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

Errata

1. pii Line 11	"Estimaion" should read "Estimation".
2. pvii	Insert the following line after line 3,

"Table 5.7 Exact Medians of $\hat{\rho}_{OLS}$ for Design Matrices $X1$, $X4$, $X5$, $X7$ and $X8$; $T = 20{,}40$ ".

3. pxv Line 10	"there is usually no analytical solutions" should read "there are usually no analytical solutions".
4. p1 Line 10	"quoting Stock (2001, p29), which" should read, "which, quoting Stock (2001, p29)".
5. p3 Line 2	"The large models" should read "Large Models".
6. p3 Line 3	"the time series model" should read "time series model".
7. p3 Line 5B	"Dielbold …" should read "Diebold …".
8. p4 Line 3	"from the IT technology" should read "of IT technology".
9. p5 Line 5	"First …" should read " The first …".
10. p5 Line 6B	Change "instead of" to "compared to".
11. p7 Line 9	Change "and are …" to "and which are …".
12. p23 Line 5	Change "In another words" to "In other words".
13. p28 Line 4B	Change "invariant for affine" to "invariant to affine".
14. p33 Line 7	Change "produce" to "producing".
15. p38 Line 5B	"the non-normal errors" should read "non-normal errors".
16. p61 Lines 3 & 13	Change "Pfanzagal" to "Pfanzagl".
17. p65 Line 4	Change "so as it is …" to "as is …".
18. p66 Line 8B	Change "Same arguments" to "The same arguments".
19. p76 Line 12	"Phillips and Andrews (1987)" should read "Andrews and Phillips (1987)".
20. p83 Line 19	"level α is set at 50%" should read, "level $\alpha/2$ is set at 50%".
21. p92 Line 11	Insert the following paragraph after Line 11,

"Previous studies related to the proposed grid inversion method include Nankervis and Savin (1996) and Beran (1997). Nankervis and Savin proposed computing MU estimates based on a grid of null values using bootstrap. Their methodology involves bootstrap *p*-values; this thesis considers simulated quantiles. It can be shown that the two methods are equivalent. For the case where a test statistic is not asymptotically pivotal under the null, a possible way of computing MU estimates is to consider the *p*-values based on the double bootstrap test proposed by Beran (1997)."

22. p107 Line 8B — Change "to count for" to "to allow for".

23. p112 Line 8 — Change "subject to the constant" to "apart from the constant".

24. p114 Line 4 — Change "interception" to "intersection".

25. p116 Line 7 — Change "impact from" to "impact of".

26. p151 Line 6 — Change "estimator (5.3)" to "estimator (5.1)".

27. p152 Line 3 — Insert the following sentence before "In particular …",

"The exact medians of the OLS estimator for different design matrices and for a sample size of 20 and 40 are also reported in Table 5.7."

28. p152 Line 9 — Change "Waston" to "Watson".

29. p165 Line 8 — Change "gird" to "grid".

30. p175 Line 4 — Insert the following paragraph after Line 4,

"Theoretically speaking, for design matrix *X1*, $\hat{\rho}_A$, $\hat{\rho}^{MU}_{s(1,0)}$ and $\hat{\rho}^{MU}_{s(1,0.5)}$ should all be exactly MU, while $\hat{\rho}^{MU}_{s(1,0)}$ and $\hat{\rho}^{MU}_{s(1,0.5)}$ are exactly MU for *X2*, *X3*, *X4* and *X5*. But due to the limited number of replications and random errors, the simulated results would not be expected to be exact. The approximation largely depends on the accuracy of the computed median functions".

31. p183 — Insert the following note beneath Table 5.2,

"Note: $\hat{\rho}_A$ should be exactly MU for this model. The inaccurate results reported were due to a grid size of 0.05 used for computing the median function of the OLS estimator. The accuracy was improved in Table 5.3a when a grid size of 0.001 was used."

32. p200 Line 3B — Change "Similar …" to "A similar …".

33. p255 Line 10 — "Dielbold …" should read "Diebold …".

MEDIAN-UNBIASED ESTIMATION IN LINEAR AUTOREGRESSIVE TIME SERIES MODELS

A Thesis Submitted for the Degree of Doctor of Philosophy

DONGHUI CHEN

BSc (Tsinghua, China), MBA (CQU)

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

MONASH UNIVERSITY

AUSTRALIA

28 April 2001

CONTENTS

Chapter 3 Some General Methods for Constructing Median Unbiased Estimators: Theory

Chapter 4 Adjusting Marginal Likelihood Scores for Median-unbiased Estimators

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 other institution, and that, to the best of my knowledge, contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

[redacted] 02/05/01

Donghui Chen

Acknowledgments

This thesis could not have been completed without the encouragement and contribution from many individuals. Among them, I owe immeasurable gratitude to Prof. Max King, whose skill as a supervisor provides indispensable inspiration and guidance. His knowledge, experience and extraordinary enthusiasm have helped to shape this thesis into its current form. Those countless discussions with him not only provided constant insights into the research problems at hand, but also saw comforting moments through his understanding of both my work and personal difficulties. I am also grateful to him for his editorial advice in the final stages of writing this thesis.

This research was funded by an International Postgraduate Research Scholarship from the Australian government and a Monash Graduate Scholarship from Monash University. I gratefully acknowledge these financial supports. The Department of Econometrics and Business Statistics, Monash University provided me with a part-time Assistant Lecturerership during my candidature, which provided not only a help financially, but also a thoroughly enjoyable experience.

The completion of this study within the given timeframe was achieved with the remarkable research support from the Department of Econometrics and Business Statistics, Monash University. Over the years, a wonderful environment that encourages research excellence has formed in the department. This thesis benefited in numerous ways from this academically stimulating atmosphere and the sometimes thought provoking challenges from colleagues. The computing facilities available to the Ph D students are second to none, and Inge Meldgaard deserves the credit for her efforts in ensuring that my computing needs are met.

I also thank the department for funding my participation at three conferences. The suggestions given to me in these conferences have helped to improve some chapters of this thesis. In particular, the audience at the Australasian Meeting of the Econometric Society in Sydney provided valuable inputs for the early development of the research reported in this thesis.

In this foreign land of Australia, the kind help from many friends constantly saw me through difficult periods of adjustment and made me feel at home. They include Kim-Leng Goh, Ajit Majumder, Feng Ying and Zen Lu. Their thoughtful kindness, care and generosity during my stay in Australia deserve my wholehearted appreciation.

My utmost gratitude goes to my parents. Without their boundless understanding and encouragement, the completion of this research would be impossible. I am equally obliged to my sister Dongen and her family. I thank them for shouldering the extra responsibilities to our parents while I was away. Their optimism when facing difficulties has been a continual source of inspiration to me. This thesis is also written in the fondest memories of my grandparents.

Last but not least, no words can express my gratefulness to my wife, Wei. My accomplishment is only made possible by her endless love and unconditional sacrifice. So often did she salvage me from the dullness of my long working hours and miraculously make all my suffering easy to forget. This was frequently done by putting my interest in the top priority while sacrificing hers in the process. The arrival of the present from heaven - our little girl Iris at the end of 1998 brought enormous joy into my life. The sight of her has unfailingly allayed any pain and frustration that may dwell upon me.

Therefore I dedicate this thesis to my wife and daughter, two beautiful ladies who have been and will always be the most valuable treasures all my life.

Abstract

This thesis combines computer-intensive techniques with theoretical propositions to develop some general methods for constructing (approximately) median-unbiased (MU) estimators in small samples.

The first method proposed involves adjusting an estimating equation towards median-unbiasedness. It is shown that a median unbiased estimating equation is more likely to produce a median-unbiased estimator compared with a mean-unbiased estimating equation. If we subtract the original estimating equation by its median function, and if the resulting new estimating function satisfies certain conditions, the solution to the adjusted estimating equation will be median-unbiased. As there is usually no analytical solutions to the adjusted equation, iterative algorithms are needed. The proposed method is shown to be analogous to two of the existing mean-bias reduction methods. The median-unbiased estimator so constructed will generally have the same asymptotic properties as the solution to the original estimating equation.

This method is shown to be effective when applied to estimation of the linear regression with AR(1) disturbances. If the marginal likelihood score is adjusted towards median-unbiasedness, the solution to this adjusted estimating equation is almost exactly median-unbiased. Its root mean square error (RMSE) is significantly smaller than that of the least squares (LS) and maximum likelihood (ML) estimators. The method can be revised slightly to account for nuisance parameter problems in estimating a first-order dynamic regression model. An iterative algorithm delivers an approximately median-unbiased estimator, which successfully corrects the small sample bias of the LS estimator. The bias correction in this case is less accurate than in the previous example.

The second recommended method for constructing median-unbiased estimators is to invert a test statistic at the 50% significance level. We point out that there is a direct link between the power of a test and the small sample performance of the estimator based on inverting its median function. When the median function of a test is not

monotonic, this method may break down. We propose a 'grid inversion' algorithm to overcome the non-monotonicity problem associated with the median function of a single test statistic. The method is based on the 'median-envelope' of a series of tests and is shown to be able to better explore the good power properties over the whole parameter space.

We give a counter example of Andrews' (1993) MU estimator in a linear regression with AR(1) or random walk disturbances. His estimator breaks down for some design matrices, because the test statistic his estimator is based on lacks in power in small samples. We propose to use the point optimal invariant (POI) test instead. When the limiting power of a single test when the autoregressive parameter goes to -1 is not zero, the median function of the POI test is monotonic. Otherwise the median envelope method should be used. We outline a simple way of calculating this limiting power. The MU estimator based on POI test statistics is shown to be median-unbiased for all design matrices examined, including the cases where other methods fail. The RMSEs of the new estimators are significantly smaller than their biased counterparts. The proposed MU estimator based on the median envelope is also found to be quite robust to non-normal errors and other forms of error misspecifications.

Finally, we show that the median-unbiased estimators have the capability of improving the small sample performance of hypothesis testing and forecasting procedures. In particular, Wald tests based on the MU estimators successfully correct the local biasedness and non-monotonic power plaguing the Wald tests based on the biased estimators in the linear regression with AR(1) disturbances and the dynamic linear regression model. The power curves of the modified tests are properly centred and monotonic. As for prediction, the forecasting errors of the predictors based on the proposed median-unbiased estimators are shown to be significantly smaller than those based on the conventional estimators for these autoregressive models.

Chapter 1

Introduction

1.1 Some Emerging Trends in Econometrics

It is commonly agreed (see e.g., Darnell, 1984, 1994 and Bjerkholt, 1995) that R. Frisch first coined the term 'econometrics', which was envisaged by him as a new discipline intermediate between mathematics, statistics and economics, and more importantly, a powerful unification of the three. The mission of econometrics, as described by Frisch (1933), was 'to turn pure economics, as far as is possible, into a science in the strict sense of the word' (Chipman et al., 1971, p386). The foundation of the Econometric Society in 1930 formally marked econometrics emerging as a distinct subject independent of either statistics or economic theory. The next important milestone in the short history of econometrics ought to be the inauguration of the Cowles Commission shortly after World War II, quoting Stock (2001, p29), which

> '... over the course of a few years, developed a research agenda that structured macroeconometrics for the second half of the 20th century. The central vision of this research program was simple: the development of a mathematical model of the macroeconomy with grounding in economic theory, with parameters estimated using sound statistical methods, tested against and thus consistent with empirical evidence.'

Since then, economists have relied more and more on econometric models for testing economic theories, for macroeconomic forecasting, and for advising policymakers. Recently, several prominent econometricians contributed to an open forum organised by the *Journal of Econometrics* to mark the publication of its 100th volume, in which they assessed the current status of econometrics and identified some important trends that may influence the future development of the subject. The motivation of this thesis is largely rooted in some of these emerging themes of econometrics.

One such trend is the rapid development in the analysis of non-stationary time series. Since the early 1980s, a good deal of time-series econometrics has dealt with nonstationarity. As Phillips (2001, p21) remarked,

> 'the preoccupation has steadily become a central concern and it seems destined to continue, if only for the good empirical reason that most macroeconomic aggregates and financial time series are dominated by trend-like and random wandering behaviour. Behaviour, it should be said, that is very imperfectly understood.'

So, the study of trends brings together empirical-quantitative and theory-quantitative aspects of modelling and has, in turn, been empowered by that synergy. The literature is already vast and continues to grow swiftly, involving a full spread of participants and engaging a wide sweep of academic journals.

However, the focus on modelling non-stationary time series also draws criticism from various econometricians, see Heckman (2001), Granger (2001) and Maasoumi (2001) among others. In particular, Maasoumi (2001) observed that the currently popular dynamic models in which the own past history dominates, seem rather barren transformations of more economically interesting distributed lag processes. They are useful and economical curve-fitting media, but they do not reveal much about the behaviour or working of economies. The fact that the multi-trillion dollar quantities like the US GDP appear to move like a random walk is a numerical artefact, not a causal model which can play a role in policy analysis. Hence, we need to specify structural models that take account of the time series properties of the variables involved.

Therefore a combination of the dynamic specification and structural model building is necessary. Originally, the large models were not very dynamic, in contrast with the 'time series analysis' which concentrated on dynamics, paid little or no attention to economic theory and built models involving only a few variables. As

Granger (2001) remarked, over the years these two approaches have interacted with each other, one side learning from and being influenced by the other. The large models became more dynamic and involved unit roots and cointegration, the time series models considered size, that is, the number of variables used more seriously and payed more attention to the use of economic theory. The same view was also held by Hsiao (1997), Krishnakumar (2001) and Phillips (2001).

A second new development in econometric theory during the past two decades is the shift of focus from asymptotic theory to finite sample results. Traditionally, there was little econometricians could do about the exact finite sample distributions of estimators and test statistics. Therefore most inference procedures rely on first-order asymptotics. In science it is widely accepted that the first term of Taylor's series approximations, already merely locally valid, can be very poor. First-order asymptotic expansions, already less solid as an approximating concept, delivers even less. The Monte Carlo evidence provides frequent embarrassing evidence against first order asymptotic results. Many studies show that, in many settings, first-order asymptotic theory provides poor approximations to the finite sample distribution, and thus provides a poor basis for inference in applications.

This shift towards more accurate and more reliable small sample inference is largely facilitated by the rapid increase in the computing capacity. As King (1987b, p170) remarked,

> 'advances at all levels have reduced the costs of computing to such an extent that highly computational procedures are becoming more and more feasible. ... It appears that we now reaching the state where we should be asking: what kind of inference procedure would we wish to use if computation were not a constraint?'

Dielbold (2001) agrees that the pervasive effects of advances in processing speed have produced an irreversible and ongoing shift away from closed-form analytic methods and toward algorithmic numerical and simulation methods. This move is obvious in all aspects of econometric modelling. In particular, resampling techniques, such as bootstrap, are wonderfully simple tools for mimicking the sampling inference methods. They have been widely used to explore higher-order approximations to unknown distributions of econometric statistics. Also commenting on the impact from the IT technology on the development of econometrics, Maasoumi (2001, p85) went even further and predicted that

> 'perhaps there will be a development towards more realistic models and inference techniques, accommodating more sophisticated adaptive behaviour and learning,.... Perhaps even model specification will be recast into a learning and adaptive procedure, ... which will truly acknowledge that all models are approximations/misspecifications.'

1.2 Motivation and Direction

The research reported in this thesis was motivated by the on-going pursuit of exact small sample estimation procedures by econometricians. Since Fisher (1925) advocated the use of maximum likelihood (ML) based inference procedures, ML estimation has generally become very popular. This popularity is mainly based on asymptotic optimality properties and on computational convenience, as commented by Kiviet and Dufour (1997). Given present-day computer speed and facilities, however, practitioners can and should bring more aspects into their statistical utility function than just ease of computation and behaviour in infinitely large samples. Nowadays, a more challenging and appropriate objective is to employ procedures which optimise the actual efficiency and accuracy from the finite set of sample data at hand. As far as estimation procedures are concerned, it is no longer the only requirement today for the estimator to be asymptotically unbiased and relatively efficient. Instead, these days, the profession can be much better served if provided with estimators that have adequately characterised approximation levels and performance guarantees in small samples.

The focus on small sample properties of inference procedures is particularly important due to the non-experimental nature of economic data. As explained by Haavelmo (1944), by non-experimental, one means data cannot be generated by providing appropriate stimuli to elicit changes in the response variable, as in the ideal

type of experiment, but rather, a researcher can only play the role of a passive observer and gather data the way they are churned out by nature. This is a prominent feature distinguishing econometrics from its parent discipline, statistics. King (1996) pointed out two potential problems in inference procedures due to the non-experimental nature of the discipline. First is that a model cannot be specified with the level of certainty of that in disciplines where data are the direct product of experiments. This highlights the necessity to specify a model as a stochastic process to reflect the uncertainty involved. The other problem is that the effects of different factors in a model cannot be isolated in 'controlled' experiments as in the disciplines of natural science. In this less than ideal situation, econometric models often contain large numbers of parameters. As such, econometricians are often encountered with the difficult situation in which a large stochastic model has to be identified given only a limited amount of data. Therefore the precision of the estimators and the power of tests in small samples become crucial in the model building process.

To develop exact finite sample procedures usually involves sacrificing generality in exchange for a gain in small sample efficiency, see discussions in King (1987b). Estimators that are applicable generally for a class of models sometimes fail to explore fully the information contained in the data structure at hand. One way to improve on this is to design procedures that depend on the design structure. A prominent problem is the correction of the small sample bias of an estimator. Bias-reduction becomes important as many estimators are only unbiased asymptotically. Traditional bias-correction methods usually involve approximating the first-order bias function by asymptotic expansions and subtracting it from the original estimator. The finite sample bias-reduction approach, however, instead of using the uniform asymptotic adjustment factor, allows the correction factor to depend on the data. This will draw extra computational cost as the factor has to be computed differently each time, but it will also provide results that are more consistent with the data instead of asymptotic distributions.

The search for procedures efficient in small samples has been greatly enhanced by Monte Carlo methodology in the past two decades. Monte Carlo experiments are often employed to simulate random processes mimicking real life happenings with the use of random numbers, usually with the aim of investigating or estimating properties of statistics that are analytically intractable. Such studies help to decide on the best approach to use, or provide input directly needed, for modelling, estimation, testing and prediction. The simulation experiments, however, are never an end product in econometrics, their results await to be applied to non-experimental situations.

This thesis is mainly concerned with developing some 'operational exact techniques' (Kiviet and Dufour, 1997) for point estimation in small samples. Our attention is given to the impartiality of the estimator. In pursuit of unbiasedness in finite samples, we need bias-reduction techniques that are different from those based on asymptotic theories. Computer-intensive methods are used to explore the exact distributions of estimators and test statistics. To correct bias by computer simulations is not a new idea. For example, in an interview (Phillips, 1988), James Durbin revealed that in the early 1950s, he attempted to study bias-correction based on computer-intensive methods but eventually gave up the idea because the computers available at that time could not provide the speed and capacity required by the research.

The focus of this thesis is on the concept of median-unbiasedness. Unbiased estimators play an important role in point estimation theory. Because a uniformly optimal estimator is almost impossible to find, unbiasedness is usually the first prerequisite for a 'good' estimator. Unbiasedness makes sure that no one or more values of a parameter are too strongly favoured at the cost of neglecting other possible values (Lehmann, 1983, p5). Among different definitions of unbiasedness, median-unbiasedness has not drawn as much attention as mean-unbiasedness. Different researchers (e.g., Brown, 1947, Birnbaum, 1961 and van Haart, 1962), however have criticised the inappropriateness of mean-unbiasedness in some situations. They showed that median-unbiasedness possesses some attractive features and may be preferred as an impartiality measure under some circumstances. This thesis adds to the literature that advocates the use of median-unbiased (MU) estimators.

The main contribution of our research is to supply some generally applicable methods for constructing (approximately) MU estimators. We found that there is a

lack of such systematic discussions on how to construct MU estimators in the literature. Our proposed methods will serve as alternative bias-reduction techniques to the existing ones that were all designed to correct mean-bias.

However, as pointed out by Firth (1993), it is not an assumption of this thesis that unbiased estimation is always desirable and therefore bias-reduction is always necessary. The merits of bias-reduction in any particular problem will depend on a number of factors, including the skewness of the distribution of the estimator and any sacrifice in precision that might result. Sometimes, the variance of the bias-corrected estimator might be inflated significantly, thus offsetting any benefit of correcting the bias. Although in the examples examined in this thesis, the proposed estimators do not suffer from this problem, it is always a critical issue that researchers have to look out for before adopting any bias-reduction techniques in their studies.

The models we apply our proposed method to are the linear autoregressive models with exogenous variables. The bias in the estimation of these models without explanatory variables are well documented, therefore it is natural for us to compare the proposed techniques with the existing ones. When extra regressors are added into the model, the factors that affect the bias become more complicated and the analytical methods become even more intractable. Although practically popular, models containing exogenous regressors have not been studied systematically in terms of small sample bias-correction. We apply the proposed methods to these models and derive some estimators which are almost median-unbiased, with smaller overall risks in small samples and are applicable for most regressor structures.

1.3 Scope and Plan

The thesis is organised into seven chapters. Following this introductory chapter, Chapter 2 reviews the literature related to unbiased estimation and bias-reduction techniques. In a general point estimation setting, we start the survey by comparing several different definitions of unbiasedness. The link between unbiasedness and the risk function is highlighted. The two most frequently used unbiasedness criteria, mean-unbiasedness and median-unbiasedness are deliberated upon. In the section that follows, we aim to reveal the circumstances under which median-unbiasedness might be a more appropriate measure of the impartiality of a point estimator. This is followed by a thorough discussion of the properties of the MU estimators both asymptotically and in small samples. The focus is given to the optimality results derived by various researchers. Several optimality measures that are different from the ones usually used for mean-unbiased estimators are explored. The applications of MU estimation in different econometric models are then surveyed. Attention is given to the rationale of using such estimators and the methods used to derive them. It is found that the most important applications are in estimating autoregressive models. In the second half of the chapter, we examine a different yet closely related topic – small sample bias-reduction methods. These methods are classified into two categories: bias-correction and bias prevention. The two major bias-correction methods: the analytical approach and the bootstrap approach are compared, followed by a detailed discussion of previous studies that aimed to adjust estimating equations to prevent the bias in small samples. Finally, the research questions emerging from the literature review are summarised, as they will be dealt with in the proceeding chapters.

Chapter 3 gives the theoretical outline of the methods developed in this thesis which can be used to construct MU estimators. Lehmann's (1959) work implicitly indicated that MU estimators could be constructed by inverting the median functions of the sufficient statistics. We first review this approach and point out why it has not been widely adopted. Based on this fundamental result, we propose two different methods for constructing MU estimators. The first one is analogous to the bias-prevention technique in the mean-unbiased estimation context. The estimating equations are adjusted towards median-unbiasedness in order to correct the median bias in the estimator. The second method is based on the duality between a test statistic and a confidence interval. But a slightly different method of inverting the median function is proposed to suit point estimation. The link between the proposed methods and the existing bias-reduction techniques are explored. Both methods are extended to the multivariate case in which nuisance parameters have to be eliminated.

In particular, Chapter 3 outlines the main theoretical contributions this thesis makes. Constructing MU estimators by adjusting the estimating equations, as far as we know, has not been discussed in any previous study. We argue that there is a stronger link between a MU estimating equation and a MU estimator compared with their mean-unbiased counterparts. Therefore it is more likely to correct the (median-) bias by adjusting the estimating equation towards unbiasedness. The second approach, i.e., inverting the median function of a test statistic, is not new. However, we propose a different algorithm – grid inverting, together with the use of the point optimal tests, which greatly enhance the efficiency of such an approach. We expect these two methods could be applied to a range of econometric models and deliver (median-) unbiased estimators in small samples. They will serve as useful alternative bias-reduction techniques in situations where mean-unbiasedness is hard to achieve or robustness becomes crucial.

Chapter 4 provides empirical illustrations of the first method developed in Chapter 3. We choose to adjust the marginal likelihood scores towards median-unbiasedness in the linear regression with AR(1) disturbances and the first-order dynamic linear regression model. These two models are probably the most studied time series models in econometrics. We attempt to show that the proposed method can correct the small sample bias as well as or even better than the existing techniques. In the first example, the method is used directly as the median function is invariant to nuisance parameters. In the second example, however, this invariance does not hold. So the method is revised to account for the nuisance parameters. As a result, an iterative algorithm is developed to produce an approximately MU estimator. The new estimators are compared with their more conventional counterparts via Monte Carlo studies. The small sample bias and root mean squared errors (RMSEs) are examined.

Chapter 5 explores the relationship between the power of a test and the performance of the MU estimator based on inverting its median function. This issue is largely ignored by previous studies. As an example, we show that Andrews' (1993) MU estimator breaks down when extended to models with certain exogenous regressors. The reason lies in the lack of power of the test he chose to invert. In the linear regression with AR(1) and random walk disturbances, we examined the median functions of several frequently used tests and particularly the point optimal invariant (POI) tests proposed by King (1985a, 1987b) and Dufour and King (1991). We attempt to show that the POI tests are good candidates for constructing MU estimators due to their excellent small sample power properties. Monte Carlo studies are conducted to compare the performance of the estimators based on inverting different tests. In particular we compare the estimator based on the POI tests and Andrews' estimator. Finally, the robustness of the proposed estimators to non-normal errors is examined.

Chapter 6 studies hypothesis testing and forecasting procedures based on the MU estimators proposed in the previous chapters. We attempt to correct the small sample deficiencies of the Wald test by using an unbiased estimator in the test. The tests examined include a test of autocorrelated disturbances, a test for random walk disturbances and a test for the lagged dependent variable coefficient. We expect that, by correcting the bias in the point estimator, we should also be able to correct the small sample bias in the Wald test. The simulated power curves of the Wald tests based on different estimators are compared. The second half of the chapter is concerned with forecasting. It is revealed that the risk of a predictor is closely linked to the bias in the point estimator. Therefore Monte Carlo studies are conducted to compare the prediction risks based on different estimators.

The final chapter of the thesis summarises the major findings from the previous chapters in terms of satisfying the goals we set up in this introduction. Possible research questions which require further research are also identified.

1.4 Computations

All calculations reported in this thesis were performed using GAUSS System Version 3.2.11 (Aptech Systems, Inc.) on IBM-compatible personal computers. Random numbers were generated by the built-in random number generators (function RNDS) of the software for Monte Carlo simulation experiments. In the case of optimisation problems, the Constrained Optimisation module of the GAUSS System was utilised.

Chapter 2

Median Unbiased Estimation in Econometrics: Literature Review

2.1 Introduction

Point estimation is one of the most common forms of statistical inference. It involves procedures to specify a plausible value for an unknown parameter based on some observed data. The quality of an estimator is then measured by criteria such as impartiality, efficiency and robustness. Due to the non-experimental nature of economic data, it is even more crucial for econometricians to rely on estimators that have sound performance not only asymptotically but also in small samples. However, as there is typically no unique, convincing definition of optimality, the optimal estimation procedure depends heavily on the assumed utility (or risk) function. In most cases, unbiasedness is always one of the most frequently used quality measures for a point estimator. Median-unbiasedness is one of several unbiasedness definitions. Compared with mean unbiasedness, which is its much more popular alternative, median-unbiasedness possesses some unique features that can be attractive under different circumstances. Many early researchers, including Brown (1947), Lehmann (1951, 1959) and Birnbaum (1961, 1964), emphasised the importance and plausibility of the concept of median-unbiasedness. But since then, mean unbiasedness has always dominated point estimation theory until recently when MU estimators were successfully used in several time series models, which to some degree reminds researchers of the importance of the class of MU estimators.

This chapter provides a review of MU estimation in econometrics. The scope covers not only MU estimators and their applications, but also some of the related robust methods and bias-correction techniques. Our review inevitably focuses on the facets of the literature that offer insights into and direct exposition of the research problems of concern in this thesis, and therefore is by no means exhaustive. The main objective of this chapter is to synthesise the messages contained in previous studies in order to clarify and highlight areas in the literature where research questions remain open or further work is required. The succeeding chapters will address these research questions identified from this survey.

This review is organised as follows. General principles of unbiased estimation as well as different definitions of unbiasedness are deliberated upon in Section 2.2. Focus is given to the link between unbiasedness and the risk function. Section 2.3 compares median-unbiasedness with mean unbiasedness and reviews several circumstances in which median-unbiasedness is more plausible. Small sample and asymptotic properties of MU estimators are also discussed. Section 2.4 reviews applications of MU estimation in econometrics from early contributions to recent developments. Some robust methods, including least absolute deviation (LAD) estimation, which are closely related to MU estimation, are reviewed in Section 2.5. Section 2.6 looks at different small sample bias-correction techniques developed by econometricians. Although all of these techniques are aimed at reducing the mean-bias of an estimator, they provide some guidelines for us to apply similar methods in the context of MU estimation. Section 2.8 concludes the review by identifying some research gaps left unfilled in the literature.

2.2 Theory of Unbiased Estimation

2.2.1 Risk Function and Unbiasedness

We consider the general setting of statistical estimation. The observed data are postulated to be a random variable X (typically vector-valued) taking on values in the sample space Ξ, according to a distribution P, which is known to belong to a family Π. The distributions are indexed by an unknown parameter θ, taking values in a set Ω. A real valued function g is defined over the parameter space Ω, whose value at θ is to be estimated; we shall call $g(\theta)$ (in many cases, $g(\theta)=\theta$) the *estimand*. The estimation problem is then the determination of a suitable *estimator*, that is, a function δ defined over the sample space, of which it is hoped that $\delta(X)$

will tend to be close to the unknown $g(\theta)$. The value $\delta(x)$ taken on by the estimator for the observed value x of X is then called the *estimate* of $g(\theta)$.

We then, following Lehmann (1959, pp5-6), define a loss function $L(\theta,d)$, to measure the consequences of estimating $g(\theta)$ by a value d. The loss function is usually required to satisfy the following conditions,

$$L(\theta,d)\geq 0 \text{ for all } \theta,\ d, \tag{2.1}$$

and

$$L(\theta,g(\theta))=0 \text{ for all } \theta. \tag{2.2}$$

The accuracy of an estimator δ is then measured by the *risk function*

$$R(\theta,\delta)=E_\theta\{L[\theta,\delta(X)]\}, \tag{2.3}$$

the long term average loss resulting from the use of δ. One would like to find a δ which minimises the risk for all values of θ. However, except for a constant parameter, there exists no uniformly best estimator (Lehmann, 1959), which minimises (2.3) simultaneously for all values of θ.

One way of avoiding this difficulty is to restrict the class of estimators by ruling out estimators that too strongly favour one or more values of θ at the cost of neglecting other possible values. This can be achieved by requiring the estimators to satisfy a condition which enforces a certain degree of impartiality, i.e. unbiasedness.

Lehmann (1951, 1959) recommended a general concept of *risk-unbiasedness*, which is a theoretical definition of unbiasedness depending on the loss function L. An estimator δ of $g(\theta)$ is said to be risk-unbiased if it satisfies

$$E_\theta L[\theta,\delta(X)]\leq E_\theta L[\theta',\delta(X)] \tag{2.4}$$

for all $\theta' \neq \theta$. If one interprets $L(\theta,d)$ as measuring how far the estimated value d is from the estimand $g(\theta)$, then risk-unbiasedness states that $\delta(X)$ is on average at least as close to the true value $g(\theta)$ as it is to any false value $g(\theta')$.

The importance of this general definition is shown by its connection with the principle of invariance. An estimator $\delta(X)$ is said to be invariant to a 1:1 transformation family H, if for any $h \in H$,

$$\delta(h(X)) = h(\delta(X)). \tag{2.5}$$

Lehmann's (1951) results state that whenever among all risk-unbiased estimation procedures there exists a unique one that uniformly minimises the risk, then it is almost invariant. And under certain restrictions on the transformation group, the converse statement is also true, i.e., if among all invariant procedures there exists one that uniformly minimises the risk, then it is unbiased in the sense of (2.4).

2.2.2 Median as a Location Estimator

We begin by stating the definition of the median and some of its properties as a location parameter estimate before we define median-unbiasedness. A real number m is a median for the random variable Y, if

$$\Pr\{Y \geq m\} \geq \tfrac{1}{2} \text{ and } \Pr\{Y \leq m\} \geq \tfrac{1}{2}. \tag{2.6}$$

This definition of a median allows for non-uniqueness, and as a matter of fact, Lehmann (1983) showed that the medians of the same random variable always form a closed set. If a median m of Y is not a probability mass point, then condition (2.6) reduces to

$$\Pr\{Y > m\} = \Pr\{Y < m\} = \tfrac{1}{2}. \tag{2.7}$$

For a population with underlying distribution density $f(y)$ that has its median at μ, the sample median m is shown, among others, by Chu (1956) to be asymptotically normal with

$$\sqrt{T}(m-\mu) \rightarrow N(0,[4f^2(\mu)]^{-1}). \tag{2.7}$$

The efficiency of the median compared to the mean is 0.637 when the underlying distribution is normal, which is somewhat low. But its efficiency increases and exceeds unity as the tails of the parent distribution become heavy (Rao, 1988). The median is well known for its robustness properties. Huber (1987) showed that the median minimises the maximum asymptotic bias that can be caused by asymmetric data contamination. It is also the simplest so-called 'high-breakdown' estimate. Therefore it is usually used as a good initial value in iterative estimation by a robust method. Another important property is that it is invariant to monotonic transformations. The median of a function of a random variable is the function of the median of the random variable, provided the function is one-on-one.

2.2.3 Different Definitions of Unbiasedness

Depending on the choice of the loss function, different concepts of unbiasedness can be defined. The most frequently used are the following two:

1、*Mean-unbiasedness.* If the loss function is squared error, (2.4) becomes

$$E_\theta[\delta(X)-g(\theta')]^2 \geq E_\theta[\delta(X)-g(\theta)]^2 \tag{2.8}$$

for all $\theta' \neq \theta$. The left side of (2.8) is minimised by $g(\theta') = E_\theta\delta(X)$ and the condition of risk-unbiasedness therefore reduces to the mean-unbiasedness condition

$$E_\theta\delta(X) = g(\theta). \tag{2.9}$$

The condition of mean-unbiasedness ensures that in the long run, the amounts by which the estimator over- and under-estimates $g(\theta)$ will balance, so that the estimated value will be correct 'on the average'.

2. *Median-unbiasedness.* If the loss function is absolute error, (2.4) becomes

$$E_\theta|\delta(X) - g(\theta')| \geq E_\theta|\delta(X) - g(\theta)| \tag{2.10}$$

for all $\theta' \neq \theta$. The left side of (2.10) is minimised by any median of $\delta(X)$. It follows that the risk-unbiasedness condition reduces to

$$med_\theta \delta(X) = g(\theta). \tag{2.11}$$

An estimator satisfying (2.11) is said to be *median-unbiased* (MU). Different from mean-unbiasedness, median-unbiasedness ensures that the *frequency* but not the *amount* of over- and under-estimation of an estimator should balance, i.e., the probability of over-estimating is equal to probability of under-estimating in the long run.

There are other attempts to define unbiasedness in a more general sense. Brown (1947) suggested the use of likelihood-unbiasedness instead of mean-unbiasedness. An estimator is likelihood-unbiased if estimates in the neighbourhood of a given parameter value θ would occur more frequently when the true value is θ itself than when it differs from θ. Hence if we assume the estimator $\hat{\theta}$ of θ has probability density $h(\hat{\theta}|\theta)$, then $\hat{\theta}$ is likelihood-unbiased if

$$h(\theta|\theta') \leq h(\theta|\theta). \tag{2.12}$$

This definition has the advantage of being invariant under simultaneous one-to-one transformations of the parameter and the estimate, which is an important property not possessed by mean-unbiasedness.

Van der Haart (1962) attempted to generalise the definition of median-unbiasedness. For an estimator $\delta(X)$ of a parameter ϑ, he introduced a comparing value γ and a comparing estimator $c(X)$. The estimator $\delta(X)$ is called γ-unbiased (or more generally distribution-unbiased) if

$$\Pr\{\delta(X) \leq \gamma\} = \Pr\{c(X) \geq \gamma\}. \tag{2.13}$$

It is apparent that median-unbiasedness becomes a special case of this definition because if we choose $\gamma = \theta$ and $c(X) = \delta(X)$, the above condition becomes the definition of median-unbiasedness.

2.3 Properties of MU Estimators

2.3.1 Comparison of Mean-unbiasedness and Median-unbiasedness

According to Van der Haart (1962), Laplace (1774) was most likely the first who worked with the idea of MU estimators. He virtually rejected the use of arithmetic means of observations as location estimates, and therefore the concept of mean-unbiasedness. However, in the history of statistics and econometrics, much interesting work has been devoted to mean-unbiased estimators. Yet it is hard to find the requirement of mean-unbiasedness justified in print. According to Lehmann (1983, p4), it might be Gauss who first advocated the square of error as the measure of loss or inaccuracy. But even Gauss himself admitted that, should someone object to this specification as arbitrary, he was in complete agreement. He defends his choice by an appeal to mathematical simplicity and convenience. Among the infinite variety of possible loss functions for measuring the departure of the estimate from the true parameter, the square loss function is the simplest and is therefore preferable. This view was corroborated by Fraser (1956, p839), who observed that median-unbiasedness does not lend itself to the mathematical analysis needed to find minimum risk estimates, and hence has found little application. Still on the popularity of mean-unbiasedness, Birnbaum (1961) commented that mean-unbiasedness is merely a technically useful property of the classical estimators in the linear estimation problem, which, at least in the case of normal errors, could equally well or preferably be justified on the basis of median-unbiasedness.

The critics of mean-unbiasedness mainly focus on three aspects: 1. it is not robust towards extreme values and heavy tail distributions; 2. it is not invariant under one-to-one transformations; and 3. it is ill-defined when the parameter space is a closed set. In contrast, MU estimators are largely immune to these problems. Therefore under many circumstances, median-unbiasedness is a better measure of impartiality than mean-unbiasedness. We review these 3 aspects together with some of the examples that have emerged in the literature and illustrate those circumstances in which median-unbiasedness can be more relevant.

2.3.1.1 Robustness

While the mean is not robust when used as an estimator of a location parameter, estimators produced by squared error loss often are as uncomfortably sensitive to outlying observations and to the tail behaviour of the assumed distribution of the observed random variable, as pointed out by Lehmann (1983). Kendall and Stuart (1967, Section 17.9) reported that Girshick et al. (1946), Halmos (1946) and Savage (1954, p20) all provided examples that the available mean-unbiased estimators in certain situations can be even inferior to any single observation. On the other hand, MU estimators are much more robust in this sense. For example, Cox and Hinkley (1974) and Andrews and Phillips (1987) both reported that the MU estimator of the error variance in a linear regression is much more robust to non-normal errors than the mean-unbiased estimator. Andrews (1993) examined the robustness of his MU estimator of the first-order autoregressive model under different error structures and concluded that the MU estimator is very robust to skewed or heavy-tailed error distributions.

As an extreme case, the moments of the estimator may not exist in some situations. If this happens, mean-unbiasedness will not be an appropriate measure while median-unbiasedness is still applicable. Jensen (1979) observed that when the disturbances in a simple linear regression model are Cauchy instead of Gaussian, the

usual least squares estimator $\hat{\beta}$ is still median-unbiased (and optimal among all MU estimators for all bounded loss functions), but mean-unbiasedness is not applicable as the estimator does not have finite moments. Both Zaman (1981) and Fiebig (1985) also reported cases of estimators without finite sample moments. The parameter of interest was the reciprocal of a normal mean in a simultaneous equation model. In both cases, as observed by Zaman (1981), the use of quadratic loss as a criterion can conflict with the objective of obtaining an estimator that has high probabilities of being close to the true value.

2.3.1.2 Invariance to 1:1 Transformations

An important feature of the maximum likelihood estimator is its invariance to one-to-one transformations, i.e., if $\hat{\theta}$ is the ML estimator of θ and $g(\bullet)$ is a 1:1 transformation, then $g(\hat{\theta})$ is also the ML estimator of $g(\theta)$. But this property does not hold for mean-unbiased estimators. This is one of the most important arguments used by early critics of mean-unbiasedness. One simple example, as pointed out by Andrews and Phillips (1987), is that although the sample variance s^2 is a mean-unbiased estimator of the population variance σ^2, the sample standard deviation s is a biased estimator of the population standard deviation σ. In contrast, median-unbiasedness is invariant to any one-to-one transformations (see e.g., Brown, 1947 and Van der Haart, 1962). So given a MU estimator of the population variance (an example of this is given in the next section), its square root is also a MU estimator of the population standard deviation.

This invariance property of median-unbiasedness was explored by Andrews and Chen (1994) in the estimation of the impulse response function (IRF) in the AR(p) model. A scalar measure, the cumulative impulse response (CIR) was used to measure the persistence of the series instead of the whole IRF. The CIR, which is the sum of the IRF over all time horizons, equals $1/(1-\alpha)$, where α is the sum of the AR coefficients. Thus a MU estimator of α will lead to a MU estimator of the CIR. A similar approach towards estimating impulse response functions was also taken by

Kilian (1998) and Wright (2000a) but the focus there was mainly on interval estimation.

2.3.1.3 In a Restricted Parameter Space

When the parameter space is subject to restrictions, median-unbiasedness becomes a preferable measure of impartiality. Both Fuller (1996) and Andrews (1993) pointed out that when the parameter space is bounded and closed, it is impossible to have a global mean-unbiased estimator because all estimators are biased at extreme boundary points, while the MU estimator is immune to this problem. In time series regression models, when unit roots are taken into consideration, the parameter space for the autoregressive coefficient ρ is a bounded set $[-1,1]$. Therefore a uniformly unbiased estimator of ρ has to be median-unbiased instead of mean-unbiased. Another important situation of the parameter space being restricted occurs when the parameters are restricted by nonlinear constraints. This is frequently encountered in practice. Andrews and Phillips (1987) examined the linear regression model in which the coefficients are subject to nonlinear constraints. They concluded that the mean-unbiasedness condition becomes more restrictive than median-unbiasedness because estimators that take advantage of the restrictions on the parameters generally are mean-biased. MU estimators, however, can be adjusted to take account of restrictions without losing their property of median-unbiasedness.

2.3.2 Optimality Measures of MU Estimators

The most important optimality result associated with MU estimators is due to Lehmann (1959, pp80-83). It states that, for the family of densities that have monotone likelihood ratios (explained in the next paragraph), there exists a unique optimal MU estimator, which among all MU estimators, minimises $EL(\hat{\theta},\theta)$ for any loss function which for fixed θ has a minimum of 0 at $\hat{\theta}=\theta$ and is nondecreasing as $\hat{\theta}$ moves away from θ in either direction. If we take the loss function as the finite

sample concentration measure used in Pfazangl (1970), i.e., $L(\hat{\theta},\theta)=0$ if $|\hat{\theta}-\theta|\leq\Delta$ and $L(\hat{\theta},\theta)=1$ otherwise, it is seen that among all MU estimates, $\hat{\theta}$ minimises the probability of differing from θ by more than any given amount; more generally it maximises the probability

$$\Pr\{-\Delta_1\leq\theta-\hat{\theta}\leq\Delta_2\} \text{ for any } \Delta_1,\Delta_2\geq 0. \tag{2.14}$$

The real-parameter family of densities $p_\theta(x)$ is said to have a monotone likelihood ratio if there exists a real-valued function $T(x)$ such that for any $\theta<\theta'$ the distributions P_θ and $P_{\theta'}$ are distinct, and the ratio $p_{\theta'}(x)/p_\theta(x)$ is a nondecreasing function of $T(x)$. An important class of families of distributions that have monotone likelihood ratios are the one-parameter exponential families.

Pfanzagl (1979) extended Lehmann's (1959) results to exponential families with nuisance parameters and defined the conditions for an unique optimal MU estimator to exist. Brown, Cohen and Strawderman (1976) showed that within the class of MU estimators, the one based on the minimal sufficient statistic has the smallest risk for a wide class of losses, including but not limited to a convex loss function, which is a direct corollary from Lehmann's (1959, p80) results.

An important application of Lehmann's (1959) and Pfanzagl's (1979) results is the analogue to the Gauss-Markov theorem in the linear regression model derived by Andrews and Phillips (1987). In the linear regression model with 'consistent elliptically symmetrical' errors, the generalised least squares estimator (GLS) is shown to be the unique best MU estimator in the sense of uniformly minimum risk for any monotone loss function. More importantly, if the parameters are restricted in a known (possibly infinite) interval that does not depend on the true parameter value, the restricted GLS estimator is the unique best MU estimator. In contrast, in this case, similar optimality results do not hold for mean-unbiasedness.

The optimality of the MU estimators is also closely linked to another criterion used quite frequently to assess the quality of a point estimator, namely, the Pitman closeness (PC) criterion (Pitman, 1937): For estimating a parameter θ, consider two rival estimators δ_1 and δ_2; then δ_1 is said to be Pitman-closer than δ_2 if

$$\Pr_\theta\{|\delta_1-\theta|\le|\delta_2-\theta|\}\ge \tfrac{1}{2}, \tag{2.15}$$

for all $\theta\in\Omega$. According to the PC criterion, rival estimators are usually compared two at a time. But it is also important to obtain the Pitman-closest estimators within reasonable classes of estimators. Keating and Gupta (1984) and Keating and Mason (1985) considered the case of the general-scale family of distributions (with pdf of the form $\sigma^{-1}f(x/\sigma)$, $\sigma>0$). They compared the maximum likelihood, uniformly minimum variance, MU, and mean absolute deviation estimates of σ according to the PC criterion. They concluded that the MU estimator is the Pitman-closest among the four.

Ghosh and Sen (1989) extended this result and showed that the MU estimator in this case is also Pitman-closest within the scale-equivariant estimators under a wide class of loss functions. Ghosh and Sen also presented a general result in the location-scale families of distributions to link the median-unbiasedness to the PC criterion. Let T_1 and T_2 (nonnegative) be MU estimators of the location and scale parameter, respectively, and Z_1 and Z_2 be two ancillary statistics independent of T_1 and T_2, respectively, then T_1 is Pitman-closest among all estimators with the form $U=T_1+Z_1$, while T_2 is Pitman-closest among all estimators with the form $V=T_2(1+Z_2)$.

2.3.3 Asymptotic Concentration of MU Estimators

Pfanzagl (1970) suggested using the asymptotic and finite sample 'concentration' to measure the efficiency of a MU estimator. For a sequence $\{\hat{\theta}_t\}$ of estimators of θ, the asymptotic concentration is defined as

$$Lim_{t\to\infty}\Pr_\theta\{\theta-\delta_1/\sqrt{t}<\hat{\theta}_t<\theta+\delta_2/\sqrt{t}|\theta\} \tag{2.16}$$

for positive constants δ_1 and δ_2. The limit (2.16) exists for asymptotically normal sequences of estimators, but also for a broader class of estimators. Pfanzagl (1970) gives an example of a sequence of MU estimators $\{\hat{\theta}_t\}$ for which the distributions of $\sqrt{t}(\hat{\theta}_t-\theta)$ do not converge.

It immediately follows that a sequence $\{\hat{\theta}_t\}$ of estimators is asymptotically optimal in a class C if

$$\begin{aligned}&Lim_{t\to\infty}\Pr_\theta\{\theta-\delta_1/\sqrt{t}<\hat{\theta}_t<\theta+\delta_2/\sqrt{t}|\theta\}\\ &\ge Lim_{t\to\infty}\Pr_\theta\{\theta-\delta_1/\sqrt{t}<\hat{\hat{\theta}}_t<\theta+\delta_2/\sqrt{t}|\theta\}\end{aligned} \tag{2.17}$$

for any other sequence $\{\hat{\hat{\theta}}_t\}$ of estimators in C, for all δ_1, $\delta_2>0$ and all θ in Ω. In another words, in a class of estimators, if $\hat{\theta}$ has greater asymptotic concentration than any other estimators in the class for all δ_1, $\delta_2>0$, $\hat{\theta}$ is asymptotically optimal.

Pfanzagl (1971) proved that any sequence $\{\hat{\theta}_t\}$ of estimators which is asymptotically normal is also asymptotically MU in the sense that

$$Lim_{t\to\infty}\Pr(\hat{\theta}_t\le\theta)=Lim_{t\to\infty}\Pr(\hat{\theta}_t\ge\theta)=1/2. \tag{2.18}$$

He also showed that under certain regularity conditions, an asymptotically efficient sequence of estimators can be adjusted to give a sequence of asymptotically optimal MU estimators having the same asymptotic behaviour, and in which the adjustment amount tends to 0 as $t\to\infty$.

More importantly, Pfanzagl (1970) gives an upper bound (roughly speaking) to the asymptotic concentration for a sequence of MU estimators. Under certain regularity conditions,

$$\limsup_{t\to\infty} \Pr\{\theta-\delta_1/\sqrt{t} \le \hat{\theta}_t \le \theta+\delta_2/\sqrt{t}\}$$
$$\le \Phi(\delta_2 I(\theta)^{1/2}) - \Phi(\delta_1 I(\theta)^{1/2}) \tag{2.19}$$

for all δ_1, $\delta_2 > 0$ and all $\theta \in \Omega$, where $\Phi(\bullet)$ is the standard normal density function and

$$I(\theta) = E_\theta(-\frac{\partial^2 \ln L(x,\theta)}{\partial\theta^2}) \tag{2.20}$$

is the Fisher information matrix component. This result holds true without any assumptions concerning the convergence rate of the sequence, in particular, asymptotic normality is not required. Under suitable regularity conditions, this maximal asymptotic concentration is achieved by the sequence of maximum likelihood estimates. As maximum likelihood estimates are not MU in general, Pfanzagl (1971) defines conditions under which MU estimates with maximal asymptotic concentration exist. The results imply that the MU estimates with maximal asymptotic concentration exist for all exponential families fulfilling certain regularity conditions. Also see Strasser (1978). This is consistent with the fact that families with monotone likelihood ratios admit MU estimates with strong optimum properties, as suggested by Lehmann (1959, p83).

Pfanzagl (1971) went on to derive a similar efficiency bound in finite samples. Under certain regularity conditions, for every $\theta \in \Omega$ and t_1, $t_2 > 0$, there exists a number $c(t_1,t_2)$ such that for every sample size T and any MU estimator $\hat{\theta}_T$, we have

$$\Pr\{\theta-t_1/\sqrt{T} \le \hat{\theta}_T \le \theta+t_1/\sqrt{T}\}$$
$$\le \Phi(t_2 I(\theta)^{1/2}) - \Phi(t_1 I(\theta)^{1/2}) + c(t_1,t_2)T^{-1/2} \tag{2.21}$$

Michel (1973) provided a similar efficiency bound but gave a slightly different proof from that of Pfanzagl (1971). As an example, Pfanzagl (1979) applied his results to the use of a sample quantile $\hat{q}_{\alpha,t}$ as an estimate of the corresponding population

quantile q_α. The sequence $\{\hat{q}_{\alpha,t}\}$ is asymptotically median-unbiased. And if the shape of the population distribution is not known or if q_α is not the median of a symmetric distribution, the sequence is shown to attain the maximal concentration.

2.3.4 Efficiency Bounds for MU Estimators

Another approach for assessing the relative efficiency and optimality of a MU estimator, is by considering its 'diffusivity' as suggested by Sung (1988, 1990). Based on this measure, a generalised Cramer-Rao analogue for MU estimators having continuous density functions was derived for both a scalar parameter and vector-valued parameters.

Following Sung (1988), to assess the variation of a MU estimator, instead of using variance, we define the diffusivity of a MU estimator $\hat{\theta}$ by the reciprocal of twice of its density at its median point, i.e.,

$$d = 1/[2g_{\hat{\theta}}(\theta_0)] \tag{2.22}$$

where $g_{\hat{\theta}}(\bullet)$ is the density function of $\hat{\theta}$ and θ_0 is the true value of θ. Then under certain regularity conditions (see Sung, 1990), the diffusivity of a MU estimator in an exponential family is bounded by:

$$d \ge 1/E_\theta|U(x,\theta)| \tag{2.23}$$

where

$$U(x,\theta) = \partial l(x;\theta)/\partial\theta \tag{2.24}$$

is the score function and $l(x;\theta)$ is the log-likelihood function. This is a direct analogue of the Cramer-Rao inequality for mean-unbiased estimators, only with the

Fisher information replaced by the absolute moment of the sample score and the variance of the estimator replaced by diffusivity.

Sung (1990) extended the efficiency bound to the multidimensional parameter case. Let $\hat{\theta}=(\hat{\theta}_1,\ldots,\hat{\theta}_k)=(\delta_1(x),\ldots,\delta_k(x))$ be a MU estimator of a vector-valued parameter $\theta=(\theta_1,\ldots,\theta_k)$, i.e., each component $\delta_i(x)$ is MU for θ_i, and assume $\hat{\theta}$ has a continuous joint density function $g(\theta_1,\ldots,\theta_k)$. Then the joint diffusity of $\hat{\theta}$ is defined as

$$d=\left[2^k g(\theta_1,\ldots,\theta_k)\big|_{\theta_0}\right]^{-1}. \tag{2.25}$$

Similarly, under certain regularity conditions, the joint diffusity of $\hat{\theta}$ is bounded by the absolute moment of the score, i.e.

$$\frac{1}{2^k g(\theta_1,\ldots,\theta_k)\big|_{\theta_0}} \geq E_\theta\left|\frac{\partial^k l(x;\theta)}{\partial\theta_1\ldots\partial\theta_k}\right|. \tag{2.26}$$

As a simple example, Sung (1990) showed that in a multivariate normal distribution, the sample mean is a MU estimator for the population mean and its diffusity also attains the lower bound. In contrast, the MU estimator of the population variance does not attain the lower bound.

