which excelled in its performance) when the MA coefficient is positive but not when it is negative. For the cases where POI tests cannot be constructed, they constructed APOI tests and compared their performance with those for the LM, PS and prediction tests. The APOI tests worked extremely well for the former problem in terms of size and power properties, but performed poorly for the latter problem. In particular, for the latter problem, the APOI test sizes are generally closer to zero in absolute value and powers are less than or equal to the nominal size. We note that, for the former problem, the probability of the form (2.2.7) is closer to zero, while this is not true for the latter problem. This suggests that the APOI test may work well if it is nearly a PO test, otherwise it is not always reliable.

2.4.2 Comparing Tests for Composite Nested Problems

King (1989) applied the APOI test to the problem of testing for fourth-order autoregressive (AR(4)) disturbances in the linear regression model in the presence of AR(1) disturbances. For this case, the POI test (and APOI test) can be constructed as follows. Consider the linear regression model (2.4.1). Suppose the error term u is generated by a joint AR(1) and simple AR(4) process,

$$(1 - \rho_1 L)(1 - \rho_4 L^4)\, u_t = e_t, \tag{2.4.11}$$

in which $0 \leq \rho_4 < 1$ and $0 \leq \rho_1 < 1$ are unknown parameters, L is the lag operator such that $Lu_t = u_{t-1}$, and $e = (e_1, \ldots, e_n)' \sim N(0, \sigma^2 I_n)$. Note that (2.4.11) is the AR(5) process,

$$u_t = \rho_1 u_{t-1} + \rho_4 u_{t-4} - \rho_1 \rho_4 u_{t-5} + e_t. \tag{2.4.12}$$

Suppose we wish to test

$$H_0: \rho_4 = 0 \text{ against } H_a: \rho_4 > 0. \tag{2.4.13}$$

This testing problem can be simplified by using invariance arguments. Now King's POI test and APOI test can be constructed as outlined in section 2.4.1. This involves first constructing the test for the simpler problem of testing

$$H_0^1: (\rho_1, \rho_4)' = (\rho_{10}, 0)' \tag{2.4.14}$$

against

$$H_a^1: (\rho_1, \rho_4)' = (\rho_{11}, \rho_{41})' \tag{2.4.15}$$

where $0 \leq \rho_{10} \leq 0.999$, $0 \leq \rho_{11} \leq 0.999$ and $0 < \rho_{41} < 1$ are known and fixed. Let $\Omega_1(\rho_1, 0)$ and $\Omega_2(\rho_1, \rho_4)$ be disturbance covariance matrices under the null and alternative hypotheses, respectively, and $\Delta_1 = \Omega_1(\rho_{10}, 0)$ and $\Delta_2 = \Omega_2(\rho_{11}, \rho_{41})$. Now the FNPL provides the most powerful test based on critical region of the form

$$s(\rho_{10}, \rho_{11}, \rho_{41}) = \hat{u}'\Delta_2^{-1}\hat{u} / \tilde{u}'\Delta_1^{-1}\tilde{u} < c, \tag{2.4.16}$$

where $\hat{u}$ and $\tilde{u}$ are the GLS residual vectors assuming disturbance covariance matrices Δ_2 and Δ_1, respectively. For the existence of a POI test (as explained

earlier) we have to choose the critical value and ρ_{10} such that the following probabilities hold simultaneously.

$$\Pr[s(\rho_{10},\rho_{11},\rho_{41}) < c | u \sim N(0,\Omega_1(\rho_{10},0))] = \alpha \tag{2.4.17}$$

and

$$\Pr[s(\rho_{10},\rho_{11},\rho_{41}) < c | u \sim N(0,\Omega_1(\rho_1,0))] \leq \alpha . \tag{2.4.18}$$

If such ρ_{10} and c cannot be chosen then the APOI test can be obtained by searching for a ρ_{10} value which minimises

$$\alpha - \Pr[s(\rho_{10},\rho_{11},\rho_{41}) < c | u \sim N(0,\Omega_1(\rho_{10},0))] \tag{2.4.19}$$

subject to (2.4.18) holding. By construction, the POI (and APOI) tests have sizes less than or equal α for all the ρ_1 values, therefore, it is possible that their powers at some points under the alternative hypothesis will be less than their nominal size. For more information about how to select ρ_{10}, ρ_{11} and ρ_{41}, see King (1989). King considered three versions of the APOI test corresponding to three different choices of these parameters. These parameter choices are either arbitrary or satisfy some optimality criteria associated with power. Based on his empirical results, he recommended that all values of test parameters be chosen by optimality criteria associated with power. One of the APOI tests he considered is not nearly optimal and was found to be poor at optimizing power.

As mentioned earlier, existing studies show that marginal likelihood based tests perform better in finite samples. We have seen that the PO tests also have excellent finite-sample properties. Rahman and King (1994) explored which of these approaches is best in finite samples. They considered testing random regression coefficients in the presence of AR(1) errors, using APOI tests and the marginal likelihood based LM and asymptotically locally most mean powerful (ALMMP)

tests[14]. Surprisingly, they concluded that the extra computation required for the APOI tests hardly seems worthwhile. According to their empirical results, neither of the two asymptotic tests dominates the other in terms of power properties. The ALMMP test is best on average but for certain data sets in which the component scores used in the test statistic are negatively correlated, this test performs poorly. Though, the APOI tests are sometimes less powerful than their competitors, there are cases where the APOI tests dominate the asymptotic tests, particularly when their sizes are much lower than their competitors. Also, as Rahman and King noted, one of the required probabilities for the existence of an APOI test (similar to (2.4.19)) is not properly satisfied in the sense that it is typically significantly different from zero. Therefore, these APOI tests may not nearly optimize power at the chosen points under the alternative hypothesis which may be the reason for their inferior properties. Thus, an appropriate choice of APOI test could have altered Rahman and King's conclusion.

2.5 Research Findings Related to Our Study

Griliches (1967) considered an interesting testing problem, namely, testing for a static linear regression model with AR(1) errors against a dynamic linear regression model with white noise errors and suggested an ad hoc rule-of-thumb to discriminate between these two models. The null and the alternative models he considered can be written as

$$H_0: y_t = x_t'\beta + u_t, \quad u_t = \rho u_{t-1} + e_t, \ t = 1,\ldots,n, \tag{2.5.1}$$

and

$$H_a: y_t = \mu y_{t-1} + x_t'\beta + e_t, \quad e_t \sim N(0,\sigma^2), \ t = 2,\ldots,n, \tag{2.5.2}$$

[14] King and Wu (1997) introduced the ALMMP test.

respectively, where, y_t is the dependent variable at time t, x_t is a $k \times 1$ vector of non-stochastic regressors at time t, β is a $k \times 1$ vector of coefficients, $0 \leq \mu < 1$ and u_t is the disturbance term with $|\rho| < 1$ and $e_t \sim N(0, \sigma^2)$.

Griliches observed that if the true model is (2.5.1) and if (2.5.2) is fitted to the sample by OLS, it is likely to explain the data quite well. Therefore, one may mistakenly assume that the sample was generated by (2.5.2) which will lead to misleading inferences and sub-optimal forecasts. For example, King and Rankin (1993) (in the context of forecasting alone) commented that when the DW test is significant and the dynamic parameter of the dynamic linear model is large, a substantial loss in accuracy of prediction can occur if one proceeds to correct for AR(1) disturbances in a static linear model without first checking for a possibility of a dynamic linear model with white noise errors.

Observe that the null model can also be written as

$$y_t = \rho y_{t-1} + x_t'\beta - \rho x_{t-1}'\beta + e_t. \tag{2.5.3}$$

Model (2.5.3) is similar to (2.5.2), except for the regressor x_{t-1}. Obviously, these two models are not distinguishable if the regressor x_t is lag invariant (that is, the lag of the exogenous regressors added provides no extra explanation of the dependent variable). Similarly, it can be shown that these models are not distinguishable if the exogenous regressors are a constant and/or time trend only (see King (2000)). Also, the null and alternative models may not be distinguished at all if exogenous regressors are not present. Thus, this testing problem is a complicated one.

Griliches' ad hoc rule-of-thumb for discriminating between (2.5.1) and (2.5.2) is as follows. He suggested that after (2.5.2) has been estimated by OLS one should also estimate

$$y_t = \eta_1 x_t + \eta_2 y_{t-1} + \eta_3 x_{t-1} + e_t. \tag{2.5.4}$$

If $\hat{\eta}_3$ is significant and close to $-(\hat{\eta}_1 \times \hat{\eta}_2)$, then one can conclude that the sample was generated by (2.5.3) and not by (2.5.2). However, Griliches did not quantify the expression "close to". Griliches also suggested that if his proposal is followed and $\hat{\eta}_3$ is significantly positive, the true model is a generalised Koyck model. Giles (1975) noted some practical difficulties associated with this suggestion and suggested one should consider the possibility of a generalised Koyck model even if $\hat{\eta}_3$ is negative.

Maddala (1971) suggested a Bayesian posterior odds method to discriminate between the null and alternative models of interest. Giles (1975) tested (2.5.1) against (2.5.2) using Griliches' ad hoc rule-of-thumb and Maddala's Bayesian posterior odds analysis. Based on his empirical results, he concluded that the Bayesian analysis is generally superior to the ad hoc rule-of-thumb. To the best of the author's knowledge, other than Griliches (1967), Giles (1975) and Maddala (1971), so far no one has considered this complicated testing problem.

2.5.1 Non-Similar Tests

The problem with non-similar tests is that their sizes are a function of nuisance (or unknown) parameters. Consequently, it is impossible to obtain exact size critical values for which the size always equals the significance level. As noted in Chapter 1, finding exact non-similar critical values of these tests is the only promising way to proceed in this situation. Because it is extremely hard to obtain exact non-similar critical values, researchers nearly always use approximate non-similar critical values.

Inder (1985) considered testing for autocorrelated disturbances (as in (2.4.2)) in the first-order dynamic linear regression model (as in (2.4.10)). Because this testing problem is of interest in this thesis, we discuss it in detail. It is well known that, in the linear regression model with a lagged dependent variable as regressor, the assumption of uncorrelated disturbances is essential for the properties of the structural parameter estimates. If the disturbances are autocorrelated and the parameters are estimated by OLS, the estimates will be inconsistent, and may lead to misleading inferences.

Therefore, it is important to have a powerful testing procedure to detect the presence of autocorrelation.

In their seminal paper, Durbin and Watson (1950, 1951) emphasised that their DW test (which is specially designed for static linear regression models) is not applicable for models including lagged dependent variables such as the dynamic linear regression model. This is mainly because of the lack of appropriate critical values for such models. However, researchers indiscriminately used the DW test in the context of the dynamic linear regression model using upper and/or lower bounds of the critical values (intended for static linear regression models) as actual critical values. Not surprisingly this received much criticism because at the time, there was no theoretical basis for using such critical values and because of inconsistent test performance.

Later, Durbin (1970) proposed two large-sample based tests, namely, Durbin's t test and h test for models including lagged dependent variables. After carefully studying the literature on autocorrelation and dynamic linear regression models and also through Monte Carlo experiments, Inder (1985) realised that if appropriate critical values could be found then the DW test would be the preferred candidate compared to Durbin's asymptotic tests. This motivated him to derive the ASDA distribution of the DW statistic in the context of the first-order dynamic linear regression model. His ASDA distribution based critical value turns out to be the exact DW critical value for the dynamic linear regression model when the lagged dependent variable is omitted as a regressor. Therefore, his work allows one to use existing computer algorithms for calculating DW critical values or tables of bounds or further approximations appropriate for the static model.

Observing that the testing problem is not invariant to β (except for the intercept), but is invariant to rescaling such that the ratio β/σ is preserved, Inder also obtained approximate non-similar critical values by letting $\beta = 0$ and experimenting with μ values until the smallest (and largest) critical value for the DW (and Durbin's asymptotic tests) is found. In particular, he compared the performance of the DW test and Durbin's asymptotic tests based on exact size critical values (obtained via the

Monte Carlo method assuming knowledge of the nuisance parameters), approximate non-similar critical values and asymptotic critical values and found that the DW test is always the best. He also commented that the ASDA distribution based DW test performs much better than do Durbin's asymptotic tests, under the null. His empirical results indicate that the use of approximate non-similar critical values can often produce sizes well below the nominal size for all three tests. For example, when the nominal size is 0.050, sizes of the tests are frequently less than 0.010 or equal to zero. Thus, the use of approximate non-similar critical values does not lead to ideal sizes. Also, because Inder considered a limited number of nuisance parameter values in his study[15], it is not clear whether his approximate non-similar critical values are adequate enough to control the sizes of the tests over the nuisance parameter space.

After Inder's study, there seemed to be an awakening of interest in applying the DW test to dynamic linear regression models (see Grant (1987), King and Wu (1991), Rayner (1993), King and Harris (1995) and Dezhbakhsh and Thursby (1995))[16]. King and Wu (1991) and King and Harris (1995) extended Inder's study for the dynamic linear model. In particular, King and Wu (1991) observed that the exact small disturbance asymptotic distribution of the DW statistic is equivalent to the distribution of the DW statistic from the regression model with the lagged dependent variables replaced by their means (i.e., by their expected values). A shortcoming of their approach is that the expected values of the lagged dependent variables are unknown because they are functions of unknown regression coefficients. They discussed how bounds for the small disturbance critical value could be calculated. Therefore, their study gives a justification for the use of the familiar tables of bounds when applying the DW test to the dynamic linear regression model. An alternative approach would be to estimate the expected values of the variables. King and Harris (1995) adopted this approach and treated the lagged dependent variables as estimates of their means. That is, their approach involves calculating the exact critical values of the DW test with the lagged dependent variables treated as non-stochastic. Their Monte Carlo results generally support their new approach over Inder's (1985) ASDA based DW, modified PO tests and Durbin's asymptotic tests, particularly when two lags of the

[15] He used one β vector and varied it by varying σ.

[16] Grant's study is similar to Inder's.

dependent variable are present. They also found that the sizes of all five tests they considered are relatively robust to non-normality in the disturbances.

In the case of testing AR(1) errors against MA(1) errors in the static linear regression model, King and McAleer (1987) observed that the null distributions of the asymptotic tests they considered depend on ρ. Therefore, in order to compare the powers at roughly the same significance levels, they used approximate non-similar critical values for their tests. Their approach involves finding the exact critical value of an asymptotic test (via the Monte Carlo method) assuming AR(1) errors with a fixed ρ value. This is done for $\rho = 0, 0.1, \ldots, 0.9$. From these critical values, they choose the largest (absolute) value as their approximate non-similar critical value. This choice ensures that, at least for the above ρ values, the size of the test is not greater than the desired significance level. Burke et al. (1990), Silvapulle and King (1991) and Rahman and King (1998) also used this kind of approximate non-similar critical values.

Silvapulle and King (1993) (see section 2.4.1), estimated the asymptotic tests' sizes at 25 grid points, defined as $\{(\rho_1, \rho_4): \rho_1, \rho_4 = 0.1, 0.3, 0.5, 0.7, 0.9\}$. For each test, the largest of the 25 critical values was taken as the appropriate critical value, thus ensuring that, at least for these grid-points, the size of the test does not exceed the nominal size.

Grose (1998) considered testing the partial adjustment (or dynamic regressor) coefficient in the first-order dynamic linear regression model with white noise errors. In particular, she compared marginal likelihood based classical tests with conventional likelihood based counterparts. Observing that this testing problem is not invariant to β (except for the intercept), Grose obtained approximately exact non-similar critical values by doing simulations across a range of β values. For this, she considered nine β vectors, namely, (0,0,0), (0,0,1), (0,1,1), (0,1,0), (0,1,-1), (0,0,-1), (0,-1,-1), (0,-1,0) and (0,-1,1) and made β larger or smaller along each of eight directions by decreasing or increasing σ. Her approximately exact non-similar critical value of a test is the largest critical value found for all (β, σ) combinations.

Silvapulle and King's (1993) and Grose's (1998) studies indicate that obtaining approximate non-similar critical values can be quite demanding when the null distribution of the test statistic depends on more than one nuisance parameter.

2.5.2 The Simulated Annealing (SA) Algorithm

The SA algorithm is used for all the optimizations carried out in this thesis. SA is a global optimizer which is explicitly designed for optimizing functions with multiple optima. It was first introduced in thermodynamics, where one studies a system's thermal energy. To achieve a low energy state of a molten metal, it is first heated and then slowly cooled (called annealing)[17]. If the cooling is done very quickly the metal might not escape the local energy minimum and when fully cooled it may contain more energy than the desired level. This cooling schedule motivates the concept of the SA algorithm. That is, SA tries to minimize some analogue of energy in a manner similar to annealing, in order to find the global minimum. Thus, for function optimization, the objective function to be minimized corresponds to the energy of the states of the solid.

SA explores the function's entire surface and tries to optimize the function by doing both uphill and downhill moves[18]. Thus, unlike the conventional algorithms, SA is largely independent of the starting values. Also, SA assumes very little about the shape of the function to be optimized. This can be considered as another advantage of SA, because, almost all the conventional algorithms explicitly assume that the function has one optimum and often assume that the function is approximately quadratic. Thus, when faced with a function with multiple optima or a difficult function not approximately quadratic, conventional algorithms may fail. Also, because conventional algorithms assume that the function has one optimum, it is possible for them to find any local maximum as the global maximum. SA can also optimize functions that are not defined for some parameter values. Therefore, as

[17] Slowly cooling means every time the rate of heating is reduced.
[18] Conventional algorithms do not explore the entire surface, instead they head up (or down) the hill much as a blind man would (see Goffe et al. (1992)).

Goffe et al. (1992, 1994) point out, SA can easily identify corner solutions because it can snuggle up to a corner for functions that don't exist in some regions. Finally, SA can optimize functions which need not even be differentiable (see Corana et al. (1987)). All these observations led Goffe (1996) to comment that the SA algorithm is much more robust than conventional algorithms. The only undesirable aspect of SA is the greater run time it takes. Because it is more reliable than the conventional algorithms, this problem has to be stomached. Also, in an era of cheap, and increasingly cheaper computing, this may become a minor problem.

Kirkpatrick et al. (1983), and Cerny (1985) are the first to show how Metropolis et al.'s (1953) model for simulating the annealing of solids (discussed above), could be used for optimisation problems. After this, SA received much attention and has been applied to many optimisation problems occurring in areas such as statistics, computer design, image processing, molecular physics and chemistry. Some examples of applications in statistics are Corana et al. (1987), Aarts and Laarhoven (1989), Eglese (1990), Brooks and Morgan (1995), Goffe (1996) and Ali et al. (1997).

Goffe et al. (1992, 1994) introduced SA to optimize some econometric functions with great success. Goffe et al. (1992) minimised a function with two local minima, using conventional algorithms, namely, conjugate gradient, quasi-Newton and simplex methods, and SA. All algorithms were run 100 times with different starting values for the model parameters (same ones were used by all algorithms). We reproduce their results in Table 2.1, below.

Table 2.1: Goffe et al.'s comparison of four algorithms giving number of times different maxima were obtained and average time taken for 100 runs

Algorithm	Simplex	Conj. gradient	Quasi-Newton	SA
Solutions*				
20.482	40	48	48	0
16.082	60	52	52	100
Time (seconds)	0.019678	0.004083	0.006829	16.328

*Solutions are categorised by the minimum at which they terminate. The minimum with the value of 16.082 is the global minimum.

Table 2.1 shows that the conventional algorithms converge nearly half the time to the local minimum, 20.482, whereas SA always converges to the global minimum, 16.082. This example clearly shows that (for functions with multiple optima) conventional algorithms are not efficient enough to distinguish between the local and global optima. Though SA is much more efficient in this regard, it takes greater run time than its competitors[19]. This led Goffe et al. (1992, p.141) to conclude, "Though this algorithm is considerably more expensive in computational resource, its benefits in estimation of functions with multiple optima are quite obvious. Thus, its use should be considered for difficult econometric problems".

Goffe et al. (1994) compared the same algorithms (as above) for more complicated econometric functions. The first two functions they considered are, the rational expectations version of the monetary theory of exchange rate determination (minimization of which involves 14 parameters), firm production efficiency based on a system of a frontier cost function and its input share equations (minimization of which involves 62 parameters) and the third function comes from the neural network literature (minimization of which involves 35 parameters). The first function is difficult to minimize because it effectively does not exist for some parameter values. In these regions, the function value is either complex or the elements of the covariance matrix go to infinity. For such regions, they set the objective function value to about 10^{2000} to force termination of the conventional algorithms. Of the 100 runs, with the simplex and quasi-Newton algorithms, about half are observed to have terminated due to floating point errors, which indicates that the algorithms are beyond the likely region of a solution because large numbers are needed to cause a floating point error on the computer they used. The conjugate gradient algorithm performed better in this regard. However, all the conventional algorithms fail to find the global minimum.

SA also experienced some difficulty with this function. It converged to different optima for different starting values and seed values for the uniform random number generator of SA (explained later). Goffe et al. modified the function to search a

[19] The much longer time for SA is largely due to using Corana et al.'s (1987) very conservative suggestions for parameters, which are appropriate for very complicated functions. Even if these parameters are reduced, SA will take more time than conventional algorithms.

restricted region of the parameter space to find an explanation. Once SA detected the problem for this small region they enlarged the region. SA then found that the optimal value of one parameter (the interest rate elasticity of money demand) often occurs at its upper bound. This suggested that the function is decreasing in this parameter. Goffe et al. plotted several of these minima with the elasticity varying and the other parameters held constant. These plots also showed that the interest rate elasticity parameter achieved a minimum at the boundary. Thus, the function appeared to be a ditch in this parameter: as the boundary expanded, the minimum point of the function is observed to follow this wandering ditch. This might be the reason why the conventional algorithms could not minimize the function. Because SA was able to find an explanation, Goffe et al. recommended its use as a diagnostic tool for difficult functions. In order to continue the comparison of the algorithms, Goffe et al. fixed the interest rate elasticity parameter to 0.25 and repeated the experiments. Their results show that SA is much more consistent in finding the global minimum (it was 3 out of 3, while the conventional algorithms were, with generous accounting, 64 out of 300). This led Goffe et al. (1994, p 82) to comment, "While a single run of SA requires substantially greater execution time, this is ameliorated by the large number of runs with a conventional algorithm a researcher would have to make to be sure of the robustness of estimated results".

The second function they considered has steep valleys and the third function has many local minima. None of the conventional algorithms were able to optimize the second function whereas SA was able to find the global minimum. Because of its nature, it is virtually impossible to optimize the third function. For this function, SA was able to find a much better optimum than its competitors. These findings indicate that SA can optimize functions that conventional algorithms have extreme difficulty with or simply cannot optimize at all. Thus, the SA algorithm has many useful features not shared by conventional algorithms. Fortunately, GAUSS and FORTRAN programming codes for the SA algorithm are available nowadays (see Goffe (1996) and Tsionas (1995)).

Aarts and Laarhoven (1989) derived necessary and sufficient conditions to ensure that asymptotically, the SA algorithm finds a globally optimal solution with probability 1. However, they observed that these conditions cannot be satisfied in finite time, so

they advised one to find appropriate values for the parameters of the algorithm such that in finite time, near optimal solutions are returned.

The essential starting parameters involved in minimizing a function $f(x)$ are T_0, the initial temperature, x_0, the starting vector of parameters, and v_m, the step length for x_0. Here x_0 and v_m are both vectors of length a^*, the number of parameters of the model. Because v_m quickly resets from its initial value, Tsionas (1995) observed that the input of v_m is not very important. v_m gives the researchers valuable information about the function. If an element of v_m is very large, it indicates that the function is very flat in that parameter. The steps involved in minimising a function f via SA are as follows (see Goffe et al. (1994) and Brooks and Morgan (1995)).

(1) SA begins with an initial temperature T_0, and computes the function value f_0 for the initial parameter values x_0.

(2) SA randomly selects another point x in the parameter space (within the neighbourhood of the initial parameters and within the bounds set for it by the problem of interest) and calculates the corresponding function value f. Each element of x is obtained as follows.

$$x_i = x_{0,i} + a' v_{m,i},$$

where a' is a uniformly distributed random number from $[-1, 1]$ and $v_{m,i}$ is the ith element of v_m.

(3) Then SA compares the two points in terms of their function values. If f is less than f_0, then x is accepted and x_0 is set to x and the algorithm moves downhill. If this is the smallest f, it and the corresponding x are recorded as the best current value of the optimum.

If f is greater than f_0 then SA accepts x based on a criteria, known as the Metropolis criteria, as follows. Let $f^* = f - f_0$. x is accepted and SA moves uphill if and only if a random variable b', distributed uniformly over [0,1], satisfies

$$b' \leq \exp(-f^* / T)$$

where T is the current temperature. If x is accepted, as before, x_0 is updated with x and so is the corresponding function value.

(4) After repeating steps 2 - 3 N_s times, the step length vector v_m is adjusted so that 50% of all moves are accepted, where N_s is the number of cycles provided by the user. This makes SA sample values of the function widely.

(5) After repeating steps 2 - 4 N_t times, the temperature T is reduced by a factor called r_t, where, $0 \leq r_t \leq 1$. Usually r_t will be 0.5 or 0.85 (see Tsionas (1995)). The process then begins again from step 2 (taking as the initial state the point following the last iteration of the algorithm). Termination of the algorithm occurs when the optimum function value reaches a stable state (for more detail see Tsionas (1995) and Brooks and Morgan (1995)).

A lower temperature makes an uphill move less likely, therefore, the number of rejections increase and v_m declines. The smaller steps and starting at the current optimum focuses attention on the most promising area. The key parameters of SA are T, N_s, N_t and r_t. Among these, r_t and N_t greatly influence the robustness of the algorithm and number of function evaluations since they control how quickly the temperature T declines and the number of function evaluations performed at each temperature. Goffe et al. (1994) give some useful hints for determining these values. Their suggestion is to obtain the optimum results assuming small parameter values for r_t and N_t, and providing a seed value for the uniform random number generator of SA and starting values for the parameters. Thereafter, they recommend increasing r_t and N_t, and getting the optimum results once again for a different uniform random

number generator seed value and starting values (hence dealing with an entirely different sequence of sampled points). If the current optimum results are the same as before then according to Goffe et al., the global optimum is achieved, therefore, one can proceed with small r_t and N_t parameter values. If this is not the case, then it is an indication that small SA parameters are not adequate for the optimization considered and they should be increased.

2.6 Conclusion

In this chapter, we focused on contributions to PO testing in the context of composite hypotheses testing. Because all the applications considered in this thesis are non-nested, we briefly discussed some popular non-nested tests as well. Other studies relevant to this thesis were also reviewed.

King's PO tests have mainly been applied to composite non-nested testing problems rather than to composite nested testing problems. These studies show that PO tests have excellent finite-sample properties compared to existing popular non-nested tests. However, King's APO test does not seem to always be reliable. These tests appear to perform well when they are nearly optimal, otherwise there is a question mark about their reliability. For a composite non-nested testing situation, Silvapulle (1994a) preferred some asymptotic tests over an APOI test (which was not nearly optimal). Similarly, for a composite nested testing situation, Rahman and King (1994) preferred some marginal likelihood based asymptotic tests over APOI tests (which were not nearly optimal). Thus, it is important to have a reliable APO test for testing composite hypotheses. Also, it is interesting to see whether the marginal likelihood based tests still dominate an appropriate APO test in finite samples.

Because it is extremely hard to obtain exact non-similar critical values, researchers nearly always use approximate non-similar critical values in their studies. We observed that the use of approximate non-similar critical values can produce sizes well below the nominal size. Also, obtaining such critical values becomes extremely

difficult when the null distribution of the test statistics depends on more than one unknown parameter.

We observed that the SA algorithm has several advantages over conventional algorithms. In addition, it can be used as a diagnostic tool to understand why conventional algorithms fail to optimize a certain function. Goffe et al.'s (1992, 1994) studies show that SA can optimize functions that conventional algorithms have extreme difficulty with or simply cannot optimize at all. Therefore, SA can be very useful for difficult econometric problems.

CHAPTER 3

THE g TEST: AN APPROXIMATE POINT OPTIMAL TEST OF A COMPOSITE NULL HYPOTHESIS

3.1 Introduction

As discussed in Chapter 1, in the absence of UMP tests, King's (1987b) PO tests can be useful. Unfortunately, PO tests cannot always be constructed when testing a composite null hypothesis. King suggested an APO test for situations where his PO test cannot be constructed. It seems that King's APO tests are not always reliable. Therefore, it would be desirable to have another APO test for situations where the PO tests cannot be constructed. In this chapter, we propose an APO test, called the g test based on the GNPL for testing a composite null hypothesis.

Suppose the observed sample is generated by a finite number of densities under the null hypothesis. The GNPL provides a PO test for the problem of testing for a finite number of observable density functions against a single alternative density function. Now, suppose the observed sample is potentially generated by one of an infinite number of densities under the null hypothesis, the question is, can we approximate the infinite number of densities by a finite number of densities and then apply the GNPL to obtain a PO solution? The g test is based on this idea. Our question then becomes

can this type of approximation based g test work well in practice? This chapter seeks an answer to this question.

In this chapter, we construct the g test and apply it to two testing problems, namely, testing for MA(1) errors against AR(1) errors and testing for AR(1) errors against (non-stationary) IMA(1,1) errors with a negative MA coefficient in the linear regression model. Silvapulle (1991, 1994a) considered these two testing problems and recommended APOI tests for the former problem and some asymptotic tests for the latter. We compare the g test results with those for Silvapulle in terms of size and power properties.

It is well known that a significant DW statistic can result from either AR(1) or MA(1) errors. King (1983) points out that wrongly correcting for AR(1) disturbances in a model with MA(1) disturbances can lead to inefficient parameter estimates and, more importantly, misleading inferences. Therefore, powerful testing procedures are desirable to distinguish these two error processes. Informal procedures such as deciding the error process of the model based on autocorrelation and partial autocorrelation patterns of the ordinary least squares (OLS) residuals have only an asymptotic justification (see King (1983) and King and McAleer (1987)).

In practice many economic time series are non-stationary (see Dickey et al. (1986)). Knowing whether nonstationarity in the data is due to a deterministic time trend or a unit root seems to be popular in econometrics. Nelson and Plosser (1982) argue that many economic time series are better represented by unit roots than by deterministic time trends. It is well known that a unit root process can be transformed to a stationary process by differencing the series. However, differencing a dependent variable to remove non-stationarity can give rise to an MA(1) error process (see Schwert (1989)). Also, from the literature we observe that many economic time series can be adequately represented by an IMA(1,1) process (see Ermini (1993) and Huberman and Schwert (1985)).

Newbold and Davies (1978) observe that the AR(1) and IMA(1,1) error processes are equally plausible ones. However, Wichern's (1973) study, shows that the AR(1)

process cannot provide an adequate approximation to an IMA(1,1) process in general. On the other hand, a significant DW statistic can result from either error process. Newbold and Davies (1978) note that wrongly assuming an AR(1) error process, when the true process is IMA(1,1), can lead to misleading statistical inferences. Thus, we need a powerful testing procedure to distinguish these two error processes as is in the case of testing for MA(1) errors against AR(1) errors.

The plan of this chapter is as follows. The theory concerning the g test is discussed in section 3.2. This section describes the GNPL briefly and how it can be used to construct a PO test of a composite null. Section 3.3 explains how the g test can be applied to testing non-nested error covariance matrices in the linear regression model. This section also explains how testing problems of this type can be simplified using invariance methods. The theory discussed in sections 3.2 and 3.3 is applied in section 3.4 to the above mentioned testing problems. The details of the Monte Carlo experiment and its main findings are given in section 3.5. Two applications of the g test to real world data are given in section 3.6. Finally, some concluding remarks are made in section 3.7.

3.2 Theory

This section explains the basic idea behind the GNPL and how it can be used to construct a point optimal test of a composite null hypothesis (the lemma is given in the Appendix 3.1). When nuisance parameters are not present, the GNPL provides a point optimal test of

$$H_0: x \text{ is generated by one of the densities } f_1, \ldots, f_r \tag{3.2.1}$$

against

$$H_a: x \text{ is generated by the density } f_{r+1}, \tag{3.2.2}$$

where x is the observed sample. That is, under H_0, we assume that either f_1 is the true density of x, or f_2 is the true density of x, and so on. f_{r+1} is the true density at the point where we would like to maximize the power.

The point optimal test provided by the GNPL for testing hypotheses of the form (3.2.1) and (3.2.2) is the one that rejects H_0 if

$$f_{r+1}(x) > \sum_{i=1}^{r} k_i f_i(x) . \tag{3.2.3}$$

For the existence of such a test, the following r size conditions,

$$\Pr[f_{r+1}(x) > \sum_{i}^{r} k_i f_i(x) \mid x \sim f_j(x)] = \alpha_j, \quad j = 1, \ldots, r, \tag{3.2.4}$$

need to be solved simultaneously, by appropriate choices of values for $k_1, \ldots, k_r$. Here, k_i, $i = 1, \ldots, r$, are the critical values of the test and α_j is the level of significance when $f_j(x)$ is the true density. These critical values can be positive and/or negative (see Appendix 3.2).

Suppose we wish to test composite hypotheses of the form,

$$H_0: x \text{ has density } f(x,\delta), \tag{3.2.5}$$

against

$$H_a: x \text{ has density } f(x,\lambda), \tag{3.2.6}$$

where δ is a $p \times 1$ vector of parameters restricted to the set Ψ and λ is an $q \times 1$ vector of parameters restricted to the set Φ. It is assumed that any knowledge about

the possible range of parameter values has been incorporated to keep the parameter sets, Ψ and Φ, as small as possible.

In order to construct a point optimal test for this problem, let us assume that $\lambda^* \in \Phi$ is the point under the alternative hypothesis at which we wish to optimize power. Thus, the testing problem given in (3.2.5) and (3.2.6) can now be written as

$$H_0: x \text{ has density } f(x,\delta), \tag{3.2.7}$$

against

$$H_a: x \text{ has density } f(x,\lambda^*). \tag{3.2.8}$$

In order to use the GNPL for this type of testing problem, we need to approximate $f(x,\delta)$, $\delta \in \Psi$, by a finite number of densities as in (3.2.1). That is, under H_0, we need to select r separate δ points in Ψ, namely δ_1, ..., δ_r, and define the corresponding densities as

$$f_i = f(x,\delta_i),\ i = 1, \ldots, r. \tag{3.2.9}$$

We regard these finite number of densities as representative densities of $f(x,\delta)$. Now the GNPL may be used to obtain a point optimal solution for this type of approximation.

The big questions now are how to select the representative densities (this is explained in section 3.2.1) and how many of them should be selected? For the second issue, it is tempting to select the representative densities over a fine grid points of δ. However, the larger the number representative densities we choose, the greater the computing

time it takes to find the appropriate k values of the test and also (because of imposing more restrictions under the null) it may reduce the power of the test. Therefore, it is better to aim for the minimum number of representative densities and if the critical values found for them are also sufficient to control the sizes of the test over Ψ then it will result in a more desirable test. The g test we propose in this chapter is based on this idea.

3.2.1 The g Test

The g test for testing composite hypotheses of the form given in (3.2.5) and (3.2.6) is the one with the minimum number of representative densities under the null. According to our experience, in the limited case of $p = 1$ (i.e., δ is a scalar) and Ψ being a closed interval, we need at least three representative densities (i.e., $r = 3$) under the null. Therefore, to construct the g test in this case, basically we start with three representative densities and find k_i values, $i = 1, \ldots, 3$, such that the following size conditions (which are evaluated via the Monte Carlo method) hold simultaneously:

$$\Pr[f(x,\lambda^*) > \sum_{i=1}^{3} k_i f_i(x,\delta_i) \mid x \sim f_j(x,\delta_j)] = \alpha, \; j = 1, \ldots, 3. \tag{3.2.10}$$

In the case of $p = 1$ and Ψ being a closed interval, δ_1 and δ_3 can be the start and end points of Ψ, respectively, and δ_2 can be any point in between (the reasons for selecting these points will be clear when it comes to applications, see section 3.3). Also, instead of using different levels (as allowed by the GNPL), we use the same significance level for all three size conditions as a standard approach.

Next, for the same k_i values, we check whether the sizes over a fine grid points of δ, are reasonably (i.e., sizes should be not significantly different from α) controlled[20].

[20] According to our experience, two representative densities (i.e., $r = 2$ case) are not sufficient for this, which is why the $r = 2$ case is not considered.

If this is the case then we have finished constructing the g test, otherwise, we can change δ_2 (while keeping the other two fixed) and proceed as before once again. If this does not help much in terms of controlling the sizes then the number of representative densities under the null can be increased by one and the whole process repeated. This is explained in more detail in the next section. Therefore, in order to construct the g test we start with three representative densities (i.e., $r = 3$ case), but may end up with more than three representative densities (i.e., $r > 3$ case).

The critical values, k_i, $i = 1, \ldots, r$, of the g test for $r = 3$ can be obtained by the following iterative procedure and this procedure can easily be generalised to $r > 3$ if need be[21].

(1) Guess some possible values for k_1, k_2, and k_3 and obtain the sizes of the test at δ_1, δ_2, and δ_3, via the Monte Carlo simulation method.

(2) Fix the size at δ_1 as α. To do this observe the size corresponding to δ_1. If that size is greater (less) than α then increase (decrease) the k_1 value (while keeping the other two k_i values fixed) and obtain the sizes once again. Repeat this process until the size at δ_1 is α.

(3) Try to fix the size at δ_2 as α by adjusting k_2 (as explained above). While doing this, check the size at δ_1. If it has changed (i.e., become not equal to α) go to step 2. Continue this until the size at both δ_1 and δ_2 is α.

(4) Try to fix the size at δ_3 as α by adjusting k_3. If the size at δ_1 (and δ_2) changes go to step 2 (and step 3). Repeat this process until the sizes at δ_1, δ_2 and δ_3 are α.

[21] The same can be achieved more easily from the user's point of view by using SA which is explored in Chapter 5.

3.3 Testing Composite Non-Nested Disturbance Covariance Matrices in the Linear Regression Model Using the g Test: Theory

Consider the linear regression model

$$y = X\beta + u, \tag{3.3.1}$$

where y is an $n\times 1$ vector, X is an $n\times k$ nonstochastic matrix of rank $k<n$, and β is a $k\times 1$ parameter vector. Suppose we wish to test

$$H_0: u \sim N(0,\sigma^2\Pi_1(\theta_1)),\ \ 0 \le \theta_1 \le p_1 \tag{3.3.2}$$

against

$$H_a: u \sim N(0,\sigma^2\Pi_2(\theta_2)),\ \ 0 < \theta_2 \le q_1 \tag{3.3.3}$$

or

$$H_0^-: u \sim N(0,\sigma^2\Pi_1(\theta_1)),\ \ p_1 \le \theta_1 \le 0 \tag{3.3.4}$$

against

$$H_a^-: u \sim N(0,\sigma^2\Pi_2(\theta_2)),\ \ q_1 \le \theta_2 < 0, \tag{3.3.5}$$

where, Π_1 and Π_2 are $n\times n$ positive definite matrices, θ_1 and θ_2 are the parameters of interest, p_1 and q_1 are known numbers and β and σ^2 are the nuisance parameters.

Observe that this type of testing problem is invariant with respect to transformations of the form

$$y \rightarrow \upsilon_0 y + X\upsilon, \tag{3.3.6}$$

where υ_0 is a positive scalar and υ is a $k \times 1$ vector. Therefore, the test statistics we consider should also be invariant to transformations of the form (3.3.6). The vector

$$w = \frac{Pz}{(z'z)^{1/2}}, \tag{3.3.7}$$

is a maximal invariant under the group of transformations given by (3.3.6), where $z = \overline{P}_X y$ is the $n \times 1$ OLS residual vector from (3.3.1)[22], and P is an $l \times n$ matrix such that $P'P = \overline{P}_X$ and $PP' = I_l$, in which $l = n - k$ (see King (1983, 1987b)). Here the row vectors of the P matrix are the orthonormal eigenvectors corresponding to the nonzero eigenvalues of the matrix $\overline{P}_X$.

Because w is a maximal invariant, any statistic invariant to transformations of the form (3.3.6) can be expressed as a function of w and any function of w is invariant under such transformations (see King (1983, 1987b)). Therefore, we can concentrate on test statistics which treat w as the observed sample and use the density of w as the likelihood function[23]. Ara and King (1993) showed that this type of approach is equivalent to considering marginal likelihood based tests.

Now the hypotheses (3.3.2) to (3.3.5) can be shown equivalent to

$$H_0: w \text{ has density } f_0(w,\theta_1), \quad 0 \leq \theta_1 \leq p_1 \tag{3.3.8}$$

against

$$H_a: w \text{ has density } f_a(w,\theta_2), \quad 0 < \theta_2 \leq q_1 \tag{3.3.9}$$

---

[22] For any full column rank matrix K, $\overline{P}_K = I_n - K(K'K)^{-1}K'$ is called the orthogonal projector.

[23] Such tests are known as invariant tests.

or

$$H_0^-: w \text{ has density } f_0(w,\theta_1), \quad p_1 \le \theta_1 \le 0 \tag{3.3.10}$$

against

$$H_a^-: w \text{ has density } f_a(w,\theta_2), \quad q_1 \le \theta_2 < 0, \tag{3.3.11}$$

where

$$f_0(w,\theta_1)\, dw = c\,\left|P\Pi_1 P'\right|^{-1/2} (w'\,(P\Pi_1 P')^{-1}\, w\,)^{-l/2}\, dw, \tag{3.3.12}$$

and

$$f_a(w,\theta_2)\, dw = c\,\left|P\Pi_2 P'\right|^{-1/2} (w'\,(P\Pi_2 P')^{-1}\, w\,)^{-l/2}\, dw \tag{3.3.13}$$

are the probability density functions of w under the null and alternative hypotheses, respectively, $c = \frac{1}{2}\,\Gamma(l/2)\,\pi^{-l/2}$ and dw denotes the uniform measure on the surface of the unit l-sphere.

Observe that the restriction to invariant tests has removed the nuisance parameters β and σ^2, as the densities above involve only the parameters of interest. Therefore, we can say that invariance has simplified the testing problem. Also we note that the testing problem mentioned above can easily be extended to disturbances following elliptical symmetry. That is, the distribution of any statistic that is invariant to the scale of u is invariant to such a widening of the normality assumption (see King (1979)).

Now the testing problem above is similar to the one discussed in section 3.2, thus, the g test can be applied. For testing H_0 against H_a, the g test can be constructed as follows[24]. First, we have to choose three points under the null to obtain the representative densities of w, $f_1(w)$, $f_2(w)$ and $f_3(w)$. Let these three points be $\theta_1 = 0$, $\theta_1 = \theta_1^*$ (where, $0 < \theta_1^* < p_1$) and $\theta_1 = p_1$. Among these, $\theta_1 = 0$ and $\theta_1 = p_1$ are assumed to be permanent points involved in the construction of the g test and so are the densities $f_1(w)$ and $f_3(w)$. On the other hand, $f_2(w)$ can vary according to the value θ_1^* takes. Let $f_4(w)$ be the density of w under the alternative hypothesis at $\theta_{2,0}$, the point where we want to maximise power[25]. For this problem, the g test can be denoted as the $g(\theta_{2,0}; 0, \theta_1^*, p_1)$ test. However, for convenience we denote it as the $g(\theta_{2,0})$ test throughout.

As mentioned earlier, for the existence of the $g(\theta_{2,0})$ test, we need to control the sizes of the test first. This involves finding θ_1^* and k_1, k_2 and k_3 values such that the following size conditions (which are evaluated via the Monte Carlo method) hold simultaneously.

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid \theta_1 = 0] = \alpha, \tag{3.3.14}$$

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid \theta_1 = \theta_1^*] = \alpha, \tag{3.3.15}$$

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid \theta_1 = p_1] = \alpha, \tag{3.3.16}$$

and

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid 0 < \theta_1 < p_1] = \alpha^* \tag{3.3.17}$$

---

[24] The construction is similar for testing H_0^- against H_a^-.

[25] Actually all these densities should be written as $f_1(w,0)$, $f_2(w,\theta_1^*)$, $f_3(w,p_1)$ and $f_4(w,\theta_{2,0})$, but for convenience they are denoted as above throughout this chapter.

where, α is the nominal significance level, and α^* should be within the (1-α) percent confidence interval of α when evaluated by Monte Carlo methods. If for particular values of θ_1^*, k_1, k_2 and k_3, probabilities (3.3.14) to (3.3.17) hold simultaneously, then we have constructed the g($\theta_{2,0}$) test, otherwise we can try another θ_1^* value and find k_1, k_2 and k_3 values accordingly and proceed as before. If this also doesn't work, as mentioned earlier, we can add another representative density under the null (say at θ_2^*) and try to solve the following four size conditions simultaneously by appropriate choices of θ_1^*, θ_2^*, k_1, k_2, k_3, and k_4:

$$\Pr[f_5(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) + k_4 f_4(w) \mid \theta_1 = 0] = \alpha, \qquad (3.3.18)$$

$$\Pr[f_5(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) + k_4 f_4(w) \mid \theta_1 = \theta_i^*] = \alpha, \; i = 1, 2, \qquad (3.3.19)$$

$$\Pr[f_5(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) + k_4 f_4(w) \mid \theta_1 = p_1] = \alpha, \qquad (3.3.20)$$

and

$$\Pr[f_5(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) + k_4 f_4(w) \mid 0 < \theta_1 < p_1] = \alpha^*, \qquad (3.3.21)$$

where, $f_1(w)$, $f_2(w)$, $f_3(w)$ and $f_4(w)$ are the representative densities under the null, and $f_5(w)$ is the density under the alternative where the power is maximized. Even if including another representative density also doesn't work as desired, one can include yet another representative density under the null and hence solve for another size condition and so on until the desired outcome is achieved.

The reason for including $\theta_1 = 0$ in the construction of the g($\theta_{2,0}$) test and fixing the size at this point as α is to make the power curve of the test start from α. $\theta_1 = \theta_1^*$ (and $\theta_1 = \theta_i^*$, $i = 1, 2, \ldots, (r-2)$, if more than three points are considered under the null) is included in the test because it brings the sizes for $0 < \theta_1 < p_1$ reasonably

closer to the nominal size. $\theta_1 = p_1$ is included in the test because it was found to be important in the work of Silvapulle and King (1991) and Silvapulle (1994a) where $p_1 = 1$.

3.4 Applications of the g Test

3.4.1 Testing MA(1) Errors against AR(1) Errors in the Linear Regression Model

Consider the linear regression model given in (3.3.1). If the elements of the $n\times 1$ disturbance vector u follow the MA(1) process,

$$u_t = e_t + \gamma e_{t-1}, \quad |\gamma| \le 1, \quad t = 1,\ldots,n, \tag{3.4.1}$$

where $e^* = (e_0, e_1,\ldots,e_n)' \sim N(0,\sigma^2 I_{n+1})$ then $u \sim N(0,\sigma^2\Omega(\gamma))$, where $\Omega(\gamma)$ is the $n\times n$ tridiagonal matrix with $1+\gamma^2$ as the main diagonal elements and γ as the nonzero off diagonal elements. If the elements of u follow the AR(1) process,

$$u_t = \rho u_{t-1} + e_t, \quad |\rho| < 1, \quad t = 1,\ldots,n, \tag{3.4.2}$$

where $u_0 \sim N(0,\sigma^2/(1-\rho^2))$ and $e = (e_1,\ldots,e_n)' \sim N(0,\sigma^2 I_n)$, then $u \sim N(0,\sigma^2\Sigma(\rho))$ in which $\Sigma(\rho)$ is an $n\times n$ matrix whose $(i,j)th$ element is $\rho^{|i-j|}/(1-\rho^2)$.

The hypotheses of interest for this problem can be written as

$$H_0: u \sim N(0, \sigma^2\Omega(\gamma)), \qquad 0 \le \gamma \le 1, \tag{3.4.3}$$

against

$$H_a: u \sim N(0, \sigma^2\Sigma(\rho)) \qquad 0 < \rho < 1, \tag{3.4.4}$$

or

$$H_0^-: u \sim N(0, \sigma^2\Omega(\gamma)), \qquad -1 \le \gamma \le 0, \tag{3.4.5}$$

against

$$H_a^-: u \sim N(0, \sigma^2\Sigma(\rho)) \qquad -1 < \rho < 0. \tag{3.4.6}$$

Normally positive autocorrelation is more likely in econometric applications. However, negative autocorrelation is also likely if the dependent variable is differenced. Therefore, both positive and negative autocorrelation are considered in this chapter. Obviously, the hypotheses (3.4.3) to (3.4.6) are special cases of (3.3.2) to (3.3.5), respectively, in which, $\theta_1 = \gamma$, $\theta_1^* = \gamma^*$, $\theta_2 = \rho$, $\theta_{2,0} = \rho_0$, $\Pi_1 = \Omega$, $\Pi_2 = \Sigma$ and for the positive autocorrelation case, $p_1 = 1$ and $q_1 = 0.999$ and for the negative case, $p_1 = -1$ and $q_1 = -0.999$. Therefore, the g test can be constructed for this problem as outlined in section 3.3. The g tests for positive and negative autocorrelation testing are denoted as the $g(\rho_0)$ and $g(-\rho_0)$ tests, respectively, throughout. In particular, for this problem, we consider four versions of the g test, namely, the g(0.3), g(0.5), g(0.75) and g(-0.5) tests.

3.4.2 Testing AR(1) Errors against IMA(1,1) Errors in the Linear Regression Model

Here we are interested in testing the error process given in (3.4.2) against an IMA(1,1) error process,

$$(1-L)u_t = (1+\gamma L)e_t, \quad -1 \le \gamma \le 0, \quad t = 1,\ldots,n, \tag{3.4.7}$$

where L is the lag operator such that $Lu_t = u_{t-1}$, $u_1 = e_1$ and $(e_0,\ldots,e_n)' \sim N(0,\sigma^2 I_{n+1})$.

We observe that when γ = -1, equation (3.4.7) reduces to $u_t = e_t$. Consequently, the null and the alternative models become identical when $\rho = 0$ and $\gamma = -1$.

Following Silvapulle (1994a), the error process in (3.4.7) may be written as

$$\Delta u = H^* e, \tag{3.4.8}$$

where

$$\Delta = \begin{bmatrix} t_1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ -1 & 1 & & & & & & 0 \\ 0 & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & & & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & -1 & 1 \end{bmatrix} \tag{3.4.9}$$

for some unknown t_1 and

$$H^* = \begin{bmatrix} h_{11} & 0 & . & . & . & . & . & 0 \\ h_{21} & h_{22} & & & & & & . \\ 0 & . & . & & & & & . \\ . & & . & . & & & & . \\ . & & & . & . & & & . \\ . & & & & . & . & & . \\ . & & & & & . & . & 0 \\ 0 & . & . & . & . & 0 & h_{nn-1} & h_{nn} \end{bmatrix}. \qquad (3.4.10)$$

Here the non-zero elements of H^* are generated by the recursive scheme,

$$h_{11} = (1+\gamma^2)^{1/2},$$

$$h_{i,i-1} = \frac{\gamma}{h_{i-1,i-1}},$$

$$h_{i,i} = [(1+\gamma^2) - h_{i,i-1}^2]^{1/2}, \quad i = 2,...,n. \qquad (3.4.11)$$

From (3.4.8), under H_1,

$$Var(u) = \sigma^2 \Delta^{-1} H^* H^{*'} (\Delta^{-1})'$$

$$= \sigma^2 \Sigma_1(t_1, \gamma) \text{ (say)}. \qquad (3.4.12)$$

Thus, the hypotheses of interest for this problem can be written as,

$$H_0 : u \sim N(0, \sigma^2 \Sigma(\rho)), \quad 0 \le \rho < 1 \qquad (3.4.13)$$

against

$$H_a : u \sim N(0, \sigma^2 \Sigma_1(t_1, \gamma)) \quad -1 \le \gamma < 0. \qquad (3.4.14)$$

We consider these particular hypotheses because for this case, Silvapulle (1991, 1994a) could not obtain King's PO test. The hypotheses given in (3.4.13) and (3.4.14) are special cases of (3.3.2) and (3.3.5), respectively, in which $\theta_1 = \rho$, $\theta_1^* = \rho^*$, $\theta_2 = \gamma$, $\theta_{2,0} = \gamma_0$, $\Pi_1 = \Sigma$, $\Pi_2 = \Sigma_1$, $p_1 = 0.999$ and $q_1 = -1$. Therefore, the g test can be constructed for this problem as outlined in section 3.3. Here we denote this test as the $g(\gamma_0)$ test throughout and consider the g(-0.5) test only. For this problem, Silvapulle (1991, 1994a) considered only an APO test, called the s(-0.5) test and found that her APO test behaves poorly compared to the asymptotic tests she considered. This prompted us to consider the g(-0.5) test only and compare its finite-sample size and power properties with those for Silvapulle's recommended asymptotic tests.

3.5 The Monte Carlo Experiment and the Results

In order to compare our Monte Carlo results of the g test with Silvapulle's (1991, 1994a) test results, we used her design matrices and the same values of the sample size.

Details of design matrices used for the first testing problem namely testing MA(1) errors against AR(1) errors are as follows:

X1 : $n \times 3$, n = 20 and 60, the regressors are a constant, real income and relative price of spirits. This design matrix is known as Durbin and Watson's (1951) consumption of spirits example.

X2 : $n \times 3$, n = 20 and 60, the regressors are a constant dummy, the Australian quarterly consumers' price index (CPI) starting with 1959(1), and the same index lagged one quarter.

X3 : $n \times 5$, n = 20 and 60, *X3* is obtained from *X2* by adding the CPI lagged two quarters and three quarters as additional regressors.

X4 : $n \times 3$, n = 20 and 30, the regressors are a constant, logarithms of Chow's (1957, table 1) automobile stock per capita and personal money stock per capita variables for the United States 1921-1950.

X5 : $n \times 3$, n = 20 and 60, the regressors are the eigenvectors corresponding to the three smallest eigenvalues of the $n \times n$ Durbin-Watson differencing matrix A_1, which is a tridiagonal matrix whose main diagonal elements are 2 except for the top left and bottom right elements which are both 1 and whose elements in the leading off diagonals are all –1. Here, the first regressor is a constant.

For the second testing problem namely testing AR(1) errors against IMA(1,1) errors, the following design matrices were used together with the *X1*, *X2* and *X5* matrices outlined above.

X6 : $n \times 1$, n = 20, 60 and 100, the regressor is the constant dummy.

X7 : $n \times 2$, n = 20, 60 and 100, the regressors are the constant dummy and time trend.

These design matrices cover a variety of characteristics. In particular, *X1* contains smoothly evolving series (intended to reflect some typical time series data) and *X2* and *X3* exhibit realistic degrees of multicollinearity. *X5* was included because the DW test is approximately uniformly most powerful invariant for this matrix. *X4*, *X6* and *X7* have been used in previous studies.

2000 replications were used in the two Monte Carlo experiments. For the problems under consideration, the performance of the g test is invariant to the values of β and σ^2, thus, β_i, i = 1, ..., k, and σ^2 were all set to unity. Also, for the second problem, following Silvapulle (1991, 1994a), we set $t_1 = 1$. The computer programs were written using GAUSS version 3.2.11. The sizes and powers of the tests were calculated at the 5% level of significance. For the first problem and for testing positive autocorrelation, sizes were calculated at γ = 0.1, 0.3, 0.5, 0.7 and 0.9 and powers were calculated at ρ = 0.1, 0.3, 0.5, 0.7 and 0.9. For testing negative

autocorrelation, sizes and powers were calculated at γ and ρ values of –0.1, -0.3, ... , -0.9. For the second problem, sizes were calculated at ρ = 0.1, 0.3, 0.5, 0.7, 0.9 and 0.99 and powers were calculated at γ = -0.1, -0.3, -0.5, -0.7, -0.9 and –1.0. Some selected size and power results are plotted in Figures 3.1 to 3.4.

The estimated size and power properties of the APOI tests, s(0.3), s(0.5), s(0.75) and s(-0.5), the LM test and the best PS test (denoted as the $\tau_1(<)$ test[26]) were obtained from Silvapulle (1991, 1994a). For the asymptotic tests, she used approximate non-similar critical values, in order to compare asymptotic tests' powers with those of the APOI tests at approximately the same level of significance.

The γ^*, k_1, k_2, and k_3 values of the g tests for the first problem are given in Table 3.1. Unlike for the first problem, for the second problem we were often forced to consider more than three representative densities under the null. The ρ_i^*, i = 1, ... , $(r-2)$, and the k values for the second problem are given in Table 3.5.

3.5.1 Monte Carlo Results for Testing MA(1) Errors against AR(1) Errors

For this testing problem, Silvapulle (1991) and Silvapulle and King (1991) considered six versions of the APOI tests, namely, the s(0.3), s(0.5), s(0.75), s(-0.3), s(-0.5), and s(-0.75) tests. They found that their APOI tests have superior small-sample properties compared to the asymptotic tests they considered. Therefore, it is appropriate to compare our g test results with their s test results only. The size and power results for this problem are presented in Tables 3.2 to 3.4. In the following discussion, the abbreviation g (and s) test stands for both positive and negative autocorrelation tests. Similarly, g(ρ_0) (and s(ρ_0)) test stands for all the positive autocorrelation tests considered.

[26] She named this test as the $\tau_1(<)$ test because it is used for testing the null against $\gamma < 0$.

Among the g test critical values, the k_2 values are always far larger in absolute magnitude than the k_1 and k_3 values (see Table 3.1). Thus, the γ^* value seems to be important in the construction of the g test. Also, the γ^* values and γ_0^* values of the s tests are always similar (see Silvapulle (1991)). Moreover, for some cases, the k_1 and k_3 values are not very different from zero. Consequently, for such cases, the g test statistic coincides with King's POI test statistic described in Chapter 2. We now compare the performance of the g and s tests under the null.

The sizes of the s tests are always less than or equal to the nominal size whereas the sizes of the g tests are always not significantly different from the nominal size[27]. On average, the g and s test sizes are comparable though sometimes the g test sizes can be as high as 0.058 (and 0.062 on one occasion).

The sizes corresponding to the g(-0.5) test are always closer to the nominal size compared to the sizes of the s(-0.5) test and g(0.5) test for all the design matrices when $n=20$ (and for *X4* when $n = 30$). The g(-0.5) test's sizes become similar to those of the g(0.5) test when the sample size increases to 60. Because the g test sizes are successfully controlled, we now turn our attention to the power properties of the s and g tests.

The power results are encouraging. The g(0.5) and g(0.75) tests are always more powerful than the s(0.5) and s(0.75) tests, respectively. Interestingly, the g(0.5) test is always more powerful than the s(0.75) test except on two occasions when $n=20$. Similarly the g(0.3) test is always more powerful than the s(0.3) test except for a few cases when $n=20$. Even for such cases, the power differences between the tests are small. Also, the g(0.3) test is always more powerful than the s(0.75) test when n = 60. For this testing problem, the s tests (which are nearly optimal) also perform equally well, therefore the power differences between the s and g tests are small. Among all the design matrices, the power superiority of the g(ρ_0) tests over the s(ρ_0) tests ranges up to 0.052 for $n = 20$ and to 0.076 for $n = 60$.

[27] For 2000 replications, estimated sizes in the range 0.041-0.060 (0.038-0.063) are not significantly different from the nominal size of 0.050 at the five (one) percent level.

One could suggest that these power results are to be expected because the g(ρ_0) tests' sizes are generally higher. However, compared to the s(ρ_0) tests, the g(ρ_0) tests often have exact sizes and good power properties. Also, there are some cases where the g(ρ_0) tests are more powerful than the s(ρ_0) tests while the corresponding test sizes are similar.

Like the s(ρ_0) tests' powers, the g(ρ_0) tests' powers are always greater than or equal to the nominal size. Also, the g(ρ_0) tests' powers always increase with ρ and n. We would expect the g(0.3) test, g(0.5) test and g(0.75) test to be the most powerful test for small, medium, and large values of ρ, respectively. However, we find that this is not always the case, for example, for the *X3* matrix, when $n = 60$ and $\rho = 0.3$, the powers of the g(0.3) and g(0.5) tests are 0.153 and 0.164, respectively, and the corresponding average sizes are 0.052 and 0.055 respectively. Similarly, for the *X5* matrix, when $n = 60$ and $\rho = 0.7$, the powers of the g(0.5) and g(0.75) tests are 0.890 and 0.883, respectively, and the corresponding average sizes are 0.057 and 0.054, respectively. Thus, the size differences of the g(ρ_0) tests may be causing this unexpected behaviour.

Finally, we observe that the size and power results of the g(ρ_0) tests are largely unaffected by the choice of ρ_0 value. The g(0.5) test seems worth focussing on in terms of powers, but only just.

The g(-0.5) test is always more powerful than the s(-0.5) test, except for some cases where the power differences between the two tests are small. Surprisingly, for such cases, the average sizes of the g(-0.5) test are always greater than or equal to those of the s(-0.5) test. The powers of the g(-0.5) test always increase with ρ and n. Among all the design matrices, the power advantage of the g(-0.5) test over the s(-0.5) test ranges up to 0.029 and 0.067 for $n = 20$ and $n = 60$, respectively.

The power results indicate that when the s test is nearly optimal, little gain can be achieved by applying the g test.

3.5.2 Monte Carlo Results for Testing AR(1) Errors against IMA(1,1) Errors

For this problem, Silvapulle (1991, 1994a) found that the asymptotic tests such as the LM test and pure significance test (denoted as the $\tau_1(<)$ test) behave better than her s(-0.5) test (which is not nearly optimal) and the other asymptotic tests she considered. In particular, she recommended the $\tau_1(<)$ test for small to moderate values of n and the LM test for large n. Therefore, for this case, we compare our g(-0.5) test results with the LM and $\tau_1(<)$ tests results and also with the s(-0.5) test results. The size and power results for this problem are presented in Tables 3.6 to 3.7.

For all the design matrices considered, we observe that the middle k values of the $g(-0.5)$ test are always larger in absolute magnitude than the first and last k values (see Table 3.5). This indicates that the ρ_i^*, $i = 1, 2, \ldots, (r-2)$, values are important in the construction of the g(-0.5) test.

The sizes of the g(-0.5) test are always not significantly different from the nominal size, whereas, the other tests' sizes are always less than or equal to the nominal size. The $\tau_1(<)$ test has its maximum size at $\rho = 0.1$ when $n = 20$, and its sizes decrease as ρ increases. A similar pattern can be seen for $n = 60$. For $n = 100$ and $0.1 < \rho < 0.9$, the $\tau_1(<)$ test' sizes are closer to zero in absolute value and its size becomes exact when $\rho = 0.99$. Like the g(-0.5) test, the LM test tends to have its maximum size at different values of ρ for different X matrices and sample sizes. Among the tests considered by Silvapulle (1991, 1994a), generally, the LM test sizes are much closer to the nominal size. The $\tau_1(<)$ test sizes are also better for small to moderate sized samples. On the other hand, the s(-0.5) test behaves poorly. Its sizes always decrease from 0.05 to 0 and then increase to 0.05 as ρ increases from 0 to 1. Also the s(-0.5) test sizes are similar for all the design matrices.

The g(-0.5) test sizes are generally not significantly different from the nominal size, and they vary between 0.038 to 0.058. On the other hand, as noted above, the other

test sizes are sometimes closer to zero in absolute value or vary between 0.010 to 0.050. This makes it difficult for us to compare the powers of the tests at approximately same level of significance.

The power results of the g(-0.5) test are very encouraging. Its powers are always well above the powers of other tests when γ varies over -0.1 to -0.9 and becomes exactly equal to 0.050 when $\gamma = -1$. The power result at $\gamma = -1$ is not surprising, because as mentioned earlier in section 3.4.2, the null and the alternative models of interest become identical when $\rho = 0$ and $\gamma = -1$. Therefore, the power at $\gamma = -1$, should be equal to the size of the test at $\rho = 0$, which in our case is fixed as 0.050.

Among all the design matrices, the power advantage of the g(-0.5) test over the $\tau_1(<)$ test when $n = 20$ and 60 are 0.148 and 0.354, respectively. Similarly, the power advantage of the g(-0.5) test over the LM test is 0.369 when $n = 100$. Here the $\tau_1(<)$ and LM tests are the ones recommended by Silvapulle (1991, 1994a) for small to moderate and large samples, respectively. The g(-0.5) test is always more powerful than these tests, Figures 3.1, 3.2 and 3.3 clearly illustrate this. On the other hand, the s(-0.5) test performs poorly in terms of power properties, for example, the power advantage of the g(-0.5) test over the s(-0.5) test is 0.969 when $n = 100$ (see Figure 3.4). The power dominance of the g(-0.5) test over the other tests may seem due to its higher sizes. However, there are many cases where the g(-0.5) test has better sizes and higher powers than its competitors. For example, for *X6*, $\gamma = -0.3$ and $n = 60$, the powers of the $\tau_1(<)$ and g(-0.5) tests are 0.261 and 0.615, respectively, while the corresponding average sizes are 0.040 and 0.051, respectively. Similarly, for *X7*, $n = 100$ and $\gamma = -0.7$, the powers of the LM and g(-0.5) tests are 0.520 and 0.882, respectively, while the corresponding average sizes are 0.044 and 0.050, respectively.

The g(-0.5) and LM tests' powers always increase with sample size, whereas, the $\tau_1(<)$ test powers increase when n increases from 20 to 60 but not when n increases from 60 to 100. The g(-0.5) and $\tau_1(<)$ tests' powers always first increase and then

decrease as γ decreases. The same happens for the s(-0.5) test when $n = 20$ and for the LM test when $n = 60$ and 100.

Compared to other tests, the g(-0.5) test detects the fact that the null and the alternative models are the same when γ = -1 and ρ = 0, perfectly well as its power always becomes equal to 0.050 at γ = -1. The g(-0.5) test is also very good in detecting the difference between γ = -0.9 and γ = -1 compared to other tests. The distinct power difference between these two γ points is more obvious in large samples. For example, for *X6* and n = 100, the g(-0.5) test powers corresponding to γ = -0.9 and -1 are 0.506 and 0.050, respectively. The second best test in this regard for the same data matrix and sample size is the LM test, which has powers 0.178 and 0.046, respectively.

Based on our results, we strongly recommend the g(-0.5) test for this testing problem.

3.6 Applications of the g Test to Real World Data

In this section, two applications of the g test to real world data are outlined.

3.6.1 Application 1

Here, we consider the model used by Silvapulle (1991), which is a simple quarterly linear regression model for Australian real interest rates given by

$$\Delta r_t = \beta_1 + \beta_2 S_{1t} + \beta_3 S_{2t} + \beta_4 S_{3t} + u_t \qquad (3.6.1)$$

where Δ denotes first differences, $r_t = R_t - \Pi_t$ is the ex post real interest rate, R_t is the nominal interest rate measured by the 90-day bank accepted bill rate, Π_t is the annual inflation rate calculated as $\Pi_t = 400 \times [\ln(CPI_t) - \ln(L(CPI_t))]$, CPI_t is the unadjusted weighted average of consumer price indices of all eight cities in Australia

at time t, and S_{1t}, S_{2t}, and S_{3t} are quarterly seasonal dummies. The *CPI* data we used here is different from Silvapulle's because her data is not currently available. She used an unadjusted weighted average of consumer price indices of six cities in Australia excluding Darwin and Canberra,. Using quarterly Australian data for the period 1969(2) to 1987(1), estimation of (3.6.1) by OLS results in

$$\Delta\hat{r}_t = \underset{(-2.604)}{-2.991} + \underset{(4.852)}{7.878}S_{1t} + \underset{(1.497)}{2.431}S_{2t} + \underset{(1.290)}{2.096}S_{3t} \qquad (3.6.2)$$

$$R^2 = 0.274\,,\ d = 2.903.$$

The DW statistic, d, indicates significant negative first-order autocorrelation at the one percent level of significance because the one percent exact critical value is given as 2.525 (see King (1981)). Although, our OLS results are slightly different from Silvapulle's, the finding is the same. Therefore, following her, we assume that the error term in (3.6.1) follows a negative MA(1) process. An obvious alternative hypothesis is that u_t follows a negative AR(1) process.

In order to construct the g(-0.5) test for this problem, we have to find γ^*, (where, -1 < $\gamma^* < 0$) and k_1, k_2 and k_3 values such that the following size conditions (which are evaluated via Monte Carlo method) hold simultaneously.

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid \gamma = 0] = \alpha\,, \qquad (3.6.3)$$

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid \gamma = \gamma^*] = \alpha\,, \qquad (3.6.4)$$

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid \gamma = -1] = \alpha\,, \qquad (3.6.5)$$

and

$$\Pr[f_4(w) > k_1 f_1(w) + k_2 f_2(w) + k_3 f_3(w) \mid -1 < \gamma < 0] = \alpha^*\,, \qquad (3.6.6)$$

where $f_1(w)$, $f_2(w)$, and $f_3(w)$ are the densities corresponding to $\gamma = 0$, $\gamma = \gamma^*$ and $\gamma = -1$, respectively, and $f_4(w)$ is the density corresponding to $\rho = -0.5$, the point where the power is to be maximised. The solution found at $\alpha = 0.05$ is $\gamma^* = -0.4229$, $k_1 = 0.0000115$, $k_2 = 3.23$ and $k_3 = 0.000024$. For these values the calculated value of the statistic is

$$g(-0.5) = \frac{\left|P\Sigma(-0.5)P'\right|^{-1/2}(w'(P\Sigma(-0.5)P')^{-1}w)^{-t/2}}{\sum_{i=1}^{3} k_i \left|P\Omega(\gamma_i)P'\right|^{-1/2}(w'(P\Omega(\gamma_i)P')^{-1}w)^{-t/2}} = 0.009 \qquad (3.6.7)$$

where $\Omega(\gamma_i)$, $i = 1,...,3$, correspond to Ω evaluated at $\gamma = 0$, $\gamma = -0.4229$, and $\gamma = -1$, respectively. Because the calculated value of the test is less than one, we do not reject the null hypothesis that the error term follows a negative MA(1) process at the five percent level of significance. This finding coincides with Silvapulle's.

3.6.2 Application 2

For this case, we obtained data from Griffiths et al. (1993, p.516). They model the response of an area of sugarcane sown in a region of Bangladesh by the linear model,

$$\ln(A_t) = \beta_1 + \beta_2 \ln(P_{st} / P_{jt}) + e_t, \qquad (3.6.8)$$

where A is the area for sugarcane production (in thousands of hectares), P_s is the price of sugarcane (in taka/tonne) and P_j is the price of jute (in taka/tonne). Using 34 annual observations, estimation of (3.6.8) by OLS results in

$$\ln(\hat{A}_t) = \underset{(0.214)}{6.12} + \underset{(0.141)}{1.004}\ln(P_{st} / P_{jt}), \qquad (3.6.9)$$

$$R^2 = 0.614, \; d = 1.093.$$

The DW statistic, d, indicates significant positive first-order autocorrelation at the 5 percent level of significance. Griffiths et al. assume that the errors of their model follow an AR(1) error process. Therefore, here we test an AR(1) error process against an IMA(1,1) error process using the g(-0.5) test.

The parameter solution found at the 5% level of significance is $k_1 = 0.125$, $k_2 = 1.499$, $k_3 = 1.53$, $k_4 = 0.43$, $\rho_1^* = 0.38$ and $\rho_2^* = 0.74$. For these values the calculated value of the statistic is 0.055 which is less than one, therefore we do not reject the null that the error term follows an AR(1) process at the five percent level of significance. This finding coincides with Griffiths et al.'s.

3.7 Conclusions

In this chapter, we proposed a new APO test, called the g test, based on the GNPL, for testing a composite null. This new test will be helpful for situations where King's PO test cannot be constructed. The small-sample properties of the g test were investigated using Monte Carlo simulations and the results were compared with those for APOI tests and asymptotic tests of Silvapulle (1991, 1994a). In particular, we applied the g test to two testing problems, namely, testing for MA(1) disturbances against AR(1) disturbances and testing for AR(1) disturbances against IMA(1,1) disturbances with a negative MA coefficient in the context of the linear regression model. Because she could not construct King's POI test for these two composite non-nested testing problems, Silvapulle (1991, 1994a) constructed King's APOI tests which performed extremely well for the former problem but performed poorly for the latter. Our g test performed well for both testing problems, in terms of size and power properties.

The power results for the first problem indicate that when King's APO test is nearly optimal, little gain can be achieved by applying the g test. For the second problem, Silvapulle (1994a) recommended two asymptotic tests, $\tau_1(<)$ and LM, for small to moderate and large samples, respectively. Our Monte Carlo results clearly show that

the g(-0.5) test is always far superior to these two tests in terms of power properties. Therefore, we strongly recommend the g(-0.5) test for testing AR(1) errors against IMA(1,1) errors with a negative MA coefficient in the linear regression model.

The performance of the g test for these two problems is highly encouraging. In this chapter, we analysed the performance of the g test in situations where the nuisance parameters can be avoided via invariance methods. But in practice there are situations where the nuisance parameters cannot be avoided via invariance or by any other means. It is interesting to see the performance of the g test in such situations. This is explored in Chapter 5.

In the next chapter, we discuss how to obtain exact (and near exact) non-similar critical values and exact size critical values (by assuming knowledge of the unknown parameters) of general non-similar tests, via SA.

Table 3.1: k_i and γ^* values for the g tests at the 5% level for testing MA(1) vs AR(1) errors

Test	k_i	*X1*		*X2*		*X3*		*X4*		*X5*	
		$n = 20$	$n = 60$	$n = 20$	$n = 60$	$n = 20$	$n = 60$	$n = 20$	$n = 30$	$n = 20$	$n = 60$
	k_1	0.0295	0.0045	0.04809	0.0035	0.0206	0.00276	0.014	0.0113	0.0394	0.00361
g(0.3)	k_2	1.5025	2.04	1.4255	2.11	1.368	2.04	1.48	1.727	1.4112	2.04178
	k_3	-0.000987	0.0000001	-0.001922	9.0×10^{-9}	0.000681	7.8×10^{-7}	0.00007	-0.000052	-0.000594	7.5×10^{-8}
	γ^*	0.28	0.277	0.28	0.28	0.27	0.28	0.27	0.278	0.27	0.277
	k_1	0.0105	0.0000845	0.01894	0.000148	0.025	0.00016	0.01986	0.00183	0.023	0.00023
g(0.5)	k_2	2.37	2.872	2.287	2.95	2.03	2.843	2.315	2.798	2.13	2.803
	k_3	0.0066	0.000026	-0.00712	0.00009	-0.0009	0.00023	-0.0044	0.00251	-0.00092	0.000027
	γ^*	0.405	0.405	0.405	0.4105	0.405	0.414	0.405	0.405	0.39	0.406
	k_1	0.00249	6.2×10^{-7}	-0.0005	4.1×10^{-7}	-0.0032	4.4×10^{-7}	0.0083	0.00029	-0.0002	1.08×10^{-6}
g(0.75)	k_2	3.1	0.8	2.78	0.614	2.64	0.682	3	2.496	2.8	0.835
	k_3	0.001	0.000065	0.095	0.000208	0.107	0.00048	0.027	0.009	0.025	0.00008
	γ^*	0.5	0.495	0.48	0.4995	0.48	0.505	0.499	0.499	0.46	0.492
	k_1	0.00045	0.00003	0.0051	0.000045	-0.00582	0.000045	0.002	0.00144	-0.00295	0.000018
g(-0.5)	k_2	2.828	3.1	2.541	3.18	1.815	2.94	2.755	3.162	2.0085	3.21
	k_3	0.2	0.0029	0.286	0.001	0.824	0.0024	0.145	0.1056	0.82	0.004
	γ^*	-0.5	-0.44	-0.5	-0.439	-0.5	-0.439	-0.5	-0.46	-0.5	-0.44

Table 3.2: Sizes and powers of the s(ρ_0) tests and g(ρ_0) tests at the 5% level for $n = 20$, when testing for MA(1) errors against AR(1) errors

		s(0.3)	s(0.5)	s(0.75)	g(0.3)	g(0.5)	g(0.75)
X1	$\gamma = 0.1$	0.046	0.044	0.044	0.053	0.055	0.053
	0.3	0.045	0.043	0.043	0.049	0.051	0.049
	0.5	0.046	0.045	0.044	0.042	0.054	0.050
	0.7	0.047	0.048	0.046	0.042	0.053	0.051
	0.9	0.048	0.049	0.048	0.047	0.050	0.050
	$\rho = 0.1$	0.052	0.051	0.051	0.056	0.060	0.054
	0.3	0.079	0.076	0.075	0.085	0.088	0.082
	0.5	0.116	0.163	0.159	0.168	0.190	0.182
	0.7	0.335	0.337	0.336	0.323	0.365	0.360
	0.9	0.515	0.526	0.536	0.508	0.556	0.562
X2	$\gamma = 0.1$	0.048	0.047	0.046	0.056	0.055	0.049
	0.3	0.044	0.045	0.044	0.046	0.052	0.052
	0.5	0.046	0.046	0.045	0.044	0.043	0.049
	0.7	0.049	0.048	0.048	0.042	0.047	0.053
	0.9	0.049	0.049	0.049	0.049	0.050	0.053
	$\rho = 0.1$	0.052	0.051	0.050	0.058	0.057	0.050
	0.3	0.076	0.077	0.073	0.089	0.082	0.074
	0.5	0.156	0.156	0.154	0.160	0.170	0.166
	0.7	0.311	0.324	0.324	0.310	0.338	0.349
	0.9	0.504	0.529	0.540	0.518	0.548	0.578
X3	$\gamma = 0.1$	0.045	0.046	0.047	0.052	0.056	0.049
	0.3	0.044	0.045	0.045	0.051	0.058	0.046
	0.5	0.044	0.045	0.045	0.052	0.049	0.049
	0.7	0.046	0.047	0.048	0.052	0.053	0.052
	0.9	0.048	0.049	0.050	0.050	0.051	0.048
	$\rho = 0.1$	0.052	0.053	0.053	0.056	0.060	0.052
	0.3	0.070	0.071	0.071	0.082	0.088	0.075
	0.5	0.130	0.133	0.132	0.154	0.158	0.149
	0.7	0.252	0.261	0.264	0.276	0.285	0.295
	0.9	0.418	0.435	0.447	0.453	0.472	0.480
X4	$\gamma = 0.1$	0.046	0.045	0.046	0.051	0.055	0.053
	0.3	0.044	0.044	0.045	0.050	0.053	0.053
	0.5	0.044	0.045	0.045	0.048	0.048	0.050
	0.7	0.046	0.048	0.047	0.050	0.046	0.050
	0.9	0.048	0.049	0.048	0.049	0.051	0.049
	$\rho = 0.1$	0.054	0.053	0.053	0.054	0.056	0.055
	0.3	0.076	0.076	0.074	0.077	0.081	0.088
	0.5	0.152	0.155	0.150	0.158	0.164	0.170
	0.7	0.293	0.310	0.317	0.304	0.312	0.329
	0.9	0.438	0.484	0.539	0.459	0.493	0.562
X5	$\gamma = 0.1$	0.046	0.045	0.045	0.052	0.055	0.048
	0.3	0.045	0.043	0.043	0.052	0.054	0.054
	0.5	0.046	0.045	0.044	0.050	0.049	0.050
	0.7	0.049	0.048	0.047	0.048	0.049	0.053
	0.9	0.050	0.050	0.049	0.049	0.049	0.051
	$\rho = 0.1$	0.053	0.052	0.053	0.055	0.059	0.053
	0.3	0.075	0.073	0.072	0.085	0.091	0.082
	0.5	0.141	0.139	0.137	0.150	0.160	0.150
	0.7	0.250	0.251	0.249	0.248	0.266	0.267
	0.9	0.343	0.348	0.350	0.345	0.361	0.363

Table 3.3: Sizes and powers of the $s(\rho_0)$ tests and $g(\rho_0)$ tests at the 5% level for $n = 60^*$, when testing for MA(1) errors against AR(1) errors

		s(0.3)	s(0.5)	s(0.75)	g(0.3)	g(0.5)	g(0.75)
X1	$\gamma = 0.1$	0.047	0.044	0.046	0.056	0.055	0.056
	0.3	0.042	0.039	0.039	0.050	0.049	0.050
	0.5	0.044	0.042	0.042	0.053	0.054	0.049
	0.7	0.046	0.044	0.044	0.054	0.058	0.055
	0.9	0.048	0.047	0.047	0.054	0.056	0.055
	$\rho = 0.1$	0.054	0.052	0.051	0.065	0.067	0.065
	0.3	0.131	0.122	0.119	0.159	0.156	0.147
	0.5	0.434	0.426	0.419	0.489	0.500	0.486
	0.7	0.839	0.842	0.845	0.877	0.890	0.889
	0.9	0.975	0.979	0.984	0.984	0.990	0.992
X2	$\gamma = 0.1$	0.046	0.044	0.043	0.056	0.057	0.058
	0.3	0.043	0.042	0.041	0.049	0.052	0.053
	0.5	0.045	0.043	0.042	0.047	0.050	0.050
	0.7	0.048	0.047	0.046	0.049	0.055	0.055
	0.9	0.049	0.049	0.049	0.050	0.052	0.053
	$\rho = 0.1$	0.053	0.052	0.051	0.061	0.065	0.065
	0.3	0.136	0.132	0.126	0.148	0.154	0.154
	0.5	0.455	0.453	0.442	0.486	0.512	0.513
	0.7	0.870	0.875	0.872	0.892	0.908	0.908
	0.9	0.990	0.991	0.992	0.993	0.995	0.996
X3	$\gamma = 0.1$	0.045	0.044	0.044	0.055	0.059	0.056
	0.3	0.043	0.041	0.041	0.048	0.054	0.050
	0.5	0.045	0.043	0.042	0.050	0.053	0.050
	0.7	0.048	0.047	0.047	0.054	0.060	0.058
	0.9	0.049	0.049	0.050	0.052	0.050	0.055
	$\rho = 0.1$	0.058	0.056	0.057	0.064	0.069	0.062
	0.3	0.132	0.127	0.123	0.153	0.164	0.151
	0.5	0.441	0.436	0.429	0.493	0.512	0.500
	0.7	0.859	0.862	0.862	0.891	0.901	0.894
	0.9	0.989	0.989	0.990	0.991	0.994	0.994
X4	$\gamma = 0.1$	0.044	0.045	0.044	0.051	0.052	0.054
	0.3	0.043	0.043	0.042	0.049	0.049	0.048
	0.5	0.045	0.044	0.043	0.044	0.049	0.051
	0.7	0.048	0.048	0.048	0.047	0.053	0.053
	0.9	0.049	0.049	0.050	0.048	0.053	0.052
	$\rho = 0.1$	0.052	0.054	0.054	0.061	0.058	0.061
	0.3	0.094	0.092	0.090	0.101	0.107	0.112
	0.5	0.243	0.244	0.238	0.254	0.274	0.278
	0.7	0.543	0.538	0.538	0.546	0.584	0.582
	0.9	0.778	0.792	0.803	0.786	0.818	0.837
X5	$\gamma = 0.1$	0.045	0.043	0.042	0.054	0.062	0.056
	0.3	0.043	0.042	0.040	0.049	0.054	0.048
	0.5	0.044	0.043	0.042	0.051	0.054	0.051
	0.7	0.047	0.047	0.047	0.055	0.061	0.057
	0.9	0.049	0.050	0.050	0.055	0.054	0.056
	$\rho = 0.1$	0.058	0.056	0.054	0.065	0.071	0.066
	0.3	0.133	0.130	0.122	0.159	0.164	0.149
	0.5	0.435	0.434	0.423	0.486	0.508	0.492
	0.7	0.840	0.844	0.842	0.876	0.890	0.883
	0.9	0.977	0.979	0.979	0.985	0.989	0.988

* For *X4*, sample size, $n = 30$ was used

Table 3.4: Sizes and powers of the s(-0.5) and g(-0.5) tests at the 5% level, when testing for MA(1) errors against AR(1) errors

		$n = 20$		$n = 60$*	
		s(-0.5)	g(-0.5)	s(-0.5)	g(-0.5)
X1	γ = -0.1	0.048	0.048	0.051	0.057
	-0.3	0.045	0.048	0.041	0.050
	-0.5	0.046	0.050	0.043	0.051
	-0.7	0.048	0.052	0.047	0.056
	-0.9	0.050	0.050	0.049	0.053
	ρ = -0.1	0.051	0.051	0.053	0.062
	-0.3	0.083	0.085	0.135	0.165
	-0.5	0.216	0.215	0.499	0.553
	-0.7	0.538	0.527	0.930	0.947
	-0.9	0.883	0.877	0.999	1.000
X2	γ = -0.1	0.047	0.050	0.047	0.057
	-0.3	0.043	0.051	0.041	0.050
	-0.5	0.045	0.050	0.045	0.053
	-0.7	0.048	0.052	0.049	0.056
	-0.9	0.050	0.051	0.050	0.052
	ρ = -0.1	0.050	0.056	0.053	0.064
	-0.3	0.081	0.099	0.136	0.173
	-0.5	0.217	0.246	0.508	0.553
	-0.7	0.546	0.553	0.933	0.947
	-0.9	0.889	0.894	0.999	1.000
X3	γ = -0.1	0.047	0.049	0.047	0.055
	-0.3	0.045	0.048	0.043	0.049
	-0.5	0.046	0.050	0.044	0.052
	-0.7	0.049	0.050	0.048	0.055
	-0.9	0.050	0.050	0.050	0.050
	ρ = -0.1	0.050	0.052	0.052	0.064
	-0.3	0.078	0.086	0.134	0.175
	-0.5	0.188	0.203	0.482	0.549
	-0.7	0.484	0.489	0.919	0.940
	-0.9	0.855	0.849	0.999	0.999
X4	γ = -0.1	0.049	0.052	0.048	0.050
	-0.3	0.044	0.052	0.042	0.051
	-0.5	0.043	0.050	0.044	0.050
	-0.7	0.045	0.049	0.046	0.050
	-0.9	0.047	0.051	0.047	0.050
	ρ = -0.1	0.053	0.055	0.051	0.057
	-0.3	0.082	0.092	0.096	0.115
	-0.5	0.217	0.235	0.299	0.332
	-0.7	0.544	0.559	0.711	0.731
	-0.9	0.887	0.887	0.967	0.962
X5	γ = -0.1	0.049	0.049	0.050	0.052
	-0.3	0.045	0.052	0.042	0.047
	-0.5	0.047	0.050	0.043	0.051
	-0.7	0.049	0.049	0.047	0.055
	-0.9	0.050	0.051	0.049	0.051
	ρ = -0.1	0.052	0.054	0.059	0.060
	-0.3	0.083	0.093	0.137	0.160
	-0.5	0.216	0.234	0.501	0.547
	-0.7	0.540	0.531	0.931	0.949
	-0.9	0.885	0.878	0.999	1.000

* For *X4*, sample sizes $n = 20$ and $n = 30$ were used

Table 3.5: k_i and ρ_i^* values of the g(-0.5) test at the 5% level for testing AR(1) errors against IMA(1,1) errors

n	k_1	k_2	k_3	k_4	k_5	ρ_1^*	ρ_2^*	ρ_3^*
				X1				
20	0.1185	2.09	0.285	-	-	0.4	-	-
60	0.01435	2.21	0.8725	0.0443	-	0.43	0.74	-
				X2				
20	0.0656	1.2865	1.2234	0.0096	-	0.32	0.69	-
60	0.0052	1.399	1.48	0.029	-	0.42	0.73	-
				X5				
20	0.571	1.188	0.00938	-	-	0.4	-	-
60	0.0104	1.93	1.75	0.0075	-	0.36	0.64	-
				X6				
20	0.195	2.71	0.57	-	-	0.595	-	-
60	0.00131	0.646	1.32	0.0416	-	0.454	0.799	-
100	7.7×10^{-6}	0.0346	0.192	0.054	0.00264	0.404	0.643	0.848
				X7				
20	0.414	1.996	0.026	-	-	0.5	-	-
60	0.00042	1.169	1.98	0.0118	-	0.37	0.72	-
100	0.00011	0.192	0.663	0.045	0.00018	0.402	0.6461	0.853

Table 3.6: Sizes and powers of the s(-0.5) and g(-0.5) tests at the 5% level for *X1*, *X2* and *X5*, when testing for AR(1) errors against IMA(1,1) errors

		$n = 20$				$n = 60$			
		$\tau_1(<)$	LM	s(-0.5)	g(-0.5)	$\tau_1(<)$	LM	s(-0.5)	g(-0.5)
X1	$\rho = 0.1$	0.050	0.050	0.029	0.047	0.050	0.050	0.024	0.053
	0.3	0.049	0.039	0.009	0.047	0.044	0.043	0.010	0.049
	0.5	0.040	0.029	0.000	0.052	0.046	0.034	0.000	0.052
	0.7	0.031	0.022	0.001	0.052	0.043	0.023	0.000	0.055
	0.9	0.018	0.019	0.033	0.048	0.032	0.015	0.020	0.056
	0.99	0.020	0.018	0.046	0.050	0.018	0.014	0.044	0.052
	$\gamma = -0.1$	0.031	0.012	0.045	0.085	0.050	0.021	0.040	0.145
	-0.3	0.067	0.018	0.056	0.163	0.239	0.089	0.038	0.551
	-0.5	0.100	0.025	0.064	0.172	0.425	0.193	0.029	0.727
	-0.7	0.073	0.037	0.059	0.116	0.312	0.128	0.029	0.518
	-0.9	0.051	0.045	0.054	0.056	0.085	0.054	0.035	0.118
	-1		0.057	0.051	0.050	0.052	0.049	0.045	0.050
X2	$\rho = 0.1$	0.048	0.050	0.030	0.044	0.023	0.050	0.022	0.046
	0.3	0.045	0.043	0.010	0.051	0.022	0.040	0.008	0.044
	0.5	0.050	0.038	0.001	0.048	0.021	0.032	0.000	0.056
	0.7	0.040	0.033	0.001	0.048	0.036	0.022	0.000	0.050
	0.9	0.038	0.024	0.034	0.050	0.040	0.016	0.015	0.051
	0.99	0.028	0.020	0.046	0.052	0.050	0.011	0.045	0.049
	$\gamma = -0.1$	0.059	0.016	0.045	0.082	0.111	0.007	0.043	0.143
	-0.3	0.126	0.018	0.056	0.166	0.350	0.065	0.040	0.538
	-0.5	0.170	0.029	0.063	0.188	0.519	0.188	0.030	0.752
	-0.7	0.128	0.034	0.059	0.127	0.379	0.129	0.030	0.593
	-0.9	0.061	0.049	0.055	0.061	0.068	0.044	0.038	0.131
	-1	0.051	0.052	0.052	0.050	0.039	0.055	0.045	0.050
X5	$\rho = 0.1$	0.050	0.049	0.032	0.054	0.050	0.034	0.024	0.053
	0.3	0.042	0.050	0.007	0.056	0.045	0.033	0.012	0.049
	0.5	0.042	0.048	0.002	0.050	0.050	0.036	0.000	0.055
	0.7	0.035	0.049	0.003	0.045	0.045	0.041	0.000	0.049
	0.9	0.025	0.046	0.033	0.048	0.034	0.050	0.024	0.045
	0.99	0.023	0.048	0.044	0.048	0.027	0.045	0.047	0.046
	$\gamma = -0.1$	0.046	0.032	0.046	0.071	0.090	0.017	0.045	0.117
	-0.3	0.085	0.026	0.053	0.107	0.327	0.051	0.039	0.404
	-0.5	0.092	0.031	0.064	0.101	0.465	0.101	0.030	0.580
	-0.7	0.077	0.041	0.059	0.068	0.301	0.051	0.038	0.394
	-0.9	0.054	0.052	0.055	0.053	0.077	0.027	0.045	0.087
	-1	0.055	0.049	0.053	0.050	0.050	0.030	*	0.050