2.3.5 Multi-dimensional Median and Median-unbiasedness

It is a difficult problem to extend the definition of median-unbiasedness to the multi-parameter case. The reason is caused by the well-known difficulty in defining a multi-dimensional median. Small (1990) has surveyed previous efforts to extend the definition of median to the multivariate case. Rao (1988) also reviewed some of the multidimensional medians in the context of L1-norm inference. Both these reviews concluded that, unlike for the expectation operator, how to define a multivariate median is still an open question. Therefore the discussions of median-unbiasedness

so far are mainly restricted to the estimation of a scalar parameter. Here we examine some of the prominent definitions of a multivariate median before we give our definition of median-unbiasedness for a vector-valued parameter.

2.3.5.1 Marginal Median

The concept of marginal median is a genuine extension from the univariate median. First consider the discrete case. Let $x_1,\ldots,x_T$ be a sample of size T from a $p-$variate population, then the sample marginal median is defined as $m=(m_1,\ldots,m_p)'$ that satisfies:

$$\Pr[x_i \geq m_i]=\Pr[x_i \leq m_i]=1/2. \tag{2.27}$$

Therefore each component of the median is the median of the corresponding component of the variable. Rao (1988) showed that this median is the solution to the minimisation problem,

$$\min_{\mu}\sum_{1}^{T}\|x_i-\mu\|, \tag{2.28}$$

where the distance $\|\bullet\|$ is the L_1-norm distance given by,

$$\|x-y\|_1=|x_1-y_1|+\ldots+|x_p-y_p|. \tag{2.29}$$

The sample marginal median $m=(m_1,\ldots,m_p)'$ is then treated as an estimate of the population vector of marginal medians $\mu=(\mu_1,\ldots,\mu_p)'$.

Let f_i be the marginal density of each component x_i at its population marginal median μ_i, and define an association matrix by

$$\Gamma = \begin{bmatrix} \frac{\gamma_{11}}{f_1^2} & \frac{\gamma_{12}}{f_1 f_2} & \cdots & \frac{\gamma_{1p}}{f_1 f_p} \\ \vdots & \vdots & & \vdots \\ & & \ddots & \\ \frac{\gamma_{p1}}{f_p f_1} & \frac{\gamma_{p2}}{f_p f_2} & \cdots & \frac{\gamma_{pp}}{f_p^2} \end{bmatrix} . \tag{2.30}$$

where $\gamma_{ij} = \Pr\{x_i \le \mu_i, x_j \le \mu_j\} - \frac{1}{4}$. Let $m = (m_1, \ldots, m_p)'$ be the sample marginal median vector based on an i.i.d. sample from the population, then Babu and Rao (1988) and Rao (1988) showed that the sample marginal median vector is asymptotically normal with the limiting distribution given by,

$$\sqrt{T}(m - \mu) \to N_p(0, \Gamma). \tag{2.31}$$

Therefore testing the hypothesis $H_0 : \mu = \mu_0$ can be based on the test statistic

$$s^2 = T(m - \mu_0)' \hat{\Gamma}^{-1} (m - \mu_0), \tag{2.32}$$

which is asymptotically χ^2 with p degrees of freedom provided Γ is positive definite. The difficulty lies in finding a consistent estimator $\hat{\Gamma}$, see discussions in McKean and Schrader (1984) and Rao and Babu (1988).

The advantage of using the marginal median vector is its computational simplicity. It is also invariant to shifts in the individual components of the sample vectors but not invariant for affine transformations of the sample vectors. (Following Small (1990), affine invariance is defined as follows. If δ is an affine invariant estimate based on x_1, ..., x_T, then $A\delta A'$ should be the estimate based on $Ax_1, \ldots, Ax_T$ for any affine transformation matrix A).

2.3.5.2 Spatial Median

The spatial median was first considered by Haldane (1948). The median $m_f(x)$ of a p-variate probability distribution $f(x)$ is the solution to the minimisation problem:

$$E\|f(x) - m_f(x)\| = \inf_{\phi \in R^p} E\|f(x) - \phi\|, \tag{2.33}$$

where $\|\bullet\|$ is the usual Euclidean norm. This median is sometimes called the L1-median or the geometric median while the vector of marginal medians is sometimes referred to as the arithmetic median. In the special case where $p = 1$, the L1-median reduces to the standard univariate median. Lopuhaa and Rousseeuw (1987) reported that this median is very robust to data contamination (with a breakdown point of 50%). Brown (1983) studied the asymptotic properties of the L1 median and found it is asymptotically normal. If the population is multivariate normal, the asymptotic relative efficiency of the L1 median increases to one as $p \to \infty$. Another advantage is that it is unique when $T \ge 2$ provided the points are not all on the same plane. Some asymptotic dispersion measures of the L1 median were studied by Bose (1995).

It is subject to the same criticism that it is not invariant to affine transformations. Observing this, Rao (1988) proposed a generalised spatial median which minimises

$$\tfrac{T}{2}|S| + \sum_{1}^{T} [(x_i - m)' S^{-1} (x_i - m)]^{\frac{1}{2}}, \tag{2.34}$$

where S is the sample covariance matrix. It was shown that the solution is invariant to affine transformations. But little is known about its other properties.

2.3.5.3 Other Multivariate Medians

Oja (1983) proposed a different affine invariant median. In the discrete sample case, his median θ_m is found by solving the minimisation problem:

$$\min \gamma(\theta) = E[\Delta(x_1, x_2, ..., x_k, \theta)], \tag{2.35}$$

where

$$\Delta(x_1, x_2, ..., x_k, \theta) = \frac{1}{k!}\begin{vmatrix} 1 & 1 & \cdots & 1 & 1 \\ x_{11} & x_{21} & \cdots & x_{k1} & \theta_1 \\ x_{12} & x_{22} & \cdots & x_{k2} & \theta_2 \\ \vdots & & & & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{kk} & \theta_k \end{vmatrix} \tag{2.36}$$

is the volume of the simplex $(x_1, x_2, ..., x_k, \theta)$.

The advantage of Oja's median is its invariance to affine transformations. Its disadvantages are: 1. it is usually not unique; 2. its robustness to outliers is doubtful; and 3. it is very hard to calculate when the sample size gets large and when X becomes a continuous random variable.

Apart from the definitions mentioned above, there are a few others, such as the halfspace median suggested by Tukey (1975) and Donoho (1982), the simplicial depth median of Liu (1988, 1990), and the median definition based on convex hull stripping and related methods suggested by Seheult et al. (1976) and Green (1981). But none of these definitions has gone very far in the practice of statistical inference for different reasons.

As each definition has its benefits and drawbacks, in this thesis, we adopt the marginal median as the definition of a multidimensional median mainly for its simplicity. Another justification is that when assessing the impartiality of an estimator, it is usually more important to ensure the unbiasedness of each coordinate of the estimator (as each component alone is usually an estimate of a separate scalar parameter) rather than pursuing some hard-to-define global unbiasedness. Therefore a multi-variate estimator $\hat{\theta} = \delta(x) = \{\delta_1(x), \delta_2(x), ..., \delta_p(x)\}'$ is called median-unbiased if each component is median-unbiased, i.e.

$$med[\hat{\theta}] = \{m[\delta_1(x)], ..., m[\delta_p(x)]\}' = \theta_0, \tag{2.37}$$

where $m(\bullet)$ is the usual univariate median function. The definition of marginal median-unbiasedness was also adopted by Sung (1990), Rudebusch (1992), Andrews and Chen (1994) and Fair (1996).

2.4 Applications of MU Estimators

In this section, we survey the applications of MU estimators in different econometric models that have appeared in the literature. Attention is given to the circumstances in which MU estimators were requested and the methods used by previous researchers to construct these MU estimators.

2.4.1 Error Variance

Let $x_1, ..., x_T$ be a sample from a normal distribution $N(\mu, \sigma^2)$. Then it is well known that the sample variance

$$S^2 = (T-1)^{-1}\sum_{i=1}^{T}(x_i - \bar{x})^2 \tag{2.38}$$

is the best mean-unbiased estimator in the sense of uniformly minimum risk for any convex loss function (see Lehmann 1983, p185). However, S^2 is not median-unbiased, and neither is S as an estimator of σ. Eisenhart and Martin (1948) showed that S^2 is negatively median-biased. Eisenhart (1949) pointed out that the median-bias of S^2, and hence of S, is of interest in quality control. Cox and Hinkley (1974) suggested a MU estimator given by

$$S_{MU}^2 = \sum_{i=1}^{T}(x_i - \bar{x})^2 \Big/ med(\chi_{T-1}^2), \tag{2.39}$$

where the denominator stands for the median of a standard chi-squared random variable with $T-1$ degrees of freedom, which can be obtained from published tables or via numerical integration. Ghosh and Sen (1989) extended this estimator to the general location-scale family of distributions with pdf $\sigma^{-1}f((x-\theta)/\sigma)$ and showed that S_{MU} is not only median-unbiased, but also the Pitman-closest estimator of σ within a certain class of equivariant estimators. Andrews and Phillips (1987) applied a similar estimator to the estimation of error variance in a linear regression, i.e.

$$y = X\beta + u\,,\ E(u) = 0,\ \mathrm{cov}(u) = \sigma^2\Sigma, \tag{2.40}$$

where y is the dependent variable, X is a $T\times k$ matrix of fixed regressors and u is the vector of random errors; The estimator proposed for σ^2 was

$$S^2_{MU} = (y - X\tilde{\beta})'(y - X\tilde{\beta})/med(\chi^2_{T-k}), \tag{2.41}$$

where $\tilde{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y$ is the feasible GLS estimator. The same estimator was also defined in Pfanzagl (1979). Andrews and Phillips (1987) concluded that S^2_{MU} is the best MU estimator of σ^2 for any monotone loss function. In contrast to the optimality results for the mean-unbiased estimator S^2, this result holds even when the regression parameter β is subject to restrictions, provided that the parameter space of β has a non-empty interior. As expected, S^2_{MU} is always slightly larger than S^2, and the difference in the denominators of the two is approximately 0.66 when $T-k$ is between 8 and 50.

2.4.2 Autoregressive Models

Measuring the persistence of shocks to macroeconomic time series variables might be the single most prominent problem in econometrics during the past two decades. Since the seminal papers of Dickey and Fuller (1979, 1981) and the empirical studies by Nelson and Plosser (1982), much of the literature has focused on testing for whether the largest autoregressive root of a series is one. However, the emphasis on

unit root tests has been criticised by many researchers, see for example Campbell and Mankiw (1987), Cochrane (1988), Stock (1991, 1994)), Andrews (1993) and Hansen (1999). They argued that reporting only unit root tests is unsatisfying as a description of the data as this fails to convey information about the most likely model that is consistent with the observed data. In particular, mounting evidence, both theoretical and empirical, suggests that the unit root tests usually have low power for the alternatives where power is needed most, therefore frequently produce unconvincing results. All these point to the importance of having a reliable and impartial point estimator in small samples. We review the effort made by various researchers to construct MU estimators for the first-order autoregressive model, as this seems to be the most prominent application of MU estimation in econometrics.

Three models of first-order autoregression are considered. These models can be defined as follows,

Model 1: $y_t = \rho y_{t-1} + \varepsilon_t$,

Model 2: $y_t = \mu + \rho y_{t-1} + \varepsilon_t$,

Model 3: $y_t = \mu + \beta t + \rho y_{t-1} + \varepsilon_t$,

where $\varepsilon_t \sim IN(0,\sigma^2)$ and $\rho \in [-1,1]$. The initial conditions used by different researchers do vary, e.g., see Pantula et al. (1994). In the following discussion, we adopt the simple conditions that require y_0 to be stationary if $|\rho|<1$ and to equal zero if $|\rho|=1$ (a fixed start-up).

2.4.2.1 Asymptotically MU Estimators

This approach involves using local-to-unity asymptotic theory to construct confidence intervals or compute asymptotically MU estimators based on the limiting distribution of a test statistic. Developed by Bobkoski (1983), Cavanagh (1985), Phillips (1987) and Chan and Wei (1987), the local-to-unity reparametrization models the true value of ρ as being in a decreasing neighbourhood of one, specifically $\rho = 1 + c/T$, where c is a fixed constant (the Pitman drift) and T is the sample size. Cavanagh (1985) first applied this theory to construct confidence intervals based on the t statistic for Model 1.

Stock (1991) constructed two sets of confidence intervals based on the augmented Dickey-Fuller (1979) t statistic (ADF) and a modification of Sargan and Bhargava's (1983) locally most powerful (LMP) test statistic. As a special case, how to compute MU estimates was also described. Among the two test statistics examined, the ADF statistic was shown to deliver more reliable intervals and estimates. Therefore we review the results based on the ADF test for Model 3. The model is rearranged to yield the usual Dickey-Fuller regression and the ADF test statistic is given by,

$$t(\rho) = (\hat{\rho}_{OLS} - 1)/s(\hat{\rho}_{OLS}) \tag{2.42}$$

As $T \to \infty$, the limiting distribution of the ADF test is non-standard. Both Stock (1991) and Hansen (1999) derived it for Model 3. The limiting distribution is given by,

$$t(\rho) \to (\int_0^1 W_c dW) \Big/ (\int_0^1 W_c^2)^{1/2}, \tag{2.43}$$

where W_c is a detrended diffusion process and W is the standard Wiener process, for details see Chan (1988) and Stock (1991). Therefore the limiting distribution depends only on the local-to-unity parameter c and is continuous in c. The median function of this limiting distribution is then computed, i.e.,

$$m(c) = med[(\int_0^1 W_c dW) \Big/ (\int_0^1 W_c^2)^{1/2} \Big| c], \tag{2.44}$$

and a MU estimator of c is given by,

$$\hat{c} = m^{-1}[t(\rho)]. \tag{2.45}$$

The estimator for ρ is then retrieved from $\hat{c}$. The method of computing confidence intervals proposed by Hansen (1999) was slightly different from that of Stock's. It allows the null value in the test to vary and simulate the quantile functions of the limiting distributions of the series of test statistics. Confidence intervals are then computed by solving the equations for c_α,

$$t(\rho_0(c_\alpha)) = q_\alpha(\rho_0(c_\alpha)), \tag{2.46}$$

in which α is the pre-selected confidence level and q_α is the corresponding quantile function of the limiting distribution.

Recently, Elliot (1999) inverted the quantile functions of the limiting distribution of the efficient unit root tests proposed by Elliot et al. (1996) to construct confidence intervals in the same model. His tests were shown to be closer to the asymptotic power envelope of the testing problem and therefore more powerful than the ADF test in some circumstances. His method of constructing confidence intervals is very similar to the ones discussed above. But point estimation was not discussed in his studies.

From the above examples, we can criticise the asymptotic approach on two grounds. First, its median function is computed for the limiting distribution, therefore the quality of the final estimator really depends on whether this asymptotic distribution is a good approximation of the true distribution of the test statistic in finite samples. When the sample size is small, there has to be increasing discrepancies between the asymptotic and exact distributions. As a result, the estimator is only median-unbiased asymptotically. Second, both Stock (1991) and Hansen (1999) reported that the quantile functions of the limiting distribution are not monotonic for a certain range of c values. If the focus is on confidence intervals, this is not a serious problem as it will only cause disjoint or empty confidence intervals at times. But if a point estimate is to be calculated, this non-monotonicity will lead to multiple solutions, which is a serious problem plaguing the use of this method.

Different from all the above efforts, Fuller (1996, p579) proposed another asymptotically MU estimator based on the median function of the limiting distribution of the weighted symmetrical estimator. In model 1, the weighted symmetric (WS) estimator is given by,

$$\hat{\rho}_{WS} = (\sum_{t=2}^{T-1} y_t^2 + n^{-1}\sum_{t=1}^{T} y_t^2)^{-1}\sum_{t=2}^{T} y_{t-1}y_t \,. \tag{2.47}$$

The t statistic based on $\hat{\rho}_{WS}$ testing for a unit root is then given by

$$\hat{\tau}_{WS} = [\sigma_{WS}^2(\sum_{t=2}^{T-1} y_t^2 + n^{-1}\sum_{t=1}^{T} y_t^2)^{-1}]^{-\frac{1}{2}}(\hat{\rho}_{WS} - 1). \tag{2.48}$$

Fuller's estimator is then based on the median function of the limiting distribution of $\hat{\tau}_{WS}$, i.e.,

$$\hat{\tau}_{WS} \to \tfrac{1}{2}(\int_0^1 W^2(t)dt)^{-\frac{1}{2}}(W^2(1)-1) - (\int_0^1 W^2(t)dt)^{\frac{1}{2}}, \tag{2.49}$$

where $W(t)$ is the standard Wiener process. As the median of the limiting distribution (2.49) at $\rho = 1$ is approximately -1.20, Fuller's estimator is defined as

$$\tilde{\rho} = \hat{\rho}_{WS} + c(\hat{\tau}_{WS})[\hat{V}(\hat{\rho}_{WS})]^{\frac{1}{2}}, \tag{2.50}$$

where $\hat{V}(\hat{\rho}_{WS})$ is the estimated variance of $\hat{\rho}_{WS}$ given in the denominator of $\hat{\tau}_{WS}$ in (2.48) and the smooth function $c(\hat{\tau}_{WS})$ is chosen as

$$c(\hat{\tau}_{WS}) = \begin{cases} -\hat{\tau}_{WS} & \text{if } \hat{\tau}_{WS} \geq -1.2 \\ 0.035672(\hat{\tau}_{WS} + 7.0)^2 & \text{if } -7.00 < \hat{\tau}_{WS} \leq -1.2 \\ 0 & \text{if } \hat{\tau}_{WS} \leq -7.00 \end{cases}. \tag{2.51}$$

The nature of this estimator is similar to the one studied by Stock (1991). Both are based on inverting the median function of the limiting distribution of the t statistic except that Fuller's estimator is based on the WS estimator while Stock's

estimator was based on the usual ADF test statistic. Another difference is that estimator (2.51) coincides with $\hat{\rho}_{WS}$ if ρ is small (when $\hat{\tau}_{WS} \leq -7.00$) because for small ρ, no bias-correction is needed. This cut off point is roughly at $\hat{\rho}_{WS} = (T+49)^{-1}(T-49)$.

For Models 2 and 3, the same procedure can be applied with y_t demeaned or detrended first. Fuller (1996) showed that this estimator is approximately median-unbiased when the sample size is as small as 50. However, it suffers from an increased variance compared with $\hat{\rho}_{WS}$ for ρ values differing from one by more than $8/T$. As a result, the mean squared error of $\hat{\rho}_{WS}$ is smaller than that of $\tilde{\rho}$ for all ρ values that are different from one by more than four standard errors of the estimator. However, Fuller's (1996, p580) simulation results showed that the percentage difference in the mean squared error between the two estimators is modest. Fuller's estimator was used by Enders and Falk (1998) to examine the validity of purchasing power parity. Five exchange rates series were examined. While the DF unit root test fails to reject the unit-root null for all but one series, the OLS point estimates implied stationarity in all five cases. They showed that Fuller's estimator was more conservative than the unit-root test approach in assigning unit-roots but was less conservative than the OLS estimator.

2.4.2.2 Exactly MU Estimators

The first exactly MU estimator of the AR(1) coefficient was proposed by Hurwicz (1950). He observed that in Model 1, if the errors are normally distributed, then every ratio y_t/y_{t-1}, $t = 2, 3, \ldots, T$, is a MU estimator of ρ. In this case, each ratio has a Cauchy distribution, and is not an efficient estimator. Hurwicz (1950) conjectured that the median of the ratios, i.e.

$$\hat{\rho} = med(\frac{y_2}{y_1}, \frac{y_3}{y_2}, \ldots, \frac{y_T}{y_{T-1}}) \tag{2.52}$$

would be a more efficient estimator and perhaps also a median-unbiased one. This claim was partly verified by Zilelinski (1999). He showed that in Model 1, if the errors have median zero and $\Pr\{y_t = 0\} = 0$ for all t, then the estimator (2.52) is exactly median-unbiased for ρ. Boldin (1994) also examined the same estimator and established its asymptotic normality. But there is no efficiency comparison available in his paper between this estimator and the single ratio originally proposed by Hurwicz (1950). In fact, apart from median-unbiasedness, little is known about other properties of this estimator. Another serious drawback of these estimators is that they cannot be extended to models that contain an intercept and/or a time trend (Model 2 and 3). It is commonly agreed that Model 1 itself is not very useful in economic modelling (see discussions in Andrews (1993)).

Andrews (1993) proposed an exactly MU estimator for Models 1 – 3. His results were obtained by inverting the median function of the OLS estimator of ρ within the parameter space, i.e.,

$$\hat{\rho}_A = \begin{cases} 1 & \text{if } \hat{\rho}_{OLS} > m(1) \\ m^{-1}(\hat{\rho}_{OLS}) & \text{if } m(-1) < \hat{\rho}_{OLS} \le m(1), \\ -1 & \text{if } \hat{\rho}_{OLS} \le m(-1) \end{cases} \qquad (2.53)$$

where $m(\rho)$ is the unique median of $\hat{\rho}_{OLS}$ when ρ is the true parameter value. As the distribution of $\hat{\rho}_{OLS}$ is invariant to the nuisance parameters and the initial conditions in these models, this estimator was shown to be exactly median-unbiased for all sample sizes and all ρ values. It effectively corrected the serious downward biases of $\hat{\rho}_{OLS}$ especially for Model 2 and 3 and for small sample sizes. This estimator can be used to construct a MU estimator of the impulse response function, which is a monotonic function of ρ, and a MU model-selection procedure. Andrews (1993) also demonstrated the robustness of his MU estimator to the non-normal errors and other error misspecifications.

Compared with the asymptotic approach, Andrews' approach is an exact small-sample one. His estimator avoids using the asymptotic limits as an approximation of the small sample distributions. The median functions were shown

to be monotonic for all three models. However, Andrews' method requires the computation (or simulation) of the median function for every different sample size, which increases the computational cost of the procedure. Another drawback is that, as admitted in Andrews (1993) and Andrews and Chen (1994), it is not known if their MU estimators are optimal in any sense.

Recently, So and Shin (1999) proposed a Cauchy estimator to estimate Model 1 – 3 in small samples. For Model 1, the estimator is

$$\hat{\rho}_c = \sum_{t=2}^{T} sign(y_{t-1}) y_t \Big/ \sum_{t=2}^{T} |y_{t-1}|, \qquad (2.54)$$

where $sign(y_{t-1}) = 1$ if $y_{t-1} \ge 0$ and $= -1$ otherwise. This estimator has a history as long as the OLS estimation of Gauss. According to So and Chin, in 1836, Cauchy first considered such an estimator in the simple linear regression model. The attractive property of this estimator is its asymptotic normality, which is not possessed by the OLS estimator for $\rho = 1$. The estimator was also shown to be approximately median-unbiased for all values of ρ. Compared with Andrews' (1993) estimator, it does not require evaluation of the median function, hence is easier to compute. However, extending the use of this estimator to Model 2 and 3 is not a trivial exercise. It requires a specially designed algorithm of recursive mean adjustment and/or recursive detrending. This estimator also needs theoretical justification for its (approximate) median-unbiasedness. So and Shin (1999, 2000) applied the estimator in constructing tests for a unit root and seasonal unit roots.

2.4.2.3 Extensions to the AR(p) Model

Andrews' (1993) method of constructing MU estimators was extended to estimate an AR(p) model by Andrews and Chen (1994). A similar approach was adopted by Rudebusch (1992) in re-examining the Nelson-Plosser (1982) series. In both studies, iterative algorithms were developed to invert the median functions of the parameters one at a time, and with other parameters replaced by their estimates from the previous round when evaluating the median function. The process is repeated until convergence. As the median function of the estimator of each coefficient is not invariant to other coefficients, the final estimator will only be approximately median-unbiased. Andrews and Chen (1994) showed, however, that the approximation is good for moderate sample sizes. Fair (1996) theoretically outlined an algorithm essentially the same as the one suggested by Andrews and Chen but extended its use to the simultaneous equations model which contains endogenous variables and their lags. Based on the approximately MU estimates, both Rudebusch (1992) and Andrews and Chen (1994) reported that the persistence in the Nelson-Plosser series were seriously underestimated by the OLS estimators and overstated by the usual unit root tests. However, Fair (1996) concluded that the use of MU estimates did not improve the forecasting accuracy in the macroeconomic models he examined.

2.4.3 Other Applications

2.4.3.1 Time Varying Parameter Models

Stock and Watson (1998) developed some asymptotically MU estimators for the time-varying parameter model. A special case of the general model considered in their paper is the so-called 'local level' unobserved components model, which was also studied by Harvey (1985), Shepherd and Harvey (1990) and Shepherd (1993) among others. The model can be specified as,

$$y_t = \beta_t' x_t + u_t, \qquad \beta_t = \beta_{t-1} + v_t, \tag{2.55}$$

where y_t is the dependent variable (observed at time t), x_t is a $k \times 1$ vector of fixed regressors, β_t are the time-varying coefficients. For simplicity, we assume $u_t \sim IN(0, \sigma^2)$, $v_t \sim IN(0, \tau^2)$ and u_t and v_t are independent. The parameter of interest is the scale parameter τ. The ML estimators implemented by Kalman filter suffer from the undesirable property that if τ is small, it has the so-called pile-up

problem, i.e. the ML estimator has a probability mass at 0. This problem was also reported in Shepherd and Harvey (1990).

The asymptotic approach was taken by Stock and Watson (1998) by considering the nesting $\tau = \lambda/T$, which is very similar to the local-to-unity reparametrisation in the unit root models. Three test statistics testing the hypothesis $\lambda = 0$ were considered: Nybolm's (1986, 1989) locally most powerful invariant (LMPI) test L_t, the sequential GLS Chow statistic $F_t(s)$, and the point optimal invariant (POI) test suggested by Shively (1988) among others. As the F_t statistic is an empirical process, three scalar functionals were used: the Quandt (1960) maximum F_t (QLR), the mean Wald test (MW) and the exponential Wald test (EW) of Andrews and Ploberger (1994). The asymptotic distributions of these test statistics were derived. The median function of the limiting distribution of a test statistic is denoted by $m_D(\bullet)$, where D is the matrix of nuisance parameters. For example, based on one of the functionals of F_t, λ can be estimated by

$$\hat{\lambda}_g = m_D^{-1}(g(F_t)). \tag{2.56}$$

In practice, the unknown nuisance parameters in D have to be replaced by some consistent estimates, which do not alter the asymptotic distributions and therefore still ensure the asymptotic median-unbiasedness of $\hat{\lambda}$.

This estimator is subject to the same criticism as Stock's (1991) method, that the limiting distribution may not be a good approximation in small samples, and as a result, the estimator may not be MU when the sample size is small. The estimation procedure is also computationally cumbersome as it requires computing the inverse median function m_D^{-1} for every set of estimates $\hat{D}$. Stock and Watson (1998) conducted Monte Carlo studies to compare the pile-up probabilities and the asymptotic relative efficiencies of these MU estimators together with the ML estimator. It was found that the pile-up problem plaguing the ML estimator was properly controlled by all MU estimators for small τ and the MU estimators also have good asymptotic relative efficiencies for small to moderate amounts of parameter variability. Particularly, the MU estimators based on the QLR and the POI(17) test statistics (the POI test that maximises the power at $\lambda = 17$, see Saikkonen and Luukonen, 1993) were the best among the six estimators considered.

2.4.3.2 Binary Choice Models

Hirji, Tsiatis and Mehta (1989) developed a MU estimator for a logistic regression model with two binary covariates. The model relating the response of a patient to treatment and age, was used for assessment of the treatment effect while adjusting for the effect of age. It can be written as

$$\Pr\{Y_i = 1 | x_i\} = \{1 + \exp(-\beta' x_i)\}^{-1}, \tag{2.57}$$

where $\beta = (\beta_0, \beta_1, \beta_2)'$ and $x_i = (1, x_{1i}, x_{2i})'$. The parameter of interest is β_1, which is the relative log odds of response for treatment 1 versus treatment 2, and can be considered a measure of the magnitude of the treatment effect while controlling for the effect of age. The MU estimator is constructed based on the conditional distribution of the sufficient statistics for β. The vector of sufficient statistics for β is given by $T = X'y$ and if we write $T = (T_0, T_1, T_2)'$, the conditional distribution of T_1 can be worked out. The MU estimator of β is then computed by solving

$$\Pr\{T_1 \le t_1 | t_0, t_2, \beta_1\} = \tfrac{1}{2}. \tag{2.58}$$

This involves first evaluating the conditional distribution for each different set of β and numerically solving the above equation. An efficient binary-search procedure is needed to find the final estimates. Their Monte Carlo studies revealed that the MU estimator is uniformly more accurate than the ML estimator for small to moderately large sample sizes and a broad range of parameter values. As a result, Hirji, Tsiatis and Mehta recommended the use of the MU estimator as an alternative to the ML estimator especially when the sample size is not large or when the data structure is sparse.

2.5 Least Absolute Deviation Method

A class of estimators that are related to the MU estimators are the least absolute deviation (LAD) estimators. In this section, we briefly review the major results presented in the L1-norm literature, with focus given to the link between the LAD criteria and median-unbiasedness.

In a general regression model

$$y_t = f(x_t, \beta) + \varepsilon_t, \tag{2.62}$$

where y_t and x_t, $t = 1, \ldots, T$ are the observed data of the dependent and exogenous variables, respectively, $f(\bullet)$ is a known continuous function and ε is the error vector specified by some known distribution function or moment conditions. The LAD estimator of the regression coefficient β is then the solution to the minimisation problem,

$$\min_{\beta \in \Omega} \sum_{t=1}^{T} |y_t - f(x_t, \beta)|. \tag{2.63}$$

According to Rao (1988), the LAD method dates back to Laplace and Gauss. Unfortunately, its applications were restricted by computational difficulties and lack of asymptotic studies, until in the last two decades, when the breakthrough in the computing technology and the development of asymptotic theory prompted a great deal of interest in LAD estimators. See Amemiya (1985) or Rao (1988) for a review and Narula and Wellington (1982) for a survey in the context of regression models. The LAD estimator is a special case of a general class of robust methods based on minimising an expression of the type

$$\sum_{1}^{T} \rho(y_t, f(x_t, \beta)), \tag{2.64}$$

where $\rho(\bullet)$ is a properly chosen convex loss function; see Ronchetti (1982) and Huber (1987) for reviews of these robust methods.

From the definition of median-unbiasedness given in (2.10) and (2.11), it is clear that the univariate median is a LAD estimator of the location parameter of a univariate population. But a LAD estimator is in general not a MU (or a mean-unbiased) estimator. As an example, we examine the link between these two criteria in model (2.62) when $f(\bullet)$ is linear.

As pointed out by Andrews (1986) and Rao (1988), the LAD estimator $\hat{\beta}$ may not be unique. Fisher (1985) and Rao (1988) suggested a method of picking a unique value of $\hat{\beta}$. Under regularity conditions (e.g., Bai et al., 1988), the LAD estimator is shown by Bassett and Koenker (1978) and Bai et al. (1988) among others, to be asymptotically normal, with the limiting distribution given by,

$$2\sqrt{T} f(0) S_T^{1/2}(\hat{\beta}-\beta) \rightarrow N(0, I_m), \tag{2.65}$$

where $f(0)$ is the density function of the errors ε_t evaluated at zero, while the matrix S_T is given by,

$$S_T = \tfrac{1}{T}(x_1 x_1' + \ldots + x_T x_T'). \tag{2.66}$$

Therefore the LAD estimator $\hat{\beta}$ is asymptotically median-unbiased under fairly general conditions. Angelis et al. (1993) discussed some analytical and bootstrap approximations to the distribution of the LAD estimator in finite samples. Hypothesis testing procedures analogous to those based on the least squares estimator are developed by Koenker and Bassett (1982), namely the LR, Wald and LM type tests. They found that the tests based on the LAD estimator have the usual chi-squared limiting distribution but are generally less efficient than their classic counterparts when the errors are normally distributed.

Provided that the LAD estimator $\hat{\beta}$ is unique (e.g, pick a unique estimate using the method suggested by Fisher (1985) and Rao (1988)), and if the errors ε_i are symmetrically distributed, it can be shown that the distribution of $\hat{\beta}$ is symmetrical, using the general results developed by Andrews (1986). His results are based on the observation that if a random variable is an odd function of some symmetrically distributed errors, it also has a symmetric distribution. Therefore for any estimator $\tilde{\beta}$ that maximises an objective function $r(y-f(X,\beta))$, where $X=(x_1, x_2, \ldots, x_T)'$ as in model (2.62), and if $r(\bullet)$ is an even function of $y-f(X,\beta)$, $\tilde{\beta}$ will possess the property that $\tilde{\beta}-\beta_0$ is an odd function of the error vector ε, therefore has a symmetric distribution function. The LAD estimator satisfies these criteria if it is uniquely defined. Therefore in the usual Gauss-Markov set up, LAD estimators are both mean-unbiased and median-unbiased.

Apart from the linear regression model, LAD estimation has also been used in many other estimation situations. For example, Amemiya (1982) proposed a class of two stage LAD estimators for the estimation of the parameters of a structural equation in the simultaneous equations model. The performance of LAD estimators was compared with least squares estimators in the simultaneous equations model by Glahe and Hunt (1970). Powell (1984) suggested using the LAD method in the censored regression ('Tobit') model as an alternative to the MLE estimator. The estimator was found to be robust to heteroscedastic errors. A 'trimmed' LAD estimator for the Tobit model was proposed by Honore (1992). LAD estimation was also used in nonlinear dynamic models with neither independent nor identically distributed errors by Weiss (1991). Recently, Bai (1995) applied the LAD approach to the estimation of a shift in linear regressions.

An important example of linking LAD estimation to MU estimation is provided by Campbell and Honore (1993). They considered a panel data censored regression model (so-called a dynamic Tobit model) with individual specific fixed effects, which can be written as

$$y_{tk} = \max\{0, \alpha_k + x_{tk}\beta + \varepsilon_{tk}\},\ k = 1,\ 2 \text{ and } t = 1, \ldots, T. \tag{2.59}$$

Honore (1992) proposed a split sample LAD estimator by minimising

$$Q_T(b) = \sum_{t=1}^{T} q(y_{t1}, y_{t2}, (x_{t1} - x_{t2})b), \tag{2.60}$$

where

$$q(y_1, y_2, \delta) = \begin{cases} 0 & \text{if } \delta \le -y_2,\ y_1 = 0 \\ g(y_1) - (y_2 + \delta)g^-(y_1) & \text{if } \delta \le -y_2,\ y_1 > 0 \\ g(y_1 - y_2 - \delta) & \text{if } -y_2 < \delta < y_1 \\ g(-y_2) - (\delta - y_1)g^+(-y_2) & \text{if } \delta \ge y_1,\ y_2 > 0 \\ 0 & \text{if } \delta \ge y_1,\ y_2 = 0 \end{cases}, \tag{2.61}$$

and $g(\bullet)$ is an even (symmetric) convex loss function with right derivative g^+, left derivative g^- and $g(0) = 0$. Particularly, if we let $g(d) = |d|$, the above estimator becomes the so-called 'trimmed' LAD estimator. Campbell and Honore (1993) showed that if only one parameter is estimated, the proposed estimator is median-unbiased. This result can be obtained even though the estimator is not symmetrically distributed.

Although the asymptotic properties of LAD estimators have been established for most of the above reviewed applications, little is known about their finite sample properties. In particular, not much effort has been made to examine the unbiasedness of these LAD estimators, as in most cases, the attention is usually given to their robustness to outliers and data contamination, and their relative efficiencies to least squares counterparts.

2.6 Bias-reduction Techniques in Econometrics

Bias is frequently encountered by econometricians in many estimation situations. As median-unbiasedness is one of several definitions of unbiasedness, the pursuing of MU estimation belongs to the broad area of unbiased estimation, in which bias-reduction plays a central role. In this section, we review some of these techniques.

The econometric literature on bias-correction is vast. Therefore it is impossible for us to provide an exhaustive review. Our main interest is to outline the generally applicable approaches towards bias correction mainly in small samples, as these methods may lend us some guidelines and techniques that can be used in our search for MU estimators. Most of these studies are concerned with bias-correction towards mean-unbiasedness (as far as we know, MacKinnon and Smith (1998) is the only study that directly refers to median-bias correction). However, we would expect that the same principles should apply to bias-correction in the context of median-unbiasedness.

2.6.1 Analytical and Bootstrap Bias-correction

We start our discussions with maximum likelihood estimators (MLEs). It is well known that MLEs are often biased in finite samples. This bias may come from two sources,

1). The non-linearity or curvature of the score function. Box (1971) attempts to quantitatively assess these biases. A measure closely related to Beale's (1960) measures of nonlinearity was developed to link the curvature of the estimation problem and the bias to the MLE estimator. Efron (1975) further elaborated on the concept of curvature of a statistical problem. Cook, Tsai and Wei (1986) also explored the relationship between bias and curvature in the context of nonlinear regression.

2). The effect of the nuisance parameters. MLE procedures usually involve eliminating nuisance parameters by replacing them with estimates. Although, asymptotically, it does not alter the properties of the estimator, it may bring bias into the score and lead to small sample bias in the estimator.

Many studies are committed to reducing bias in a MLE. These efforts can be classified into two categories: bias-correction and bias-prevention. The first approach involves removing the bias after the initial estimate is calculated, while the latter attempts to prevent the bias beforehand. In the first category, Cox and Hinkley

(1974) pointed out the two frequently used bias-correction methods: 1. analytical correction, i.e., subtracting the (estimated or approximated) bias function from the original estimator, and 2. correction via resampling schemes such as the jackknife or bootstrap.

For the analytical approach towards bias-correction, knowledge about the bias function is essential. The bias function of an estimator $\hat{\theta}$ is defined as

$$B(\theta) = E(\hat{\theta}) - \theta, \tag{2.67}$$

and following Ferrari and Criberi-Neto (1998), under mild regularity conditions, the bias function can be written as,

$$B(\theta) = \frac{B_1(\theta)}{T} + \frac{B_2(\theta)}{T^2} + \cdots, \tag{2.68}$$

where $B_1(\theta)$, $B_2(\theta)$, in a maximum likelihood estimation set up, are functions of cumulants of log-likelihood derivatives with respect to θ for a single observation. For example, Cox and Hinkley (1974) gave a general formula for $B_1(\theta)$,

$$B_1(\theta) = \frac{k_{11}(\theta) + k_{30}(\theta)}{2(i(\theta))^2}, \tag{2.69}$$

where

$$k_{11}(\theta) = E[U(\theta)\{U'(\theta) + i(\theta)\}],$$
$$k_{30}(\theta) = E[\{U(\theta)\}^3],$$
$$U(\theta) = \partial \log f(x|\theta)/\partial\theta,$$
$$i(\theta) = -E[U'(\theta)].$$

Mardia, Southworth and Taylor (1999) gave a simplified expression,

$$B_1(\theta) = \frac{1}{2i(\theta)^2}[2E\{U(\theta)U'(\theta)\} + E\{U''(\theta)\}]. \tag{2.70}$$

Formulas for bias of higher orders were also developed via asymptotic expansions in many specific models: e.g., just to name a few, in the dynamic linear regression (Kiviet and Philipps, 1993, 1994, 1996), in generalised linear models (Cordeiro and McCullagh, 1991), in ARMA models (Cordeiro and Klein, 1994) and in the one-parameter exponential family (Ferrari et al., 1998).

All these efforts lead to a natural way of correcting the bias in the MLE, namely, subtracting the approximated bias from the original estimator. Ferrari et al. (1996, 1998) showed that in order to get an estimator bias-free to order T^{-1}, we can use the corrected estimator given by

$$\tilde{\theta}_1 = \hat{\theta} - B_1(\hat{\theta})/T, \tag{2.71}$$

where $B_1(\theta)$ can be replaced by the dominant term in (2.69) or (2.70). If the bias is to be removed to order n^{-2}, the following correction is needed,

$$\tilde{\theta}_2 = \hat{\theta} - \frac{B_1(\hat{\theta})}{T} - \frac{B_2^*(\hat{\theta})}{T^2}, \tag{2.72}$$

where

$$B_2^*(\theta) = B_2(\theta) - B_1(\theta)B_1'(\theta) - \tfrac{1}{2}V_1(\theta)B_1''(\theta), \tag{2.73}$$

in which $V_1(\theta)$ is the first term of the expansions of the variance of the MLE (similar to the expansion in (2.68)), while $B_2(\theta)$ is the second term in (2.68), of which $B_1'(\theta)$ and $B_1''(\theta)$ are its first and second derivative, respectively. These terms are usually functions of cumulants of log-likelihood derivatives with respect to θ for a single observation.

Ferrari et al. (1998) commented that the analytical bias-correction based on studying the form of the bias function entails a great deal of algebra but has the nice feature that the final expressions are usually simple enough that they can be easily used by practitioners. Therefore it dominated the bias-correction literature for decades until the bootstrap methods became prominent.

The bootstrap was originally proposed by Efron (1979), and research in the theoretical development and empirical applications of the method has flourished since then. A number of review articles and books have appeared in the past decade, see, e.g. Efron and Tibshirani (1993), Hall (1992, 1994), Gonzalez et al. (1994), Shao and Tu (1995) and Davidson and Hinkley (1996). Bias-correction is also an important theme in the bootstrap literature. Much effort has been focused on correcting bias in the bootstrap confidence intervals, see, e.g. Efron (1987) and Efron and Tibshirani (1986). Bias correction for point estimators was addressed quite extensively in Hall (1992). It was recently discussed more systematically by MacKinnon and Smith (1998) and was compared with the analytical approach by Ferrari et al. (1998). The essence of the bootstrap bias-correction involves, instead of explicitly working out the bias function, using resampling schemes to achieve the same correction implicitly.

Following the notation in Hall (1992), the parameter of interest is written as the functional $\theta = \theta(F_0)$, where F_0 stands for the true population distribution. The MLE of θ is then denoted by $\hat{\theta} = \theta(F_1)$ as it is based on a sample $Y_1, \ldots, Y_n$ drawn from F_0 with θ as its true parameter. The MLE of θ from a sample (called a bootstrap sample) generated from F_0 with $\theta = \hat{\theta}$ is denoted by $\theta^* = \theta(F_2)$ and similarly the MLE of θ from a sample (called a double bootstrap sample) with $\theta = \theta^*$ is denoted by $\theta^{**} = \theta(F_3)$. The bias-corrected estimator after the first round of bootstrapping is given by,

$$\hat{\theta}_1 = 2\hat{\theta} - E\{\theta^* | F_1\}, \tag{2.74}$$

while the bias-corrected estimator after a double bootstrap is given by,

$$\hat{\theta}_2 = 3\hat{\theta} - 3E\{\theta^* | F_1\} + E\{\theta^{**} | F_1\}. \tag{2.75}$$

Hall (1988, 1994) and Ferrari et al. (1998) all showed that the bootstrap corrected estimators (2.74) and (2.75) achieved the same order (T^{-1} and T^{-2} respectively) of accuracy as the analytically corrected estimators in (2.72) and (2.73). They argued that the bootstrap approach avoids the need for messy algebraic derivations.

MacKinnon and Smith (1998) discussed the above method in a more practical setting by examining the accuracy of the correction for different forms of bias function. It is noteworthy that they concluded that reducing bias may increase the variance, or even the mean squared error of an estimator. Whether it does so depends on the shape of the bias function. The bias correction was extended to multiparameter estimation problems in both Ferrari et al. (1998) and MacKinnon and Smith (1998).

2.6.2 Bias-prevention Methods

Instead of correcting the bias after the estimator has been calculated, different methods have been proposed to prevent the bias beforehand. In the MLE context, this is usually done by adjusting the likelihood or the score function. If the original profile log-likelihood function is $l_p(\theta)$, these adjustments replace it with a new objective function,

$$l_{ap}(\theta) = l_p(\theta) + r(\theta), \tag{2.76}$$

for a suitably chosen additive adjustment function $r(\theta)$. Various researchers, including Bartlett (1955), Barndorff-Nielsen (1983, 1994), Barndorff-Nielsen and Cox (1984), Cox and Reid (1987, 1993), McCullagh and Tibishirani (1990), DiCiccio and Stern (1993) and Stern (1997), have suggested specific adjustment functions to the profile log-likelihood which have the effect of reducing the score bias. The properties of these adjustments were discussed further by Liang (1987), Levin and Kong (1990), Cox and Reid (1993), Ghosh and Mukerjee (1994) and Stern (1997). We review the two prominent examples of such adjustments, the conditional profile likelihood of Cox and Reid (1987) and the modified profile likelihood of Barndorff-Nielsen (1986).

Assume θ is vector-valued and it can be partitioned into $\theta' = (\psi', \lambda')$, where the subvectors ψ and λ are the parameters of interest and nuisance parameter respectively. Let $\hat{\theta}' = (\hat{\psi}', \hat{\lambda}')$ be the overall ML estimate and let $\hat{\psi}_\lambda$ be the ML estimate of ψ with λ fixed and similarly $\hat{\lambda}_\psi$. If the parameter of interest ψ is a scalar, the conditional profile likelihood (Cox and Reid, 1987) is defined by,

$$l_{cp}(\psi) = l(\psi, \hat{\lambda}_\psi) - \tfrac{1}{2}\log\{\det nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)\} \qquad (2.77)$$

where $j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)$ is the observed information per observation for the λ components. This definition requires ψ and λ to be orthogonal in the sense defined by Cox and Reid, i.e., $E(-\partial^2 l/\partial\psi\partial\lambda_i) = 0$ for all λ_i. As observed by McCullagh and Tibshirani (1990), the interpretation of the correction term is that it penalises values of ψ for which the information about λ is relatively large. Application of this adjustment will typically require an initial reparameterisation of the nuisance parameters or orthogonality. But unfortunately, as criticised by Stern (1997), such a reparametrization can only be guaranteed to exist when ψ is a scalar.

The modified profile likelihood (Barndorff-Nielsen, 1986) is defined by

$$l_{mp}(\psi) = l(\psi, \hat{\lambda}_\psi) - \tfrac{1}{2}\log\{\det nj_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) + \log\{\det(d\hat{\lambda}_\psi / d\hat{\lambda})\}. \qquad (2.78)$$

This definition does not require the orthogonality of ψ and λ; the last term on the right-hand side can be thought of as a correction for non-orthogonality. In this term, $\hat{\lambda}_\psi$ is regarded as a function of $\hat{\lambda}$ and the ancillary statistic. Thus, application of the modified profile likelihood requires explicit knowledge of an ancillary statistic, which may be difficult to obtain in practice. When ψ is a scalar, Barndorff-Nielsen (1994) developed an approximation to the above adjustment which does not require explicit knowledge of an ancillary statistic.

Apart from the conditional and the modified profile likelihood, there are several other modifications to the likelihood function, such as the use of marginal likelihood and expected likelihood. For a comprehensive review of these modifications, see Laskar and King (1997, 1998). All these efforts are aimed to nullify the effects of unknown nuisance parameters. Most of the applications of these modified likelihoods are aimed at improving the small sample performance of test procedures (particularly the LR and LM tests), see for example, Bellhouse (1978), Cruddas et al. (1989), Tunnicliffe-Wilson (1989) and Laskar and King (1998). Less attention has been given to comparing the small sample performance of the point estimators that maximise various modified likelihood functions.

As an important alternative to the above two modifications, McCullagh and Tibshirani (1990) attempted to adjust the profile likelihood score function instead of the likelihood itself. If the original score is $U(\psi)$, the aim of their adjustment was to force the following two conditions onto the adjusted score given by

$$U^*(\psi) = \{U(\psi) - m(\psi)\}w(\psi), \qquad (2.79)$$

such that

$$E_{\psi,\hat{\lambda}_\psi}[U^*(\psi)] = 0 \qquad (2.80)$$

and

$$\mathrm{var}_{\psi,\hat{\lambda}_\psi}[U^*(\psi)] = -E_{\psi,\hat{\lambda}_\psi}[\partial U^*(\psi)/\partial\psi]. \qquad (2.81)$$

The solutions to the above two conditions are

$$m(\psi) = E_{\psi,\hat{\lambda}_\psi}[U(\psi)] \qquad (2.82)$$

and

$$w(\psi)=\{\mathrm{var}_{\psi,\hat{\lambda}_\psi}[U(\psi)]\}^{-1}\{-E_{\psi,\hat{\lambda}_\psi}[\partial^2 l_p(\psi)/\partial\psi^2]+\partial m(\psi)/\partial\psi\}. \quad (2.83)$$

Therefore the nature of the adjustment is to subtract its expectation from the original score and to impose a second derivative term to make sure the information unbiasedness of the adjusted score is held. This adjustment helps to alleviate the bias problems inherent in the use of the profile likelihood.

The justification of concentrating on the score rather than the likelihood is found in the theory of optimal estimating equations. The score functions are examples of estimating functions defined by Godambe (1960, 1997). As remarked in McCullagh and Tibshirani (1990, p341), under fairly mild regularity conditions, unbiasedness of the estimating equation essentially guarantees consistency, while the condition of information unbiasedness ensures asymptotic optimality within a class of estimating functions, see also Godambe and Thompson (1974) and Godambe (1997). And there is a strong link between the optimality of an estimating equation and the optimality of the corresponding estimator.

McCullaugh and Tibshirani (1990) argued that centring the score function should improve estimation accuracy. This claim was illustrated empirically by Mahmood (2000). He simply subtracted the expectation from the original score without imposing the second order term penalty. It was shown that this adjusted score is able to deliver estimates that are less biased than those based on the original profile likelihood score in a dynamic linear regression model. As the score of a profile log-likelihood is usually biased (i.e., its expectation at the true parameter value is not zero), both McCullaugh and Tibishirani (1990) and Mahmood (2000) advocated adjusting the score function towards unbiasedness. Mahmood (2000) showed that an unbiased score equation can reduce the bias in an estimator from solving a biased score equation. But the extent of the bias-prevention differs from model to model. In the dynamic linear regression, for example, the bias correction is not completely satisfactory, the solution of the adjusted score equation can still be biased in small samples.

Another adjustment to the score function that more directly addresses bias-reduction in the ML estimator was suggested by Firth (1993). This adjustment is based on a 'proper' score function, which satisfies the unbiasedness and information unbiasedness conditions. This differs from the scores of the profile log-likelihood considered in the previous studies that are usually biased. Firth's argument is that although the score is unbiased, the resulting estimator is not necessarily unbiased. Therefore in order to remove the $o(n^{-1})$ bias from the estimator, an appropriate bias term is subtracted from the score function. If the original score is $U(\theta)$ for a scalar parameter θ, the adjusted score is given by

$$U^*(\theta)=U(\theta)-i(\theta)b(\theta), \quad (2.84)$$

where $i(\theta)=-U'(\theta)$ is the local gradient of the score and $b(\theta)$ is the bias function of the MLE estimator $\hat{\theta}$, which is usually approximated by its n^{-1} term. Firth (1993) applied this method in several generalised linear model estimation problems. However, this method is subject to the same criticism that it requires the knowledge about the bias function $b(\theta)$, which in many cases are not readily available. Another criticism is that this method is not easy to be extended to a multi-parameter case.

Finally, we notice that Lele (1991) considered resampling the estimating equations instead of the estimates by jackknife to improve the quality of the estimator. Let

$$G(X,\theta)=\sum_{t=1}^{T} g_t(X,\theta)=0 \quad (2.85)$$

be the original estimating equation and $\hat{\theta}_n$ be the estimate from it. Lele's method consists of deleting one estimating equation at a time the thus obtaining the pseudo-values. To see this, let

$$G^j(X,\theta)=\sum_{t\neq j} g_t(X,\theta)=0 \quad (2.86)$$

and $\hat{\theta}_{T,-j}$ be the estimate therefrom. The jackknife estimate of θ is then given by

$$\hat{\theta}_T^{JK} = \hat{\theta}_T - \frac{T-1}{T}\sum_j (\hat{\theta}_{T,-j} - \hat{\theta}_T). \tag{2.87}$$

Under certain regularity conditions, Lele (1991) showed that this estimator is weakly consistent and asymptotically normal. This method is attractive when the data are correlated observations. In these situations, the usual jackknife has to remove data segments from a serially correlated sequence of data causing difficulties. Jacknifing estimating equations, however can avoid this problem as the information in x_k is used conditionally but not unconditionally. But the small sample properties (including unbiasedness) of this estimator remain unclear as no empirical studies have been conducted to examine its small sample performance.

2.7 Concluding Remarks

This chapter reviews the literature related to MU estimation and bias-reduction methods. The attractive features of MU estimators compared with those of mean-unbiased estimators were highlighted. The applications of MU estimators in different econometric models were surveyed. In a broad context, some important bias-correction and bias-prevention techniques were also discussed. The review exposed some research questions and voids, treatment of which will form the main theme of this thesis. In what follows, we provide our observations on a few key aspects which may require further attention.

1. There seems to be a lack of a systematic approach towards constructing MU estimators. The examples of these estimators in the literature usually failed to provide generally applicable guidelines on how to find MU estimators for a given model. The clue to solving this may lie in two aspects. First, can we borrow the existing bias-reduction techniques that are designed to achieve mean-unbiasedness, and modify them to construct MU estimators? Second, much effort has been put into constructing confidence intervals by inverting the critical value

functions of a test statistic. Can we apply the same method to constructing a MU estimator? What are the problems that may arise when point estimation is the purpose of inverting a test?