*This value is not reported in Silvapulle (1991)

Table 3.7 : Sizes and powers of the s(-0.5) and g(-0.5) tests at the 5% level for *X6* and *X7*, when testing for AR(1) errors against IMA(1,1) errors

	$\tau_1(<)$	LM	s(-0.5)	g(-0.5)		$\tau_1(<)$	LM	s(-0.5)	g(-0.5)
				n = 20, *X6*					
$\rho = 0.1$	0.050	0.050	0.028	0.054	$\gamma = -0.1$	0.041	0.019	0.043	0.091
0.3	0.045	0.050	0.008	0.058	-0.3	0.083	0.025	0.055	0.220
0.5	0.037	0.048	0.000	0.049	-0.5	0.170	0.042	0.065	0.318
0.7	0.028	0.044	0.002	0.046	-0.7	0.152	0.039	0.058	0.253
0.9	0.022	0.046	0.032	0.053	-0.9	0.070	0.044	0.051	0.078
0.99	0.032	0.032	0.045	0.049	-1	0.055	0.047	*	0.050
				n = 20, *X7*					
$\rho = 0.1$	0.050	0.050	0.018	0.052	$\gamma = -0.1$	0.041	0.022	0.048	0.085
0.3	0.049	0.048	0.012	0.055	-0.3	0.084	0.015	0.053	0.163
0.5	0.045	0.042	0.001	0.050	-0.5	0.124	0.016	0.067	0.176
0.7	0.031	0.050	0.003	0.050	-0.7	0.075	0.030	0.055	0.125
0.9	0.024	0.038	0.025	0.050	-0.9	0.050	0.045	0.048	0.059
0.99	0.019	0.040	0.048	0.050	-1	0.050	0.049	0.050	0.050
				n = 60, *X6*					
$\rho = 0.1$	0.048	0.041	0.021	0.055	$\gamma = -0.1$	0.064	0.037	0.041	0.160
0.3	0.047	0.046	0.009	0.044	-0.3	0.261	0.248	0.039	0.615
0.5	0.050	0.045	0.000	0.053	-0.5	0.549	0.519	0.029	0.880
0.7	0.044	0.050	0.000	0.054	-0.7	0.619	0.434	0.030	0.802
0.9	0.025	0.043	0.018	0.048	-0.9	0.168	0.061	0.035	0.260
0.99	0.031	0.031	0.041	0.054	-1	0.052	0.041	0.042	0.050
				n = 60, *X7*					
$\rho = 0.1$	0.049	0.041	0.027	0.038	$\gamma = -0.1$	0.069	0.026	0.047	0.138
0.3	0.046	0.044	0.009	0.043	-0.3	0.317	0.125	0.035	0.475
0.5	0.050	0.043	0.000	0.058	-0.5	0.515	0.255	0.026	0.735
0.7	0.045	0.050	0.003	0.050	-0.7	0.403	0.152	0.034	0.574
0.9	0.031	0.047	0.017	0.045	-0.9	0.096	0.036	0.043	0.113
0.99	0.019	0.035	0.043	0.049	-1	0.049	0.042	*	0.050
				n = 100, *X6*					
$\rho = 0.1$	0.002	0.046	0.018	0.055	$\gamma = -0.1$	0.027	0.067	0.030	0.209
0.3	0.002	0.046	0.003	0.048	-0.3	0.222	0.548	0.024	0.856
0.5	0.001	0.049	0.000	0.057	-0.5	0.536	0.882	0.015	0.984
0.7	0.001	0.050	0.000	0.052	-0.7	0.521	0.802	0.017	0.962
0.9	0.010	0.045	0.015	0.051	-0.9	0.033	0.178	0.021	0.506
0.99	0.050	0.040	0.042	0.050	-1	0.002	0.046	0.034	0.050
				n = 100, *X7*					
$\rho = 0.1$	0.001	0.049	0.020	0.056	$\gamma = -0.1$	0.004	0.046	0.038	0.192
0.3	0.002	0.049	0.002	0.045	-0.3	0.086	0.400	0.025	0.769
0.5	0.002	0.045	0.000	0.055	-0.5	0.323	0.697	0.013	0.946
0.7	0.006	0.050	0.000	0.048	-0.7	0.257	0.520	0.014	0.882
0.9	0.009	0.045	0.014	0.046	-0.9	0.068	0.055	0.023	0.264
0.99	0.050	0.031	0.043	0.049	-1	0.001	0.047	0.035	0.050

*These values are not reported in Silvapulle (1991)

Figure 3.1: Comparing the PS and g(-0.5) tests for *X6* with n = 20, when testing for AR(1) errors against IMA(1,1) errors

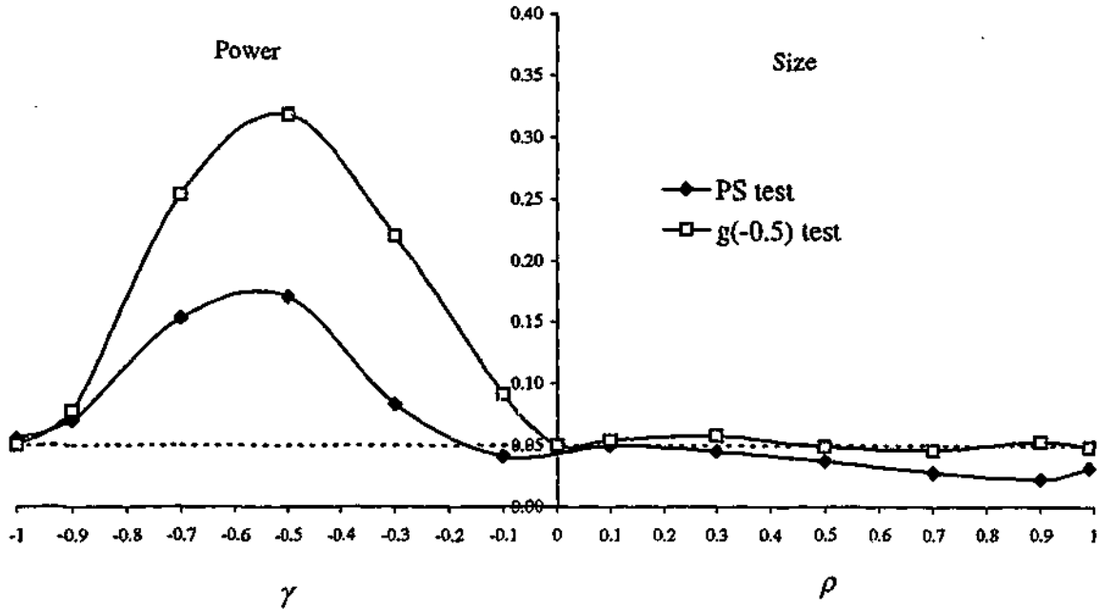


Figure 3.2: Comparing the PS and g(-0.5) tests for *X6* with n = 60, when testing for AR(1) errors against IMA(1,1) errors

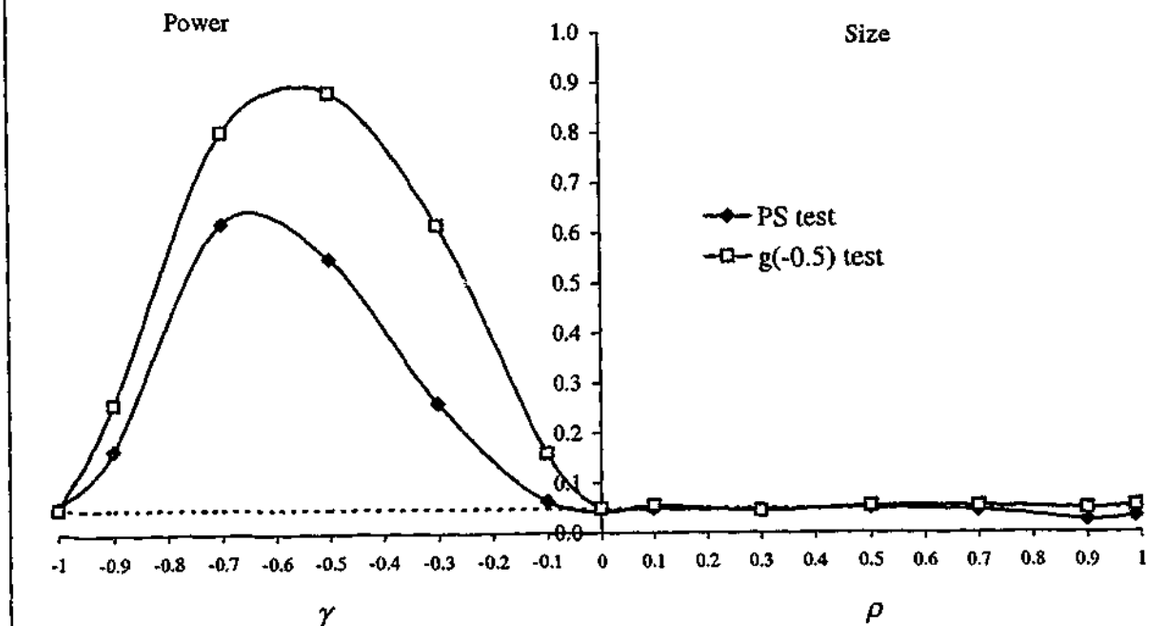


Figure 3.3: Comparing the LM and g(-0.5) tests for *X7* with n = 100, when testing for AR(1) errors against IMA(1,1) errors

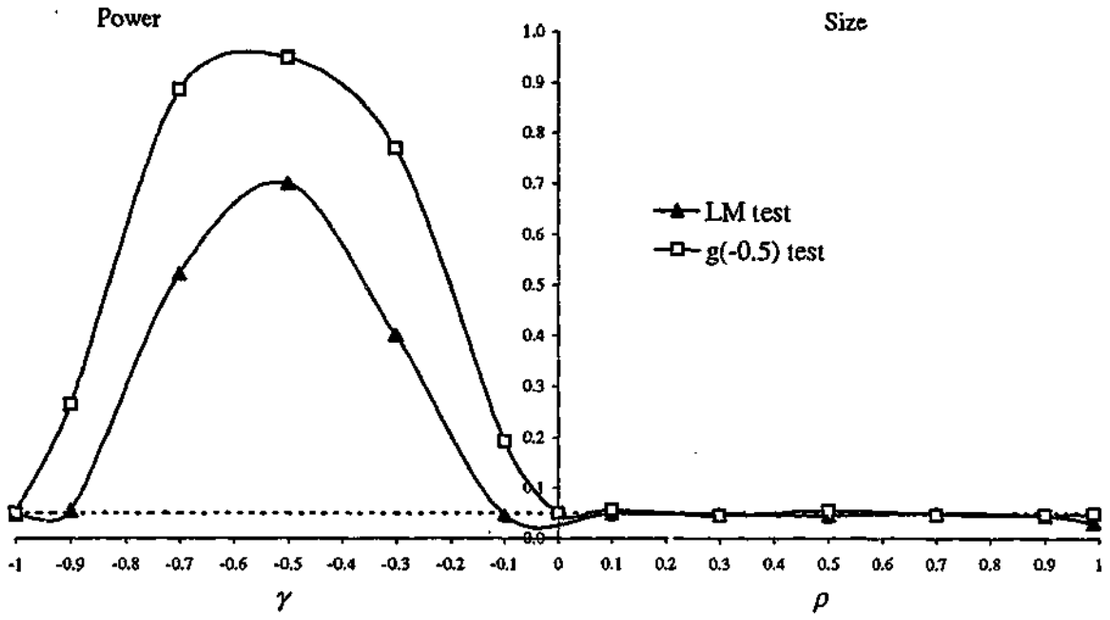
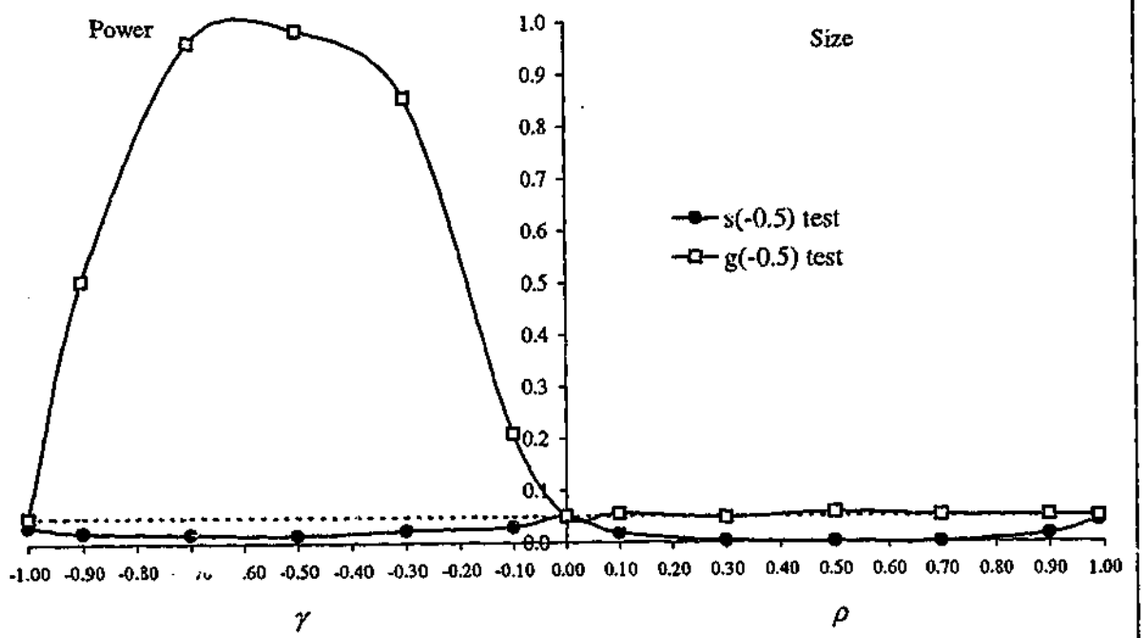


Figure 3.4: Comparing the s(-0.5) and g(-0.5) tests for *X6* with n = 100, when testing for AR(1) errors against IMA(1,1) errors



APPENDIX 3.2

The Generalised Neyman-Pearson Lemma (GNPL)

Here we give the theorem only. For full mathematical details see Lehman (1986).

Theorem (Lehman (1986, p. 96)): Let $f_1, \ldots, f_{r+1}$ be real valued functions defined on a Euclidean space A and integrable μ, and suppose that for given constants $\alpha_1, \ldots, \alpha_r$ there exists a critical function ϕ satisfying

$$\int \phi f_i d\mu = \alpha_i, \quad i = 1,\ldots,r. \tag{A3.1}$$

Let b be the class of critical functions ϕ for which (A3.1) holds.

1) Among all members of b there exists one that maximises

$$\int \phi f_{r+1} d\mu. \tag{A3.2}$$

2) A sufficient condition for a member of b to maximise (A3.2) is the existence of constants $k_1, \ldots, k_r$ such that

$$\begin{aligned} \phi(x) &= 1 \text{ when } \quad f_{r+1}(x) > \sum_{i=1}^{r} k_i f_i(x) \\ \phi(x) &= 0 \text{ when } \quad f_{r+1}(x) < \sum_{i=1}^{r} k_i f_i(x). \end{aligned} \tag{A3.3}$$

3) If a member of b satisfies (A3.3) with $k_1, \ldots, k_r \geq 0$ then it maximises (A3.2) among all critical functions satisfying

$$\int \phi f_i d\mu \leq \alpha_i, \quad i = 1, ..., r. \tag{A3.4}$$

4) The set R of points in $r-$dimensional space whose coordinates are

$$(\int \varphi f_1 d\mu, ..., \int \phi f_r d\mu)$$

for some critical function ϕ, is convex and closed. If α_i, $i = 1, ..., r$, is an inner point of R, then there exist constants $k_1, ..., k_r$, and a test ϕ satisfying (A3.1) and (A3.3), and a necessary condition for a member of b to maximise (A3.2) is that (A3.3) holds.

Notes

The size conditions (A3.1) can also be written more simply as

$$\Pr[\text{ reject } H_0 \mid f_i] = \alpha_i, \quad i = 1, ..., r, \tag{A3.5}$$

where the α_i's are ones preferred levels of significance, like, 0.05, 0.01 etc. Generally, we would like to fix the nominal size at one level. However, the GNPL does allow flexibility in this area when testing a composite null hypothesis.

The power of the test (A3.2) can also be written more simply as

$$\Pr[\text{ reject } H_0 \mid f_{r+1}]. \tag{A3.6}$$

The GNPL states that provided one finds appropriate critical values k_i, $i = 1, ..., r$, such that the r size conditions (A3.1) hold simultaneously, then for those critical values, the test (given in A3.3) will be the most powerful test among the tests of size equal to α_i, $i = 1, ..., r$. If all the critical values satisfying (A3.1) happen to be

positive then the test will be the most powerful test among all tests of size less than or equal to α_i, $i = 1, \ldots, r$.

CHAPTER 4

EXACT NON-SIMILAR CRITICAL VALUES FOR GENERAL NON-SIMILAR TESTS

4.1 Introduction

In the previous chapter, we considered two testing problems for which the nuisance parameters can be avoided by using invariance methods. However, in practice one often has to work in the presence of unavoidable nuisance parameters. The presence of nuisance parameters can make tests non-similar.

The classical approach to non-similar tests is to find exact non-similar critical values, for which sizes are never greater than the nominal size for all possible values of the nuisance parameters. Because there are no analytical tools proposed in the literature to derive or approximate these critical values for particular tests, the critical values have to be obtained by using the Monte Carlo method. Consequently, there are few existing studies based on non-similar critical values (see Inder (1985), Grant (1987), King and McAleer (1987) and Silvapulle (1991)).

As noted in Chapter 2, previous studies based on non-similar critical values (except studies involving PO and APO tests) are actually based on approximate non-similar critical values. Approximate non-similar critical values are obtained by varying the key nuisance parameters (those which if we knew their value we could apply a standard exact test) only, keeping the other nuisance parameters constant. Researchers (who used approximate non-similar critical values) checked whether their critical values are adequate to control the sizes over a part of nuisance parameter space only. Therefore, we can say that the performance of approximate non-similar critical values over the entire nuisance parameter space is unknown. Also, these critical values may not work well for all testing problems, tests and design matrices. Hence, the need for a method of finding exact non-similar critical values.

In this chapter, we propose a new approach to obtaining exact non-similar critical values based on the simulated annealing (SA) algorithm. This involves an iterative process as follows. First we allow SA to find values for the nuisance parameters (over the nuisance parameter space) such that the size of a non-similar test is at its maximum[28]. Then for those values of the nuisance parameters, we obtain the exact size critical value. Using SA and the new exact size critical value, we again obtain the maximum size of the non-similar test. If the maximum size obtained at this stage is equal to the nominal size then we stop this process, otherwise we obtain the exact size critical value for the new parameter solution and proceed as before, until the maximum size obtained is acceptable. This iterative procedure will eventually lead one to an exact non-similar critical value of a non-similar test. If ones aim is to obtain sizes that are always less than or equal to the nominal size, one can follow this iterative process. However, this can be an extremely computer intensive procedure. Therefore, a more practical approach might be to stop the iterative process after one full round of the procedure and hope the exact critical value obtained at this stage is close to the exact non-similar critical value (this assumes, little change in the values of the nuisance parameters that maximise size for different critical values). Critical values obtained this way can be regarded as near exact non-similar critical values. In this chapter, we investigated the performance of near exact non-similar critical values.

[28] The same could not be achieved by using the Newton-Raphson method, not surprisingly, because it is not designed to optimize a step function such as the size function estimated via simulation.

The exact non-similar critical values (involving Monte Carlo methods) mentioned above can be thought of as approximate because they are subject to sampling error. The size of this error can be controlled by the number of iterations taken. As noted in Chapter 1, our SA based approach will not only lead us to exact (and near exact) non-similar critical values but also allows us to assess whether an approximate distribution is a good approximation to the exact distribution of the test statistic under the null hypothesis. We also show how SA can be used to obtain exact size critical values of non-similar tests, assuming knowledge of the nuisance parameters.

In this chapter, we apply the SA based approach to two non-similar tests, namely, the DW test and Durbin's t test in the context of the dynamic linear regression model. Because Durbin's h test is known to perform poorly in finite samples and because it cannot be defined sometimes, we did not include it. The critical values for the DW and t tests are obtained from the approximate small disturbance asymptotic (ASDA) distribution and large-sample distribution of the statistics, respectively. Therefore, for this case, our SA based approach can be expected to indicate which asymptotic approach is best. The SA based approach can also be used to check, for example, whether the standard normal distribution or Student's t distribution is appropriate for the null distribution of Durbin's t test statistic in finite samples. Also, in this chapter, we compare SA based near exact non-similar critical values with approximate non-similar critical values obtained following Inder (1985). Further, an extensive Monte Carlo study is conducted to see whether SA based near exact non-similar critical values are indeed working well in terms of controlling the sizes of the tests over the nuisance parameter space.

The plan of this chapter is as follows. The theory including how SA can be effectively used to obtain exact (and near exact) non-similar critical values and exact size critical values of tests is discussed in section 4.2. This theory is applied in section 4.3 to the problem of testing for autocorrelation in the dynamic linear regression model. Section 4.4 presents the details of the Monte Carlo experiment and its main findings. Finally, some concluding remarks are given in section 4.5.

4.2 Theory

Let y be an observable $n\times 1$ vector which has probability density

$$f(y;\vartheta,\phi,\varepsilon), \tag{4.2.1}$$

where ϑ, ϕ and ε are $u\times 1$, $v\times 1$ and $w\times 1$ vectors of unknown parameters.

Suppose we wish to test

$$H_0: \vartheta = \vartheta_0$$

against

$$H_a: \vartheta > \vartheta_0 \text{ (or } \vartheta < \vartheta_0 \text{ or } \vartheta \neq \vartheta_0), \tag{4.2.2}$$

where ϑ_0 is a known $u\times 1$ vector. Then ϕ and ε are vectors of nuisance parameters. Suppose we have a test statistic $T(y)$ whose null distribution is invariant with respect to ϕ but depends on ε. In other words, the $T(y)$ test sizes vary with values of ε, thus, the test is non-similar. In sections 4.2.2 and 4.2.3, we discuss how to obtain exact size critical values (by assuming knowledge of the unknown parameters) and exact (and near exact) non-similar critical values, respectively, for such tests.

4.2.1 Finding the Maximum Size of a Non-Similar Test via SA

In order to apply SA, the function to be optimized first has to be defined. Usually one is able to provide this function explicitly. However, our case is different, because, the function, namely, the size function, has to be estimated via simulation. Therefore, this SA based approach may seem unattractive in terms of computation, however, this should not be the case, especially with the highly advanced computing facilities

available nowadays. Also, a computational gain in terms of time can be achieved by noticing which parameters need to be varied and which do not over the nuisance parameter space.

To find the maximum size of a non-similar test over the nuisance parameter space we can define the function to be minimized as either

$$f = 10000 \times s$$

or

$$f = 10000 \times s^2, \tag{4.2.3}$$

where $s = (1 - \varsigma)$ and ς = size of the test (over the nuisance parameter space) at the α nominal level of significance. The rationale behind (4.2.3) is simple. We want SA to find values for the nuisance parameter vector, ε, such that s is as small as possible. In other words, we want SA to find values for the nuisance parameters such that the test size is as big as possible. Therefore, by multiplying s by a big number (such as 10000) we are in a way forcing the SA algorithm more towards minimizing s. Here one is included in s, because we wish SA to find a size (if any) closest or equal to one.

4.2.2 A New Approach to Obtaining Exact Size Critical Values via SA

By assuming knowledge of the nuisance parameters, the exact size critical value of a non-similar test can be found via Monte Carlo methods. This involves taking repeated samples under the null hypothesis and finding the value for which the correct percentage of statistics are in the rejection region. Here we apply SA to do the latter part. That is, our method involves calculating the test statistic values for the number of Monte Carlo replications, and then applying SA to find the exact size critical value for the calculated test statistic values. That is, SA's role here is to vary the critical value of the test until the test size becomes equal to the nominal size.

If we wish to achieve an exact size critical value of a test at the α level of significance, the function to be minimized can be provided as

$$f = 10000 \times s^2, \tag{4.2.4}$$

where, $s = (\alpha - \varsigma)$. Here, we strictly prefer to work with s^2, instead of s, because s can also take negative values in this case. The rationale behind (4.2.4) is as explained above, that is, we want SA to find a critical value such that (4.2.4) is as small as possible, thus by multiplying by a big number this is ensured. In other words, we want SA to find a critical value such that the test size equals to α, which is why α is included in the function to be optimized.

4.2.3 A New Approach to Obtaining Exact (and Near Exact) Non-Similar Critical Values via SA

In this section, we provide the steps involved in obtaining an exact non-similar critical value of a non-similar test.

(1) Start the iterative process with (probably an asymptotic) critical value of a non-similar test.

(2) Apply SA (as explained in section 4.2.1) to find the nuisance parameter values such that the test size is at its maximum.

(3) Apply SA (as explained in section 4.2.2) to find exact critical value at these values of nuisance parameters.

(4) Using the new critical value repeat steps 2 and 3 continuously until convergence (i.e. repeat this process until maximum size obtained is equal to the nominal size).

As mentioned earlier, following all four steps will lead to an exact non-similar critical value and following the first three steps will lead to a near exact non-similar critical value of a non-similar test.

4.3 Testing for Autocorrelation in the Dynamic Linear Regression Model

Consider the dynamic linear regression model

$$y_t = \mu y_{t-1} + x_t'\beta + u_t, \quad t = 2,\ldots,n, \tag{4.3.1}$$

where y_t is the tth observation on the dependent variable, x_t is a $k \times 1$ vector of observations on the exogenous variables at time t, μ and β are unknown parameters with $|\mu| < 1$ (as noted in Chapter 2, μ is believed to be non-negative in economic applications) and u_t is a stochastic disturbance term which follows an AR(1) process,

$$u_t = \rho u_{t-1} + e_t, \quad t = 2,\ldots,n, \tag{4.3.2}$$

where, $|\rho| < 1$, and $e_t \sim IN(0,\sigma^2)$.

Suppose we wish to test

$$H_0: \rho = 0$$

against

$$H_1: \rho > 0. \tag{4.3.3}$$

As a result of the dynamic nature of the model, it is necessary to make further assumptions about y_1 and u_1 before the model is completely defined. Following Inder (1985) and King (1996), we make the following two assumptions.

(1) The mean of y_t is stable at $t = 1$. That is, $E(y_1) = E(y_0)$.

(2) The variance of y_t is the same for all $t = 1,...,n$.

According to these two assumptions, it can be shown that

$$y_1 = \frac{x_1'\beta}{(1-\mu)} + d_1 u_1, \tag{4.3.4}$$

where $d_1^2 = \frac{(1+\mu\rho)}{(1-\mu\rho)(1-\mu^2)}$ and $u_1 = \frac{e_1}{(1-\rho^2)^{1/2}}$ (see Inder (1985)). Thus, the model is now completely specified by equations (4.3.1), (4.3.2) and (4.3.4).

For this testing problem, μ, β and σ are nuisance parameters. However, this testing problem is invariant with respect to the constant coefficient, β_1, and is affected by σ only in the same way as a scaling of the $\beta^* = (\beta_2,...,\beta_k)'$ vector (see Inder (1985)). That is, the testing problem is invariant to rescaling such that the ratio β^* / σ is preserved. This means, β^* can be fixed and σ can be varied or vice versa. Because σ can take only positive values, it is better to vary β^*. Therefore, for the testing problem, μ and β^* (or σ) are unavoidable nuisance parameters. Consequently, tests applied to this problem will be non-similar in nature and the critical values of the tests can be obtained following section 4.2.3.

It will be useful to represent the model in vector notation. For any time series z_t, define $z = (z_2, z_3,..., z_n)'$ and $z_{-1} = (z_1, z_2,..., z_{n-1})'$. For example, y stands for $(y_2,..., y_n)'$ and y_{-1} for $(y_1,..., y_{n-1})'$. Similarly for any matrix H, $H' = (h_2,..., h_n)$, and $H_{-1}' = (h_1,..., h_{n-1})$, where h_t' is the row of H representing the t th observations. For example, X' stands for $(x_2,..., x_n)$ and X_{-1}' for $(x_1,..., x_{n-1})$.

Following Inder (1985), the OLS estimates of μ and β, and the OLS residual vector from (4.3.1) can be written as

$$\hat{\mu} = \frac{y_{-1}'\overline{P}_X y}{y_{-1}'\overline{P}_X y_{-1}}, \tag{4.3.5}$$

$$\hat{\beta} = (X'X)^{-1}X'(y - y_{-1}\hat{\mu}), \tag{4.3.6}$$

and

$$\hat{u} = \overline{P}_X(y - y_{-1}\hat{\mu}), \tag{4.3.7}$$

respectively, where $\overline{P}_X$ is the orthogonal projector of X.

4.3.1 The Tests

Durbin and Watson (1950, 1951) proposed the famous DW test for AR(1) disturbances in the context of the static linear regression model, based on the statistic

$$d = \frac{\hat{u}'A_1\hat{u}}{\hat{u}'\hat{u}}, \tag{4.3.8}$$

where A_1 is the first differencing matrix as defined in section 3.5. For the testing problem considered, the null hypothesis is rejected for small values of d.

The construction of Durbin's t test for testing H_0: $\rho = 0$ amounts to the following. First $\hat{u}$ has to be regressed on y_{-1}, X, and $\hat{u}_{-1}$, thereafter, the significance of the coefficient of $\hat{u}_{-1}$ is tested using the usual t test, which is called Durbin's t test. For any time series vector, w_t, by defining $w^- = (w_3, w_4, \ldots, w_n)'$ and $w_{-1}^- = (w_2, w_3, \ldots, w_{n-1})'$, Inder (1985) showed that Durbin's t statistic can be written as

$$t = \frac{\hat{u}_{-1}^{-\prime} \overline{P}_Z \hat{u}^- \sqrt{n-k-4}}{\sqrt{\hat{u}^{-\prime} \overline{P}_Z \hat{u}^- . \hat{u}_{-1}^{-\prime} \overline{P}_Z \hat{u}_{-1}^- - (\hat{u}_{-1}^{-\prime} \overline{P}_Z \hat{u}^-)^2}}, \tag{4.3.9}$$

where $Z = [y_{-1}^-, X^-]$. The t test, under regularity conditions outlined by Durbin (1970), follows the standard normal distribution asymptotically.

4.4 Monte Carlo Experiment Design and Main Findings

An extensive Monte Carlo study was conducted to see whether SA based near exact non-similar critical values work well in terms of controlling the sizes over the nuisance parameter space. For this a variety of design matrices with different characteristics were used, namely:

X8: $n \times 4$, $n = 20$ and 60, the X matrix is the same as *X5*. The σ values used for $n =$ 20 and 60 are $\sigma = 0.7$, 2, and 40 and $\sigma = 0.64$, 1.35, and 8, respectively.

X9: $n \times 4$, $n = 20$ and 60, the regressors are a constant and eigenvectors associated with the largest characteristic roots of A_1. The σ values used for $n = 20$ and 60 are $\sigma = 0.25$, 0.54, and 6 and $\sigma = 0.4$, 0.86, and 5, respectively.

X10: $n \times 4$, $n = 20$ and 60, regressors are a constant, the natural log of quarterly observations on three Australian series: nominal interest rates, CPI and GDP starting 1969(3). The σ values used for $n = 20$ and 60 are $\sigma = 0.2$, 0.5, and 3 and $\sigma = 0.3$, 0.8, and 4, respectively.

X11: $n \times 2$, $n = 32$ and 76, the regressors are a constant dummy, and Maddala and Rao's (1973) GNP data. The σ values used for $n = 32$ and 76 are $\sigma = 26.5$, 53, and 210 and $\sigma = 57$, 116, and 465, respectively.

X12: This design matrix is identical to *X11* except that the GNP series is replaced by another series with less serial correlation generated by adding a random variable o_t to

the GNP series, where $o_t \sim IN(0,1600)$. The σ values used for $n = 32$ and 76 are $\sigma = 31$, 64.2, and 250 and $\sigma = 72$, 147, and 580, respectively.

X13: $n \times 5$, $n = 30$ and 60, the regressors are a constant, three quarterly seasonal dummy variables, and the quarterly Australian CPI commencing 1959(1). The σ values used for $n = 30$ and 60 are $\sigma = 0.5$, 5, and 150 and $\sigma = 10$, 20, and 250, respectively.

X14: $n \times 3$, $n = 30$ and 60, the X matrix is the same as *X1*. The σ values used for $n = 30$ and 60 are $\sigma = 0.035$, 0.081, and 0.68 and $\sigma = 0.085$, 0.185, and 0.95, respectively.

Among these, *X8* to *X10* were used by Grant (1987) and the rest by Inder (1985). These design matrices cover a variety of characteristics. In particular, *X8* and *X9* (as noted in Chapter 3) reflect some extreme data sets, *X10* is typical economic data, *X11*, *X13* and *X14* are smoothly evolving series, and *X12* possesses a high degree of randomness.

In order to check whether the SA based near exact non-similar critical values are adequately controlling the sizes, sizes were estimated via simulation for a variety of β and σ values. Accordingly, β vectors over a range of directions were used, for example, for *X14*, the β vectors used were: $(0,0,0)'$, $(0,1,1)'$, $(0,-1,-1)'$, $(0,-1,1)'$, $(0,1,-1)'$, $(0,1,0)'$, $(0,-1,0)'$, $(0,0,1)'$ and $(0,0,-1)'$, and β was made larger or smaller along each of eight directions by decreasing or increasing σ. The complete list of β vectors used for each design matrix is given in the footnotes below each table of results. For each of these β directions and σ values, sizes were estimated for $\mu = 0.1$, 0.3, 0.5, 0.7, 0.9, and 0.99. When $\beta = 0$, results are invariant with respect to σ, therefore, for this case, size calculations were done for only one σ value. All other studies conducted in a similar setting, calculated sizes for only one β vector and for some σ values (see Inder (1985) and Grant (1987)).

To ensure that reasonable values of σ were used, three σ values that made the average R^2 (coefficient of determination) approximately 0.3, 0.7 and 0.9, respectively, were chosen. The average R^2 for all β directions used was obtained by assuming $\mu = 0.5$ and $\rho = 0$ and for each β vector with 100 iterations being used.

A nominal significance level of five percent and 5000 iterations were used throughout. The ASDA distribution based critical values for the DW test are obtained from Inder (1985) and Grant (1987). Because the testing problem of interest is invariant with respect to β_1 and is affected by σ only as a scaling of $\beta^* = (\beta_2,...,\beta_k)'$ vector, we set $\beta_1 = 0$ and $\sigma = 1$ throughout. Here σ is fixed as one, purposely, because it allows us to see the size changes with respect to true β^* (over the nuisance parameter space). The random normal numbers of the model were generated using seed value 98726679. The uniform random numbers of SA were generated using seed value 6696. The computer programs were written in GAUSS for windows NT/95 version 3.2.35 and Tsionas's (1995) program code for the SA algorithm was used.

For the testing problem considered, following Inder (1985), approximate non-similar critical values are obtained by setting β^* to zero and experimenting with μ values until the smallest (and largest) critical value for the DW (and t) test is found. Eleven values of μ between 0 and 1, namely, $\mu = 0.001, 0.1, 0.2, 0.3, \ldots, 0.9$ and 0.999 were used in a grid search. On the other hand, SA based near exact non-similar critical values are obtained by varying β^* and μ over the nuisance parameter space.

Before applying SA to find values for the nuisance parameters, β^* and μ, we should note an important issue. That is, Tsionas's program code is not intended for optimizing functions which need to be evaluated via the Monte Carlo method, such as, the size functions. Perhaps because of this or may be this is a problem specific to GAUSS, we observed that the SA algorithm gets confused when there is a seed value

present inside the main procedure in which the function to be optimized is evaluated[29] and tends to consider only that seed value and ignore any other seed value present outside the main procedure such as the one used for generating uniform random numbers. Consequently, the experiment cannot be repeated if it needs to be[30]. Therefore, in order to avoid this problem, we were forced to generate the normal random numbers of the model beforehand, outside of the main procedure where the function to be optimised is evaluated. This means we had to generate, an $n \times 5000$ matrix of random normal numbers (using seed value, 98726679) beforehand[31]. This effectively avoids the presence of a seed value inside the main procedure, which makes SA work properly. The need to generate random numbers beforehand is the only undesirable aspect of the SA based approach because this may cause trouble if one wants to consider an even larger number of replications. For a large number of replications, this amounts to using a vast amount of computer memory, which will slow down the computer. This hurdle may be overcome in the future, because, Tsionas (1995) commented that he is going to write another programming code for SA and others might do this too. If this problem is rectified then it will increase the use of SA in econometrics.

In order to obtain the maximum sizes of the tests, we optimized (4.2.3) by setting the SA parameters as, $N_t = 2$, $r_t = 0.5$, $N_s = 2$ and $T = 2$. The starting values for the parameters provided were, $\beta^* = 0$ and $\mu = 0.5$. For these parameter values, SA quickly finds the optimum solution. We observed that the maximum time it takes to provide the results is about one and a half hours (for $n = 76$) and the minimum time it takes is about three minutes (for $n = 20$) when using Pentium II 400 or Pentium III 500 machines.

In order to check whether SA has indeed achieved the global maximum size, following Goffe et al.'s (1994) suggestion, we obtained the optimum solution once again for a different uniform random number generator seed value, namely, 123

[29] To obtain sizes via the Monte Carlo method, naturally, a seed value for the random number generator of the model has to be provided.
[30] Even if the seed values involved in programming, are defined separately as seed1 and seed2, and even if they are used in separate procedures, the SA algorithm still considers one seed value and ignores the other.
[31] If a seed value is permissible inside the main procedure, we would have generated an $n \times 1$ vector of random numbers for each iteration.

(because SA is largely independent of starting values, the same starting values were used). If the maximum sizes obtained for both the seed values were roughly the same then we stopped, if not, we increased the SA parameters and proceeded as before. We observed that, overall, the small SA parameter values (as above) are adequate to find the global maximum size, except for a few cases. For such cases, increasing the SA parameter values led to the global maximum size, except for *X12*, $n = 76$ and the DW test. Also, for both the uniform random number generator seed values, we noted that approximately the same maximum sizes occur, but for different parameter values. This is possible with size functions because they are step functions. That is, for different parameter values it is possible to obtain the same size. To check why SA appeared to have failed for *X12*, $n = 76$ and the DW test, we obtained sizes via simulation for a variety of nuisance parameter values. These sizes suggest that (for this case) the size function is almost always bounded below by 0.057 and reaches 0.085 only when $\beta = 0$ and $\mu \geq 0.9$. Therefore, for the starting values we used, there will be no significant change in the function value for quite some time, consequently SA stops at the local maximum size, 0.057. For this case, by increasing the SA parameter values further or by providing starting values close to $\beta = 0$ and $\mu > 0.9$, we could have achieved the global maximum size. However, for this case, we used the local maximum size in order to see its impact on the performance of an near exact non-similar critical value.

In order to obtain exact size critical values (by assuming knowledge of the unknown parameters), we optimised (4.2.4) by setting $N_t = 2$, $N_s = 6$, $r_t = 0.5$ and $T = 2$. According to our experience, these parameters are adequate to obtain the exact size critical values. If they are not adequate, obviously, one can increase these parameters slightly and proceed. For these parameters, SA will take less than three seconds to produce the result.

The maximum sizes of the tests over the nuisance parameter space and the corresponding parameter values are given in Table 4.1. Tables 4.2 and 4.3 report approximate non-similar critical values and SA based near exact non-similar critical values of the DW test and Durbin's t test, respectively. The size results using near exact non-similar critical values are given in Tables 4.4 through to 4.17.

4.4.1 The Results

Our study clearly indicates that neither the ASDA distribution based DW test nor the large-sample distribution based t test is best under the null because both tests can have significantly higher than nominal sizes over the nuisance parameter space (see Table 4.1). Sometimes, both tests have approximately the same (higher than nominal) maximum sizes. For example, for *X10* and $n = 20$, the DW test and Durbin's t test have maximum sizes of 0.109 and 0.114, respectively, over the nuisance parameter space. The DW test behaves markedly better than Durbin's t test only on three occasions. For example, for *X9* and $n = 20$, the maximum sizes observed for the DW test and Durbin's t test are 0.093 and 0.281, respectively. Similarly, Durbin's t test outperforms the DW test on two occasions. Across all the design matrices, the highest sizes observed for the DW test and Durbin's t test over the nuisance parameter space are, 0.205 and 0.281, respectively.

Sometimes, the performance of Durbin's t test improves with sample size. For example, for *X9*, the maximum sizes observed for $n = 20$ and 60 are 0.281 and 0.057, respectively. Apart from this example, strangely, there is no notable improvement in the t test's maximum sizes with respect to the sample size. In some cases, the maximum size increases with sample size as well. The DW test also has similar behaviour.

The DW test's maximum size generally occurs for small β^* values and for μ between 0.84 to 0.98, however, on one occasion, the same happens for big β^* values and μ closer to zero. The maximum size of Durbin's t test occurs for small β^* values and for μ between 0.27 to 0.92.

For the DW test, approximate non-similar critical values and near exact non-similar critical values are approximately the same, except for a few cases. For Durbin's t test, approximate non-similar critical values are generally higher than near exact non-similar critical values, except for *X12* when $n = 32$.

For the problem of interest, (any) β approaches 0 when σ increases, therefore, the size results for $\beta = 0$ and for big σ can be expected to be alike. Thus, in the discussion below any findings reported for $\beta = 0$ also apply to big σ.

The near exact non-similar critical values we proposed are working well in terms of controlling the sizes for almost all the design matrices considered. Only for *X12*, $n = 76$, $\mu \geq 0.9$ and the DW test, do we have sizes that are slightly above the nominal size. However, these sizes are never greater than 0.075 (see Table 4.8). Except for this case, the SA based approach works extremely well for all the design matrices and sample sizes. Recall that only for *X12*, $n = 76$, and the DW test, we used the local maximum size in order to achieve the near exact non-similar critical value. Therefore, according to our results, if we make sure that the SA algorithm achieves the global maximum size at step 2 of section 4.2.3, then near exact non-similar critical values can be expected to work well.

The sizes based on near exact non-similar critical values are generally less than or equal to the nominal size and on some occasions marginally above the nominal size. If the tests were to be ranked on the basis of their sizes, overall Durbin's t test is superior. Of the 1668 cases, the t test has sizes that are not significantly different from the nominal size in 688 cases. Similarly, the DW test sizes are not significantly different from the nominal size in 319 cases. If we consider sizes above 0.030 the superiority of Durbin's t test is even more marked.

As noted above, the near exact non-similar critical values work well for Durbin's t test compared to the DW test. Recall that the approximate non-similar critical values of Durbin's t test are almost always greater than the near exact non-similar critical values, therefore, if we use the former critical values they will unnecessarily make the test under reject. This in turn will make the test's sizes low and hence reduce the powers. Therefore, the approximate non-similar critical values do not seem suitable for Durbin's t test. The DW test seems to be less affected in this regard.

Inder's (1985) results based on approximate non-similar critical values, show that the DW test performs better than Durbin's asymptotic tests. However, his study also

shows that Durbin's t test always has reasonable powers which are not far behind those of the DW test. The above observation led us to wonder whether the inferiority of Durbin's t test is due to the approximate non-similar critical values Inder used.

4.5 Conclusions

How to successfully deal with non-similar tests in finite samples is an important problem that econometricians often face. This is because the sizes of such tests vary with nuisance parameter values. The classical approach to this problem is to use exact non-similar critical values, for which the size is never greater than the nominal size for all possible values of the nuisance parameters. Because these critical values are extremely hard to compute, researchers nearly always use approximate non-similar critical values. In this chapter, we proposed a new approach (based on the SA algorithm) to obtain exact non-similar critical values of general non-similar tests. Our SA based approach involves controlling the maximum size of a non-similar test over the nuisance parameter space, therefore this approach also allows us to assess any approximate distributions under the null. Because this approach is extremely computer intensive, in this chapter, we also suggested and investigated the performance of near exact non-similar critical values. In addition, we showed that SA can be used to obtain exact size critical values (by assuming knowledge of the unknown parameters) of non-similar tests.