2. Most of the applications of MU estimation are concerned with estimating the first order autoregressive model. Can we extend the model to include exogenous regressors? As in practice, most models will contain explanatory variables and this usually causes the properties of the inference procedures to depend on the structure of the design matrix. It would be useful if MU estimation could be applied to these more general models, such as the linear regression with autoregressive disturbances, and the dynamic linear regression model, which are two of the most frequently used time series models in econometrics.

3. So far the applications of MU estimation are mainly for a scalar parameter. Can we construct MU estimators when nuisance parameters are present? It would be useful to apply the existing techniques to account for nuisance parameters, such as marginal likelihood methods and invariant tests, to the MU estimation of the parameters of interest.

4. Not many researchers have applied MU estimators in other inference procedures such as hypothesis testing and forecasting. Most studies stopped at examining the properties of these estimators. Are these estimators able to improve the small sample performance of hypothesis testing and forecasting? More importantly can they correct the deficiencies plaguing some of these procedures that are possibly caused by small sample bias in the estimators?

These are some of the open research questions that emerge from our literature review. The remaining chapters of this thesis consider each of these themes according to the layout detailed in Chapter 1.

Chapter 3

Some General Methods for Constructing Median-unbiased Estimators: Theory [1]

3.1 Introduction

Median-unbiasedness is an important alternative to mean-unbiasedness when assessing the impartiality of an estimator. More importantly, it can enjoy some indispensable advantages in some situations when mean-unbiasedness is not achievable or robustness is highly desirable. However, as seen from Chapter 2, median-unbiased (MU) estimators have not been widely used in econometrics apart from a few well-known examples. Part of the reason for this is that they are not as easily found as mean-unbiased estimators. There seems to be a lack of systematic methods for constructing MU estimators in the literature. Unlike mean-unbiasedness, with separate books devoted to how to construct mean-unbiased estimators in different models (e.g., Voinov and Nikulin, 1993a, 1993b), there are only a few applications of MU estimators scattered in the literature, while there is not much guidance of any generally applicable approach towards MU estimation.

In this chapter, we attempt to establish a theoretical framework for constructing a MU estimator. Two general methods are developed and formalised, which are applicable in different parametric models. Although a few applications of the second method can be found, the theoretical discussion and formalisation (especially the application of estimating equations and the optimal invariant tests) is our contribution to the literature. In particular, we attempt to address some of the problems existing with the current examples in the literature, which may explain why MU estimators have not been used more widely.

[1] Some of the material contained in this chapter was published in a conference proceedings, see Chen and King (1998). It was also presented at the Annual Meeting of the Econometric Society, Sydney, June 1998.

Many results in this chapter can be traced back to a lemma developed by Lehmann (1959) and Pfanzagl's work (1970, 1971, 1979), which explored the link between the existence of an optimal MU estimator in the exponential family and the distribution function of the sufficient statistics. However, the use of sufficient statistics is not the recommended approach in this thesis, as the original form of this approach is usually quite hard to implement. Instead, our main interest in this chapter is to develop and formalise two of its more practical derivatives:

1. Adjusting estimating equations to correct median bias, or

2. Inverting the median function of a test statistic at the 50% significance level for a MU estimator.

The first method parallels the bias-reduction and bias-prevention techniques for mean-unbiased estimation. But it does not require the knowledge of the derivative of the estimating function (such as the second derivatives of the likelihood function) nor does it require the exact or approximate form of the bias function. The latter is usually needed in most analytical mean-bias correction methods. We also discuss the link between the proposed method and an analytical and a bootstrap bias-reduction technique. An iterative algorithm to solve the adjusted estimating equations is developed when the exact solution is not available. The method is also extended to the multi-parameter case, in which we need to adjust the equations recursively while replacing the unknown parameters in the equations by their estimates from the previous replication.

The second method is a special case of the familiar duality of a confidence interval and a significance test. But it is not as trivial as expected when test inversion is applied to point estimation. A new problem that is not associated with interval estimation arises in point estimation, namely, non-unique estimates due to a non-monotonic median function of the test. This problem is not properly addressed in most of the existing applications of MU estimation. We define some conditions for this method to work and also discuss the importance of selecting a 'good' test to invert in order to get a good point estimate. In particular, the use of two classes of optimal invariant tests is considered. If inverting the median function of a single test

statistic is not applicable, we propose a 'gird inversion' method as the remedy, which is more likely to produce reliable estimates.

The chapter is organised as follows: Section 3.2 sets out the fundamental results of Lehmann and Pfanzagl, and some reas why this original method is rarely used directly in practice. Section 3.3 shows that a MU estimating equation is more likely to produce a MU estimator compared with mean-unbiased counterparts. A generally applicable adjustment to the original estimating equation is defined. The link between this adjustment and the existing bias-reduction methods is disclosed. Section 3.4 discusses the general approach of inverting the median function of a test statistic to compute MU estimators. In particular, it addresses the issue of which test we should invert and how to invert. It is shown that grid inversion rather than fixed-point inversion is more likely to produce accurate estimates. The chapter ends with some concluding remarks in Section 3.5.

3.2 MU Estimators Based on Sufficient Statistics

The importance of sufficient statistics in test construction is well known. If a UMP test exists, it is usually a function of the sufficient statistics. As a special case, MU estimators can also be constructed based on the conditional distribution function of sufficient statistics in the family of distributions with monotone likelihood ratios. We first review the results due to Lehmann (1959, Corollary 3, p80 and p83):

Lemma 3.2.1 Let the family of densities $p_\theta(x)$, $\theta \in \Omega$ have monotone likelihood ratio in $T(x)$ and suppose that the cumulative distribution function $F_\theta(t)$ of the sufficient statistic $T = T(x)$ is a continuous function of t for each fixed θ, then

(i) if x denotes the observed values of X and $t = T(x)$, and if the equation $F_\theta(t) = \frac{1}{2}$ has a solution $\theta = \hat{\theta}$ in Ω, this solution must be unique and it is a MU estimator of θ;

(ii) among all MU estimates, $\hat{\theta}$ is optimal in the sense that it minimises $EL(\theta,\hat{\theta})$ for any nondecreasing loss function L.

Based on Lehmann's results, Pfanzagal (1970, 1971 and 1979) further specified the conditions of the existence of such an optimal MU estimator in the exponential family and also defined its asymptotic properties. His results are summarised in Lemma 3.2.2:

Lemma 3.2.2 For exponential families with density function of the form

$$f(x) = C(\theta,\eta)h(x)\exp[a(\theta)T(x) + \sum_1^p a_i(\theta,\eta)S_i(x)],$$

with $(\theta,\eta) \subset \Theta \times H$, $\Theta \subset R$, $a(\bullet)$ increasing and continuous, there exists a MU estimator $\hat{\theta}$, and

(i) it is of minimal risk for any monotone loss function in the class of all MU estimators;

(ii) under regularity conditions (Pfanzagal, 1979), $\hat{\theta}$ is asymptotically normally distributed.

These two lemmas not only define the existence of a MU estimator when the distribution function of the sufficient statistics satisfies certain conditions, but also provide some quite strong optimality results. Andrews and Phillips (1987) applied this optimality property to prove that the generalised least squares estimators are optimal MU estimators of the linear regression coefficients for all bounded loss functions.

The proof of these lemmas provided by Lehmann (1959) and Pfangzal (1979) also implicitly pointed out a natural way of constructing MU estimators based on the distribution function of the sufficient statistics. This was illustrated more clearly in Birnbaum (1961, 1964) and Read (1985). We state it here as a corollary to the above lemmas:

Corollary 3.2.1 If a sufficient statistic $T(X)$ (scalar) exists such that the values of $\hat{\theta}$ vary monotonically with $T(X)$, then for any observed value of $T = t$, $\hat{\theta}$ that satisfies

$$\Pr\{T \le t | \theta = \hat{\theta}\} \ge 1/2 \text{ and } \Pr\{T \ge t | \theta = \hat{\theta}\} \le 1/2$$

is a MU estimator of θ.

Although seemingly straightforward, this method is rarely used in its original form in practice. The only examples we could find in the literature are: estimating the binomial and Poisson distributions (Birnbaum, 1964) and estimating the logit regression model (Hirji et al., 1989). There are a few reasons why this method has not been popular in practice:

1. For a complicated model, in which nuisance parameters exist, the classic sufficient statistic for a single parameter may be difficult to define. We may need to rely on other definitions of sufficiency, such as L-sufficiency or G-sufficiency (see Ara, 1995 for a review). Just as UMP tests are rarely available, a sufficient statistic that is monotonic is hard to find except for a few simple models. Therefore, this direct approach is rarely available, and we have to look for less optimal statistics to construct MU estimators.

2. Even if the sufficient statistic is well defined as in the example of Hirji et al. (1989), its conditional distribution function is usually non-standard, not computable or even hard to simulate. We not only need its conditional distribution, but also its conditional median function. So when searching for the estimates that satisfy the two inequalities in Corollary 3.2.1, the computational burden is heavy.

3. The requirement of co-monotonicity of $\hat{\theta}$ and sufficient statistics $T(X)$ is usually not easy to satisfy and hard to verify. Because of this, it is difficult to control the practical problem of multiple solutions or empty solution when this monotonicity condition is violated.

In line with the idea of exploring information contained in the sufficient statistics, we develop two methods that are more practical to use and deliver more reliable estimates. Instead of inverting the conditional median function of a sufficient statistic itself, we study those of estimating functions and significance tests. Intuitively, good significance tests and estimating functions should explore the information contained in the sufficient statistics. Ideally, they should be functions of the minimal sufficient statistics where available. Their distribution properties are usually easier to analyse than the sufficient statistics themselves. This could greatly simplify the proposed estimation procedures.

3.3 Adjusting Estimating Equations for MU Estimators

3.3.1 Estimating Equations

Most procedures for point estimation of an unknown real scalar parameter θ can be viewed as solving an equation of the form

$$Q(\theta; y) = 0, \tag{3.1}$$

Q being a real function (which is sometimes called an *estimating function*) with arguments θ and the observed value of the corresponding random variable y. The equation (3.1) is then called an *estimating equation*.

The formal definition of an estimating equation is due to Godambe (1960). Commonly used estimating equations include the normal equations in least squares (LS) estimation, the score equations in maximum likelihood (ML) estimation, and the conditional moment conditions in generalised method of moments (GMM) estimation. The concept is also used heavily in the indirect inference literature, such as simulated methods of moments and empirical likelihood methods. The properties of estimating equations and their impact on the quality of the resulting estimators have been studied by Godambe (1976, 1980, 1984, 1985), Godambe and Thompson

(1974, 1984), Ferreira (1982) and Crowder (1987) among others. Liang (1987) and Godambe (1997) provided surveys on the estimating equation methodology. Interestingly, Vinod (1997) pointed out that the 'main lesson' from the estimating equation theory is to deemphasize the estimates (roots) and focus on the underlying equations.

An estimating equation is said to be unbiased (mean-unbiased) if it satisfies the condition:

$$E_{\theta_0}(Q(\theta; y)) = 0, \tag{3.2}$$

where θ_0 is the true value of the parameter. Here E_{θ_0} indicates that the expectation is taken with θ_0 treated as the true parameter underlying the data generating process. Most commonly used estimating equations (such as the score functions and the conditional moment conditions) are usually unbiased. But the unbiasedness of the estimating equations does not generally lead to the solution to the equation being necessarily an unbiased estimator. Therefore although asymptotically unbiased under fairly general conditions (see Godambe and Thompson (1974)), the LS, ML or GMM estimators are not uncommonly biased in small samples.

As a very special case, Durbin (1960) and Lieberman (1998) both considered estimating equations linear in θ, i.e.,

$$Q(\theta; y) = T_1\theta - T_2, \tag{3.3}$$

where T_1 and T_2 are two random functions of the data. Lieberman (1998) showed that if this estimating equation is unbiased, we need an extra condition which states that T_2/T_1 and T_1 are uncorrelated, in order to get a mean-unbiased estimator by solving (3.3).

A good counter example is the first order autoregressive model given by $y_t = \rho y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim IN(0, \sigma^2)$. Now the normal equation for the LS estimation of ρ is given by,

$$Q(\rho;y)=\sum_{t=2}^{T}y_t y_{t-1}-\rho\sum_{t=2}^{T}y_{t-1}^2 \qquad (3.4)$$

which is linear in ρ and also unbiased, i.e.

$$E_{\rho_0}[\sum_{t=2}^{T}y_t y_{t-1}-\rho\sum_{t=2}^{T}y_{t-1}^2]=0. \qquad (3.5)$$

But because $\sum y_t y_{t-1}/\sum y_{t-1}^2$ and $\sum y_{t-1}^2$ are obviously correlated, so as it is well known, the resulting estimator $\hat{\rho}$ is not mean-unbiased in small samples. From this example, it can be seen that except for a few very simple cases, even linear mean-unbiased estimating equations do not directly produce mean-unbiased estimators.

More importantly, $Q(\bullet;y)$ in most cases are nonlinear functions of θ and this almost certainly leads to bias in the estimator $\hat{\theta}$, at least in small samples. This is partly due to the simple fact that the expectation of a nonlinear function is not equal to the function of the expectation. As Firth (1993) observed, generally speaking, a convex estimating function, combined with its mean-unbiasedness, will cause a downward mean-bias in the estimator, while a concave function leads to an upward bias.

Our purpose is to avoid this bias problem associated with the expectation operator and non-linear estimating equations by considering the median. This is based on the fact that the median of a monotone function is the function of the median. Intuitively, if we work with the median, a properly constructed estimating equation is more likely to deliver a MU estimator.

3.3.2 Adjusting Estimating Equations Towards Median-unbiasedness

If an estimating equation satisfies median-unbiasedness, the consequent estimator (the solution to the equation) will be median-unbiased under conditions

more general than those for mean-unbiasedness. These conditions are stated in the following theorem.

Theorem 3.3.1 If an estimating equation $Q(\theta)=0$ satisfies:

1. $med[Q(\theta)|\theta_0]=0$, where θ_0 is the true parameter value; and
2. $Q(\theta)$ is a continuous function monotonic in θ,

then the solution to the estimating equation, $\hat{\theta}$, is a MU estimator, i.e., $med(\hat{\theta})=\theta_0$.

Proof. As $\hat{\theta}$ is the solution of the estimating equation, we have

$$Q(\hat{\theta})=0. \qquad (3.6)$$

From the median-unbiasedness condition in 1,

$$\Pr\{Q(\theta)\geq 0|\theta_0\}=1/2. \qquad (3.7)$$

Therefore if we combine (3.6) and (3.7), we have

$$\Pr\{Q(\theta)\geq Q(\hat{\theta})|\theta_0\}=1/2. \qquad (3.8)$$

And from the monotonicity of $Q(\bullet)$, $\Pr\{\hat{\theta}\geq\theta_0\}=1/2$ must hold. Same arguments can be used to show $\Pr\{\hat{\theta}\leq\theta_0\}=1/2$. Therefore, $\hat{\theta}$ is MU.

Corollary 3.3.1 If a linear estimating equation $Q(\theta)=T_1-\theta T_2=0$ is median-unbiased, the resulting estimator must be median-unbiased.

Compared with mean-unbiasedness, among the MU estimating equations, we have included all the linear estimating equations and all the monotonic nonlinear estimating functions, which form a possibly broader class of estimating equations for us to compute MU estimators. In particular, for a linear MU estimating equation, no other constraints (such as the independence between T_1/T_2 and T_2 in mean-unbiased case) is required for it to produce a MU estimator. Therefore this corollary can be seen as an extension of the result by Lieberman (1998).

The monotonicity requirement is more general than the linearity condition in the mean-unbiasedness case, but it is still sometimes too restrictive. From the empirical results, we know that global monotonicity is probably not required. The condition can be further relaxed to a wider class of functions, which is given in the following lemma:

Lemma 3.3.1 Given a random variable X with its median at m_X, and a continuous function $f: \Xi \to \Theta$ defined on R^1, which satisfies:

1. $\forall x \in \Xi$, if $f(x) \neq f(m_X)$, then $x \neq m_X$, and

2. $\forall x_1 < m_X$ and $x_2 > m_X$, $\mathrm{sgn}(f(x_1) - f(m_X)) \bullet \mathrm{sgn}(f(x_2) - f(m_X)) = -1$,

then we have

$$med(f(X)) = f(m_X). \tag{3.9}$$

The major difficulty that the proposed method usually faces lies in the fact that in most cases, unlike mean-unbiasedness, the estimating functions are not MU, therefore we need to adjust them towards median-unbiasedness in order to solve for a MU estimator. Now we introduce the proposed adjusted estimating equations. In order to compare with mean-unbiased estimating equations, we consider the score function as our example in the following discussions. Without loss of generality, the same arguments can be applied to other forms of estimating equations.

We assume the score function used to estimate parameter θ is $U(\theta)$. In order to force it to be MU, we use the following adjustment:

$$U^{MU}(\theta) = U(y;\theta) - med[U(y;\theta) | y \sim f(\theta)] = 0, \tag{3.10}$$

where $f(\bullet)$ is the assumed distribution function that y follows while the median is computed as if θ were the true parameter. It is quite clear that this new score, $U^{MU}(\theta)$, is a direct analogue of the adjusted profile likelihood score in McCullagh and Tibshirani (1990) with expectation replaced by median and without the second-order term penalty. The functional form of the density $f(\bullet)$ is usually required if the median function is to be computed via numerical integration.

Asymptotically, the solution to this adjusted estimating equation is equivalent to the solution to the original estimating equation. This can be seen from the asymptotic normality of $U(\theta)$, which is satisfied by most estimating equations in practice. If the initial estimator is the ML estimator, for example, under the usual regularity conditions, the score is asymptotically normal. Therefore asymptotically, the median of the score vanishes and the proposed adjusted score is equivalent to the original score. The variance of the left-hand side of the adjusted equation should converge to the corresponding information matrix component as the median term tends to zero. Hence the adjusted equation (3.10) should produce an estimator with the same asymptotic properties held by the solution to the original estimating equation $U(\theta) = 0$.

The adjusted estimating equation will be MU by construction. Based on Theorem 3.1.1, if the left-hand side of the equation is monotonic in θ, or satisfies conditions 1 and 2 in Lemma 3.1.3, the solution of equation (3.10) should be a MU estimator of θ. However, in practice, the adjusted equation, just like the original estimating equation, can rarely be solved analytically. Therefore the output of the proposed method is usually an approximately MU estimator. In the next section, we are concerned with how to find a solution to the adjusted estimating equation.

3.3.3 An Iterative Algorithm to Solve MU Estimating Equations

In practice, it is usually hard to analytically solve the adjusted estimation equations as the median function of the original score rarely has an explicit form. First we construct a sequence $\{\hat{\theta}_{(r)}\}$ by the iterative definition:

$$U(y,\hat{\theta}_{(r)}) = med\{U(y,\hat{\theta}_{(r)})|\hat{\theta}_{(r+1)}\} \tag{3.11}$$

where $\hat{\theta}_{(0)}$ can be an arbitrarily picked initial value. Therefore, for a given $\hat{\theta}_{(r)}$, $\hat{\theta}_{(r+1)}$ is obtained by solving (3.11). The conditional median $med\{\bullet|\hat{\theta}_{(r+1)}\}$ in (3.11) denotes the median of the random function $U(y,\hat{\theta}_{(r)})$, in which y is generated by its distribution function with $\hat{\theta}_{(r+1)}$ as the true parameter. We introduce a function to represent the RHS of (3.11):

$$g(\tilde{\theta},\theta) = m\{U(y,\theta)|\tilde{\theta}\}. \tag{3.12}$$

Now (3.11) becomes

$$U(y,\hat{\theta}_{(r)}) = g(\hat{\theta}_{(r+1)},\hat{\theta}_{(r)}). \tag{3.13}$$

We rely on the following recursive algorithm to update $\{\hat{\theta}_{(r)}\}$ and solve equation (3.10):

Step 1 Pick a starting value $\hat{\theta}_{(0)}$;

Step 2 Simulate (or in some cases, compute) the conditional median of the score, i.e., $g(\hat{\theta}_{(0)},\theta)$ in (3.13), which is equivalent to computing $med[U(y;\theta)|y \sim f(\hat{\theta}_{(0)})]$ for a grid of θ values;

Step 3 Solve (3.13) for θ, i.e., $U(y;\theta) = g(\hat{\theta}_{(0)},\theta)$ via algorithms such as the Secant method, and denote the solution by $\hat{\theta}_{(1)}$;

Step 4 Use $\hat{\theta}_{(1)}$ in the place of $\hat{\theta}_{(0)}$ and go back to step 2; Continue the procedure until converging (i.e., the difference between the two consecutive estimates is smaller than a pre-determined margin);

Our procedure is the replication of a two step process, which is similar to the well-known EM algorithm developed by Dempster et al. (1977). The difference is that we replace its expectation step by the conditional median computation in Step 2. Mak (1993) provided an efficient iterative procedure to solve non-linear mean-unbiased estimating equations of a similar nature. The proposed algorithm in this chapter is analogous to Mak's approach. However, unlike Mak's algorithm, an analytic proof for the convergence of the proposed algorithm is not available due to the difficulty in defining the conditions for taking derivatives inside the median operator. We will illustrate this algorithm in two practical examples in Chapter 4, and in both examples the algorithm is shown to converge at least as fast as the usual non-linear optimisation routines used in most ML estimation procedures.

3.3.4 Link to Other Bias-reduction Methods

In this section, we demonstrate that the proposed adjusted estimating equation is closely linked to two existing mean bias-reduction methods: the bootstrap bias-correction suggested by MacKinnon and Smith (1998) and the ML bias-reduction method of Firth (1993).

MacKinnon and Smith (1998) treated the initial estimator as a sum of the true parameter value, a bias function and a random deviation from the mean. By using a bootstrap procedure, the bias term to the order of n^{-1} can be removed. Here we use their ideas in the context of median-bias correction. Similar to their definition of the mean-bias function, we define the median-bias function $b(\theta_0,T)$ of an estimator $\hat{\theta}$ as

$$b(\theta_0,T) = med\{\hat{\theta}\} - \theta_0, \tag{3.15}$$

where θ_0 is the true value of the parameter, T is the sample size and $med\{\bullet\}$ is the usual median function. We then can express the initial estimator $\hat{\theta}$ in the following way:

$$\hat{\theta}=\theta_0+b(\theta_0,T)+d(\theta_0,T), \tag{3.16}$$

where $d(\theta_0,n)$ is the random deviation of $\hat{\theta}$ from its median. MacKinnon and Smith (1998) showed that if we want to correct the bias of $\hat{\theta}$, we simply solve the equation

$$\tilde{\theta}+b(\tilde{\theta},T)=\hat{\theta} \tag{3.17}$$

for $\tilde{\theta}$. Although explicitly, the analytical form of the bias function is required in this method, bootstrap can be used to approximate the left-hand side of (3.17) as a single function to avoid the derivation of the bias function, see MacKinnon and Smith (1998) for details.

As reviewed in Chapter 2, Firth (1993) proposed another intuitively very attractive way of reducing the mean-bias of a ML estimator. If the original score $U(\theta)$ leads to the MLE $\hat{\theta}$, we can use the adjusted score:

$$U^*(\theta)=U(\theta)-i(\theta)b(\theta), \tag{3.18}$$

and equate it to zero to get a bias-corrected estimator. $i(\theta)$ and $b(\theta)$ are the Fisher information and the mean bias function of $\hat{\theta}$, respectively. In practice, the exact form of the bias function $b(\theta)$ is usually unknown, so either the term to a certain order (usually the T^{-1} term) in the expansion of $b(\theta)$ is used instead, or the empirical bias function $b(\theta)$ is estimated by bootstrap.

The proposed approach in this section is in line with the idea of preventing bias by adjusting the estimating equations. But we try to achieve median-unbiasedness instead of mean-unbiasedness. We now show that the proposed adjustment defined by (3.10) can be seen as equivalent to an analogy of Firth's

method and will also lead to the bias-corrected estimator from solving equation (3.17).

First we prove that by simply replacing the bias function in Firth's method with the median-bias function defined by (3.15), and by using the estimated information component in (3.18), the (approximate) median-bias correction can be achieved, i.e., the bias-corrected estimator which solves equation (3.17) should also solve the adjusted estimating equation (3.18) at least approximately.

The initial estimator $\hat{\theta}$ satisfies the original estimating equation, so we have $U(\hat{\theta})=0$. We denote the bias-corrected estimator based on MacKinnon and Smith's method by $\hat{\theta}_{MU}$. From (3.17), $\hat{\theta}_{MU}$ should satisfy

$$\hat{\theta}_{MU}=\hat{\theta}-b(\hat{\theta}_{MU}). \tag{3.19}$$

On the other hand, we assume the bias-corrected estimator $\hat{\theta}_{MU}$ is also the solution of a different (unknown) estimating equation:

$$U^*(\hat{\theta}_{MU})=0. \tag{3.20}$$

We combine (3.19) and (3.20) and use a Taylor's expansion:

$$\begin{aligned}
U(\hat{\theta}_{MU}) &= U(\hat{\theta}-b(\hat{\theta}_{MU}))\\
&\approx U(\hat{\theta})-\left.\frac{\partial U}{\partial\theta}\right|_{\hat{\theta}} b(\hat{\theta}_{MU})\\
&= U(\hat{\theta})+\hat{I}(\hat{\theta})b(\hat{\theta}_{MU})\\
&= \hat{I}(\hat{\theta})b(\hat{\theta}_{MU})
\end{aligned}$$

where $I(\theta)$ is the observed information (i.e., the corresponding diagonal component in the Hessian matrix of the likelihood function). Therefore we have

$$U(\hat{\theta}_{MU})-\hat{I}(\hat{\theta})b(\hat{\theta}_{MU})=0. \tag{3.21}$$

Hence if we use

$$U^*(\theta) = U(\theta) - \hat{I}(\hat{\theta})b(\theta) \tag{3.22}$$

as the adjusted estimation equation, it should lead to the approximately MU estimator $\hat{\theta}_{MU}$. Therefore the corrections given by (3.17) and (3.18) are equivalent only subject to the difference caused by the estimated information component evaluated at different points. The unbiasedness of both methods is only approximate because the bias function is usually approximated to a certain order. Firth (1993) showed that the expected information $i(\theta)$ is preferred to the observed information $I(\theta)$ to be used in the adjusted score. In our proof, the derivative of the estimating equation at the original estimate is used to estimate the information matrix component.

Next we show that the proposed MU adjustment to the estimating equations given by (3.10) is an analogy of Firth's adjustment, i.e., the two adjusted estimating equations (3.10) and (3.18) are equivalent to each other to the order of n^{-1}, provided $U(y;\theta)$ satisfies the conditions defined in Lemma 3.3.1. To see this, we apply a Taylor's expansion to the adjusted estimating function (3.10),

$$\begin{aligned}
U^*(\hat{\theta}_{MU}) &= U(\hat{\theta}_{MU}) - med\{U(y;\hat{\theta}_{MU})|\hat{\theta}_{MU}\} \\
&= U(\hat{\theta}_{MU}) + [U(y;\hat{\theta}) - U(y;med(\hat{\theta}_{MU}))] \\
&\approx U(\hat{\theta}_{MU}) + \left.\frac{\partial U}{\partial \theta}\right|_{\hat{\theta}} (med(\hat{\theta}_{MU}) - \hat{\theta}) \\
&= U(\hat{\theta}_{MU}) - \hat{I}(\hat{\theta})b(\hat{\theta}_{MU}).
\end{aligned}$$

Notice that $U(y;\hat{\theta}) = 0$ and $U(\bullet)$ is assumed to satisfy the conditions for it to be invariant to the median operator. The proposed adjustment to the estimating equation is equivalent to Firth's correction apart from the difference in estimating the information component.

From the above analysis, we prove from a different point of view that by solving the proposed MU estimating equation (3.10) for $\hat{\theta}_{MU}$, with some constraints

on $U(\theta)$, we achieve the same median bias correction as the existing methods. The advantage of the proposed method compared with Firth's method is that knowledge about the second derivative of the likelihood is not required. Compared with MacKinnon and Smith (1998), we do not need to derive or approximate the bias function. The possible disadvantage of the proposed approach is clearly the monotonic requirement of the estimating function. Nevertheless, in Chapter 4, we will show that this method works well for the marginal likelihood score in the linear regression with first order autoregressive disturbances and the first order dynamic linear regression model. But let us first illustrate this method with several simple theoretical examples.

3.3.5 Some Examples

3.3.5.1 The GLS Estimator in the Gaussian Linear Regression

Consider the classic linear regression model, $y = X\beta + u$, where y is the dependent variable, X is the matrix of regressors and u is a vector of random errors. Assume $E(u) = 0$ and $Var(u) = \sigma^2\Sigma$ for some positive definite Σ. The normal equation for the GLS estimator (without the normality error assumption, this is not necessarily the score function) is given by:

$$-(X'\Sigma^{-1}X)\beta + X'\Sigma^{-1}y = 0. \tag{3.23}$$

If we denote the left-hand side of this equation by $U(\beta)$, then it is apparent that

$$U(\beta) = X'\Sigma^{-1}u$$

is an odd function of the symmetrically distributed errors. By Andrews (1986), $U(\beta)$ must have a symmetric distribution around 0 provided that u follows a symmetric distribution. Hence $med_\beta[U(\beta)] = 0$. In other words, $U(\beta) = 0$ is a MU estimating equation and it is also monotonic in β. Therefore the usual GLS estimator

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \tag{3.24}$$

is the solution to both the original and the adjusted estimation equation (3.10). Based on Theorem 3.2.1, $\hat{\beta}_{GLS}$ is MU. As a matter of fact, Phillips and Andrews (1987) proved that $\hat{\beta}_{GLS}$ is the best linear MU estimator for a large family of error distributions and loss functions.

3.3.5.2 Error Variance in the Simple Linear Regression

We consider the simple linear regression model $y = X\beta + u$ with $u \sim N(0, \sigma^2 I)$, and the parameter of interest is the error variance σ^2. The log-likelihood, concentrated score and ML estimator are:

$$l = -\frac{T}{2}\ln\sigma^2 - \frac{(y-X\beta)'(y-X\beta)}{2\sigma^2} + const.,$$

$$U(\sigma^2) = -\frac{T}{2\sigma^2} + \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{2\sigma^4}, \tag{3.25}$$

and

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

$$\hat{\sigma}^2_{MLE} = \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{T}. \tag{3.26}$$

First we use the analog of Firth's adjustment given by (3.21). It is easy to show that

$$b(\sigma^2) = med(\hat{\sigma}^2_{MLE}) - \sigma^2 = med(\chi^2_{T-k})\frac{\sigma^2}{T} - \sigma^2, \tag{3.27}$$

where $med(\chi^2_{T-k})$ is the median of a chi-squared distribution with $T-k$ degrees of freedom. Therefore, the adjusted score (3.21) is given by:

$$\begin{aligned} U^*(\sigma^2) &= U(\sigma^2) - I(\sigma^2)b(\sigma^2) \\ &= U(\sigma^2) - I(\sigma^2)(med(\hat{\sigma}_{MLE}{}^2) - \sigma^2). \end{aligned} \tag{3.28}$$

Here we use the expected Fisher information in place of $I(\sigma^2)$,

$$I^E(\sigma^2) = \frac{T}{2\sigma^4}, \tag{3.29}$$

So equation (3.28) can be written as,

$$\begin{aligned} U^*(\sigma^2) &= U(\sigma^2) - \frac{T}{2\sigma^4}\left(\frac{med(\chi^2_{T-k})\sigma^2}{T} - \sigma^2\right) \\ &= -\frac{T}{2\sigma^2} + \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{2\sigma^4} - \frac{med(\chi^2_{T-k})\sigma^2}{2\sigma^2} + \frac{T}{2\sigma^2}. \end{aligned}$$

Solve $U^*(\sigma^2) = 0$ for σ^2 and we get the MU estimator

$$\hat{\sigma}^2_{MU} = \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{med(\chi^2_{T-k})}, \tag{3.30}$$

which was discussed in Cox and Hinkley (1974) and also in Phillips and Andrews (1987).

Now we use the proposed adjusted score function (3.10). Because

$$med(U(\tilde{y};\sigma^2)) = -\frac{T}{2\sigma^2} + \frac{med(\chi^2_{T-k})}{2\sigma^2}, \tag{3.31}$$

we have

$$\begin{aligned} U^{MU}(\sigma^2) &= U(\bar{y};\sigma^2) - med(U(\bar{y};\sigma^2)) \\ &= [-\frac{T}{2\sigma^2} + \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{2\sigma^4}] - [\frac{med(\chi^2_{T-k})}{2\sigma^2} - \frac{T}{2\sigma^2}] \\ &= \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{2\sigma^4} - \frac{med(\chi^2_{T-k})}{2\sigma^2}. \end{aligned}$$

Therefore, if we solve the adjusted score equation $U^{MU}(\sigma^2) = 0$, the same $\hat{\sigma}^2{}_{MU}$ in (3.30) is derived. This example illustrates the equivalence between the proposed adjustment in (3.10) and the analogy to Firth's adjustment in (3.21).

3.3.5.3 First-order Autoregression

We consider a simple first-order autoregressive model without an intercept, which was studied by many other researchers. We will attempt to show that the MU estimator proposed by Andrews (1993) can be derived by solving the proposed adjusted score equation (3.10). The model of interest is:

$$\begin{aligned} y_t &= \rho y_{t-1} + u_t\,, \\ u_t &\sim IN(0,\sigma^2), \end{aligned}$$

where $y_0 = \frac{1}{\sqrt{1-\rho^2}} u_0$, for $|\rho| < 1$, and $y_0 = 0$, if $\rho = 1$. The OLS estimator, which coincides with the ML estimator, is given by:

$$\hat{\rho}_{OLS} = \sum_{t=2}^{T} y_t y_{t-1} \Big/ \sum_{t=2}^{T} y_{t-1}{}^2. \tag{3.32}$$

Andrews (1993) showed that

$$\hat{\rho}_{MU} = m^{-1}_{OLS}(\hat{\rho}_{OLS}) \tag{3.33}$$

is exactly MU, where

$$m_{OLS}(\rho) = med[\hat{\rho}_{OLS}|\rho], \tag{3.34}$$

is the median function of $\hat{\rho}_{OLS}$. Now we prove that the adjusted score function (3.10) will lead to the same $\hat{\rho}_{MU}$. The ML score function is given by

$$U(\rho) = \rho\sum_{t=2}^{T} y^2_{t-1} - \sum_{t=2}^{T} y^2{}_{t-1}. \tag{3.35}$$

We adjust (3.35) according to (3.10) towards median-unbiasedness, i.e.

$$U^{MU}(\rho) = U(\rho) - m_U(\rho), \tag{3.36}$$

where

$$m_U(\rho) = med[U(\rho)|\rho] \tag{3.37}$$

stands for the median function of the score. Now if we denote the solution to the adjusted score equation $U^{MU}(\rho) = 0$ by $\tilde{\rho}$, then we have

$$med[\rho\sum_{t=2}^{T} y^2_{t-1} - \sum_{t=2}^{T} y_t y_{t-1} \Big| \tilde{\rho}] = \tilde{\rho}\sum_{t=2}^{T} y^2_{t-1} - \sum_{t=2}^{T} y_t y_{t-1}. \tag{3.38}$$

From (3.32) and (3.35), the above equation is equivalent to

$$\Pr\{(\hat{\rho}_{OLS} - \rho)\sum_{t=2}^{T} y^2_{t-1} \geq (\hat{\rho}_{OLS} - \tilde{\rho})\sum_{t=2}^{T} y^2_{t-1} \Big| \tilde{\rho}\} = 1/2. \tag{3.39}$$

If we denote the median function of $\hat{\rho}_{OLS} - \rho$ by $m'_{OLS}(\rho)$, which is given by

$$m'_{OLS}(\rho) = med[(\hat{\rho}_{OLS} - \rho)|\rho], \tag{3.40}$$

then from (3.39) and based on the definition of the median function, we have

$$m'_{OLS}(\tilde{\rho}) = \hat{\rho}_{OLS} - \tilde{\rho}. \tag{3.41}$$

Notice that the two median functions $m_{OLS}(\rho)$ given by (3.34) and $m'_{OLS}(\rho)$ given by (3.40) are linked by

$$m_{OLS}(\rho) = m'_{OLS}(\rho) - \rho, \tag{3.42}$$

Therefore we have

$$m_{OLS}(\tilde{\rho}) = \hat{\rho}_{OLS}. \tag{3.43}$$

or equivalently, the solution to the adjusted estimation equation (3.38), $\tilde{\rho}$, coincides with the exactly MU estimator (3.30) proposed by Andrews (1993).

We also notice that in the simple AR(1) model, the least squares estimator is identical to the minimal sufficient statistic of ρ (e.g. see Hurwitz, 1950). Therefore, this is a good example of the proposed method in some cases being a special case of the general approach set out in Section 3.2.

3.3.6 Extension to the Multi-parameter Case

Consider the problem of estimating a multi-variate parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$. Usually we have k estimating equations to solve for $\hat{\theta}$. Denote these equations by

$$U(\theta) = [U_1(\theta), \ldots, U_k(\theta)]' = 0. \tag{3.44}$$

In Chapter 2, we reviewed several definitions of median-unbiasedness for a multi-variate estimator and adopted the idea of marginal median, i.e., we are trying to achieve median-unbiasedness for each coordinate of the estimator $\hat{\theta}$, such that

$$med[\hat{\theta}_i] = \theta_i \text{ for } i = 1, \ldots, k. \tag{3.45}$$

We argue that in many cases, each single parameter θ_i is of its own importance and the impartiality of each $\hat{\theta}_i$ is sometimes more relevant than the hard-to-define overall unbiasedness of the vector $\hat{\theta}$. Now we propose the following algorithm to adjust the estimating equations (3.44) iteratively for an approximately MU estimator:

1. Pick a starting point $\hat{\theta}^{(1)} = (\hat{\theta}_1^{(1)}, \ldots, \hat{\theta}_k^{(1)})'$.

2. In the first equation $U_1(\theta) = 0$, treat θ_1 as the only unknown parameter and replace all the other parameters by their corresponding values in $\hat{\theta}^{(1)}$. Adjust this equation by using the method proposed in the previous section, i.e. solve for θ_1 from

$$U_1(\theta_1, \hat{\theta}_2^{(1)}, \ldots, \hat{\theta}_k^{(1)}) - med[U_1(\theta) | y \sim f(\theta_1, \hat{\theta}_2^{(1)}, \ldots, \hat{\theta}_k^{(1)})] = 0. \tag{3.46}$$

   Denote the solution by $\hat{\theta}_1^{(2)}$.

3. Replace $\hat{\theta}_1^{(1)}$ in $\hat{\theta}^{(1)}$ by $\hat{\theta}_1^{(2)}$, and treat θ_2 as the unknown parameter. Adjust $U_2(\theta)$ and solve for $\hat{\theta}_2^{(2)}$; Repeat the process for every $U_i(\theta)$ and denote the solutions from this replication by $\hat{\theta}^{(2)} = (\hat{\theta}_1^{(2)}, \ldots, \hat{\theta}_k^{(2)})'$.

4. Treat $\hat{\theta}^{(2)}$ as the new starting point and repeat Step 2 and Step 3. If the desired accuracy is achieved, stop; otherwise continue the replications until converging (i.e., the difference between the two consecutive estimates becomes less than a pre-determined error of margin).

When we adjust the estimating equation each time in Step 2, we replace the other parameters by their previous estimates. The median function is computed based on the distribution of $U_i(\theta)$ with estimated θ. This introduces error into the median function. Therefore the adjustment in Step 2 will only be approximate. As a result, the final estimates $\hat{\theta}^{(r)}$ will only be approximately MU.

It should be noted that the median function has to be evaluated for each set of estimates in the replications. This makes the computational burden heavy. We also point out that the proposed method is similar to the one discussed in Fair (1996) and the one used in Andrews and Chen (1994), although they did not discuss the method in the context of estimating equations. An analytical proof for the convergence of this algorithm in a general setting is not available. Empirical evidence provided in Chapter 4 suggests that the algorithm converges fast at least for the linear dynamic linear regression model. We notice that the convergence of the similar algorithms developed by Rudebusch (1992), Andrews and Chen (1994) and Fair (1996) were also only illustrated empirically.

3.4 Inverting Significance Tests for MU Estimators

The method discussed in Section 3.3 was based on estimating equations. It may face difficulties when these equations are complicated and contain nuisance parameters. For example, most of the score functions in MLE and Quasi-maximum likelihood procedures are nonlinear functions that usually cannot be solved analytically. Most of the scores also involve multiple parameters in the same equation. This can make it hard to find the adjustment defined in (3.10) because of the difficulty of computing the conditional median function of the score needed in the adjustment. More importantly it is also subject to the same criticism encountered by the sufficient statistics, i.e., the monotonicity of the difference between the estimating function and its conditional median function can be hard to satisfy.

To avoid these difficulties, we take a similar yet different approach in this section. Instead of looking at estimating equations, we consider hypothesis test statistics. Compared with estimating equations, test statistics associated with a single parameter are easier to find, and their distribution functions are in most cases easy to work out, so it is more likely that we are able to verify the monotonicity of their median functions.

We do not claim that inverting a test statistic at the 50% significance level for a MU estimator is a new method. As a matter of fact, since being regarded as a special case of the application of the duality between confidence bounds and significance tests discussed in Lehmann (1959), it has been used to construct MU estimators in autoregressive time series models. Examples include Stock (1991), Andrews (1993) and Watson and Stock (1998). Interestingly, all these examples are concerned with estimating near non-stationary time series.

However, in this thesis we address some of the problems existing with current applications, which may prevent it, in its current form, being extended to other estimation problems as it may fail to deliver reliable estimates. In particular, we discuss the importance of choosing a 'good' test statistic to invert, which has largely been ignored by other researchers. Because of the difference between interval estimation and point estimation, we also develop a more reliable test inversion method – grid inversion, which is different from the one used in most current examples – fixed-point inversion. We show that the proposed grid inversion method is more likely to be immune to the problems that the fixed-point inversion method may suffer from. We start our discussion by reviewing the well-known relationship between a significance test and a confidence interval.

3.4.1 Duality of Significance Tests and Confidence Intervals

It is generally accepted that a confidence interval and a significance test can be treated as the two sides of the same coin, see e.g., discussions in Lehmann (1959). Several recent papers appearing in the unit root literature explored this duality in constructing confidence intervals, and are related to our research: Dufour (1990) inverted the Durbin-Watson statistic to compute exact confidence sets in the linear regression model with AR(1) disturbances. Kiviet and Phillips (1992) and Kiviet and Dufour (1997) inverted a modified t statistic in the dynamic linear regression model for confidence intervals. Ahtola and Tiao (1984) inverted the square root of the score test statistic to compute confidence intervals in non-stationary autoregressive models. Stock (1991) also presented an empirical comparison of the different confidence intervals based on inverting two different test statistics. Carpenter (1999) and Hansen

(1999) combined test inversion with bootstrap and provided a theoretical justification for the accuracy of the intervals so constructed. Recently Wright (2000b) considered confidence intervals based on test inversion in the cointegration model.

The close link between the power of a test and the accuracy of the corresponding confidence set is also well known. Following Lehmann (1959, Ch. 5), for each $\theta_0 \in \Omega$, if $A(\theta_0)$ is the acceptance region of a level-α test for testing $H(\theta_0)$: $\theta = \theta_0$, then

$$S(x) = \{\theta : x \in A(\theta), \theta \in \Omega\} \tag{3.47}$$

is a family of confidence sets for θ at the $1-\alpha$ level. If for all θ_0, $A(\theta_0)$ is UMP for testing H_0 at level α against the alternatives $K(\theta_0)$, these intervals are most accurate in the sense that for each $\theta_0 \in \Omega$, $S(x)$ minimizes $\Pr_\theta\{\theta_0 \in S(x)\}$, $\forall \theta \in K(\theta_0)$.

If we set the confidence level at 50%, instead of a confidence interval, we get a MU point estimate when inverting the test statistic. This was highlighted in Lehmann (1959) and more recently discussed by Stock (1994).

A very simple example of this is the MU estimator of the error variance in a simple linear regression model. The two-sided confidence interval at level-α is given by,

$$\Pr\{\frac{s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{s^2}{\chi^2_{1-\alpha/2}}\} = 1-\alpha, \tag{3.48}$$

and if the significance level α is set at 50%, the confidence interval shrinks to a single point, which is a MU estimator of σ^2 and given by,

$$\hat{\sigma}^2 = \frac{s^2}{med(\chi^2_{T-k})}, \tag{3.49}$$

where s^2 is the sum of squares of the OLS residuals and $med(\chi^2_{T-k})$ is the median of a chi-square random variable with $T-k$ degrees of freedom. This coincides with the error variance estimator reviewed in Chapter 2.

3.4.2 Inverting a Test Statistic for a MU Estimator

We first define the conditions for a test statistic to be inverted to produce a MU estimator. Let x be the observed data and $t(x)$ be a test statistic testing a null hypothesis involving a scalar parameter θ against some one-sided alternative hypothesis. We define the median function of $t(x)$ related to the true value of θ by:

$$m(\theta) = med[t(x)|y \sim f(\theta)], \tag{3.50}$$

where $f(\bullet)$ is the assumed distribution under which the observed data were generated. Some test statistics (such as the t test) are based on a consistent estimator $\hat{\theta}$, but this is not always the case. Hence it is important to emphasize that $m(\bullet)$ is defined to be a function of θ, not $\hat{\theta}$. We define an estimator by

$$\hat{\theta}_{MU} \equiv m^{-1}[t(x)] \tag{3.51}$$

within the parameter space of θ. The following theorem defines the conditions for $m^{-1}(\bullet)$ to exist.

Theorem 3.4.1 If $t(x)$ is continuous in θ and $m(\bullet)$ is non-decreasing or non-increasing monotone in θ, then $\hat{\theta}_{MU}$ in (3.51) is exactly MU.

Proof.

$$\Pr\{\hat{\theta}_{MU} \ge \theta_0\} = \Pr\{m(\hat{\theta}_{MU}) \ge m(\theta_0)\} = \Pr\{t(x) \ge m(\theta_0)|\theta_0\} = 1/2,$$

where θ_0 is the true value of the parameter. The three equalities are based on the monotonicity of $m(\bullet)$, the definition of $\hat{\theta}_{MU}$ and the definition of $m(\bullet)$, respectively. $\Pr\{\hat{\theta}_{MU} \le \theta_0\} = 1/2$ can be derived similarly. Therefore $\hat{\theta}_{MU}$ is a MU estimator.

It is important to point out that the condition requires the monotonicity of the median function, not the test statistic itself. This is sometimes misunderstood in the literature, see for example Hirji et al. (1989). Our experience shows that the monotonicity of a test statistic doesn't lead to the monotonicity of its quantile functions. This was also confirmed by both Andrews (1993) and Hansen (1999). They showed that although the OLS estimator and t-statistics are usually monotonic in θ, their quantile functions can be non-monotonic in the neighbourhood of a unit root. We will revisit this point in Chapter 5.

When the median function depends on nuisance parameters and/or only the limit of the test statistic under the null is tractable, we may choose to invert the median function of the limiting distribution of the test statistic. As a result, the estimator will only be MU asymptotically. The asymptotic approach avoids computing the median function for each sample size and each set of the nuisance parameter estimates, but the performance of the asymptotic MU estimators in small samples is not guaranteed. We state this asymptotic approach generally as a corollary of Theorem 3.4.1.

Corollary 3.4.1 Suppose that a sequence of functions (e.g., test statistics regarded as functions of some consistent estimators) f_t and a sequence of estimators $\hat{\theta}_t$ satisfies $\hat{\theta}_t \to \hat{\theta}$ in probability and $f_t(\hat{\theta}_t) \to f(\hat{\theta})$ in distribution, respectively, as $t \to \infty$. Define

$$\tilde{\theta}_t = m^{-1}[f_t(\hat{\theta}_t)]. \tag{3.52}$$

and if $m(\theta) = med_\theta[f(\hat{\theta})]$ is continuous and monotonic in θ, then $\tilde{\theta}_t$ is asymptotically MU, i.e., $\Pr\{\tilde{\theta}_t \le \theta_0\} \to \frac{1}{2}$ and $\Pr\{\tilde{\theta}_t \ge \theta_0\} \to \frac{1}{2}$ as $n \to \infty$.

Proof.

$$\begin{aligned}
&\Pr\{\tilde{\theta}_t \ge \theta_0\} = \Pr\{m^{-1}[f_t(\hat{\theta}_t)] \ge \theta_0\} \\
&\to \Pr\{m^{-1}[f(\hat{\theta})] \ge \theta_0\} \\
&= \Pr\{f(\hat{\theta}) \ge m(\theta_0) | \theta_0\} = \tfrac{1}{2}.
\end{aligned}$$

The continuous mapping theorem underlies the second step of the proof. $\Pr_\theta\{\tilde{\theta}_t \le \theta_0\} \to \frac{1}{2}$ can be derived similarly. Therefore then $\tilde{\theta}_t$ is asymptotically MU.

The best example of this approach is estimating the unit root model based on a local-to-unity parameterisation, in which the limiting null distribution depends on the drift parameter. Many researchers have adopted this approach to construct confidence intervals for the autoregressive parameter. In particular, Stock (1991) compared his confidence intervals based on the asymptotic distribution with those based on numerical approximation of the finite sample distribution and found the asymptotic intervals have good coverage probabilities but are usually wider. In this thesis, we avoid the asymptotic approach and attempt to achieve median-unbiasedness in finite samples.

3.4.3 Test Performance and Efficiency of MU Estimator

Although the method is straightforward to understand, two important questions remain unanswered. First, which test statistics can deliver reliable MU estimates? In other words, which tests have well-behaved median functions that can be inverted? Second, which test statistics produce the most efficient MU estimator among tests that can be used? In this section we try to answer these two questions.

Apart from that of a UMP test, the non-monotonicity of the quantile function of a test statistic is quite usual. Non-monotonic quantile functions have been reported by several authors. For example, for the first-order autoregressive model with a drift and a time trend, Andrews (1993) found that the 95% quantile function of the OLS estimator is not monotonic; Stock (1991) observed that the quantile functions (include the median function) of both the DW test statistic and the Sargan-Bhargava test statistics (Sargan and Bhargava, 1983) are not monotonic; Hansen (1999) reported that the quantile functions of the t statistics (Dickey-Fuller test in the case of testing the unit root hypothesis) are also not monotonic. If only confidence intervals are required, the nonmonotonicity of the quantile functions will only lead to disjoint confidence intervals or empty intervals from time to time. Dufour (1990, 1997)) reported such intervals and argued that they are meaningful and should not be discarded. But if we are interested in point estimation, this non-monotonicity will lead to multiple solutions (estimates) for a single parameter value, and hence fail to produce reliable estimates. Therefore we need to find test statistics which have a monotonic median function.

If a test is locally biased, i.e., its power drops below its size for some local alternatives, or if it has a non-monotonic power curve, its median function is likely to be non-monotonic. Tests suffering from these problems in small samples are quite common and have been reported by many researchers. For example, for the linear regression with AR(1) or random walk disturbances, this was reported by Tillman (1975), King (1985a), Kramer (1985), Kramer and Zeisel (1990) and Bartels (1992), while Goh and King (1999) studied small sample deficiencies of the tests in the dynamic linear regression model. It should be cautioned that the monotonicity of the quantile functions at different significance levels may be different. Andrews (1993) observed that the median function of the OLS estimator in a first order autoregressive model with an intercept and a time trend is always monotonic for all the sample sizes examined, while the 95%-quanitle function is not monotonic in the neighbourhood of 1 and for some sample sizes. Therefore it is pertinent that the median function should be examined for that particular sample size and design matrix before it is inverted. We also point out that the monotonicity of the power curve of a test is not a sufficient or necessary condition for the monotonicity of its quantile functions, due to the fact that the distributions of the test statistic under the alternative hypotheses are usually affected by factors other than the magnitude of the parameter of interest.

Now we examine the relationship between the power of a test and the efficiency of the MU estimator based on it if it has a monotonic median function.

First of all, if a UMP unbiased test exists, the optimality of the MU estimator can be derived based on Lemann's (1959, p220) results:

Lemma 3.4.1 Within the family of distributions with monotone likelihood ratios, if $T(x)$ is the test statistic for a UMP unbiased test of H_0 against a two-sided alternative H_1, then:

(i) $m(\theta) = med_\theta[T(x)]$ is strictly increasing;

(ii) $\tilde{\theta} = m^{-1}[T(x)]$ is the optimal MU estimator for θ, in the sense that it minimises $EL(\theta,\hat{\theta})$ for any monotonic loss function L.

But in most examples of practical interest, only tests less optimal than UMP unbiased are available. It is well known that more powerful tests will lead to more accurate confidence intervals. Is there a similar link between the choice of a test and the performance of the estimator based on inverting the test? Stuart and Ord (1991, pp956-958) established the equivalence of the asymptotic relative efficiency of a consistent estimator and the test based on it. We extend Kendall's result to link the asymptotic power properties of the tests to the asymptotic efficiency of the estimators based on inverting their median functions.