In the case of testing for autocorrelation in the dynamic linear regression model, our Monte Carlo results almost always support the new SA based near exact non-similar critical values. Only for one design matrix, $\mu \geq 0.9$ and the DW test, do we have sizes that are slightly above the nominal size. Apart from this case, none of the (near exact non-similar critical values based) sizes are significantly above the nominal size. Therefore, near exact non-similar critical values seem to be a good approximation to the exact non-similar critical values. For this study, we used the ASDA distribution based DW test and the large-sample based Durbin's t test.

Our study clearly indicated that neither the DW test nor the t test is best under the null in small and large samples. On many occasions, both the tests have approximately the same (higher than nominal) maximum sizes over the nuisance parameter space. In some cases, the DW test seemed better than Durbin's t test and vice-versa. Therefore, Inder's (1985) claim that the ASDA based approach is better than the large-sample based approach is questionable.

Based on our size results using near exact non-similar critical values, we conjecture that approximate non-similar critical values are not ideal for Durbin's t test because they may make the test's sizes unnecessarily low, and hence reduce the powers.

It is ideal to use an exact non-similar critical value for a non-similar test. However, it may be extremely difficult to obtain an exact non-similar critical value as outlined in this chapter. Our study in the case of the DW test and Durbin's t test suggests that near exact non-similar critical values are a good approximation to the exact non-similar critical values. Also, it seems that if we make sure that the SA algorithm achieves the global maximum size at step 2 of section 4.2.3, then we can expect near exact non-similar critical values to be successful in terms of controlling the sizes over the nuisance parameter space. It takes about the same time to calculate approximate non-similar critical values and SA based near exact non-similar critical values. In addition, the SA based approach allows one to assess the accuracy of any approximate distribution of the test statistic under the null. Furthermore, there may be design matrices or tests for which approximate non-similar critical values may not work well, whereas, near exact non-similar critical values may be helpful. Also as noted above, for some tests, the use of approximate non-similar critical values may unnecessarily make the test sizes low and hence reduce the powers. Therefore, based on our results, we recommend the use of SA based near exact non-similar critical values. Because, the SA based approach worked well for this problem, we apply it to a more complicated testing situation in the next chapter.

APPENDIX 4.1

Tables of Results for the DW test and Durbin's t test

Table 4.1: Maximum sizes of the DW and t tests together with the nuisance parameter values at this maximum size

		X8		*X9*		*X10*		*X11*	
		DW	t	DW	t	DW	t	DW	t
n_1	Max Size	0.199	0.232	0.093	0.281	0.109	0.114	0.090	0.128
	β_2	-0.651	0.006	-111.654	-0.029	0.134	0.157	0.009	-0.001
	β_3	-0.118	-0.014	9.792	-0.108	-0.974	-0.330	-	-
	β_4	0.137	-0.001	199.417	-0.840	0.987	0.221	-	-
	β_5	-	-	-	-	-	-	-	-
	μ	0.841	0.767	0.003	0.506	0.872	0.700	0.864	0.515
n_2	Max Size	0.205	0.190	0.073	0.057	0.105	0.114	0.104	0.099
	β_2	-0.680	0.368	-0.408	0.153	-0.591	0.157	0.000	-0.001
	β_3	0.356	0.039	0.002	0.180	0.713	-0.330	-	-
	β_4	0.323	-0.052	-	-0.151	-0.332	0.221	-	-
	β_5	-	-	-	-	-	-	-	-
	μ	0.946	0.911	0.975	0.272	0.963	0.700	0.979	0.900

		X12		*X13*		*X14*	
		DW	t	DW	t	DW	t
n_1	Max Size	0.077	0.085	0.082	0.065	0.127	0.121
	β_2	0.001	0.001	-0.041	0.304	0.441	-0.504
	β_3	-	-	1.043	0.307	-0.457	0.184
	β_4	-	-	-0.315	0.017	-	-
	β_5	-	-	0.005	0.000	-	-
	μ	0.975	0.513	0.899	0.831	0.952	0.500
n_2	Max Size	0.057	0.089	0.096	0.118	0.124	0.124
	β_2	0.501	0.000	1.904	0.809	0.143	0.044
	β_3	-	-	-2.396	-0.738	-0.007	-0.594
	β_4	-	-	8.211	3.298	-	-
	β_5	-	-	-0.008	-0.020	-	-
	μ	0.063	0.645	0.987	0.802	0.953	0.635

n_1 and n_2 are the small and large sample sizes, respectively

Table 4.2: Comparison of near exact and approximate non-similar critical values of the DW test

Critical values	*X8*	*X9*	*X10*	*X11*	*X12*	*X13*	*X14*
		$n = 20$		$n = 32$		$n = 30$	
Approx	1.4452	0.9663	1.3351	1.3827	1.3857	1.3017	1.3971
Near exact	1.4451	0.8949	1.3378	1.3823	1.3850	1.3020	1.4024
		$n = 60$		$n = 76$		$n = 60$	
Approx	1.5214	1.5886	1.5425	1.5771	1.5901	1.4961	1.5446
Near exact	1.5271	1.5887	1.5486	1.5797	1.6330	1.4954	1.5361

Table 4.3: Comparison of near exact and approximate non-similar critical values of Durbin's t test

Critical values	*X8*	*X9*	*X10*	*X11*	*X12*	*X13*	*X14*
		$n = 20$		$n = 32$		$n = 30$	
Approx	2.8808	2.8206	2.1864	2.1614	1.9293	2.1515	2.1919
Near exact	2.8350	2.8087	2.1635	2.1467	2.1286	2.1468	2.1772
		$n = 60$		$n = 76$		$n = 60$	
Approx	2.4480	1.6918	2.0512	1.9922	1.9240	2.1114	2.1224
Near exact	2.4500	1.6901	2.0361	1.9736	1.9453	2.0752	2.0932

Table 4.4: Sizes of the DW test at the 5% level for *X8*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
20	0.7	0.1	0.003	0.008	0.008	0.008	0.007	0.008	0.008	0.003	0.004
		0.3	0.008	0.013	0.012	0.012	0.011	0.013	0.013	0.009	0.009
		0.5	0.019	0.014	0.017	0.013	0.016	0.016	0.017	0.021	0.019
		0.7	0.036	0.010	0.017	0.009	0.017	0.010	0.013	0.030	0.034
		0.9	0.049	0.010	0.033	0.007	0.021	0.011	0.023	0.020	0.041
		0.99	0.050	0.006	0.008	0.044	0.007	0.032	0.018	0.020	0.011
	2	0.1		0.005	0.005	0.004	0.004	0.005	0.005	0.003	0.004
		0.3		0.008	0.008	0.010	0.009	0.008	0.008	0.009	0.009
		0.5		0.019	0.019	0.019	0.018	0.020	0.019	0.018	0.019
		0.7		0.032	0.034	0.028	0.031	0.031	0.032	0.036	0.035
		0.9		0.036	0.047	0.027	0.045	0.039	0.044	0.043	0.047
		0.99		0.024	0.025	0.049	0.023	0.048	0.040	0.042	0.039
	40	0.1		0.004	0.004	0.003	0.004	0.004	0.004	0.003	0.003
		0.3		0.008	0.008	0.008	0.009	0.008	0.008	0.008	0.008
		0.5		0.019	0.019	0.019	0.018	0.019	0.019	0.019	0.019
		0.7		0.036	0.037	0.036	0.035	0.037	0.037	0.036	0.036
		0.9		0.048	0.048	0.049	0.048	0.048	0.049	0.049	0.049
		0.99		0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
60	0.64	0.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5	0.002	0.003	0.003	0.005	0.005	0.003	0.003	0.002	0.002
		0.7	0.014	0.019	0.019	0.018	0.018	0.019	0.019	0.016	0.015
		0.9	0.046	0.015	0.019	0.011	0.017	0.013	0.014	0.033	0.043
		0.99	0.054	0.010	0.022	0.010	0.012	0.010	0.019	0.014	0.027
	1.35	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5		0.003	0.003	0.002	0.002	0.003	0.002	0.002	0.002
		0.7		0.017	0.017	0.017	0.016	0.017	0.017	0.015	0.014
		0.9		0.029	0.034	0.022	0.033	0.028	0.030	0.044	0.046
		0.99		0.015	0.040	0.013	0.026	0.022	0.035	0.031	0.042
	8	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5		0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		0.7		0.014	0.014	0.015	0.015	0.015	0.014	0.014	0.014
		0.9		0.046	0.046	0.047	0.047	0.046	0.046	0.046	0.046
		0.99		0.047	0.053	0.048	0.052	0.049	0.053	0.052	0.053

β_1, ..., β_9 are (0,0,0,0)', (0,1,1,1)', (0,-1,1,1)', (0,-1,1,-1)', (0,0,-1,-1)', (0,0,0,1)', (0,-1,0,1)', (0,1,-1,0)', (0,1,1,0)', respectively

Table 4.5: Sizes of the DW test at the 5% level for $X9$, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
20	0.25	0.1	0.001	0.011	0.012	0.002	0.008	0.005	0.007	0.004	0.003
		0.3	0.002	0.010	0.011	0.003	0.008	0.004	0.005	0.005	0.004
		0.5	0.004	0.013	0.014	0.005	0.011	0.005	0.007	0.009	0.009
		0.7	0.010	0.017	0.016	0.013	0.017	0.012	0.014	0.016	0.017
		0.9	0.027	0.013	0.014	0.022	0.015	0.028	0.025	0.018	0.017
		0.99	0.032	0.012	0.013	0.015	0.013	0.032	0.022	0.013	0.014
	0.54	0.1		0.003	0.004	0.002	0.003	0.001	0.003	0.003	0.001
		0.3		0.003	0.005	0.002	0.004	0.002	0.003	0.003	0.003
		0.5		0.007	0.010	0.005	0.007	0.004	0.005	0.006	0.005
		0.7		0.014	0.017	0.011	0.015	0.011	0.011	0.013	0.014
		0.9		0.018	0.018	0.026	0.018	0.027	0.027	0.025	0.023
		0.99		0.014	0.014	0.023	0.016	0.032	0.027	0.019	0.019
	6	0.1		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.3		0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		0.5		0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
		0.7		0.010	0.011	0.010	0.011	0.010	0.010	0.010	0.011
		0.9		0.028	0.026	0.028	0.027	0.027	0.028	0.028	0.027
		0.99		0.031	0.031	0.031	0.032	0.031	0.031	0.031	0.032
60	0.4	0.1	0.000	0.026	0.028	0.008	0.027	0.018	0.021	0.017	0.022
		0.3	0.001	0.026	0.027	0.011	0.027	0.019	0.021	0.019	0.023
		0.5	0.002	0.027	0.027	0.014	0.027	0.020	0.023	0.021	0.025
		0.7	0.013	0.027	0.027	0.021	0.028	0.023	0.025	0.024	0.026
		0.9	0.039	0.032	0.032	0.036	0.031	0.033	0.033	0.034	0.036
		0.99	0.049	0.030	0.030	0.044	0.031	0.041	0.040	0.037	0.035
	0.86	0.1		0.016	0.019	0.000	0.018	0.004	0.008	0.006	0.009
		0.3		0.018	0.021	0.002	0.020	0.007	0.011	0.010	0.012
		0.5		0.019	0.022	0.007	0.023	0.011	0.014	0.013	0.014
		0.7		0.022	0.024	0.016	0.025	0.018	0.019	0.018	0.020
		0.9		0.035	0.036	0.039	0.035	0.037	0.038	0.038	0.039
		0.99		0.038	0.037	0.049	0.038	0.045	0.045	0.044	0.043
	5	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.001
		0.5		0.004	0.004	0.002	0.004	0.003	0.003	0.003	0.003
		0.7		0.013	0.013	0.014	0.014	0.013	0.014	0.014	0.013
		0.9		0.040	0.039	0.039	0.039	0.038	0.041	0.040	0.039
		0.99		0.050	0.048	0.049	0.048	0.049	0.049	0.050	0.049

$\beta_1, \ldots, \beta_9$ are (0,0,0,0)', (0,1,1,1)', (0,-1,-1,-1)', (0,0,1,0)', (0,-1,0,-1)', (0,-1,1,1)', (0,1,-1,0)', (0,1,0,0)', (0,0,-1,-1)', respectively

Table 4.6: Sizes of the DW test at the 5% level for $X10$, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
20	0.2	0.1	0.006	0.016	0.014	0.010	0.007	0.008	0.014	0.015	0.016
		0.3	0.010	0.021	0.017	0.016	0.011	0.010	0.017	0.021	0.021
		0.5	0.020	0.022	0.018	0.019	0.021	0.022	0.018	0.023	0.021
		0.7	0.039	0.016	0.017	0.017	0.039	0.038	0.016	0.018	0.015
		0.9	0.051	0.012	0.013	0.013	0.045	0.049	0.013	0.014	0.012
		0.99	0.052	0.013	0.013	0.013	0.032	0.050	0.013	0.013	0.013
	0.5	0.1		0.007	0.009	0.007	0.005	0.006	0.009	0.009	0.008
		0.3		0.014	0.014	0.011	0.011	0.010	0.015	0.014	0.015
		0.5		0.024	0.021	0.020	0.019	0.021	0.021	0.024	0.025
		0.7		0.031	0.028	0.031	0.038	0.037	0.029	0.031	0.030
		0.9		0.031	0.031	0.035	0.049	0.052	0.029	0.034	0.029
		0.99		0.028	0.034	0.034	0.046	0.053	0.029	0.036	0.026
	3	0.1		0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.005
		0.3		0.010	0.009	0.010	0.010	0.010	0.009	0.009	0.009
		0.5		0.020	0.020	0.019	0.020	0.021	0.020	0.020	0.020
		0.7		0.036	0.038	0.038	0.039	0.039	0.037	0.037	0.037
		0.9		0.051	0.050	0.050	0.052	0.051	0.051	0.052	0.052
		0.99		0.048	0.052	0.053	0.051	0.053	0.052	0.049	0.049
60	0.3	0.1	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001
		0.3	0.000	0.002	0.002	0.002	0.001	0.001	0.002	0.003	0.003
		0.5	0.004	0.009	0.008	0.007	0.005	0.004	0.009	0.010	0.010
		0.7	0.019	0.024	0.023	0.023	0.020	0.020	0.022	0.024	0.023
		0.9	0.047	0.024	0.032	0.034	0.028	0.046	0.028	0.032	0.024
		0.99	0.054	0.019	0.021	0.038	0.019	0.030	0.021	0.025	0.019
	0.8	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.001	0.001	0.000	0.000	0.000	0.001	0.001	0.001
		0.5		0.005	0.003	0.004	0.004	0.003	0.004	0.005	0.005
		0.7		0.021	0.021	0.022	0.018	0.018	0.021	0.020	0.021
		0.9		0.038	0.044	0.044	0.046	0.047	0.042	0.043	0.038
		0.99		0.019	0.034	0.050	0.021	0.046	0.034	0.043	0.020
	4	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5		0.003	0.004	0.003	0.003	0.003	0.004	0.003	0.003
		0.7		0.019	0.020	0.020	0.019	0.019	0.020	0.019	0.019
		0.9		0.047	0.047	0.047	0.047	0.048	0.047	0.047	0.047
		0.99		0.041	0.052	0.053	0.045	0.053	0.051	0.051	0.044

$\beta_1, \ldots, \beta_9$ are (0,0,0,0)', (0,1,1,1)', (0,-1,1,1)', (0,-1,1,-1)', (0,0,-1,-1)', (0,0,0,1)', (0,-1,0,1)', (0,1,-1,0)', (0,1,1,0)', respectively

Table 4.7: Sizes of the DW test at the 5% level for *X11*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3
32	26.5	0.1	0.001	0.001	0.000
		0.3	0.001	0.002	0.002
		0.5	0.005	0.011	0.009
		0.7	0.019	0.026	0.023
		0.9	0.046	0.019	0.018
		0.99	0.050	0.015	0.015
	53	0.1		0.001	0.000
		0.3		0.001	0.001
		0.5		0.006	0.005
		0.7		0.021	0.024
		0.9		0.028	0.026
		0.99		0.018	0.018
	210	0.1		0.001	0.001
		0.3		0.001	0.001
		0.5		0.005	0.005
		0.7		0.019	0.019
		0.9		0.043	0.044
		0.99		0.039	0.039
76	57	0.1	0.000	0.000	0.000
		0.3	0.000	0.000	0.000
		0.5	0.001	0.001	0.002
		0.7	0.011	0.016	0.013
		0.9	0.038	0.031	0.032
		0.99	0.051	0.020	0.020
	116	0.1		0.000	0.000
		0.3		0.000	0.000
		0.5		0.001	0.001
		0.7		0.013	0.011
		0.9		0.036	0.038
		0.99		0.024	0.024
	465	0.1		0.000	0.000
		0.3		0.000	0.000
		0.5		0.001	0.001
		0.7		0.011	0.011
		0.9		0.038	0.039
		0.99		0.044	0.044

$\beta_1, \beta_2, \beta_3$ are (0,0)', (0,1)', (0,-1)', respectively

Table 4.8: Sizes of the DW test at the 5% level for *X12*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3
32	31	0.1	0.001	0.031	0.036
		0.3	0.001	0.028	0.034
		0.5	0.005	0.027	0.032
		0.7	0.022	0.027	0.033
		0.9	0.046	0.031	0.033
		0.99	0.051	0.032	0.036
	64.2	0.1		0.016	0.020
		0.3		0.019	0.023
		0.5		0.022	0.024
		0.7		0.028	0.031
		0.9		0.037	0.038
		0.99		0.038	0.043
	250	0.1		0.002	0.002
		0.3		0.003	0.003
		0.5		0.008	0.007
		0.7		0.022	0.023
		0.9		0.045	0.043
		0.99		0.051	0.050
76	72	0.1	0.000	0.020	0.024
		0.3	0.000	0.026	0.027
		0.5	0.004	0.036	0.036
		0.7	0.028	0.042	0.043
		0.9	0.067	0.044	0.043
		0.99	0.075	0.042	0.042
	147	0.1		0.004	0.005
		0.3		0.008	0.010
		0.5		0.022	0.021
		0.7		0.042	0.039
		0.9		0.046	0.048
		0.99		0.042	0.043
	580	0.1		0.000	0.000
		0.3		0.001	0.000
		0.5		0.005	0.007
		0.7		0.031	0.029
		0.9		0.062	0.063
		0.99		0.058	0.060

$\beta_1, \beta_2, \beta_3$ are (0,0)', (0,1)', (0,-1)', respectively

Table 4.9: Sizes of the DW test at the 5% level for *X13* , and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
30	0.5	0.1	0.000	0.018	0.016	0.018	0.018	0.000	0.016	0.015	0.018
		0.3	0.001	0.021	0.020	0.022	0.021	0.001	0.020	0.021	0.021
		0.5	0.005	0.023	0.022	0.023	0.022	0.005	0.022	0.022	0.022
		0.7	0.020	0.021	0.019	0.020	0.020	0.022	0.019	0.019	0.021
		0.9	0.045	0.020	0.020	0.020	0.021	0.036	0.021	0.020	0.020
		0.99	0.050	0.021	0.021	0.020	0.021	0.027	0.021	0.021	0.021
	5	0.1		0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.001
		0.3		0.001	0.002	0.001	0.001	0.001	0.002	0.002	0.001
		0.5		0.007	0.007	0.007	0.007	0.005	0.007	0.008	0.007
		0.7		0.024	0.024	0.025	0.022	0.021	0.024	0.024	0.023
		0.9		0.034	0.032	0.033	0.035	0.045	0.034	0.031	0.034
		0.99		0.025	0.023	0.023	0.027	0.048	0.024	0.022	0.027
	150	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.5		0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
		0.7		0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020
		0.9		0.045	0.045	0.045	0.045	0.045	0.045	0.046	0.045
		0.99		0.050	0.049	0.050	0.050	0.050	0.049	0.049	0.050
60	10	0.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5	0.001	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001
		0.7	0.013	0.014	0.014	0.014	0.014	0.012	0.014	0.014	0.014
		0.9	0.043	0.038	0.035	0.037	0.038	0.043	0.036	0.035	0.038
		0.99	0.049	0.028	0.029	0.028	0.028	0.051	0.029	0.028	0.028
	20	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.7		0.013	0.013	0.013	0.013	0.012	0.013	0.013	0.013
		0.9		0.040	0.039	0.040	0.041	0.043	0.039	0.038	0.040
		0.99		0.040	0.040	0.039	0.042	0.050	0.040	0.040	0.041
	250	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		0.7		0.012	0.012	0.012	0.012	0.013	0.012	0.012	0.012
		0.9		0.043	0.043	0.043	0.043	0.043	0.043	0.042	0.043
		0.99		0.050	0.050	0.050	0.049	0.049	0.050	0.049	0.050

β_1, ..., β_9 are (0,0,0,0,0)', (0,1,1,1,1)', (0,-1,-1,-1,-1)', (0,-1,1,1,1)', (0,1,-1,-1,1)', (0,1,0,1,0)', (0,-1,0,-1,-1)', (0,0,-1,-1,-1)', (0,1,0,0,1)', respectively

Table 4.10: Sizes of the DW test at the 5% level for *X14*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
30	0.035	0.1	0.000	0.006	0.005	0.002	0.004	0.003	0.002	0.003	0.002
		0.3	0.002	0.011	0.011	0.007	0.009	0.007	0.005	0.007	0.006
		0.5	0.007	0.017	0.018	0.015	0.017	0.016	0.016	0.014	0.015
		0.7	0.025	0.018	0.020	0.022	0.022	0.021	0.024	0.021	0.022
		0.9	0.048	0.015	0.016	0.019	0.019	0.017	0.018	0.020	0.020
		0.99	0.051	0.014	0.014	0.016	0.014	0.014	0.014	0.016	0.016
	0.081	0.1		0.002	0.001	0.001	0.001	0.000	0.001	0.001	0.000
		0.3		0.004	0.003	0.003	0.003	0.002	0.002	0.003	0.003
		0.5		0.014	0.012	0.009	0.011	0.010	0.010	0.009	0.008
		0.7		0.025	0.024	0.027	0.026	0.026	0.025	0.024	0.025
		0.9		0.019	0.020	0.036	0.034	0.026	0.029	0.032	0.034
		0.99		0.014	0.015	0.028	0.024	0.016	0.017	0.024	0.023
	0.68	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		0.5		0.008	0.007	0.007	0.008	0.008	0.007	0.008	0.007
		0.7		0.025	0.025	0.026	0.026	0.026	0.026	0.025	0.025
		0.9		0.046	0.046	0.048	0.048	0.047	0.047	0.048	0.047
		0.99		0.045	0.045	0.051	0.050	0.048	0.047	0.050	0.050
60	0.085	0.1	0.000	0.001	0.001	0.001	0.001	0.000	0.000	0.001	0.001
		0.3	0.000	0.005	0.004	0.005	0.004	0.001	0.001	0.003	0.003
		0.5	0.002	0.014	0.012	0.013	0.013	0.005	0.005	0.011	0.012
		0.7	0.017	0.019	0.017	0.020	0.018	0.021	0.019	0.020	0.018
		0.9	0.046	0.018	0.017	0.018	0.016	0.020	0.021	0.019	0.017
		0.99	0.045	0.015	0.015	0.016	0.015	0.015	0.015	0.016	0.015
	0.185	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.001
		0.5		0.006	0.004	0.006	0.005	0.003	0.003	0.005	0.005
		0.7		0.021	0.018	0.021	0.020	0.019	0.017	0.021	0.018
		0.9		0.022	0.024	0.024	0.023	0.030	0.029	0.028	0.027
		0.99		0.016	0.016	0.019	0.017	0.016	0.017	0.020	0.018
	0.95	0.1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.3		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		0.5		0.003	0.003	0.003	0.002	0.002	0.002	0.003	0.002
		0.7		0.017	0.016	0.017	0.017	0.017	0.017	0.017	0.017
		0.9		0.043	0.040	0.042	0.043	0.044	0.043	0.043	0.042
		0.99		0.033	0.032	0.043	0.043	0.041	0.040	0.044	0.042

$\beta_1, \ldots, \beta_9$ are (0,0,0)', (0,1,1)', (0,-1,-1)', (0,-1,1)', (0,1,-1)', (0,1,0)', (0,-1,0)', (0,0,1)', (0,0,-1)', respectively

Table 4.11: Sizes of Durbin's t test at the 5% level for $X8$, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
20	0.7	0.1	0.012	0.005	0.007	0.005	0.006	0.006	0.007	0.012	0.011
		0.3	0.025	0.007	0.010	0.005	0.010	0.007	0.008	0.022	0.020
		0.5	0.040	0.004	0.010	0.002	0.009	0.004	0.005	0.031	0.026
		0.7	0.047	0.002	0.009	0.000	0.006	0.002	0.003	0.022	0.021
		0.9	0.050	0.001	0.016	0.000	0.003	0.001	0.008	0.002	0.006
	2	0.1		0.014	0.015	0.012	0.013	0.015	0.015	0.012	0.011
		0.3		0.021	0.023	0.021	0.022	0.022	0.023	0.026	0.025
		0.5		0.032	0.035	0.025	0.031	0.032	0.032	0.039	0.037
		0.7		0.031	0.043	0.020	0.039	0.034	0.035	0.045	0.041
		0.9		0.015	0.044	0.009	0.033	0.027	0.040	0.034	0.030
	40	0.1		0.012	0.012	0.011	0.012	0.011	0.011	0.011	0.012
		0.3		0.024	0.025	0.026	0.027	0.024	0.024	0.025	0.025
		0.5		0.040	0.040	0.039	0.039	0.039	0.040	0.040	0.038
		0.7		0.047	0.047	0.048	0.048	0.048	0.048	0.047	0.047
		0.9		0.050	0.050	0.053	0.053	0.050	0.049	0.051	0.051
60	0.64	0.1	0.022	0.020	0.021	0.015	0.016	0.021	0.021	0.021	0.023
		0.3	0.036	0.030	0.033	0.029	0.030	0.031	0.031	0.037	0.038
		0.5	0.039	0.030	0.030	0.028	0.030	0.031	0.032	0.041	0.042
		0.7	0.042	0.018	0.023	0.013	0.019	0.020	0.022	0.039	0.040
		0.9	0.049	0.003	0.005	0.001	0.004	0.002	0.003	0.018	0.040
		0.99	0.051	0.001	0.007	0.001	0.001	0.001	0.004	0.001	0.010
	1.35	0.1		0.022	0.022	0.020	0.019	0.022	0.021	0.021	0.023
		0.3		0.036	0.036	0.037	0.036	0.035	0.036	0.037	0.037
		0.5		0.037	0.039	0.038	0.037	0.039	0.040	0.041	0.040
		0.7		0.035	0.036	0.033	0.036	0.036	0.036	0.041	0.041
		0.9		0.016	0.024	0.008	0.023	0.015	0.017	0.039	0.049
		0.99		0.002	0.028	0.001	0.009	0.005	0.020	0.012	0.029
	8	0.1		0.022	0.022	0.023	0.022	0.023	0.022	0.021	0.023
		0.3		0.036	0.036	0.038	0.037	0.037	0.036	0.037	0.036
		0.5		0.038	0.038	0.039	0.039	0.038	0.039	0.040	0.039
		0.7		0.042	0.043	0.042	0.041	0.042	0.042	0.042	0.042
		0.9		0.047	0.048	0.047	0.049	0.047	0.047	0.050	0.050
		0.99		0.043	0.050	0.038	0.048	0.045	0.049	0.048	0.050

$\beta_1, \ldots, \beta_9$ are (0,0,0,0)', (0,1,1,1)', (0,-1,1,1)', (0,-1,1,-1)', (0,0,-1,-1)', (0,0,0,1)', (0,-1,0,1)', (0,1,-1,0)', (0,1,1,0)', respectively

Table 4.12: Sizes of Durbin's t test at the 5% level for $X9$, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
20	0.25	0.1	0.041	0.021	0.015	0.031	0.016	0.019	0.019	0.020	0.028
		0.3	0.050	0.031	0.029	0.041	0.027	0.031	0.032	0.032	0.041
		0.5	0.049	0.033	0.034	0.046	0.033	0.040	0.036	0.037	0.046
		0.7	0.045	0.017	0.022	0.040	0.025	0.038	0.034	0.033	0.039
		0.9	0.040	0.005	0.005	0.023	0.008	0.032	0.028	0.012	0.013
		0.99	0.036	0.004	0.004	0.009	0.004	0.030	0.017	0.005	0.005
	0.54	0.1		0.032	0.026	0.040	0.027	0.027	0.031	0.034	0.035
		0.3		0.042	0.043	0.046	0.040	0.043	0.040	0.043	0.045
		0.5		0.042	0.042	0.048	0.043	0.046	0.045	0.046	0.049
		0.7		0.038	0.036	0.045	0.038	0.044	0.042	0.042	0.045
		0.9		0.016	0.015	0.037	0.022	0.040	0.038	0.031	0.031
		0.99		0.005	0.006	0.021	0.010	0.036	0.031	0.014	0.015
	6	0.1		0.044	0.042	0.042	0.043	0.044	0.043	0.042	0.042
		0.3		0.046	0.051	0.050	0.049	0.051	0.047	0.047	0.049
		0.5		0.049	0.051	0.048	0.050	0.048	0.050	0.049	0.049
		0.7		0.046	0.046	0.045	0.046	0.046	0.044	0.045	0.045
		0.9		0.040	0.041	0.041	0.040	0.040	0.040	0.040	0.041
		0.99		0.035	0.036	0.034	0.036	0.035	0.034	0.035	0.036
60	0.4	0.1	0.049	0.018	0.016	0.030	0.016	0.020	0.020	0.022	0.022
		0.3	0.050	0.018	0.016	0.026	0.016	0.017	0.017	0.018	0.019
		0.5	0.037	0.018	0.017	0.022	0.016	0.017	0.016	0.017	0.018
		0.7	0.031	0.018	0.017	0.021	0.018	0.019	0.016	0.017	0.019
		0.9	0.039	0.020	0.020	0.029	0.021	0.025	0.023	0.024	0.025
		0.99	0.040	0.018	0.019	0.032	0.020	0.028	0.028	0.026	0.024
	0.86	0.1		0.021	0.025	0.039	0.026	0.030	0.030	0.029	0.033
		0.3		0.017	0.023	0.035	0.023	0.022	0.024	0.025	0.027
		0.5		0.017	0.019	0.029	0.020	0.019	0.019	0.021	0.022
		0.7		0.017	0.019	0.025	0.020	0.020	0.019	0.021	0.023
		0.9		0.025	0.026	0.036	0.026	0.030	0.030	0.031	0.032
		0.99		0.026	0.025	0.038	0.027	0.035	0.034	0.033	0.033
	5	0.1		0.040	0.044	0.047	0.047	0.046	0.047	0.049	0.050
		0.3		0.039	0.043	0.049	0.043	0.045	0.045	0.046	0.049
		0.5		0.030	0.033	0.037	0.033	0.035	0.032	0.032	0.037
		0.7		0.027	0.029	0.031	0.030	0.030	0.030	0.030	0.030
		0.9		0.038	0.037	0.040	0.037	0.040	0.039	0.039	0.038
		0.99		0.039	0.037	0.040	0.039	0.040	0.040	0.040	0.040

$\beta_1, \ldots, \beta_9$ are (0,0,0,0)', (0,1,1,1)', (0,-1,-1,-1)', (0,0,1,0)', (0,-1,0,-1)', (0,-1,1,1)', (0,1,-1,0)', (0,1,0,0)', (0,0,-1,-1)', respectively

Table 4.13: Sizes of Durbin's t test at the 5% level for $X10$, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
20	0.2	0.1	0.014	0.007	0.004	0.007	0.011	0.011	0.004	0.006	0.006
		0.3	0.024	0.008	0.004	0.008	0.016	0.019	0.004	0.008	0.008
		0.5	0.039	0.006	0.004	0.007	0.028	0.028	0.003	0.007	0.005
		0.7	0.052	0.001	0.003	0.004	0.033	0.038	0.002	0.003	0.002
		0.9	0.047	0.000	0.002	0.003	0.025	0.038	0.002	0.001	0.000
		0.99	0.042	0.000	0.002	0.003	0.012	0.032	0.002	0.002	0.000
	0.5	0.1		0.013	0.009	0.008	0.013	0.011	0.009	0.014	0.014
		0.3		0.026	0.015	0.016	0.023	0.023	0.016	0.023	0.024
		0.5		0.030	0.020	0.024	0.038	0.036	0.020	0.027	0.028
		0.7		0.026	0.021	0.026	0.048	0.047	0.021	0.026	0.026
		0.9		0.016	0.018	0.022	0.044	0.046	0.014	0.023	0.016
		0.99		0.009	0.021	0.021	0.032	0.041	0.014	0.026	0.008
	3	0.1		0.012	0.012	0.014	0.014	0.012	0.012	0.012	0.012
		0.3		0.025	0.024	0.024	0.025	0.024	0.024	0.024	0.024
		0.5		0.041	0.036	0.038	0.039	0.038	0.036	0.040	0.041
		0.7		0.050	0.047	0.049	0.050	0.053	0.047	0.051	0.051
		0.9		0.048	0.047	0.048	0.048	0.047	0.046	0.048	0.048
		0.99		0.041	0.042	0.044	0.042	0.041	0.044	0.043	0.042
60	0.3	0.1	0.028	0.023	0.017	0.018	0.023	0.026	0.019	0.023	0.023
		0.3	0.043	0.032	0.022	0.026	0.034	0.038	0.022	0.031	0.030
		0.5	0.051	0.034	0.022	0.029	0.045	0.044	0.022	0.032	0.029
		0.7	0.050	0.024	0.020	0.026	0.047	0.047	0.020	0.024	0.021
		0.9	0.046	0.008	0.016	0.022	0.017	0.044	0.014	0.020	0.008
		0.99	0.043	0.004	0.005	0.023	0.004	0.013	0.006	0.010	0.004
	0.8	0.1		0.027	0.025	0.026	0.026	0.030	0.025	0.025	0.025
		0.3		0.041	0.037	0.040	0.041	0.043	0.036	0.041	0.041
		0.5		0.046	0.044	0.046	0.048	0.049	0.043	0.047	0.046
		0.7		0.046	0.043	0.047	0.049	0.049	0.042	0.044	0.045
		0.9		0.025	0.037	0.038	0.038	0.045	0.036	0.039	0.025
		0.99		0.005	0.017	0.039	0.007	0.036	0.021	0.028	0.005
	4	0.1		0.027	0.029	0.029	0.028	0.028	0.029	0.027	0.027
		0.3		0.043	0.044	0.045	0.043	0.043	0.043	0.041	0.042
		0.5		0.051	0.050	0.050	0.050	0.051	0.051	0.049	0.050
		0.7		0.051	0.050	0.051	0.050	0.050	0.050	0.051	0.051
		0.9		0.044	0.045	0.046	0.047	0.045	0.046	0.046	0.044
		0.99		0.025	0.042	0.044	0.034	0.043	0.042	0.042	0.028

$\beta_1, \ldots, \beta_9$ are (0,0,0,0)', (0,1,1,1)', (0,-1,1,1)', (0,-1,1,-1)', (0,0,-1,-1)', (0,0,0,1)', (0,-1,0,1)', (0,1,-1,0)', (0,1,1,0)', respectively

Table 4.14: Sizes of Durbin's t test at the 5% level for *X11*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3
32	26.5	0.1	0.031	0.022	0.021
		0.3	0.044	0.029	0.027
		0.5	0.048	0.024	0.023
		0.7	0.050	0.017	0.015
		0.9	0.047	0.003	0.004
		0.99	0.043	0.002	0.002
	53	0.1		0.028	0.023
		0.3		0.040	0.038
		0.5		0.041	0.037
		0.7		0.033	0.032
		0.9		0.011	0.010
		0.99		0.003	0.003
	210	0.1		0.028	0.026
		0.3		0.047	0.042
		0.5		0.051	0.045
		0.7		0.052	0.047
		0.9		0.042	0.043
		0.99		0.025	0.025
76	57	0.1	0.039	0.033	0.035
		0.3	0.048	0.043	0.047
		0.5	0.051	0.044	0.049
		0.7	0.051	0.040	0.041
		0.9	0.051	0.024	0.025
		0.99	0.048	0.010	0.009
	116	0.1		0.035	0.037
		0.3		0.047	0.049
		0.5		0.050	0.050
		0.7		0.047	0.049
		0.9		0.037	0.039
		0.99		0.013	0.014
	465	0.1		0.038	0.040
		0.3		0.049	0.049
		0.5		0.052	0.050
		0.7		0.050	0.053
		0.9		0.050	0.048
		0.99		0.039	0.039

β_1, β_2, β_3 are (0,0)', (0,1)', (0,-1)', respectively

Table 4.15: Sizes of Durbin's t test at the 5% level for *X12*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3
32	31	0.1	0.022	0.008	0.010
		0.3	0.028	0.008	0.009
		0.5	0.031	0.008	0.009
		0.7	0.031	0.008	0.008
		0.9	0.029	0.009	0.009
		0.99	0.028	0.007	0.009
	64.2	0.1		0.010	0.012
		0.3		0.011	0.012
		0.5		0.012	0.013
		0.7		0.012	0.013
		0.9		0.014	0.015
		0.99		0.015	0.015
	250	0.1		0.020	0.021
		0.3		0.026	0.024
		0.5		0.030	0.024
		0.7		0.027	0.026
		0.9		0.028	0.028
		0.99		0.025	0.026
76	72	0.1	0.037	0.017	0.018
		0.3	0.044	0.016	0.020
		0.5	0.048	0.017	0.019
		0.7	0.047	0.017	0.019
		0.9	0.047	0.014	0.016
		0.99	0.040	0.013	0.013
	147	0.1		0.021	0.025
		0.3		0.025	0.027
		0.5		0.025	0.026
		0.7		0.025	0.026
		0.9		0.019	0.021
		0.99		0.013	0.015
	580	0.1		0.035	0.036
		0.3		0.044	0.044
		0.5		0.043	0.045
		0.7		0.043	0.043
		0.9		0.038	0.040
		0.99		0.026	0.026

$\beta_1, \beta_2, \beta_3$ are (0,0)', (0,1)', (0,-1)', respectively

Table 4.16: Sizes of Durbin's t test at the 5% level for *X13*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
30	0.5	0.1	0.022	0.012	0.011	0.012	0.012	0.022	0.011	0.011	0.012
		0.3	0.032	0.009	0.009	0.011	0.010	0.033	0.009	0.009	0.010
		0.5	0.041	0.007	0.007	0.007	0.006	0.042	0.007	0.007	0.006
		0.7	0.044	0.005	0.005	0.005	0.005	0.045	0.005	0.005	0.005
		0.9	0.051	0.004	0.005	0.004	0.005	0.031	0.005	0.005	0.004
		0.99	0.048	0.005	0.005	0.005	0.005	0.010	0.005	0.005	0.005
	5	0.1		0.021	0.024	0.021	0.020	0.022	0.024	0.024	0.020
		0.3		0.033	0.032	0.034	0.034	0.032	0.032	0.031	0.034
		0.5		0.039	0.040	0.039	0.037	0.041	0.041	0.041	0.038
		0.7		0.039	0.040	0.040	0.038	0.045	0.040	0.039	0.039
		0.9		0.021	0.018	0.017	0.022	0.050	0.018	0.018	0.022
		0.99		0.008	0.009	0.007	0.010	0.049	0.009	0.008	0.009
	150	0.1		0.022	0.023	0.022	0.022	0.022	0.023	0.023	0.022
		0.3		0.033	0.032	0.033	0.033	0.032	0.032	0.032	0.033
		0.5		0.041	0.041	0.041	0.041	0.041	0.041	0.040	0.041
		0.7		0.044	0.043	0.044	0.044	0.044	0.043	0.044	0.044
		0.9		0.050	0.049	0.048	0.050	0.050	0.049	0.049	0.050
		0.99		0.046	0.049	0.048	0.047	0.047	0.049	0.049	0.046
60	10	0.1	0.026	0.031	0.025	0.030	0.031	0.026	0.025	0.025	0.031
		0.3	0.035	0.036	0.034	0.037	0.036	0.035	0.034	0.034	0.036
		0.5	0.042	0.040	0.040	0.042	0.040	0.042	0.039	0.040	0.040
		0.7	0.045	0.044	0.042	0.043	0.044	0.045	0.043	0.042	0.044
		0.9	0.053	0.034	0.035	0.034	0.036	0.053	0.035	0.034	0.035
		0.99	0.048	0.018	0.017	0.017	0.018	0.050	0.017	0.017	0.018
	20	0.1		0.028	0.025	0.028	0.028	0.026	0.025	0.025	0.028
		0.3		0.039	0.034	0.038	0.038	0.035	0.034	0.034	0.039
		0.5		0.043	0.042	0.043	0.043	0.042	0.043	0.042	0.043
		0.7		0.045	0.045	0.046	0.045	0.045	0.046	0.045	0.045
		0.9		0.047	0.047	0.047	0.047	0.053	0.047	0.047	0.047
		0.99		0.032	0.030	0.030	0.032	0.050	0.030	0.030	0.033
	250	0.1		0.026	0.026	0.026	0.026	0.026	0.026	0.026	0.026
		0.3		0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035
		0.5		0.043	0 042	0.043	0.042	0.042	0.042	0.042	0.042
		0.7		0.045	0.044	0.045	0.045	0.045	0.044	0.044	0.045
		0.9		0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053
		0.99		0.049	0.048	0.048	0.049	0.048	0.048	0.048	0.049

$\beta_{1,\ldots,}\beta_9$ are (0,0,0,0,0)', (0,1,1,1,1)', (0,-1,-1,-1,-1)', (0,-1,1,1,1)', (0,1,-1,-1,1)', (0,1,0,1,0)', (0,-1,0,-1,-1)', (0,0,-1,-1,-1)', (0,1,0,0,1)', respectively

Table 4.17: Sizes of Durbin's t test at the 5% level for *X14*, and selected β, σ, and μ values, using near exact non-similar critical values

n	σ	μ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
30	0.035	0.1	0.025	0.010	0.009	0.013	0.011	0.013	0.014	0.014	0.012
		0.3	0.041	0.008	0.008	0.014	0.012	0.015	0.015	0.015	0.014
		0.5	0.051	0.007	0.007	0.010	0.011	0.012	0.011	0.014	0.014
		0.7	0.046	0.005	0.005	0.009	0.008	0.006	0.007	0.012	0.010
		0.9	0.041	0.002	0.003	0.005	0.004	0.002	0.003	0.006	0.006
		0.99	0.037	0.002	0.002	0.003	0.003	0.002	0.002	0.003	0.003
	0.081	0.1		0.017	0.014	0.022	0.018	0.021	0.020	0.020	0.019
		0.3		0.020	0.018	0.031	0.028	0.029	0.031	0.029	0.030
		0.5		0.017	0.018	0.032	0.030	0.028	0.030	0.031	0.034
		0.7		0.012	0.013	0.028	0.027	0.021	0.022	0.029	0.028
		0.9		0.005	0.004	0.019	0.018	0.008	0.007	0.017	0.016
		0.99		0.002	0.003	0.007	0.007	0.003	0.004	0.008	0.006
	0.68	0.1		0.025	0.025	0.026	0.025	0.026	0.026	0.025	0.026
		0.3		0.040	0.039	0.043	0.039	0.039	0.038	0.041	0.039
		0.5		0.048	0.046	0.048	0.049	0.047	0.048	0.051	0.051
		0.7		0.043	0.046	0.047	0.046	0.046	0.046	0.046	0.047
		0.9		0.035	0.036	0.042	0.038	0.039	0.040	0.039	0.041
		0.99		0.025	0.026	0.035	0.036	0.031	0.032	0.035	0.034
60	0.085	0.1	0.031	0.016	0.013	0.015	0.016	0.023	0.016	0.019	0.017
		0.3	0.046	0.017	0.015	0.017	0.018	0.031	0.025	0.019	0.017
		0.5	0.050	0.016	0.014	0.015	0.014	0.030	0.027	0.018	0.016
		0.7	0.052	0.011	0.011	0.011	0.009	0.022	0.021	0.013	0.011
		0.9	0.050	0.005	0.006	0.006	0.007	0.008	0.009	0.007	0.006
		0.99	0.040	0.004	0.004	0.005	0.004	0.004	0.005	0.004	0.004
	0.185	0.1		0.022	0.019	0.022	0.022	0.029	0.024	0.027	0.025
		0.3		0.029	0.027	0.030	0.029	0.041	0.035	0.033	0.033
		0.5		0.029	0.029	0.029	0.030	0.044	0.039	0.033	0.032
		0.7		0.024	0.025	0.024	0.025	0.042	0.040	0.030	0.030
		0.9		0.011	0.012	0.012	0.010	0.018	0.020	0.015	0.014
		0.99		0.005	0.005	0.007	0.007	0.005	0.006	0.008	0.008
	0.95	0.1		0.030	0.030	0.030	0.032	0.031	0.030	0.029	0.031
		0.3		0.044	0.045	0.044	0.050	0.047	0.045	0.043	0.047
		0.5		0.050	0.050	0.047	0.051	0.051	0.050	0.049	0.050
		0.7		0.051	0.048	0.049	0.051	0.053	0.052	0.050	0.049
		0.9		0.043	0.042	0.044	0.046	0.049	0.048	0.045	0.047
		0.99		0.021	0.021	0.034	0.033	0.029	0.031	0.033	0.031

$\beta_1,\ldots,\beta_9$ are (0,0,0)', (0,1,1)', (0,-1,-1)', (0,-1,1)', (0,1,-1)', (0,1,0)', (0,-1,0)', (0,0,1)', (0,0,-1)', respectively

CHAPTER 5

TESTING FOR A STATIC LINEAR REGRESSION MODEL WITH AR(1) ERRORS AGAINST A DYNAMIC LINEAR REGRESSION MODEL WITH WHITE NOISE ERRORS

5.1 Introduction

In this chapter, we deal with a more complicated testing problem, namely, testing for a static linear regression model with AR(1) errors against a dynamic linear regression model with white noise errors. If the DW statistic for autocorrelation is significant in the context of a linear regression model, some prefer to work with static linear regression models with autocorrelated errors while others favour dynamic linear regression models with white noise errors. That is, the dynamic part of the model can be incorporated into the model through the error term or through lagged dependent variable regressors[32]. Both the approaches are plausible. However, it would be desirable to have a powerful testing procedure to distinguish these two approaches because a correct model specification is important for forecasting purposes and also

[32] This choice is similar to one discussed in the unit root testing literature. That is, some researchers test for unit roots in errors and others test for it in the mean of their models (see Stock (1994) and Silvapulle (1992)).

for the purpose of further inference. As noted in Chapter 2, King and Rankin (1993) point out that when the DW test in a static linear model is significant, and the true model is the dynamic linear model with a large autoregressive parameter, a substantial loss in accuracy of prediction can occur if one proceeds to correct for AR(1) disturbances without first checking for the possibility of a dynamic linear model with white noise errors.