Theorem 3.4.2 Assume T_1 and T_2 are two test statistics testing the same set of hypotheses about a scalar parameter θ (with null value at θ_0), and their limits are given by τ_1 and τ_2 as $T \to \infty$. Let $\tilde{\theta}_1$ and $\tilde{\theta}_2$ be asymptotically MU estimators defined according to (3.52) and based on T_1 and T_2, respectively. Then in a neighbourhood of the true (null) value θ_0, we have:

$$ARE(\tilde{\theta}_1/\tilde{\theta}_2) = ARE(T_1/T_2),$$

where, according to Kendall and Stuart (1967), the asymptotic relative efficiency (ARE) of two estimators $\tilde{\theta}_1$ and $\tilde{\theta}_2$ is defined by

$$ARE(\tilde{\theta}_2/\tilde{\theta}_1)=\lim_{T\to\infty}\left(\frac{\mathrm{var}_{\theta_0}[\tilde{\theta}_1]}{\mathrm{var}_{\theta_0}[\tilde{\theta}_2]}\right)^{1/2\delta}, \tag{3.58}$$

while the ARE of two test statistics T_1 and T_2 is defined by,

$$ARE(T_2/T_1)=\lim_{T\to\infty}\left\{\frac{\{\partial E_{\theta_0}(T_2)/\partial\theta\}^2/\mathrm{var}_{\theta_0}[T_2]}{\{\partial E_{\theta_0}(T_1)/\partial\theta\}^2/\mathrm{var}_{\theta_0}[T_1]}\right\}^{1/\delta}, \tag{3.59}$$

where δ is the order of magnitude in n of the variances of the estimators or the test statistics. For example, corresponding to an estimation variance of order T^{-1}, $\delta=\frac{1}{2}$.

Proof. The consistency of $\tilde{\theta}_i$ $(i=1, 2)$ implies $\tilde{\theta}_i\to\theta_0$ in probability as $T\to\infty$. From Corollary 3.4.1,

$$\tilde{\theta}_i=m_i^{-1}(T_i), \tag{3.60}$$

where $m_i(\bullet)$ is the median function of τ_i, i.e., the limit (in distribution) of T_i as $T\to\infty$. Based on the continuous mapping theorem, we have $\theta_0=m_i^{-1}(\tau_i)$. Expanding (3.60) about τ_i by Taylor's theorem, we have,

$$\tilde{\theta}_i=\theta_0+(T_i-\tau_i)\left[\frac{\partial m^{-1}(T_i)}{\partial\tau_i}\right]_{T_i=T_i^*},$$

where T_i^* is intermediate in value between T_i and τ_i. It is obvious that as $T\to\infty$, $T_i^*\to\tau_i$, and $E_{\theta_0}(T_i^*)\to\tau_i$. So

$$\left[\frac{\partial m^{-1}(T_i)}{\partial T_i}\right]_{T_i=T_i^*}=\left.\frac{\partial\tilde{\theta}_i}{\partial E_{\theta_0}(T_i)}\right|_{\theta_0}+o_p(T).$$

Therefore,

$$\mathrm{var}_{\theta_0}[\tilde{\theta}_i]=\mathrm{var}_{\theta_0}[T_i]\Bigg/\left[\left.\frac{\partial E_{\theta_0}(T_i)}{\partial\tilde{\theta}_i}\right|_{\theta_0}\right]^2+o_p(T)\ . \tag{3.61}$$

If 2δ is the order of magnitude in n of the variances of $\tilde{\theta}_i$, based on (3.58) and (3.61), the ARE of $\tilde{\theta}_1$ and $\tilde{\theta}_2$ is given by

$$ARE(\tilde{\theta}_2/\tilde{\theta}_1)=\lim_{n\to\infty}\left(\frac{\mathrm{var}_{\theta_0}[\tilde{\theta}_1]}{\mathrm{var}_{\theta_0}[\tilde{\theta}_2]}\right)^{1/2\delta}=\lim_{T\to\infty}\left\{\frac{\{\partial E_{\theta_0}(T_2)/\partial\theta\}^2/\mathrm{var}_{\theta_0}[T_2]}{\{\partial L_{\theta_0}(T_1)/\partial\theta\}^2/\mathrm{var}_{\theta_0}[T_1]}\right\}^{1/2\delta},$$

and by definition (3.59), the theorem follows.

The above theorem only links locally the ARE of two tests to the ARE of the resulting estimators. Although for a test, the ARE is a function of the slope of the power curve in the neighbourhood of the null value, the asymptotic efficiency sometimes is not a good measure of the finite sample performance of an estimator. Therefore it is still not clear if there is any direct relationship between the finite sample power of a test and the finite sample bias and/or efficiency of the estimator based on it. In Chapter 5, we provide some empirical evidence for this possible link between the power of a test and the small sample performance of the corresponding MU estimator in the linear regression model with AR(1) or random walk disturbances.

We conclude this section by reiterating the main results: 1. Inverting a significance test statistic at its 50% level will produce a (sometimes asymptotically, if inverting the median function of the asymptotic distribution of the test statistic) MU estimator provided its (asymptotic) median function is monotonic. 2. The asymptotic efficiency of the consequent MU estimator directly depends on the power performance of the test being inverted. In the next section, we examine the situation when the median function of a test statistic is not monotonic.

3.4.4 Fixed-point Inversion and Grid Inversion

If the test statistic is pivotal, i.e., its quantile function is independent of the null value of the parameter and if the median function is monotonic, we can simply invert it at the fixed point of the calculated test statistic given the sample data. We call this method *fixed-point inversion*. But as discussed in the previous section, what happens more commonly in time series models is that the critical value function of a test statistic varies with the null value and/or its median function may not be monotonic. When these problems occur, the fixed-point inversion method breaks down due to non-unique estimates. This leads to the need for a different way of inverting the test statistic. In this section, we define a *grid inversion* method based on the median envelope of a series of test statistics. We contrast the new method with the fixed-point inversion.

Method 1 (Fixed-point invertion)

Test statistic:	$T(\theta_0;y)$, where θ_0 is the fixed null point.
Hypotheses:	H_0: $\theta=\theta_0$
	H_1: $\theta<\theta_0$.
Median function:	$m(\theta)=med[T(\theta_0;y)\|\theta]$.
Estimation procedure:	$\hat{\theta}^{MU}=m^{-1}[T(\theta_0;\tilde{y})]$, where $\tilde{y}$ is the observed sample data.

Method 2 (Grid invertion)

Test statistic:	$T(\theta_*;y)$, where θ_* is the null point which can vary within the parameter space of θ.
Hypotheses:	H_0: $\theta=\theta_*$
	H_1: $\theta<\theta_*$.
Median envelope:	$m(\theta_*)=med[T(\theta_*;y)\|\theta_*]$.
Estimation procedure:	solve $m(\theta_*)=T(\theta_*;\tilde{y})$ for $\hat{\theta}^{MU}$.

The grid inversion method allows the null value of the test statistic to vary in the parameter space. Instead of the median function of a single test statistic with a fixed null value, the median function used in the grid inversion method corresponds to the medians of a series of test statistics each evaluated at the corresponding null value. This approach avoids the possible non-monotonicity of the median function of a single test by considering a different median function

$$m_e(\theta)=med[T(\theta,y)|\theta], \tag{3.62}$$

which we call a '*median envelope*'. The term borrows from the idea of a power envelope which measures the maximum attainable power for a testing problem at each alternative parameter value. The difference is that the median envelope is a quantile function while the power envelope is a probability function.

We use the simple case of a t test (a one-sided Wald test) to illustrate these two ways of constructing a MU estimator. Assume test statistic $t=(\hat{\theta}-\theta_0)/SE(\hat{\theta})$ is used to test $H_0{:}\theta=\theta_0$ against a one-sided alternative $H_1{:}\theta<\theta_0$. The fixed-point inversion method then involves computing (or simulating) the median function for a t-test statistic at a fixed null point θ_0, i.e. $m(\theta)=med[(\hat{\theta}-\theta_0)/SE(\hat{\theta})|\theta]$ and then the MU estimate is given by $\hat{\theta}_{MU}=m^{-1}[(\hat{\theta}-\theta_0)/SE(\hat{\theta})]$. In contrast, the grid inversion method requires computing (or simulating) the 'median envelope', i.e. $m_E(\theta)=med[(\hat{\theta}-\theta)/SE(\hat{\theta})|\theta]$ and then the MU estimate is obtained by solving the equation $(\hat{\theta}-\theta)/SE(\hat{\theta})=m_E(\theta)$ for θ. In some simple cases, the two methods coincide with each other. For example, if the test statistic $\hat{\theta}-\theta_0$ is used, the fixed-point inversion and the grid inversion will lead to the same estimate. But in most other cases, the two methods are different.

If we compare the two methods of inverting a test statistic, the fixed-point inversion method implicitly makes the assumption that the median function is of the same shape and parallel to each other for different null values, which is true in many simple test procedures. Therefore it will lead to the same estimate if a test statistic for a different null hypothesis is inverted. In time series models, when the t test statistic is not distributed as Student's t and asymptotic normality does not approximate the null distribution well (e.g. see Nankervis and Savin, 1985, 1987, 1988b), the median function may not be of the same shape for different parameter values. In these situations, the grid inversion method is based on the more realistic assumption that the median function depends on the null value and therefore is more likely to deliver accurate estimates.

More importantly it is quite usual to find a test with non-monotonic median functions in small samples, and as a result, the fixed-point inversion will fail to produce unique estimates. The grid inversion may provide a remedy in this situation. In Chapter 5, we will show that the fixed-point inversion fails in many of the examples we examine while grid inversion may still work well.

After this chapter was first drafted, we became aware of the grid bootstrap suggested by Hansen (1999), which is similar to the proposed Method 2. We coined the term grid inversion similar to the terminology used in his paper. He applied his grid bootstrap method to confidence interval calculation in the first-order autoregressive model with a local-to-unity parameterisation and showed that the interval based on the usual bootstrap method fails to cover the true value at the nominal level (even asymptotically) while the interval based on grid bootstrap is correct to the order of n^{-1}. However, the quantile functions of the limiting distribution of the t test were found to be non-monotonic. In Chapter 5, we show that his method may not be able to deliver unique point estimates for the autoregressive parameter.

However, the fixed-point inversion method does enjoy the advantage of computational simplicity. It does not require solving equations. A simple tabulation or graph can be used to find the point estimates, as shown in Stock (1991). This is also why this method dominates the test inversion confidence intervals in the literature.

3.4.5 The Use of Optimal Invariant Tests

It is now clear that when choosing a test statistic to construct MU estimators, it is important to choose one that has good power properties in small samples. In this section, we consider two classes of tests that were shown to have such properties when a UMP test does not exist, namely, the point optimal invariant (POI) tests developed by King (1985a, 1987b) and the locally best invariant (LBI) tests advocated by King and Hillier (1985). Both these classes of tests possess some optimality properties that make them attractive in some circumstances especially in small samples. One would expect that the good small sample power properties of these tests will make them good candidates when we choose a test statistic to construct a MU estimator.

Generally speaking, the POI test is designed to maximise the power within the class of invariant tests in a neighbourhood of a preselected alternative point while the LBI test is aimed at maximising the power in the neighbourhood of the null hypothesis. As well as being most powerful at some points in the alternative hypothesis parameter space, these tests may also have optimum power at a number of other points and indeed be uniformly most powerful when such a test exists (see examples provided in King (1987b) and Hillier and King (1985)). MU estimators can be constructed based on both the LBI tests and POI tests based on the methods discussed in the previous section.

Consider the linear regression model

$$y = X\beta + u, \tag{3.63}$$

where y is the dependent variable, X is a $n\times k$ matrix of observed values of the exogenous regressors, β is a $k\times 1$ vector of fixed coefficients, and u is a $n\times 1$ vector of random disturbances. We are interested in the hypothesis testing problem that involves testing

$$H_0{:}u \sim N(0,\sigma^2 I_n) \text{ against } H_1{:}u \sim N(0,\sigma^2\Sigma(\theta)), \tag{3.64}$$

where Σ is a positive definite matrix known subject to the parameter vector of interest θ, which is to be estimated. As pointed out by King and Hillier (1985), this problem is interesting because a large number of hypothesis testing problems in linear regression analysis can be parameterized in this manner. In addition to AR(1) disturbances problems, these include simple pth order autoregressive disturbances, first-order moving average disturbances, various parametric forms of heteroscedasticity in the disturbances, random regression coefficients under different assumptions and various error component models. The estimation procedures outlined below can potentially be applied to all these models.

King (1980) showed that this testing problem is invariant to transformations of the form

$$y \rightarrow \eta_0 y + X\eta, \tag{3.65}$$

where η_0 is a positive scalar and η is a $k \times 1$ vector. Under this group of transformations, the maximal invariant vector is given by

$$v = Pz/(z'P'Pz)^{1/2} \tag{3.66}$$

where z is the OLS residual vector and P is an $m \times n$ matrix such that $PP' = I_m$ and $P'P = M$, with $M = I_m - X(X'X)^{-1}X'$ and $m = n-k$.

For the problem of testing H_0: $\theta = 0$ against the specific alternative H_1: $\theta = \theta_1 > 0$, the Neyman-Pearson lemma applied to the density of the maximal invariant (3.66) yields critical regions of the form

$$v'(P_1\Sigma(\theta_1)P_1')^{-1}v < c_1, \tag{3.67}$$

where c_1 is a constant. King (1980, Lemma 2) shows that this test can also be written as

$$s = \tilde{u}'\Sigma(\theta_1)^{-1}\tilde{u}/z'z, \tag{3.68}$$

where $\tilde{u}$ is the GLS residual vector assuming covariance matrix $\Sigma(\theta_1)$.

In most cases, the uniformly most powerful invariant test for H_0: $\theta = 0$ against H_1: $\theta > 0$ does not exist as the critical region (3.67) usually depends on θ_1. There are two ways to proceed in order to construct tests with well-defined optimality properties: to construct LBI tests or POI tests. The former class of tests were shown by King and Hillier (1985) to exist as long as $\Sigma(\theta)$ has a first-order derivative at θ_0. They are LBI in the sense that the power function has the maximum slope at the origin among all invariant tests. The second approach is designed to maximise the power at a pre-selected alternative point. The existence of such tests depends on the form of $\Sigma(\theta)$.

For different models, in order to compute test statistics such as (3.68), it is often appropriate to consider the transformed model

$$y(\theta) = X(\theta)\beta + u(\theta) \tag{3.69}$$

where $y(\theta) = \Sigma(\theta)^{-1/2}y$, $X(\theta) = \Sigma(\theta)^{-1/2}X$ and $u(\theta) = \Sigma(\theta)^{-1/2}u \sim N(0,\ \sigma^2 I)$. It was shown in King and Hillier (1985), Shively et al. (1990) and Dufour (1990) that the LBI tests can be expressed as

$$s_{LBI}(\theta_0) = \hat{u}(\theta_0)'\Sigma^{1/2}(\theta_0)'A\Sigma^{1/2}(\theta_0)\hat{u}(\theta_0)/\hat{u}(\theta_0)'\hat{u}(\theta_0), \tag{3.70}$$

where $A(\theta) = -\left.\dfrac{\partial\Sigma^{-1}(\theta)}{\partial\theta}\right|_{\theta_0}$, while the POI tests can be expressed as

$$s_{POI}(\theta_0, \theta_1) = \hat{u}(\theta_1)'C\hat{u}(\theta_1)/\hat{u}(\theta_0)'D\hat{u}(\theta_0) \tag{3.71}$$

where C and D are fixed matrices (possibly functions of θ_0 or θ_1) depending on the testing problem, and $\hat{u}(\theta_i) = (I - X(\theta_i)(X(\theta_i)'X(\theta_i))^{-1}X(\theta_i)')y(\theta_i)$, for $i = 0,\ 1$.

While LBI tests have optimal power in the neighbourhood of the null hypothesis, they may have poor power away from the null. In terms of most statistical loss functions (except the 0-1 loss), accepting the null when the parameter is far from the null is most damaging. This drop in power phenomenon was reported in Kramer (1985), Kramer and Ziesel (1990) and Bartels (1992) among others for the linear regression model with AR(1) disturbances. The argument was also supported by Dufour and King (1991), who, based on Monte Carlo evidence, concluded that when testing for correlated disturbances or random walk disturbances in the linear regression model, the POI tests are generally more powerful in small samples than the LBI tests over the whole parameter space under the alternative hypothesis. Therefore, in the sections that follow, we focus our attention on the POI tests. But the outlined method of inverting the median function or the median envelope also applies to the LBI tests.

The POI tests have been used effectively in testing for autocorrelation in the linear regression model (King, 1985a), testing for random walk disturbances (Dufour and King, 1991), testing for heteroscedastic disturbances (Evans and King, 1985, 1988), testing for fourth-order autoregressive disturbances (King, 1984), testing for deterministic trend and seasonal components (Franzini and Harvey, 1983 and Nyblom, 1989), testing for Hildreth-Houck random coefficients in the linear regression model (Milan, 1984, King, 1987c), testing for random walk coefficient (Brooks, 1993, Brooks and King, 1994, Shively, 1988) and testing for MA(1) errors against AR(1) errors in the linear regression model (King, 1983, 1985b, King and McAleer, 1987, Silvappulle and King, 1991). Therefore the method of constructing MU estimators outlined here can be potentially applied to estimating the error covariance structure in all these models. For example, Stock and Watson (1998) considered the POI tests and inverted their asymptotic quantile functions to construct MU estimators in the time-varying coefficient model in which the coefficient follows a random walk.

Another interesting feature of the class of POI test is that they can be used to trace out the finite sample maximum attainable power envelope for a certain class of hypothesis testing problems, thus providing a benchmark against which test

procedures can be evaluated. This is sometimes more accurate and meaningful than the asymptotic power envelope. See for example, Podivinsky and King (2000) and also Elliot et al. (1996). This helps us explain why the gird inversion based on the median envelope is able to better explore the small sample power advantage of the POI test when constructing a MU estimator.

When we apply the two methods of inverting a test statistic to the POI test, it is usually more convenient to consider the alternative value rather than the null value when we decide whether to use the fixed-point inversion or the grid inversion. Assume that the covariance matrix Σ is indexed by a scalar parameter θ, then a MU estimator of θ can be constructed in the following two ways. If the median function of the POI test is monotonic, we apply the fixed-point inversion for a pre-selected alternative point θ_1, i.e.,

$$\hat{\theta}_{MU} = m^{-1}(\hat{u}(\theta_1)' A(\theta_1)\hat{u}(\theta_1)/\hat{u}(\theta_0)' B(\theta_0)\hat{u}(\theta_0)). \tag{3.72}$$

The median function $m(\bullet)$ is given by,

$$m(\theta) = med[s_{POI}(\theta_0,\theta_1)|u \sim N(0,\sigma^2\Sigma(\theta)]. \tag{3.73}$$

Based on the good small-sample power properties of the POI tests, we would expect they are more likely to have monotonic median functions according to Theorem 3.4.1. However, if the median function of a single POI test is not monotonic, we can apply the grid inversion method to the series of POI tests which allow the *alternative value* θ_1 to vary within the alternative parameter space. Hence we solve for $\hat{\theta}_{MU}$ in the equation,

$$\begin{aligned}&\hat{u}(\theta)' A(\theta)\hat{u}(\theta)\big/\hat{u}(\theta_0)' B(\theta_0)\hat{u}(\theta_0)\\&-med[s_{POI}(\theta_0,\theta)|u \sim N(0,\sigma^2\Sigma(\theta)] = 0.\end{aligned} \tag{3.74}$$

The median function in (3.74) is also a *median envelope* of the POI statistics for testing for hypotheses (3.64). We will further examine the properties of the median envelope in Chapter 5 for the estimation of the linear regression model with

AR(1) or random walk disturbances. We show that the grid inversion of POI tests is a more reliable way to construct MU estimators than inverting a single POI test statistic for some design matrices.

Although not an idea pursued here, confidence intervals can also be constructed based on the procedures we developed. Two approaches are available: 1. Replacing the median function in the procedures by the $\alpha/2$- and $(1-\alpha)/2$-quantile functions of the test statistics and solving for the two confidence limits. Confidence intervals based on test inversion have been used frequently in econometrics (e.g., see Stock (1991), Andrews (1993), Kabaila (1993a) and Carpenter (1999)). 2. Bootstrap confidence intervals based on the proposed estimators, which involves approximating the distribution function of the estimator and constructing intervals based on the approximated quantiles (e.g., see discussions in Beran (1987), Hall (1988, 1994) and Efron and Tibshirani (1993)). We choose the percentile-*t* method and apply it to the dynamic linear regression model in Chapter 4, as it avoids the difficulty of computing the quantile functions needed in the first method.

3.4.6 Nuisance Parameters and Computation Issues

Nuisance parameters exist in most hypothesis testing problems. Popular methods to eliminate these parameters when constructing tests are those of similar tests and invariant tests. If such tests are available for the parameter of interest, the distribution of the test statistic, and therefore the median function of the test statistic will be invariant to nuisance parameters. We then only need to apply the methods described above to compute MU estimators. But in many cases, it is difficult to find either similar tests or invariant tests. In this situation, we need to use algorithms similar to the one outlined in Section 3.3.6, to iteratively invert non-similar test statistics while replacing the nuisance parameters in the process by their estimates.

Assume the parameter space can be partitioned into $\theta=(\beta,\gamma)'$, and a test $s(\theta)$ is designed to test hypotheses about β only, i.e., γ is the nuisance parameter.

Without loss of generality, we assume that given an estimate of β, there is an explicit way to compute a corresponding estimate of γ, i.e.

$$\hat{\gamma}=\hat{\gamma}(\hat{\beta}). \tag{3.75}$$

Start with a starting point $\hat{\theta}_1=(\hat{\beta}_1,\hat{\gamma}(\hat{\beta}_1))'$, invert the test statistic as if $\hat{\gamma}(\hat{\beta}_1)$ were the true value of γ, i.e.

$$\hat{\beta}_2=m_1^{-1}[s(\theta)] \tag{3.76}$$

where

$$m_1(\beta)=med[s(\theta)|y\sim f(\beta,\hat{\gamma}(\hat{\beta}_1))]. \tag{3.77}$$

We then replace the starting value by $\hat{\theta}_2=(\hat{\beta}_2,\hat{\gamma}(\hat{\beta}_2)$ and repeat the steps. The process is continued till convergence (i.e., the difference between two consecutive estimates becomes less than a pre-determined margin of error). As a result of this, the final estimator will only be approximately MU.

In the proposed procedure, the median function has to be calculated and inverted separately for each required sample size. If nuisance parameters are to be replaced by their consistent estimators, the median function has to be calculated for each different set of estimates used. This can be quite computationally cumbersome. Although not recommended, if the limiting distribution of $s(\theta)$ still depends on the parameter of interest, then based on Corollary 3.4.1, the procedure could be simplified by just inverting the median function of the limiting distribution, so the median function needs to be calculated just once. In the models we examine in this thesis, the median functions of many test statistics can be calculated exactly by using Imhof's (1961) algorithm to any desired level of accuracy.

3.5 Concluding Remarks

In this chapter, we outlined two general methods of constructing MU estimators in econometric models. One is based on adjusting the estimating equations and the second is based on inverting the median function of a significance test. Both methods can be regarded as different ways of exploring the relationship between the sufficient statistics (or less optimal statistics when sufficient statistics are intractable) and MU estimators, which was established by Lehmann (1959).

When an estimating equation generates a biased estimator, one can effectively adjust this equation to reduce the estimation bias. The conditions for a MU estimating equation to deliver a MU estimator are more general than those for a mean-unbiased estimating equation to produce a mean-unbiased estimator. So for a given estimating equation, we suggest subtracting its median function from the original estimating function and if the difference is a monotonic function, we will get a MU estimator. No analytical or simulated bias function is required, although sometimes it is hard to verify the monotonicity of the new estimating function. The relationship between this proposed bias prevention method and two other bias reduction techniques was disclosed. In Chapter 4, we give two examples of applying this method to the marginal likelihood score function in the linear regression model with AR(1) disturbances and the dynamic linear regression model.

In case it is too complicated to adjust the estimating equations or the monotonicity of the adjusted estimating function does not hold, a MU estimator can be constructed by inverting a significance test at the 50% significance level. Depending on whether the median function of the test statistic is monotonic, two different methods are considered: fixed-point inversion and grid inversion. The latter is theoretically more reliable but does involve extra computational costs. We also recommend inverting the POI test statistics mainly because of the sound small sample power properties. In Chapter 5, we will apply this approach to the linear regression model with AR(1) or random walk disturbances. The relationship between the power performance of a test and the efficiency of the MU estimator based on inverting its median function will be examined more thoroughly.

Chapter 4

Adjusting Marginal Likelihood Scores for Median-unbiased Estimators [1]

4.1 Introduction

In Chapter 3, we outlined the method for constructing MU estimators based on adjusting estimating equations. In this chapter, this method is applied to two commonly used time series models: the linear regression model with first-order autoregressive disturbances and the dynamic linear regression model. For the first model, it is shown that the proposed estimator of the autoregressive parameter is almost free of small sample median bias. In the second model, we extend the use of the proposed method to cover the case in which nuisance parameters cannot be eliminated from estimating equations through invariance arguments. The method is slightly revised to overcome this difficulty and an approximately MU estimator is derived.

Bias has always been a serious problem in the estimation of autoregressive time series models. Many bias-correction techniques have been proposed (e.g., see Quenouille, 1949, Orcutt and Winokur, 1969, Shaman and Stine, 1988, Fuller, 1996, MacKinnon and Smith, 1998 and Patterson, 2000, among others). The proposed MU estimator will serve as an alternative bias-prevention device for these models. It is also expected that through these examples, we will show that constructing a MU estimator by adjusting estimating equations can be a simple, yet effective way of bias-correction in the estimation of time series models.

MU estimation has already been applied to a first-order autoregressive model with only a constant and/or a linear time trend as the regressors. We extend the application by including other exogenous regressors. It is well known that the

[1] Some of the findings contained in this chapter were published in two conference proceedings, see Chen and King (1998, 2000).

properties (e.g., small sample bias) of the estimators will then depend on the structure of the design matrix as well as the magnitude of the autoregressive coefficient. A bias correction that works for different design matrix structures would be highly desirable.

The focus on point estimation in our study may need some justification. In the model of concern, especially when the persistence in the time series is very strong, i.e., when the autoregressive parameter or the lagged dependent variable coefficient is a large positive value close to 1, hypothesis testing may suffer from size distortion and/or low power problem (e.g., see Dickey and Fuller, 1981, Evans and Savin, 1981, 1984, Nankervis and Savin, 1985, 1987, 1988b, Schwert, 1989 and Magee, 1989), while confidence intervals may have bad coverage probabilities (e.g., see Stock, 1991 and Hansen, 1999) at the same time. In this situation, the impartiality of a point estimator becomes a crucial issue (see also discussions in Andrews, 1993, Stock, 1994 and Maddala and Kim, 1998). In some cases, it can provide an important insight into the validity of the model specification, and therefore straightforward evidence for researchers to choose the most likely model based on the data, when other inference tools are not reliable. Therefore median-unbiasedness is not only relevant, but can be indispensable in these circumstances.

In both models, we choose to adjust the marginal likelihood score equations, as they enjoy some distinctive advantages over the traditional profile likelihood approach both theoretically and empirically in small samples (see for example, Tunnicliffe-Wilson, 1989, Ara, 1995, Grose, 1998 and Mahmood, 2000). But due to the likelihood function being discontinuous in the neighbourhood of unity of the autoregressive parameter, we only consider the stationary case in this chapter. The unit root case will be treated in Chapter 5.

We aim to show that the proposed MU estimators enjoy two important advantages over the existing bias-reduction techniques. 1. Unlike other MU estimation methods (e.g., Andrews, 1993, Stock, 1994 and Fuller, 1996), the proposed method can be applied to models including a wide range of exogenous regressors. 2. The reduction in bias will not be offset by an increase in the variance. Hence the proposed MU estimators have root mean squared errors (RMSE) lower

than or similar to those of the conventional ML estimators in most cases. This is in contrast to some mean-bias-corrected estimators that may suffer from a trade-off between a reduced bias and an increased RMSE (e.g., see discussions in Orcutt and Winokur, 1969, MacKinnon and Smith, 1998 and Patterson, 2000).

Throughout the chapter, we focus on the small sample behaviour of the estimators. We attempt to achieve median-unbiasedness or approximate median-unbiasedness in small samples. Due to the difficulty in solving non-linear estimating equations and also the lack of analytical forms for the median functions, Monte Carlo simulations are used to investigate and compare the properties of different estimators. This has become a more and more popular tool used by econometricians especially when the finite sample feature is the major concern in research. Fortunately, existing algorithms often facilitate exact numerical calculations of the median functions in the models considered, which helps to alleviate the variability that may be associated with the results solely based on simulations.

The chapter is organised as follows. In Section 4.2, we specify the linear regression model with stationary AR(1) disturbances and briefly review the existing MU estimators in the literature. We then seek to adjust the marginal likelihood score towards median-unbiasedness. A Monte Carlo study is conducted to compare the new estimator with the conventional alternatives. In Section 4.3, we consider the dynamic linear regression model. An iterative bias-correction algorithm is developed to estimate the lagged dependent variable coefficient. The small sample properties of the new estimator are compared with those of the OLS estimator via Monte Carlo simulations. The chapter ends with some concluding remarks in Section 4.4.

4.2 MU Estimation of the Linear Regression Model with AR(1) Disturbances

4.2.1 Model Specification and Existing MU Estimators

There are two reasons why we choose the linear regression model with AR(1) disturbances: First it seems to have been the most popular time series model used in economic applications. A good understanding of this model is fundamental to studying more complicated time series models. Secondly, small-sample estimation bias has been so well documented for this model and so many remedies have been proposed, that it is natural and convenient to compare the proposed techniques with others such that we can assess the proposed approach in the broad context of bias-reduction.

This model can be stated as follows:

$$y_t = x_t'\beta + u_t, \; (t = 1, \ldots, T) \tag{4.1}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \; \varepsilon_i \text{ iid } N(0, \sigma^2), \tag{4.2}$$

where y_t is the dependent variable (observed at time t), x_t is a $k \times 1$ vector of fixed regressors, β is a $k \times 1$ vector of fixed coefficients, and u_t is a random disturbance. The coefficients β, ρ and σ^2 are unknown. We assume the disturbances follow a stationary AR(1) process with the initial condition:

$$u_0 \sim N[0, \sigma^2/(1-\rho^2)]. \tag{4.3}$$

For a model without exogenous regressors other than an intercept, the small sample bias of the least squares (LS) estimator $\hat{\rho}_{OLS}$ has been well documented. Quenouille (1949), Hurwicz (1950a), Marriott and Pope (1954), and Kendall (1954) all established the mean-bias of the LS estimator in a model with or without an intercept. They showed that the first-order mean-bias of $\hat{\rho}_{OLS}$ is $-2\rho/T$ for the model without the intercept, and $-(1+3\rho)/T$ for the model with an intercept. Le Breton and Pham (1989) calculated the exact and asymptotic biases of the same estimator in a stationary, unit root or an explosive AR(1) model without an intercept. Shaman and Stine (1988) established the first-order mean-bias of the LS estimator in a stationary AR(p) model. More generally, exact moments of the LS estimator were considered by Sawa (1978), Maekawa (1983) and Nankervis and Savin (1988a) among others.

If there are explanatory variables included in the model, no analytical approach is available to derive a general formula for the bias in the estimator as the bias also depends on the structure of the regressor matrix. Instead, many researchers conducted Monte Carlo studies to examine the small sample bias of different estimators proposed for this model. This was pioneered by the seminal paper by Orcutt and Winokur (1969). Simulation studies comparing different estimators in the model were also reported in Rao and Griliches (1969), Spitzer (1979), Park and Mitchell (1980), Kobayashi (1985), Nankervis and Savin (1988a) among others. The popular estimators of ρ compared in these studies include the Cochrane and Orcutt (1949) estimator, Durbin's (1960) estimator, Prais and Winsten's (1954) estimator and the full maximum likelihood estimator suggested by Beach and MacKinnon (1978). It was found that all these estimators, although unbiased asymptotically, are biased in small samples.

Attempts have been made to correct the small sample bias in the estimation of this model. Quenouille (1949, 1956) introduced a jackknife estimator in the model without an intercept. Orcutt and Winokur (1969) suggested using $\tilde{\rho} = (T\hat{\rho}_{OLS} + 1)/(T-3)$ as a bias-corrected estimator in a model with only an intercept. Bias reduction in autoregressive models without exogenous regressors was also considered by Shaman and Stine (1988), Rudebusch (1993) and recently by Patterson (2000). Most of these studies involve correcting the bias by subtracting the approximated bias function from the original estimate. But as observed in Rao and Griliches (1969), Orcutt and Winokur (1969), and MacKinnon and Smith (1998), sometimes it is not very effective to improve the performance by adjusting an estimator upwards for its known downward bias, because the reduction of bias may lead to an increase in variance with little or no improvement in RMSE.

A different, yet effective way of improving the estimation quality of this model is the use of the marginal likelihood (MGL) approach. This will be discussed in more detail in the next section. Ara (1995) and Laskar and King (1998), among others, provided Monte Carlo evidence and showed that the MGL estimator of the autoregressive parameter is generally less biased in small samples compared with the estimator based on the profile likelihood. Test procedures based on the MGL approach were also developed. Among the existing estimators, the MGL estimator seems to be the winner in terms of small sample bias.

So far MU estimators for this model in the literature have been mainly designed to estimate the simple first-order autoregressive model. It has been extended, with properly chosen de-meaning or de-trending procedures, to models with a drift and/or a linear time trend. As reviewed in Chapter 2, the existing MU estimators for a simple autoregressive model include: 1. Andrews's (1993) estimator which is based on inverting the median function of the OLS estimator, 2. Stock's (1991) estimator based on the limiting distribution of the Dickey-Fuller test and the Sargan-Bhargava test statistics, and 3. Fuller's (1996) estimator based on the limiting distribution of the weighted symmetric least-squares estimator.

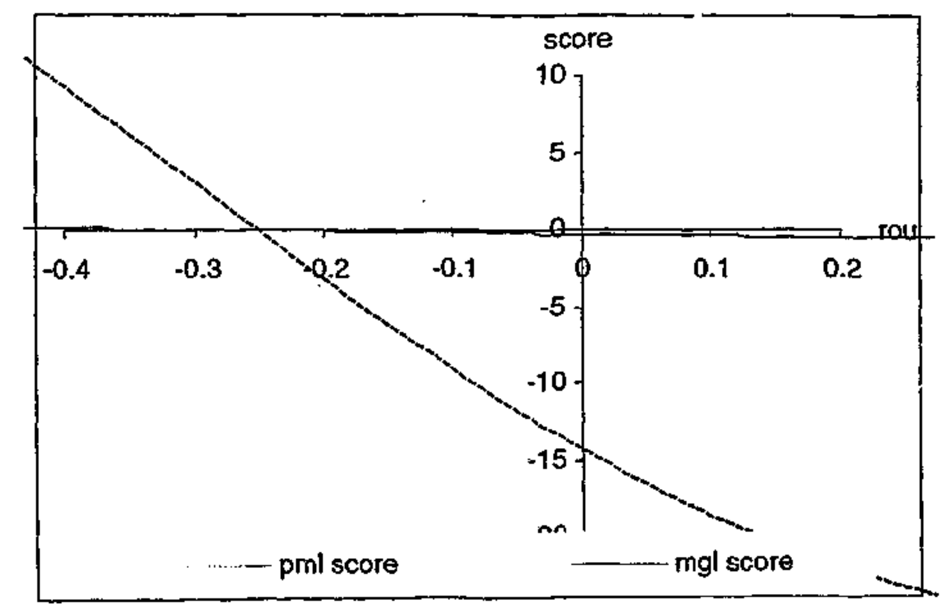
All these existing MU estimators do not address the exogenous regressor issue. When we attempt to extend their use to the linear regression model with AR(1) disturbances, it is not known what impact different design matrices might have on their small-sample properties. In the most commonly used two-step estimation procedures, people usually treat the model as an OLS regression first, and then treat the OLS residuals as if they were the true disturbances for estimating the autoregressive coefficient. But for some design matrices, e.g., Watson's X matrix, the OLS residuals can be bad approximations to the true disturbances, which will surely result in bad estimates of the autoregressive coefficients, see e.g., King (1985a). Therefore a different approach is needed to count for the impact from the exogenous regressors on the bias function of the estimator. Our aim is to find a MU estimator that is robust to different design matrices.

4.2.2 Adjusting Marginal Likelihood Score Equations

In this section, we construct a MU estimator in model (4.1) and (4.2) by adjusting an estimating equation. A natural choice of estimating equation would be the score equation based on the profile likelihood function discussed in Beach and MacKinnon (1978) among others. But in our study, we found that this score and its

median function are both very flat and close to the horizontal axis. Figure 1 shows one realisation of the scores based on the profile likelihood and the marginal likelihood at $\rho = 0$ for a model with an intercept and a time trend. Compared with the profile likelihood score, the MGL score is much steeper. Because of the flatness, when we implement the proposed adjustment to the profile likelihood score, the convergence will be very slow when solving the adjusted equation (3.10).

Figure 4.1 One Realisation of the Marginal and Profile Likelihood Scores in the Linear Regression with AR(1) Disturbances for Design Matrix $X1$; $T = 20$, $\rho = 0$



On the other hand, since Kalbfleisch and Sprott (1970, 1973) argued for the use of the marginal likelihood in time series models, evidence has been found (see for example, Tunnicliffe-Wilson (1989), Ara (1995) and Laskar and King (1998)) that the marginal likelihood approach can effectively eliminate nuisance parameters without losing information when estimating the error structure in a linear regression model. The MGL estimator was also shown to be able to reduce the small sample bias of the maximum profile likelihood estimator. Ara (1995) also showed that there is a logical relationship between marginal likelihood and sufficient statistics. Therefore using the marginal likelihood is consistent with the approach based on sufficient statistics discussed in Section 3.2. So in this section, we choose to calibrate the marginal likelihood score equation.

We start with examining the rationale behind the marginal likelihood approach in the general linear regression model

$$y = X\beta + u \tag{4.4}$$

where y is the dependent variable, X is a $T\times k$ matrix of observed values of the exogenous regressors, β is a $k\times 1$ vector of fixed coefficients, and u is a $T\times 1$ vector of random disturbances distributed as $N(0,\sigma^2\Omega(\theta))$. If $\theta=(\theta_1,\theta_2,\ldots,\theta_p)'$ are the only parameters of interest, we can construct a maximal invariant statistic d under a certain group of affine transformations (for details, see King, 1980, Ara, 1995, and Rahman and King, 1997), given by

$$y \to \eta_0 y + X\eta,$$

where η_0 is a positive scalar and η is a $k\times 1$ vector. Following Ara (1995), the often-used maximal invariant d is

$$d = Pz/(z'P'Pz)^{1/2}, \tag{4.5}$$

where $z = My$ is the vector of OLS residuals, P is a $m\times n$ matrix such that $PP' = I_m$ and $P'P = M$, with $M = I_m - X(X'X)^{-1}X$ and $m = T-k$. The density function of the maximal invariant vector is given by

$$f(d,\theta) = \frac{1}{2}\Gamma\left(\frac{m}{2}\right)\pi^{-\frac{m}{2}}\left|P\Omega(\theta)P'\right|^{-\frac{1}{2}}\left(\frac{\tilde{u}'\Omega^{-1}(\theta)\tilde{u}}{z'z}\right)^{-\frac{m}{2}}. \tag{4.6}$$

It is apparent that the distribution of d is invariant to both nuisance parameters, β and σ^2. But more importantly, as pointed out by Ara (1995), d is so-called G-sufficient for θ, in the sense that for any inference about θ, it is sufficient to study the distribution of d without information loss.

Therefore for our model (4.1) – (4.3), based on the density function (4.6), the marginal log likelihood function for ρ can be written as:

$$l_m(y;\rho) = c - \frac{1}{2}\log\left|\Omega(\rho)\right| - \frac{1}{2}\log\left|X'\Omega(\rho)^{-1}X\right| - m\log(s^2), \tag{4.7}$$

where

$$\Omega(\rho) = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix} \tag{4.8}$$

and

$$\Omega(\rho)^{-1} = \begin{bmatrix} 1 & -\rho & \cdots & & 0 \\ -\rho & 1+\rho^2 & & & 0 \\ & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & 1+\rho^2 & -\rho \\ 0 & \cdots & & -\rho & 1 \end{bmatrix} \tag{4.9}$$

is the covariance matrix of u_t and its inverse, respectively, and

$$s^2 = (y - X\hat{\beta}_\rho)'\Omega(\rho)^{-1}(y - X\hat{\beta}_\rho), \tag{4.10}$$

where

$$\hat{\beta}_\rho = (X'\Omega(\rho)^{-1}X)^{-1}X'\Omega(\rho)^{-1}y. \tag{4.11}$$

Following Rahman and King (1997), the marginal likelihood score function is given by

$$U(y;\rho)=-\frac{1}{2}tr\left[\Delta(\rho)\frac{\partial\Omega(\rho)}{\partial\rho}\right]-\frac{T-k}{2}\left[\frac{\tilde{u}'\frac{\partial\Omega(\rho)^{-1}}{\partial\rho}\tilde{u}}{\tilde{u}'\Omega(\rho)^{-1}\tilde{u}}\right], \tag{4.12}$$

where

$$\Delta(\rho)=\Omega(\rho)^{-1}-\Omega(\rho)^{-1}X(X'\Omega(\rho)^{-1}X)^{-1}X'\Omega(\rho)^{-1}, \tag{4.13}$$

and

$$\tilde{u}=y-X(X'\Omega(\rho)^{-1}X)^{-1}X'\Omega(\rho)^{-1}y. \tag{4.14}$$

It was shown in Ara (1995) that the marginal likelihood score satisfies the mean-unbiasedness condition, i.e.,

$$E[U(y;\rho)|\rho_0]=0,$$

where ρ_0 is the true parameter value. But as revealed in Chapter 3, the unbiasedness of the score does not guarantee the unbiasedness of the estimator. As a matter of fact, as reported in Ara (1995) and Laskar and King (1998), the maximum marginal likelihood estimator is still biased in small samples. One possible reason for this bias is the curvature of the non-linear score function (4.12). Ara (1995) showed that the marginal likelihood score $U(y;\rho)$ is asymptotically normally distributed with the same limiting distribution as that of the profile likelihood score. This property makes it possible for us to apply the adjustment proposed in Chapter 3.

In order to adjust the score (4.12) according to (3.10), we need to compute its median function. Notice that the marginal likelihood score in (4.12) can be expressed as the form:

$$U(y;\rho)=A(\rho)+\frac{\varepsilon'B(\rho)\varepsilon}{\varepsilon'C(\rho)\varepsilon}, \tag{4.15}$$

where

$$A(\rho)=-\frac{1}{2}tr[\Delta(\rho)\frac{\partial\Omega(\rho)}{\partial\rho}],$$

$$B(\rho)=R(\rho)'\overline{P}(\rho)'\frac{\partial\Omega(\rho)^{-1}}{\partial\rho}\overline{P}(\rho)R(\rho), \tag{4.16}$$

$$C(\rho)=R(\rho)'\overline{P}(\rho)'\Omega(\rho)^{-1}\overline{P}(\rho)R(\rho), \tag{4.17}$$

in which

$$\overline{P}(\rho)=I-X(X'\Omega(\rho)^{-1}X)^{-1}X'\Omega(\rho)^{-1}, \tag{4.18}$$

and

$$R(\rho)=\begin{bmatrix} 1/\sqrt{1-\rho^2} & 0 & \cdots & 0 \\ \rho/\sqrt{1-\rho^2} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{T-1}/\sqrt{1-\rho^2} & \rho^{T-2} & \cdots & 1 \end{bmatrix}. \tag{4.19}$$

Therefore, subject to the constant $A(\rho)$, the score is a ratio of two quadratic forms in the normal error vector ε. So the median of the score function can be calculated exactly by solving

$$\Pr[\sum_{i=1}^{T}\lambda_i\xi_i^2\le 0]=0.5 \tag{4.20}$$

for m using Imhof's (1961) algorithm, where the λ_i's are the eigenvalues (including zeros and multiple roots) of $B(\rho)-(m-A(\rho))C(\rho)$, and ξ_i^2 are independent standard χ^2- variables with one degree of freedom.

Figure 4.2 shows the shape of several sample realisations of the score and its median function for design matrix $X2$ (see Section 4.2.3), with 20 observations and the true ρ at 0.2, 0.4 and 0.8. The maximum marginal likelihood (MMGL) estimates are the interceptions between the scores and the horizontal axis, while the new estimates which solve the adjusted estimating equation (4.22), are given by the interception between the scores and the median function. It is quite apparent that the new estimator should at least correct the downward bias of the MMGL estimator in the right direction.

Figure 4.2 Three Realisations of the New Estimator: an Illustration of the Proposed Bias-prevention Method for the Linear Regression Model with AR(1) Disturbances for Design Matrix $X2$, $T=20$, $\rho=0.2,\ 0.4,\ 0.8$

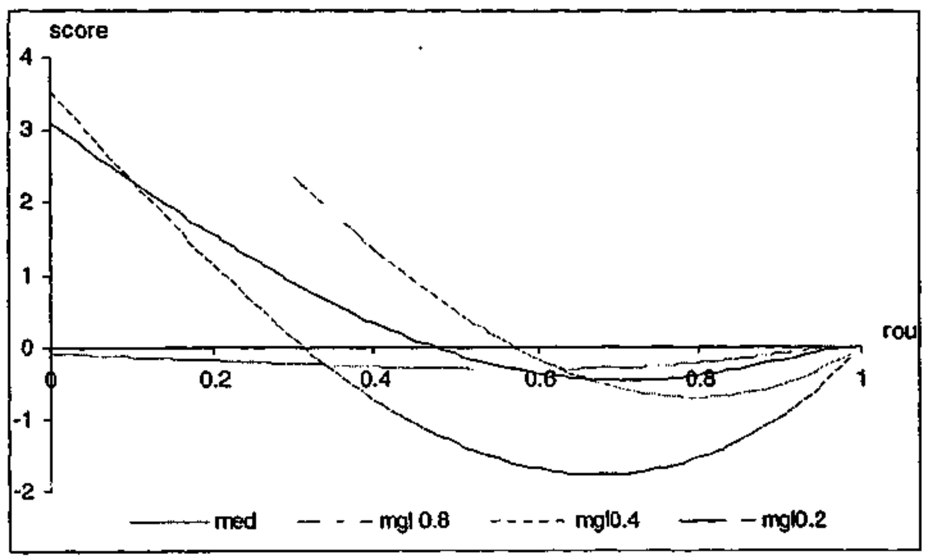
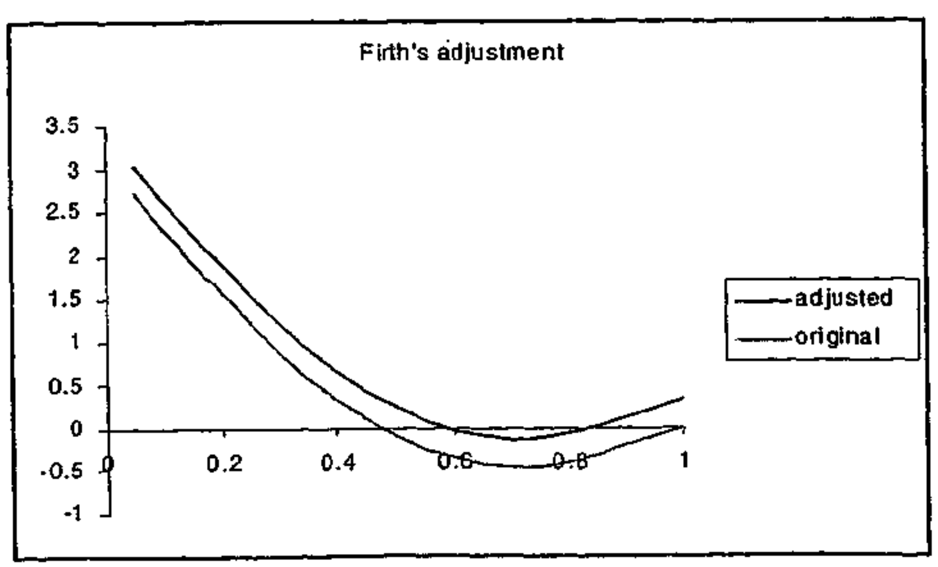


Figure 4.3 An Illustration of Firth's Bias-prevention Method for the Linear Regression Model with AR(1) Disturbances for Design Matrix $X1$, $T=20$, $\rho=0.6$



It is interesting to compare the proposed bias-correction method with Firth's (1993) method graphically. In Figure 4.3, Firth's adjusted score is plotted together with the original score. The bias reduced estimate based on his method is then the interception between the adjusted score and the horizontal axis. Intuitively, Firth's method attempts to prevent the bias by shifting the original score curve upward; while the proposed method shifts the horizontal axis downward to the position of the median function, in order to increase the value of the estimates. In Chapter 3, we proved that these two adjustments are equivalent to the order of n^{-1}. The benefit of the proposed approach, however, is that it does not require an approximation of the bias function or computation of the information matrix component.

Hence if we denote the median function of the score as

$$m(\rho) = med_{\rho}[U(y;\rho)], \tag{4.21}$$

the proposed estimator will be the solution to the adjusted estimating equation

$$U^{MU}(\rho) = U(y;\rho) - m(\rho) = 0. \tag{4.22}$$

The graph shows that the left-hand side of the above equation is not a monotonic function as required in Theorem 3.2.1. Both the score and median come across the axis at $\rho = 1$. This is partly because that we did not ignore the constant term $A(\rho)$ in the score (4.15) when solving equation (4.22). If $A(\rho)$ is discarded (which will not alter the solution to (4.22)), the median function will be monotonic and of a similar shape to the ones plotted in Figure 4.4 for the dynamic linear regression model. However, as shown in Figure 4.2, locally (in the neighbourhood of the interception), monotonicity of the LHS of (4.22) is guaranteed for the current approach. Therefore in practice, we need to select an initial value so that it falls into the region in which the adjusted estimating function is monotonic. If this is not achievable, it means there is no solution within the stationary region, and we take $\hat{\rho} = 1$ as the estimate.

As there is no closed form for $m(\rho)$, the equation (4.22) cannot be solved analytically. Therefore the iterative algorithm proposed in Section 3.3.3 has to be used. In our study we found that the speed of convergence of the proposed algorithm depends on the curvature of the score and the median function. In some cases, the two curves are almost parallel to each other and very close to the horizontal axis, so that the convergence can be very slow. But generally speaking, it should not be slower than the normal searching algorithm used for ML estimation. We will revisit this point when we discuss the estimation results in Section 4.2.4.

4.2.3 Experimental Design

An important aspect of our model specification is the structure of the exogenous regressors. We conducted our simulations based on the following two design matrices:

$X1$: An intercept and a linear time trend. This is a most commonly used design matrix in the unit root literature. We include it in order to compare the proposed MU estimator with the maximum marginal likelihood estimator, which is another bias-reduced estimator.

$X2$: The first five sets of eigenvectors corresponding to the eigenvalues (sorted in ascending order) of the matrix:

$$W = ((1+\rho^2)I - 2\rho\Theta)^{-1}, \tag{4.24}$$

where

$$2\Theta = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix},$$

These 5 eigenvectors are: $q_{i1} = \frac{1}{\sqrt{T}}$ and $q_{ij} = \sqrt{\frac{2}{T}} \cos\frac{(2i-1)(j-1)\pi}{2T}$, $i = 1, 2, \ldots, T$ and $j = 2, \ldots, 5$. This design matrix was first considered by Durbin and Watson (1950). In this case the DW test is the approximately uniformly most powerful test. King (1985a) and Dufour and King (1991) among others, used it as an extreme case in which the OLS residuals from the regression approximate the true disturbances almost as well as the GLS residuals. We included this design mainly to isolate the impact from the number of regressors in the model from the structure of the design matrix. In Chapter 5, we will include a range of other design matrices when we study the MU estimator based on inverting a significance test.

The numbers of observations used were 20 and 60. The true values of ρ were 0.95, 0.8, 0.6, 0.4, 0.2, 0, -0.2, -0.4, -0.6, -0.8 and -0.95. As the distribution of $\hat{\rho}$ is invariant to the values of β and σ^2, they were set to be one in the simulations. 2000 replications were conducted.

4.2.4 Results

The bias and risk of the OLS estimator $\hat{\rho}_{OLS}$, the profile likelihood estimator $\hat{\rho}_{FML}$, the maximum marginal likelihood estimator $\hat{\rho}_{MGL}$ and the proposed estimator $\hat{\rho}_{NEW}$ are reported in Table 4.1a and Table 4.1b.

Our results first verify the well-known bias problem associated with estimation of the autoregressive coefficient. The downward bias of the OLS estimator is more serious when there are more regressors in the model and/or for large positive ρ values. For example, for $\rho = 0.9$ and a sample size of 20, the median biases of $\hat{\rho}_{OLS}$ are -0.36 and -0.63 for $X1$ and $X2$, respectively. The bias decreases as the sample size increases but is still quite serious for a sample size of 60. It is also revealed that the estimator proposed by Beach and MacKinnon (1978) ($\hat{\rho}_{FML}$) is generally less biased than $\hat{\rho}_{OLS}$ but the difference is minimal particularly for a sample size of 20, while the marginal likelihood estimator $\hat{\rho}_{MGL}$ is able to reduce the bias of $\hat{\rho}_{OLS}$ quite significantly especially for $X1$ and a sample size of 60. But the bias is not completely eliminated. For a sample size of 20 and large positive ρ, there is still some room for improvement.