Some may argue that the correct model (from the two mentioned above) can be chosen by observing the significance of the OLS estimates and $\overline{R}^2$. However, Griliches (1967) and Giles (1975) point out that while the true model which generated the given sample is the regression model with AR(1) errors, if one fits the dynamic linear regression model with white noise errors to the sample by OLS then it is likely to explain the data rather well. Therefore, one can mistakenly select a wrong model by using the OLS approach and this may lead to misleading inferences.

In Chapter 2, we outlined some situations where the null and the alternative models of interest in this chapter may not be distinguished at all. We purposely selected this difficult problem in order to see which test succeeds in every case, if such a test exists, then we can recommend that test for other complicated testing situations with greater confidence.

For this problem, the nuisance parameters cannot be avoided, thus tests applied can be expected to be non-similar. The literature review of Chapter 2 reveals that, in the presence of nuisance parameters, marginal likelihood based tests perform better than conventional likelihood based tests in finite samples. Therefore, we consider marginal likelihood based tests in this chapter. In particular, we compare marginal likelihood based g tests with marginal likelihood based one-sided LR, LM and W tests, in terms of size and power properties obtained via Monte Carlo experiments. Because the classical tests are specially designed for nested testing, they are applied to test for the significance of the dynamic regressor coefficient of a dynamic linear regression model with AR(1) errors. To the author's knowledge, this is the first study that investigates the finite-sample performance of marginal likelihood based classical tests in the dynamic linear regression model with AR(1) errors. Grose (1998) applied

these tests to the dynamic linear model with white noise errors. The size and power comparisons of this chapter are based on near exact non-similar critical values of the tests obtained using the simulated annealing (SA) algorithm.

The plan of this chapter is as follows. The models and assumptions are discussed in section 5.2. Section 5.3 describes the marginal likelihood function of the parameters of interest and section 5.4 discusses the marginal likelihood based tests. Section 5.5 discusses how to obtain exact (and near exact) non-similar critical values of these tests. In this section, we also discuss how (4.2.4) can be generalised in order to obtain exact critical values of the g test. Section 5.6 presents the details of the Monte Carlo experiment and summarizes the main findings. Finally, concluding remarks are given in section 5.7.

5.2 Models and Hypotheses of Interest

The non-nested models (or hypotheses) of interest in this chapter are

$$H_0: y_t = x_t'\beta + u_t, \quad u_t = \rho u_{t-1} + e_t, \quad t = 1,\ldots,n, \tag{5.2.1}$$

and

$$H_a: y_t = \mu y_{t-1} + x_t'\beta + e_t, \quad t = 2,\ldots,n, \tag{5.2.2}$$

where y_t is the dependent variable at time t, x_t is a $k\times 1$ vector of non-stochastic regressors at time t, β is a $k\times 1$ vector of coefficients, u_t is the disturbance term with $0 \le \rho < 1$, $e_t \sim IN(0,\sigma^2)$ and $0 \le \mu < 1$.

Because positive autocorrelation is more likely in practice we consider this case only, however, the discussion of this chapter also applies to the negative autocorrelation

case as well. In the case of the classical tests, as mentioned earlier, we consider the dynamic linear regression model with AR(1) errors, which can be regarded as the general model. Because the models given by (5.2.1) and (5.2.2) are nested within the general model, it is appropriate to discuss the theory for the general model and this theory can be applied to each of the nested models as special cases when needed.

First consider the dynamic linear regression model with AR(1) errors,

$$y_t = \mu y_{t-1} + x_t'\beta + u_t, \quad t = 2,\ldots,n, \tag{5.2.3}$$

where $u_t = \rho u_{t-1} + e_t$, $0 \leq \rho < 1$, $e_t \sim IN(0,\sigma^2)$ and $0 \leq \mu < 1$.

As discussed in the previous chapter, y_1 and u_1 of model (5.2.3) can be written as

$$y_1 = \frac{x_1'\beta}{(1-\mu)} + d_1 u_1, \tag{5.2.4}$$

and

$$u_1 = \frac{e_1}{(1-\rho^2)^{1/2}}, \tag{5.2.5}$$

where $d_1^2 = \dfrac{(1+\mu\rho)}{(1-\mu\rho)(1-\mu^2)}$. Now model (5.2.3) is completely specified by equations (5.2.3), (5.2.4) and (5.2.5).

Model (5.2.3) can be written in matrix form as

$$\Gamma(\mu)y = X\beta + D(\mu,\rho)u, \tag{5.2.6}$$

where $\Gamma(\mu)$ is the $n \times n$ matrix

$$\Gamma(\mu)=\begin{bmatrix} (1-\mu) & 0 & 0 & . & . & . & 0 \\ -\mu & 1 & 0 & & & & . \\ 0 & -\mu & 1 & & & & . \\ . & & & . & & & . \\ . & & & & . & & . \\ . & & & & & 1 & 0 \\ 0 & & & & & -\mu & 1 \end{bmatrix}, \tag{5.2.7}$$

y is an $n\times 1$ vector, X is an $n\times k$ nonstochastic matrix of rank $k<n$, β is a $k\times 1$ parameter vector, $D(\mu,\rho)$ is the $n\times n$ diagonal matrix, defined as, diag$(d_1(1-\mu),1,1,...,1)$ and u is an $n\times 1$ vector such that, $u \sim N(0,\sigma^2\Sigma(\rho))$ in which $\Sigma(\rho)$ is an $n\times n$ matrix whose $(i,j)th$ element is $\rho^{|i-j|}/(1-\rho^2)$ (see King (1996)). Equation (5.2.6) implies that

$$y = \Gamma^{-1}(\mu)X\beta+u^*, \tag{5.2.8}$$

where the error term,

$$u^* = \Gamma^{-1}(\mu)D(\mu,\rho)u$$

$$\sim N(0,\ \sigma^2\,\Gamma^{-1}(\mu)D(\mu,\rho)\Sigma(\rho)D(\mu,\rho)\Gamma^{-1}(\mu)'). \tag{5.2.9}$$

Therefore, for the classical tests, the model and hypotheses of interest can be written as

$$y=X(\mu)\beta+u,\ u\sim N(0,\sigma^2\Sigma_1(\mu,\rho)), \tag{5.2.10}$$

and

$$H_0:\ \mu=0 \tag{5.2.11}$$

against

$$H_a: \mu > 0, \tag{5.2.12}$$

respectively, where, $0 \leq \rho < 1$ and $0 \leq \mu < 1$, $X(\mu) = \Gamma^{-1}(\mu)X$ and $\Sigma_1(\mu,\rho) = \Gamma^{-1}(\mu)D(\mu,\rho)\Sigma(\rho)D(\mu,\rho)(\Gamma^{-1}(\mu))'$.

Now, if $\mu = 0$, d_1 becomes 1 and $\Gamma(\mu)$ and $D(\mu,\rho)$ become identity matrices. Similarly, if $\rho = 0$, d_1 becomes $(1-\mu^2)^{-1/2}$ and $\Sigma(\rho)$ becomes an identity matrix. Therefore, the hypotheses of interest can be rewritten in matrix form as

$$H_0: y = X\beta + u,\ u \sim N(0, \sigma^2\Sigma(\rho)),\ 0 \leq \rho < 1, \tag{5.2.13}$$

against

$$H_a: y = X(\mu)\beta + u,\ u \sim N(0, \sigma^2\Sigma_2(\mu)),\ 0 < \mu < 1, \tag{5.2.14}$$

where y, X and $\Sigma(\rho)$ are as defined above, and $\Sigma_2(\mu) = \Gamma^{-1}(\mu)D(\mu)^2(\Gamma^{-1}(\mu))'$. Here we assume that $u_1 \sim N(0, \sigma^2/(1-\rho^2))$ and $e = (e_1,...,e_n)' \sim N(0,\sigma^2 I_n)$. In Chapter 3, following Silvapulle (1991, 1994a), u_0 was assumed to be $e_0/(1-\rho^2)^{1/2}$. Here u_1 is assumed to be $e_1/(1-\rho^2)^{1/2}$ in order to make it consistent with the alternative model assumption (5.2.5).

For the testing problem at hand, β, σ and (for the classical tests case) ρ are nuisance parameters. As discussed in the previous chapter, the testing problem is invariant with respect to the constant coefficient, β_1, and is affected by σ only in the same way as a scaling of $\beta^* = (\beta_2,...,\beta_k)'$. Therefore, one can fix σ and vary β^* or vice versa.

5.3 The Marginal Likelihood Function

If for a particular inference problem, one can factorise the sample density function into the product of two parts, namely, one depending only on the parameters of interest and the other containing no available information about the parameters of interest in the absence of knowledge of the nuisance parameters, then the former part can be interpreted as the marginal likelihood of the parameters of interest (see Grose (1998) for more detail).

In the case of the general linear regression model

$$y = Z(\xi)\beta + u,\ u \sim N(0,\sigma^2\Omega_1), \tag{5.3.1}$$

Bellhouse (1978) has shown that the marginal likelihood function for ξ and Ω_1 is equal to

$$m(\xi,\Omega_1 \mid y) = c\left|Z(\xi)'Z(\xi)\right|^{1/2}\left|\Omega_1\right|^{-1/2}\left|Z(\xi)'\Omega_1^{-1}Z(\xi)\right|^{-1/2}\left(\frac{\hat{s}}{\tilde{s}^l}\right), \tag{5.3.2}$$

where, y and u are $n\times 1$ vectors, $Z(\xi)$ is an $n\times k$ matrix dependent on a parameter vector ξ, β is a $k\times 1$ parameter vector, Ω_1 is an $n\times n$ matrix, $c = \frac{1}{2}\pi^{-l/2}\Gamma\left(\frac{l}{2}\right)$, $l = n-k$, $\hat{s}^2 = y'\overline{P}_Z y$ and $\tilde{s}^2 = y'\tilde{P}_{Z,\Omega_1} y$ are the sums of squared OLS and GLS residuals, respectively, from the regression of y on $Z(\xi)$, $\overline{P}_Z = I - Z(\xi)(Z(\xi)'Z(\xi))^{-1}Z(\xi)'$ is the orthogonal projector, and finally $\tilde{P}_{Z,\Omega_1} = \Omega_1^{-1} - \Omega_1^{-1}Z(\xi)(Z(\xi)'\Omega_1^{-1}Z(\xi))^{-1}Z(\xi)'\Omega_1^{-1}$ is the orthogonal projector for non-spherical disturbances.

The log of the marginal likelihood function for ξ and Ω_1 can be written as

$$M(\xi,\Omega_1 \mid y) = \ln c + \frac{1}{2}\ln\left|Z(\xi)'Z(\xi)\right| - \frac{1}{2}\ln\left|\Omega_1\right| - \frac{1}{2}\ln\left|Z(\xi)'\Omega_1^{-1}Z(\xi)\right| + \frac{1}{2}\ln\hat{s}^2 - \frac{l}{2}\ln\tilde{s}^2. \tag{5.3.3}$$

Model (5.3.1) covers a range of linear models including the models given in (5.2.10) to (5.2.14). For (5.2.10), $Z(\xi) = X(\mu)$ and $\Omega_1 = \Sigma_1(\mu,\rho)$, for (5.2.13), (because $\mu = 0$) $Z(\xi) = X$ and $\Omega_1 = \Sigma(\rho)$ and for (5.2.14), $Z(\xi) = X(\mu)$ and (because $\rho = 0$) $\Omega_1 = \Sigma_2(\mu)$. Therefore, for the problem of interest to this chapter, the marginal likelihood function and its log form can be denoted by $m(\mu,\Omega_1 \mid y)$ and $M(\mu,\Omega_1 \mid y)$, respectively. But for convenience they are denoted as m and M, respectively, throughout.

5.4 The Tests

The following discussion concerning tests applies to testing positive ρ and μ parameters. The tests can easily be modified to test negative ρ and μ parameters if need be.

5.4.1 The Marginal Likelihood Based g Test

For the testing problem of interest, the g test can be denoted as the $g(\mu_0; 0, \rho^*, 0.999)$ test, where μ_0 is the point under the alternative where the power is maximised and $0 < \rho^* < 0.999$. However, for convenience we denote this test as the $g(\mu_0)$ test throughout. The construction of the $g(\mu_0)$ test for testing hypotheses of the form (5.2.13) against (5.2.14) is as outlined in Chapter 3. That is, the $g(\mu_0)$ test is the one that rejects H_0 if

$$m_{r+1} > \sum_{i=1}^{r} k_i m_i \,. \tag{5.4.1}$$

For the existence of such a test, the following r size conditions,

$$\Pr[m_{r+1} > \sum_{i=1}^{r} k_i m_i \mid m_j] = \alpha, \quad j = 1, \ldots, r, \tag{5.4.2}$$

need to be solved simultaneously, by appropriate choices of values for $k_1, \ldots, k_r$. Here m_k, $k = 1, \ldots, r$, can be regarded as representative marginal likelihood functions under the null hypothesis, m_{r+1} is the marginal likelihood function corresponding to μ_0 and α is the level of significance.

As outlined in Chapter 3, we start with the $r = 3$ case and select the representative marginal likelihood functions at $\rho = 0$, $\rho = \rho^*$ and $\rho = 0.999$. Next we need to find ρ^*, k_1, k_2 and k_3 values such that the following size conditions (which are evaluated via the Monte Carlo simulation method) hold simultaneously.

$$\Pr[m_4 > k_1 m_1 + k_2 m_2 + k_3 m_3 \mid \rho = 0] = \alpha, \tag{5.4.3}$$

$$\Pr[m_4 > k_1 m_1 + k_2 m_2 + k_3 m_3 \mid \rho = \rho^*] = \alpha, \tag{5.4.4}$$

$$\Pr[m_4 > k_1 m_1 + k_2 m_2 + k_3 m_3 \mid \rho = 0.999] = \alpha, \tag{5.4.5}$$

$$\Pr[m_4 > k_1 m_1 + k_2 m_2 + k_3 m_3 \mid 0 < \rho < 0.999] = \alpha^*, \tag{5.4.6}$$

where, α^* should be within the (1 - α) percent confidence interval of α.

If for particular values of ρ^*, k_1, k_2 and k_3, probabilities (5.4.3) to (5.4.6) hold simultaneously then we have constructed the g(μ_0) test otherwise we can try some

other ρ^* value, find the k_1, k_2 and k_3 values accordingly and proceed as before. If this also doesn't work, (as mentioned in Chapter 3) we can include another representative density under the null and hence solve for another size condition and so on until the desirable test is found.

The critical values, k_1, k_2, ..., k_r of the g(μ_0) test can be found by following the iterative procedure outlined in Chapter 3. Alternatively, the same can be done by using SA as explained in section 5.5.3.

Here, we use marginal likelihood functions, m_i, instead of log marginal likelihood functions, M_i, $i = 1,\ldots, r+1$, because, the GNPL based test of the form $\log(m_{r+1})$ $(= M_{r+1}) > \log(k_1 m_1 + k_2 m_2 +, \ldots, +k_r m_r)$ is valid, but $M_{r+1} > k_1 M_1 + k_2 M_2 + \ldots + k_r M_r$ is not valid.

As discussed in Chapter 3, the GNPL provides a PO test for the problem of testing whether an observed sample is generated by a finite number of observable density functions against a single alternative density function. Thus, under the null and the alternative the observed sample is the same. For the problem of interest, under the null, the observed sample is equivalent to the maximal invariant function value (3.3.7). This is because, for the null model, Ara and King (1993) showed that the marginal likelihood based approach is equivalent to considering the density function of a maximal invariant statistic. However, the same is not true under the alternative, therefore, the observed sample is not the maximal invariant function value. Consequently, the marginal likelihood based g(μ_0) test is not perfect for the problem at hand, however, it is interesting to see its performance. In this chapter, we consider two versions of the g test, namely, the g(0.3) test and the g(0.5) test. We did not consider the g(0.75) test because when μ gets bigger it becomes easier to distinguish the null and the alternative models of interest, thus, this case is not as interesting. Also, the asymptotic tests are expected to have good power properties for this relatively easier case.

5.4.2 The Marginal Likelihood Based One-Sided LR, LM and W Tests

The one sided LR, LM and W tests for testing (5.2.11) against (5.2.12) in the presence of nuisance parameter ρ can be written as

$$LR = \text{sgn}(\hat{\mu})\,[2(\hat{M} - \tilde{M})]^{1/2}, \tag{5.4.7}$$

$$LM = q_{\mu}(I^{\mu\mu})^{-1/2}|_{\mu=0 \text{ and } \rho=\tilde{\rho}}, \tag{5.4.8}$$

and

$$W = \hat{\mu}(I^{\mu\mu})^{1/2}|_{\mu=\hat{\mu} \text{ and } \rho=\hat{\rho}}, \tag{5.4.9}$$

which under the standard regularity conditions (see White (1982)) follow a $N(0,1)$ distribution under the null asymptotically, where $\hat{\mu}$ and $\hat{\rho}$ are the unrestricted maximum marginal likelihood estimates of μ and ρ, respectively, $\tilde{\rho}$ is the restricted maximum marginal likelihood estimate of ρ under the null, $\hat{M} = M(\hat{\mu},\hat{\rho})$ and $\tilde{M} = M(0,\tilde{\rho})$ are the unrestricted and restricted log marginal likelihood functions, respectively, $\text{sgn}(\hat{\mu}) = 1, -1, 0$ according to whether $\hat{\mu} > 0$, $\hat{\mu} < 0$, or $\hat{\mu} = 0$, respectively, and q_{μ} is the score function with respect to μ. At this point, it is convenient to introduce the following notation. Define the marginal likelihood based information matrix as

$$I_f = -E\left(\frac{\partial^2 \ln m}{\partial \iota \partial \iota'}\right), \tag{5.4.10}$$

where $\iota = (\mu,\rho)'$. Let I_f be partitioned as

$$I_f = \begin{bmatrix} I_{\mu\mu} & I_{\mu\rho} \\ I_{\rho\mu} & I_{\rho\rho} \end{bmatrix}, \tag{5.4.11}$$

and let $I^{\mu\mu}$ be the block in I_f^{-1} corresponding to $I_{\mu\mu}$. That is, $I^{\mu\mu} = I_{\mu\mu} - (I_{\mu\rho} I_{\rho\rho}^{-1} I_{\mu\rho})$. Because μ and ρ are scalars, $I^{\mu\mu}$ is a scalar and $I_{\mu\rho} = I_{\rho\mu}$. Grose (1998) derived all these terms (see Appendix 5.2 for more detail). Among these, $I_{\mu\rho}$ and $I_{\rho\rho}$ have closed form solutions, whereas, $I_{\mu\mu}$ does not have a closed form solution. Grose (1998) provided two computable forms of $I_{\mu\mu}$, namely, Laplace approximated information and estimated information (see Appendix 5.2). These two approximations are asymptotically equivalent, but may differ in finite samples. Which of these approximations is best, particularly in small samples, is unknown in practice. Therefore, in this chapter, we study both tests' behaviour using both approaches.

To make the W and LM tests operational we require, in addition to the maximum marginal likelihood estimates of μ and ρ, suitable estimators of β and σ^2. The GLS estimators (conditional on μ and ρ),

$$\tilde{\beta}(\mu,\rho) = (X(\mu)'\Sigma_1(\mu,\rho)^{-1}X(\mu))^{-1}\ X(\mu)'\Sigma_1(\mu,\rho)^{-1}y \tag{5.4.12}$$

and

$$\tilde{\sigma}^2(\mu,\rho) = \frac{\tilde{e}'\Sigma_1(\mu,\rho)^{-1}\tilde{e}}{l} \tag{5.4.13}$$

are the obvious unbiased candidates, where $\tilde{e}$ is the GLS residual vector from (5.2.10).

Grose (1998, p.168) commented that "While the marginal likelihood-based score with respect to ρ is 'unbiased', in the sense that its expectation is zero, the same cannot be said of the marginal likelihood-based score with respect to the regressor parameter μ. Use of the marginal likelihood, therefore, results in unbiased estimating equations for covariance parameters, but not regressor parameters". Therefore, the marginal

likelihood based LM test may not have better finite-sample properties for the problem at hand.

5.5 Obtaining Exact (and Near Exact) Non-Similar Critical Values via SA

The exact (and near exact) non-similar critical values of the classical tests (because they have standard test statistics and known asymptotic critical values) can straightaway be obtained by using SA, as explained in Chapter 4. The g test case is different. For a given β and σ, the g test statistic is unknown until appropriate critical values and ρ^* are found such that the size conditions (5.4.2) hold simultaneously. Also, the g test has no known asymptotic distribution. Therefore, obtaining near exact non-similar critical values of the g test involves two steps (obtaining exact non-similar critical values involves an iterative process explained in section 5.5.1). The first step involves constructing the test statistic assuming values for β and σ. For this, we assume $\beta = 0$ and $\sigma = 1$ and construct the g test[33]. We denote this test the $g_0(\mu_0)$ test. The second step involves applying SA to find values for the nuisance parameters such that the $g_0(\mu_0)$ test size is at its maximum. For the parameter solution provided by SA, we construct the g test again by finding appropriate critical values and ρ^*. We denote this test the $g(\mu_0)$ test. For the classical tests, obviously, the first step is not necessary. For ease of application, we provide the steps involved in computing exact non-similar critical values of the tests in detail below. The following discussion concerning the g test applies for the $r = 3$ case, the same can easily be generalised for any $r > 3$ case if need be. Also, as in Chapter 4, β_1 and σ are set 0 and 1, respectively, throughout.

[33] As noted in the previous chapter, σ can take any value in this case.

5.5.1 Steps Involved in Obtaining Exact Non-Similar Critical Values of the $g(\mu_0)$ Test

(1) Let $\beta^* = 0$ and construct the $g_0(\mu_0)$ test following section 5.4.1. That is, ρ^* and critical values, k_1, k_2, and k_3, have to be found such that the probabilities (5.4.3) to (5.4.6) hold simultaneously.

(2) Let SA find the values for β^* and ρ over the nuisance parameter space such that the size of the $g_0(\mu_0)$ test is at its maximum.

(3) Construct the $g(\mu_0)$ test for the parameter solution provided by SA. That is, as before, new ρ^*, k_1, k_2, and k_3 values have to be found such that the probabilities (5.4.3) to (5.4.6) hold simultaneously.

(4) Again apply SA to check whether the sizes of the $g(\mu_0)$ test are reasonably controlled over the nuisance parameter space. Repeat steps 3 to 4 until the maximum size obtained is equal to the nominal size.

Observe that the critical values of the $g_0(\mu_0)$ test are obtained by fixing β vector to zero and controlling the sizes of the test over the ρ parameter space alone. Therefore, the critical values of the $g_0(\mu_0)$ test can almost be interpreted as the approximate non-similar critical values used in previous studies. Obviously, if step 2 reveals that the approximate non-similar critical values of step 1 are adequate to control the test sizes over the nuisance parameter space then steps 3 – 4 are not necessary.

The reason for fixing σ as one and varying β^* is as explained in Chapter 4. Though ρ is not a nuisance parameter in the g test case, it is also varied (over its parameter space) because the null distribution of the g test statistic depends on it as well.

5.5.2 Steps Involved in Obtaining Exact Non-Similar Critical Values of the LR, LM and W Tests

The following steps apply for all three classical tests.

(1) Let SA find the values for β^* and ρ over the nuisance parameter space such that the size of the test (based on the asymptotic critical value) is at its maximum.

(2) Apply SA (as explained in section 4.2.2) to find exact size critical value at these values of nuisance parameters.

(3) Use SA to check whether the sizes of the test (based on the exact size critical value of step 2) are reasonably controlled over the nuisance parameter space. Repeat steps 2 to 3, until the maximum size obtained is equal to the nominal size.

As explained in Chapter 4, following all four (first three) steps of section 5.5.1 and all three (first two) steps of section 5.5.2 will lead to exact non-similar critical values (and near exact non-similar critical values) of the g test and the classical tests, respectively.

5.5.3 Obtaining Exact Size Critical Values via SA

The finite-sample critical values of the classical tests (by assuming knowledge of the unknown parameters) can straightaway be obtained by minimising (4.2.4). This is because each classical test involves the finding of one critical value alone and (4.2.4) is specially designed for this. Because the g test involves r critical values, we have to generalise (4.2.4) in order to obtain them. How this can be done for the $r = 3$ case is discussed below, the same can easily be generalised for the $r > 3$ case if need be. In the following discussion, $\varsigma(*)$ means size at $*$.

For a given β^* vector and ρ^* value, we can use SA to find the appropriate critical values of the g test such that the probabilities (5.4.3) to (5.4.5) hold simultaneously. For this, we have to generalise the function (4.2.4) as,

$$f = 10000*(s_1^2 + s_2^2 + s_3^2), \tag{5.5.1}$$

where $s_1 = \alpha - \varsigma(0)$, $s_2 = \alpha - \varsigma(\rho^*)$, and $s_3 = \alpha - \varsigma(0.999)$.

Here, SA finds values for k_1, k_2, and k_3 such that (5.5.1) is minimized. In other words, SA tries to find values for k_1, k_2 and k_3 such that the sizes at $\rho = 0$, ρ^* and 0.999, are exactly equal to α. To gain in terms of computational time one can generate the marginal likelihood functions, $m_{1,i}$, $m_{2,i}$, $m_{3,i}$ and $m_{4,i}$, $i = 1, 2, \ldots, j_t$, where j_t is the number of Monte Carlo replications, for each of $\rho = 0$, ρ^* and 0.999, beforehand, and then apply SA. According to our experience, this is the most efficient way to proceed.

For the problem of interest, a possible value for ρ^* can always be guessed. That is, ρ^* will be closer to μ_0 of the g test. For example, if we are considering the g(0.3) test then ρ^* will be within 0.29 to 0.38 and most probably 0.31 or 0.32. Once ρ^* is decided, the appropriate critical values can be obtained as above. Thus, the construction of the g test is not as hard as it seems. One may wonder why ρ^* of the g test also cannot be obtained via SA. It may be possible to find ρ^* (together with appropriate critical values) via SA, but there is a practical difficulty associated with this, which is explained next.

In order to obtain ρ^* and critical values of the g test, the function to be minimised can be given as

$$f = s_1^2 + 10000*(s_2^2 + s_3^2 + s_4^2), \tag{5.5.2}$$

where s_1 = Maximum of s_i^*, where $s_i^* = (\alpha - \varsigma(\rho_i))$, $i = 1,...,19$, and $\rho_i \in (0.05, 0.1,..., 0.95)$, $s_2 = \alpha - \varsigma(0)$, $s_3 = \alpha - \varsigma(\rho^*)$, and $s_4 = \alpha - \varsigma(0.999)$.

Here, SA finds values for ρ^*, k_1, k_2 and k_3 such that (5.5.2) is minimized. s_1^2 is included in (5.5.2) in order to control the maximum size over the ρ parameter space. Because we prefer sizes at 0, ρ^* and 0.999 equal to α, (to make sure that this happens) ($s_2^2 + s_3^2 + s_4^2$) is multiplied by a big number, such as, 10000 (as above).

This type of optimization is very time consuming, because, the function f (given in (5.5.2)), involves twenty-two sizes (i.e., sizes at $\rho = 0, 0.05, 0.1,..., 0.95, 0.999$ and $\rho = \rho^*$) which have to be evaluated via the Monte Carlo simulation method. As mentioned earlier, to gain in terms of computational time, one can generate the marginal likelihood functions, $m_{1,i}$, $m_{2,i}$, $m_{3,i}$ and $m_{4,i}$, $i = 1, 2, ..., j_t$, where j_t is the number of Monte Carlo replications, for each of $\rho = 0, 0.05, 0.1, ..., 0.95$ and 0.999, beforehand, and then apply SA. However, this amounts to a large amount of computation and will not be attractive if j_t is large. Therefore, finding ρ^* via SA is not feasible in practice.

5.6 Details of the Monte Carlo Design

Extensive Monte Carlo experiments were conducted to see the performance of the marginal likelihood based g and one-sided LR, LM, and W tests in the presence of the nuisance parameters. This study was based on near exact non-similar critical values obtained via the SA based approach. We observed that this approach could be time consuming, particularly for the classical tests because their construction involves repeated maximum marginal likelihood estimation as well. On the other hand, it takes less time for the g test. In addition to this, the type of study we conducted in this chapter (which is explained below) is also time consuming, therefore, to keep the study at a manageable level, we used only two design matrices namely:

X15 : $n \times 3$, $n = 20$, and 40, the X matrix is the same as *X1*. The σ values used for $n = 20$ and 40 are $\sigma = 0.015$, 0.025, and 40 and $\sigma = 0.025$, 0.045, and 0.2, respectively.

X16 : $n \times 3$, $n = 20$ and 40, the regressors are a constant dummy and standard normal random numbers. The σ values used for $n = 20$ and 40 are $\sigma = 0.4$, 0.8, and 10 and $\sigma = 0.4$, 0.7, and 1.81, respectively.

Here, *X15* is a smoothly evolving series, while *X16* possesses a high degree of randomness. In Chapter 2, we discussed some situations where the null and the alternative models of interest may not be distinguishable, of which one is that when the design matrix is lag invariant. That is, for a particular design matrix, if the regressions of each regressor (lagged once) on the rest of regressors all have high R^2 then the corresponding design matrix can be regarded as almost lag invariant. We observe that *X15* is almost lag invariant and *X16* is not lag invariant. It will be interesting to see the tests' performance for these two design matrices.

The computer programs were written in GAUSS for windows NT/95 version 3.2.35 and Tsionas's (1995) program code for the SA algorithm was used. A nominal significance level of five percent and 2000 iterations were used throughout. The maximum marginal likelihood estimates of μ and ρ were obtained by minimizing the negative of the relevant log-marginal likelihood subject to the constraints $-1 < \mu < 1$ and $-1 < \rho < 1$, with OLS estimates as starting values and the bounds set to ± 0.999. Here, the OLS estimates were calculated for each iteration, therefore, the starting values vary for each iteration.

In order to compare the powers of the classical tests (which were applied for the dynamic linear regression model with AR(1) errors) with those of the g test (which was applied to the non-nested testing problem of interest), the y vector was generated assuming $\rho = 0$ under the alternative for the former test.

The one-sided version of a classical test statistic cannot be computable if the term for which the square root has to be taken is negative. This happened sometimes due to sampling variability. For such situations, the test statistic was set to zero, consequently the null was accepted.

In order to check whether the SA based near exact non-similar critical values are indeed working well in terms of controlling the sizes over the nuisance parameter space, we calculated the tests' sizes for a variety of β and σ values. The β vectors considered for $n = 20$ were: $(0,0,0)'$, $(0,1,1)'$, $(0,-1,-1)'$, $(0,1,-1)'$, $(0,-1,1)'$, $(0,1,0)'$, $(0,-1,0)'$, $(0,0,1)'$ and $(0,0,-1)'$, respectively, and each β was made larger or smaller along each of eight directions by decreasing or increasing σ. For each of these β directions and σ values, sizes were calculated for $\rho = 0.1$, 0.3, 0.5, 0.7, 0.9, and 0.99. When $\beta = 0$, as in Chapter 4, the size calculations were done for only one σ value. Because this type of study is quite demanding in terms of time, for $n = 40$, size calculations were done for four selected β directions only, namely, $(0,1,1)'$, $(0,-1,1)'$, $(0,1,0)'$, $(0,0,-1)'$ and $(0,0,0)'$. The size results are reported in Tables 5.4 to 5.15.

Power calculations were also done for the above β vectors (except, for $\beta = 0$) and σ values (except, for big σ of the three considered), in order to determine the most powerful test over the nuisance parameter space. The powers were not calculated for $\beta = 0$ and for big σ because for these cases, the null and the alternative models of interest become similar, consequently, powers can be expected to be close to sizes. For each β and σ, the power calculations were done for $\mu = 0.1$, 0.3, 0.5, 0.7, and 0.9. The power results are reported in Tables 5.16 to 5.21. Also, some selected power results were plotted in Figures 5.1 to 5.5.

Though, considering a variety of β and σ values takes up a lot of time, particularly in the case of the classical tests, it is worth doing so because the finite-sample performance of these tests in the context of the dynamic linear model with AR(1) errors is unknown in practice. For example, for each design matrix and $n = 20$, it takes about 12 days for the LR test, 10 days for the W test and 6 days for the LM test

to do both the size and power calculations. For similar testing scenarios, as noted in Chapter 4, researchers have calculated sizes and powers for only one β vector and for some σ values (see Inder (1985) and Silvapulle (1991)), thus, these studies can be considered inadequate because a test might behave well in terms of size and power properties in a particular β direction, but may behave poorly in another one.

The σ values were chosen as explained in Chapter 4. For this, the y vector was generated assuming $\mu = 0$ and $\rho = 0.5$ and for each β vector with 100 iterations being used. Also, in order to obtain R^2, y was regressed on y_{-1} and X. Therefore, in the g(μ_0) test case, σ values determined in this way are slightly biased because for this case, y should have been regressed on X alone. However, in order to compare the g(μ_0) test performance with that of the classical tests we use the same σ values throughout.

In order to find values for the nuisance parameters such that the size of a test is at its maximum, the key parameters of the SA algorithm, T, N_s, N_t and r_t were set as, $T = 2$, $N_s = 2$, $N_t = 2$ and $r_t = 0.5$. The same SA parameter values were used to obtain unrestricted maximum marginal likelihood estimates. Together with the other parameter values, $N_s = 3$ and $N_t = 3$ were used to obtain restricted maximum marginal likelihood estimates. Preliminary work showed that these parameter values are adequate for the maximum marginal likelihood estimation carried out in this chapter[34]. However, we could not check whether the parameter values provided are adequate to find the global maximum size of a test because it will be too time consuming, particularly for the classical tests. Also, when dealing with size functions, we cannot decide anything firmly based on a small number of Monte Carlo replications such as 50 or 100.

According to our experience, when using Pentium II 400 or Pentium III 500 machines, (for the above SA parameter values) for each design matrix, $n = 20$ and 2000 Monte Carlo replications, SA takes about 6 days for the LR test, 4 days for the W test, 2 days

[34] This observation is based on 100 iterations. We compared these small SA parameter values based optimum results with those based on increased SA parameter values (see Goffe et al. (1994)).

for the LM test and less than 5 minutes for the g test, to provide the nuisance parameter values such that the respective tests sizes are at their maximum. Obviously it takes more time for $n = 40$. The nuisance parameter values and the corresponding maximum sizes are reported in Table 5.1. The ρ^* and critical values of the $g_0(\mu_0)$ and $g(\mu_0)$ tests are given in Table 5.2. The near exact non-similar critical values of the classical tests are given in Table 5.3.

In order to obtain exact size critical values of the classical tests (by assuming knowledge of the unknown parameters), $T = 2$, $N_s = 3$, $N_t = 3$ and $r_t = 0.5$ were used. Similarly, for the g test, $T = 5$, $N_s = 5$, $N_t = 5$ and $r_t = 0.85$ were used. According to our experience, these SA parameter values are adequate for the classical tests to find the exact critical values. However, for the g tests, the SA parameter values, sometimes, have to be increased slightly. If the above SA parameter values are used, SA will take less than two seconds for the classical tests and about twenty minutes for the g test to produce the required outcome.

In order to find values for the nuisance parameters such that the size of a test is at its maximum, the starting values of the parameters provided were: $\beta^* = 0$ and $\rho = 0.5$. As mentioned earlier, the restricted (calculated under the null) and unrestricted OLS estimates of each iteration were provided as the starting values for obtaining restricted and unrestricted maximum marginal likelihood estimates, respectively. Seed values, 9662 and 63721179 were provided for the uniform random number generator of SA and the normal random number generator of the model, respectively. For reasons discussed in the previous chapter, random normal numbers were generated beforehand (i.e., outside the main procedure where the function to be optimized is being evaluated).

As mentioned earlier, in this chapter, we consider both Laplace approximated information based and estimated information based LM (and W) tests, in order to see which information based test is best in finite samples. We denote these tests as LM(L), LM(E), W(L) and W(E) tests throughout. Also, as mentioned earlier, we denote the g test corresponding to step 1 of section 5.5.1 as the $g_0(\mu_0)$ test, otherwise

we denote the g test as the $g(\mu_0)$ test. In the remaining sections of this chapter, 'j_1 is closer to j_2' means j_1 is closer to j_2 in absolute value, where j_1 and j_2 are known numbers.

5.6.1 Results Under the Null Hypothesis

Our study clearly shows the danger in using large-sample based tests, particularly the LR, W(L) and W(E) tests, in small samples. For example, for *X16* and $n = 20$, the LR, W(L) and W(E) tests can have (maximum) sizes as high as 0.748, 0.930 and 0.920, respectively, over the nuisance parameter space. The next poor performers in this regard are the $g_0(0.3)$ and $g_0(0.5)$ tests which can have sizes as high as 0.245 and 0.210, respectively. The $g_0(0.5)$ test is always slightly better than the $g_0(0.3)$ test. Among all the tests considered, the LM(L) test is the best because its size never exceeds 0.120 over the nuisance parameter space. Similarly, the LM(E) test size never exceeds 0.182, and thus can be regarded as second best. Among all the tests, the W(L) and W(E) tests always attain the highest sizes over the nuisance parameter space.

The poor performance of the large-sample critical value based classical tests, particularly, the W(L), W(E) and LR tests, may seem due to the small-sample size, $n = 20$, we used. However, this situation seems to worsen when the sample size increases to 40. For example, for *X16* and $n = 40$, the LR, W(L) and W(E) tests have (maximum) sizes as high as 0.933 and 0.981, and 0.963, respectively, over the nuisance parameter space. In fact the (maximum) sizes are increasing with sample size (see the results reported for $n = 20$ above). For the same design matrix, both the LM tests' respective maximum sizes also slightly increase with sample size. For example, the LM(E) test size increases from 0.158 to 0.182 as the sample size increases from 20 to 40. However, LM tests' (maximum) sizes seem to improve slightly for *X15*. For example, the LM(E) test size decreases from 0.169 to 0.131 when the sample size increases from 20 to 40. The $g_0(0.3)$ test's (maximum) size decreases for *X16*, but it increases for *X15*, when the sample size increases. The only

test whose (maximum) size always decreases when the sample size increases is the $g_0(0.5)$ test.

For the classical tests only, we further increased the sample size to 60 and observed that the LM tests' (maximum) sizes slightly improve. For example, for *X16*, the LM(E) test size improves from 0.182 to 0.119 when the sample size increases from 40 to 60. Similarly, for *X15*, the LM(L) test size improves from 0.102 to 0.088. For *X15*, the LR test size improves from 0.501 to 0.220, however, for *X16*, its size increases with the sample size. Our study clearly shows that the large-sample based LR, W(L) and W(E) tests are not reliable under the null even when $n = 60$. Though, the LM tests' sizes are better than their competitors, they are not ideal even when $n = 60$. Therefore, it is not particularly advisable to use large-sample critical value based classical tests in finite samples.

Generally, the LR, W(L) and W(E) tests achieve their maximum sizes for small β values and for ρ between 0.85 to 0.99. For the LM tests, the same happens for ρ between 0.3 to 0.99. While the $g_0(0.3)$ and $g_0(0.5)$ tests attain their maximum sizes for small β values and ρ between 0.6 to 0.9 or for big β values and ρ closer to zero. These observations suggest that approximate non-similar critical values used in previous studies may work well for the asymptotic tests but not for the $g(\mu_0)$ tests.

Now let us discuss the performance of the near exact non-similar critical values based tests. The near exact non-similar critical values we proposed are working remarkably well in terms of controlling sizes. Generally, the sizes are less than or equal to the nominal size and sometimes higher than the nominal size. All these tests' higher than nominal sizes are almost always less than twice the nominal size. Also we observe that such higher than nominal sizes occur more for the g(0.3) test compared to the other tests.

For the design matrices considered, all the tests' sizes have a common pattern which is as follows. The tests' sizes seem to change with β for small σ (this happens more for $n = 20$ than for $n = 40$) and become stable for big σ. For example, for *X15*, σ =

0.015 and $\rho = 0.99$, the W(L) test sizes for $\beta = (0, 1, 1)'$ and $\beta = (0, -1, 0)'$ are 0 and 0.022, respectively. Also, it can be seen that the size results corresponding to big σ are always closer to those for $\beta = 0$. This is not surprising, because, for the problem of interest, (any) β approaches 0 when σ increases, hence the result. Thus, in the discussion below, any findings reported for $\beta = 0$ also apply to big σ.

We now discuss the results for $n = 20$. For the LM(L), LM(E), g(0.3) and g(0.5) tests, use of near exact non-similar critical values often produces sizes that are not significantly different from the nominal size, while for the LR, W(L) and W(E) tests they often produce sizes well below the nominal size. For *X15*, among the asymptotic tests, the LM(E) test sizes are the best, because they generally range between 0.030 to 0.050. The g(0.5) test sizes for $\sigma = 0.015$ and 0.025 are also best in this regard. For *X16*, among the asymptotic tests, both the LM tests' sizes are better. The g(0.5) test sizes are generally closer to 0.018, but sometimes closer to zero. The g(0.3) test sizes are typically closer to 0.010, except when $\sigma = 0.8$. For both design matrices, the LR and W tests' sizes are generally closer to zero or 0.010, except when $\beta = 0$ and/or $\rho \geq 0.9$. All the asymptotic tests' sizes increase and become closer to the nominal size when σ increases, whereas, the g(0.3) and g(0.5) tests' sizes become closer to 0 or 0.01. The LR and W tests' sizes seem to increase with ρ, while other tests' sizes seem to increase with ρ sometimes and waver otherwise.

If the tests were to be ranked on the basis of their sizes, for $n = 20$, overall the LM(E) test is superior. Of the 300 cases, the LM(E) test has sizes that are not significantly different from the nominal size in 96 cases. Similarly, the LM(L), g(0.3), g(0.5), W(L), W(E) and LR tests' sizes are not significantly different from the nominal size in 90, 82, 69, 27, 25, and 21 cases, respectively.

Now let us discuss the results for $n = 40$. For *X15*, among the asymptotic tests, the LM(L) and LM(E) tests' sizes are better because they are generally closer to the nominal size. The g(0.3) test's sizes for $\sigma = 0.025$ and the g(0.5) test's sizes for $\sigma = 0.025$ and $\sigma = 0.045$, are also better in this regard. For *X16*, among the asymptotic tests, the LM(L) test sizes seem better. The LM(E) test sizes are generally less than or

equal to 0.010. The g(0.5) test sizes are generally closer to zero except when σ = 1.81, while the g(0.3) test sizes range between 0 to 0.078. For both design matrices, the LR and W tests' sizes behave as for n = 20. If the tests were to be ranked on the basis of their sizes, for n = 40, overall the LM(L) test is superior. Of the 156 cases, the LM(L), g(0.5), g(0.3), LM(E), W(E), W(L) and LR tests' sizes are not significantly different from the nominal size in 62, 57, 53, 37, 9, 8 and 2 cases, respectively.

Even if the sizes are successfully controlled, the tests will be of little use if they do not have sufficient power to reject a false null hypothesis. We compare the powers next.

5.6.2 Results Under the Alternative Hypothesis

The power comparisons of this section are based on near exact non-similar critical values. We observed that the power results for each design matrix change when β changes. For example, for *X16*, n = 20, σ = 0.8 and μ = 0.5, the g(0.3) test powers for β = (0,1,1)′ and β = (0,0,1)′ are 0.873 and 0.364, respectively, while the corresponding (average) sizes are similar. Similarly, for particular μ and σ values, the classical tests also have markedly different powers along different β directions. Therefore, it seems reasonable to compare the tests' powers along each β direction. By doing this, we observed that the main findings are largely unaffected with respect to β. The main findings are as follows.

For *X15*, n = 20 and σ = 0.015, the following pattern emerges always (regardless of β). Sometimes none of the tests are powerful at μ = 0.1. Except for these cases, the g(0.3) test is the most powerful test for 0.1 $\leq$ μ $\leq$ 0.5 and generally it continues to have reasonable powers for μ = 0.7. The g(0.5) test is the second best test for 0.1 $\leq$ μ $\leq$ 0.5, and it continues to be the most powerful test for 0.5 < μ $\leq$ 0.9 along 4 (out of 8) β directions. It may look strange that the g(0.3) test is the most powerful test at μ = 0.5 instead of the g(0.5) test. This may be happening due to the g(0.3) test's

higher sizes. On many occasions, the g(0.3) and g(0.5) tests have better sizes and powers that are almost double that of the best classical test.