When the sample size is 60, for both design matrices, the proposed estimator, $\hat{\rho}_{NEW}$, successfully corrected the bias in $\hat{\rho}_{OLS}$. The new estimator is almost exactly median-unbiased for all ρ values. For example, when $\rho = 0.9$, the median biases of $\hat{\rho}_{OLS}$ are reduced from -0.1 and -0.14 down to zero when using $\hat{\rho}_{NEW}$ instead of $\hat{\rho}_{OLS}$ for $X1$ and $X2$, respectively, while the biases in the marginal likelihood estimator $\hat{\rho}_{MGL}$ are -0.02 and -0.03. For moderate positive and negative ρ values, the new estimator does not over-correct the bias of the OLS estimator. The median biases for all negative ρ values are essentially 0.

When the sample size is 20, the new estimator also successfully removed the bias in $\hat{\rho}_{OLS}$ for $X1$, while $\hat{\rho}_{FML}$ and $\hat{\rho}_{MGL}$ remain quite biased. The bias correction of the proposed method is quite substantial for large ρ values. For example, the median biases of $\hat{\rho}_{NEW}$ for $\rho = 0.95$, 0.9, and 0.8 are -0.02, -0.02 and 0, respectively; In contrast, the biases of $\hat{\rho}_{OLS}$ in this case are -0.38, -0.36 and -0.28, respectively. We notice that in this case the biases of $\hat{\rho}_{MGL}$ are -0.1, -0.09 and -0.05 while $\hat{\rho}_{FML}$ is essentially as biased as $\hat{\rho}_{OLS}$. For $X2$, however, when ρ falls into the neighbourhood of 1 (i.e., $\rho > 0.85$), although the biases in $\hat{\rho}_{OLS}$ and $\hat{\rho}_{FML}$ are significantly reduced, a small bias still remains. For moderate ρ values, however, $\hat{\rho}_{NEW}$ is able to correct the bias successfully. For example, the median biases of $\hat{\rho}_{NEW}$ for $\rho = 0.95$, 0.9, and 0.8 are -0.12, -0.07 and -0.01, respectively. In contrast, the biases in $\hat{\rho}_{OLS}$ are -0.68, -0.63 and -0.53. The other two estimators, $\hat{\rho}_{FML}$ and $\hat{\rho}_{MGL}$, did not correct bias in $\hat{\rho}_{OLS}$ effectively in this case and remain seriously biased.

The remaining bias in the new estimator for $X2$ when $T = 20$ may be caused by the following problem: we notice that for this design matrix, when ρ is close to 1, both the score and its median function become very close to 0 (i.e., almost flat and very close to the horizontal axis). When we try to solve equation (4.22) for the new

estimator, the iterative algorithm becomes less reliable. An improved algorithm may be needed for this case. This problem goes away when the sample size goes up to 60.

Another important feature of the proposed new MU estimator is its much smaller RMSEs compared with those of other estimators. This advantage is for all positive ρ values and for both design matrices. For example, when ρ is 0.95 and 0.9, the RMSEs of $\hat{\rho}_{NEW}$ are essentially 50% of those of $\hat{\rho}_{OLS}$ and $\hat{\rho}_{FML}$ for $X1$, and less than 50% for $X2$, for a sample size of 60. The MU estimator has slightly bigger RMSEs than other estimators for negative ρ values, results for which are arguably of less interest to econometricians than those for positive serial correlation

4.3 Approximately MU Estimation of the Dynamic Linear Regression Model

4.3.1 Dynamic Linear Regression Model

It has been argued by Dufour and Kiviet (1998), among others, that the linear regression with AR(1) disturbances is a model usually too simple to capture the real dynamic nature of the economic time series. Instead, the first-order dynamic linear regression model (ARX(1)) is a more powerful modelling device which is more consistent with the data when the relationship is genuinely dynamic (e.g., see also discussion in Hendry and Mizon, 1978).

Therefore in this section, we develop a MU estimation procedure for the coefficients of the model:

$$y_t = \gamma y_{t-1} + x_t'\beta + e_t,\ t = 1,\ \ldots,\ T, \tag{4.25}$$

where y_t is the dependent variable (observed at time t), x_t is a $k \times 1$ vector of fixed regressors at time t, β is a $k \times 1$ vector of coefficients, and the disturbances $e = (e_1, \ldots, e_T)' \sim N(0, \sigma^2 I_T)$. The coefficients γ, β and σ^2 are unknown.

The first-order dynamic linear regression model is frequently used in econometric practice. Examples of such applications, to name a few, include the studies of real wages and employment by Altonji and Ashenfelter (1980) and Geary and Kennan (1982), of consumption and income by Hall (1978) and Flavin (1981), and of aggregate price and output by Froyen and Waud (1984). As pointed out by Nankervis and Savin (1987) and Kiviet and Dufour (1997), empirical researchers still largely rely on the usual inference procedures (mainly based on the LS principles), as they are still asymptotically valid under certain regularity conditions, although the magnitude of the approximation errors is unknown. Dufour (1996) showed that the approximation error can be arbitrarily bad on certain subsets of the parameter space. The only comfort researchers may find is in the thought that the committed error will get smaller as the sample size gets larger. The relevant question how large a sample is needed in order to feel confident is usually left unanswered. Monte Carlo evidence reported in Kiviet (1985) and Nankervis and Savin (1985, 1987 and 1988b) among others confirmed that the error caused by using the usual inference procedures in small samples can in fact be quite substantial.

In particular, the existing popular estimators of γ, such as the OLS estimator, the three-pass least square estimator (Taylor and Wilson, 1964), the approximate MLE and the estimator suggested by Hatanaka (1974) are all biased in small samples. Attempts have been made to approximate the bias function of the OLS estimator and to correct the bias by subtracting the estimated bias from the original estimate, see e.g., Sawa (1978) and Grubb and Symonds (1987). As the bias function can only be approximated to a certain order (usually T^{-1} or T^{-2}), these bias corrections are not exact in small samples.

Kiviet and Phillips (1990, 1992) proposed an exact inference procedure obtained by applying least squares to an augmented regression model with artificial regressors introduced. Kiviet and Dufour (1997) and Dufour and Kiviet (1998) further developed this approach and extended its use to more general models. Exact similar tests and conservative confidence intervals were developed for γ. Simultaneous inference about γ and other regression coefficients were also considered. But the performance of the point estimator constructed via this approach

has not been supported by many simulation studies or empirical applications, hence its small sample bias and efficiency remain unclear. Andrews (1993) conjectured that a MU estimator can be constructed based on the exact similar test statistic developed by Kiviet and Phillips (1992), but the idea has not been pursued any further.

The application of the marginal likelihood approach in this model provides a different way to reduce bias and improve estimation efficiency. The application of the MGL approach in this model first appeared in Levenbach (1972) and Bellhouse (1978). The marginal likelihood function was derived based on the maximal invariant principle similar to the one outlined in Section 4.2.2. Recently, Grose (1998) and Mahmood (2000) examined marginal likelihood estimation in the dynamic linear regression model via simulation. They concluded that although the MGL score is not well-behaved in small samples, the MMGL estimator is less biased than its counterpart based on the profile likelihood. But the bias is not completely eliminated. The effectiveness of bias reduction depends on the design matrix and parameter values.

An important difference when applying the adjustment given by (3.10) to this model is that, the information about the lagged dependent variable coefficient cannot be completely isolated from those about the nuisance parameters as in the linear regression model with AR(1) disturbances. As a result, the distribution of the MMGL estimator depends on nuisance parameters. As the nuisance parameters cannot be eliminated by invariance arguments, the adjustment given by (3.10) has to be revised.

4.3.2 Marginal Likelihood Score

In order to derive the marginal likelihood function of γ, it is necessary to make further assumptions about y_1 so that the distribution of y can be determined. Following Nankervis and Savin (1985), Inder (1985, 1987) and King (1996), the stationarity conditions can be stated as:

1. The stability of $E(y_t)$ at $t=1$ and 2 such that $E(y_2)=E(y_1)$, and

2. $Var(y_t)$ is the same for all $t = 1,\ldots,T$.

Inder (1985) showed that the above conditions are observationally equivalent to assuming that y_1 is generated by

$$y_1 = x_1'\beta\big/(1-\gamma) + e_1\big/\sqrt{1-\gamma^2}. \qquad (4.26)$$

It should be noticed that there are other approaches to defining a distribution for y_1 which differ from using the mean-stationarity assumptions given above. For example, Grose (1998) discussed two other approaches to generating y_1, while Dufour and Kiviet (1998) adopted a more general set of initial conditions which allows y_1 to be either fixed or follow an arbitrary distribution.

Thus assuming $|\gamma| < 1$, model (4.25) and (4.26) can be written as

$$\Gamma(\gamma)y = X\beta + D(\gamma)e, \qquad (4.27)$$

where $\Gamma(\gamma)$ is the $T \times T$ matrix

$$\Gamma(\gamma) = \begin{bmatrix} (1-\gamma) & 0 & 0 & \cdots & 0 \\ -\gamma & 1 & 0 & & \\ \vdots & -\gamma & 1 & & \vdots \\ & & & \ddots\ 1 & 0 \\ 0 & \cdots & & -\gamma & 1 \end{bmatrix},$$

and $D = diag(\sqrt{1-\gamma\big/1+\gamma}, 1, 1, \ldots, 1)$. Equation (4.27) implies that

$$y = \Gamma^{-1}(\gamma)X\beta + \Gamma^{-1}(\gamma)D(\gamma)e, \qquad (4.28)$$

where

$$\Gamma^{-1}(\gamma) = \begin{bmatrix} 1\big/(1-\gamma) & 0 & \cdots & & 0 \\ \gamma\big/(1-\gamma) & 1 & 0 & & \vdots \\ \gamma^2\big/(1-\gamma) & \gamma & 1 & \ddots & \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \gamma^{T-1}\big/(1-\gamma) & \cdots & \gamma^2 & \gamma & 1 \end{bmatrix}.$$

If we let $X(\gamma) = \Gamma^{-1}(\gamma)X$ and

$$\Omega(\gamma) = \Gamma^{-1}(\gamma)D(\gamma)^2\Gamma^{-1}(\gamma)' = \frac{1}{1-\gamma^2}\begin{bmatrix} 1 & \gamma & \gamma^2 & \cdots\gamma^{T-1} \\ \gamma & 1 & \gamma & \cdots\gamma^{T-2} \\ \gamma^2 & \ddots & \ddots & \ddots\ \vdots \\ \vdots & \cdots & & 1\ \ \gamma \\ \gamma^{T-1} & \cdots & & \gamma\ \ 1 \end{bmatrix},$$

(4.28) can then be written as the general linear model

$$y = X(\gamma)\beta + u,\ u \sim N(0, \sigma^2\Omega(\gamma)). \qquad (4.29)$$

Bellhouse (1978) showed that the marginal likelihood function for γ is proportional to

$$L_m(\gamma; y) \propto \frac{|X'(\gamma)X(\gamma)|^{1/2}\{y'y - y'X(\gamma)(X'(\gamma)X(\gamma))^{-1}X'(\gamma)y\}}{|\Omega(\gamma)|^{1/2}|X'(\gamma)\Omega^{-1}(\gamma)X(\gamma)|^{1/2}s^{n-k}}, \qquad (4.30)$$

where

$$\begin{aligned} s^2 &= y'\{\Omega^{-1}(\gamma) - \Omega^{-1}(\gamma)X(\gamma)(X(\gamma)'\Omega^{-1}(\gamma)X(\gamma))^{-1}X(\gamma)'\Omega^{-1}(\gamma)\}y \\ &= \tilde{u}'\Omega^{-1}(\gamma)\tilde{u}, \end{aligned}$$

in which $\tilde{u}$ is the GLS residual vector from the regression of (4.29) assuming covariance matrix $\Omega(\gamma)$.

Grose (1998) derived the score function for γ based on the log of (4.30), which is given by

$$Q(\gamma)=tr\{(\overline{B}-\widetilde{B})X_\gamma(\gamma)\}-\frac{1}{2}tr\{\widetilde{P}\Omega(\gamma)\}$$
$$+\frac{2m(\widetilde{u}'\Omega^{-1}(\gamma)X_\gamma(\gamma)\widetilde{\beta})-m(\widetilde{u}'\Omega_\gamma^{-1}(\gamma)\widetilde{u})}{\widetilde{u}'\Omega^{-1}(\gamma)\widetilde{u}}+\frac{2(\hat{u}'X_\gamma(\gamma)\hat{\beta})}{\hat{u}'\hat{u}}. \qquad (4.31)$$

in which

$$\overline{B}=(X(\gamma)'X(\gamma))^{-1}X(\gamma)',$$
$$\widetilde{B}=(X(\gamma)'\Omega^{-1}(\gamma)X(\gamma))^{-1}X(\gamma)'\Omega^{-1}(\gamma),$$
$$\widetilde{P}=\Omega^{-1}(\gamma)-\Omega^{-1}(\gamma)X(\gamma)(X(\gamma)'\Omega^{-1}(\gamma)X(\gamma))^{-1}X(\gamma)'\Omega^{-1}(\gamma),$$
$$X_\gamma(\gamma)=\frac{\partial\Gamma^{-1}(\gamma)}{\partial\gamma}X=-\Gamma(\gamma)\frac{\partial\Gamma(\gamma)}{\partial\gamma}X,$$
$$\Omega_\gamma^{-1}=\frac{\partial\Omega^{-1}(\gamma)}{\partial\gamma},$$

and $\hat{\beta}=\overline{B}y$ and $\widetilde{\beta}=\widetilde{B}y$ are the OLS and GLS estimators of β respectively, while $\hat{u}=y-X(\gamma)\hat{\beta}$ and $\widetilde{u}=y-X(\gamma)\widetilde{\beta}$ are their corresponding residual vectors.

Grose (1998) showed that the expectation of the marginal likelihood score for γ is not 0 in small samples, i.e., the estimating function (4.31) is not mean-unbiased. This score is also not information biased in the sense that its variance is not equal to the information matrix. The small sample deficiencies of the score may lead to a less-than-ideal MMGL estimator for some design matrices, as reported in Grose (1998). One possible improvement is to calibrate the score towards mean-unbiasedness by using corrections such as the one proposed by McCullagh and Tibshirani (1991) (see also Mahmood (2000)). But this approach is subject to the criticism that there is a lack of link between mean-unbiasedness of an estimating equation and mean-unbiasedness of the resulting estimator, as discussed in Chapter 3. We argue that correcting the median bias of the score may be more effective.

4.3.3 Accounting for Nuisance Parameters

In order to adjust the score equation according to (3.10), we need to find the median function of $Q(\gamma)$, given by

$$m(\gamma)=med([Q(y;\gamma)|y\sim N(\Gamma^{-1}(\gamma)X\beta,\sigma^2\Omega(\gamma))]. \qquad (4.32)$$

In Section 4.2, we showed that in the linear regression model with AR(1) disturbances, the marginal likelihood score in equation (4.12) can be expressed as a ratio of two quadratic forms in normal random variables, therefore Imhof's (1961) algorithm can be used to compute the median function exactly. We notice that $Q(\gamma)$ in (4.31) is the sum of two such ratios plus a constant. Hence it is impossible to use Imhof's algorithm directly to compute its median function. It might be possible to derive the characteristic function of the sum of the ratios and numerically invert it to calculate the median. But it would be computationally costly. The other method is to approximate the median function via simulation. Our simulation results showed that the median function (4.32) is seriously non-monotonic in γ and it violates the conditions needed for (3.10) to deliver a unique estimate.

To overcome this difficulty, we slightly revise the proposed method. Instead of solving equation (3.10) for varying γ, i.e.,

$$Q(y;\gamma)-med[Q(y;\gamma)|y\sim N(\Gamma^{-1}(\gamma)X\beta,\sigma^2\Omega(\gamma))]=0,$$

we consider it at a fixed point $\gamma=\gamma_0$:

$$Q(y;\gamma_0)-med[Q(y;\gamma_0)|y\sim N(\Gamma^{-1}(\gamma)X\beta,\sigma^2\Omega(\gamma))]=0, \qquad (4.33)$$

which is equivalent to inverting the median function

$$m_{\gamma_0}(\gamma)=med[Q(y;\gamma_0)|y\sim N(\Gamma^{-1}(\gamma)X\beta,\sigma^2\Omega(\gamma))] \qquad (4.34)$$

at the point $Q(y;\gamma_0)$.

This revised method admits another interpretation. As $Q(y;\gamma_0)$ can be treated as proportional to the (marginal likelihood based-) LM test for the hypothesis $H_0:\gamma=\gamma_0$ against $H_1:\gamma<\gamma_0$ (see e.g., Grose (1998)), function (4.34) represents the median function (50% critical value function) of this test statistic under the null and the alternative hypotheses depending on γ. Therefore the estimator can be regarded as the result of inverting the median function of the one-sided score test statistic at a fixed-point γ_0. This was consistent with the second method of constructing MU estimators discussed in Chapter 3, which will be our focus in Chapter 5.

If we choose $\gamma_0=0$ in (4.31), the score can be simplified as

$$\begin{aligned} Q_0(y) &= Q(\gamma;y)\big|_{\gamma=0} \\ &= const.+\frac{m(\hat{e}'L_1\hat{e})+(m-1)\hat{e}'\Gamma_\gamma(0)X\hat{\beta}}{\hat{e}'\hat{e}} \\ &= const.+\frac{y'(mML_1M+(m-1)M\Gamma_\gamma(0)X(X'X)^{-1}X')y}{y'My}, \end{aligned} \tag{4.35}$$

where $\hat{e}=\hat{u}(0)$, $M=I-X(X'X)^{-1}X'$, L_1 is the $T\times T$ matrix with the left lower-diagonal elements being 1 and all other elements being 0, and $\Gamma_\gamma(0)$ is identical to L_1 only except with the top left element being 1 instead of 0.

Different from (4.31), (4.35) is a ratio of two quadratic forms in normal random variables y plus a constant. Therefore the median function $m(\gamma)$ can be found by solving

$$\Pr[\sum_{i=1}^{T}\lambda_i\xi_i^2\le 0]=0.5 \tag{4.36}$$

for $m(\gamma)$ using Imhof's (1961) algorithm, where λ_i's are the eigenvalues of $A-m(\gamma)M$, in which

$$A=mML_1M+(m-1)M\Gamma_\gamma(0)X(X'X)^{-1}X',$$

and ξ_i^2 are independent chi-square variables with some non-central parameters δ_i^2 depending on β and given by

$$\delta_i^2=[P'\Omega^{-1/2}(0)X\beta]_i^2,\ i=1,\ \ldots,\ T,$$

in which P is the orthogonal matrix of eigenvectors of $A-m(\gamma)B$.

Figure 4.4 presents some median functions $m(\gamma)$ with β set to be vector of constants for some design matrices with 20 observations. The median functions are strictly monotonic (and almost linear) for all these designs, indicating that estimating equation (4.33) will deliver a unique solution.

In the linear regression model with AR(1) disturbances, the marginal likelihood score (4.12) can be written as a ratio of two quadratic forms in terms of the disturbances, i.e., $u\sim N(0,\sigma^2\Omega(\rho))$. Hence its median function (4.32) or (4.34) is invariant to the nuisance parameters β. But the median function of (4.35) will not be invariant to β, because the second term in the nominator in (4.35) cannot be transformed into a quadratic form in terms of u (i.e., free of the non-central parameter that depends on β). Therefore strictly speaking, $m(\gamma)$ should be written as $m_\beta(\gamma)$.

If β is known, the solution to equation (4.33) is an exactly MU estimator of γ. But in practice, when β is not known, we have to replace β by its consistent estimator and apply the method iteratively. Therefore as a result, the new estimator of γ will only be approximately MU. This iterative bias-correction towards an approximately MU estimator is similar to the approach adopted by Andrews and Chen (1994) and Fair (1996). The practical difficulty when applying the proposed method is that the median function has to be computed for every different set of estimated β values. We have designed the following iterative algorithm to save on computational costs:

Step 1. Calculate $\hat{\gamma}_{OLS}$ and $\hat{\beta}_{OLS}$ from (4.25).

Step 2. Use $\hat{\gamma}_{OLS}$ as the initial value to search for $\hat{\gamma}_2$, such that $Q_0(y) = m_{\hat{\beta}_{OLS}}(\hat{\gamma}_2)$, where $m_{\hat{\beta}_{OLS}}(\bullet)$ indicates that the median function (4.34) is computed with $\hat{\beta}_{OLS}$ treated as the true parameter.

Step 3. $\hat{\beta}_2$ is obtained by regressing $y - \hat{\gamma}_2 y_{-1}$ on X, where $y_{-1} = (y_1, \ldots, y_{T-1})'$.

Step 4. Go back to Step 2, use $\hat{\gamma}_2$ as the initial value to search for $\hat{\gamma}_3$ with $\hat{\beta}_2$ used in the median function computation.

Our experience suggests that there is no need to continue the iterations after Step 4, as the replications after that bring little change to the estimate. So $\hat{\gamma}_3$ can be used as the final estimate. In our simulation studies, we also found that the two-iteration version of the proposed method is at least as fast as the searching algorithm (such as Secant) used in the ML or MMGL estimation procedures.

4.3.4 Interval Estimation

Unlike in the linear regression model with AR(1) disturbances, in which the autoregressive parameter is usually the nuisance parameter, the lagged dependent variable coefficient is more likely to be the parameter of interest in the dynamic linear regression model. Therefore a confidence interval is often required apart from point estimation in practice. As the distributions of the estimators are not approximated very well by asymptotic normal or t distribution in small samples (e.g., see Nankervis and Savin, 1985, 1987, 1988b), we construct confidence intervals based on bootstrap principles. Efron (1985, 1987) outlined several ways of constructing bootstrap confidence intervals. The most popular ones are the percentile method and the percentile-t method, see also discussions in Beran (1987), Hall

(1988, 1994), Efron and Tibshirani (1993), Davidson and Hinkley (1996). Here we apply the percentile-t method to the OLS estimator and the proposed MU estimator.

Nankervis and Savin (1996) developed a bootstrap version of the t test in a model with an intercept and a time trend. The method of computing the confidence intervals in our study is very similar to theirs. First we approximate the distribution function of the estimator $\hat{\gamma}$ by generating B bootstrap samples and compute $\hat{\gamma}[i]$, $i = 1, \ldots, B$, for each sample via the method described above. These estimates are then normalised using the estimated mean and standard error $S\hat{E}(\hat{\gamma})$, i.e.,

$$\hat{t}[i] = (\hat{\gamma}[i] - \frac{1}{B}\sum_B \hat{\gamma}[i]) \Big/ S\hat{E}(\hat{\gamma}), \tag{4.37}$$

where the standard error is estimated by the usual bootstrap estimate,

$$S\hat{E}(\hat{\gamma}) = \{\frac{1}{B-1}\sum_B (\hat{\gamma}[i] - \frac{1}{B}\sum_B \hat{\gamma}[i])^2\}^{1/2}. \tag{4.38}$$

The percentile-t confidence interval is then given by

$$\Pr\{\hat{\gamma} - \hat{t}_{\alpha} S\hat{E}(\hat{\gamma}) \le \gamma \le \hat{\gamma} + \hat{t}_{1-\alpha} S\hat{E}(\hat{\gamma})\} \le \alpha, \tag{4.39}$$

where α is a given confidence level, and $\hat{t}_{\alpha/2}$, $\hat{t}_{1-\alpha/2}$ are the $\alpha/2$ and $(1-\alpha)/2$ quantiles of the bootstrap distribution of $\hat{t}$.

We would expect that since the small sample bias of $\hat{\gamma}_{OLS}$ is effectively corrected by $\hat{\gamma}_{MU}$, the bootstrap confidence interval based on the new estimator should also have better coverage probability than the one based on the OLS estimator.

4.3.5 Experimental Design

We conducted two sets of Monte Carlo studies to examine the small sample performance of the proposed estimator and compare it with the OLS esitmator. The first set includes the following design matrices:

$X1$: An intercept and a linear time trend.

$X3$: An intercept, a time trend and an AR(1) regressor with $\rho = 0.8$.

$X1$ and $X3$ have been studied extensively in time series literature and different methods have been suggested to correct the estimation bias. In particular, Grubb and Symons (1987) examined $X3$ and found that the bias of the OLS estimator depends on the autocorrelation factor in the regressors.

The second set includes two economic time series:

$X4$: A constant, quarterly Australian Consumer Price Index (CPI) commencing 1959(1) and the same series lagged one quarter.

$X5$: The first T observations of Durbin and Watson's (1951) example involving the annual consumption of spirits in the U.K. from 1870 to 1938 which consists of a constant, annual data on the price of spirits and household income.

These two design matrices represent economic data from two different countries. While $X5$ is based on annual data, $X4$ is comprised of quarterly data. The CPI series in $X4$ are highly correlated but smoothly evolving. In fact both design matrices were used in King (1996) and Grose (1998) among other studies.

The sample sizes were set at 20 and 40. The following γ values were used: $\gamma = 0.9,\ 0.8,\ 0.7,\ 0.6,\ 0.5,\ 0.4,\ 0.2,\ 0,\ -0.2,\ -0.5$; All results were based on 1000 estimates. For the confidence intervals, the confidence level was set at 90%. In each replication of the experiment, 200 bootstrap samples were drawn from the DGP given the estimate of $\hat{\gamma}_{MU}$ to estimate the empirical percentiles of the t statistic. The coverage probabilities were based on 1000 confidence intervals.

4.3.6 Results

We first, in Tables 4.2a and 4.2b, report a set of results assuming β known and without using the iterative correction proposed in Section 4.3.3. This experiment was designed to examine the impact of the nuisance parameter β on the effectiveness of the proposed procedure. The median functions were computed at $\beta = 0$. Table 4.2a presents the estimation results for $X1$, $X3$ and $X4$ also assuming $\beta = 0$. The OLS estimators were found to be seriously downward-biased for these designs, especially for moderate to large positive γ values. In contrast, the new estimator $\hat{\gamma}_{MU}$ successfully removed the (median-) bias in the OLS estimator completely for all γ values. The risk (i.e., RMSE) of the new estimator is also considerably smaller than that of the OLS estimator for large γ. However, this advantage is superficial as the new estimator was calculated as if β were known, which gives it an unfair advantage over the OLS estimator and makes the comparison not meaningful. It nevertheless showed that with β known, the proposed estimator (the solution to equation (4.33)) is exactly MU. We would expect when the iterative algorithm is used, in which the true β is replaced by its consistent estimator, the proposed method would at least produce an approximately MU estimator.

Table 4.2b shows that if the median function computed at one β value is used in the estimation of a model with different β values, the results differ. For $X5$, we found that the results are somewhat invariant to β. Although the median function was calculated for $\beta = 0$, the estimates computed for the models generated under other β values are also almost MU. But this is not the case for some other design matrices such as $X1$. If the β value is misspecified, the performance of the estimator without using the proposed bias-prevention algorithm deteriorates.

Estimation results based on the iterative algorithm assuming β unknown are reported in Tables 4.3a – 4.3e. The bias of the OLS estimator varies with the design matrix as well as the magnitude of γ. For many design matrices, it seems the bias is most serious for moderate positive γ (≤ 0.6), a pattern also reported in Grose (1998) and Mahmood (2000). For example, for $X1$ and $T=20$, $\hat{\gamma}_{OLS}$ has downward biases of 0.17 and 0.15 for $\gamma=0.4$ and 0.2, respectively. For $X3$ and $X5$, however, the biases of $\hat{\gamma}_{OLS}$ are quite large for all positive and moderate negative γ values. For example, $\hat{\gamma}_{OLS}$ has downward biases of 0.35 and 0.33 at $\gamma=0.9$ and 0.8 for $X5$ and $T=20$.

The new estimator based on the iterative algorithm effectively corrects the bias in $\hat{\gamma}_{OLS}$ for all these design matrices and sample sizes considered. Although the new estimator is not exactly MU as the one assuming β known, the remaining biases are minimal compared with those of $\hat{\gamma}_{OLS}$. However, the biases are not completely removed for large γ values for $X5$ and $T=20$. For example, at $\gamma=0.9$ and 0.8, $\hat{\gamma}_{MU}$ still has a bias of 0.17 and 0.16, respectively. While for $X4$, the new estimator seems to over-correct the bias in $\hat{\gamma}_{OLS}$ for moderate positive γ. For example, $\hat{\gamma}_{MU}$ has an upward bias of 0.05 and 0.03 at $\gamma=0.6$ and 0.4 for this design matrix with 20 observations. But the magnitude of all these remaining biases of $\hat{\gamma}_{MU}$ are by far smaller than those of $\hat{\gamma}_{OLS}$. When the sample size increases to 40, the new estimator becomes almost exactly MU for all γ values for $X1$ (e.g., see Table 4.3b). This indicates that if the unknown β is replaced by its estimate, the iterative algorithm produces an approximately MU estimator, and the approximation seems to be fairly accurate especially for $T=40$.

Because of the substitution of β by its estimator in the algorithm, we would expect the new estimator to have a larger standard error, and therefore a larger RMSE compared with the one assuming β known. Hence a smaller bias in $\hat{\gamma}_{MU}$ than in $\hat{\gamma}_{OLS}$ might, to some degree, be offset by this increase in standard error. However, the RMSE results in Tables 4.3a – 4.3e indicate that the total risk of $\hat{\gamma}_{MU}$ is smaller than or similar to that of the OLS estimator. For the design matrices and γ values where bias correction is significant, the reduction in RMSE by the new estimator is also substantial. For example, for $X5$ and $T=20$, the RMSE of $\hat{\gamma}_{OLS}$ is reduced by more than 25% at $\gamma=0.9$, 0.8 and 0.6.

As a by-product, we found that the β estimation results based on the new algorithm are also superior to those based on the OLS method. The biggest difference lies in the estimation of the intercept. The (median-) bias of the new estimator is much smaller than that of the OLS estimator almost for all design matrices. For example, for $X3$ and $T=20$, the new estimator of β is almost unbiased for $\gamma \leq 0.6$, while the OLS estimator (especially of the intercept) is seriously upward biased (see Table 4.3c).

The confidence interval results are reported in Table 4.4. The coverage probabilities of the percentile-t intervals based on the two estimators were compared. It is quite apparent that the intervals based on the new estimator have approximately the correct coverage rate (90%) for all design matrices and γ values, while the intervals based on the OLS estimator have coverage rates typically lower than the nominal level, especially for positive γ values. This provides a good example of improving the coverage properties of the confidence interval by correcting the bias in the point estimator. We would expect that the bias correction to the confidence interval suggested by Efron (1987, 1988) should lead to improved accuracy similar to those achieved by the proposed method.

4.4 Concluding Remarks

This chapter provides two examples of applying the method of constructing MU estimators by adjusting the estimating equations. The adjustment to the marginal likelihood score in the linear regression model with AR(1) disturbances can be computed exactly using Imhof's (1961) algorithm, and the new estimator is shown to be almost free of bias in most cases. In the dynamic linear regression model, the median function of the marginal likelihood score is not invariant to nuisance parameters, so we have to substitute these nuisance parameters by their consistent estimators and adjust the estimating equations iteratively. As a result, the new estimator is approximately MU. It was found that the remaining bias in the new estimator is minimal in most cases compared with that of the OLS estimator. The RMSE of the new estimator is generally smaller than that of the OLS estimator especially for positive parameter values. The confidence intervals based on the new estimator were shown to have better coverage probabilities than the ones based on the OLS estimator.

These two examples lead us to believe that correcting the median bias in an estimator by adjusting the estimating equations towards median-unbiasedness can be effective. The proposed method does not require knowledge of the form of the bias function. Our results show that the bias correction can be quite accurate and the overall risk of the new estimator tends to be smaller than that of the biased estimators. The drawback of the proposed method lies in the difficulty of computing the median function, and therefore the difficulty in examining criteria given in Lemma 3.3.3, which are essential for the adjusted estimating equations to deliver unique estimates. In Chapter 6, the estimators derived in this chapter will be used as inputs into hypothesis testing and forecasting procedures.

Although the MU estimator based on adjusting the marginal likelihood score equation was shown to work well in the examples we examined, we are going to change our focus in Chapter 5 to the second approach we developed in Chapter 3 for constructing MU estimators. This is for two reasons: The first is that the likelihood function and the scores are non-standard when the errors follow a random walk process. The limiting distribution of the marginal likelihood score in this case is not clear. Therefore it is not easy to extend the first method to cover the interesting case of a unit root. The second reason is that the computation burden of this approach is quite heavy and the convergence of the iterative procedure developed in Section 3.3 can be slow for some design matrices. Therefore in the next chapter, we examine the second approach to constructing MU estimators proposed in Chapter 3 – inverting the median function of a significance test statistic.

Table 4.1a

Medians and RMSEs of $\hat{\rho}_{OLS}$, $\hat{\rho}_{FML}$, $\hat{\rho}_{MML}$ and $\hat{\rho}_{new}$ in the Linear Regression with AR(1) Disturbances for Design Matrix $X1$

ρ	$T=20$				$T=60$			
	$\hat{\rho}_{OLS}$	$\hat{\rho}_{FML}$	$\hat{\rho}_{MML}$	$\hat{\rho}_{new}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{FML}$	$\hat{\rho}_{MML}$	$\hat{\rho}_{new}$
0.950	0.576	0.602	0.858	0.931	0.831	0.844	0.924	0.953
	(0.421)	(0.396)	(0.213)	(0.187)	(0.134)	(0.124)	(0.078)	(0.073)
0.900	0.539	0.579	0.813	0.878	0.802	0.811	0.882	0.903
	(0.393)	(0.361)	(0.213)	(0.193)	(0.116)	(0.110)	(0.082)	(0.083)
0.800	0.517	0.531	0.748	0.802	0.757	0.717	0.781	0.800
	(0.321)	(0.301)	(0.211)	(0.205)	(0.110)	(0.107)	(0.087)	(0.090)
0.600	0.352	0.370	0.546	0.598	0.529	0.531	0.587	0.598
	(0.277)	(0.271)	(0.222)	(0.232)	(0.106)	(0.107)	(0.096)	(0.097)
0.400	0.209	0.214	0.368	0.393	0.348	0.351	0.398	0.400
	(0.239)	(0.241)	(0.225)	(0.241)	(0.107)	(0.107)	(0.102)	(0.104)
0.200	0.055	0.056	0.187	0.201	0.146	0.145	0.186	0.199
	(0.208)	(0.211)	(0.206)	(0.221)	(0.106)	(0.106)	(0.102)	(0.104)
0.000	-0.108	-0.112	-0.001	0.001	-0.035	-0.035	-0.001	-0.000
	(0.194)	(0.198)	(0.200)	(0.212)	(0.102)	(0.102)	(0.101)	(0.103)
-0.200	-0.270	-0.276	-0.191	-0.198	-0.223	-0.225	-0.198	-0.200
	(0.171)	(0.175)	(0.183)	(0.193)	(0.100)	(0.101)	(0.102)	(0.104)
-0.400	-0.446	-0.450	-0.385	-0.400	-0.413	-0.414	-0.394	-0.400
	(0.156)	(0.159)	(0.172)	(0.180)	(0.092)	(0.092)	(0.094)	(0.096)
-0.600	-0.611	-0.617	-0.574	-0.597	-0.596	-0.597	-0.583	-0.599
	(0.141)	(0.140)	(0.158)	(0.162)	(0.082)	(0.080)	(0.084)	(0.084)
-0.800	-0.778	-0.792	-0.769	-0.797	-0.792	-0.792	-0.785	-0.796
	(0.145)	(0.112)	(0.129)	(0.127)	(0.063)	(0.061)	(0.064)	(0.063)
-0.900	-0.871	-0.885	-0.874	-0.899	-0.887	-0.891	-0.887	-0.899
	(0.097)	(0.086)	(0.098)	(0.094)	(0.052)	(0.050)	(0.052)	(0.051)
-0.950	-0.916	-0.932	-0.924	-0.945	-0.940	-0.942	-0.940	-0.952
	(0.082)	(0.068)	(0.078)	(0.087)	(0.039)	(0.035)	(0.037)	(0.035)

Notes: RMSEs are reported in the brackets beneath the medians.
All experiments are based on 2000 replications.
The MMGL estimates were computed via the Constrained Optimization module in GAUSS.

Table 4.1b

Medians and RMSEs of $\hat{\rho}_{OLS}$, $\hat{\rho}_{FML}$, $\hat{\rho}_{MML}$ and $\hat{\rho}_{new}$ in the Linear Regression with AR(1) Disturbances for Design Matrix $X2$

ρ	$T=20$				$T=60$			
	$\hat{\rho}_{OLS}$	$\hat{\rho}_{FML}$	$\hat{\rho}_{MML}$	$\hat{\rho}_{new}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{FML}$	$\hat{\rho}_{MML}$	$\hat{\rho}_{new}$
0.950	0.274	0.304	0.757	0.826	0.773	0.805	0.918	0.956
	(0.701)	(0.667)	(0.310)	(0.281)	(0.190)	(0.165)	(0.089)	(0.080)
0.900	0.268	0.303	0.740	0.829	0.746	0.765	0.874	0.903
	(0.649)	(0.621)	(0.297)	(0.270)	(0.170)	(0.150)	(0.095)	(0.094)
0.800	0.256	0.281	0.717	0.792	0.671	0.683	0.783	0.805
	(0.568)	(0.544)	(0.299)	(0.283)	(0.148)	(0.139)	(0.102)	(0.106)
0.600	0.159	0.170	0.529	0.588	0.488	0.497	0.586	0.600
	(0.457)	(0.448)	(0.307)	(0.309)	(0.136)	(0.131)	(0.102)	(0.104)
0.400	0.055	0.060	0.365	0.405	0.317	0.320	0.401	0.410
	(0.364)	(0.366)	(0.289)	(0.306)	(0.123)	(0.122)	(0.108)	(0.110)
0.200	-0.078	-0.083	0.173	0.197	0.123	0.123	0.193	0.198
	(0.305)	(0.314)	(0.271)	(0.292)	(0.125)	(0.125)	(0.114)	(0.116)
0.000	-0.202	-0.223	-0.013	-0.006	-0.060	-0.060	-0.002	-0.001
	(0.248)	(0.262)	(0.250)	(0.268)	(0.115)	(0.116)	(0.110)	(0.112)
-0.200	-0.335	-0.366	-0.185	-0.199	-0.253	-0.253	-0.206	-0.209
	(0.201)	(0.218)	(0.223)	(0.241)	(0.111)	(0.112)	(0.110)	(0.112)
-0.400	-0.485	-0.508	-0.372	-0.389	-0.429	-0.431	-0.394	-0.400
	(0.177)	(0.179)	(0.201)	(0.213)	(0.094)	(0.094)	(0.097)	(0.099)
-0.600	-0.647	-0.665	-0.574	-0.598	-0.611	-0.613	-0.588	-0.597
	(0.158)	(0.148)	(0.181)	(0.188)	(0.081)	(0.081)	(0.086)	(0.087)
-0.800	-0.798	-0.822	-0.768	-0.799	-0.797	-0801	-0787	-0.799
	(0.118)	(0.107)	(0.136)	(0.135)	(0.060)	(0.060)	(0.064)	(0.064)
-0.900	-0.880	-0.902	-0.869	-0.899	-0.893	-0.897	-0.889	-0.902
	(0.106)	(0.079)	(0.105)	(0.097)	(0.050)	(0.046)	(0.050)	(0.049)
-0.950	-0.928	-0.944	-0.924	-0.948	-0.940	-0.945	-0.941	-0.951
	(0.091)	(0.064)	(0.087)	(0.076)	(0.039)	(0.035)	(0.038)	(0.037)

Notes: RMSEs are reported in the brackets beneath the medians.
All experiments are based on 2000 replications.
The MMGL estimates were computed via the Constrained Optimization module in GAUSS.

Figure 4.4

Median Functions of the MGL Scores Evaluated at 0 ($Q_0(\gamma)$) for the Dynamic Linear Regression Model Assuming β Known

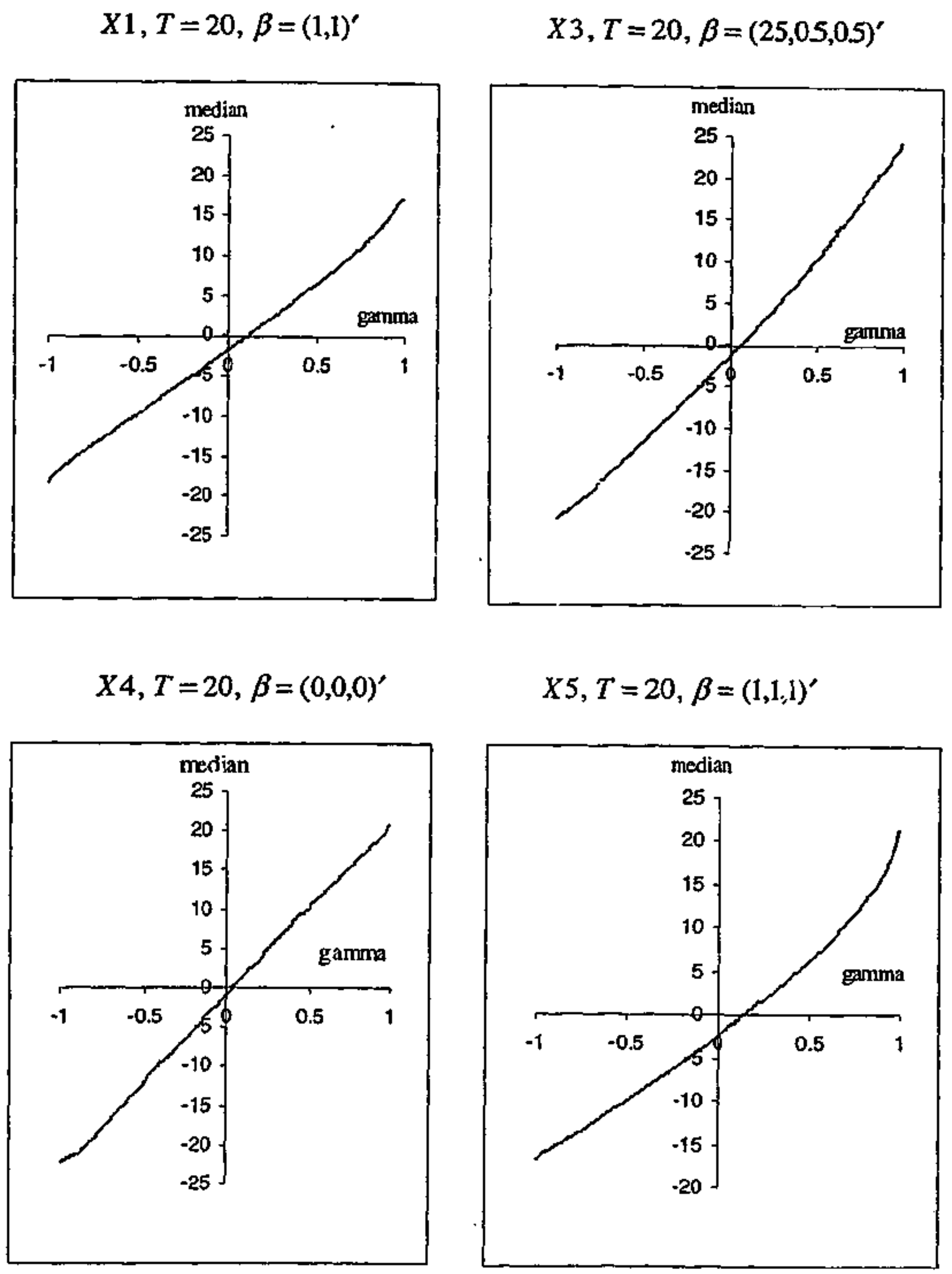


Table 4.2a

Medians and RMSEs of $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Assuming β Known, for Design Matrices $X1$, $X3$ and $X4$; $T=20$, $\beta=(0,0,0)'$

γ	$X1$		$X3$		$X4$	
	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$
0.90	0.55	0.90	0.52	0.90	0.56	0.89
	(0.43)	(0.28)	(0.47)	(0.27)	(0.43)	(0.28)
0.80	0.50	0.79	0.45	0.80	0.50	0.79
	(0.39)	(0.28)	(0.44)	(0.28)	(0.39)	(0.28)
0.60	0.37	0.59	0.31	0.61	0.36	0.60
	(0.34)	(0.30)	(0.39)	(0.30)	(0.35)	(0.31)
0.40	0.22	0.41	0.16	0.40	0.21	0.40
	(0.30)	(0.30)	(0.34)	(0.30)	(0.31)	(0.31)
0.20	0.05	0.19	0.00	0.20	0.05	0.19
	(0.27)	(0.28)	(0.30)	(0.29)	(0.28)	(0.30)
0.00	-0.11	0.00	-0.16	-0.01	-0.11	0.00
	(0.24)	(0.26)	(0.27)	(0.28)	(0.26)	(0.29)

Notes: All experiments are based on 2000 replications.
RMSE's are reported in the brackets beneath the medians.

Table 4.2b

Medians and RMSEs of $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Assuming β Known, for Design Matrix $X5$; $T=20$

γ	$\beta=(0,0,0)'$		$\beta=(1,1,1)'$		$\beta=(25,0.1,0.01)'$	
	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$
0.90	0.57	0.89	0.56	0.90	0.56	0.89
	(0.45)	(0.28)	(0.45)	(0.29)	(0.45)	(0.30)
0.80	0.50	0.80	0.50	0.79	0.50	0.79
	(0.46)	(0.29)	(0.42)	(0.29)	(0.42)	(0.29)
0.60	0.33	0.60	0.34	0.59	0.34	0.59
	(0.38)	(0.31)	(0.38)	(0.31)	(0.38)	(0.31)
0.40	0.17	0.39	0.18	0.40	0.18	0.39
	(0.34)	(0.30)	(0.34)	(0.31)	(0.34)	(0.31)
0.20	0.02	0.20	0.00	0.19	0.00	0.19
	(0.30)	(0.29)	(0.30)	(0.29)	(0.30)	(0.29)
0.00	-0.14	0.00	-0.15	-0.01	-0.15	-0.01
	(0.27)	(0.28)	(0.26)	(0.28)	(0.26)	(0.28)

Notes: All experiments are based on 2000 replications.
RMSE's are reported in the brackets beneath the medians.

Table 4.3a

Medians, Means and RMSEs of $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Assuming β Unknown, for Design Matrix $X1$; $T=20$, $\beta=(1,1)'$

γ Estimation Results

γ	Median $\hat{\gamma}_{OLS}$	Median $\hat{\gamma}_{MU}$	Mean $\hat{\gamma}_{OLS}$	Mean $\hat{\gamma}_{MU}$	RMSE $\hat{\gamma}_{OLS}$	RMSE $\hat{\gamma}_{MU}$
0.90	0.58	0.84	0.56	0.82	0.43	0.30
0.80	0.55	0.75	0.54	0.75	0.39	0.28
0.60	0.42	0.56	0.41	0.55	0.27	0.15
0.40	0.23	0.38	0.22	0.37	0.28	0.22
0.20	0.05	0.19	0.05	0.18	0.27	0.25
0.10	-0.01	0.12	-0.01	0.11	0.25	0.26
0.00	-0.12	-0.02	-0.11	-0.01	0.25	0.26
-0.10	-0.19	-0.11	-0.19	-0.10	0.23	0.25
-0.20	-0.25	-0.18	-0.25	-0.18	0.22	0.25
-0.40	-0.45	-0.40	-0.43	-0.38	0.20	0.23
-0.60	-0.61	-0.60	-0.59	-0.57	0.18	0.22
-0.80	-0.79	-0.80	-0.76	-0.76	0.15	0.18

β Estimation Results (Median)

γ	$\hat{\beta}_{OLS}$ $\hat{\beta}_1$	$\hat{\beta}_{OLS}$ $\hat{\beta}_2$	$\hat{\beta}_{NEW}$ $\hat{\beta}_1$	$\hat{\beta}_{NEW}$ $\hat{\beta}_2$
0.90	2.009	1.080	0.987	1.015
0.80	1.948	1.215	0.731	1.219
0.60	2.037	1.453	0.878	1.113
0.40	2.181	1.288	1.003	1.045
0.20	2.164	1.182	1.008	1.028
0.10	2.098	1.117	0.994	0.989
0.00	2.101	1.110	0.991	1.008
-0.10	2.086	1.081	0.998	1.000
-0.20	2.058	1.044	1.004	0.981
-0.40	2.000	1.024	0.967	0.990
-0.60	1.965	0.998	0.961	0.984
-0.80	1.974	0.978	0.992	0.979

Notes: All experiments are based on 1000 replications.

Table 4.3b

Medians, Means and RMSEs of $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Assuming β Unknown, for Design Matrix $X1$; $T=40$, $\beta=(1,1)'$

γ Estimation Results

γ	Median $\hat{\gamma}_{OLS}$	Median $\hat{\gamma}_{MU}$	Mean $\hat{\gamma}_{OLS}$	Mean $\hat{\gamma}_{MU}$	RMSE $\hat{\gamma}_{OLS}$	RMSE $\hat{\gamma}_{MU}$
0.[illegible]0	0.69	0.89	0.69	0.89	0.28	0.2[illegible]
0.80	0.65	0.79	0.67	0.79	0.26	0.16
0.60	0.51	0.59	0.50	0.58	0.17	0.11
0.40	0.31	0.40	0.30	0.39	0.18	0.15
0.20	0.12	0.19	0.12	0.19	0.18	0.17
0.10	0.03	0.09	0.04	0.10	0.17	0.17
0.00	-0.06	0.00	-0.06	0.00	0.17	0.17
-0.10	-0.15	-0.10	-0.15	-0.10	0.16	0.16
-0.20	-0.23	-0.19	-0.23	-0.19	0.15	0.17
-0.40	-0.42	-0.40	-0.41	-0.39	0.14	0.16
-0.60	-0.60	-0.60	-0.59	-0.58	0.13	0.14
-0.80	-0.79	-0.80	-0.77	-0.78	0.11	0.12

β Estimation Results (Median)

γ	$\hat{\beta}_{OLS}$ $\hat{\beta}_1$	$\hat{\beta}_{OLS}$ $\hat{\beta}_2$	$\hat{\beta}_{NEW}$ $\hat{\beta}_1$	$\hat{\beta}_{NEW}$ $\hat{\beta}_2$
0.9	1.911	1.028	0.706	1.080
0.8	1.860	1.106	0.915	1.036
0.6	1.956	1.253	0.941	1.047
0.4	2.071	1.159	0.991	1.020
0.2	2.081	1.100	1.001	1.013
0.1	2.074	1.070	1.010	1.001
0.0	2.049	1.056	0.993	1.001
-0.1	2.024	1.042	0.978	0.999
-0.2	2.035	1.023	1.006	0.990
-0.4	2.019	1.007	1.007	0.991
-0.6	2.007	0.994	1.011	0.989
-0.8	1.991	0.984	1.007	0.988

Notes: All experiments are based on 1000 replications.

Table 4.3c
Medians, Means and RMSEs of $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Assuming β Unknown, for Design Matrix $X3$; $T = 20$, $\beta = (25,0.5,0.5)'$

γ Estimation Results

γ	Median $\hat{\gamma}_{OLS}$	Median $\hat{\gamma}_{MU}$	Mean $\hat{\gamma}_{OLS}$	Mean $\hat{\gamma}_{MU}$	RMSE $\hat{\gamma}_{OLS}$	RMSE $\hat{\gamma}_{MU}$
0.90	0.786	0.897	0.757	0.890	0.284	0.206
0.80	0.646	0.797	0.603	0.791	0.351	0.194
0.60	0.310	0.561	0.292	0.573	0.385	0.220
0.40	0.156	0.381	0.144	0.382	0.377	0.270
0.20	0.021	0.188	0.020	0.198	0.319	0.259
0.10	-0.054	0.082	-0.061	0.079	0.302	0.242
0.00	-0.128	0.015	-0.113	0.009	0.284	0.256
-0.10	-0.199	-0.098	-0.189	-0.089	0.269	0.251
-0.20	-0.282	-0.200	-0.270	-0.191	0.221	0.246
-0.40	-0.443	-0.394	-0.429	-0.376	0.194	0.228
-0.60	-0.616	-0.596	-0.595	-0.576	0.173	0.213
-0.80	-0.789	-0.802	-0.761	-0.767	0.155	0.193

β Estimation Results (Median)

γ	$\hat{\beta}_{OLS}$ $\hat{\beta}_1$	$\hat{\beta}_{OLS}$ $\hat{\beta}_2$	$\hat{\beta}_{OLS}$ $\hat{\beta}_3$	$\hat{\beta}_{NEW}$ $\hat{\beta}_1$	$\hat{\beta}_{NEW}$ $\hat{\beta}_2$	$\hat{\beta}_{NEW}$ $\hat{\beta}_3$
0.9	92.334	0.836	0.057	65.431	0.598	0.226
0.8	63.186	1.013	0.154	39.854	0.613	0.413
0.6	45.235	0.797	0.258	26.664	0.528	0.494
0.4	36.647	0.657	0.255	25.722	0.515	0.517
0.2	31.599	0.564	0.212	25.058	0.501	0.505
0.1	30.468	0.542	0.200	25.599	0.509	0.495
0.0	28.730	0.515	0.205	24.747	0.498	0.514
-0.1	27.976	0.496	0.169	24.738	0.496	0.504
-0.2	26.253	0.448	0.181	27.567	0.500	0.504
-0.4	27.632	0.624	0.183	25.868	0.503	0.501
-0.6	32.200	0.511	0.196	26.768	0.502	0.501
-0.8	28.640	0.509	0.266	25.273	0.502	0.501

Notes: All experiments are based on 1000 replications.