None of the classical tests are reliable when $\mu = 0.1$ in terms of power. Among the classical tests, the LM(E) test is the best for $\mu = 0.3$ and the LM(L) test is second best, however, their powers are much less than those for the g(0.3) test. Here, it is difficult to compare power because the LM(E) tests' sizes are almost always higher than those of the LM(L) test, still there are a few cases where the latter test demonstrates superiority over the former. For example, for $\beta = (0,-1,-1)'$ and $\mu = 0.5$, the LM(L) and LM(E) tests' powers are 0.722 and 0.687, respectively, while the corresponding average sizes are 0.015 and 0.030, respectively. It is worth noting that the LR test's power for this case is 0.905, whereas, its corresponding average size is 0.010. Among the classical tests, the LR test is the most powerful test for $\mu \geq 0.5$, except for two β directions in which the LM(E) test is more powerful at $\mu = 0.5$. Among all the tests, generally, the LR test is the most powerful test for $\mu \geq 0.7$.

When σ increases, β decreases, therefore for this case, it will be harder to distinguish the null and the alternative models when μ and ρ are small. This can be described as the most complicated testing situation. In addition, if the design matrix is lag invariant then it will be impossible to distinguish the null and the alternative models of interest. Such a situation exists when $\sigma = 0.025$ and $\mu = 0.1$. Though none of the tests are powerful in this case, the g(0.3) and g(0.5) tests manage to have better powers along two (out of eight) β directions. When μ increases, the g(0.3) and g(0.5) tests often have powers that are more than double those of their competitors. The LR test overtakes the g(0.5) test only when $\mu = 0.9$ (see Figure 5.1). Surprisingly, the W tests sometimes outperform the LM tests at $\mu = 0.9$. Similarly, there are cases where the g(0.3) and g(0.5) tests outperform the LM and W tests at $\mu = 0.9$.

For *X16*, $n = 20$ and $\sigma = 0.4$, generally the g(0.3) test is the most powerful test at $\mu = 0.1$. Because of their higher sizes, sometimes, the LM tests outperform the g(0.3)

test at $\mu = 0.1$. The g(0.3) test is the most powerful test at $\mu = 0.3$ and it continues to have better powers for $\mu \leq 0.7$. Whenever, the g(0.3) test sizes are higher than those of the g(0.5) test, the former test turns out to be the most powerful test at $\mu = 0.5$ and vice-versa. The g(0.5) test is the most powerful test at $\mu = 0.7$, even when its sizes are closer to zero.

Generally, the asymptotic tests' powers are not far behind those for the g(0.3) and g(0.5) tests. Among the asymptotic tests, the LM(L) test is most powerful for $0.1 \leq \mu \leq 0.3$ and the LM(E) test is second best for $\mu = 0.1$. The LM(L) test clearly dominates the LM(E) test on many occasions. For example, for $\beta = (0,1,1)'$ and $\mu = 0.5$, the LM(L) and LM(E) tests' powers are 0.967 and 0.587, respectively, while both tests' sizes are comparable. Among the asymptotic tests, the LR test is second best for $\mu = 0.3$ along some β directions and for the other directions the LM(E) test is second best. Similarly, for $\mu = 0.5$, the LR test is the most powerful test along some β directions and for the other directions the LM(L) test dominates. The LR test possesses reasonable power for $\mu = 0.7$ and it turns out to be the most powerful test for $\mu = 0.9$.

When σ increases to 0.8, the only test that always successfully overcomes the most complicated testing situation (outlined above) is the g(0.3) test. None of the classical tests are reliable at $\mu = 0.1$. The g(0.3) test is the most powerful test for $\mu \leq 0.3$ and it continues to have reasonable powers for $\mu = 0.5$ and $\mu = 0.7$. The g(0.5) test is the second best test for $\mu \leq 0.3$ and the most powerful test for $0.3 < \mu \leq 0.7$. Both the g(0.3) and g(0.5) tests have very high powers compared to their competitors and the LR test overtakes the g(0.5) test only when $\mu > 0.7$ (see Figure 5.3). There are cases where the g(0.5) test outperforms the LM and W tests at $\mu = 0.9$ (see Figure 5.3). Similarly, the W tests sometimes outperform the LM tests at $\mu = 0.9$ (see Figure 5.4). Among the asymptotic tests, the LM(L) test is the best for $0.1 < \mu \leq 0.5$ and the LR test overtakes thereafter.

The W(E) and W(L) tests are generally the worst performers. Their powers are zero or closer to zero for $0.1 \leq \mu < 0.7$ and sometimes (as noted above) they catch up with their competitors at $\mu = 0.9$ (see Figure 5.1 and 5.4). This may be due to their lower sizes. The W(L) test seems more powerful than the W(E) test for *X15* and the converse is true for *X16*.

The power results at $\mu = 0.9$ are worth mentioning. The LR test is always the most powerful test at $\mu = 0.9$ and its powers approach one, whereas, the LM and W tests' are not always reliable. The g(0.5) test powers approach one along some β directions and for the other β directions the test powers approach zero. Generally, the g(0.3) test is the least powerful test at $\mu = 0.9$ (see Figure 5.4), however, it occasionally outperforms the W(L) and W(E) tests. This shows that the g(0.3) and g(0.5) tests are always reliable in the neighbourhood of the point where the power is maximised and not always reliable some distance from that point. As expected, compared to *X15* (which is nearly lag invariant), all the tests' powers are generally higher for *X16* (which is not lag invariant).

When the sample size increases to 40, none of the classical tests are recommendable for *X15* when $\mu = 0.1$ and $\sigma = 0.025$, however, the g(0.3) test manages to have better powers along three (out of four) β directions. This test is the most powerful test for $\mu \leq 0.5$ and the g(0.5) test is second best. As expected, asymptotic tests' powers improve with increased sample size, although, they are less powerful than the g(0.3) and g(0.5) tests for the above parameter range. The LR test overtakes the g(0.5) test only when $\mu \geq 0.7$. However, the g(0.5) test powers are the second best at $\mu = 0.7$. All the asymptotic tests powers approach one at $\mu = 0.9$, whereas, the g(0.3) and g(0.5) tests' powers approach zero (see Figure 5.2). The LM(L) and LM(E) tests' powers are almost similar, whereas, the W(L) test is more powerful than the W(E) test (see Figure 5.2).

When σ increases to 0.045, none of the tests are advisable at $\mu = 0.1$. The rest of the findings are similar to those above, except now the LR test overtakes the g(0.5)

test only when $\mu = 0.9$. Also, the LR test powers (at $\mu = 0.9$) always approach one, whereas, the other asymptotic tests' powers do the same along some β directions only. Moreover, compared to the LM tests, the W tests are more powerful at $\mu = 0.9$.

For *X16* (which is not lag invariant) and $n = 40$, all the tests' powers increase remarkably. When $\sigma = 0.4$, the LM(L) test is the most powerful test at $\mu = 0.1$. For this case, the LM(L) test always possesses higher average sizes than the g(0.3) test. The g(0.3) test is the most powerful test for $0.1 < \mu \leq 0.7$ and its powers are closer to one. The g(0.5) test is also equally powerful for $0.5 \leq \mu \leq 0.7$. The LM(L) test's powers closely follow the g(0.3) test's powers for $0.1 < \mu \leq 0.5$. For this design matrix, the LM(L) test's sizes are always higher than those for the LM(E) test, which makes the latter test less powerful than the former. Among the asymptotic tests, the LR test is the most powerful test for $\mu \geq 0.5$ and its powers always approach one. All the asymptotic tests' powers at $\mu = 0.9$ approach one, while the g(0.5) test also manages to have better powers along three (out of four) β directions.

When σ increases to 0.7, the g(0.3) test is the most powerful test for $\mu \leq 0.5$ and it continues to have better powers for $\mu = 0.7$. The g(0.5) test is the second best test for $\mu = 0.5$ and the most powerful test for $\mu = 0.7$. Among the asymptotic tests, the LM(L) test is the best for $\mu \leq 0.5$ and the LR test is the best for $\mu > 0.5$ (see Figure 5.5). For both the design matrices, the W(L) and W(E) tests are the worst performers.

Our study supports the LM(L) test over the LM(E) test, particularly when the design matrix is not lag invariant and $n = 20$. When the sample size increases, as expected, the LM(L) and LM(E) tests tend to behave similarly. In the case of the W(L) and W(E) tests, it is not clear which test is better. For *X15* , the W(L) test seems better than the W(E) test in terms of power, but the converse is true for *X16*. Because both the W(L) and W(E) tests are the worst performers always, neither of them are advisable for the testing problem considered.

5.7 Conclusion

In this chapter, we dealt with a complicated testing problem, for which the nuisance parameters cannot be avoided. Because marginal likelihood based tests are known to perform better in the presence of nuisance parameters, we considered marginal likelihood based tests for this problem. In particular, we compared marginal likelihood based g tests with marginal likelihood based one-sided LR, LM and W tests in terms of size and power properties, using near exact non-similar critical values obtained via SA. Both the Laplace approximated and estimated information based LM and W tests were considered for this study.

This study clearly indicated that the large-sample based classical tests, such as, the LR and W tests cannot be trusted under the null even when $n = 60$. Among all the tests considered, the LM(L) test is found to be the best test because it always achieves the least maximum size over the nuisance parameter space. The LM(E) test is the second best in this regard.

The size and power calculations were made for a variety of β and σ values in order to determine the best test over the nuisance parameter space. The size results indicate that, the SA based near exact non-similar critical values are generally reliable over the nuisance parameter space. We also observed that approximate non-similar critical values used in previous studies are not useful for the g(μ_0) test, whereas, our SA based near exact non-similar critical values seem useful.

The power study conducted along each β direction revealed that the g(0.3) test is the most powerful test for $\mu < 0.5$ and the g(0.5) test is generally the second best test for this range and the most powerful test for $0.5 \leq \mu \leq 0.7$. Whenever, the g(0.3) test sizes are higher than those of the g(0.5) test, the former test turns out to be the most powerful test at $\mu = 0.5$. The g(0.3) and g(0.5) tests often have powers that are more than double of their competitors. The g(0.3) test is generally the least powerful test at $\mu = 0.9$. The g(0.5) test's power sometimes approaches zero when μ approaches zero or one. Therefore, these tests seem always reliable in the neighbourhood of the

point where the power is maximised and not always reliable some distance from that point. Consequently, the g(0.75) test might be suitable for $\mu > 0.7$. However, this test may lack power for small values of μ.

Among the asymptotic tests, the LM tests behave better for $\mu \leq 0.5$ and the LR test overtakes thereafter. The only test whose power always approaches one at $\mu = 0.9$ is the LR test. The W(L) and W(E) tests are generally the worst performers. The LM(L) test seems better than the LM(E) test, whereas, this is not clear in the W tests' case.

We observed that when μ increases it becomes easier to distinguish the null and the alternative models of interest. Therefore, for the testing problem considered, the g(0.3) and g(0.5) tests should be commended because they are the ones doing a tougher job. In addition, the only test that successfully overcomes some very tough testing situations typically is the g(0.3) test. Even though the marginal likelihood based g tests are not perfect for the problem of interest (see section 5.4.1), they excel in their performance. The LM tests are also not expected to have better finite-sample properties (see section 5.4.2), yet, they are the best asymptotic tests for small values of μ. Another, notable performer is the LR test whose sizes are generally closer to zero than those of other asymptotic tests, yet it possesses reasonable powers.

Based on our results, we recommend the g(0.3) and g(0.5) tests for $\mu < 0.5$ and $0.5 \leq \mu \leq 0.7$, respectively. The critical values of these tests can be obtained using SA, as explained in this chapter. The asymptotic tests, particularly the LR test seems suitable for $\mu > 0.7$. Recall that, in the context of the static linear regression model, Rahman and King (1994) preferred some marginal likelihood based asymptotic tests over King's APO tests. For a more complicated testing situation, we have seen that the $g(\mu_0)$ test (which is an APO test) behaves better than the marginal likelihood based classical tests.

APPENDIX 5.1

Tables of Results of the Monte Carlo Experiments

Table 5.1: Maximum sizes of the asymptotic and g_0 tests together with the nuisance parameter values at this maximum size

		LR	W(L)	W(E)	LM(L)	LM(E)	$g_0(0.3)$	$g_0(0.5)$
					$n = 20$			
X15	Max Size	0.373	0.525	0.627	0.120	0.169	0.245	0.210
	β_2	-0.234	-0.258	0.189	-1.500	-0.955	-72.998	-58.800
	β_3	-0.831	-0.261	0.140	-1.175	0.932	-13.420	-8.590
	ρ	0.993	0.853	0.983	0.922	0.996	0.002	0.002
X16	Max Size	0.748	0.930	0.920	0.094	0.158	0.198	0.131
	β_2	-0.003	0.016	-0.005	-0.055	-0.015	-1.068	-0.618
	β_3	0.006	0.001	0.000	-0.006	-0.077	0.308	0.266
	ρ	0.949	0.986	0.982	0.685	0.523	0.912	0.817
					$n = 40$			
X15	Max Size	0.501	0.847	0.803	0.102	0.131	0.268	0.151
	β_2	-0.013	0.051	0.421	-1.172	-0.675	-12.375	-6.914
	β_3	0.030	0.507	3.183	2.324	-0.422	-18.411	-27.567
	ρ	0.955	0.949	0.992	0.895	0.999	0.025	0.381
X16	Max Size	0.933	0.981	0.963	0.109	0.182	0.152	0.087
	β_2	0.000	-0.029	0.024	-0.021	0.042	-0.525	-0.005
	β_3	0.004	-0.033	0.000	-0.082	0.031	0.494	0.556
	ρ	0.960	0.994	0.901	0.502	0.373	0.024	0.696
					$n = 60$			
X15	Max Size	0.220	0.961	0.860	0.088	0.107		
	β_2	-1.078	0.135	0.135	-0.292	2.531		
	β_3	0.191	-0.132	-0.130	-0.350	-0.180		
	ρ	0.739	0.934	0.934	0.729	0.976		
X16	Max Size	0.953	0.968	0.946	0.099	0.119		
	β_2	0.000	0.034	-0.016	0.017	0.000		
	β_3	0.000	0.016	0.017	-0.005	0.011		
	ρ	0.961	0.843	0.807	0.353	0.500		

Table 5.2: Critical values and ρ^* values of the g_0 and g tests at the 5% level

		$g_0(0.3)$		g(0.3)		$g_0(0.5)$		g(0.5)	
		X15	*X16*	*X15*	*X16*	*X15*	*X16*	*X15*	*X16*
$n=20$	k_1	-0.02494	-0.02328	0.0312	-0.081	-0.01151	-0.016799	-0.07332	-0.017
	k_2	1.4576	2.1277	3.0512	5.35	0.675	3.5839	3.015	6.97
	k_3	0.696	0.000223	0.14138	0	4.4591	0.047263	2.40945	0.022
	ρ^*	0.32	0.32	0.31	0.37	0.42	0.52	0.44	0.58
$n=40$	k_1	0.0155	0.02233	-0.018	0.049	-0.00022	0.00035	-0.0021	-0.00023
	k_2	1.479	2.636	3.89	4.810	1.83	5.3745	3.73	9.2
	k_3	0.064	0	0.00405	0	3.715	0	0.92	0
	ρ^*	0.335	0.353	0.31	0.353	0.525	0.55	0.495	0.561

Table 5.3: Near exact non-similar critical values of the classical tests

Test	*X15*		*X16*	
	$n=20$	$n=40$	$n=20$	$n=40$
LR	2.5656	2.7121	2.8538	3.3016
LM(L)	2.0789	2.0172	1.8869	1.9874
LM(E)	2.3953	2.1960	2.6734	2.7839
W(L)	10.1000	10.1983	16.9175	30.5770
W(E)	10.0200	11.2518	15.5300	22.7321

Table 5.4: Sizes of all the tests at the 5% level for *X15*, selected β, σ and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.015	0.1	0.007	0.027	0.012	0.000	0.000	0.068	0.042
		0.3	0.007	0.028	0.009	0.000	0.000	0.068	0.038
		0.5	0.008	0.027	0.012	0.000	0.000	0.061	0.035
		0.7	0.008	0.024	0.012	0.000	0.001	0.046	0.025
		0.9	0.010	0.025	0.017	0.000	0.000	0.039	0.025
		0.99	0.014	0.031	0.018	0.000	0.000	0.045	0.021
	0.025	0.1	0.009	0.029	0.014	0.000	0.001	0.051	0.049
		0.3	0.009	0.035	0.013	0.000	0.000	0.050	0.051
		0.5	0.009	0.031	0.016	0.000	0.000	0.044	0.048
		0.7	0.013	0.029	0.018	0.001	0.001	0.042	0.038
		0.9	0.017	0.023	0.023	0.001	0.003	0.037	0.033
		0.99	0.021	0.027	0.029	0.002	0.006	0.047	0.038
	40	0.1	0.002	0.056	0.013	0.004	0.009	0.000	0.004
		0.3	0.003	0.067	0.027	0.004	0.011	0.000	0.003
		0.5	0.007	0.059	0.036	0.006	0.030	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.044	0.011	0.015
		0.9	0.032	0.043	0.049	0.042	0.061	0.019	0.021
		0.99	0.051	0.050	0.057	0.049	0.047	0.023	0.031
(0, -1, -1)'	0.015	0.1	0.007	0.030	0.012	0.000	0.000	0.071	0.039
		0.3	0.008	0.031	0.015	0.000	0.000	0.068	0.035
		0.5	0.011	0.035	0.014	0.000	0.000	0.067	0.032
		0.7	0.009	0.026	0.011	0.000	0.000	0.058	0.030
		0.9	0.012	0.028	0.015	0.000	0.000	0.049	0.024
		0.99	0.013	0.033	0.021	0.000	0.000	0.052	0.024
	0.025	0.1	0.006	0.031	0.013	0.000	0.000	0.045	0.047
		0.3	0.007	0.035	0.017	0.000	0.000	0.047	0.049
		0.5	0.010	0.034	0.016	0.000	0.000	0.048	0.047
		0.7	0.014	0.030	0.020	0.000	0.001	0.049	0.047
		0.9	0.018	0.031	0.023	0.000	0.003	0.048	0.042
		0.99	0.026	0.036	0.031	0.005	0.007	0.052	0.041
	40	0.1	0.003	0.056	0.013	0.004	0.010	0.000	0.004
		0.3	0.003	0.067	0.027	0.004	0.015	0.000	0.003
		0.5	0.008	0.059	0.035	0.006	0.029	0.004	0.007
		0.7	0.016	0.044	0.040	0.017	0.041	0.011	0.015
		0.9	0.032	0.043	0.049	0.040	0.066	0.019	0.021
		0.99	0.051	0.050	0.057	0.049	0.043	0.023	0.032
(0, -1, 1)'	0.015	0.1	0.006	0.026	0.009	0.000	0.000	0.045	0.013
		0.3	0.009	0.026	0.013	0.001	0.001	0.064	0.025
		0.5	0.010	0.033	0.022	0.001	0.001	0.086	0.036
		0.7	0.013	0.036	0.028	0.003	0.002	0.096	0.043
		0.9	0.018	0.040	0.033	0.003	0.003	0.091	0.038
		0.99	0.022	0.039	0.032	0.005	0.005	0.081	0.039

Table 5.5: Sizes of all the tests at the 5% level for *X15*, selected β, σ and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, -1, 1)'	0.025	0.1	0.004	0.026	0.012	0.001	0.001	0.024	0.016
		0.3	0.007	0.031	0.016	0.002	0.001	0.044	0.028
		0.5	0.008	0.035	0.027	0.002	0.002	0.062	0.042
		0.7	0.017	0.036	0.030	0.007	0.005	0.068	0.052
		0.9	0.024	0.037	0.034	0.014	0.015	0.066	0.051
		0.99	0.030	0.038	0.040	0.022	0.018	0.068	0.054
	40	0.1	0.002	0.056	0.013	0.004	0.009	0.000	0.004
		0.3	0.003	0.067	0.026	0.004	0.019	0.000	0.003
		0.5	0.009	0.059	0.036	0.006	0.025	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.042	0.011	0.015
		0.9	0.031	0.043	0.049	0.042	0.060	0.019	0.021
		0.99	0.050	0.050	0.057	0.049	0.049	0.023	0.032
(0, 1, -1)'	0.015	0.1	0.005	0.021	0.007	0.000	0.000	0.038	0.012
		0.3	0.007	0.029	0.010	0.000	0.000	0.067	0.017
		0.5	0.008	0.030	0.017	0.001	0.000	0.087	0.029
		0.7	0.011	0.028	0.017	0.001	0.000	0.098	0.038
		0.9	0.016	0.033	0.025	0.003	0.003	0.087	0.044
		0.99	0.017	0.035	0.031	0.005	0.006	0.089	0.045
	0.025	0.1	0.004	0.027	0.011	0.000	0.000	0.017	0.011
		0.3	0.005	0.033	0.018	0.000	0.000	0.040	0.019
		0.5	0.006	0.034	0.024	0.002	0.001	0.062	0.041
		0.7	0.011	0.033	0.027	0.004	0.002	0.071	0.053
		0.9	0.024	0.035	0.034	0.016	0.016	0.065	0.055
		0.99	0.029	0.038	0.037	0.023	0.023	0.072	0.053
	40	0.1	0.003	0.056	0.013	0.004	0.010	0.000	0.004
		0.3	0.004	0.067	0.027	0.004	0.015	0.000	0.003
		0.5	0.007	0.059	0.035	0.006	0.032	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.043	0.011	0.015
		0.9	0.032	0.044	0.049	0.041	0.064	0.019	0.021
		0.99	0.049	0.050	0.057	0.049	0.050	0.023	0.031
(0, 1, 0)'	0.015	0.1	0.008	0.034	0.015	0.000	0.000	0.036	0.046
		0.3	0.008	0.035	0.015	0.000	0.000	0.039	0.048
		0.5	0.012	0.029	0.021	0.000	0.000	0.044	0.049
		0.7	0.022	0.026	0.023	0.003	0.007	0.048	0.052
		0.9	0.031	0.033	0.028	0.007	0.011	0.043	0.046
		0.99	0.035	0.034	0.031	0.011	0.016	0.042	0.048
	0.025	0.1	0.006	0.044	0.015	0.000	0.000	0.007	0.017
		0.3	0.006	0.043	0.019	0.000	0.000	0.013	0.023
		0.5	0.010	0.034	0.027	0.001	0.001	0.016	0.029
		0.7	0.014	0.030	0.029	0.005	0.005	0.022	0.034
		0.9	0.035	0.035	0.038	0.009	0.013	0.026	0.042
		0.99	0.046	0.041	0.041	0.012	0.018	0.031	0.040

Table 5.6: Sizes of all the tests at the 5% level for *X15*, selected β, σ and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 0)'	40	0.1	0.003	0.056	0.013	0.004	0.010	0.000	0.004
		0.3	0.004	0.067	0.027	0.004	0.016	0.000	0.003
		0.5	0.007	0.059	0.035	0.006	0.028	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.045	0.011	0.015
		0.9	0.032	0.043	0.049	0.041	0.056	0.019	0.021
		0.99	0.050	0.050	0.057	0.049	0.043	0.023	0.031
(0, -1, 0)'	0.015	0.1	0.008	0.034	0.012	0.000	0.001	0.031	0.042
		0.3	0.007	0.032	0.015	0.001	0.001	0.035	0.045
		0.5	0.013	0.034	0.018	0.000	0.000	0.034	0.043
		0.7	0.020	0.030	0.025	0.001	0.002	0.045	0.044
		0.9	0.031	0.031	0.027	0.007	0.016	0.044	0.047
		0.99	0.033	0.034	0.033	0.012	0.022	0.051	0.051
	0.025	0.1	0.004	0.042	0.014	0.000	0.000	0.006	0.017
		0.3	0.005	0.046	0.020	0.000	0.001	0.011	0.019
		0.5	0.006	0.037	0.024	0.000	0.000	0.014	0.027
		0.7	0.018	0.033	0.027	0.003	0.004	0.024	0.033
		0.9	0.034	0.037	0.039	0.010	0.016	0.029	0.040
		0.99	0.050	0.041	0.044	0.016	0.020	0.037	0.051
	40	0.1	0.003	0.056	0.013	0.004	0.011	0.000	0.004
		0.3	0.004	0.067	0.026	0.004	0.015	0.000	0.003
		0.5	0.007	0.059	0.035	0.006	0.030	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.040	0.011	0.015
		0.9	0.032	0.043	0.049	0.040	0.062	0.019	0.021
		0.99	0.050	0.050	0.057	0.050	0.048	0.023	0.032
(0, 0, 1)'	0.015	0.1	0.004	0.029	0.011	0.001	0.001	0.049	0.024
		0.3	0.006	0.030	0.012	0.001	0.001	0.061	0.028
		0.5	0.007	0.028	0.018	0.001	0.001	0.072	0.034
		0.7	0.008	0.030	0.021	0.001	0.002	0.068	0.040
		0.9	0.016	0.033	0.025	0.001	0.003	0.066	0.036
		0.99	0.021	0.033	0.031	0.003	0.005	0.063	0.038
	0.025	0.1	0.005	0.034	0.013	0.001	0.001	0.025	0.023
		0.3	0.006	0.030	0.015	0.001	0.001	0.033	0.027
		0.5	0.009	0.035	0.023	0.000	0.000	0.044	0.032
		0.7	0.009	0.033	0.026	0.001	0.003	0.045	0.040
		0.9	0.021	0.031	0.033	0.006	0.010	0.050	0.041
		0.99	0.027	0.034	0.028	0.006	0.009	0.054	0.045
	40	0.1	0.003	0.056	0.013	0.004	0.010	0.000	0.004
		0.3	0.004	0.067	0.026	0.004	0.015	0.000	0.003
		0.5	0.007	0.059	0.036	0.006	0.019	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.044	0.011	0.015
		0.9	0.032	0.043	0.049	0.042	0.062	0.019	0.021
		0.99	0.050	0.050	0.057	0.049	0.049	0.023	0.032

Table 5.7: Sizes of all the tests at the 5% level for *X15*, selected β, σ, and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 0, -1)'	0.015	0.1	0.005	0.029	0.013	0.000	0.000	0.047	0.024
		0.3	0.004	0.030	0.017	0.000	0.000	0.062	0.028
		0.5	0.008	0.031	0.020	0.000	0.000	0.078	0.038
		0.7	0.010	0.028	0.019	0.000	0.000	0.081	0.047
		0.9	0.014	0.031	0.024	0.000	0.002	0.076	0.042
		0.99	0.019	0.036	0.032	0.002	0.004	0.074	0.045
	0.025	0.1	0.002	0.034	0.016	0.000	0.000	0.022	0.019
		0.3	0.003	0.035	0.022	0.000	0.000	0.034	0.029
		0.5	0.007	0.035	0.023	0.001	0.001	0.050	0.037
		0.7	0.014	0.030	0.023	0.003	0.003	0.056	0.049
		0.9	0.022	0.032	0.032	0.009	0.012	0.053	0.050
		0.99	0.026	0.041	0.041	0.010	0.016	0.060	0.050
	40	0.1	0.003	0.056	0.013	0.004	0.011	0.000	0.004
		0.3	0.004	0.067	0.027	0.004	0.015	0.000	0.003
		0.5	0.007	0.059	0.035	0.006	0.029	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.046	0.011	0.015
		0.9	0.033	0.043	0.049	0.040	0.061	0.019	0.021
		0.99	0.049	0.050	0.057	0.049	0.050	0.023	0.031
(0, 0, 0)'	0.015	0.1	0.002	0.056	0.013	0.004	0.010	0.000	0.004
		0.3	0.004	0.067	0.026	0.004	0.015	0.000	0.003
		0.5	0.007	0.059	0.035	0.006	0.021	0.004	0.007
		0.7	0.015	0.044	0.040	0.017	0.043	0.011	0.015
		0.9	0.033	0.043	0.049	0.041	0.055	0.019	0.021
		0.99	0.051	0.050	0.057	0.049	0.044	0.023	0.032

Table 5.8: Sizes of all the tests at the 5% level for *X15*, selected β, σ, and ρ values with $n = 40$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.025	0.1	0.004	0.022	0.018	0.000	0.000	0.048	0.035
		0.3	0.007	0.019	0.016	0.000	0.000	0.052	0.038
		0.5	0.007	0.020	0.018	0.000	0.000	0.046	0.039
		0.7	0.006	0.025	0.026	0.000	0.001	0.041	0.031
		0.9	0.009	0.026	0.029	0.003	0.007	0.046	0.029
		0.99	0.017	0.032	0.034	0.007	0.015	0.051	0.031
	0.045	0.1	0.004	0.020	0.018	0.000	0.000	0.019	0.041
		0.3	0.005	0.025	0.021	0.000	0.001	0.020	0.046
		0.5	0.007	0.025	0.028	0.000	0.001	0.022	0.048
		0.7	0.008	0.027	0.032	0.001	0.005	0.023	0.049
		0.9	0.019	0.036	0.043	0.010	0.017	0.040	0.048
		0.99	0.037	0.042	0.050	0.023	0.043	0.051	0.057
	0.2	0.1	0.001	0.040	0.019	0.001	0.001	0.000	0.006
		0.3	0.003	0.037	0.030	0.001	0.001	0.000	0.001
		0.5	0.004	0.033	0.037	0.002	0.003	0.000	0.005
		0.7	0.007	0.035	0.044	0.003	0.007	0.005	0.028
		0.9	0.031	0.039	0.047	0.017	0.022	0.031	0.050
		0.99	0.078	0.047	0.060	0.046	0.070	0.045	0.068
(0, -1, 1)'	0.025	0.1	0.004	0.031	0.027	0.000	0.000	0.038	0.031
		0.3	0.004	0.032	0.030	0.001	0.001	0.049	0.040
		0.5	0.004	0.034	0.032	0.000	0.001	0.057	0.055
		0.7	0.010	0.030	0.036	0.001	0.002	0.061	0.057
		0.9	0.013	0.031	0.035	0.004	0.006	0.059	0.049
		0.99	0.019	0.034	0.044	0.015	0.026	0.056	0.043
	0.045	0.1	0.002	0.032	0.027	0.001	0.001	0.008	0.021
		0.3	0.002	0.029	0.028	0.001	0.001	0.017	0.029
		0.5	0.003	0.031	0.035	0.001	0.002	0.022	0.047
		0.7	0.008	0.038	0.042	0.003	0.004	0.027	0.057
		0.9	0.023	0.036	0.045	0.012	0.021	0.045	0.057
		0.99	0.046	0.046	0.052	0.038	0.055	0.050	0.058
	0.2	0.1	0.001	0.047	0.023	0.001	0.001	0.000	0.005
		0.3	0.002	0.038	0.032	0.002	0.002	0.000	0.001
		0.5	0.002	0.032	0.037	0.002	0.004	0.000	0.005
		0.7	0.004	0.034	0.042	0.008	0.011	0.006	0.026
		0.9	0.028	0.043	0.050	0.018	0.027	0.028	0.057
		0.99	0.080	0.048	0.058	0.046	0.080	0.042	0.069

Table 5.9: Sizes of all the tests at the 5% level for *X15*, selected β, σ, and ρ values with $n = 40$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 0)'	0.025	0.1	0.006	0.028	0.022	0.000	0.000	0.033	0.043
		0.3	0.008	0.029	0.025	0.000	0.001	0.035	0.050
		0.5	0.009	0.025	0.025	0.000	0.001	0.038	0.059
		0.7	0.008	0.026	0.029	0.001	0.004	0.041	0.053
		0.9	0.014	0.037	0.041	0.005	0.013	0.049	0.051
		0.99	0.029	0.040	0.050	0.021	0.035	0.051	0.047
	0.045	0.1	0.003	0.028	0.023	0.000	0.000	0.003	0.023
		0.3	0.006	0.028	0.029	0.001	0.001	0.007	0.030
		0.5	0.007	0.026	0.033	0.000	0.002	0.009	0.036
		0.7	0.009	0.032	0.038	0.002	0.004	0.013	0.049
		0.9	0.024	0.041	0.051	0.011	0.016	0.041	0.056
		0.99	0.062	0.047	0.057	0.035	0.051	0.049	0.068
	0.2	0.1	0.001	0.049	0.025	0.000	0.000	0.000	0.006
		0.3	0.003	0.038	0.028	0.002	0.002	0.000	0.001
		0.5	0.003	0.032	0.037	0.004	0.004	0.000	0.004
		0.7	0.007	0.036	0.045	0.006	0.009	0.006	0.025
		0.9	0.026	0.040	0.047	0.017	0.023	0.029	0.052
		0.99	0.080	0.047	0.058	0.052	0.078	0.049	0.072
(0, 0, -1)'	0.025	0.1	0.004	0.025	0.021	0.000	0.000	0.024	0.032
		0.3	0.004	0.025	0.023	0.000	0.000	0.028	0.038
		0.5	0.003	0.026	0.029	0.001	0.001	0.034	0.048
		0.7	0.008	0.025	0.027	0.002	0.002	0.048	0.063
		0.9	0.021	0.036	0.044	0.005	0.012	0.052	0.052
		0.99	0.034	0.044	0.050	0.020	0.036	0.051	0.045
	0.045	0.1	0.002	0.025	0.022	0.001	0.001	0.004	0.013
		0.3	0.002	0.029	0.028	0.000	0.001	0.006	0.020
		0.5	0.002	0.031	0.033	0.001	0.002	0.011	0.031
		0.7	0.003	0.030	0.037	0.002	0.003	0.025	0.052
		0.9	0.024	0.039	0.047	0.011	0.018	0.039	0.058
		0.99	0.056	0.043	0.055	0.033	0.053	0.051	0.060
	0.2	0.1	0.000	0.050	0.025	0.000	0.000	0.000	0.005
		0.3	0.002	0.044	0.033	0.001	0.002	0.000	0.001
		0.5	0.002	0.032	0.036	0.003	0.004	0.002	0.004
		0.7	0.005	0.033	0.040	0.007	0.008	0.007	0.026
		0.9	0.028	0.041	0.050	0.018	0.024	0.030	0.054
		0.99	0.081	0.048	0.059	0.046	0.069	0.047	0.066
(0, 0, 0)'	0.025	0.1	0.000	0.045	0.022	0.000	0.000	0.000	0.005
		0.3	0.002	0.043	0.031	0.002	0.002	0.000	0.001
		0.5	0.002	0.032	0.037	0.004	0.006	0.000	0.003
		0.7	0.006	0.034	0.042	0.009	0.015	0.005	0.022
		0.9	0.030	0.040	0.050	0.018	0.028	0.029	0.052
		0.99	0.082	0.048	0.059	0.048	0.082	0.048	0.069

Table 5.10: Sizes of all the tests at the 5% level for *X16*, selected β, σ, and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.4	0.1	0.007	0.021	0.021	0.000	0.000	0.012	0.001
		0.3	0.009	0.023	0.023	0.000	0.000	0.009	0.000
		0.5	0.009	0.028	0.026	0.000	0.000	0.009	0.000
		0.7	0.009	0.025	0.024	0.000	0.000	0.008	0.000
		0.9	0.005	0.018	0.023	0.000	0.000	0.015	0.000
		0.99	0.005	0.018	0.021	0.000	0.000	0.021	0.000
	0.8	0.1	0.005	0.022	0.021	0.000	0.000	0.045	0.017
		0.3	0.008	0.024	0.028	0.000	0.000	0.044	0.013
		0.5	0.009	0.027	0.031	0.000	0.000	0.041	0.015
		0.7	0.009	0.023	0.025	0.000	0.000	0.042	0.014
		0.9	0.008	0.019	0.025	0.000	0.000	0.040	0.018
		0.99	0.008	0.016	0.026	0.001	0.001	0.042	0.019
	10	0.1	0.003	0.047	0.026	0.001	0.001	0.009	0.027
		0.3	0.005	0.052	0.035	0.002	0.002	0.006	0.024
		0.5	0.010	0.045	0.040	0.003	0.003	0.004	0.018
		0.7	0.015	0.029	0.044	0.008	0.006	0.002	0.016
		0.9	0.035	0.014	0.046	0.023	0.020	0.003	0.016
		0.99	0.064	0.017	0.052	0.055	0.052	0.002	0.013
(0, -1, -1)'	0.4	0.1	0.003	0.025	0.023	0.000	0.000	0.009	0.001
		0.3	0.003	0.030	0.029	0.000	0.000	0.006	0.000
		0.5	0.004	0.027	0.027	0.000	0.000	0.005	0.000
		0.7	0.005	0.023	0.026	0.000	0.000	0.006	0.000
		0.9	0.005	0.021	0.024	0.000	0.000	0.013	0.000
		0.99	0.006	0.022	0.023	0.000	0.000	0.018	0.000
	0.8	0.1	0.004	0.024	0.022	0.000	0.000	0.053	0.015
		0.3	0.004	0.032	0.033	0.000	0.000	0.047	0.011
		0.5	0.005	0.027	0.032	0.000	0.000	0.042	0.011
		0.7	0.007	0.025	0.031	0.000	0.000	0.038	0.011
		0.9	0.008	0.023	0.029	0.000	0.000	0.035	0.017
		0.99	0.006	0.021	0.030	0.001	0.001	0.039	0.019
	10	0.1	0.003	0.047	0.019	0.000	0.000	0.009	0.023
		0.3	0.005	0.055	0.040	0.001	0.000	0.004	0.020
		0.5	0.007	0.052	0.049	0.002	0.001	0.004	0.019
		0.7	0.016	0.034	0.052	0.010	0.008	0.003	0.015
		0.9	0.038	0.017	0.051	0.032	0.029	0.003	0.014
		0.99	0.069	0.015	0.061	0.060	0.056	0.002	0.011
(0, -1, 1)'	0.4	0.1	0.004	0.017	0.016	0.000	0.000	0.008	0.001
		0.3	0.005	0.023	0.023	0.000	0.000	0.008	0.000
		0.5	0.004	0.022	0.023	0.000	0.000	0.010	0.000
		0.7	0.005	0.026	0.027	0.000	0.000	0.017	0.000
		0.9	0.005	0.024	0.025	0.000	0.000	0.029	0.000
		0.99	0.005	0.022	0.022	0.000	0.000	0.032	0.001

Table 5.11: Sizes of all the tests at the 5% level for *X16*, selected β, σ, and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, -1, 1)'	0.8	0.1	0.002	0.019	0.016	0.000	0.000	0.046	0.018
		0.3	0.006	0.022	0.024	0.000	0.000	0.044	0.016
		0.5	0.008	0.022	0.023	0.000	0.000	0.045	0.017
		0.7	0.007	0.024	0.026	0.000	0.000	0.052	0.023
		0.9	0.008	0.021	0.026	0.000	0.000	0.046	0.025
		0.99	0.008	0.019	0.023	0.000	0.000	0.049	0.023
	10	0.1	0.004	0.044	0.023	0.000	0.000	0.009	0.023
		0.3	0.004	0.058	0.039	0.000	0.000	0.007	0.022
		0.5	0.007	0.045	0.042	0.002	0.002	0.004	0.018
		0.7	0.016	0.027	0.041	0.010	0.008	0.002	0.014
		0.9	0.036	0.014	0.046	0.029	0.026	0.003	0.012
		0.99	0.063	0.014	0.058	0.051	0.048	0.004	0.012
(0, 1, -1)'	0.4	0.1	0.002	0.015	0.014	0.000	0.000	0.006	0.001
		0.3	0.003	0.015	0.014	0.000	0.000	0.005	0.000
		0.5	0.003	0.017	0.017	0.000	0.000	0.006	0.001
		0.7	0.003	0.016	0.020	0.000	0.000	0.013	0.001
		0.9	0.005	0.017	0.018	0.000	0.000	0.020	0.001
		0.99	0.003	0.018	0.019	0.000	0.000	0.029	0.000
	0.8	0.1	0.002	0.016	0.013	0.000	0.000	0.046	0.013
		0.3	0.002	0.017	0.013	0.000	0.000	0.039	0.009
		0.5	0.004	0.018	0.018	0.000	0.000	0.040	0.010
		0.7	0.006	0.016	0.022	0.000	0.000	0.043	0.015
		0.9	0.006	0.016	0.023	0.000	0.000	0.045	0.017
		0.99	0.006	0.014	0.021	0.001	0.001	0.052	0.023
	10	0.1	0.002	0.044	0.020	0.000	0.900	0.005	0.022
		0.3	0.005	0.056	0.039	0.002	0.001	0.004	0.014
		0.5	0.010	0.047	0.050	0.005	0.004	0.004	0.015
		0.7	0.021	0.029	0.050	0.012	0.011	0.004	0.014
		0.9	0.038	0.013	0.050	0.033	0.028	0.002	0.014
		0.99	0.064	0.016	0.056	0.057	0.053	0.001	0.011
(0, 1, 0)'	0.4	0.1	0.006	0.022	0.020	0.000	0.000	0.024	0.002
		0.3	0.006	0.023	0.022	0.000	0.000	0.024	0.001
		0.5	0.005	0.022	0.025	0.000	0.000	0.024	0.001
		0.7	0.008	0.021	0.023	0.000	0.000	0.031	0.002
		0.9	0.007	0.023	0.024	0.000	0.000	0.053	0.007
		0.99	0.005	0.021	0.027	0.000	0.000	0.082	0.008
	0.8	0.1	0.005	0.025	0.023	0.000	0.000	0.052	0.030
		0.3	0.007	0.024	0.027	0.000	0.000	0.050	0.028
		0.5	0.008	0.024	0.030	0.000	0.000	0.047	0.029
		0.7	0.008	0.022	0.030	0.000	0.000	0.047	0.034
		0.9	0.012	0.019	0.029	0.001	0.001	0.053	0.044
		0.99	0.010	0.021	0.030	0.006	0.005	0.067	0.057

Table 5.12: Sizes of all the tests at the 5% level for *X16*, selected β, σ, and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 0)'	10	0.1	0.003	0.047	0.023	0.001	0.000	0.006	0.025
		0.3	0.004	0.051	0.037	0.002	0.001	0.004	0.021
		0.5	0.010	0.046	0.045	0.003	0.003	0.003	0.018
		0.7	0.017	0.027	0.048	0.010	0.009	0.003	0.015
		0.9	0.035	0.014	0.045	0.028	0.022	0.003	0.013
		0.99	0.064	0.017	0.055	0.058	0.054	0.002	0.013
(0, -1, 0)'	0.4	0.1	0.004	0.019	0.015	0.000	0.000	0.017	0.001
		0.3	0.006	0.023	0.022	0.000	0.000	0.016	0.000
		0.5	0.006	0.026	0.026	0.000	0.000	0.026	0.001
		0.7	0.005	0.026	0.027	0.000	0.000	0.038	0.001
		0.9	0.006	0.023	0.026	0.000	0.000	0.054	0.006
		0.99	0.007	0.022	0.026	0.000	0.000	0.082	0.010
	0.8	0.1	0.003	0.017	0.014	0.000	0.000	0.050	0.023
		0.3	0.005	0.026	0.026	0.000	0.000	0.050	0.020
		0.5	0.008	0.026	0.030	0.000	0.000	0.054	0.025
		0.7	0.009	0.026	0.038	0.000	0.000	0.059	0.036
		0.9	0.008	0.022	0.031	0.001	0.001	0.056	0.045
		0.99	0.009	0.024	0.036	0.002	0.001	0.063	0.056
	10	0.1	0.004	0.051	0.022	0.000	0.000	0.009	0.022
		0.3	0.003	0.055	0.039	0.000	0.000	0.005	0.021
		0.5	0.006	0.049	0.046	0.002	0.002	0.003	0.017
		0.7	0.018	0.034	0.049	0.010	0.009	0.003	0.014
		0.9	0.035	0.015	0.046	0.031	0.029	0.003	0.013
		0.99	0.068	0.016	0.059	0.054	0.055	0.003	0.013
(0, 0, 1)'	0.4	0.1	0.003	0.024	0.023	0.000	0.000	0.042	0.009
		0.3	0.003	0.024	0.023	0.000	0.000	0.032	0.008
		0.5	0.004	0.021	0.025	0.000	0.000	0.023	0.005
		0.7	0.003	0.019	0.021	0.000	0.000	0.015	0.004
		0.9	0.002	0.017	0.016	0.000	0.000	0.010	0.000
		0.99	0.002	0.015	0.018	0.001	0.001	0.007	0.000
	0.8	0.1	0.005	0.028	0.023	0.000	0.000	0.064	0.056
		0.3	0.007	0.028	0.024	0.000	0.000	0.057	0.051
		0.5	0.007	0.022	0.026	0.001	0.000	0.037	0.040
		0.7	0.006	0.019	0.025	0.002	0.002	0.021	0.030
		0.9	0.006	0.013	0.021	0.007	0.006	0.014	0.016
		0.99	0.008	0.012	0.021	0.016	0.016	0.008	0.012
	10	0.1	0.004	0.044	0.023	0.000	0.000	0.008	0.024
		0.3	0.005	0.056	0.042	0.001	0.000	0.004	0.020
		0.5	0.009	0.040	0.042	0.004	0.003	0.003	0.018
		0.7	0.014	0.030	0.045	0.010	0.008	0.003	0.015
		0.9	0.033	0.016	0.045	0.031	0.025	0.002	0.011
		0.99	0.061	0.017	0.056	0.055	0.054	0.002	0.011