Table 4.3d
Medians, Means and RMSEs of $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Assuming β Unknown, for Design Matrix $X4$; $T = 20$, $\beta = (1,1,1)'$

γ Estimation Results

γ	Median $\hat{\gamma}_{OLS}$	Median $\hat{\gamma}_{MU}$	Mean $\hat{\gamma}_{OLS}$	Mean $\hat{\gamma}_{MU}$	RMSE $\hat{\gamma}_{OLS}$	RMSE $\hat{\gamma}_{MU}$
0.90	0.725	0.865	0.710	0.826	0.240	0.210
0.80	0.615	0.820	0.592	0.760	0.260	0.180
0.60	0.480	0.630	0.464	0.570	0.230	0.170
0.40	0.220	0.430	0.200	0.420	0.300	0.180
0.20	0.000	0.210	-0.010	0.190	0.320	0.230
0.10	-0.050	0.110	-0.050	0.110	0.280	0.240
0.00	-0.150	0.000	-0.140	0.000	0.280	0.240
-0.10	-0.230	-0.090	-0.220	-0.080	0.270	0.250
-0.20	-0.310	-0.200	-0.300	-0.180	0.250	0.250
-0.40	-0.490	-0.400	-0.470	-0.380	0.220	0.230
-0.60	-0.630	-0.570	-0.610	-0.550	0.190	0.230
-0.80	-0.800	-0.790	-0.770	-0.760	0.150	0.190

β Estimation Results (Median)

γ	$\hat{\beta}_{OLS}$ $\hat{\beta}_1$	$\hat{\beta}_{OLS}$ $\hat{\beta}_2$	$\hat{\beta}_{OLS}$ $\hat{\beta}_3$	$\hat{\beta}_{NEW}$ $\hat{\beta}_1$	$\hat{\beta}_{NEW}$ $\hat{\beta}_2$	$\hat{\beta}_{NEW}$ $\hat{\beta}_3$
0.90	8.574	2.453	-0.522	-0.114	1.018	0.967
0.80	9.012	2.433	-0.435	0.513	1.015	1.006
0.60	10.798	2.320	0.035	0.003	1.116	0.726
0.40	9.392	2.489	0.083	0.750	1.031	0.895
0.20	7.663	2.643	-0.193	0.393	0.996	1.036
0.10	8.593	2.583	-0.326	1.322	0.991	0.984
0.00	8.519	2.617	-0.413	1.268	1.017	0.977
-0.10	7.980	2.613	-0.463	0.842	1.014	0.956
-0.20	7.566	2.622	-0.531	0.556	1.028	0.948
-0.40	8.222	2.537	-0.524	1.395	0.986	0.981
-0.60	8.193	2.446	-0.513	1.466	0.96	0.974
-0.80	7.425	2.477	-0.579	0.415	1.021	0.937

Notes: All experiments are based on 1000 replications.

Table 4.3e
Medians, Means and RMSEs of $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Assuming β Unknown, for Design Matrix $X5$; $T=20$, $\beta=(1,1,1)'$

γ Estimation Results

γ	Median		Mean		RMSE	
	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$	$\hat{\gamma}_{OLS}$	$\hat{\gamma}_{MU}$
0.9	0.553	0.731	0.524	0.687	0.454	0.333
0.8	0.470	0.636	0.445	0.625	0.444	0.318
0.6	0.317	0.486	0.310	0.492	0.38	0.287
0.4	0.175	0.359	0.171	0.345	0.339	0.281
0.2	0.010	0.193	0.012	0.175	0.299	0.266
0.1	-0.050	0.108	-0.046	0.104	0.278	0.266
0	-0.163	-0.024	-0.154	-0.017	0.272	0.262
-0.1	-0.229	-0.105	-0.218	-0.093	0.245	0.255
-0.2	-0.308	-0.208	-0.297	-0.192	0.233	0.255
-0.4	-0.475	-0.406	-0.457	-0.382	0.202	0.235
-0.6	-0.637	-0.595	-0.603	-0.561	0.179	0.226
-0.8	-0.800	-0.790	-0.770	-0.760	0.150	0.190

β Estimation Results (Median)

γ	$\hat{\beta}_{OLS}$			$\hat{\beta}_{NEW}$		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.9	13.924	3.283	1.316	8.443	2.041	1.349
0.8	4.953	2.084	2.269	2.984	1.333	1.798
0.6	1.704	2.354	1.151	1.218	1.174	1.390
0.4	1.496	1.260	1.440	1.240	0.660	1.417
0.2	0.650	1.550	1.246	0.662	0.982	1.265
0.1	1.768	0.840	1.169	1.735	0.643	0.952
0.0	0.805	2.185	0.372	0.849	1.632	0.524
-0.1	0.258	1.377	1.301	0.513	1.047	1.191
-0.2	1.164	1.632	0.525	1.248	0.951	0.899
-0.4	0.610	1.223	1.099	0.787	0.536	1.515
-0.6	0.761	2.112	0.085	1.001	1.157	0.787
-0.8	7.425	2.477	-0.579	0.415	1.021	0.937

Notes: All experiments are based on 1000 replications.

Table 4.4
Coverage Probabilities of the Bootstrap Confidence Intervals at the 90% Confidence Level Based on $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the Dynamic Linear Regression Model for Design Matrices $X1$, $X4$, $X5$

γ	0.9	0.8	0.6	0.4	0.2	0.1	0	-0.1	-0.2	-0.4	-0.6	-0.8
						$X1$, $T=20$						
OLS	0.880	0.823	0.718	0.779	0.820	0.854	0.875	0.862	0.902	0.899	0.895	0.856
MU	0.894	0.890	0.878	0.896	0.888	0.903	0.896	0.901	0.895	0.905	0.884	0.856
						$X1$, $T=40$						
OLS	0.882	0.852	0.777	0.830	0.850	0.855	0.871	0.888	0.899	0.912	0.894	0.846
MU	0.880	0.898	0.882	0.900	0.898	0.899	0.902	0.896	0.903	0.898	0.886	0.865
						$X4$, $T=20$						
OLS	0.904	0.888	0.786	0.741	0.789	0.831	0.818	0.849	0.877	0.900	0.907	0.857
MU	0.903	0.895	0.908	0.900	0.894	0.899	0.900	0.898	0.894	0.896	0.877	0.833
						$X5$, $T=20$						
OLS	0.504	0.572	0.662	0.727	0.805	0.840	0.839	0.844	0.872	0.897	0.890	0.876
MU	0.769	0.870	0.894	0.898	0.890	0.901	0.903	0.900	0.891	0.887	0.869	0.872

Chapter 5

Inverting Point Optimal Invariant Tests for Median-unbiased Estimators

5.1 Introduction

In Chapter 4, we constructed a median-unbiased (MU) estimator for the linear regression model with stationary AR(1) disturbances by adjusting the marginal likelihood score equation. In this chapter, we extend the model to include the important case of random walk disturbances. We apply the second method discussed in Chapter 3, i.e., constructing a MU estimator by inverting the median function of a significance test statistic to this model. Some new MU estimators of the autoregressive parameter are developed and their small sample biases and risks are compared with those of the conventional estimators via Monte Carlo simulations.

Because random walk disturbances are included in the model specification, the autoregressive coefficient is now restricted to a closed set [-1,1]. In this case, global mean-unbiasedness is not achievable as the mean of any estimator would bias towards the boundary for large positive or negative parameter values (Andrews, 1993). Therefore median-unbiasedness becomes a very important measure of impartiality of the estimators, see discussions in Andrews (1993), Stocks (1994), Fuller (1996), and Maddala et al. (1998). On the other hand, due to the discontinuous likelihood function when ρ moves from the stationary region to its boundary, many other bias-correction methods may not work over the whole parameter space. We show that the proposed method is not affected by this problem and it produces a MU estimator for all values in the parameter space.

When applying this method, there are many different test statistics that we can choose to invert. They can either be tests for serial correlation or tests for random walk disturbances. It is well known that when constructing a confidence interval by inverting a test statistic, the risk of the interval depends on the power of the test. We attempt to use this model to illustrate a similar relationship between the small sample power properties of a test and the small sample performance of the resulting MU estimator.

In particular, we point out that Andrews' (1993) estimator breaks down for some design matrices due to the non-monotonic median function of the OLS estimator. We propose to invert the point optimal invariant (POI) test instead, as a remedy to this problem. After the exact median functions of different test statistics are examined via Imhof's (1961) algorithm, it was found that the POI test statistic has a monotonic median function for positive parameter values and for all design matrices considered, while other tests do not. Therefore the POI test statistic is recommended for constructing a MU estimator in this model.

The POI test can have non-monotonic power for certain design matrices as $\rho \to -1$ when testing the random walk disturbance hypothesis. We derive the conditions for the test to have non-monotonic power, and hence a non-monotonic median function. In this situation, we suggest using a median-envelope that is based on the grid inversion method developed in Chapter 3, in place of the median function of a single POI test. It is shown that the median-envelope approach can overcome the difficulty and produce reliable estimates. We also identify an easy-to-use criterion of when to use which method given a design matrix.

The chapter is organised as follows. In Section 5.2, we specify the initial conditions of the model and set out the design matrices in our study. In Section 5.3, we point out the problem that Andrews' estimator may suffer from. By examining the median functions of several tests, we show that these tests can also encounter the same problem. In Section 5.4, we study the small-sample power properties of POI tests and derive the conditions for them to have non-monotonic power curves. We also introduce the concept of the median-envelope. In Section 5.5, we compare the performance of different estimators based on different test statistics and disclose the relationship between the power of a test and the property of the resulting estimator. In Section 5.6, we examine the robustness of the proposed estimator to non-normal errors. The chapter ends with some concluding remarks in Section 5.7.

5.2 Model Specification

The model of concern in this chapter is essentially the same as the one in Section 4.1, except that we now need to specify a different initial condition for the random walk case. We restate the model here,

$$y_t = x_t'\beta + u_t, \ (t = 1,\dots,T)$$
$$u_t = \rho u_{t-1} + \varepsilon_t, \ \varepsilon_t \text{ iid } N(0,\sigma^2),$$

where y_t is the dependent variable (observed at time t), x_t is a $k \times 1$ vector of fixed regressors, β is a $k \times 1$ vector of fixed coefficients, and u_t is a random disturbance. The coefficients β, ρ and σ^2 are unknown.

The initial condition plays an important role in studying inference procedures in random walk disturbances models. Methods that avoid the problem by discarding the first observation or condition on the first observation may cause a loss of efficiency in the estimation as discussed in Beach and MacKinnon (1978). On the other hand, a fixed starting point, if chosen too far away from the deterministic trend line, might have an adverse effect on the finite sample results of the estimation and testing in the unit root regression, see for example Pantula et al. (1994) for a discussion. Therefore we do not restrict the initial error to be a constant.

When specifying the initial conditions, one would expect the distribution of the estimators of concern to have a smooth transition when ρ moves from the stationary region to its boundary. The most popular choices of initial conditions are the following two sets:

Assumption 1: $u_0 \sim N[0, \sigma^2/(1-\rho^2)]$ if $|\rho| < 1$, and

Assumption 2.1: u_0 is an arbitrary constant or with an arbitrary distribution which is independent of $\varepsilon_1,\dots,\varepsilon_T$ if $|\rho| = 1$, or

Assumption 2.2: if $|\rho| = 1$, $u_0 \sim N(0, d^2)$ and independent of $\varepsilon_1, \ldots, \varepsilon_T$, where d is an unknown constant.

Andrews (1993), Hansen (1999) and many others used assumption 1 and 2.1, while Berenblut and Webb (1973), Dufour (1990) and Dufour and King (1991), among others used assumption 1 and assumption 2.2. In this chapter, we try to be consistent with Dufour and King (1991) and choose assumption 1 and 2.2, in order to use some of their results about POI tests in our study. This set of assumptions is more restrictive than 2.1, but the calculation of the exact distributions in our study requires the specification of the distribution of the start-up value. However, in Section 5.6, we will show that the estimation procedures developed in this chapter are robust to error misspecifications and therefore the actual distribution of the initial error is not crucial.

It is also well known that the small sample performance of the estimators and tests depend on the number of regressors and choice of regressors. We attempt to include a range of design matrices that cover most of the important cases considered in the previous studies in terms of estimation bias and power of tests. In our study, we choose 8 different design matrices, as representatives of typical economic time series. Throughout this chapter and Chapter 6, we will refer to them frequently, therefore we specify them here. These design matrices can be classified into two sets.

The aim of the first set is to examine the possible extremes of differences in power between the DW test and the POI test, in order to detect the impact of this difference on the performance of the resulting MU estimator. We would expect that in those design matrices that favour the POI test in terms of power should also favour the MU estimator based on the POI test statistics in terms of unbiasedness and efficiency. This set includes the following 5 $T \times k$ design matrices with $T = 20$, 40, 60 and $k = 2$, 3, 4, 5:

$X1$: An intercept and a linear time trend. For stationary serial correlation testing, the POI test does not have a clear power advantage over the DW test for this design. We include it in order to compare the proposed MU estimator with Andrews'

estimator and other bias-reduced estimators. It also allows us to apply the proposed new estimator in testing for a unit root in Chapter 6.

$X2$: An intercept, a linear trend and a stationary $AR(1)$ regressor. The autoregression coefficient in the third regressor is set at 0.8.

$X3$: An intercept, a linear trend and a random walk regressor. This design matrix and $X2$ are frequently encountered by economic researchers. They have also been used in studying the estimation and testing of ρ, for example, by Spitzer (1979), Park and Mitchel (1981), Nankervis and Savin (1987) and Atukorala (2000).

$X7$& $X8$: Watson's X matrix with 3 and 5 regressors respectively. The regressors are a_1, $(a_1 + a_T)/\sqrt{2}$, ... , $(a_k + a_{T-k+2})/\sqrt{2}$, where $a_1, \ldots, a_T$ are the eigenvectors corresponding to the eigenvalues of A_1 (given by (4.24)) arranged in ascending order. This design matrix is a well known extreme case in which the OLS residuals are poor estimates of the real disturbances and the DW test has poor small sample powers compared with those of the POI tests (e.g., see King, 1985a). These regressors were utilised in the Monte Carlo studies by King (1985a), Kramer and Ziesel (1990), Dufour and King (1991) and Bartels (1992) among others to investigate the small-sample properties of the autocorrelation tests in the linear regression model, and by King (1996) and Goh and King (1999) to investigate tests on the lagged dependent variable in the linear dynamic regression model.

The second set of design matrices are comprised of some typical economic data. All series are quarterly data that are associated with various seasonality features. These features may have a big impact on the performance of estimators and hypothesis tests.

$X4$: A constant dummy, the quarterly Australian Consumer Price Index (ACPI) commencing 1959(1) and the same index lagged one quarter up to $k-2$ quarters as regressors.

$X5$: A constant dummy, quarterly Australian private capital movements and quarterly Australian Government capital movements commencing 1968(1) and the government capital movement lagged one quarter.

$X6$: $X5$ with quarterly Australian retail trade as an additional regressor.

Among these 3 designs, the quarterly ACPI is a weakly seasonal series while the two capital movement series are strongly seasonal with two seasonal peaks per year. The capital movement series also exhibit some large fluctuations, while the quarterly Australian retail trade series is much more well behaved, showing moderate seasonality. The seasonal patterns of the regressors of $X6$ are not too distant from those of the 'seasonal' components of Waston's X matrix.

5.3 Which Test Statistic to Invert

5.3.1 Andrews' Estimator and Its Problem

Andrews (1993) proposed a MU estimator for the first-order autoregressive/unit root model with a drift and/or a time trend by inverting the median function of the OLS estimator, i.e.

$$\hat{\rho}_A = m^{-1}(\hat{\rho}_{OLS}), \tag{5.1}$$

where

$$\hat{\rho}_{OLS} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} \Big/ \sum_{t=2}^{T} \hat{u}_{t-1}^2, \tag{5.2}$$

$$\hat{u} = (I - X(X'X)^{-1}X')y, \tag{5.3}$$

and

$$m(\rho) = median[\hat{\rho}_{OLS} | u \sim N(0, \sigma^2 \Sigma(\rho))]. \tag{5.4}$$

Andrews' method was originally designed for the simple autoregessive model without any exogenous regressors, But if extra regressors are included in the model, similar procedure can be used with the observed series in the calculation replaced by the least squares residuals. The estimator in this case is the one-step Cochrane-Orcutt estimator, but we still refer to it as $\hat{\rho}_{OLS}$ as it involves the OLS regression of the residuals. As pointed out by MacKinnon and Smith (1998), estimator (5.3) can be seen as inverting the median function of the non-pivotized t test, i.e., $\hat{\rho}_{OLS} - \rho_0$.

The median function $m(\rho)$ can be computed exactly using algorithms such as Imhof (1961), as we can write the OLS estimator as a ratio of two quadratic forms in terms of the normal errors:

$$\begin{aligned} \hat{\rho}_{OLS} &= \sum_{2}^{T} \hat{u}_t \hat{u}_{t-1} \Big/ \sum_{2}^{T} \hat{u}_{t-1}^2 \\ &= u'M'AMu / u'M'BMu \\ &= \varepsilon'R'M'AMR\varepsilon / \varepsilon'R'M'BMR\varepsilon \end{aligned} \tag{5.5}$$

where

$$A = \begin{bmatrix} 0 & \frac{1}{2} & \cdots & 0 \\ \frac{1}{2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \frac{1}{2} \\ 0 & \cdots & \frac{1}{2} & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix},$$

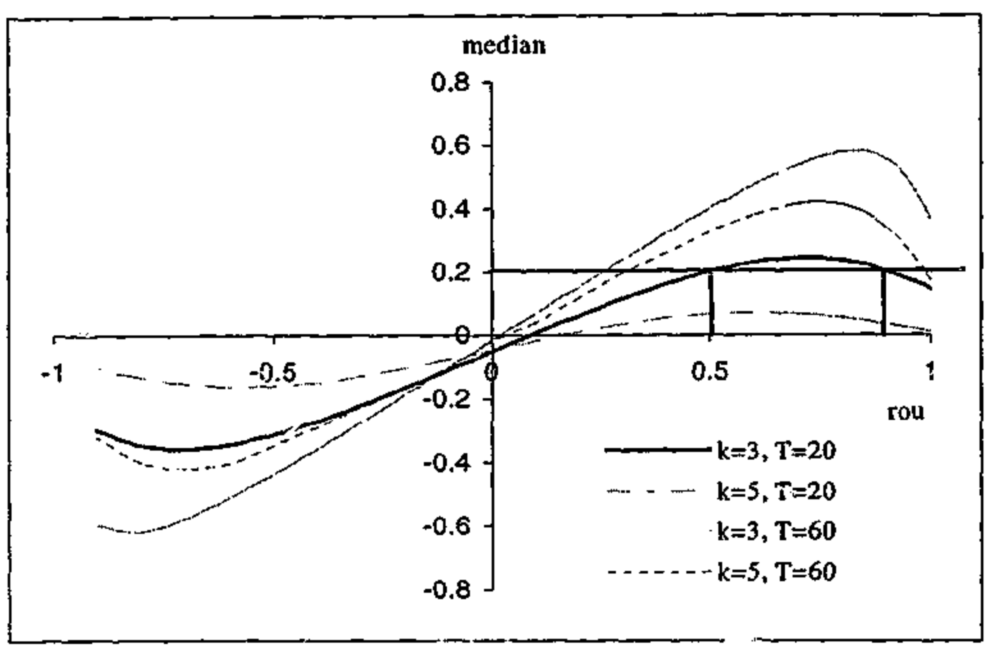
$$M = I - X(X'X)^{-1}X', \tag{5.8}$$

$$R(\rho) = \begin{bmatrix} 1/\sqrt{1-\rho^2} & 0 & \cdots & 0 \\ \rho/\sqrt{1-\rho^2} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{T-1}/\sqrt{1-\rho^2} & \rho^{T-2} & \cdots & 1 \end{bmatrix}. \tag{5.9}$$

Andrews (1993) conjectured that his method could be applied to the models with exogenous regressors. Now we show that his method can break down for some design matrices. As a counter example, consider Watson's X matrix: $X7$ and $X8$.

Figure 5.1 shows that the median functions of the OLS estimators are seriously non-monotonic for both positive and negative ρ values. In particular, if we apply Andrews' method here for Watson's X matrix with 3 regressors and a sample size of 20, and if the original OLS estimate is $\hat{\rho}_{OLS}$= 0.2, say, then Andrews' MU estimator $\hat{\rho}_A = m^{-1}(\hat{\rho}_{OLS})$ can either be 0.5 or 0.9, as shown in Figure 5.1. Based on the sample information, it would be impossible to choose one from these two estimates. This shows that Andrews' estimator is inapplicable in this circumstance. The median functions of the OLS estimator for different sample sizes and different numbers of regressors of Waston's X matrix are also presented in Figure 5.1. For a sample size of 60, the median function is still not monotonic.

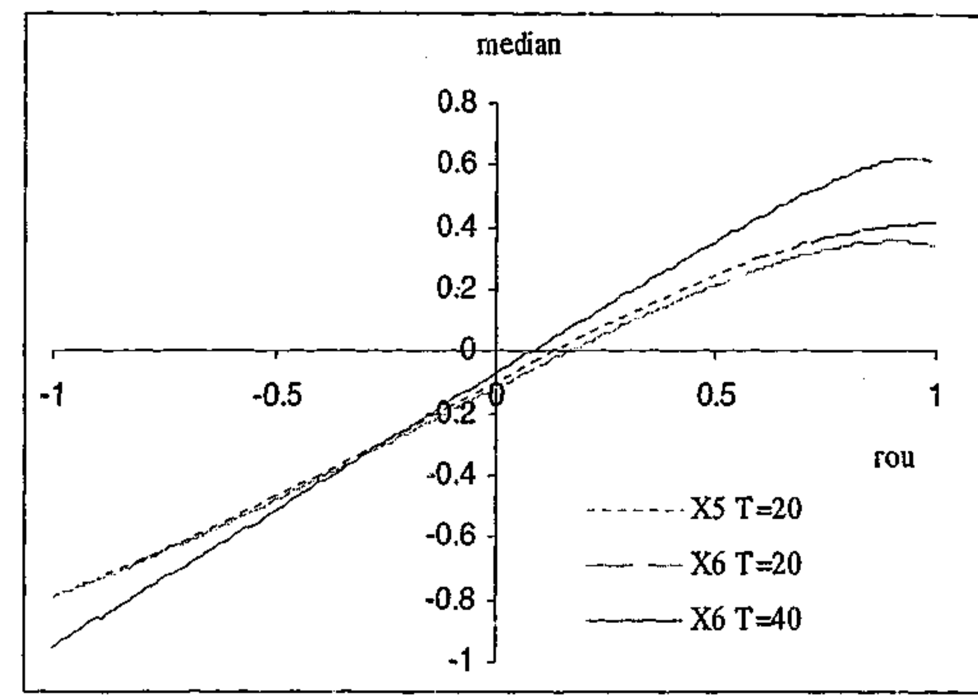
Figure 5.1 Computed Median Functions of $\hat{\rho}_{OLS}$ for Watson's X Matrix



The median functions of the OLS estimator for some other design matrices are presented in Figure 5.2. While the median function is monotonic for design matrices $X1$, $X2$, $X3$, $X4$ and $X5$, it is also non-monotonic on both the positive and negative side for design matrix $X6$ and for $T = 20$, 40. Another important feature is that unlike $X1$, with just an intercept and a time trend as regressors, for which the median function is almost linear, for design matrices $X4$, $X5$ and $X6$, the median function bends downwards when ρ is close to 1. As the median function of the OLS

estimator is just the 50% critical value function of the test statistic $\hat{\rho}-\rho$, it is natural to link the non-monotonicity of the median function to the non-monotonic power curve of this test (the unpivotized version of the t test). For Watson's X matrix and $X6$, the power curve of this test drops in the neighbourhood of both $\rho = 1$ and $\rho = -1$. On the other hand, for design matrices such as $X4$ and $X5$, the test lacks in power for large positive ρ. That is why the median function for these design matrices, though still monotonic, goes flat when ρ goes to 1. We revisit this convex curvature of the median function in Section 5.5 and relate it to the performance of Andrews' estimator for different design matrices.

Figure 5.2 Computed Median Functions of $\hat{\rho}_{OLS}$ for $X5$ and $X6$, $T = 20$, 40



The non-monotonic median function associated with the OLS estimator leads us to examine other available tests and look for one that has a monotonic median function (or equivalently, power curve) for small samples and for all design matrices.

5.3.2 Median Functions of Different Tests

When we construct a MU estimator of ρ by inverting a test statistic at its 50% significance level, the test can either be a test for serial correlation (testing $H_0:\rho=0$ against autocorrelation alternatives) or a test for random walk disturbances (testing $H_0:\rho=1$ against stationary alternatives). There is a large literature on these tests. For a thorough review on testing for serial correlation, see King (1987c), and for a complete literature survey on unit root testing, see Phillips and Xiao (1999). Based on Theorem 3.3.1, only test statistics with a monotonic median function can be inverted at a fixed point. Apart from this criterion, ease of computation of the median function can also be an attractive feature. Many of these tests can be expressed as quadratic forms of the normal errors or ratios of two quadratic forms. We can apply the popular Imhof (1961) algorithm to evaluate the median function with any desired level of accuracy. Otherwise we have to approximate the median functions by simulation.

We briefly review three popular tests designed to test for serial correlation or for random walk errors. For each test, where applicable, we consider the fixed-point inversion method and the grid inversion method developed in Section 3.5. The median functions for these test statistics are computed or simulated for different design matrices.

5.3.2.1 Durbin-Watson Test.

The DW test is the most popular test statistic for testing for autocorrelation in the disturbances against the null of independent errors. Many researchers have examined the small-sample power properties of the DW test. Non-monotonic power problem in small samples has been reported in several studies. Tillman (1975), Kramer (1985), King (1985), Zeisel (1989), Kramer and Zeisel (1990) and Bartels (1992) all examined the power functions of the DW test for different design matrices. In particular, Kramer and Zeisel (1990) and Bartels (1992) investigated its limiting distribution as $\rho\to 1$ or -1. They reported that the limiting power could drop to 0

for some design matrices (e.g., $X7$ and $X8$). This non-monotonic power seems to occur more often on the positive side of the origin. Criteria based on the limiting power of the test were developed by Kramer and Zeisel (1990) and Bartels (1992) to determine the appropriateness of using the DW test for a given design matrix. Dufour (1991) considered inverting the quantile functions of the DW statistic to construct exact confidence intervals. We are more interested in the median function of the DW statistic.

Similar to the OLS estimator, the DW test statistic can also be expressed as a ratio of two quadratic forms in terms of ε, i.e.,

$$\begin{aligned} DW &= \sum_{t=2}^{T}(z_t-z_{t-1})^2 \Big/ \sum_{t=1}^{T} z_t^2 \\ &= \varepsilon' R' M' A_{DW} M R \varepsilon / \varepsilon' R' M R \varepsilon, \end{aligned}$$

where z is the OLS residual vector and

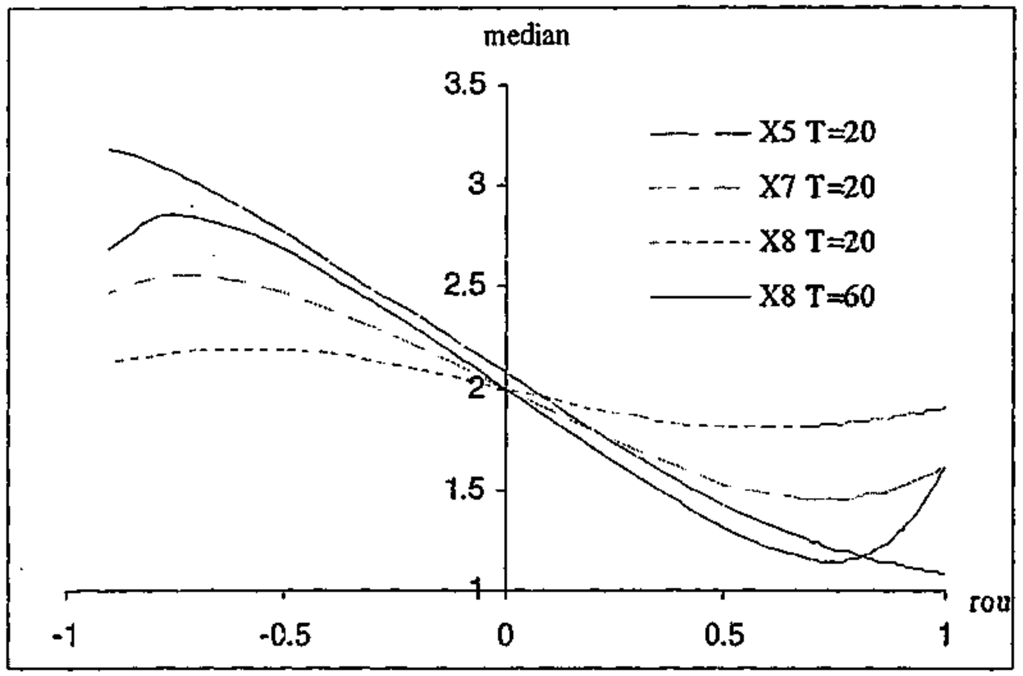
$$A_{DW}=\begin{bmatrix} 1 & -1 & \cdots & 0 \\ -1 & 2 & \ddots & \vdots \\ \vdots & \ddots & 2 & -1 \\ 0 & \cdots & -1 & 1 \end{bmatrix}.$$

As the DW test is usually used to test the hypothesis at $\rho_0=0$, we only apply the proposed Method I (fixed-point inversion). Therefore the MU estimator based on the DW statistic will be:

$$\hat{\rho}_{DW}^{MU}=m_{DW}^{-1}[\sum_{2}^{T}(z_t-z_{t-1})^2 \Big/ \sum_{2}^{T} z_t^2]. \tag{5.12}$$

The median functions of the DW test for different design matrices are presented in Figure 5.3. They depict a similar pattern to those of the OLS estimator. The median functions are not monotonic for design matrices $X6$, $X7$ and $X8$. Therefore the estimator (5.12) is not reliable for these design matrices. It reminds us that although inverting the DW test will produce accurate confidence intervals (see e.g., Dufour, 1990), it may not produce reliable point estimates for some design matrices.

Figure 5.3 Computed Median Functions of the *DW* Statistic for $X5$, $X7$ and $X8$, $T=20$, 60



Based on discussions in Sections 3.3 and 3.4, our results on the median functions reported in this section verify the findings reported in Kramer and Zeisel (1990) and Bartels (1992), that the DW test may have a non-monotonic power curve in small samples for some design matrices.

5.3.2.2 *t* Statistic

Theoretically, the *t* statistic:

$$t=\hat{V}(\hat{\rho}_{OLS})^{-1/2}(\hat{\rho}_{OLS}-\rho_0) \tag{5.13}$$

can also be used to test hypotheses about ρ. It is well documented that the null distribution of this *t* statistic will not be Student *t* when $\rho_0=1$, and that asymptotic

normality can be a poor approximation for large positive ρ values even when the sample size is fairly large (e.g., see Dickey and Fuller, 1979 and Nankervis and Savin, 1985, 1987, 1988b). For models with just a drift and/or a linear time trend, Dickey and Fuller (1979) derived its limiting distributions and tabulated its quantiles. Nankervis and Savin (1987, 1988b) analysed the discrepancies between the finite sample distribution of the *t* statistic and the Student *t* distribution approximation. Stock (1991) inverted the median function of the limiting distribution of this test to construct confidence intervals and an asymptotically MU estimator for ρ. Hansen (1999) took a similar approach to constructing confidence intervals that have asymptotically correct coverage probabilities.

We cannot express the *t* statistic as a ratio of two quadratic forms of normal errors. Therefore simulation is needed to approximate its median function. On the other hand, it is possible to use both fixed-point inversion and grid inversion based on the *t* statistic. The two MU estimators so constructed are given by:

$$\hat{\rho}^{MU}_{t(\rho_0)}=m^{-1}_{t(\rho_0)}[(\hat{\rho}_{OLS}-\rho_0)V(\hat{\rho}_{OLS})^{-1/2}], \tag{5.14}$$

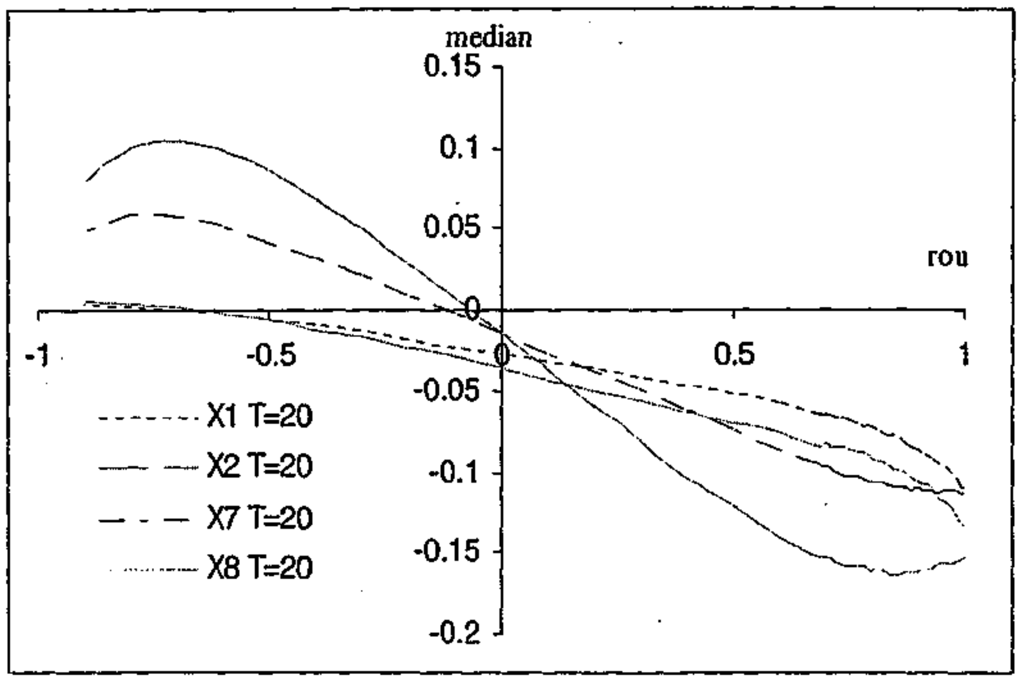
where the null value ρ_0 is fixed, and $\hat{\rho}^{MU}_{t(\rho)}$, which solves the following equation for ρ:

$$m_{t(\rho)}(\rho)=(\hat{\rho}_{OLS}-\rho)V(\hat{\rho}_{OLS})^{-1/2}. \tag{5.15}$$

Here $m_{t(\rho_0)}(\rho)$ and $m_{t(\rho)}(\rho)$ stand for the median function of the *t* statistic for a fixed null value (Method 1) and a median envelope for changing null values (Method 2), respectively. However, based on the results reported in Nankervis and Savin (1985) and Hansen (1999), the distribution of the *t* statistic depends on the null value especially in the neighbourhood of unity. Therefore the fixed-point inversion will not deliver reliable estimates for the current model, as the median functions of the statistics under different null hypotheses will not be of the same shape and parallel to each other. Therefore we only report the grid-simulated median functions for different design matrices in Figure 5.4.

Our results showed that the median functions of the t test statistic are not monotonic for most of the design matrices we examined. This non-monotonicity was verified in Stock (1991) and Hansen (1999) for $X1$. In their studies, only confidence intervals were considered, so this non-monotonicity only caused disjoint or empty intervals in some cases. But if we want to avoid non-unique point estimates, we should not use the t test statistic for constructing a MU estimator for these design matrices. We will revisit the non-monotonic power of the t test in Chapter 6 and propose a remedy for the problem.

Figure 5.4 Simulated Median Functions of the t Statistic for $X1$, $X2$, $X7$ and $X8$, $T=20$



5.3.2.3 LM Test.

LM tests can have good power properties when used for diagnostic testing, see e.g. Godfrey (1988) for a survey. LM tests for the unit root hypothesis or the random walk error hypothesis were considered by Sargan and Bhargava (1983) and Schmidt and Phillips (1992) among others. The one-sided LM test is also closely linked to the LBI test developed by King and Hillier (1985). Here we examine the median functions of the one-sided version of the score test. We discard the

information component term, as it is a constant and can be ignored in the proposed algorithm. So the test statistic is simply the score $s(\rho)|_{\rho_0}$, where ρ_0 is the pre-determined null value.

Following Beach and MacKinnon (1978), the score function of the profile likelihood is:

$$Sc(\rho)=\rho^3+a\rho^2+b\rho+c, \qquad (5.16)$$

where

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

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

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

We can also express the test statistic as a ratio of two quadratic forms of normal errors:

$$Sc(\rho)=\varepsilon'R'\overline{P}'A_{Sc}\overline{P}R\varepsilon/\varepsilon'R'\overline{P}'B_{Sc}\overline{P}R\varepsilon \qquad (5.17)$$

in which the two middle matrices are given by

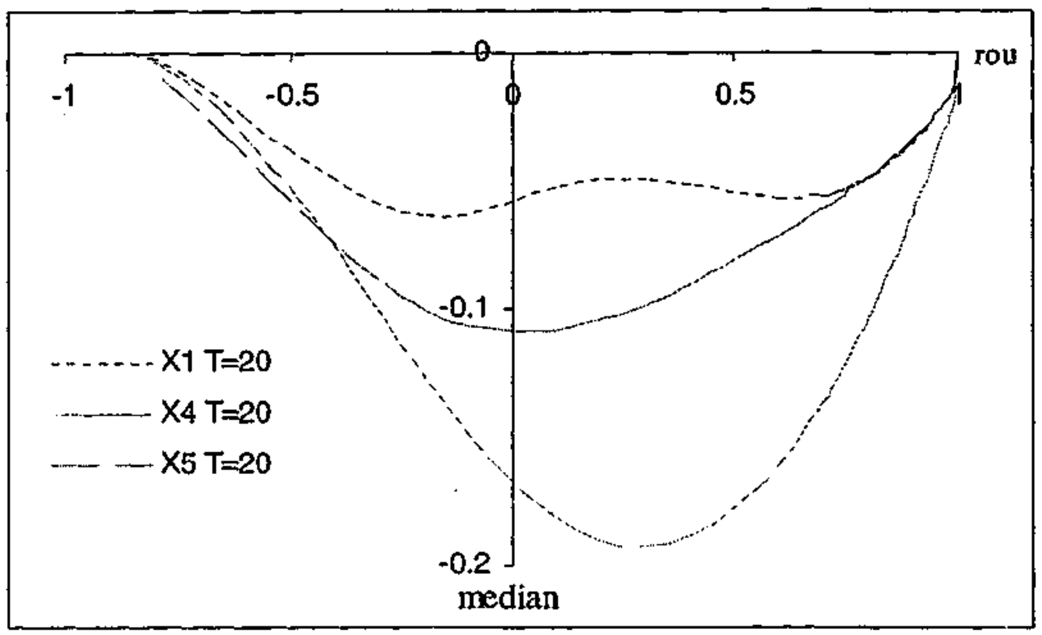
$$A_{Sc}=\begin{bmatrix} -\rho & 0 & \cdots & & 0 \\ (T-1)-\rho(T-2) & (T-1)\rho^3-(T+1)\rho & \ddots & & \vdots \\ 0 & \ddots & \ddots & & \\ \vdots & & & (T-1)\rho^3-(T+1)\rho & 0 \\ 0 & \cdots & 0 & (T-1)-\rho(T-2) & -\rho \end{bmatrix}$$

$$B_{Sc}=\begin{bmatrix}0 & 0 & \cdots & & 0\\ 0 & T-1 & \ddots & & \vdots\\ \vdots & \ddots & \ddots & & \\ & & & T-1 & 0\\ 0 & \cdots & & 0 & 0\end{bmatrix},$$

$$\widetilde{P}=I-X(X'\Sigma^{-1}(\rho)X)^{-1}X'\Sigma^{-1}(\rho). \qquad (5.18)$$

Figure 5.5 presents the median functions computed by the grid inversion method, for the one-sided LM test statistics for different design matrices. Similar to the picture of the t test, these median functions are mostly non-monotonic. Therefore the LM test is not recommended for constructing a MU estimator in the current model.

Figure 5.5 Computed Median Functions of the LM Test Statistic for $X1$, $X4$ and $X5$, $T=20$



We conclude this section by reiterating our main findings, that the median functions of the OLS estimator, DW test, t test and LM test statistics can all be non-monotonic for some design matrices, which means that none of these tests is a good candidate for constructing a MU estimator that can deliver unique estimates for all design matrices. We are motivated by these findings to consider more powerful tests.

5.4 Inverting Point Optimal Invariant Test

5.4.1 Testing for Random Walk Disturbances by POI Tests

As noted by Honda (1989), there are two approaches to 'optimal tests' in hypothesis testing when nuisance parameters exist. The first is the use of conditional distributions, which usually leads to similar tests, and the second appeals to the argument of invariance. Point optimal tests are most often used in line with the second approach. In the linear regression model with autocorrelated disturbances, many researchers have reported that the POI test can have a distinctive power advantage over other tests in small samples at least for some design matrices. These include testing for serial correlation (King, 1985a, Honda, 1989, Kramer and Zeisel, 1990) and testing for random walk disturbances (Dufour and King, 1991, Phillips and Xiao, 1998). King (1985a) and Dufour and King (1991) showed that in the current model, when the design matrix is made up of the eigenvectors of the error covariance matrix, the POI test is uniformly most powerful invariant. Otherwise, it is the locally most powerful test in the neighbourhood of the pre-selected alternative point. We would expect the superior small-sample performance of this test can carry over to the MU estimator based on it.

In order to test hypotheses:

$$H_0:\rho=\rho_0 \text{ against } H_1:\rho<\rho_0, \qquad (5.19)$$

the POI test involves rejecting H_0 for small values of

$$p(\rho_0,\rho_1)=\hat{u}_1'\Sigma^{-1}(\rho_1)\hat{u}_1/\hat{u}_0'\Sigma^{-1}(\rho_0)\hat{u}_0, \qquad (5.20)$$

in which ρ_1 is a pre-selected alternative point at which the power is to be maximised, and $\hat{u}_1$ and $\hat{u}_0$ are the GLS regression residual vectors for error covariance matrices $\Sigma(\rho_1)$ and $\Sigma(\rho_0)$, respectively.

This is a special case of the POI tests defined in Dufour and King (1991) which account for more general initial conditions. Their test is based on the assumption that $u_1 = d_1\varepsilon_1$, where d_1 is an unknown parameter. As d_1 is not consistently estimable, a preselected d_1 has to be used when constructing the POI test. If we want to have a POI test which is free of this extra parameter, the transformation group under which a maximal invariant can be derived for the construction of the POI test has to be enlarged. This not only makes the POI test more complicated but also affects its small sample power performance. Therefore, in this thesis, we fix the variance of the initial disturbance to be $d_1 = 1$. This is less general but not uncommon in the unit root literature and studies of the linear regression model with random walk disturbances, see Pantula et al. (1994) for a discussion.

Another issue is the choice of the alternative point ρ_1. The general approach for picking the pre-selected alternative point where the power is to be maximised was outlined by King (1987b). People usually pick the point such that the maximised power is approximately 0.5 or 0.8. King (1985a) and Shively (1988) both reported that the POI test that optimizes power at 0.5 is favoured because of the overall closeness of its power curve to the power envelope. When testing for autocorrelation, it was shown by King (1985a) that $\rho_1 = 0.5$ and 0.75 are both good choices depending on whether power is required most for weak autocorrelation or for strong autocorrelation. For the random walk hypothesis, no clear indication was given in Dufour and King (1991) on what ρ_1 should be. But applying the same principle as in the autocorrelation test case, if we want to maximise the power for alternatives close to H_0 (i.e., for large positive ρ), $\rho_1 = 0.5$ should be a good candidate. On the other hand, if we fix the alternative point at $\rho_1 = 0$, we would expect the power advantage is spread more evenly over the parameter space under the alternative hypothesis.

We also need to evaluate the median functions of the POI test statistics. It was shown in King (1985a) and King (1987b) that the POI test statistics in our model can be expressed as a ratio of two quadratic forms in normal errors:

$$\begin{aligned} s(\rho_0,\rho_1) &= u'\overline{P}_1'\Sigma^{-1}(\rho_1)\overline{P}_1 u \Big/ u'\overline{P}_0'\Sigma^{-1}(\rho_0)\overline{P}_0 u \\ &= \varepsilon' R'\overline{P}_1'\Sigma^{-1}(\rho_1)\overline{P}_1 R\varepsilon \Big/ \varepsilon' R'\overline{P}_0'\Sigma^{-1}(\rho_0)\overline{P}_0 R\varepsilon \\ &= \varepsilon' A(\rho_1,\rho)\varepsilon \big/ \varepsilon' B(\rho_0,\rho)\varepsilon, \end{aligned} \tag{5.21}$$

where $\overline{P}_1$ and $\overline{P}_0$ are given by (5.18) with ρ replaced by ρ_0 and ρ_1, respectively, and R_0 and R_1 are the transformation matrix given by (5.9) under the null and the alternative hypothesis, respectively.

Hence the median function of the test statistic can be calculated exactly by solving

$$\Pr\{\sum_{i=1}^{T} \lambda_i \xi_i^2 < 0\} = 0.5, \tag{5.22}$$

for $m(\rho)$, where $\lambda_1,\ldots,\lambda_T$ are the eigenvalues (including zeroes and multiple roots) of

$$A(\rho_1,\rho) - m(\rho)B(\rho_0,\rho), \tag{5.23}$$

and $\xi_1^2,\ldots,\xi_T^2$ are independent chi-squared variates with one degree of freedom. This probability can be evaluated using Imhof's (1961) algorithm.

5.4.2 Fixed-point Inversion and Grid Inversion

As discussed in Chapter 3, there are two ways of inverting a significance test statistic to construct MU estimators, the fixed-point inversion and the grid inversion. We apply each of these two methods to the POI tests for our model and compare their effectiveness via simulations. The two estimation procedures are given by:

Method 1: We fix the null value ρ_0 at 1, and the alternative value ρ_1 at 0.5 or 0. To get an estimate, we simply calculate $s(1,0.5)$ or $s(1,0)$ and use the ρ for which the median function $m(\rho)$ is equal to the sample statistics, i.e.,

$$\hat{\rho}^{MU}_{s(1,0.5)} = m^{-1}_{s(1,0.5)}[s(1,0.5)], \tag{5.24}$$

or

$$\hat{\rho}^{MU}_{s(1,0)} = m^{-1}_{s(1,0)}[s(1,0)], \tag{5.25}$$

where the median function $m_{s(1,\rho_1)}(\rho)$ is defined as

$$m_{s(1,\rho_1)}(\rho) = median[s(1,\rho_1) | u \sim N(0, \sigma^2\Sigma(\rho))]. \tag{5.26}$$

Method 2: We fix the null point at 1, but allow the alternative value to change from 1 to −1, and calculate the median functions of the series of POI tests $s(1,\rho)$ each at the corresponding alternative point ρ. We denote the new median function by $m_e(\rho)$, which is given by,

$$m_e(\rho) = median[s(1,\rho) | u \sim N(0, \sigma^2\Sigma(\rho))]. \tag{5.27}$$

It is slightly different from the general method discussed in Chapter 3, in which we allowed the null value to vary. (5.27) is a more straightforward analogue of the construction of the exact power envelope by using the POI tests, as discussed in King (1987b), Elliot (1999) and Podivinsky and King (2000) among others. When the POI tests are used to construct the power envelope, the power envelope is computed by

$$\pi(\rho) = \Pr\{s(1,\rho) \le c_\rho | u \sim N(0, \sigma^2\Sigma^{-1}(\rho))\}, \tag{5.28}$$

where c_ρ is the critical value corresponding to the POI test statistic $s(1,\rho)$ that satisfies

$$\Pr\{s(1,\rho) \le c_\rho | H_0\} = \alpha, \tag{5.29}$$

where α is the preselected significance level. If we let α be 50%, the median function in (5.27) corresponds to the 50% quantile function of the POI test statistic under the series of alternative hypotheses. For convenience, we call the new median function $m_e(\rho)$ a median envelope for the testing problem. In Section 5.4, we will discuss the nature of the median envelope in more detail. It is apparent that the median envelope can be evaluated exactly via the same algorithm used to compute the median function of a single POI test.

Hence the MU estimator $\hat{\rho}^{MU}_{p(1,\rho)}$, which is constructed by the gird inversion method proposed in Chapter 3, will be the solution to the equation

$$s(1,\rho) - m_e(\rho) = 0. \tag{5.30}$$

5.4.3 Median Functions and Median Envelope of the POI Tests

In order to choose a better method from the two for a given design matrix, we first compute and plot the median functions of the POI test statistics for a fixed alternative value and for a grid of alternative values for all design matrices. The median functions for the POI test with an alternative value at 0 are plotted in Figures 5.5 and 5.6. The median functions of the $s(1,0.5)$ test depict a very similar pattern, and therefore are not presented.

The fixed-point median functions are monotonic for most design matrices. Different from the median functions of other test statistics previously examined, the median functions of the POI test statistic are concave for large positive ρ. Heuristically speaking, this indicates a steeper power curve (which can be seen as the mirror image of the median function) when ρ moves away from the null hypothesis. It verifies the proposition that the POI test enjoys better small sample power properties than other tests for these design matrices. We would expect for $X1$ - $X6$, this power advantage of the POI test should lead to a better MU estimator compared with the ones based on less powerful tests, when the fixed-point inversion method is applied. This conjecture will be examined in the Monte Carlo studies reported in Section 5.6.

Figure 5.6 Computed Median Functions of the $s(1,0)$ Test Statistic for $X1$, $X4$, $X5$ and $X6$

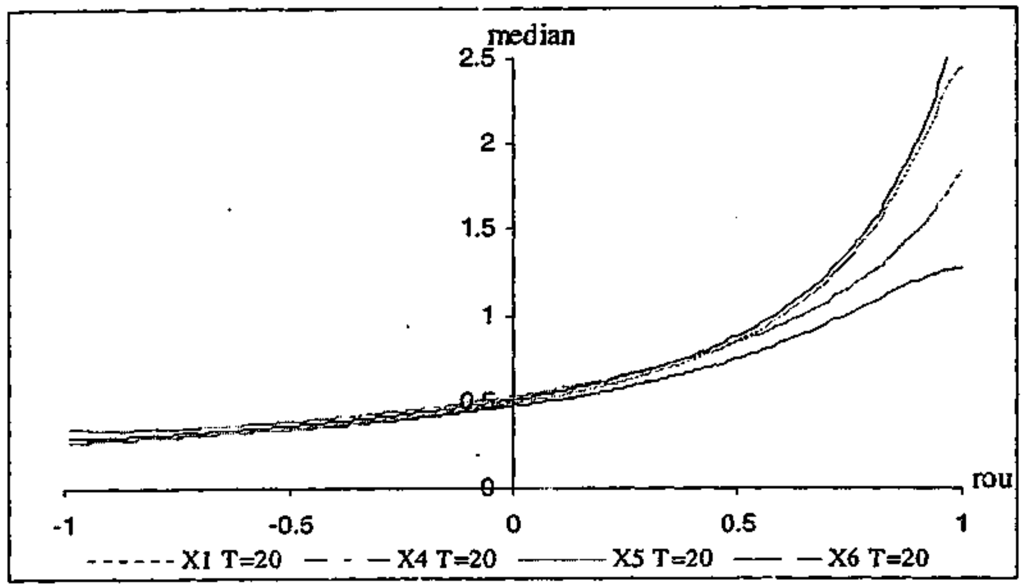
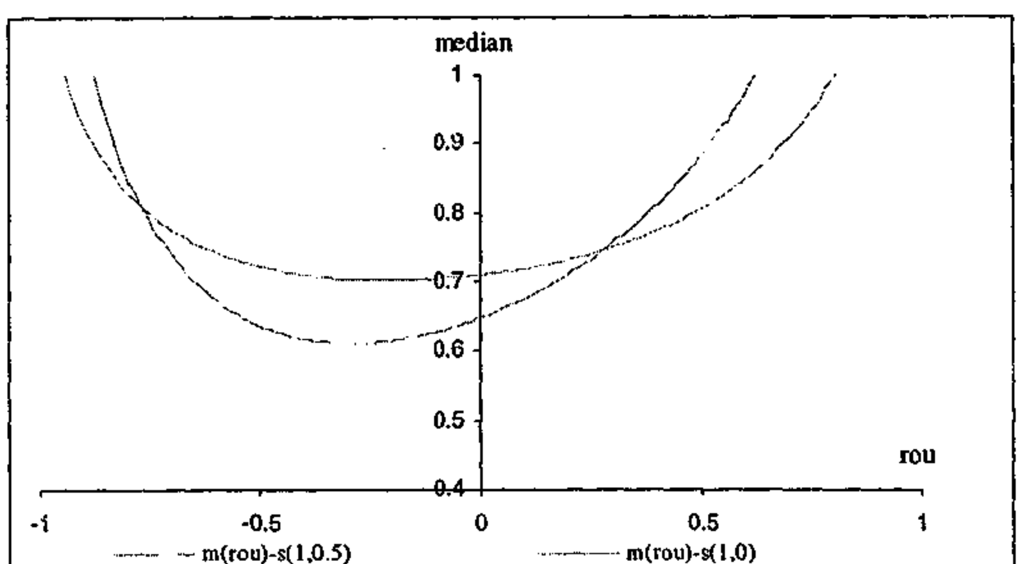


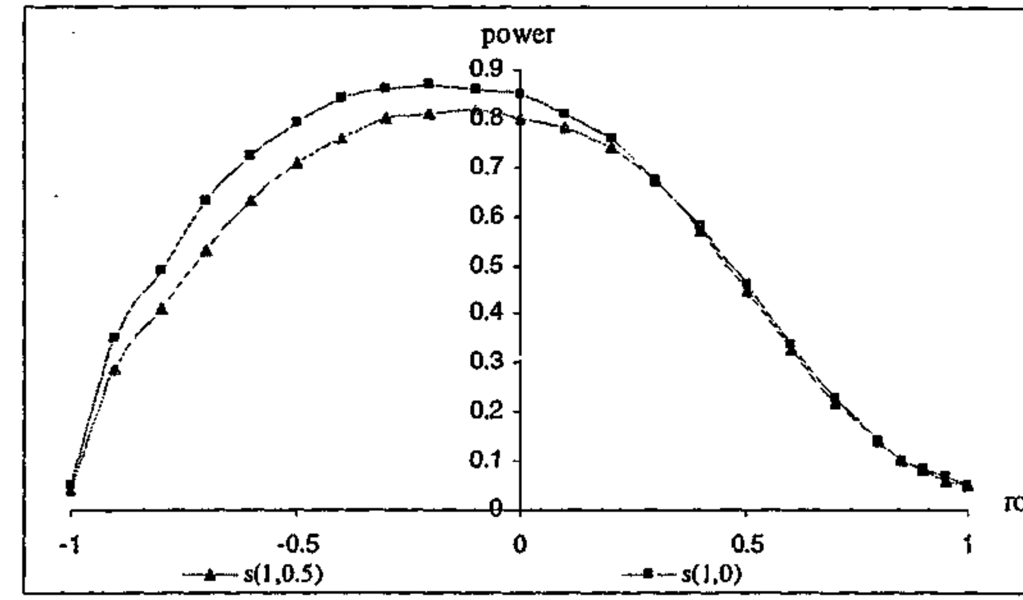
Figure 5.7 Computed Median Functions of the $s(1,0)$ and $s(1,0.5)$ Test Statistics for $X8$, $T=20$



For $X7$ and $X8$, however, the median functions of the $s(1,0)$ test statistic are not monotonic. It seems for these designs, a strong positive autocorrelation is not differentiated from a strong negative autocorrelation. There is a similarity in the two

ends of the parameter space in the median function. For example, this is shown in Figure 5.7 for $X8$ and $T=20$ and 40. We believe this is caused by the non-monotonic power curve of the $s(1,0)$ test on the negative side of H_0 for these design matrices.