Table 5.13: Sizes of all the tests at the 5% level for *X16*, selected β, σ, and ρ values with $n = 20$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 0, -1)'	0.4	0.1	0.003	0.022	0.018	0.000	0.000	0.039	0.010
		0.3	0.003	0.021	0.019	0.000	0.000	0.030	0.005
		0.5	0.004	0.023	0.026	0.000	0.000	0.023	0.001
		0.7	0.004	0.022	0.027	0.000	0.000	0.012	0.001
		0.9	0.005	0.021	0.026	0.000	0.000	0.007	0.002
		0.99	0.006	0.021	0.025	0.001	0.001	0.005	0.001
	0.8	0.1	0.003	0.026	0.020	0.000	0.000	0.060	0.054
		0.3	0.005	0.025	0.025	0.000	0.000	0.049	0.047
		0.5	0.005	0.025	0.029	0.001	0.000	0.036	0.038
		0.7	0.007	0.021	0.031	0.001	0.001	0.019	0.026
		0.9	0.007	0.017	0.031	0.006	0.006	0.008	0.013
		0.99	0.005	0.018	0.030	0.018	0.018	0.008	0.011
	10	0.1	0.004	0.044	0.021	0.000	0.000	0.008	0.022
		0.3	0.005	0.052	0.042	0.001	0.000	0.003	0.018
		0.5	0.009	0.049	0.053	0.004	0.003	0.003	0.015
		0.7	0.019	0.030	0.053	0.011	0.008	0.003	0.012
		0.9	0.037	0.015	0.051	0.033	0.028	0.001	0.010
		0.99	0.066	0.016	0.057	0.055	0.055	0.001	0.011
(0, 0, 0)'	0.4	0.1	0.003	0.039	0.023	0.000	0.000	0.006	0.021
		0.3	0.005	0.050	0.044	0.001	0.000	0.004	0.017
		0.5	0.009	0.044	0.046	0.004	0.003	0.003	0.013
		0.7	0.019	0.034	0.051	0.011	0.009	0.003	0.012
		0.9	0.036	0.014	0.051	0.031	0.026	0.002	0.009
		0.99	0.065	0.015	0.059	0.055	0.053	0.001	0.009

Table 5.14 : Sizes of all the tests at the 5% level for *X16*, selected β, σ, and ρ values with $n = 40$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.4	0.1	0.001	0.011	0.023	0.000	0.000	0.001	0.000
		0.3	0.001	0.013	0.027	0.000	0.000	0.000	0.000
		0.5	0.001	0.013	0.024	0.000	0.000	0.001	0.000
		0.7	0.001	0.013	0.025	0.000	0.000	0.002	0.000
		0.9	0.000	0.009	0.029	0.000	0.000	0.004	0.000
		0.99	0.000	0.007	0.032	0.000	0.000	0.018	0.015
	0.7	0.1	0.001	0.012	0.024	0.000	0.000	0.014	0.001
		0.3	0.002	0.014	0.028	0.000	0.000	0.015	0.000
		0.5	0.001	0.014	0.027	0.000	0.000	0.018	0.000
		0.7	0.001	0.009	0.027	0.000	0.000	0.019	0.001
		0.9	0.000	0.009	0.030	0.000	0.000	0.025	0.002
		0.99	0.000	0.009	0.033	0.002	0.001	0.033	0.018
	1.81	0.1	0.001	0.013	0.023	0.000	0.000	0.056	0.030
		0.3	0.001	0.015	0.035	0.000	0.000	0.055	0.032
		0.5	0.002	0.013	0.036	0.000	0.000	0.048	0.031
		0.7	0.004	0.009	0.041	0.000	0.000	0.042	0.031
		0.9	0.003	0.008	0.038	0.017	0.003	0.032	0.030
		0.99	0.002	0.009	0.040	0.086	0.030	0.034	0.036
(0, -1, 1)'	0.4	0.1	0.000	0.006	0.016	0.000	0.000	0.000	0.000
		0.3	0.001	0.009	0.017	0.000	0.000	0.000	0.000
		0.5	0.001	0.008	0.022	0.000	0.000	0.001	0.000
		0.7	0.001	0.009	0.025	0.000	0.000	0.006	0.000
		0.9	0.001	0.008	0.022	0.000	0.000	0.029	0.000
		0.99	0.001	0.008	0.023	0.002	0.001	0.060	0.016
	0.7	0.1	0.000	0.006	0.018	0.000	0.000	0.016	0.001
		0.3	0.001	0.008	0.018	0.000	0.000	0.014	0.000
		0.5	0.001	0.010	0.025	0.000	0.000	0.019	0.001
		0.7	0.001	0.009	0.029	0.000	0.000	0.033	0.003
		0.9	0.001	0.008	0.026	0.004	0.002	0.064	0.011
		0.99	0.001	0.009	0.028	0.034	0.021	0.078	0.038
	1.81	0.1	0.000	0.010	0.021	0.001	0.000	0.053	0.036
		0.3	0.000	0.013	0.030	0.001	0.000	0.055	0.037
		0.5	0.001	0.007	0.035	0.002	0.000	0.046	0.037
		0.7	0.003	0.005	0.038	0.007	0.000	0.050	0.034
		0.9	0.006	0.005	0.036	0.067	0.013	0.045	0.044
		0.99	0.007	0.007	0.044	0.155	0.060	0.047	0.058

Table 5.15: Sizes of all the tests at the 5% level for *X16*, selected β, σ, and ρ values with $n = 40$, using near exact non-similar critical values

β	σ	ρ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 0)'	0.4	0.1	0.000	0.009	0.020	0.000	0.000	0.003	0.000
		0.3	0.000	0.009	0.024	0.000	0.000	0.002	0.000
		0.5	0.002	0.010	0.026	0.000	0.000	0.004	0.000
		0.7	0.001	0.010	0.027	0.000	0.000	0.006	0.000
		0.9	0.001	0.008	0.026	0.000	0.000	0.015	0.000
		0.99	0.001	0.007	0.026	0.002	0.000	0.028	0.015
	0.7	0.1	0.001	0.008	0.020	0.000	0.000	0.030	0.003
		0.3	0.000	0.010	0.026	0.000	0.000	0.032	0.003
		0.5	0.002	0.010	0.030	0.000	0.000	0.029	0.002
		0.7	0.002	0.011	0.031	0.000	0.000	0.030	0.005
		0.9	0.002	0.009	0.029	0.003	0.000	0.037	0.012
		0.99	0.002	0.007	0.033	0.023	0.012	0.043	0.025
	1.81	0.1	0.000	0.013	0.025	0.000	0.000	0.047	0.043
		0.3	0.000	0.014	0.034	0.000	0.000	0.048	0.044
		0.5	0.002	0.012	0.040	0.003	0.001	0.042	0.039
		0.7	0.005	0.011	0.045	0.009	0.000	0.031	0.036
		0.9	0.010	0.008	0.041	0.044	0.011	0.028	0.038
		0.99	0.013	0.006	0.045	0.178	0.087	0.034	0.043
(0, 0, -1)'	0.4	0.1	0.002	0.009	0.020	0.000	0.000	0.015	0.001
		0.3	0.001	0.008	0.023	0.000	0.000	0.014	0.001
		0.5	0.000	0.007	0.026	0.000	0.000	0.018	0.000
		0.7	0.002	0.007	0.024	0.000	0.000	0.029	0.001
		0.9	0.002	0.006	0.022	0.000	0.000	0.061	0.006
		0.99	0.001	0.007	0.023	0.005	0.000	0.077	0.026
	0.7	0.1	0.001	0.009	0.020	0.000	0.000	0.048	0.014
		0.3	0.001	0.010	0.025	0.000	0.000	0.048	0.013
		0.5	0.001	0.008	0.030	0.000	0.000	0.049	0.010
		0.7	0.002	0.007	0.031	0.000	0.000	0.061	0.017
		0.9	0.002	0.006	0.027	0.005	0.000	0.068	0.031
		0.99	0.002	0.006	0.026	0.030	0.004	0.073	0.055
	1.81	0.1	0.000	0.022	0.025	0.001	0.000	0.035	0.044
		0.3	0.000	0.022	0.038	0.001	0.000	0.036	0.041
		0.5	0.002	0.014	0.047	0.001	0.000	0.034	0.042
		0.7	0.004	0.011	0.046	0.005	0.001	0.037	0.042
		0.9	0.009	0.006	0.040	0.022	0.002	0.026	0.041
		0.99	0.016	0.006	0.038	0.093	0.032	0.033	0.048
(0, 0, 0)'	0.4	0.1	0.001	0.044	0.034	0.002	0.000	0.006	0.017
		0.3	0.001	0.051	0.045	0.005	0.000	0.006	0.017
		0.5	0.004	0.029	0.052	0.006	0.000	0.006	0.019
		0.7	0.006	0.014	0.048	0.007	0.000	0.006	0.016
		0.9	0.031	0.007	0.050	0.047	0.005	0.002	0.012
		0.99	0.069	0.005	0.046	0.134	0.045	0.017	0.022

Table 5.16 : Powers of all the tests at the 5% level for *X15*, selected β, σ, and μ values with $n = 20$, using near exact non-similar critical values

β	σ	μ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.015	0.1	0.024	0.078	0.037	0.000	0.000	0.158	0.090
		0.3	0.239	0.383	0.255	0.001	0.001	0.637	0.432
		0.5	0.895	0.710	0.711	0.017	0.025	0.991	0.955
		0.7	1.000	0.981	0.986	0.478	0.553	0.996	1.000
		0.9	1.000	1.000	1.000	0.986	0.983	0.790	1.000
	0.025								
		0.1	0.016	0.059	0.025	0.000	0.001	0.082	0.083
		0.3	0.076	0.177	0.109	0.000	0.001	0.337	0.285
		0.5	0.438	0.421	0.316	0.003	0.005	0.805	0.722
		0.7	0.961	0.728	0.677	0.084	0.123	0.928	0.981
		0.9	1.000	0.999	0.999	0.815	0.848	0.687	0.963
(0, -1, -1)'	0.015	0.1	0.022	0.088	0.032	0.000	0.000	0.169	0.099
		0.3	0.261	0.418	0.294	0.000	0.000	0.654	0.458
		0.5	0.905	0.687	0.722	0.015	0.021	0.989	0.956
		0.7	1.000	0.978	0.991	0.491	0.571	0.992	1.000
		0.9	1.000	1.000	1.000	0.991	0.988	0.811	0.999
	0.025								
		0.1	0.010	0.057	0.023	0.000	0.000	0.087	0.082
		0.3	0.079	0.200	0.112	0.000	0.000	0.367	0.310
		0.5	0.462	0.449	0.354	0.001	0.003	0.802	0.734
		0.7	0.959	0.710	0.682	0.091	0.125	0.933	0.974
		0.9	1.000	1.000	1.000	0.816	0.854	0.717	0.963
(0, -1, 1)'	0.015	0.1	0.016	0.062	0.025	0.000	0.000	0.102	0.036
		0.3	0.119	0.224	0.130	0.003	0.004	0.509	0.276
		0.5	0.491	0.403	0.345	0.048	0.072	0.882	0.803
		0.7	0.907	0.713	0.728	0.480	0.589	0.319	0.800
		0.9	0.999	0.945	0.959	0.952	0.975	0.001	0.015
	0.025	0.1	0.009	0.046	0.020	0.001	0.000	0.048	0.029
		0.3	0.035	0.110	0.064	0.002	0.002	0.221	0.145
		0.5	0.153	0.173	0.134	0.012	0.016	0.505	0.462
		0.7	0.468	0.265	0.276	0.149	0.211	0.385	0.571
		0.9	0.851	0.473	0.537	0.614	0.737	0.024	0.076
(0, 1, -1)'	0.015	0.1	0.017	0.057	0.023	0.000	0.000	0.104	0.028
		0.3	0.122	0.219	0.132	0.001	0.001	0.512	0.256
		0.5	0.504	0.400	0.332	0.060	0.078	0.889	0.802
		0.7	0.900	0.719	0.727	0.463	0.563	0.329	0.799
		0.9	0.997	0.938	0.956	0.943	0.970	0.001	0.016
	0.025	0.1	0.010	0.048	0.019	0.000	0.000	0.043	0.022
		0.3	0.031	0.111	0.061	0.000	0.000	0.221	0.152
		0.5	0.160	0.168	0.134	0.013	0.018	0.512	0.465
		0.7	0.464	0.254	0.271	0.155	0.210	0.372	0.580
		0.9	0.837	0.482	0.533	0.616	0.727	0.026	0.082

Table 5.17 : Powers of all the tests at the 5% level for *X15*, selected β, σ, and μ values with $n = 20$, using near exact non-similar critical values

β	σ	μ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 0)'	0.015	0.1	0.013	0.052	0.022	0.000	0.000	0.058	0.073
		0.3	0.050	0.125	0.067	0.000	0.001	0.199	0.200
		0.5	0.228	0.311	0.200	0.006	0.008	0.563	0.547
		0.7	0.826	0.589	0.512	0.031	0.046	0.887	0.909
		0.9	1.000	0.932	0.972	0.249	0.357	0.962	0.995
	0.025	0.1	0.008	0.049	0.020	0.000	0.000	0.013	0.026
		0.3	0.021	0.081	0.042	0.000	0.000	0.050	0.083
		0.5	0.064	0.143	0.081	0.002	0.006	0.175	0.233
		0.7	0.382	0.286	0.197	0.011	0.019	0.497	0.569
		0.9	0.981	0.685	0.625	0.030	0.051	0.833	0.904
(0, -1, 0)'	0.015	0.1	0.010	0.053	0.020	0.000	0.001	0.058	0.077
		0.3	0.043	0.135	0.075	0.001	0.001	0.231	0.238
		0.5	0.257	0.329	0.222	0.002	0.004	0.583	0.574
		0.7	0.841	0.599	0.532	0.020	0.046	0.873	0.911
		0.9	1.000	0.938	0.969	0.257	0.368	0.965	0.996
	0.025	0.1	0.004	0.052	0.018	0.000	0.001	0.009	0.025
		0.3	0.013	0.085	0.048	0.000	0.000	0.049	0.081
		0.5	0.071	0.149	0.094	0.002	0.004	0.196	0.273
		0.7	0.406	0.309	0.221	0.005	0.011	0.523	0.591
		0.9	0.977	0.674	0.620	0.036	0.058	0.832	0.905
(0, 0, 1)'	0.015	0.1	0.016	0.058	0.032	0.001	0.001	0.113	0.056
		0.3	0.131	0.226	0.148	0.001	0.002	0.491	0.300
		0.5	0.593	0.409	0.352	0.022	0.033	0.909	0.831
		0.7	0.979	0.819	0.812	0.382	0.470	0.672	0.960
		0.9	1.000	1.000	1.000	0.960	0.970	0.016	0.293
	0.025	0.1	0.007	0.047	0.021	0.001	0.001	0.047	0.037
		0.3	0.033	0.124	0.074	0.001	0.002	0.216	0.171
		0.5	0.195	0.211	0.164	0.005	0.009	0.534	0.497
		0.7	0.666	0.306	0.306	0.101	0.147	0.602	0.758
		0.9	0.992	0.844	0.863	0.629	0.712	0.103	0.348
(0, 0, -1)'	0.015	0.1	0.011	0.061	0.029	0.000	0.000	0.115	0.055
		0.3	0.130	0.251	0.161	0.000	0.000	0.502	0.324
		0.5	0.610	0.439	0.366	0.028	0.041	0.909	0.837
		0.7	0.975	0.815	0.811	0.387	0.473	0.692	0.956
		0.9	1.000	1.000	1.000	0.964	0.973	0.018	0.316
	0.025	0.1	0.005	0.051	0.023	0.000	0.000	0.042	0.033
		0.3	0.036	0.123	0.071	0.000	0.000	0.231	0.176
		0.5	0.206	0.239	0.169	0.003	0.007	0.575	0.523
		0.7	0.679	0.325	0.332	0.110	0.150	0.623	0.770
		0.9	0.987	0.834	0.851	0.627	0.720	0.126	0.362

Table 5.18 : Powers of all the tests at the 5% level for *X15*, selected β, σ, and μ values with $n = 40$, using near exact non-similar critical values

β	σ	μ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.025	0.1	0.013	0.063	0.052	0.000	0.000	0.132	0.088
		0.3	0.233	0.421	0.380	0.000	0.001	0.615	0.441
		0.5	0.920	0.487	0.476	0.011	0.025	0.989	0.973
		0.7	1.000	0.977	0.977	0.876	0.951	0.602	0.988
		0.9	1.000	1.000	1.000	1.000	1.000	0.000	0.000
	0.045	0.1	0.005	0.038	0.033	0.000	0.000	0.044	0.076
		0.3	0.042	0.155	0.137	0.001	0.001	0.255	0.281
		0.5	0.397	0.373	0.355	0.001	0.006	0.779	0.761
		0.7	0.939	0.253	0.273	0.149	0.284	0.722	0.899
		0.9	1.000	1.000	1.000	1.000	1.000	0.000	0.002
(0, -1, 1)'	0.025	0.1	0.012	0.053	0.048	0.000	0.000	0.077	0.053
		0.3	0.100	0.249	0.227	0.000	0.000	0.446	0.326
		0.5	0.618	0.412	0.394	0.002	0.009	0.942	0.870
		0.7	0.970	0.451	0.472	0.382	0.590	0.453	0.845
		0.9	1.000	1.000	1.000	1.000	1.000	0.000	0.000
	0.045	0.1	0.003	0.043	0.038	0.001	0.001	0.018	0.035
		0.3	0.022	0.095	0.092	0.001	0.001	0.126	0.159
		0.5	0.149	0.199	0.190	0.001	0.002	0.507	0.534
		0.7	0.607	0.131	0.138	0.028	0.076	0.637	0.725
		0.9	0.995	0.742	0.777	0.966	0.990	0.002	0.027
(0, 1, 0)'	0.025	0.1	0.012	0.051	0.043	0.000	0.000	0.074	0.078
		0.3	0.086	0.221	0.202	0.001	0.001	0.384	0.331
		0.5	0.611	0.432	0.413	0.004	0.006	0.910	0.860
		0.7	0.979	0.485	0.508	0.389	0.583	0.567	0.892
		0.9	1.000	1.000	1.000	1.000	1.000	0.000	0.000
	0.045	0.1	0.004	0.039	0.032	0.000	0.000	0.012	0.040
		0.3	0.016	0.088	0.084	0.000	0.001	0.086	0.157
		0.5	0.138	0.208	0.199	0.002	0.004	0.450	0.533
		0.7	0.702	0.107	0.113	0.033	0.085	0.691	0.762
		0.9	1.000	0.987	0.991	0.999	0.999	0.000	0.006
(0, 0, -1)'	0.025	0.1	0.006	0.045	0.036	0.000	0.000	0.047	0.058
		0.3	0.036	0.169	0.156	0.000	0.000	0.270	0.277
		0.5	0.354	0.323	0.306	0.001	0.003	0.763	0.725
		0.7	0.876	0.275	0.290	0.108	0.205	0.743	0.869
		0.9	1.000	0.992	0.994	0.997	0.999	0.000	0.038
	0.045	0.1	0.004	0.036	0.029	0.000	0.000	0.006	0.021
		0.3	0.006	0.075	0.068	0.000	0.001	0.043	0.087
		0.5	0.045	0.134	0.130	0.001	0.002	0.214	0.364
		0.7	0.348	0.128	0.133	0.015	0.043	0.519	0.630
		0.9	0.959	0.447	0.491	0.712	0.844	0.073	0.151

Table 5.19 : Powers of all the tests at the 5% level for *X16*, selected β, σ, and μ values with $n = 20$, using near exact non-similar critical values

β	σ	μ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.4	0.1	0.060	0.193	0.198	0.000	0.000	0.132	0.007
		0.3	0.750	0.533	0.885	0.000	0.000	0.935	0.506
		0.5	0.989	0.587	0.967	0.055	0.042	1.000	0.998
		0.7	0.999	0.972	0.988	0.863	0.840	0.999	1.000
		0.9	1.000	1.000	1.000	1.000	1.000	0.047	0.001
	0.8	0.1	0.018	0.070	0.072	0.000	0.000	0.153	0.057
		0.3	0.171	0.341	0.372	0.000	0.000	0.558	0.389
		0.5	0.569	0.502	0.577	0.001	0.001	0.873	0.855
		0.7	0.876	0.410	0.508	0.159	0.137	0.828	0.969
		0.9	0.994	0.629	0.878	0.992	0.991	0.038	0.048
(0, -1, -1)'	0.4	0.1	0.057	0.196	0.198	0.000	0.000	0.131	0.005
		0.3	0.758	0.560	0.877	0.001	0.000	0.933	0.509
		0.5	0.989	0.614	0.963	0.061	0.049	1.000	0.997
		0.7	0.999	0.966	0.987	0.881	0.857	1.000	1.000
		0.9	1.000	1.000	1.000	1.000	1.000	0.043	0.000
	0.8	0.1	0.017	0.071	0.070	0.000	0.000	0.156	0.054
		0.3	0.175	0.335	0.368	0.000	0.000	0.571	0.377
		0.5	0.582	0.490	0.579	0.003	0.003	0.871	0.855
		0.7	0.873	0.395	0.493	0.155	0.139	0.842	0.971
		0.9	0.993	0.628	0.876	0.993	0.993	0.039	0.058
(0, -1, 1)'	0.4	0.1	0.044	0.173	0.174	0.000	0.000	0.144	0.009
		0.3	0.638	0.558	0.811	0.000	0.000	0.893	0.559
		0.5	0.964	0.540	0.970	0.010	0.008	1.000	0.987
		0.7	0.998	0.816	0.994	0.144	0.135	0.992	1.000
		0.9	1.000	0.984	0.994	0.657	0.697	0.332	0.957
	0.8	0.1	0.012	0.062	0.058	0.000	0.000	0.148	0.073
		0.3	0.128	0.282	0.293	0.001	0.001	0.493	0.390
		0.5	0.407	0.470	0.531	0.003	0.002	0.720	0.760
		0.7	0.702	0.523	0.591	0.023	0.020	0.600	0.878
		0.9	0.893	0.437	0.572	0.132	0.145	0.162	0.616
(0, 1, -1)'	0.4	0.1	0.049	0.174	0.179	0.000	0.000	0.140	0.011
		0.3	0.646	0.558	0.811	0.001	0.000	0.894	0.543
		0.5	0.961	0.536	0.975	0.006	0.005	0.999	0.991
		0.7	0.997	0.806	0.991	0.148	0.137	0.993	1.000
		0.9	1.000	0.979	0.994	0.654	0.697	0.315	0.954
	0.8	0.1	0.015	0.071	0.067	0.000	0.000	0.147	0.075
		0.3	0.126	0.284	0.293	0.000	0.000	0.493	0.386
		0.5	0.403	0.475	0.523	0.004	0.004	0.731	0.748
		0.7	0.703	0.514	0.572	0.027	0.022	0.608	0.880
		0.9	0.896	0.417	0.562	0.140	0.155	0.157	0.614

Table 5.20 : Powers of all the tests at the 5% level for *X16*, selected β, σ, and μ values with $n = 20$, using near exact non-similar critical values

β	σ	μ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 0)'	0.4	0.1	0.035	0.122	0.121	0.000	0.000	0.139	0.019
		0.3	0.411	0.497	0.593	0.000	0.000	0.796	0.429
		0.5	0.849	0.631	0.746	0.005	0.005	0.992	0.961
		0.7	0.966	0.697	0.721	0.306	0.274	0.980	0.993
		0.9	1.000	0.815	0.943	0.995	0.994	0.032	0.062
	0.8	0.1	0.012	0.054	0.052	0.000	0.000	0.119	0.066
		0.3	0.069	0.199	0.189	0.000	0.000	0.354	0.289
		0.5	0.247	0.262	0.290	0.001	0.001	0.563	0.628
		0.7	0.527	0.174	0.234	0.026	0.021	0.500	0.811
		0.9	0.843	0.123	0.329	0.584	0.569	0.058	0.275
(0, -1, 0)'	0.4	0.1	0.030	0.122	0.119	0.000	0.000	0.151	0.016
		0.3	0.402	0.480	0.575	0.000	0.000	0.803	0.427
		0.5	0.865	0.633	0.741	0.009	0.007	0.993	0.961
		0.7	0.964	0.700	0.726	0.294	0.267	0.982	0.993
		0.9	0.999	0.820	0.941	0.997	0.996	0.039	0.075
	0.8	0.1	0.008	0.062	0.053	0.000	0.000	0.128	0.071
		0.3	0.067	0.173	0.185	0.001	0.001	0.356	0.293
		0.5	0.257	0.252	0.282	0.002	0.002	0.565	0.632
		0.7	0.525	0.174	0.250	0.028	0.024	0.506	0.814
		0.9	0.843	0.129	0.341	0.592	0.579	0.069	0.280
(0, 0, 1)'	0.4	0.1	0.024	0.097	0.095	0.000	0.000	0.177	0.061
		0.3	0.282	0.481	0.562	0.000	0.000	0.698	0.494
		0.5	0.770	0.667	0.878	0.003	0.002	0.942	0.920
		0.7	0.974	0.807	0.899	0.160	0.147	0.887	0.995
		0.9	0.999	0.876	0.968	0.993	0.992	0.016	0.217
	0.8	0.1	0.010	0.058	0.050	0.000	0.000	0.126	0.115
		0.3	0.050	0.186	0.189	0.000	0.000	0.265	0.299
		0.5	0.199	0.329	0.366	0.003	0.001	0.364	0.544
		0.7	0.506	0.324	0.400	0.028	0.025	0.298	0.686
		0.9	0.861	0.166	0.380	0.625	0.616	0.045	0.307
(0, 0, -1)'	0.4	0.1	0.024	0.103	0.107	0.000	0.000	0.177	0.063
		0.3	0.291	0.482	0.559	0.000	0.000	0.702	0.484
		0.5	0.779	0.664	0.861	0.006	0.004	0.949	0.926
		0.7	0.975	0.786	0.878	0.153	0.145	0.896	0.996
		0.9	0.999	0.866	0.964	0.996	0.996	0.018	0.203
	0.8	0.1	0.011	0.062	0.047	0.000	0.000	0.122	0.113
		0.3	0.054	0.187	0.193	0.001	0.001	0.250	0.303
		0.5	0.201	0.305	0.349	0.005	0.004	0.355	0.538
		0.7	0.498	0.317	0.389	0.032	0.027	0.291	0.686
		0.9	0.859	0.168	0.365	0.638	0.617	0.045	0.297

Table 5.21 : Powers of all the tests at the 5% level for *X16*, selected β, σ, and μ values with $n = 40$, using near exact non-similar critical values

β	σ	μ	LR	LM(E)	LM(L)	W(E)	W(L)	g(0.3)	g(0.5)
(0, 1, 1)'	0.4	0.1	0.107	0.367	0.519	0.000	0.000	0.067	0.000
		0.3	0.992	0.705	0.999	0.000	0.000	0.999	0.635
		0.5	1.000	0.806	1.000	0.019	0.003	1.000	1.000
		0.7	1.000	1.000	1.000	0.949	0.631	1.000	1.000
		0.9	1.000	1.000	1.000	1.000	1.000	0.423	0.990
	0.7	0.1	0.019	0.114	0.192	0.000	0.000	0.146	0.008
		0.3	0.596	0.766	0.865	0.000	0.000	0.926	0.525
		0.5	0.979	0.930	0.981	0.000	0.000	1.000	0.995
		0.7	0.999	0.973	0.993	0.206	0.027	1.000	1.000
		0.9	1.000	0.998	1.000	1.000	1.000	0.176	0.733
(0, -1, 1)'	0.4	0.1	0.057	0.241	0.365	0.000	0.000	0.112	0.001
		0.3	0.935	0.872	0.984	0.000	0.000	0.990	0.627
		0.5	1.000	0.898	1.000	0.000	0.000	1.000	1.000
		0.7	1.000	0.999	1.000	0.227	0.023	1.000	1.000
		0.9	1.000	1.000	1.000	0.994	0.903	0.920	1.000
	0.7	0.1	0.011	0.078	0.148	0.000	0.000	0.181	0.023
		0.3	0.332	0.551	0.688	0.000	0.000	0.849	0.489
		0.5	0.858	0.752	0.852	0.000	0.000	0.997	0.977
		0.7	0.988	0.739	0.870	0.005	0.000	0.999	1.000
		0.9	1.000	0.819	0.949	0.517	0.170	0.680	0.965
(0, 1, 0)'	0.4	0.1	0.033	0.176	0.275	0.000	0.000	0.114	0.003
		0.3	0.826	0.884	0.968	0.000	0.000	0.972	0.596
		0.5	0.997	0.936	0.998	0.000	0.000	1.000	0.999
		0.7	1.000	0.998	0.999	0.167	0.018	1.000	1.000
		0.9	1.000	1.000	1.000	1.000	0.999	0.024	0.640
	0.7	0.1	0.007	0.059	0.111	0.000	0.000	0.161	0.032
		0.3	0.210	0.452	0.606	0.000	0.000	0.771	0.470
		0.5	0.729	0.706	0.815	0.000	0.000	0.982	0.943
		0.7	0.956	0.666	0.833	0.008	0.001	0.982	0.996
		0.9	0.997	0.689	0.895	0.899	0.700	0.082	0.501
(0, 0, -1)'	0.4	0.1	0.018	0.105	0.185	0.000	0.000	0.141	0.010
		0.3	0.594	0.734	0.847	0.000	0.000	0.930	0.501
		0.5	0.981	0.909	0.960	0.000	0.000	1.000	0.994
		0.7	0.999	0.970	0.993	0.254	0.039	1.000	1.000
		0.9	1.000	1.000	1.000	1.000	1.000	0.124	0.971
	0.7	0.1	0.005	0.048	0.084	0.000	0.000	0.154	0.046
		0.3	0.114	0.289	0.420	0.000	0.000	0.664	0.394
		0.5	0.555	0.416	0.591	0.000	0.000	0.953	0.894
		0.7	0.904	0.384	0.610	0.011	0.000	0.968	0.993
		0.9	0.993	0.592	0.839	0.972	0.851	0.084	0.683

Figure 5.1: Comparison of powers for *X15*, $\beta = (0,1,1)'$, $\sigma = 0.025$ and $n = 20$, using near exact non-similar critical values

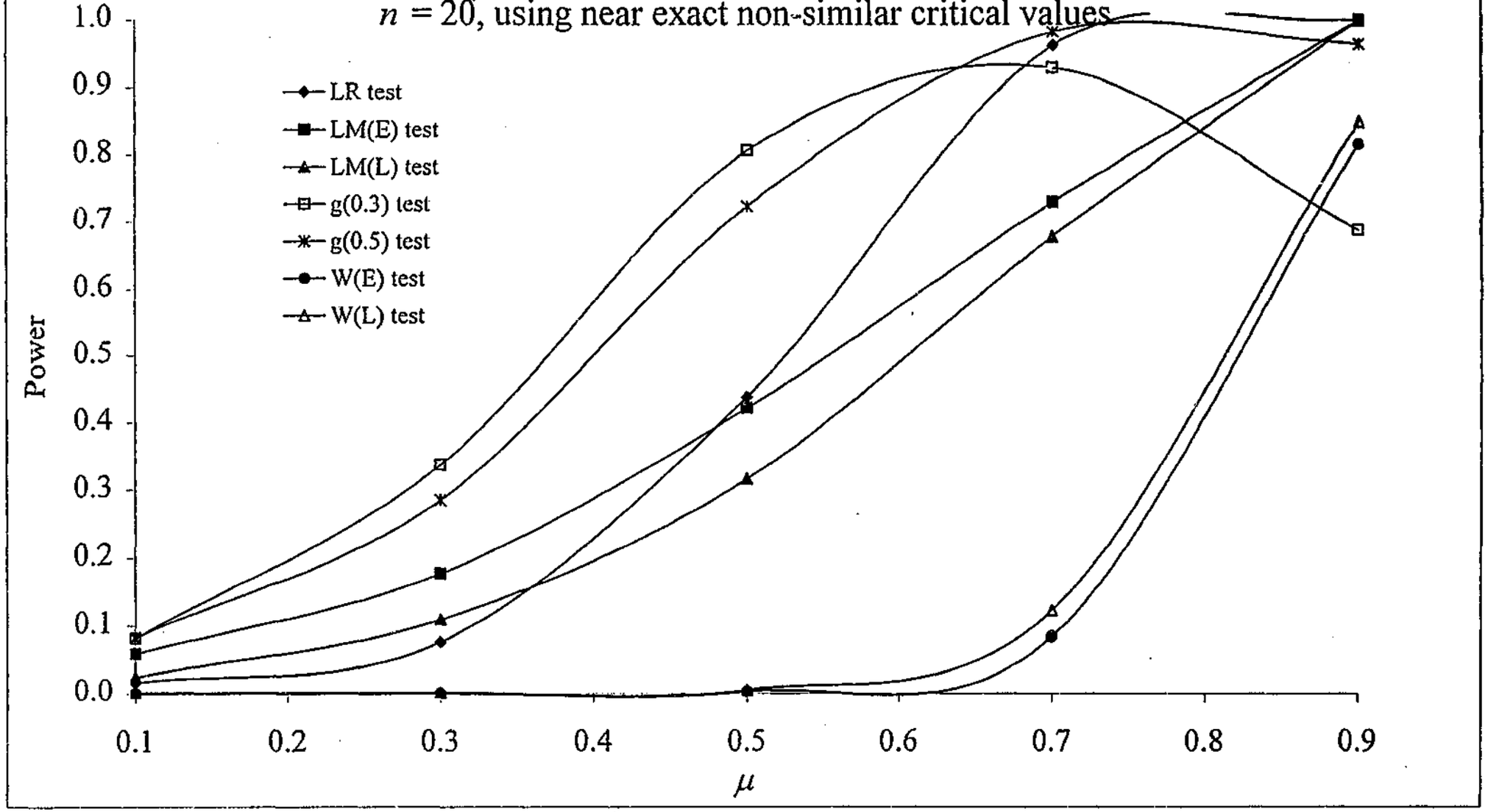


Figure 5.2: Comparison of powers for $X15$, $\beta = (0,1,1)'$, $\sigma = 0.025$ and $n = 40$, using near exact non-similar critical values

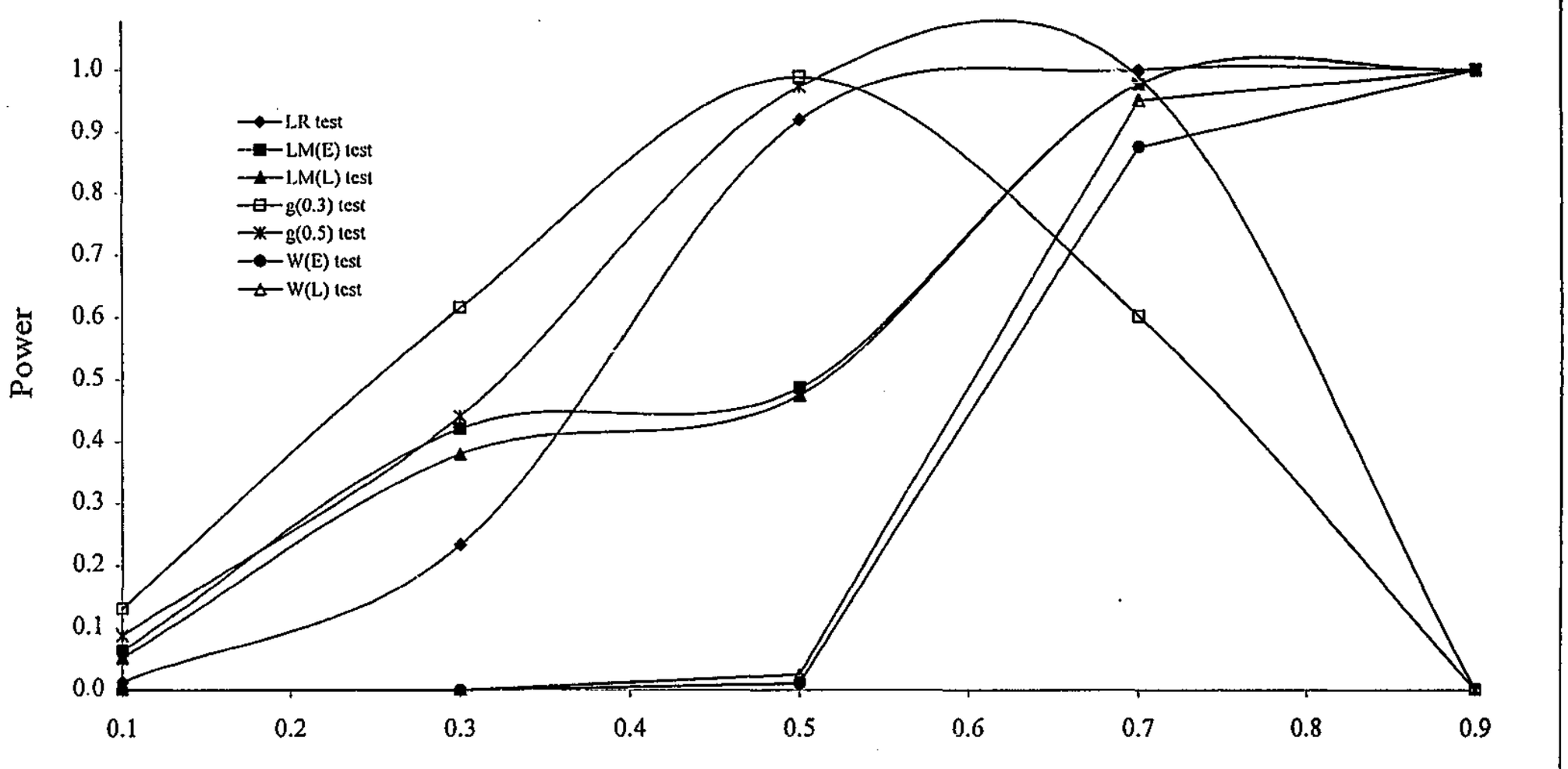


Figure 5.3: Comparison of powers for *X16*, $\beta = (0, -1, 1)'$, $\sigma = 0.8$ and $n = 20$, using near exact non-similar critical values

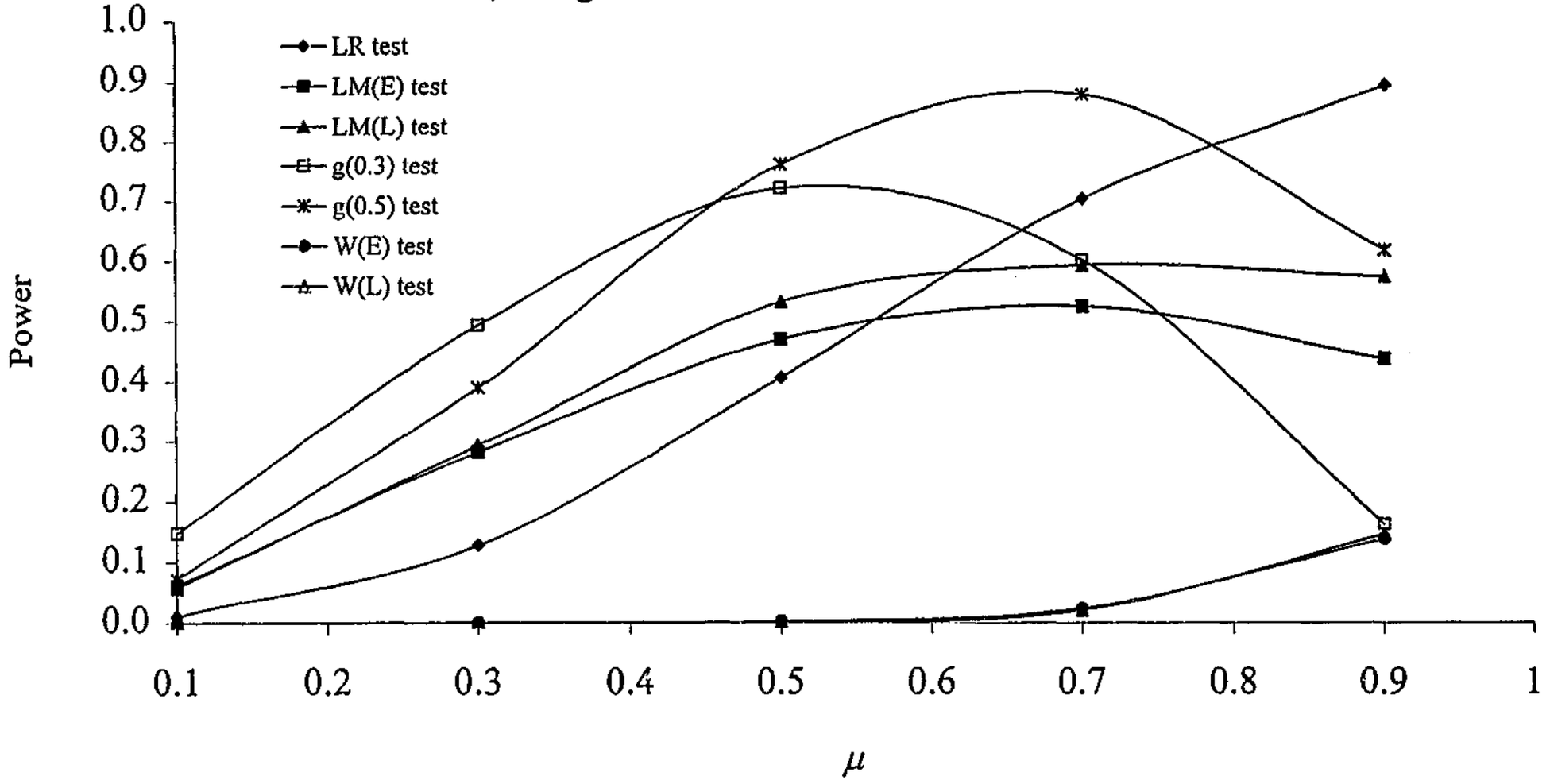


Figure 5.4: Comparison of powers for $X16$, $\beta = (0,-1,0)'$, $\sigma = 0.8$ and $n = 20$, using near exact non-similar critical values