Figure 5.8 Computed Power Curves of Two POI Tests for $X8$, $T=20$.



In Figure 5.8, we plot the exact power curves of two POI tests for $X8$ and $T=20$. The power of these tests drop to 0 as $\rho \to -1$. This phenomenon makes it necessary to apply the grid inversion method to these design matrices. In Section 5.4.4, we define some simple criteria to decide, given a design matrix, whether the fixed-point inversion method can be applied or the median envelope approach is needed.

Figure 5.9 presents the median envelopes for different design matrices. The median envelope tells a different story from that of the median function of a single POI test. In contrast with all the other tests considered in this chapter, the median envelope is strictly monotonic for all design matrices. In particular, for design matrices $X7$ and $X8$, the median envelope is the only method known to us that produces a monotonic median function. This will at least guarantee that by inverting the median envelope of the POI test via the grid inversion approach, we can get reliable point estimates.

Figure 5.9 Computed Median Envelopes of the POI Tests for Different Design Matrices, $T = 20$

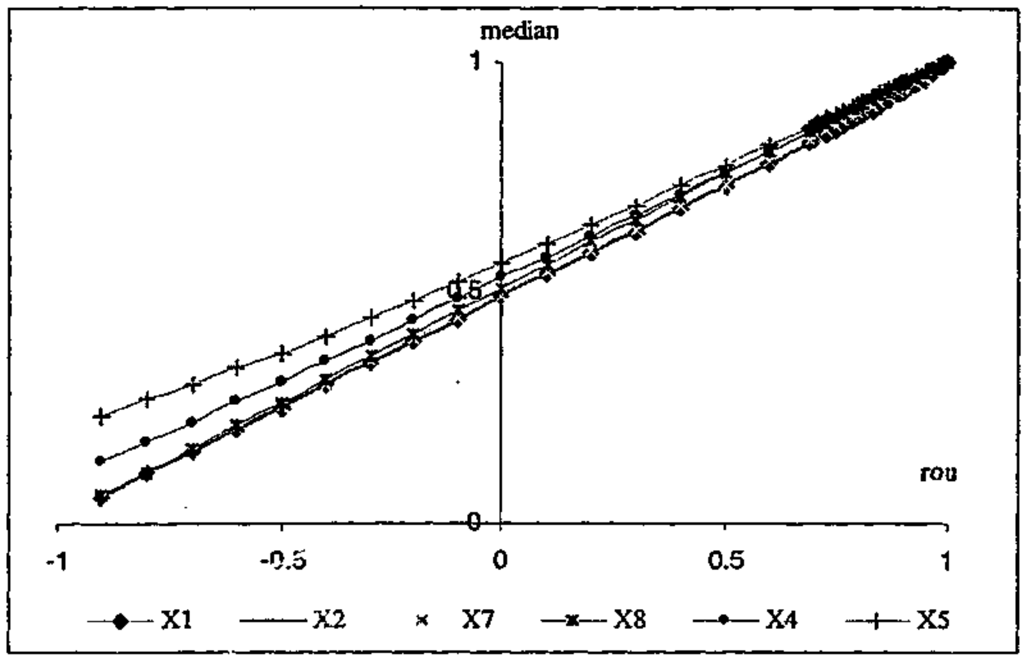


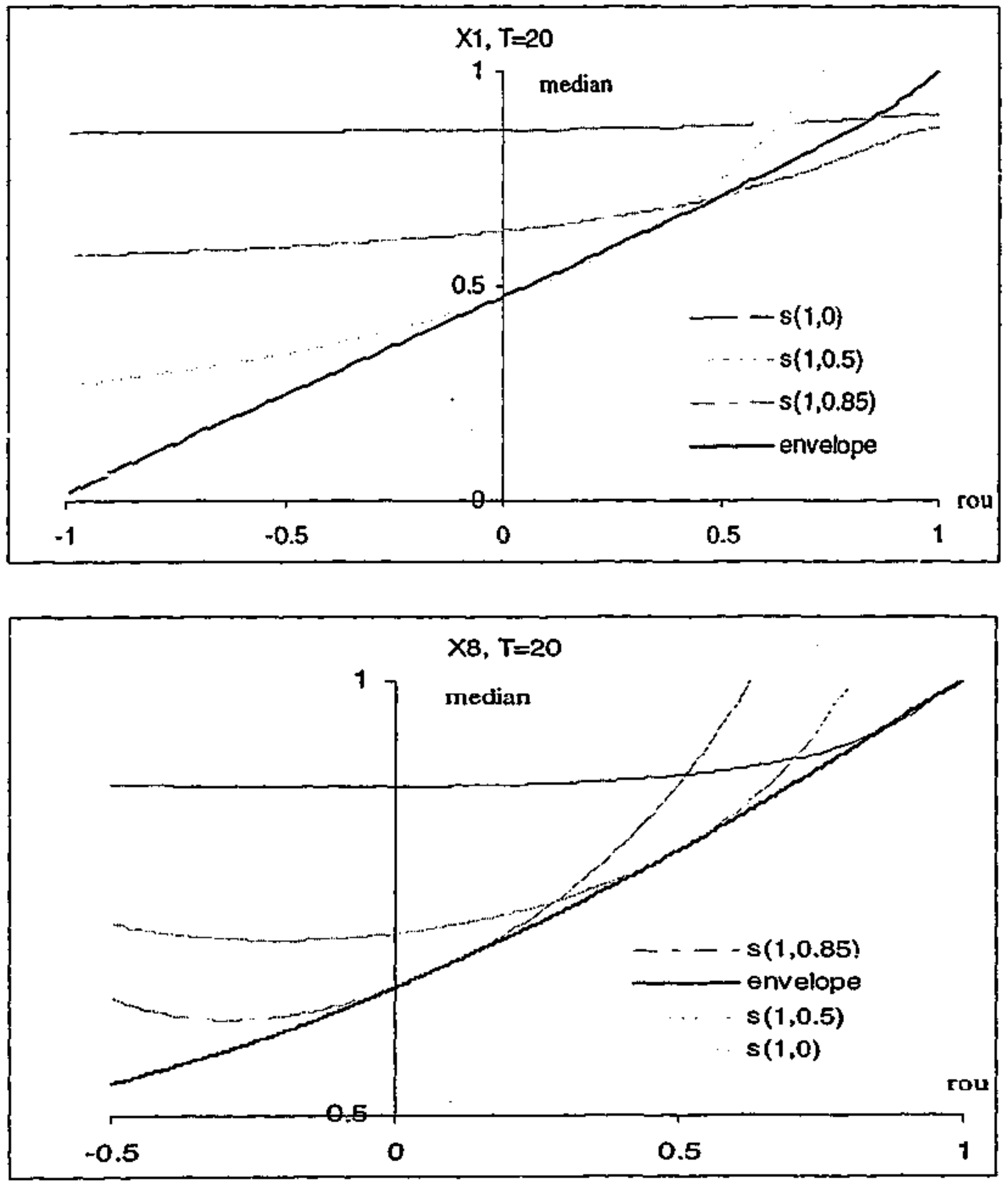
Figure 5.9 also shows that the median envelopes of the POI tests are almost linear in ρ. This is another advantage of the POI test over other tests. Because in practice, we can only calculate or simulate the median functions over a grid of ρ values, for other values not included, interpolation is needed. So if the median function is approximately linear, it will make this interpolation more reliable. But for statistics with non-linear median functions, the interpolation is less accurate if the grid is not fine enough. Sometimes curve fitting is needed to get a reliable median function. For example, Hansen (1999) applied Kernel estimation to the quantile functions of the t-test. This will certainly increase the computational cost.

The sharp difference in the shape of the median envelope and a single median function can be explained by the nature of the median envelope. As each single POI test $s(1,\rho)$ reaches the maximal attainable power at ρ, intuitively, we would expect the median-envelope to be tangent from below to the median function of a POI test with a fixed alternative point. This is equivalent to:

$$med[s(1,\rho_1)|u \sim N(0,\sigma^2\Sigma(\rho_1))] \leq med[s(1,\rho)|u \sim N(0,\sigma^2\Sigma(\rho_1))] \quad (5.31)$$

for any $\rho \neq \rho_1$. The relationship between the median functions of single POI tests and the median envelopes for $X1$ and $X8$ are illustrated in Figure 5.10. The graph gives us some visual justification for the use of the term 'median envelope'.

Figure 5.10 Median Functions and Median Envelopes of the POI Tests; $T = 20$



5.4.4 When to Use Which Method

It is important for us to define the criteria to determine which method is applicable given a particular design matrix. The question we need to answer is, given a design matrix, does the POI test $s(1,\rho_1)$ have a monotonic median function? If the answer is yes, we would use the fixed-point inversion method to compute the MU estimator; otherwise we have to rely on the median envelope approach. We believe the question is equivalent to the following problem: given a design matrix, does the limiting power of $s(1,\rho_1)$ as $\rho \to -1$ drop below its highest level? This is similar to the limiting power problem of the autocorrelation tests such as the DW test when $\rho \to 1$ (or -1), which was considered by Kramer (1985), Zeisel (1987), Kramer and Zeisel (1990) and Bartels (1992) among others. These studies were concerned with testing the null hypothesis of $\rho = 0$. We adopt a similar approach in our thesis to examine the limiting power of the $s(1,0)$ test as $\rho \to -1$.

We consider the one-sided test against $H_a: \rho < 1$, which rejects for small values of $s(1,0)$. The rejection probability can then be expressed as

$$\begin{aligned}&\Pr\{u'(A - c_\alpha B)u < 0\}\\ &= \Pr\{\varepsilon' R'(A - c_\alpha B)R\varepsilon < 0\}\\ &= \Pr\left\{\sum_{i=1}^{T} \gamma_i \xi_i^2 < 0\right\},\end{aligned} \tag{5.33}$$

where

$$\begin{aligned}B &= \overline{P}_1' \, \Sigma^{-1}(1) \overline{P}_1,\\ A &= I - X(X'X)^{-1}X',\\ \overline{P}_1 &= I - X(X'\Sigma^{-1}(1)X)^{-1}X'\Sigma^{-1}(1),\end{aligned}$$

in which R was given by (5.9) and $\Sigma^{-1}(1)$ was given by (5.20). γ_i are the eigenvalues (including zero roots and multiple roots) of

$$\Gamma = R'(A - c_\alpha B)R, \tag{5.34}$$

where c_α is the critical value at a pre-selected significance level, which satisfies

$$\Pr\{s(1,0) < c_\alpha \mid \rho = 1\} = \alpha. \tag{5.35}$$

As noticed by Kramer and Zeisel (1990), if we let

$$V = \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{bmatrix}, \tag{5.36}$$

γ_i are then also the eigenvalues of

$$V^{1/2}\Gamma V^{-1/2} = V(A - c_\alpha B), \tag{5.37}$$

and as $\rho \to -1$, the limit of V is given by

$$V^- = \begin{bmatrix} 1 & -1 & 1 & \cdots(-1)^{T-1} \\ -1 & 1 & -1 & \cdots(-1)^{T-2} \\ 1 & -1 & 1 & \cdots(-1)^{T-3} \\ & & \ddots & \vdots \\ (-1)^{T-1} & & & 1 \end{bmatrix}. \tag{5.38}$$

Following Kramer and Zeisel (1990), if $V^-(A - c_\alpha B) \neq 0$, it must have a rank of one, and therefore only one nonzero eigenvalue. If this eigenvalue is positive, the limiting power will be zero, otherwise the limiting power will be one. For the more common case in which $V^-(A - c_\alpha B) = 0$, in order to derive the limiting rejection probability, we replace the eigenvalues γ_i in (5.33) by

$$\tilde{\gamma}_i = \gamma_i / (1+\rho), \tag{5.39}$$

since this does not affect the rejection probability (5.33). The $\tilde{\gamma}_i$ are the eigenvalues of

$$\frac{V}{1+\rho}(A-c_\alpha B), \tag{5.40}$$

As $\rho \to -1$,

$$\begin{aligned}\lim_{\rho\to-1}(1+\rho)^{-1}V(A-c_\alpha B) &= \lim(1+\rho)^{-1}(V-V^{+})(A-c_\alpha B)\\ &= W^{-}(A-c_\alpha B),\end{aligned} \tag{5.41}$$

where W^- is given by,

$$W^{-} = \lim_{\rho\to-1}(1+\rho)^{-1}(V-V^{-})$$
$$= \begin{bmatrix} 0 & 1 & -2 & \cdots(T-1)(-1)^{T-1} \\ 1 & 0 & 1 & \cdots(T-2)(-1)^{T-2} \\ -2 & 1 & 0 & \cdots(T-3)(-1)^{T-3} \\ \vdots & & \ddots & \vdots \\ (T-1)(-1)^{T-1} & & \cdots & 0 \end{bmatrix}. \tag{5.42}$$

Therefore the nonzero eigenvalues of the matrix

$$W^{-}(A-c_\alpha B) \tag{5.43}$$

completely determine the limiting power of the test $s(1,0)$ as $\rho \to -1$.

Hence given a design matrix, in order to decide whether to invert the median function of a fixed POI test ($s(1,0)$, say), or to invert the median envelope, we only need to examine the limiting power of the test as $\rho \to -1$. If the limit is 1, it indicates that the median function of the single POI test is monotonic for this design matrix and we can use the fixed-point inversion method. If the probability is a small number (or sometimes 0), then it is safer to use the grid inversion of the median envelope. In

practice, we only need to examine the smallest eigenvalue of (5.43). As a rule of thumb, if this eigenvalue is small in absolute value (compared with the largest positive one), the limiting power will be small, and the median function is likely to be non-monotonic. If the smallest eigenvalue is negative but large in absolute value, the median function is likely to be monotonic.

Table 5.1
The Smallest and Largest Eigenvalue of Matrix (5.43) and the Limiting Power of the s(1,0) Test as $\rho \to -1$

	$X1$ $T=20$	$X7$ $T=15$	$X7$ $T=20$	$X8$ $T=15$	$X8$ $T=20$
Smallest Eigenvalue	-45606	-39	-287	-3.9	-10.88
Largest Eigenvalue	1.277	8716	2960	24492	24578
Power	0.995	0.05	0.085	0.02	0.027

Notes: The limiting powers are based on 20,000 simulations.

Although it is more accurate to work out the exact limiting probability (5.33), we found for the design matrices considered in this study, this practical criterion works quite well. For example, the smallest and the largest eigenvalues of the limiting matrix (5.43) together with the limiting power of the POI tests for some design matrices are reported in Table 5.1. For $X1$, the smallest and largest eigenvalues are −45606, 1.277, respectively. This leads to a limiting power of 0.995, which indicates a monotonic power curve and a monotonic median function of the $s(1,0)$ test. The smallest and largest eigenvalues for $X8$ and $T=20$ are −10.88, 24578, respectively, and the $s(1,0)$ test has a limiting power of 0.027. In this case, the median function of a single POI test is not monotonic, as shown in Figure 5.7.

5.5 Comparing the Estimators Based on Different Tests

5.5.1 Experimental Design

In the sections that follow, Monte Carlo studies designed to examine the small sample performance of the MU estimators based on different test statistics are outlined. We also compare these MU estimators with the more conventional counterparts in terms of bias and total risk. We conducted three sets of experiments:

1. In the first experiment, we focused on the fixed-point inversion method and examined the relationship between the power of a test and the performance of the estimator by inverting its median function. For $X1$, all tests have monotonic median functions and estimators can be constructed by inverting any of them. We used the following tests: DW test, score tests ($Sc(0)$ and $Sc(0.5)$) and the POI tests ($s(0,0.5)$ and $s(1,0.5)$) together with Andrews' estimator. As these tests have different power properties in small samples, it is interesting to compare the performance of the MU estimators based on them. We compared their small-sample bias, variance and RMSE for different ρ values.

2. In the second experiment, we applied both fixed-point inversion and median envelope inversion, and compared the two methods, in order to further disclose the relationship between the power of a test and the properties of the resulting estimator. For $X1$ - $X6$, MU estimators can be computed by either inverting a single POI test statistic or inverting the median envelope. We compared these two methods. Andrews' estimator was also calculated as the median function of the OLS estimator is monotonic for these designs. These MU estimators were also compared with the more conventional counterparts: $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$.

3. In the third experiment, we focused on the median envelope method when other methods cannot be used. For $X7$ and $X8$, Andrews' method breaks down. Inverting the POI envelope becomes the only option for constructing MU estimators. We compared the new estimator with $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$.

The sample sizes used were 20, 40 and 60. 2000 estimates were calculated for all the estimators. Their small sample bias and efficiency were compared for $\rho = 1$, 0.95, 0.9, 0.8, 0.6, 0.4, 0.2, 0, -0.4, and -0.8. As the estimation of ρ is invariant to β and σ^2, they were set to be vector of ones and one, respectively.

5.5.2 Estimation Results

The MU estimators based on inverting different test statistics for $X1$ and $T = 20$ are presented in Table 5.2. King (1985a) among others, reported that for this design matrix, the power advantage of the POI tests over the DW test is minimal compared with some other design matrices. This similarity of the powers of the tests is accurately reflected in the performance of the estimators based on them. Each of these estimators successfully corrects the bias in $\hat{\rho}_{OLS}$ for $\rho = 0.95$, 0.8, 0.6. The 6 MU estimators behave rather similarly in terms of mean, variance and RMSE, although the one based on the DW test appears to have slightly smaller RMSEs compared with the other five estimators. This is consistent with the findings in King (1985a) and Honda (1989), that for this design matrix, the DW test has good small sample power. The RMSEs of $\hat{\rho}_{OLS}$ have been reduced significantly by all MU estimators. For example, it is reduced from 0.48 to 0.28 by the MU estimator based on the DW test statistic for $\rho = 0.95$, and from 0.40 down to 0.28 for $\rho = 0.8$. This set of results show that if the tests have similar power properties, the MU estimators based on inverting their median functions will also have similar small sample bias and RMSE. For the more interesting case of different tests behaving differently, we turn to other design matrices.

The results of the second experiment – comparison of $\hat{\rho}_{OLS}$, $\hat{\rho}_{MLE}$, Andrews' estimator ($\hat{\rho}_A$), the MU estimator based on inverting the median envelope of the POI tests ($\hat{\rho}_E^{MU}$) and the two MU estimators by inverting two single POI test statistics ($\hat{\rho}_{s(1,0)}^{MU}$ and $\hat{\rho}_{s(1,0.5)}^{MU}$) for $X1$ - $X6$ and $T = 20$, 40 are reported in Tables 5.3a – 5.3f. The estimation results for $T = 60$ are presented in Table 5.4.

The conventional estimators, $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$, as shown by many previous studies, are biased for all designs and sample sizes, especially for large positive ρ values. The magnitude of the downward bias not only depends on ρ, but also on the design matrix structure. For $X6$ and $T=20$, for example, the biases of $\hat{\rho}_{OLS}$ at $\rho=1$ and 0.8 reached –0.66 and –0.49, respectively. $\hat{\rho}_{MLE}$ has biases similar to those of $\hat{\rho}_{OLS}$ for $X1$, $X2$ and $X3$, while for $X4$, $X5$ and $X6$, it is less biased than $\hat{\rho}_{OLS}$. The biases of these two estimators become smaller for $T=60$.

The two estimators based on inverting a single POI test statistic, $\hat{\rho}^{MU}_{s(1,0)}$ and $\hat{\rho}^{MU}_{s(1,0.5)}$, behave rather similarly for positive ρ. Both estimators are essentially MU for all these design matrices and sample sizes on the positive side of 0. The bias correction is very effective. For example, for $X6$ and $T=20$, the biases in $\hat{\rho}^{MU}_{s(1,0)}$ at $\rho=1$ and 0.8 are both 0, while the biases in $\hat{\rho}_{OLS}$ are –0.66 and –0.49, respectively. For sample sizes of 20, 40 and 60, the two new estimators show almost no bias for all design matrices. For large negative ρ values ($\rho=-0.8$ and -0.95) however, $\hat{\rho}^{MU}_{s(1,0)}$ outperforms $\hat{\rho}^{MU}_{s(1,0.5)}$ for $T=20$ and for $X3$ and $X5$ in terms of bias-correction. This could be explained by the construction of the POI tests. While $s(1,0)$ is designed to maximise the power at $\rho=0$, $s(1,0.5)$ is designed to maximise its power at $\rho=0.5$, therefore for negative ρ values, we would expect the power of $s(1,0.5)$ drops more than that of $s(1,0)$. The remaining bias in $\hat{\rho}^{MU}_{s(1,0.5)}$ is probably due to its lack in power for large negative ρ value compared with $s(1,0)$. When sample size increases, the difference between the two estimators becomes minimal.

This bias-correction by using MU estimators also results in a significant reduction in RMSE. Compared with the biased estimators, especially $\hat{\rho}_{OLS}$, the RMSEs of the two MU estimators are much smaller for large positive ρ values. For example, the RMSEs of $\hat{\rho}_{OLS}$ at $\rho=1$, 0.9, 0.8 for $X5$ and $T=20$ are 0.62, 0.54 and 0.48, while the corresponding RMSEs of $\hat{\rho}^{MU}_{s(1,0)}$ are 0.23, 0.24 and 0.26, respectively. When the sample size goes up to 40 and 60, the RMSEs of the new MU estimators are also smaller than those of $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$, but by a smaller margin. For negative ρ values, however, the MU estimators appear to have RMSEs similar to or slightly

higher than those of $\hat{\rho}_{MLE}$. For example the RMSEs of $\hat{\rho}^{MU}_{s(1,0)}$ at $\rho=-0.6$ and -0.8 for $X4$ and $T=20$ are 0.25 and 0.21, compared with those of $\hat{\rho}_{MLE}$ in this case, of 0.17 and 0.14, respectively. The RMSEs of $\hat{\rho}^{MU}_{s(1,0.5)}$ for large negative ρ values and $T=20$ are larger than those of $\hat{\rho}^{MU}_{s(1,0)}$, as expected. Based on this comparison, we recommend $\hat{\rho}^{MU}_{s(1,0)}$ if a fixed-point inversion method can be used for a given design matrix, especially when there is no information about the possible magnitude and direction of the autocorrelation in the disturbances.

When we compare the proposed MU estimators with Andrews' estimator, we find the results favour $\hat{\rho}^{MU}_{s(1,0)}$ for most cases. For $X1$, $\hat{\rho}_A$ performs in a very similar way to $\hat{\rho}^{MU}_{s(1,0)}$ in terms of unbiasedness. But for all other design matrices and for $T=20$ and 40, $\hat{\rho}_A$ does not eliminate the bias in $\hat{\rho}_{OLS}$ as effectively as $\hat{\rho}^{MU}_{s(1,0)}$ for moderate positive ρ values. In the worst case of $X4$, $X5$ and $X6$ and $T=20$, $\hat{\rho}_A$ is still quite biased for $\rho=0.6$, 0.5, 0.4, with biases of -0.16, -0.18, -0.17, respectively. The RMSEs of $\hat{\rho}_A$ are generally larger than those of $\hat{\rho}^{MU}_{s(1,0)}$ for positive ρ values, while similar for negative ρ values. The difference is most apparent for $\rho\geq0.2$ and for $X4$ and $X5$. We attribute the inferior performance of $\hat{\rho}_A$ to the lack of power of the test $\hat{\rho}_{OLS}-\rho_0$ as a test for random walk disturbances or a test for autocorrelation for these design matrices. Although the median functions are still monotonic, indicating the power curves are monotonic, these median functions are non-linear and convex on the positive side of 0, as shown in Section 5.3. In contrast to this test, the POI test $s(1,0)$ has a concave median function in this case, which produces a less biased estimator. The results from this experiment lead us to believe that there is a direct link between the power of a test and the efficiency of the point estimator based on inverting the test statistic's median function. Therefore, when we choose a test statistic to construct MU estimators, it is crucial to choose the test statistic with the best power properties in small samples.

Compared with $\hat{\rho}^{MU}_{s(1,0)}$, the estimator $\hat{\rho}^{MU}_{E}$, which was computed by inverting the median envelope of the POI tests also performs well for $T=40$ and 60, as it is essentially MU for all design matrices and all ρ values. For negative ρ values, $\hat{\rho}^{MU}_{E}$

appears to correct the bias more accurately than $\hat{\rho}^{MU}_{s(1,0)}$ for $X5$ and $X6$. The RMSEs of $\hat{\rho}^{MU}_{E}$ are also generally slightly smaller than those of $\hat{\rho}^{MU}_{s(1,0)}$, indicating $\hat{\rho}^{MU}_{E}$ is a very reliable estimator to use for these sample sizes. But for $T=20$, $\hat{\rho}^{MU}_{E}$ is not exactly MU for large positive ρ. The remaining bias is much smaller than those of $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$, but still quite apparent especially for $X1$, $X2$ and $X3$. For example, the biases of $\hat{\rho}^{MU}_{E}$ for $X1$ and $T=20$ at $\rho=1$, 0.9 and 0.8 are -0.14, -0.08 and -0.06 respectively. However, $\hat{\rho}^{MU}_{E}$ outperforms $\hat{\rho}^{MU}_{s(1,0)}$ for negative ρ values, with less bias and smaller RMSEs for all design matrices. The remaining bias in $\hat{\rho}^{MU}_{E}$ for large positive ρ might be linked to the phenomenon described in Section 5.4, namely, the odd behaviour of the median envelope for $X1$ and $T=20$. The median envelope is not tangent from below to the median functions of the POI tests in the neighbourhood of 1, although it was expected to be. The bias in $\hat{\rho}^{MU}_{E}$ is much less for $X6$, for example, and the median envelope for $X6$ is 'well-behaved' compared with that for $X1$. Therefore, if there is strong evidence for highly persistent disturbances, then $\hat{\rho}^{MU}_{s(1,0)}$ is the preferable MU estimator for small sample sizes; otherwise $\hat{\rho}^{MU}_{E}$ can also be used as a reliable estimator with little bias and low risk regardless of design matrices.

The results from experiment 3 are reported in Tables 5.9 and 5.10, in which we compare the estimator based on inverting the median envelope of the POI tests, $\hat{\rho}^{MU}_{E}$ with $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$ for $X7$ and $X8$, and for $T=20$, 40, and 60. This design is usually used as an extreme case in favour of the POI tests, as the OLS residuals become very poor estimates of the true disturbances and therefore tests based on them perform badly. This is precisely reflected in our estimation results. The bias in $\hat{\rho}_{OLS}$ is appalling. For $T=60$ and $\rho=1$, it still has a bias of -0.64 for $X7$ and -0.83 for $X8$. Ironically, in this case when bias-correction is most needed, most bias-correction methods do not work. As mentioned earlier, $\hat{\rho}_{A}$ fails to deliver unique estimates due to the non-monotonic median function of $\hat{\rho}_{OLS}$. The fixed-point inversion of a single POI test is not applicable as the criteria we set out in Section 5.5 are not met. The limiting power of these POI tests as $\rho\to-1$ drop below one. Therefore the proposed estimator $\hat{\rho}^{MU}_{E}$ becomes our only choice. Fortunately, the results endorse its performance. For $T=60$, $\hat{\rho}^{MU}_{E}$ is almost exactly MU for both $X7$

and $X8$ and for all ρ values, while for smaller sample sizes, $\hat{\rho}^{MU}_{E}$ also effectively reduces the bias in other estimators significantly, but the bias is not completely removed for $\rho\geq 0.85$, especially for $T=20$. For example, for $X7$ and $T=20$, $\hat{\rho}^{MU}_{E}$ has biases of -0.07, -0.05 and -0.05 at $\rho=1$, 0.95 and 0.9, respectively. We argue that compared with the bias in other estimators, the remaining bias in $\hat{\rho}^{MU}_{E}$ is ignorable. In the same example, the biases of $\hat{\rho}_{OLS}$ are -0.85, -0.79 and -0.69. $\hat{\rho}^{MU}_{E}$ also has RMSEs that are less than one-third of those of $\hat{\rho}_{OLS}$ for large positive ρ values and for $T=20$. The reduction in RMSE is also significant for moderate positive ρ values and for larger sample sizes. The only exception is when $\rho=0$ or 0.2 and $T=20$, the RMSEs of $\hat{\rho}^{MU}_{E}$ are slightly higher than those of $\hat{\rho}_{OLS}$. We conclude that for Watson's matrix, where the bias problem in estimating the autocorrelation coefficient is at its extreme, $\hat{\rho}^{MU}_{E}$ is a good remedy for this deficiency.

To summarise our major findings from the Monte Carlo studies, we notice that the estimators based on inverting the POI test statistics effectively correct the serious bias associated with $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$ for all design matrices. When constructing an estimator by inverting a test statistic, the bias and risk of the estimator directly depends on small-sample power properties of the chosen test. The more powerful POI test, for example, does produce a better MU estimator than other test statistics for most design matrices. The combination of the fixed-point inversion and the grid inversion can overcome the difficulty encountered by Andrews' estimator, and produce a reliable estimator regardless of design matrix structure. In practice, we recommend that given a design matrix, one examines the limiting power of the POI tests as described in Section 5.5. If the criterion is satisfied, $\hat{\rho}^{MU}_{s(1,0)}$ is the preferable MU estimator, and if the criterion is not met, $\hat{\rho}^{MU}_{E}$ is to be used.

5.6 Robustness to Non-normal Errors

In this section, we study the robustness of the proposed estimator to the disturbances that are non-normal. The robustness of estimators in autoregressive models have been studied by many researchers. Fiebig et al. (1991) examined the robustness of LS estimators for a more general covariance structure in the linear regression model. In the unit root literature, it is well known that the OLS estimator and the related Dickey-Fuller-type tests are not robust to misspecification of the error structure. In particular, Phillips and Peron (1987) and De Jong et al. (1992b) pointed out that if there is a MA component or if the order of autoregression is misspecified, the power of these unit root tests can be very low. On the other hand, a very important feature of the median as a location estimator is its robustness to non-normal errors. We would expect this property to carry over to the proposed MU estimators. For $X1$, Andrews (1993) examined his MU estimator and found it to be quite robust to non-normal errors.

To examine the robustness of $\hat{\rho}_{MU}$, we do not need to actually compute the estimates under different error distributions. Instead, we only need to study the median functions in these circumstances. We focus on the estimator based on the median envelope in our discussion. The same result applies to the estimators based on the fixed-point inversion method. Because $\hat{\rho}^{E}_{MU}$ is the solution to

$$s(1,\rho) - med[s(1,\rho)|u \sim f(0,\sigma^2\Sigma(\rho))] = 0, \qquad (5.44)$$

where $f(\bullet)$ is the distribution function of the errors ε_t, and the form of $\Sigma(\bullet)$ reflects the specification of the error structure. If the median function under different $f(\bullet)$ and $\Sigma(\bullet)$ are similar to the one under Gaussian AR(1) disturbances, then the solution to the above equation under different error structures will be similar to the estimates computed assuming normal AR(1) errors. In other words, the estimator is a robust one.

Tables 5.6a and 5.6b and Figures 5.11a and 5.11b presents the median functions of $\hat{\rho}_{OLS}$ and the POI median envelopes defined in Section 5.4 for different design matrices with 20 observations, under different error structures. The error structures considered include: 1. Skewed distributions: χ^2_2-distribution and the log-normal distribution; 2. Heavy tail distributions: Student t-distribution with 3 degrees of freedom and Cauchy distribution; 3. Different error structures: AR(2) errors,

MA(1) errors and the first-order *ARCH* errors with the autoregressive parameter set to 0.85, which is consequently denoted by *ARCH*(0.85).

The results show that, apart from the AR(2) and MA errors, the median envelopes of the POI tests are very robust to all these error misspecifications for both design matrices with a sample size of 20. Without knowing the error structure, it is impossible to tell if the median envelope is computed assuming the normal distribution. The envelopes are consistently monotonic and almost linear on the positive side of 0. This leads us to believe that the new MU estimator proposed in this chapter is very robust to different types of error misspecifications. However, with MA errors present or if the order of autoregression is misspecified in the disturbances, the median function is significantly different from the one under Gaussian AR(1) disturbances. Therefore it is important to test for such misspecifications in the error structure before applying the proposed estimation procedure. This precaution was echoed in many previous studies in the unit root literature, such as Schwert (1987), Nankervis and Savin (1988b), Phillips and Perron (1988), DeJong et al. (1992b) and Kiviet and Dufour (1997).

Andrews' estimator, however is not as robust as $\hat{\rho}_{MU}$ for these misspecifications. For large positive ρ, the median functions of $\hat{\rho}_{OLS}$ under Cauchy distributions and ARCH errors both depart quite a distance from the ones under normal errors for most design matrices. For $X2$ and $X5$, the median function under χ^2_2 are also not close to the ones under normal errors. In these circumstances, Andrews' estimator is more sensitive to the error structures than the proposed one.

5.7 Concluding Remarks

We applied the second approach proposed in Chapter 3 – inverting the median function of a significance test to the linear regression model with AR(1) or random walk disturbances in this chapter. Andrews' (1993) estimator breaks down for some design matrices due to the problem of non-unique estimates, which is caused by the non-monotonic median function of the OLS estimator. The same problem also plagues the DW test, t test and the LM test for various design matrices. It was shown that the POI test has a monotonic and convex median function for most design matrices on the positive side of ρ. This reflects its good small sample power properties. An easy-to-use criterion based on the limiting power of the test is given to determine if the median function of a single POI test statistic is non-monotonic given a particular design matrix.

If the median function of a single POI test is monotonic for a design matrix, the MU estimator based on the fixed-point inversion method is almost exactly MU for all ρ values and has smaller RMSEs compared with other estimators. It generally performs better than Andrews' estimator except in the model with only an intercept and a time trend as the regressors.

For the design matrices that a single POI test fails to deliver a monotonic median function, inverting the median envelope of a series of POI tests is the recommended method of constructing a MU estimator. It was shown that the proposed estimator almost eliminates the bias present in the OLS and MLE estimators. The bias correction is substantial for these designs, as they usually represent the extreme cases in which the small sample biases of the conventional estimators are most serious.

Finally, we examined the robustness of the proposed estimator to non-normal errors and error structure misspecifications. It was found that the new estimator is more robust than the one based on the OLS estimator. It also performs well under different error structures, except for errors with an MA component or generated by a higher order autoregressive process.

Table 5.2
Medians, Means, Variances and RMSEs of the MU Estimators Based on Different Test Statistics in the Linear Regression with AR(1) Disturbances for Design Matrix $X1$; $T=20$

	$\hat{\rho}_{OLS}$	$\hat{\rho}_{A}$	$\hat{\rho}_{DW}$	$\hat{\rho}_{s(0,0.5)}$	$\hat{\rho}_{s(1,0.5)}$	$\hat{\rho}_{Sc(0)}$	$\hat{\rho}_{Sc(0.5)}$
				$\rho=0.95$			
median	0.555	0.893	0.98	0.935	0.969	0.962	0.965
mean	0.518	0.786	0.811	0.801	0.809	0.809	0.810
variance	0.044	0.066	0.061	0.063	0.062	0.063	0.062
RMSE	0.480	0.304	0.284	0.292	0.286	0.288	0.285
				$\rho=0.8$			
median	0.484	0.761	0.794	0.774	0.792	0.793	0.791
mean	0.457	0.719	0.740	0.730	0.738	0.737	0.739
variance	0.045	0.075	0.073	0.074	0.074	0.075	0.074
RMSE	0.403	0.286	0.277	0.281	0.279	0.282	0.279
				$\rho=0.6$			
median	0.364	0.585	0.601	0.588	0.603	0.601	0.602
mean	0.347	0.579	0.592	0.584	0.589	0.589	0.592
variance	0.045	0.082	0.080	0.080	0.081	0.081	0.080
RMSE	0.331	0.287	0.284	0.284	0.285	0.285	0.283
				$\rho=0.4$			
median	0.274	0.315	0.403	0.386	0.403	0.401	0.401
mean	0.249	0.299	0.394	0.384	0.391	0.388	0.393
variance	0.045	0.082	0.082	0.083	0.083	0.082	0.083
RMSE	0.302	0.284	0.288	0.288	0.286	0.288	0.282

Notes: All experiments are based on 2000 replications.

Table 5.3a

Medians and RMSEs of MU Estimators Based on the POI Test Statistics in the Linear Regression with AR(1) Disturbances for Design Matrix $X1$

ρ		$T=20$						$T=40$					
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$
1.00	median	0.58	0.61	0.97	0.86	1.00	0.99	0.78	0.80	0.97	0.92	0.99	0.98
	RMSE	0.50	0.48	0.31	0.31	0.29	0.30	0.28	0.26	0.17	0.17	0.16	0.16
0.90	median	0.55	0.58	0.90	0.82	0.93	0.89	0.73	0.75	0.90	0.87	0.91	0.90
	RMSE	0.43	0.42	0.28	0.27	0.26	0.27	0.23	0.22	0.16	0.15	0.15	0.15
0.80	median	0.49	0.52	0.78	0.74	0.81	0.78	0.67	0.67	0.80	0.79	0.81	0.80
	RMSE	0.40	0.38	0.28	0.27	0.27	0.28	0.20	0.20	0.16	0.15	0.15	0.16
0.70	median	0.43	0.45	0.68	0.67	0.71	0.69	0.58	0.58	0.69	0.69	0.71	0.70
	RMSE	0.36	0.35	0.29	0.27	0.28	0.29	0.20	0.20	0.18	0.17	0.17	0.17
0.60	median	0.36	0.37	0.58	0.57	0.60	0.58	0.49	0.50	0.60	0.60	0.61	0.60
	RMSE	0.34	0.34	0.30	0.28	0.29	0.29	0.19	0.19	0.17	0.16	0.16	0.17
0.50	median	0.29	0.30	0.49	0.48	0.51	0.48	0.40	0.40	0.50	0.50	0.51	0.49
	RMSE	0.31	0.31	0.29	0.29	0.29	0.29	0.18	0.18	0.17	0.16	0.16	0.17
0.40	median	0.22	0.22	0.39	0.39	0.41	0.38	0.31	0.31	0.40	0.40	0.41	0.39
	RMSE	0.30	0.30	0.29	0.29	0.29	0.29	0.18	0.18	0.17	0.18	0.17	0.17
0.20	median	0.06	0.06	0.20	0.21	0.22	0.20	0.13	0.13	0.20	0.20	0.21	0.19
	RMSE	0.26	0.27	0.27	0.29	0.27	0.28	0.17	0.17	0.17	0.20	0.17	0.18
0.00	median	-0.11	-0.11	0.00	0.01	0.01	0.00	-0.05	-0.06	0.00	-0.01	0.01	0.00
	RMSE	0.24	0.24	0.26	0.29	0.27	0.28	0.16	0.16	0.17	0.22	0.17	0.18
-0.40	median	-0.44	-0.44	-0.39	-0.39	-0.38	-0.39	-0.43	-0.43	-0.41	-0.40	-0.39	-0.40
	RMSE	0.19	0.20	0.23	0.24	0.25	0.29	0.14	0.14	0.15	0.15	0.16	0.18
-0.80	median	-0.78	-0.79	-0.80	-0.80	-0.78	-0.78	-0.79	-0.79	-0.80	-0.79	-0.78	-0.78
	RMSE	0.15	0.15	0.18	0.18	0.20	0.27	0.11	0.11	0.11	0.12	0.13	0.17

Notes: All experiments are based on 2000 replications.

Table 5.3b

Medians and RMSEs of MU Estimators Based on the POI Test Statistics in the Linear Regression with AR(1) Disturbances for Design Matrix $X2$

ρ		$T=20$						$T=40$					
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$
1.00	median	0.42	0.55	0.64	0.88	0.99	0.94	0.75	0.80	1.00	0.92	1.00	1.00
	RMSE	0.65	0.57	0.55	0.34	0.35	0.36	0.30	0.26	0.26	0.17	0.14	0.15
0.90	median	0.40	0.51	0.60	0.80	0.89	0.86	0.71	0.75	0.96	0.87	0.98	0.92
	RMSE	0.56	0.50	0.49	0.29	0.31	0.32	0.25	0.22	0.26	0.15	0.14	0.15
0.80	median	0.35	0.44	0.48	0.72	0.81	0.75	0.63	0.67	0.77	0.78	0.84	0.81
	RMSE	0.52	0.48	0.48	0.30	0.32	0.34	0.23	0.21	0.28	0.15	0.17	0.17
0.70	median	0.31	0.37	0.37	0.66	0.71	0.66	0.55	0.58	0.66	0.69	0.74	0.70
	RMSE	0.47	0.44	0.45	0.29	0.31	0.34	0.22	0.20	0.27	0.16	0.18	0.17
0.60	median	0.25	0.29	0.32	0.56	0.56	0.59	0.47	0.49	0.57	0.60	0.63	0.61
	RMSE	0.43	0.41	0.43	0.30	0.32	0.36	0.21	0.20	0.25	0.17	0.18	0.18
0.50	median	0.20	0.22	0.26	0.48	0.47	0.48	0.38	0.40	0.45	0.50	0.53	0.49
	RMSE	0.39	0.38	0.40	0.31	0.32	0.36	0.20	0.20	0.23	0.17	0.18	0.18
0.40	median	0.13	0.14	0.16	0.39	0.39	0.35	0.29	0.31	0.36	0.41	0.43	0.40
	RMSE	0.36	0.36	0.38	0.32	0.31	0.37	0.20	0.19	0.22	0.19	0.18	0.19
0.20	median	-0.01	-0.01	0.00	0.21	0.19	0.15	0.10	0.11	0.16	0.20	0.21	0.18
	RMSE	0.30	0.31	0.32	0.32	0.31	0.38	0.19	0.19	0.22	0.22	0.18	0.20
0.00	median	-0.16	-0.17	-0.12	0.01	0.01	-0.07	-0.06	-0.07	-0.04	-0.01	0.01	-0.01
	RMSE	0.26	0.27	0.28	0.33	0.30	0.39	0.17	0.17	0.19	0.24	0.18	0.20
-0.40	median	-0.46	-0.48	-0.44	-0.38	-0.40	-0.47	-0.43	-0.44	-0.41	-0.40	-0.40	-0.43
	RMSE	0.19	0.20	0.20	0.28	0.27	0.40	0.14	0.14	0.15	0.16	0.17	0.21
-0.80	median	-0.77	-0.81	-0.77	-0.79	-0.74	-0.68	-0.79	-0.80	-0.77	-0.80	-0.82	-0.85
	RMSE	0.15	0.14	0.16	0.20	0.23	0.42	0.10	0.10	0.11	0.11	0.14	0.18

Notes: All experiments are based on 2000 replications.

Table 5.3c

Medians and RMSEs of MU Estimators Based on the POI Test Statistics in the Linear Regression with AR(1) Disturbances for Design Matrix $X3$

ρ		$T=20$						$T=40$					
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$
1.00	median	0.47	0.58	0.92	0.98	0.99	1.00	0.69	0.77	0.96	0.91	0.93	0.93
	RMSE	0.61	0.54	0.51	0.32	0.34	0.32	0.36	0.30	0.33	0.18	0.19	0.18
0.90	median	0.45	0.54	0.77	0.83	0.91	0.94	0.67	0.73	0.89	0.87	0.88	0.88
	RMSE	0.53	0.48	0.48	0.29	0.31	0.29	0.29	0.25	0.31	0.16	0.16	0.16
0.80	median	0.39	0.47	0.57	0.74	0.79	0.81	0.60	0.63	0.76	0.78	0.78	0.78
	RMSE	0.49	0.45	0.47	0.29	0.31	0.30	0.26	0.24	0.31	0.16	0.17	0.17
0.70	median	0.34	0.40	0.41	0.67	0.69	0.71	0.52	0.55	0.62	0.68	0.68	0.68
	RMSE	0.45	0.42	0.45	0.29	0.31	0.32	0.24	0.23	0.30	0.17	0.17	0.18
0.60	median	0.27	0.31	0.31	0.58	0.59	0.59	0.45	0.47	0.53	0.60	0.59	0.59
	RMSE	0.42	0.40	0.44	0.29	0.31	0.33	0.22	0.22	0.27	0.18	0.18	0.18
0.50	median	0.21	0.23	0.24	0.50	0.51	0.50	0.36	0.37	0.44	0.50	0.49	0.49
	RMSE	0.38	0.38	0.42	0.30	0.31	0.34	0.21	0.21	0.24	0.19	0.17	0.18
0.40	median	0.13	0.15	0.16	0.40	0.40	0.39	0.28	0.28	0.36	0.40	0.40	0.40
	RMSE	0.36	0.36	0.39	0.31	0.30	0.34	0.20	0.20	0.24	0.21	0.18	0.19
0.20	median	0.00	0.00	0.00	0.22	0.19	0.19	0.09	0.10	0.15	0.19	0.19	0.19
	RMSE	0.30	0.31	0.33	0.32	0.29	0.35	0.19	0.19	0.23	0.23	0.18	0.20
0.00	median	-0.15	-0.16	-0.13	0.02	-0.01	-0.01	-0.07	-0.07	-0.04	0.01	0.01	0.00
	RMSE	0.26	0.27	0.28	0.32	0.29	0.35	0.17	0.17	0.20	0.25	0.18	0.22
-0.40	median	-0.46	-0.48	-0.47	-0.38	-0.40	-0.38	-0.44	-0.44	-0.44	-0.40	-0.40	-0.41
	RMSE	0.19	0.20	0.21	0.27	0.26	0.35	0.14	0.14	0.15	0.16	0.17	0.25
-0.80	median	-0.77	-0.81	-0.77	-0.78	-0.74	-0.61	-0.79	-0.80	-0.80	-0.80	-0.81	-0.87
	RMSE	0.15	0.14	0.16	0.20	0.22	0.41	0.10	0.10	0.11	0.11	0.13	0.22

Notes: All experiments are based on 2000 replications.

Table 5.3d

Medians and RMSEs of MU Estimators Based on the POI Test Statistics in the Linear Regression with AR(1) Disturbances for Design Matrix $X4$

ρ		$T=20$						$T=40$					
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$
1.00	median	0.54	0.64	1.00	0.90	1.00	1.00	0.73	0.81	1.00	0.95	1.00	0.99
	RMSE	0.55	0.49	0.48	0.30	0.28	0.29	0.32	0.25	0.32	0.16	0.15	0.15
0.90	median	0.48	0.58	0.77	0.84	0.92	0.90	0.68	0.76	0.88	0.88	0.91	0.90
	RMSE	0.51	0.46	0.47	0.28	0.27	0.28	0.28	0.22	0.33	0.15	0.15	0.15
0.80	median	0.41	0.48	0.52	0.75	0.81	0.78	0.62	0.68	0.72	0.80	0.81	0.80
	RMSE	0.48	0.44	0.46	0.29	0.29	0.30	0.24	0.20	0.32	0.16	0.15	0.16
0.70	median	0.35	0.41	0.41	0.67	0.71	0.69	0.53	0.59	0.63	0.70	0.71	0.70
	RMSE	0.44	0.41	0.44	0.29	0.29	0.30	0.24	0.21	0.30	0.18	0.17	0.17
0.60	median	0.27	0.31	0.32	0.57	0.60	0.58	0.45	0.49	0.52	0.60	0.61	0.60
	RMSE	0.41	0.40	0.43	0.30	0.30	0.31	0.22	0.20	0.26	0.19	0.17	0.17
0.50	median	0.22	0.24	0.25	0.49	0.51	0.48	0.36	0.39	0.44	0.50	0.50	0.49
	RMSE	0.37	0.37	0.40	0.30	0.30	0.31	0.21	0.19	0.23	0.19	0.17	0.17
0.40	median	0.13	0.15	0.16	0.39	0.40	0.38	0.28	0.30	0.37	0.40	0.41	0.39
	RMSE	0.36	0.36	0.38	0.32	0.31	0.32	0.20	0.20	0.22	0.20	0.18	0.18
0.20	median	0.01	0.01	0.02	0.21	0.22	0.20	0.11	0.11	0.20	0.20	0.21	0.19
	RMSE	0.30	0.31	0.33	0.33	0.30	0.32	0.18	0.19	0.22	0.24	0.18	0.19
0.00	median	-0.15	-0.16	-0.12	0.01	0.01	0.00	-0.07	-0.07	-0.04	-0.01	0.01	0.00
	RMSE	0.26	0.27	0.29	0.32	0.29	0.32	0.17	0.17	0.20	0.24	0.18	0.19
-0.40	median	-0.46	-0.48	-0.44	-0.39	-0.38	-0.39	-0.43	-0.44	-0.40	-0.40	-0.39	-0.41
	RMSE	0.19	0.20	0.21	0.27	0.27	0.33	0.14	0.14	0.15	0.16	0.17	0.21
-0.80	median	-0.78	-0.81	-0.79	-0.80	-0.79	-0.78	-0.78	-0.80	-0.76	-0.80	-0.79	-0.79
	RMSE	0.15	0.14	0.16	0.18	0.21	0.28	0.11	0.10	0.12	0.12	0.15	0.22

Notes: All experiments are based on 2000 replications.