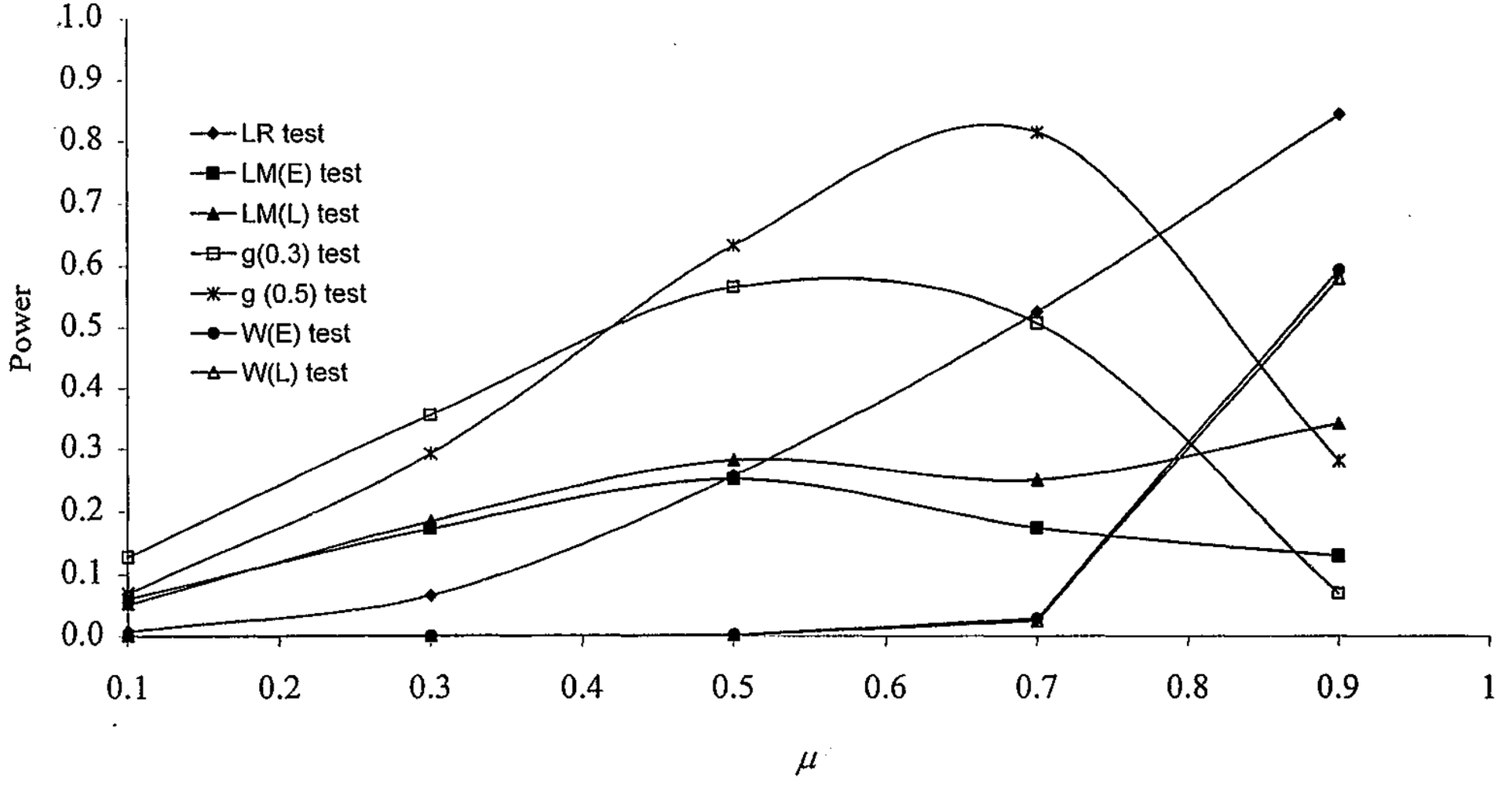
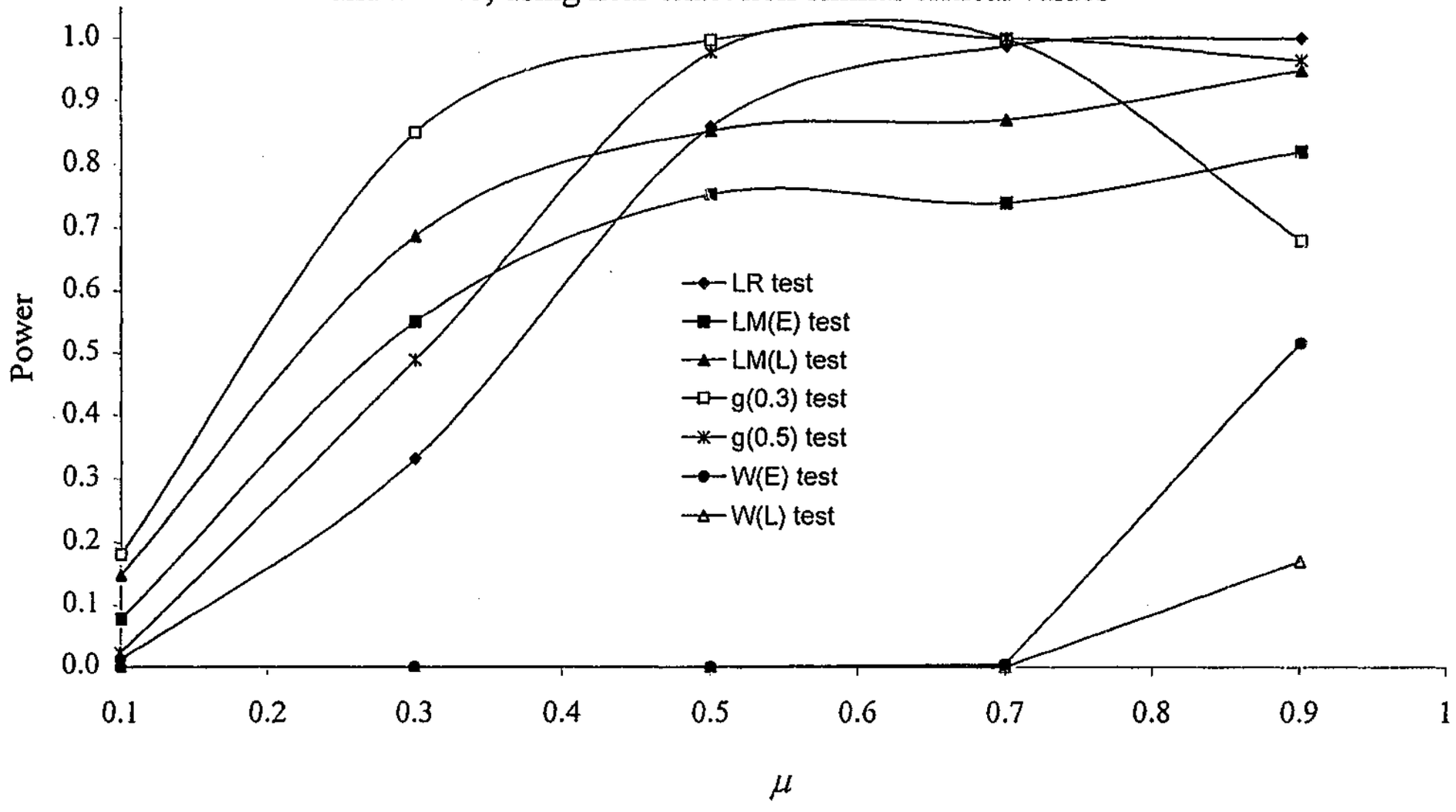


Figure 5.5: Comparison of powers for *X16*, $\beta = (0,-1,1)'$, $\sigma = 0.7$ and $n = 40$, using near exact non-similar critical values



APPENDIX 5.2

Terms Involved in Marginal Likelihood Based One-Sided LR, LM and W Tests

Following Grose (1998), we define the following terms, in which Ω and Z stand for $\Sigma_1(\mu,\rho)$ and $X(\mu)$, respectively. Similarly C, Γ, D and Σ stand for $C(\mu,\rho)$, $\Gamma(\mu)$, $D(\mu,\rho)$, and $\Sigma(\rho)$, respectively.

$$I_{\rho\rho} = \frac{1}{2(l+2)}\{l\,\mathrm{tr}(\Omega_\rho \tilde{P}_{Z,\Omega})^2 - [\mathrm{tr}(\Omega_\rho \tilde{P}_{Z,\Omega})]^2\}, \tag{A5.1}$$

$$I_{\mu\rho} = \mathrm{tr}[Z_\mu \tilde{B}_{Z,\Omega}\Omega_\rho \tilde{P}_{Z,\Omega}] + \frac{1}{2(l+2)}\{l\,\mathrm{tr}(\Omega_\mu \tilde{P}_{Z,\Omega}\Omega_\rho \tilde{P}_{Z,\Omega}) - \mathrm{tr}(\Omega_\mu \tilde{P}_{Z,\Omega})\mathrm{tr}(\Omega_\rho \tilde{P}_{Z,\Omega})\}, \tag{A5.2}$$

and

$$I^L_{\mu\mu} = q + \frac{l^2 - l + 2}{l(l-2)}\{\mathrm{tr}((Z'\Omega^{-1}Z)^{-1}Z'_\mu \tilde{P}_{Z,\Omega}Z_\mu) + \beta' Z'_\mu \tilde{P}_{Z,\Omega}Z_\mu \beta/\sigma^2\}$$

$$+ \frac{\{\mathrm{tr}((\overline{P}_Z)_\mu \Omega)\}^2 + 2\mathrm{tr}((\overline{P}_Z)_\mu \Omega)^2 + 4\beta' Z'_\mu \overline{P}_Z \Omega \overline{P}_Z Z_\mu \beta/\sigma^2}{2\{\mathrm{tr}(\overline{P}_Z\Omega)\}^2}$$

$$- \frac{\mathrm{tr}((\overline{P}_Z)_{\mu\mu}\Omega) + \beta' Z'_\mu \overline{P}_Z Z_\mu \beta/\sigma^2}{\mathrm{tr}(\overline{P}_Z\Omega)}, \tag{A5.3}$$

$$\tilde{I}_{\mu\mu} = q - 2\frac{\hat{e}'\{Z_\mu \bar{B}_Z Z_\mu - 0.5 Z_{\mu\mu}\}\hat{\beta}}{\hat{e}'\hat{e}} - \frac{\hat{\beta}' Z_\mu' \bar{P}_Z Z_\mu \hat{\beta}}{\hat{e}'\hat{e}} + \frac{\hat{e}' Z_\mu (Z'Z)^{-1} Z_\mu' \hat{e}}{\hat{e}'\hat{e}}$$

$$+ 2\left\{\frac{\hat{e}' Z_\mu \hat{\beta}}{\hat{e}'\hat{e}}\right\}^2 + \frac{\tilde{\beta}' Z_\mu' \tilde{P}_{Z,\Omega} Z_\mu \tilde{\beta}}{\tilde{\sigma}^2} - \frac{2}{l}\left\{\frac{\tilde{e}'\Omega^{-1} Z_\mu \tilde{\beta}}{\tilde{\sigma}^2}\right\}^2, \qquad \text{(A5.4)}$$

where $I_{\mu\mu}^L$ and $\tilde{I}_{\mu\mu}$ are the Laplace approximated information and estimated information, respectively, $\bar{B}_Z = (Z'Z)^{-1} Z'$, $\tilde{B}_{Z,\Omega} = (Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1}$ and

$$q = -\mathrm{tr}\{(\bar{B}_Z - \tilde{B}_{Z,\Omega}) Z_{\mu\mu}\} - \mathrm{tr}\{(Z'Z)^{-1} Z_\mu' \bar{P}_Z Z_\mu\} + \mathrm{tr}\{(\bar{B}_Z Z_\mu \bar{B}_Z - \tilde{B}_{Z,\Omega} Z_\mu \tilde{B}_{Z,\Omega}) Z_\mu\}$$

$$+ 2\,\mathrm{tr}(Z_\mu \tilde{B}_{Z,\Omega} \Omega_\mu \tilde{P}_{Z,\Omega}) + \frac{1}{2(l+2)}\left\{l\,\mathrm{tr}(\Omega_\mu \tilde{P}_{Z,\Omega})^2 - [\mathrm{tr}(\Omega_\mu \tilde{P}_{Z,\Omega})]^2\right\}. \qquad \text{(A5.5)}$$

The first derivatives of Ω with respect to ρ and μ are given below, in which, $R_1(\delta)$ stands for the diagonal matrix with the scalar δ in the top-left corner, zeros otherwise.

$$\Omega_\rho = C_\rho \Sigma C' + C\Sigma_\rho C' + C\Sigma C_\rho' \qquad \text{(A5.6)}$$

in which, $C = \Gamma^{-1} D$, $C_\rho = -C C_\rho^{-1} C$, $C_\rho^{-1} = R_1(\xi_\rho)$, $\xi_\rho = -d_1^{-1}\left\{\frac{\mu}{1-\mu^2\rho^2}\right\}$,

$d_1 = \sqrt{\frac{1+\mu\rho}{(1-\mu\rho)(1-\mu^2)}}$, $\Sigma_\rho = -\Sigma\Sigma_\rho^{-1}\Sigma$, $\Sigma_\rho^{-1} = 2\rho(I_n - S_1) - (L_1 + L_1')$, S_1 is a $n \times n$ diagonal matrix with diagonal elements (1,0,...,0,1), L_1 is the $n \times n$ matrix with ones on the first lower off-diagonal, zeros otherwise,

$$\Omega_\mu = -\{C C_\mu^{-1} \Omega + \Omega C_\mu'^{-1} C'\}, \qquad \text{(A5.7)}$$

in which, $C_\mu^{-1} = R_1(\xi_\mu) - L_1$, $\xi_\mu = -d_1^{-1}\left(\frac{\rho}{1-\mu^2\rho^2} + \frac{\mu}{1-\mu^2}\right)$.

The first and second derivatives of Z with respect to μ are as follows.

$$Z_\mu = \Psi(\mu)Z, \tag{A5.8}$$

and

$$Z_{\mu\mu} = 2\Psi(\mu)^2 Z. \tag{A5.9}$$

where $\Psi(\mu) = \Gamma^{-1}(R_1 + L_1)$.

By letting $P_Z = Z\overline{B}_Z$, the first and second derivatives of $\overline{P}_Z$ with respect to μ can be written as

$$(\overline{P}_Z)_\mu = -\{\overline{P}_Z\Psi(\mu)P_Z + P_Z\Psi(\mu)'\overline{P}_Z\}, \tag{A5.10}$$

$$(\overline{P}_Z)_{\mu\mu} = 2\{\overline{P}_Z\overline{Q}_1P_Z + P_Z\overline{Q}_1'\overline{P}_Z + P_Z\Psi(\mu)'\overline{P}_Z\Psi(\mu)P_Z - \overline{P}_Z\Psi(\mu)P_Z\Psi(\mu)'\overline{P}_Z\}, \tag{A5.11}$$

where, $\overline{Q}_1 = -\Psi(\mu)\overline{P}_Z\Psi(\mu)$.

The score function with respect to μ can be written as

$$q_\mu = \mathrm{tr}\{(\overline{B}_z - \tilde{B}_{z,\Omega})Z_\mu\} - \frac{1}{2}\mathrm{tr}\{\tilde{P}_{Z,\Omega}\Omega_\mu\} + \frac{1}{2}Q_\mu, \tag{A5.12}$$

where, $Q_\mu = -2\dfrac{\hat{e}'Z_\mu\tilde{\beta}}{\hat{e}'\hat{e}} + 2l\dfrac{\tilde{e}'\Omega^{-1}Z_\mu\tilde{\beta}}{\tilde{e}'\Omega^{-1}\tilde{e}} - l\dfrac{\tilde{e}'\Omega_\mu^{-1}\tilde{e}}{\tilde{e}'\Omega^{-1}\tilde{e}}$ and $\hat{e} = (I - Z\overline{B}_Z)y$ and $\tilde{e} = (I - Z\tilde{B}_{Z,\Omega})y$ are the OLS and GLS residual vectors, respectively.

CHAPTER 6

SUMMARY AND CONCLUSIONS

In this thesis, we have proposed a new APO test, called the g test, for testing composite null hypotheses and a new approach to obtain exact (and near exact) non-similar critical values of general non-similar tests. The g test is based on the generalised Neyman-Pearson lemma (GNPL) and it involves the finding of multiple critical values. We have also outlined two methods for finding these critical values. The new approach to obtain exact (and near exact) non-similar critical values involves controlling the maximum size of a non-similar test over the nuisance parameter space, thus it also allows one to assess whether an approximate distribution is a good approximation to the exact distribution of the test statistic under the null hypothesis. In Chapter 1, we discussed the motivations for constructing the g test and for a new approach to finding exact non-similar critical values. In this chapter, we briefly summarise the main observations and conclusions which can be drawn from this thesis. These conclusions include a recommendation regarding the best test procedure

for the testing problems considered. Some issues which require further research are also discussed.

A review of hypothesis testing was presented in Chapter 2. This chapter mainly focuses on studies involving King's (1987b) PO tests in the context of composite hypothesis testing. Because all the applications considered in this thesis are non-nested, some popular non-nested tests were also briefly discussed. In addition, other studies that are relevant to this thesis, such as, studies based on non-similar critical values and those based on SA were also discussed. This chapter reveals the importance of developing tests which have excellent finite-sample properties, such as, PO tests, rather than relying on large-sample based tests. PO tests cannot always be constructed when testing a composite null. For situations where his PO test cannot be constructed, King suggested an APO test. APO tests seem to perform well when they are nearly optimal, otherwise there is a question mark about their reliability. For example, in the context of composite nested testing, Rahman and King (1994) preferred some marginal likelihood based asymptotic tests over the APO tests (which were not nearly optimal). Similarly, in the context of composite non-nested testing, Silvapulle (1994a) preferred some asymptotic tests over an APO test (which was not nearly optimal). Studies like these indicate the need for a reliable APO test for testing a composite null.

Also in Chapter 2, we observed that approximate non-similar critical values can be difficult to obtain when the number of nuisance parameters is large. Because researchers conducting simulation studies always used a limited number of nuisance parameter values, the reliability of approximate non-similar critical values over the nuisance parameter space is largely unknown. Studies on SA confirm that the SA algorithm is much more robust than conventional algorithms in finding the global maximum and less likely to fail on difficult functions. Also, SA has several advantages compared to the conventional algorithms. It can be used to optimize functions with multiple optima and functions that are not defined for some parameter values, it assumes very little about the shape of the function to be optimized and it is largely independent of the starting values. In addition, Goffe et al. (1994) recommend that SA to be used as a diagnostic tool to understand how conventional algorithms fail. The only undesirable aspect of the SA algorithm is that it takes greater run time.

Because it is much more reliable than the conventional algorithms this problem has to be stomached.

In Chapter 3, we discussed the theory behind the g test and how it can be applied to testing composite non-nested disturbance covariance matrices in the linear regression model. The g test was applied to the problems of testing for MA(1) errors against AR(1) errors and AR(1) errors against IMA(1,1) errors with a negative MA coefficient in the linear regression model. For the first problem, we considered four versions of the g test, namely, the g(0.3), g(0.5), g(0.75) and g(-0.5) tests and for the second problem, we considered the g(-0.5) test only. Silvapulle (1991, 1994a) investigated the testing of these two problems using King's APO tests and some asymptotic tests. Based on her results, she recommended APO tests for the former problem and some asymptotic tests for the latter. We compared our g test results with those of Silvapulle. Also, the use of the g test has been illustrated by its application to two real world data sets.

For both testing problems considered, the g tests performed better than the APO tests. For the first problem, the performance of the g test was largely unaffected by the choice of the point under the alternative where the power is maximised. For this problem, the APO tests (which were nearly optimal) also performed equally well, therefore, the power advantage of the g tests over the APO tests was small in magnitude. But for the second problem, the g(-0.5) test always has a significant power advantage over its competitors. In particular, the g(-0.5) test always outperformed Silvapulle's recommended asymptotic tests in terms of size and power properties. On many occasions, the g(-0.5) test is observed to have exact sizes and higher powers than its competitors. Therefore, based on our results, we strongly recommend the g(-0.5) test for the second problem. The power results for the first problem indicate that when King's APO tests are nearly optimal, little gain can be achieved by applying the g test.

In Chapter 4, we explained how SA can effectively be used to obtain exact size critical values (by assuming knowledge of the unknown parameters) and exact (and near exact) non-similar critical values of general non-similar tests. In this chapter, we investigated the performance of near exact non-similar critical values. For this study,

we used two non-similar tests, namely, the DW test and Durbin's t test in the context of the dynamic linear regression model. In order to obtain the near exact non-similar critical values of these tests, we used approximate small disturbance asymptotic (ASDA) and large-sample based critical values, respectively, and SA. In this chapter, we calculated the sizes of the tests for a variety of nuisance parameter values and design matrices, in order to check whether the SA based near exact non-similar critical values are indeed working well in terms of controlling the sizes over the nuisance parameter space. We also compared near exact non-similar critical values with approximate non-similar critical values used in previous studies.

Our study clearly showed that neither the large-sample based t test nor the ASDA distribution based DW test is best under the null because both tests can have approximately the same (higher than nominal) maximum sizes over the nuisance parameter space. For the nuisance parameter values considered, the SA based near exact non-similar critical values controlled the sizes remarkably well, indicating that these critical values are reliable over the nuisance parameter space. Also, the use of near exact non-similar critical values often produced sizes that are insignificantly different from the nominal size. Based on our size results, we conjecture that approximate non-similar critical values are not ideal for Durbin's t test because they may make the test's sizes unnecessarily low, and hence reduce the powers. Also, for the tests considered, it takes about the same time to calculate near exact non-similar critical values and approximate non-similar critical values. In addition, (as mentioned earlier) the SA based approach has an advantage, it can be used to assess any approximate distributions under the null. Therefore, we recommend SA based near exact non-similar critical values over approximate non-similar critical values. Our study suggests that if SA achieves the global (instead of local) maximum size then near exact non-similar critical values can be expected to work well in terms of controlling the sizes. Therefore, the SA parameters need to be chosen carefully, otherwise SA might find the local maximum size.

In Chapter 5, we explored the problem of testing for a static linear model with AR(1) errors against a dynamic linear model with white noise errors. Here, we compared marginal likelihood based g tests with marginal likelihood based one-sided LR, LM and W tests. Both Laplace approximated information based and estimated

information based W (and LM) tests were considered in order to see which test is best in finite samples. Two versions of the g test, namely the g(0.3) and g(0.5) tests were considered. The size and power comparisons of this chapter were based on near exact non-similar critical values obtained via SA. Here, the size and power calculations were made for a variety of nuisance parameter values in order to determine the best test over the nuisance parameter space.

When the exogenous coefficient vector, β, is close to zero and the dynamic coefficient, μ, and the autocorrelation coefficient, ρ, are similar in magnitude, it may be impossible to distinguish the null and alternative models of interest. In addition, if exogenous regressors are lag invariant, it is virtually impossible to distinguish between the models. We regard these testing situations as the toughest testing situations. On the other hand, distinguishing between the null and alternative models becomes easier when μ approaches one.

Our study clearly showed that the large-sample based LR, W(L) and W(E) tests are not reliable under the null even when the sample size is 60. In particular, these tests can have maximum sizes close to one, over the nuisance parameter space. Of all the tests considered, the LM(L) test is found to be the best in this regard and the LM(E) test is the second best. For the nuisance parameter values considered, near exact non-similar critical values generally controlled the sizes remarkably well, but sometimes produced sizes that are approximately twice the nominal size. This indicates that these critical values are generally reliable over the nuisance parameter space. For the g(0.3), g(0.5), LM(L) and LM(E) tests, the use of near exact non-similar critical values often produced sizes that are not significantly different from the nominal size, while for the LR, W(L) and W(E) tests, they produced sizes well below the nominal size. Obtaining near exact non-similar critical values of the asymptotic tests (via SA) can be quite demanding in terms of time, because the tests involve repeated maximum marginal likelihood estimation. On the other hand, it takes much less time for the g tests.

Power results using near exact non-similar critical values led to the following conclusions. The g(0.3) and g(0.5) tests generally outperformed the classical tests

when $\mu < 0.9$. In many situations, the g tests possess powers that are more than double those of their competitors. Generally, the g(0.3) test is the least powerful test at $\mu = 0.9$. The g(0.5) test's powers also sometimes approach zero when $\mu = 0.1$ and $\mu = 0.9$. These observations suggest that the g tests are always reliable in the neighbourhood of the point where the power is maximised, but not always reliable some distance from that point. Therefore, it seems that the g(0.75) test might be suitable for $\mu > 0.7$. However, this test might lack power for small values of μ.

Among the asymptotic tests, the LM(L) test was found to be the best for $\mu < 0.7$, though its powers are lower than those for the g tests. The LR test outperforms the g tests only when μ is closer to one. Both the W(L) and W(E) tests were found to be the worst performers always. As expected, when the sample size increases, the asymptotic tests' powers improve, but they are still lower than those for the g tests, particularly for $\mu \leq 0.5$.

The only tests that generally overcome the toughest testing situations (discussed above) are the g tests, particularly the g(0.3) test. The asymptotic tests seem trustworthy only when μ approaches one (which is a relatively easier testing situation as noted above). Therefore, compared to the asymptotic tests the g tests should be commended on their performance. Also, the overall performance of the g tests seems largely unaffected with respect to the nuisance parameter values. For reasons discussed in Chapter 5, marginal likelihood based g tests are not perfect for the problem of interest, still they perform better than the asymptotic tests for $\mu < 0.9$. Similarly, the LM tests are also not expected to have better finite-sample properties, yet, they are found to be the best asymptotic tests for small values of μ. Based on our results, we recommend the g(0.3) test for $\mu < 0.5$, the g(0.5) test for $0.5 \leq \mu \leq 0.7$ and the LR test for $\mu > 0.7$.

The overall recommendation of this thesis is that, if a PO test is suitable for a particular testing problem it should be used. If the null hypothesis is composite and the PO test cannot be constructed then the g test should be the next preferred candidate. However, our study suggests that when King's APO test is nearly optimal,

there is little gain in applying the g test in terms of power properties. According to our study, the g test is always reliable when the nuisance parameters can be avoided via the invariance method. When these nuisance parameters cannot be avoided, the g test is always reliable in the neighbourhood of the point where the power is maximised and not necessarily always reliable some distance from that point. In the context of the dynamic linear model, the large-sample based LR, W(L) and W(E) tests seem unreliable under the null even when the sample size is 60. Of all the tests, the LM(L) test seems always reliable in this regard. The LM(L) test seems better than the LM(E) test in finite samples, whereas, this is not clear in the case of the W tests. Our study also supports SA based near exact non-similar critical values over approximate non-similar critical values. We observed that approximate non-similar critical values are not useful for the g test, whereas, near exact non-similar critical values are useful. This indicates that the SA based non-similar critical values may be helpful for situations where approximate non-similar critical values fail to control the sizes.

Finally, there are a few areas in which further research could be undertaken. In this thesis, we applied the g test to three composite non-nested testing situations, with encouraging results. How well the g test works for composite nested testing situations is clearly an interesting area which requires further research. By construction the g test has sizes not significantly different from the nominal size, therefore, for composite nested testing situations, it is possible that its powers at some points under the alternative hypothesis will be not significantly different from the nominal size.

The only problems we have investigated in this thesis are those were the parameter under test is reduced to an interval (after using invariance method or due to economic beliefs). The g tests performed well for this case. The question is, can this test be applied for testing problems involving null hypotheses defined over an open interval, such as, $(-\infty, 0)$, $(0, \infty)$ or $(-\infty, \infty)$? In order to apply the g test for this case, we might have to use a function that can map an infinite interval to a finite interval. For this, a logistic function might be a useful way to map an infinite interval to a finite interval $[0, 1]$. How well the g test performs for this type of testing problems is an interesting question which requires further research.

All the testing problems we have investigated in this thesis are one dimensional in nature and the g test performed well. It would be interesting to see how well this approach extends to two and higher dimensional cases. For this, the general ideas developed in Chapter 3 and Chapter 5 could be extended to cover (1) testing higher order MA against higher order AR processes, (2) testing higher order non-stationary error processes involving IMA(1, q), $q > 1$, (3) testing higher order non-stationary against higher order stationary disturbances and (4) testing higher order non-nested processes in the dynamic linear regression model. Similarly it would be interesting to apply the g test to higher order composite nested testing problems as well. Finally robustness of the g test to non-normal errors can also be explored.

Our study clearly showed that SA is very useful for optimizing difficult functions, like a size function, which is a step function when obtained via simulation. This indicates that SA should be used for difficult econometric functions. Also, for the problems considered, SA effectively found near exact non-similar critical values and exact size critical values. How well this result translates to more complicated problems is clearly an another interesting area for further research.

REFERENCES

Aarts, E.H.L. and P.J.M. Laarhoven (1989), Simulated annealing: An introduction, *Statistica Neerlandica*, 43, 31-52.

Ali, M., A. Törn and S. Viitanen (1997), A direct search simulated annealing algorithm for optimization involving continuous variables, Turku Centre for Computer Science TUCS Technical Report No 97.

Andrews, D.W.K. and W. Ploberger (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

Ara, I. (1995), Marginal likelihood based tests of regression disturbances, Unpublished Ph.D. thesis, Monash University.

Ara, I. and M.L. King (1993), Marginal likelihood based tests of regression disturbances, Paper presented at the 1993 Australasian Meeting of the Econometric Society, Sydney.

Arthanari, T.S. and Y. Dodge (1981), *Mathematical Programming in Statistics*, John Wiley & Sons, New York.

Atkinson, A.C. (1970), A method for discriminating between models, *Journal of the Royal Statistical Society*, B, 32, 323-353.

Baltagi, B.H. and Q. Li (1995), Testing AR(1) against MA(1) disturbances in an error component model, *Journal of Econometrics*, 68, 133-151.

Beggs, J.J. (1988), Diagnostic testing in applied econometrics, *Economic Record*, 64, 81-101.

Bellhouse, D.R. (1978), Marginal likelihood methods for distributed lag models, *Statistische Hefte*, 19, 2-14.

Berenblut, I.I. and G.I. Webb (1973), A new test for autocorrelated errors in the linear regression model, *Journal of the Royal Statistical Society*, B, 35, 33-50.

Bhatti, M.I. (1992), Testing regression models based on sample survey data, Unpublished Ph.D. thesis, Monash University.

Bhatti, M.I. and M.L. King (1994), Testing for subblock effects in multi-stage linear regression models, Paper presented at the European Meeting of the Econometric Society, Maastricht, Netherlands.

Berens, H.J. (1988), Towards a theory of point optimal testing: Comment, *Econometric Reviews*, 6, 231-233.

Brooks, R.D. (1993), Alternative point optimal tests for regression coefficient stability, *Journal of Econometrics*, 57, 365-376.

Brooks, R.D. (1995), The robustness of point optimal testing for Rosenberg random regression coefficients, *Econometric Reviews*, 14, 35-53.

Brooks, R.D. and M.L. King (1994), Testing Hildreth-Houck against return to normalcy random regression coefficients, *Journal of Quantitative Economics*, 10, 33-52.

Brooks, S.P. and B.J.T. Morgan (1995), Optimization using simulated annealing, *The Statistician*, 44, 241-257.

Burke, S.P., L.G. Godfrey and A.R. Tremayne (1990), Testing AR(1) against MA(1) disturbances in the linear regression model: An alternative approach, *Review of Economic Studies*, 57, 135-145.

Cerny, V. (1985), Thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm, *Journal of Optimization Theory and Applications*, 45, 41-51.

Chatfield, C. (1996), *The Analysis of Time Series: An Introduction*, Chapman & Hall, London.

Chow, G.C. (1957), *Demand for Automobiles in the United States*, North-Holland, Amsterdam.

Corana, A., M. Marchesi, C. Martini and S. Ridella (1987), Minimising mutimodal functions of continuous variables with the simulated annealing algorithm, *ACM Transactions on Mathematical Software*, 13, 262-280.

Cox, D.R. (1961), Tests of separate families of hypotheses, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 105-123.

Cox, D.R. (1962), Further results on tests of separate families of hypotheses, *Journal of the Royal Statistical Society*, B, 24, 406-424.

Cox, D.R. and D.V. Hinkley (1974), *Theoretical Statistics*, John Wiley & Sons, New York.

Dastoor, N.K. and G. Fisher (1988), The theory and practice of point optimal testing, *Econometric Reviews*, 6, 219-229.

Davidson, R. and J.G. MacKinnon (1981), Several tests for model specification in the presence of alternative hypotheses, *Econometrica*, 49, 781-793.

Davidson, R. and J.G. MacKinnon (1982), Some non-nested hypotheses tests and the relations among them, *Review of Economic Studies*, 49, 551-565.

Davidson, R. and J.G. MacKinnon (1996), The size distortion of bootstrap tests, Discussion paper 936, Queen's University.

Davies, R.B. (1980), Algorithm AS155: The distribution of a linear combination of χ^2 random variables, *Applied Statistics*, 29, 323-333.

Deaton, A.S. (1982), Model selection procedures, or, Does the consumption function exist?, in G.C. Chow and P. Corsi (eds.), *Evaluating the Reliability of Macroeconomic Models*, John Wiley & Sons, New York, 43-65.

Dezhbakhsh, H. and J.G. Thursby (1995), A Monte Carlo comparison of tests based on the Durbin-Watson statistic with other autocorrelation tests in dynamic models, *Econometric Reviews*, 14, 347-365.

Dickey, D.A., W.R. Bell and R.B. Miller (1986), Unit roots series model: Tests and implications, *American Statistician*, 40, 12-26.

Dufour, J.M. (1990), Exact tests and confidence sets in linear regressions with autocorrelated errors, *Econometrica*, 58, 475-494.

Dufour, J.M. and M.L. King (1991), Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or non-stationary AR(1) errors, *Journal of Econometrics*, 47, 115-143.

Durbin, J. (1970), Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, *Econometrica*, 38, 410-421.

Durbin, J. and G.S. Watson (1950), Testing for serial correlation in least squares regression I, *Biometrika*, 37, 409-428.

Durbin, J. and G.S. Watson (1951), Testing for serial correlation in least squares regression II, *Biometrika*, 38, 159-178.

Eglese, R.W. (1990), Simulated annealing: A tool for operational research, *European Journal of Operational Research*, 46, 271-281.

Ermini, L. (1993), Effects of transitory consumption and temporal aggregation on the permanent income hypothesis, *The Review of Economics and Statistics*, 75, 736-740.

Evans, G. and A. Deaton (1980), Testing linear versus logarithmic regression models, *The Review of Economic Studies*, 47, 275-291.

Evans, M.A. and M.L. King (1985), A point optimal test for heteroscedastic disturbances, *Journal of Econometrics*, 27, 163-178.

Evans, M.A. and M.L. King (1988), A further class of tests for heteroscedasticity, *Journal of Econometrics*, 37, 265-276.

Fan, Y. and Q. Li (1995), Bootstrapping J-type tests for non-nested regression models, *Economic Letters*, 48, 107-112.

Fisher, G.R. and M. McAleer (1981), Alternative procedures and associated tests of significance for non-nested hypotheses, *Journal of Econometrics*, 16, 103-119.

Franses, P.H. (1992), A model selection test for an AR(1) versus an MA(1) model, *Statistics and Probability Letters*, 15, 281-284.

Franzini, L. and A.C. Harvey (1983), Testing for deterministic trend and seasonal components in time series models, *Biometrika*, 70, 673-682.

Giles, D.E.A. (1975), Discriminating between autoregressive forms: A Monte Carlo comparison of Bayesian and ad hoc methods, *Journal of Econometrics*, 3, 229-248.

Gill, L., M. McAleer and C.R. MacKenzie (1986), Simple procedures for testing autoregressive versus moving average errors in regression models, Paper presented to the Time Series Conference of the Statistical Society of Australia, Canberra.

Godfrey, L.G. (1992), *The Implementation and Constructive Use of Misspecification Tests in Econometrics*, Manchester University Press, Manchester and New York.

Godfrey, L.G. (1998), Tests of non-nested regression models: Some results on small sample behaviour and the bootstrap, *Journal of Econometrics*, 84, 59-74.

Godfrey, L.G. and M.H. Pesaran (1983), Tests of non-nested regression models: Small sample adjustments and Monte Carlo evidence, *Journal of Econometrics*, 21, 133-154.

Godfrey, L.G. and A.R. Tremayne (1988), Checks of model adequacy for univariate time series models and their application to econometric relationships, *Econometric Reviews*, 7, 1-42.

Godfrey, L.G. and A.R. Tremayne (1992), Misspecification tests for univariate time series models and their applications in econometrics, in: L.G. Godfrey (ed), *The Implementation and Constructive Use of Misspecification Tests in Econometrics*, Manchester University Press, Manchester and New York, 1-66.

Goffe, W.L. (1996), SIMANN: A global optimization algorithm using simulated annealing, *Studies in Nonlinear Dynamics and Econometrics*, 1, 169-176.

Goffe, W.L., G.D. Ferrier and J. Rogers (1992), Simulated annealing: An initial application in econometrics, *Computer Science in Economics and Management*, 5, 133-146.

Goffe, W.L., G.D. Ferrier and J. Rogers (1994), Global optimization of statistical functions with simulated annealing, *Journal of Econometrics*, 60, 65-99.

Gourieroux, C. and A. Monfort (1995), Testing, encompassing, and simulating dynamic econometric models, *Econometric Theory* , 11, 195-228.

Gourieroux, C. and A. Monfort (1997), *Time Series and Dynamic Models*, Cambridge University Press.

Grant, M. (1987), The robustness of tests for autocorrelation in the dynamic linear regression model, Unpublished M.Ec. thesis, Monash University.

Green, W.H. (1997), *Econometric Analysis*, Third Edition, Macmillan Publishing Company, New York.

Griffiths, W.E., R.C. Hill and G.G. Judge (1993), *Learning and Practicing Econometrics*, John Wiley & Sons, New York.

Griliches, Z. (1967), Distributed lags: A survey, *Econometrica*, 35, 16-49.

Grose, S. (1998), Marginal likelihood methods in econometrics, Unpublished Ph.D. thesis, Monash University.

Hall, A.D. and M. McAleer (1989), A Monte Carlo study of some tests of model adequacy in time series analysis, *Journal of Business and Economic Statistics*, 7, 95-106.

Horowitz, J.L. (1997), Bootstrap methods in econometrics: Theory and numerical performance, in: D.M. Kreps and K.F. Wallis (eds), *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, 188-222.

Huberman, G. and G.W. Schwert (1985), Information aggregation, inflation and the pricing of indexed bonds, *Journal of Political Economy*, 93, 92-114.

Imhof, P.J. (1961), Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48, 419-426.

Inder, B.A. (1985), Testing for first order autoregressive disturbances in the dynamic linear regression model, Unpublished Ph.D. thesis, Monash University.

Inder, B.A. (1986), An approximation to the null distribution of the Durbin-Watson statistic in models containing lagged dependent variables, *Econometric Theory*, 2, 413-428.

King, M.L. (1979), Some aspects of statistical inference in the linear regression model, Unpublished Ph.D. thesis, University of Canterbury, Christchurch.

King, M.L. (1981), The Durbin-Watson test for serial correlation: Bounds for regressions with trend and/or seasonal dummy variables, *Econometrica*, 49, 1571-1581.

King, M.L. (1983), Testing for autoregressive against moving average errors in the linear regression model, *Journal of Econometrics*, 21, 35-51.

King, M.L. (1985a), A point optimal test for autoregressive disturbances, *Journal of Econometrics*, 27, 21-37.

King, M.L. (1985b), A point optimal test for moving average regression disturbances, *Econometric Theory*, 1, 211-222.

King, M.L. (1987a), Testing for autocorrelation in linear regression models: A survey, in: M.L. King and D.E.A. Giles (eds), *Specification Analysis in the Linear Model*, Routledge and Kegan Paul, London, 19-73.

King, M.L. (1987b), Towards a theory of point optimal testing, *Econometric Reviews*, 6, 169-218.

King, M.L. (1989), Testing for fourth-order autocorrelation in regression disturbances when first-order autocorrelation is present, *Journal of Econometrics*, 41, 285-301.

King, M.L. (1994), Recent developments in hypothesis testing from an econometrician's point of view: The editor's introduction, *Pakistan Journal of Statistics*, 10, 255-259.

King, M.L. (1996), Hypothesis testing in the presence of nuisance parameters, *Journal of Statistical Planning and Inference*, 50, 103-120.

King, M.L. (2000), Serial correlation, in B. Baltagi (ed), *Companion to Theoretical Econometrics*, Basil Blackwell, Oxford (forthcoming).

King, M.L. and D.C. Harris (1995), The application of the Durbin-Watson test to the dynamic regression model under normal and non-normal errors, *Econometric Reviews*, 14, 487-510.

King, M.L. and M. McAleer, (1987), Further results on testing AR(1) against MA(1) disturbances in the linear regression model, *Review of Economic Studies*, 54, 649-663.

King, M.L. and M. Rankin, (1993), Pre-test strategies for time series forecasting in the linear regression model, Monash University, Department of Econometrics, Research Working Paper, 17/93.

King, M.L. and T.S. Shively (1993), Locally optimal testing when a nuisance parameter is present only under the alternative, *The Review of Economics and Statistics*, 75, 1-7.

King, M.L. and M.D. Smith (1986), Joint one-sided tests of linear regression coefficients, *Journal of Econometrics*, 32, 367-383.

King, M.L. and P.X. Wu (1991), Small-disturbance asymptotics and the Durbin-Watson and related tests in the dynamic regression model, *Journal of Econometrics*, 47, 145-152.

King, M.L. and P.X. Wu (1997), Locally optimal one-sided tests for multiparameter hypotheses, *Econometric Reviews*, 16, 131-156.

Kirkpatrick, S., C.D. Gelatt. Jr. and M.P. Vecchi (1983), Optimization by simulated annealing, *Science*, 220, 671-679.

Koerts, J. and A.P.J. Abrahamse (1969), *On the Theory and Application of the General Linear Model*, Rotterdam University Press, Rotterdam.

Lehmann, E.L. (1959), *Testing Statistical Hypotheses*, John Wiley & Sons, New York.

Lehmann, E.L. (1986), *Testing Statistical Hypotheses*, Second Edition, John Wiley & Sons, New York.

Maddala, G.S. (1971), Some notes on discrimination between different distributed lag models, CORE Discussion Paper no. 7109, Catholic University of Louvain, Belgium.

Maddala, G.S. and A.S. Rao (1973), Tests for serial correlation in regression models with lagged dependent variables and serially correlated errors, *Econometrica*, 41, 761-774.

McAleer, M. (1987), Specification tests for separate models: A survey, in M.L. King and D.E.A. Giles (eds), *Specification Analysis in the Linear Model*, Routledge and Kegan Paul, London, 146-196.

McAleer, M. (1988), Some comments on testing time series models, *Econometric Reviews*, 7, 49-57.

McAleer, M. (1995), The significance of testing empirical non-nested models, *Journal of Econometrics*, 67, 149-171.

McAleer, M., C.R. McKenzie and A.D. Hall (1988), Testing separate time series models, *Journal of Time Series Analysis*, 9, 169-189.

McAleer, M. and M.H. Pesaran (1986), Statistical inference in non-nested econometric models, *Applied Mathematics and Computation*, 20, 271-311.

McAleer, M., M.H. Pesaran and A.K. Bera (1990), Alternative approaches to testing non-nested models with autocorrelated disturbances: An application to models of US unemployment, *Communication in Statistics: Theory and Methods*,19, 3619-3644.

McKenzie, C.R., M. McAleer and L. Gill (1999), Simple procedures for testing autoregressive versus moving average errors in regression models, *Japanese Economic Review*, 50, 239-252.

Metropolis, N., N. Rosenbluth, A. Rosenbluth, M. Teller and E. Teller (1953), Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, 21, 1087-1090.

Mizon, G.E. and J.F. Richard (1986), The encompassing principle and its application to non-nested hypotheses, *Econometrica*, 54, 657-678.

Nelson, C.R. and C.I. Plosser (1982), Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics*, 10, 139-162.

Newbold, P. and N. Davies (1978), Error mis-specifications and spurious regressions, *International Economic Review*, 2, 513-519.

Palm, F.C. and J.M. Sneek (1984), Significance tests and spurious correlation in regression models with autocorrelated errors, *Statistische Hefte*, 25, 87-105.

Pesaran, M.H. (1974), On the general problem of model selection, *Review of Economic Studies*, 41, 153-171.

Pesaran, M.H. and A.S. Deaton (1978), Testing non-nested nonlinear regression models, *Econometrica*, 46, 677-694.

Pesaran, M.H. and B. Pesaran (1993), A simulation approach to the problem of computing Cox's statistic for testing non-nested models, *Journal of Econometrics*, 57, 377-392.

Pesaran, M.H. and M. Weeks (2000), Non-nested hypothesis testing: An overview, in B. Baltagi (ed), *Companion to Theoretical Econometrics*, Basil Blackwell, Oxford (forthcoming).

Potscher, B.M. (1988), Towards a theory of point optimal testing: Comment, *Econometric Reviews*, 6, 235-239.

Rahman, S. and M.L. King (1994), A comparison of marginal likelihood based and approximate point optimal tests for random regression coefficients in the presence of autocorrelation, *Pakistan Journal of Statistics*, 10, 375-394.

Rahman, S. and M.L. King (1998), Marginal likelihood score-based tests of regression disturbances in the presence of nuisance parameters, *Journal of Econometrics*, 82, 81-106.

Rayner, R. (1993), Testing for serial correlation in regression models with lagged dependent variables, *The Review of Economics and Statistics*, 75, 716-721.

Saikkonen, P. and R. Luukkonen (1993), Point optimal tests for testing the order of differencing in ARIMA models, *Econometric Theory*, 9, 343-362.

Sargan, J.D. and A. Bhargava (1983), Testing residuals from least squares regression for being generated by the Gaussian random walk, *Econometrica*, 51, 153-174.

Schwert, W.S. (1989), Tests for unit roots: A Monte Carlo investigation, *Journal of Business and Economic Statistics*, 2, 147-159.

Shively, T.S. (1986), Issues in the analysis of stochastic coefficient regression models, Unpublished Ph.D. thesis, University of Chicago.

Shively, T.S. (1988a), An analysis of tests for regression coefficient stability, *Journal of Econometrics*, 39, 367-386.

Shively, T.S. (1988b), An exact test for a stochastic coefficient in a time series regression model, *Journal of Time Series Analysis*, 9, 81-88.

Shively, T.S. (1993), Testing for autoregressive disturbances in a time series regression with missing observations, *Journal of Econometrics*, 57, 233-255.

Shively, T.S., C.F. Ansley, and R. Kohn (1990), Fast evaluation of the distribution of the Durbin-Watson and other invariant test statistics in time series regression, *Journal of the American Statistical Association*, 85, 676-685.

Silvapulle, P. (1991), Testing non-nested disturbances, Unpublished Ph.D. thesis, Monash University.

Silvapulle, P. (1992), Testing for AR(p) against IMA(1,q) disturbances in the linear regression model, *Economics Letters*, 40, 257-261.

Silvapulle, P. (1994a), Testing for AR(1) against IMA(1,1) disturbances in the linear regression model, *Communication in Statistics: Theory and Methods*, 23, 701-720.

Silvapulle, P. (1994b), Testing AR(1) against MA(1) disturbances in the dynamic linear regression model, *Pakistan Journal of Statistics*, 10, 395-422.

Silvapulle, P. and M.L. King, (1991), Testing moving average against autoregressive disturbances in the linear regression model, *Journal of Business and Economic Statistics*, 9, 329-335.

Silvapulle, P. and M.L. King, (1993), Non-nested testing for autocorrelation in the linear regression model, *Journal of Econometrics*, 58, 295-314.

Smith, M.D. and A.R. Tremayne (1990), Finite sample theory for testing AR(1) against MA(1) time series models, a paper presented at the World Congress of the Econometric Society, Spain.

Stock, J.H. (1994), Unit roots and trend breaks in econometrics, in: R.F. Engle and D. McFadden (eds), *Handbook of Econometrics*, 4, North-Holland, Amsterdam, 2740-2841.

Tsionas, E.G. (1995), GAUSS version of Bill Goffe's simulated annealing program for global optimization, Unpublished paper, Internet address: http://gurukul.ucc.american.edu/econ/gaussres/optimize/anneal.src.

Walker, A.M. (1967), Some tests of separate families of hypotheses in time series analysis, *Biometrika*, 54, 39-68.

White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-25.

Wichern, D.W. (1973), The behaviour of the sample autocorrelation function for an integrated moving average process, *Biometrika*, 60, 235-239.

Wu, P.X. and M.L. King (1994), One-sided hypothesis testing in econometrics: A survey, *Pakistan Journal of Statistics*, 10, 261-300.