Table 5.3e

Medians and RMSEs of MU Estimators Based on the POI Test Statistics in the Linear Regression with AR(1) Disturbances for Design Matrix $X5$

ρ		$T=20$						$T=40$					
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$
1.00	median	0.41	0.83	0.99	0.93	1.00	1.00	0.79	0.91	1.00	0.96	1.00	1.00
	RMSE	0.62	0.34	0.44	0.24	0.23	0.22	0.26	0.15	0.31	0.12	0.11	0.10
0.90	median	0.40	0.75	0.88	0.86	0.92	1.00	0.73	0.83	0.84	0.88	0.91	0.94
	RMSE	0.54	0.34	0.42	0.24	0.24	0.26	0.22	0.15	0.33	0.12	0.12	0.12
0.80	median	0.37	0.65	0.73	0.76	0.81	0.86	0.66	0.74	0.69	0.79	0.81	0.83
	RMSE	0.48	0.35	0.41	0.25	0.26	0.30	0.21	0.15	0.32	0.13	0.14	0.15
0.70	median	0.34	0.56	0.60	0.68	0.71	0.75	0.58	0.65	0.56	0.70	0.71	0.72
	RMSE	0.42	0.35	0.41	0.27	0.27	0.34	0.20	0.16	0.29	0.15	0.15	0.17
0.60	median	0.29	0.45	0.44	0.57	0.60	0.61	0.49	0.55	0.55	0.60	0.61	0.60
	RMSE	0.38	0.35	0.41	0.29	0.29	0.38	0.19	0.17	0.25	0.15	0.16	0.18
0.50	median	0.25	0.36	0.32	0.49	0.51	0.50	0.40	0.44	0.44	0.49	0.51	0.48
	RMSE	0.34	0.34	0.40	0.29	0.29	0.41	0.19	0.18	0.21	0.17	0.16	0.20
0.40	median	0.18	0.26	0.23	0.40	0.41	0.37	0.30	0.34	0.37	0.40	0.40	0.37
	RMSE	0.31	0.34	0.39	0.32	0.31	0.46	0.19	0.19	0.22	0.18	0.17	0.22
0.20	median	0.05	0.07	0.10	0.21	0.22	0.12	0.13	0.14	0.19	0.20	0.21	0.13
	RMSE	0.26	0.32	0.33	0.33	0.31	0.50	0.18	0.19	0.22	0.22	0.18	0.26
0.00	median	-0.10	-0.13	-0.10	0.01	0.01	-0.19	-0.05	-0.05	-0.03	-0.01	0.01	-0.13
	RMSE	0.24	0.30	0.29	0.34	0.32	0.52	0.16	0.17	0.19	0.23	0.18	0.30
-0.40	median	-0.39	-0.48	-0.39	-0.39	-0.39	-0.78	-0.41	-0.44	-0.40	-0.40	-0.39	-0.66
	RMSE	0.20	0.23	0.22	0.31	0.33	0.44	0.14	0.15	0.15	0.16	0.17	0.35
-0.80	median	-0.67	-0.82	-0.67	-0.80	-0.83	-0.87	-0.76	-0.80	-0.76	-0.79	-0.79	-0.88
	RMSE	0.21	0.14	0.22	0.23	0.28	0.28	0.12	0.10	0.14	0.12	0.16	0.17

Notes: All experiments are based on 2000 replications.

Table 5.3f

Medians and RMSEs of MU Estimators Based on the POI Test Statistics in the Linear Regression with AR(1) Disturbances for Design Matrix $X6$

ρ		$T=20$						$T=40$					
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{A}$	$\hat{\rho}_{E}^{MU}$	$\hat{\rho}_{s(1,0)}^{MU}$	$\hat{\rho}_{s(1,0.5)}^{MU}$
1.00	median	0.34	0.67	0.76	0.92	1.00	0.99	0.61	0.90	0.88	0.96	1.00	0.99
	RMSE	0.72	0.55	0.56	0.29	0.29	0.29	0.44	0.19	0.39	0.13	0.13	0.12
0.90	median	0.35	0.62	1.00	0.84	0.92	0.89	0.61	0.82	1.00	0.87	0.91	0.90
	RMSE	0.60	0.47	0.48	0.30	0.28	0.30	0.34	0.19	0.33	0.13	0.14	0.13
0.80	median	0.33	0.53	0.73	0.75	0.80	0.77	0.57	0.72	0.77	0.79	0.82	0.80
	RMSE	0.53	0.43	0.45	0.32	0.29	0.33	0.28	0.19	0.33	0.14	0.15	0.15
0.70	median	0.31	0.46	0.64	0.68	0.71	0.69	0.50	0.61	0.61	0.69	0.71	0.70
	RMSE	0.46	0.40	0.44	0.35	0.30	0.35	0.26	0.21	0.33	0.17	0.17	0.17
0.60	median	0.26	0.37	0.45	0.58	0.60	0.58	0.43	0.51	0.52	0.59	0.61	0.59
	RMSE	0.42	0.39	0.44	0.39	0.32	0.38	0.24	0.20	0.30	0.18	0.17	0.18
0.50	median	0.21	0.30	0.32	0.49	0.51	0.48	0.35	0.40	0.41	0.49	0.51	0.49
	RMSE	0.37	0.37	0.43	0.44	0.32	0.40	0.22	0.20	0.25	0.20	0.17	0.20
0.40	median	0.15	0.20	0.19	0.40	0.40	0.38	0.27	0.31	0.33	0.40	0.40	0.39
	RMSE	0.34	0.36	0.42	0.51	0.34	0.44	0.21	0.21	0.23	0.22	0.18	0.21
0.20	median	0.03	0.03	0.04	0.21	0.22	0.18	0.10	0.11	0.15	0.20	0.20	0.19
	RMSE	0.28	0.33	0.36	0.62	0.34	0.47	0.19	0.20	0.22	0.37	0.19	0.25
0.00	median	-0.12	-0.15	-0.09	0.02	0.01	-0.03	-0.07	-0.08	-0.05	-0.01	0.01	-0.02
	RMSE	0.24	0.30	0.30	0.73	0.35	0.52	0.17	0.18	0.19	0.59	0.19	0.28
-0.40	median	-0.40	-0.49	-0.40	-1.00	-0.39	-0.44	-0.43	-0.45	-0.42	-1.00	-0.40	-0.42
	RMSE	0.19	0.22	0.22	0.73	0.37	0.56	0.14	0.14	0.15	0.60	0.20	0.38
-0.80	median	-0.68	-0.82	-0.69	-1.00	-0.81	-0.89	-0.78	-0.80	-0.79	-1.00	-0.79	-0.88
	RMSE	0.20	0.14	0.21	0.53	0.31	0.54	0.11	0.10	0.13	0.22	0.19	0.38

Notes: All experiments are based on 2000 replications.

Table 5.4a
Medians and RMSEs of MU Estimators Based on the Median Envelopes of the POI Tests in the Linear Regression with AR(1) Disturbances for Design Matrix $X7$

ρ		$T=20$			$T=40$			$T=60$		
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$
1	Median	0.15	0.78	0.93	0.30	0.91	0.96	0.36	0.94	0.98
	RMSE	0.84	0.74	0.22	0.70	0.23	0.11	0.65	0.10	0.08
0.95	Median	0.16	0.75	0.90	0.38	0.87	0.92	0.48	0.90	0.94
	RMSE	0.78	0.67	0.23	0.59	0.17	0.12	0.51	0.10	0.08
0.9	Median	0.21	0.73	0.85	0.43	0.83	0.87	0.55	0.85	0.88
	RMSE	0.70	0.54	0.23	0.50	0.16	0.12	0.40	0.10	0.09
0.85	Median	0.22	0.69	0.80	0.47	0.79	0.83	0.58	0.81	0.84
	RMSE	0.64	0.50	0.24	0.43	0.15	0.12	0.32	0.10	0.09
0.8	Median	0.23	0.66	0.78	0.48	0.74	0.79	0.58	0.75	0.79
	RMSE	0.58	0.45	0.23	0.37	0.15	0.13	0.27	0.11	0.10
0.6	Median	0.22	0.49	0.58	0.41	0.56	0.60	0.47	0.56	0.60
	RMSE	0.42	0.40	0.27	0.24	0.16	0.15	0.18	0.12	0.12
0.4	Median	0.16	0.33	0.41	0.29	0.37	0.40	0.32	0.37	0.40
	RMSE	0.30	0.35	0.28	0.19	0.18	0.17	0.15	0.13	0.13
0.2	Median	0.06	0.11	0.20	0.13	0.17	0.20	0.15	0.18	0.20
	RMSE	0.24	0.34	0.33	0.16	0.19	0.19	0.13	0.14	0.14
0	Median	-0.06	-0.10	-0.01	-002	-0.02	0.01	-0.01	-0.02	0.00
	RMSE	0.20	0.32	0.34	0.15	0.19	0.20	0.13	0.14	0.15

Notes: All experiments are based on 2000 replications.

Table 5.4b
Medians and RMSEs of MU Estimators Based on the Median Envelopes of the POI Tests in the Linear Regression with AR(1) Disturbances for Design Matrix $X8$

ρ		$T=20$			$T=40$			$T=60$		
		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$
1	Median	0.01	0.07	0.93	0.12	0.91	0.96	0.17	0.94	0.98
	RMSE	0.98	1.25	0.26	0.86	0.44	0.11	0.80	0.19	0.07
0.95	Median	0.02	0.59	0.90	0.17	0.87	0.92	0.25	0.90	0.94
	RMSE	0.91	1.09	0.24	0.77	0.28	0.12	0.69	0.09	0.07
0.9	Median	0.03	0.60	0.85	0.21	0.84	0.88	0.34	0.86	0.89
	RMSE	0.86	1.00	0.27	0.68	0.19	0.12	0.57	0.09	0.08
0.85	Median	0.04	0.61	0.80	0.26	0.79	0.83	0.39	0.82	0.84
	RMSE	0.80	0.89	0.26	0.60	0.19	0.13	0.48	0.09	0.08
0.8	Median	0.04	0.55	0.76	0.27	0.74	0.78	0.40	0.76	0.79
	RMSE	0.75	0.85	0.30	0.54	0.17	0.13	0.41	0.10	0.10
0.6	Median	0.06	0.45	0.59	0.28	0.57	0.60	0.39	0.58	0.60
	RMSE	0.54	0.65	0.34	0.34	0.20	0.16	0.25	0.13	0.13
0.4	Median	0.04	0.22	0.40	0.21	0.38	0.40	0.23	0.38	0.40
	RMSE	0.38	0.56	0.40	0.24	0.22	0.19	0.17	0.15	0.13
0.2	Median	0.01	0.03	0.20	0.09	0.16	0.20	0.13	0.18	0.20
	RMSE	0.23	0.46	0.46	0.17	0.24	0.20	0.14	0.16	0.16
0	Median	-0.05	-0.17	0.02	-0.03	-0.05	-0.00	-0.02	-0.02	-0.00
	RMSE	0.15	0.42	0.47	0.14	0.23	0.24	0.13	0.16	0.17

Notes: All experiments are based on 2000 replications.

Table 5.5 Medians and RMSEs of MU Estimators Based on the Median Envelopes of the POI Tests; $T = 60$

		$X1$			$X2$			$X4$			$X5$		
ρ		$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{MLE}$	$\hat{\rho}_{MU}$
1	Median	0.85	0.86	0.95	0.82	0.86	0.95	0.81	0.88	0.96	0.83	0.94	0.98
	RMSE	0.19	0.18	0.11	0.22	0.18	0.11	0.22	0.16	0.10	0.21	0.11	0.08
0.95	Median	0.83	0.84	0.92	0.80	0.84	0.93	0.80	0.85	0.93	0.80	0.90	0.94
	RMSE	0.16	0.15	0.10	0.19	0.16	0.10	0.19	0.15	0.10	0.19	0.11	0.08
0.90	Median	0.80	0.80	0.88	0.76	0.80	0.88	0.77	0.81	0.88	0.75	0.85	0.89
	RMSE	0.15	0.15	0.10	0.19	0.15	0.10	0.18	0.14	0.10	0.19	0.12	0.09
0.80	Median	0.71	0.72	0.79	0.67	0.71	0.79	0.69	0.72	0.80	0.68	0.75	0.79
	RMSE	0.14	0.14	0.11	0.18	0.16	0.11	0.16	0.14	0.11	0.17	0.11	0.09
0.60	Median	0.53	0.53	0.60	0.50	0.51	0.59	0.51	0.53	0.59	0.51	0.55	0.60
	RMSE	0.14	0.14	0.12	0.17	0.16	0.13	0.16	0.14	0.14	0.15	0.13	0.12
0.40	Median	0.34	0.34	0.40	0.31	0.32	0.40	0.31	0.32	0.39	0.33	0.35	0.40
	RMSE	0.14	0.14	0.13	0.15	0.15	0.13	0.50	0.15	0.14	0.14	0.14	0.14
0.20	Median	0.15	0.15	0.20	0.14	0.14	0.21	0.14	0.14	0.20	0.15	0.16	0.20
	RMSE	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.14	0.14	0.15

Table 5.6a
Median Functions of $\hat{\rho}_{OLS}$ and Median Envelope of the POI Tests Under Different Error Processes for Design Matrix $X1$; $T = 20$

	$N(0,1)$	$AR(0.8)$	$MA(0.2)$	$ARCH$	χ_2^2	t_3	log-normal	Cauchy
				Median Envelope of the POI Tests				
ρ								
0.9	0.943	0.972	0.961	0.939	0.940	0.941	0.943	0.942
0.8	0.895	0.964	0.936	0.887	0.897	0.896	0.895	0.894
0.7	0.857	0.942	0.93	0.847	0.857	0.848	0.848	0.847
0.6	0.809	0.937	0.902	0.797	0.811	0.821	0.803	0.813
0.4	0.721	0.909	0.861	0.724	0.736	0.715	0.713	0.708
0.2	0.621	0.842	0.830	0.619	0.621	0.633	0.619	0.628
0.0	0.536	0.752	0.743	0.535	0.532	0.545	0.522	0.518
				Median Function of $\hat{\rho}_{OLS}$				
ρ								
0.9	0.391	0.500	0.496	0.400	0.409	0.387	0.399	0.397
0.8	0.381	0.514	0.462	0.367	0.390	0.371	0.385	0.383
0.7	0.341	0.488	0.490	0.362	0.341	0.351	0.330	0.339
0.6	0.280	0.489	0.462	0.274	0.293	0.307	0.292	0.303
0.4	0.199	0.393	0.376	0.208	0.172	0.173	0.185	0.184
0.2	0.029	0.304	0.298	0.029	0.050	0.061	0.043	0.059
0.0	-0.101	0.182	0.183	-0.088	-0.093	-0.084	-0.116	-0.104

Notes: All medians are computed based on 20,000 simulations.

Table 5.6b
Median Functions of $\hat{\rho}_{OLS}$ and Median Envelope of the POI Tests Under Different Error Processes for Design Matrix $X7$; $T = 20$

	$N(0,1)$	$AR(0.8)$	$MA(0.2)$	$ARCH$	χ_2^2	t_3	log-normal	Cauchy
			Median envelope of the POI tests					
ρ								
0.9	0.944	0.974	0.958	0.936	0.940	0.940	0.943	0.937
0.8	0.894	0.952	0.930	0.885	0.888	0.899	0.892	0.888
0.7	0.848	0.963	0.901	0.834	0.842	0.852	0.845	0.844
0.6	0.810	0.932	0.877	0.793	0.813	0.810	0.806	0.803
0.4	0.716	0.893	0.861	0.717	0.714	0.726	0.726	0.722
0.2	0.648	0.857	0.805	0.648	0.641	0.654	0.636	0.644
0.0	0.564	0.797	0.740	0.540	0.573	0.568	0.560	0.567
			Median function of $\hat{\rho}_{OLS}$					
ρ								
0.9	0.205	0.270	0.274	0.225	0.225	0.226	0.249	0.256
0.8	0.264	0.326	0.325	0.247	0.266	0.262	0.254	0.279
0.7	0.261	0.342	0.353	0.237	0.249	0.261	0.260	0.273
0.6	0.235	0.368	0.358	0.219	0.254	0.258	0.247	0.261
0.4	0.147	0.330	0.318	0.177	0.155	0.173	0.161	0.182
0.2	0.074	0.270	0.250	0.077	0.067	0.059	0.042	0.077
0.0	-0.050	0.169	0.152	-0.064	-0.049	-0.040	-0.051	-0.031

Notes: All medians are computed based on 20,000 simulations.

Figure 5.11a
Median Functions of $\hat{\rho}_{OLS}$ and Median Envelopes of the POI tests Under Different Error Structures Using Design Matrix $X1$; $T = 20$

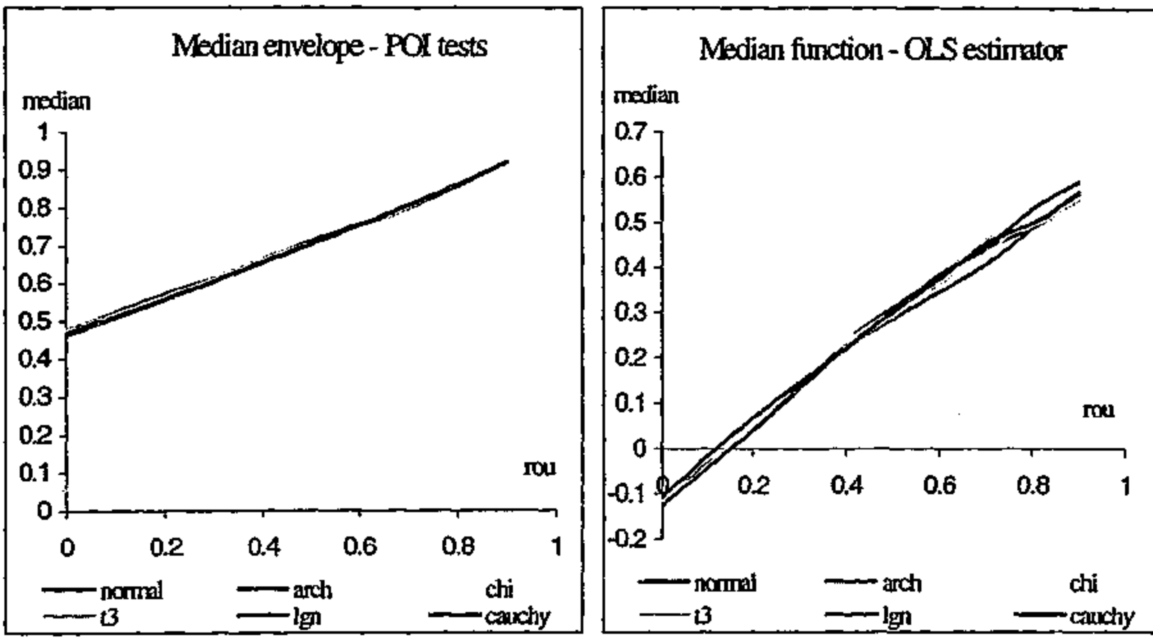
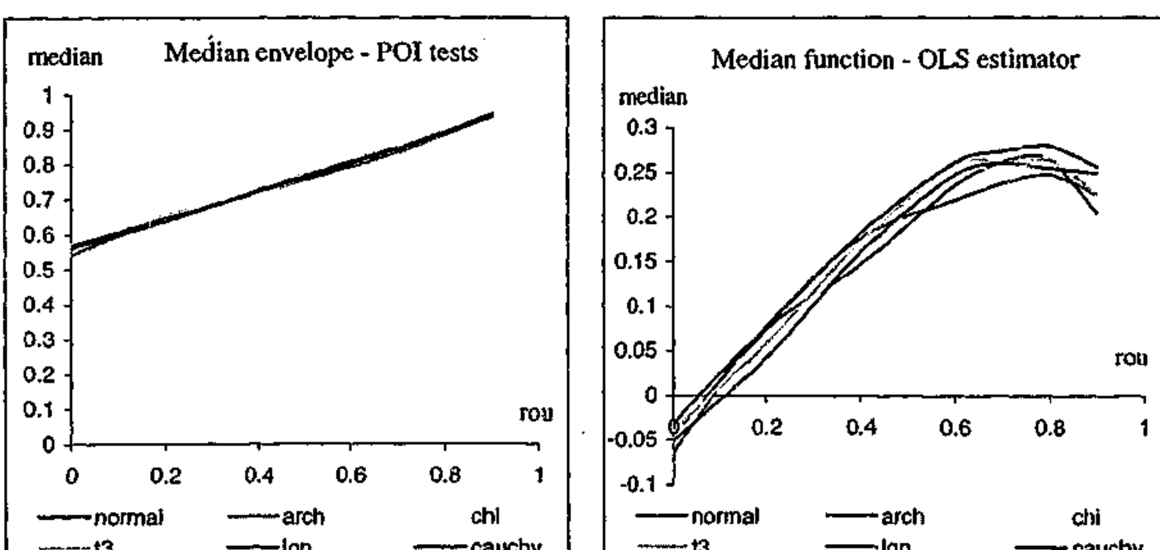


Figure 5.11b
Median Functions of $\hat{\rho}_{OLS}$ and Median Envelopes of the POI tests Under Different Error Structures Using Design Matrix $X7$; $T = 20$



Chapter 6

Hypothesis Testing and Forecasting Based on Median-unbiased Estimators [1]

6.1 Introduction

Point estimation, though important in its own right, often serves as the first stage of a statistical inference process. People rely more on procedures such as hypothesis testing and forecasting for decision making and policy recommendations in practice. Therefore, for any point estimation procedure developed, it is important to examine the usefulness of other inference procedures based on it. In Chapters 4 and 5, we developed some new (approximately) median-unbiased (MU) estimators for two time series models. We apply these estimators in hypothesis testing and forecasting procedures in this chapter. We show that the improved small sample performance of these estimators can also improve the small sample efficiency of the corresponding inference procedures.

We first examine the small sample power properties of the Wald test. It is well known that the Wald test, although efficient asymptotically, can suffer from size distortion, local biasedness and non-monotonic power in small samples. Much effort has been made to provide remedies for these problems. We show that by using the estimators proposed in Chapter 4 and 5, we can correct the small sample bias of the Wald test when testing for autocorrelated disturbances in the linear regression model, and the non-monotonic power problem when testing for random walk disturbances. In contrast, the likelihood ratio test seems to be less affected by the choice of estimators.

In the dynamic linear regression model, Nankervis and Savin (1985) suggested adjusting the moments of the t statistic so that it is better approximated by

[1] A paper based on the results of this chapter was presented to a departmental seminar at the Department of Econometrics and Business Statistics, Monash University, in April 2001.

the Student t distribution. We conjecture that this can be achieved by a simple correction in the bias of the estimator while controlling the variance. Goh and King (1999) proposed a bootstrap correction to the Wald test in order to correct its local biasedness. We show that the (approximately) MU estimator developed in Chapter 4 is equivalent to the implicit bias-corrected estimator in their test. The power curve is properly centred and tightened when the MU estimator is used in place of the OLS estimator. Compared with other methods, the computational cost of the proposed approach is lower.

The rest of the chapter is concerned with forecasting. We review the formulae for computing the root mean square prediction errors for the linear regression model with AR(1) disturbances and the dynamic linear regression model and discuss the relationship between the prediction risk and the estimation risk. Via Monte Carlo simulations, we compare the estimated one-step-ahead prediction risks based on different estimators. We find that the small sample bias in the conventional estimators of the autoregressive/lagged dependent variable coefficient usually lead to bigger prediction errors compared with those for the proposed MU estimators.

The chapter is organised as follows: Section 6.2.1 defines the Wald test and summarises the small sample deficiencies it may suffer from. The corrected Wald test based on MU estimators is discussed in Section 6.2.3 with a comparison made with other corrections in the literature. Their asymptotic validity is also addressed. Section 6.2.6 provides three sets of evidence on the effectiveness of the proposed correction. The small sample powers of the Wald test based on different estimators are compared in three different test situations for the linear regression model: 1. testing for autocorrelation; 2. testing for random walk disturbances; and 3. testing the lagged dependent variable coefficient. We move on to prediction in Section 6.3. The relationship between the bias in an estimator and the prediction error is discussed. The prediction error based on different estimators are compared in two models. The chapter ends with some concluding remarks in Section 6.4.

6.2 Wald-type Tests Based on MU Estimators

In this section, we construct Wald-type tests based on the MU estimators we proposed in Chapter 4 and 5. The modified Wald-tests are shown to be able to correct the small sample deficiencies that the conventional Wald tests suffer from in these models. We start this section by reviewing these deficiencies.

6.2.1 Small Sample Deficiencies of the Wald Test

The Wald test plays an important role in the theory of likelihood-based hypothesis testing. Let θ be a $k\times 1$ unknown parameter vector and y_t, $t=1, 2, \ldots, T$, be T observations generated independently from the implicit probability density function $f(y_t|x_t,\theta)$ in which x_t is the explanatory variable vector. The log-likelihood function is given by

$$l(\theta)=\sum_t \ln f(y_t|x_t,\theta)\ . \tag{6.1}$$

Suppose the parameter θ is partitioned into two sub-vectors $\theta=(\beta',\gamma')'$ where only β $(r\times 1)$ is the parameter vector of interest while γ $(k-r\times 1)$ is treated as a nuisance parameter vector. We are interested in testing the hypotheses

$$H_0:\beta=\beta_0 \text{ against } H_1:\beta\neq\beta_0 \tag{6.2}$$

where β_0 is a vector of known constants. The Wald test is then based on the unconstrained maximum likelihood (ML) estimator defined in the parameter space Θ and given by,

$$\hat{\theta}=\arg\max_{\theta\in\Theta} l(\theta) \tag{6.3}$$

where $\hat{\theta}=(\hat{\beta}',\hat{\gamma}')$. Let $\vartheta(\theta)=-E\left(\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right)$ be the Fisher information matrix evaluated at θ. The Wald test rejects H_0 for large values of

$$W=(\hat{\beta}-\beta_0)'\left[R\vartheta(\hat{\theta})^{-1}R'\right](\hat{\beta}-\beta_0), \tag{6.4}$$

where $R=(I_r:0)$ is an $r\times k$ matrix and I_r is an r-dimensional identity matrix. In practice, people more often use a different version of the Wald test which rejects H_0 for large values of

$$W=(\hat{\beta}-\beta_0)'A(\hat{\theta})^{-1}(\hat{\beta}-\beta_0), \tag{6.5}$$

where $A(\theta)=RV(\theta)R'$ and $V(\theta)$ is an estimator which converges stochastically to $\vartheta(\theta)^{-1}$ in an open neighbourhood of the true value of θ (e.g., see discussions in Stroud, 1971).

Under the standard regularity conditions (see for example Amemiya, 1985, Chapter 4), the asymptotic distribution of the Wald test statistic under the null hypothesis is χ^2_r, a central chi-square distribution with r degrees of freedom. While under the alternative hypothesis, the Wald test asymptotically follows a non-central chi-square distribution $\chi^2_r(\mu^2)$, with a non-centrality parameter μ. For details, see Godfrey (1988, Chapter 1) and Hendry (1995, Chapter 13).

Tests constructed using the Wald principle (i.e., tests that take the form of (6.5)), but not in the context of classical ML estimation, are often referred to as Wald-type tests. In such tests, the ML estimators are replaced by a broader class of asymptotically normal estimators, such as the generalised method of moment (GMM) estimators. For a review on Wald-type tests, see Burguete et al. (1982). These tests were shown to be also asymptotically optimal, similar to the classical Wald test.

The Wald test is a consistent test asymptotically. Within the class of asymptotically unbiased tests, the procedure is also asymptotically most powerful against local alternatives (see, e.g., Cox and Hinkley, 1974, Chapter 9). However these properties may not hold in small samples. Three small sample problems, namely, local biasedness, power non-monotonicity and non-invariance of the Wald test to the reparameterisation of the null hypothesis, have been identified and studied by many researchers. This chapter is mainly concerned with the first two problems, which are briefly reviewed below.

6.2.1.1 Local Biasedness

Peers (1971) studied the power function of the Wald test with a simple null hypothesis against a two-sided alternative using asymptotic expansions. He found that the Wald test can be locally biased for finite sample sizes, as the power can drop below its size in the neighbourhood of H_0 when the first-order derivative of the power function evaluated at the null value becomes negative. It was also found that the power function can be asymmetrical in the neighbourhood of H_0. This bias disappears as $n\to\infty$. Hayakawa (1975) and Hayakawa and Puri (1985) extended the analysis for tests of composite null hypotheses. Examples of this local biasedness have been reported by Magdalinos (1990), Oya (1997) and Goh and King (1999) among others for various models.

6.2.1.2 Non-monotonic Power

The power function of a test is said to be non-monotonic if the power first increases but eventually decreases (sometimes to zero) as the distance between the true parameter value and the null value increases. Hauck and Donner (1977) first reported the non-monotonic power behaviour of the Wald test for testing a single parameter in a binomial logit model. Similar phenomenon has been reported for different models by Mantel (1987), Nelson and Savin (1988, 1990) and Laskar and King (1997) among others. One possible reason for this anomaly is that, for alternatives sufficiently far away from the null hypothesis, $A(\hat{\theta})$ has a tendency to increase faster than $(\hat{\beta}-\beta_0)^2$ as the departure from the null gets larger; e.g, see discussions in Goh and King (2000).

6.2.1.3 Existing Remedies

Many researchers have attempted to improve the small-sample performance of the Wald test. These remedies can be classified into 3 categories:

1. Correcting the asymptotic critical value so as to control the size of the test. This can be done by either employing higher-order (usually second or third order) asymptotic expansions or by bootstrap. The two methods were shown to be equivalent by Hall (1992). Examples of the analytical approach include Rothenberg (1988), Magdalinos (1990) and Phillips and Park (1988), while the bootstrap critical values were discussed by Nankervis and Savin (1996) and Horowitz and Savin (1998) among many others. Nankervis and Savin (1985) and Cribari-Neto and Cordeiro (1996), on the other hand, attempted to correct the test statistic to make it more consistent with the asymptotic critical values.

2. Correcting the small sample bias in the estimator used in the test. Goh and King (1999) provided a good example of this approach. The proposed method in this chapter also falls into this category. Ferrari and Cribari-Neto (1993) adopted a similar approach in correcting the Wald test of nonlinear restrictions.

3. Using alternative estimators for the covariance matrix. For example, Mantel (1988), Laskar and King (1996) and Goh and King (2000) advocated the so-called null-Wald test to correct the non-monotonic power problem. Instead of using the variance estimator evaluated at the estimate, the null-Wald test replaces it by the variance estimated at the null value. For example, Laskar and King (1997) reported that the null-Wald test is able to remove the power non-monotonicity when testing for MA(1) errors in a linear regression model.

6.2.2 Bias-corrected Wald Tests

In the original Wald test, all parameters are estimated by unconstrained ML estimators. Goh and King (1999) conjectured that biasedness of the Wald test is a direct result of poor centring of its power curve, and one possible cause of this is the small sample biasedness of the ML estimates in the presence of nuisance parameters. They suggested an implicit correction factor for this possible bias in the estimation of β. Instead of using $\hat{\beta}$, they suggest using $\hat{\beta}_{CW}=\hat{\beta}-c_w$ when constructing the Wald test. The corrected Wald test is given by

$$CW=(\hat{\beta}_{CW}-\beta_0)'A(\hat{\theta}_{CW})^{-1}(\hat{\beta}_{CW}-\beta_0) \tag{6.6}$$

where $\hat{\theta}_{CW}=(\hat{\beta}_{CW}',\hat{\gamma}(\hat{\beta}_{CW})')'$. The correction factor c_w and the critical value d_{CW} are found by numerically solving the following two equations:

$$\begin{cases}\pi_{CW}(\beta)\big|_{H_0}=\alpha\\ (\partial\pi_{CW}(\beta)/\partial\beta)\big|_{H_0}=0\end{cases} \tag{6.7}$$

where $\pi_{CW}(\beta)=\Pr[CW>d_{CW}|\beta]$ is the power of the test at β. The first equation controls the size of the test while the second equation enforces the local unbiasedness. Because there is no analytical expression for the derivative of the power function, the second equation is approximated by

$$\pi_{CW}(\beta)\big|_{H_0^+}-\pi_{CW}(\beta)\big|_{H_0^-}=0 \tag{6.8}$$

where H_0^+ and H_0^- are local alternatives on the two sides of H_0. Goh and King (1999) designed a parametric bootstrap procedure to find c_w and d_{CW}.

We propose a different method of correcting local biasedness of the Wald test in this chapter. Instead of implicitly correcting $\hat{\beta}$ by bootstrap, we explicitly correct the small sample bias in $\hat{\beta}$ by replacing it with one of the MU estimators we developed in Chapters 4 and 5. Our version of the corrected Wald test is given by

$$W_{MU} = (\hat{\beta}_{MU} - \hat{\beta}_0)' A(\hat{\theta}_{MU})^{-1} (\hat{\beta}_{MU} - \hat{\beta}_0), \tag{6.9}$$

where $\hat{\theta}_{MU}$ is the approximately MU estimator constructed by iteratively correcting the bias of each parameter using the methods proposed in Chapter 3. By doing so, if the MU estimator can eliminate the small sample bias in $\hat{\beta}$, the corrected Wald test should correct the local biasedness of the original test.

The rationale underlying the suggested correction lies in the observation that, as it seems in many cases, the poor centring of the power curve of the Wald test appears to be linked to the wrong location (mean) of the test statistic (e.g. see Nankervis and Savin, 1985, 1988b, Goh and King, 1999). A shift in location towards the origin for the entire power curve should be able to alleviate the local biasedness and better centre the power curve. We argue that this can be done by shifting the estimator used in the test in the right direction (towards the true value), while keeping the magnitude of the variance of the estimator under control. From the simulation results reported in Chapters 4 and 5, we believe that the proposed MU estimators effectively corrected the downward bias of the LS estimators, while at the same time not substantially increasing its variance. Therefore we would expect by applying these MU estimators, we should be able to correct the local biasedness of the Wald tests.

6.2.3 Construction of Wald Tests Based on MU Estimators

6.2.3.1 Testing for Autocorrelated Errors

We examine the same model studied in Chapter 4, i.e. the linear regression model with stationary AR(1) disturbances. The testing problem considered in this section is that of testing the existence of autocorrelation in the disturbances. This is probably the most extensively studied hypothesis testing problem by econometricians. For a

comprehensive review, see King (1987a). The purpose of our study is to use this model to illustrate the effectiveness of correcting the local biasedness of a two-sided Wald test by applying MU estimators, rather than to propose a new test. Therefore it is not our interest to have exhaustive power comparisons of all available tests. Instead, we simply concentrate on the Wald tests based on three different estimators: 1. the two-step OLS estimator ($\hat{\rho}_{OLS}$); 2. the full maximum likelihood estimator ($\hat{\rho}_{ML}$); and 3. the MU estimator ($\hat{\rho}_{MU}$) proposed in Chapter 4, which is constructed by adjusting and then solving the marginal likelihood score equations.

The first two estimators, $\hat{\rho}_{OLS}$ and $\hat{\rho}_{ML}$, are asymptotically equivalent, with the same asymptotic variance:

$$AV(\hat{\rho}) = (1-\rho^2)/T \tag{6.10}$$

where $\hat{\rho}$ can be either $\hat{\rho}_{OLS}$ or $\hat{\rho}_{ML}$. But in order to improve its efficiency in small samples, we use the finite sample estimate of the estimator variance in the Wald test. This variance estimator is given by,

$$V(\hat{\rho}) = \{(T-3)^{-1}\{(1-\hat{\rho}^2)\hat{u}_1^2 + \sum_{t=2}^{T}(\hat{u}_t - \hat{\rho}\hat{u}_{t-1})^2\} \Big/ \sum_{t=2}^{T}\hat{u}_{t-1}^2, \tag{6.11}$$

where $\hat{u}_t$ are the GLS residuals for covariance matrix $\Sigma(\hat{\rho})$. Therefore the Wald test for testing

$$H_0: \rho = 0 \text{ against } H_a: \rho \neq 0 \tag{6.12}$$

is given by

$$W = \hat{\rho}^2 / V(\hat{\rho}). \tag{6.13}$$

The corrected Wald-test based on $\hat{\rho}_{MU}$, however, needs some justification. Recall that $\hat{\rho}_{MU}$ is the solution to the following estimating equation:

$$U(y;\rho)-med[U(y;\rho)|y\sim N(X\beta,\sigma^2\Omega(\rho))]=0, \tag{6.14}$$

where

$$U(y;\rho)=-\frac{1}{2}tr\left[\Delta(\rho)\frac{\partial\Omega(\rho)}{\partial\rho}\right]-\frac{T-k}{2}\left[\frac{\hat{u}'\frac{\partial\Omega(\rho)^{-1}}{\partial\rho}\hat{u}}{\hat{u}'\Omega(\rho)^{-1}\hat{u}}\right] \tag{6.15}$$

is the marginal likelihood score function, in which

$$\Delta(\rho)=\Omega^{-1}(\rho)-\Omega^{-1}(\rho)X(X'\Omega^{-1}(\rho)X)^{-1}X'\Omega^{-1}(\rho).$$

Ara (1995) showed that this score is mean-unbiased, i.e.

$$E_\rho[U(y;\rho)]=0. \tag{6.16}$$

Therefore we can derive the asymptotic distribution of the score function in the same way as in the classical MLE context, by expanding the score function at the true parameter value. If we denote the diagonal component in the information matrix corresponding to parameter ρ by

$$I(\rho)=E\{-\partial^2\ln L(y;\rho)/\partial\rho^2\}=Var[U(y;\rho)], \tag{6.17}$$

Ara (1995) showed that the marginal likelihood score $U(y;\rho)$ is asymptotically normally distributed, i.e.,

$$I(\rho)^{-1/2}U(y;\rho)\xrightarrow{d}N(0,I_m). \tag{6.18}$$

Therefore asymptotically $\hat{\rho}_{MU}$ is equivalent to the maximum marginal likelihood estimator because the median of the score vanishes as $T\to\infty$ in equation (6.15), and the variance of the adjusted score tends to the expected information matrix component. Hence the adjusted Wald test based on $\hat{\rho}_{MU}$ could use the same

information matrix component as the Wald-test based on the maximum marginal likelihood estimator. Under appropriate regularity conditions (e.g., see Ara, 1995), the corrected Wald-test will have the usual asymptotic chi-square distribution as the classical Wald-test, i.e.,

$$I(\hat{\rho}_{MU})(\hat{\rho}_{MU}-\rho_0)^2\xrightarrow{d}\chi_1^2 \tag{6.19}$$

where the diagonal component corresponding to ρ in the MGL-based information matrix, $I(\rho)$, is given by

$$I(\rho)$$
$$=\frac{1}{2(m+2)}\left\{m\times tr\left[\Delta(\rho)\frac{\partial\Omega(\rho)}{\partial\rho}\Delta(\rho)\frac{\partial\Omega(\rho)}{\partial\rho}\right]-\left[tr(\Delta(\rho)\frac{\partial\Omega(\rho)}{\partial\rho}\right]^2\right\}. \tag{6.20}$$

Hence for testing hypotheses (6.12), the corrected Wald-test based on the MU estimator is given by

$$W_{MU}=\hat{I}(\hat{\rho}_{MU})^{-1}\hat{\rho}_{MU}^2. \tag{6.21}$$

6.2.3.2 Testing for Random Walk Disturbances

In this section, we consider the one-sided test problem

$$H_0:\rho=1 \text{ against } H_a:\rho<1 \tag{6.22}$$

in the linear regression model with stationary AR(1) or random walk disturbances. The model specifications are given in Section 5.1. This testing problem has also been studied extensively by econometricians. If there are no exogenous regressors other than a time trend, this is the familiar testing for a unit root problem. The literature on such tests is vast, see Phillips and Xiao (1998) for a recent survey. The difficulty of this problem is that the distributions of the tests are non-standard even asymptotically. For the models with exogenous regressors, Berenblut and Webb (1973), Sargan and Bhargava (1983), King (1987b) and Dufour and King (1991) among others considered testing for random walk disturbances. It was found that among the tests available, the test statistic developed by Berenblut and Webb (1973), which is a special case of the POI tests proposed by Dufour and King (1991), generally has good small sample power properties (e.g, see Phillips and Xiao (1998)) against stationary alternatives. The test also has the advantage of being an exact test.

In Chapter 5, we reported that many tests, including the DW test, t test and the POI tests may suffer from non-monotonic power problem for some design matrices on the negative side of the alternative ρ values. A similar problem was reported in Kramer and Zeisel (1990) and Bartels (1992) for testing the null hypothesis of zero correlation. For the random walk null hypothesis, we propose the Wald-test based on the MU estimators developed in Chapter 5 as a remedy to this problem. We show that this test is able to correct the non-monotonic power problem when $\rho \to -1$. It can also correct the local biasedness suffered by other tests for some deign matrices.

The one-sided Wald test is given by

$$W_i = (\hat{\rho}_i - 1) / V(\hat{\rho}_i)^{1/2} \tag{6.23}$$

where $\hat{\rho}_i$ can be $\hat{\rho}_{GLS}$, $\hat{\rho}_{ML}$, $\hat{\rho}_A$ and $\hat{\rho}_{MU}$ from Chapter 5, and the estimated variance is given by

$$V(\hat{\rho}_i) = \hat{\sigma}^2 \Big/ \sum_{2}^{T} \hat{u}_{t-1}^2 \tag{6.24}$$

where

$$\hat{\sigma}^2 = (1-\rho_i^2)\left[\hat{u}'\Omega(\rho_i)^{-1}\hat{u}/(T-k)\right], \tag{6.25}$$

$$\hat{u} = [I - X(X'\Omega(\rho_i)^{-1}X)^{-1}X'\Omega(\rho_i)^{-1}]y. \tag{6.26}$$

6.2.3.3 Testing the Lagged Dependent Variable Coefficient

Consider the dynamic linear regression model (4.25) examined in Section 4.3. We are interested in testing the significance of the lagged dependent variable coefficient, i.e.

$$H_0: \gamma = 0 \text{ against } H_a: \gamma \neq 0. \tag{6.27}$$

The t statistic is routinely used to test these hypotheses. The exact distribution of the t statistic is complicated. In a model with no other exogenous regressors other than an intercept and/or a time trend, Nankervis and Savin (1985, 1987, 1988b) studied the exact distribution of the t statistic under the null hypothesis and clearly established that the Student t distribution is not a satisfactory approximation for sample sizes typical in economic applications. Monte Carlo evidence reported in their papers and also in Tanaka (1983) and Rayner (1990) confirmed the inadequacy of using the t statistic in this model. This may cause size distortions when the asymptotic critical values are used, and low power when the level-corrected critical values are used. An alternative approach is to use the bootstrap method to obtain asymptotically valid critical values for the test. Beran (1988) showed that the test with bootstrap-based critical values can provide better control over the rejection probability than the test that uses the asymptotic critical value. Nankervis and Savin (1996) examined the level and power of the bootstrap t test in the model with an intercept and a time trend as regressors and concluded that the bootstrap test has essentially the same power as the empirically level-corrected asymptotic-theory test reported in Nankervis and Savin (1988b).

An important finding reported in Nankervis and Savin (1988b) is that for small sample sizes, the reason for the bad approximation of the t statistic under the null hypothesis by the Student's t distribution is not the shape of the distribution but its location, i.e., the mean of the distribution of the t statistic is located substantially to the left of zero when the autoregressive parameter is near unity. Nankervis and Savin (1988b) suggested adjusting the t statistic to have the correct mean and variance. Let τ be a random variable which is distributed as Student's t with $T-k$ degrees of freedom. The adjusted t statistic is

$$t_A = (t - E(t))/\omega, \tag{6.28}$$

where $\omega^2 = [\mathrm{var}(\tau)/\mathrm{var}(t)]$. It was shown in Nankervis and Savin (1988b) that Student's t distribution accurately approximates the null distribution of the adjusted t statistic. The problem of this approach is that for models with exogenous regressors, the moments of the t statistic depend on nuisance parameters – the regression coefficients, and thus make it hard to compute the adjusted statistic (6.28).

We consider the Wald-test based on the approximately MU estimator developed in Chapter 4. Instead of correcting the moments of the t statistic, we correct the location (bias) of the estimator. Roughly speaking, the bias-correction of the estimator should result in adjusting the location of the t statistic, provided that the variance of the bias-corrected estimator is similar to that of the original estimator. Note that our correction factor also depends on the regression coefficients, and therefore is consistent with the approach discussed in Nankervis and Savin (1988b, p142).

Grose (1998) and Mahmood (2000) observed that, unlike for the linear regression model with AR(1) disturbances, the marginal likelihood score for the lagged dependent variable is not mean-unbiased, i.e., its expectation at the true parameter value is not zero for all sample sizes. The score is also not information unbiased, i.e., its variance is not identical to the corresponding information matrix component. Therefore it is not as straightforward to derive the asymptotic distribution of the Wald-type test for this model as in the previous section.

However, for large sample sizes, under certain conditions on the order of magnitude of X and Ω, Grose (1998) showed that the marginal likelihood score $U(y;\gamma)$ is asymptotically unbiased and information unbiased, i.e., as $T \to \infty$,

$$\lim T^{-1/2} U(y;\gamma) = 0 \tag{6.29}$$

$$\lim T^{-1} Var[U(y;\gamma)] = \lim T^{-1} \vartheta_{\gamma\gamma}, \tag{6.30}$$

where $\vartheta_{\gamma\gamma}$ is the diagonal component corresponding to γ in the marginal likelihood-based information matrix. For details, see Grose (1998). Therefore asymptotically the marginal score is still normal and the same arguments used in Section 6.2.2 can be used here to show that the Wald-test is asymptotically valid, with the usual chi-square distribution as the approximately MU estimator should also be asymptotically equivalent to the maximum marginal likelihood estimator (which is also asymptotically equivalent to the global maximum likelihood estimator).

But according to Grose (1998), the information matrix component $\vartheta_{\gamma\gamma}$ contains second moment terms that cannot be resolved analytically. If we apply the Wald-test based on $\vartheta_{\gamma\gamma}$, it has to be evaluated either via a Laplace approximation (as in Grose, 1998 and Mahmood, 2000) or by numerical integration, which are both computationally costly. To overcome this difficulty, we use the corresponding component in the Hessian, namely $h_{\gamma\gamma}$, which is a consistent estimator of $\vartheta_{\gamma\gamma}$. Grose also derived the asymptotic equivalence of the MGL-based likelihood and the profile likelihood. Therefore the Hessian of the profile likelihood can be used to further simplify the construction of the Wald-type test.

Following Goh and King (1999), let $\theta = (\gamma, \beta', \sigma^2)'$. If we have an estimator of γ, $\hat{\gamma}_i$, which can either be the OLS estimator $\hat{\gamma}_{OLS}$ or the approximately MU estimator $\hat{\gamma}_{MU}$, then the corresponding estimators of the other two parameters are given by

$$\hat{\beta}(\hat{\gamma}_i) = \left\{ \left(\frac{1+\hat{\gamma}_i}{1-\hat{\gamma}_i} \right) x_1 x_1' + \sum_{t=2}^{T} x_t x_t' \right\}^{-1} \times \left\{ (1+\hat{\gamma}_i) x_1 y_1 + \sum_{t=2}^{T} x_t (y_t - \hat{\gamma}_i y_{t-1}) \right\}, \tag{6.31}$$

and

$$\hat{\sigma}^2(\hat{\gamma}_i) = \frac{1}{T-k-1}\{(1-\hat{\gamma}_i^2)\left(y_1 - \frac{x_1'\hat{\beta}(\hat{\gamma}_i)}{1-\hat{\gamma}_i}\right)^2 + \sum_{t=2}^{T}[y_t - \hat{\gamma}_i y_{t-1} - x_t'\hat{\beta}(\hat{\gamma}_i)]^2\}. \tag{6.32}$$

In this case, the Wald statistic is

$$W_i = \hat{\gamma}_i^2[RV(\hat{\theta}_i)R']^{-1}, \tag{6.33}$$

where $R = \begin{bmatrix}1 & 0_{k-1}'\end{bmatrix}$ and $\hat{\theta}_i = (\hat{\gamma}_i, \hat{\beta}(\hat{\gamma}_i)', \hat{\sigma}^2(\hat{\gamma}_i))'$, while the variance matrix is given by

$$V(\theta) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12}' & a_{22} & 0_{k-2} \\ a_{13} & 0_{k-2}' & a_{33} \end{bmatrix}^{-1}, \tag{6.34}$$

where

$$a_{11} = \frac{1}{\sigma^2}\left\{\frac{(1+\gamma)(x_1'\beta)^2}{(1-\gamma)^3} + \sum_{t=2}^{T} m_{t-1}^2\right\} + \frac{T-2}{1-\gamma^2}, \tag{6.35}$$

$$a_{12} = \frac{1}{\sigma^2}\left\{\frac{(1+\gamma)}{(1-\gamma)^2}\beta' x_1 x_1' + \sum_{t=2}^{T} m_{t-1} x_t'\right\},$$

$$a_{13} = \frac{\gamma}{\sigma^2(1-\gamma^2)},$$

$$a_{22} = \frac{1}{\sigma^2}\left\{\frac{(1+\gamma)}{(1-\gamma)} x_1 x_1' + \sum_{t=2}^{T} x_t x_t'\right\},$$

$$a_{33} = \frac{T}{2\sigma^4},$$

$$m_1 = \frac{x_1'\beta}{1-\gamma} \text{ and } m_t = \gamma m_{t-1} + x_t'\beta,\ t = 2,\ldots,n. \tag{6.36}$$

6.2.4 Monte Carlo Tests

Throughout our simulation studies that follow, we adopt the technique of the Monte Carlo test originally suggested by Dwass (1957) and Barnard (1963). The main reason is that the test statistics obtained in this chapter have fairly complex null distributions which are difficult to compute analytically. Applying a Monte Carlo test involves simulating the critical value under the null hypothesis and using this critical value to assess the test's power properties. As pointed out by Kiviet and Dufour (1997) and Dufour and Kiviet (1998), although the Monte Carlo tests are related to tests based on a parametric bootstrap, Monte Carlo tests have the important advantage of being valid in finite samples even when the number of replications used is small; see also discussions in Jockel (1986), Horowitz (1994) and Horowitz and Savin (2000). In the dynamic linear regression model, Nankervis and Savin (1996) reported the equivalence between the Monte Carlo tests and the bootstrap-based tests in terms of small sample power. Monte Carlo tests are usually computationally more efficient than the bootstrap tests.

However, Horowitz and Savin (2000) criticised the Monte Carlo test methodology on the grounds that the level-corrected critical values used in these tests are irrelevant to the empirical testing problems, as they are artificial and only valid for the particular set of simulation experiments. We join Kiviet and Dufour (1997) and Dufour and Kiviet (1998) and argue that, nevertheless, the Monte Carlo test is a powerful tool for examining the small sample properties of tests which have null distributions not possibly tractable analytically. As pointed out by Goh (1998), Monte Carlo techniques provide an important device for econometricians to evaluate and choose sound inference procedures awaiting to be used in practice. Therefore Monte Carlo tests should be deemed as an important device for us to assess the candidates of tests before they are applied to non-experimental settings.

6.2.5 Experimental Design

For the linear regression model with AR(1) disturbances, we used the same eight design matrices in Chapter 5, as they cover a range of both artificial and empirical time series. The sample sizes were 20, 40 and 60. The tests compared were the Wald tests based on the two-step OLS estimator ($\hat{\rho}_{OLS}$), the ML estimator ($\hat{\rho}_{MLE}$) Andrews' estimator ($\hat{\rho}_A$) and the proposed MU estimator ($\hat{\rho}_{MU}$). For the first two test statistics, the variance of the estimator was estimated by the profile likelihood variance estimate given by (6.11).

When testing for autocorrelation, these tests were also compared with the DW test and the $s(1,0.5)$ test (King, 1985b), which were used as the power benchmark. The POI test, $s(1,0.5)$ (Dufour and King, 1991), was used as the benchmark when testing for random walk disturbances.

For the first-order dynamic linear regression model, we also used the same design matrices as in Section 4.4. The Wald-test based on the proposed estimator was compared with the one based on the OLS estimator. The variance of the estimator in both tests is estimated by the estimator given by (6.36).

2000 replications were conducted for the linear regression model with AR(1) disturbances and 1000 replications were used for the dynamic linear regression model. The quantiles of the simulated test statistics under the null hypothesis were used as the critical values. The rejection probabilities were reported for the parameter values under the alternative hypothesis. A significance level of 10% was used for testing autocorrelated disturbances and the LDV coefficient, while 5% was used for testing for random walk disturbances.

6.2.6 Results

6.2.6.1 Testing for Autocorrelated Errors

The empirical power (rejection probabilities) of the Wald tests based on different estimators are reported in Tables 6.1a – 6.1c for $X1$ - $X6$, and Tables 6.2a – 6.2b for $X7$ and $X8$. The corresponding power curves are presented in Figure 6.1.

Using simulated critical values, local biasedness is exhibited in the Wald-tests based on the OLS estimator and the MLE estimator on the right side of H_0. This is particularly serious for a sample size of 20. The powers of the W_{OLS} and W_{MLE} tests drop below 0.05 at $\rho = 0.1$ and 0.2 for all design matrices except $X7$ and $X8$. For example, the powers of the W_{OLS} and W_{MLE} tests at $\rho = 0.2$ for $X2$ and $T = 20$ are 0.02 and 0.03, respectively. As a result, the power curves of these two tests are poorly centred. Although still apparent, the local biasedness becomes less serious for $T = 40$. When ρ moves further away from H_0 to non-local alternatives, the W_{OLS} test also suffers from non-monotonic power. Its power drops in both tails for design matrices $X6$, $X7$ and $X8$ with 20 and 40 observations. For example, for $X8$ and $T = 20$, the powers of the W_{OLS} test at $\rho = 1$ and $\rho = -0.95$ are 0.06 and 0.1, respectively. Both problems plaguing these tests disappear for a sample size of 60. As a benchmark for power comparisons, we notice that the point optimal test $s(0,0.5)$ enjoys significant power superiority on the right side of H_0 over the W_{OLS} and W_{MLE} tests. It does not suffer from either local biasedness or non-monotonic power. However, its power on the left side of H_0 tends to be lower than those of the Wald tests, as $s(0,0.5)$ was designed to maximise the power in the neighbourhood of $\rho = 0.5$.

We first examine the performance of the Wald-test based on the proposed MU estimator $\hat{\rho}_{MU}$ for design matrices $X1$ - $X6$, where local biasedness is the main concern. As exhibited in Figure 6.1, the new test W_{MU} successfully corrected the local bias in the W_{OLS} and W_{MLE} tests for all these design matrices with a sample size of 20 and 40. The power curves of the W_{MU} test are properly centred at the null point and also tightened on the positive side of the parameter space. The power gain on the right side of H_0 is quite significant. For example, for $X4$ and $T=20$, the power of W_{MU} at $\rho=0.2$, 0.6, 0.8 are 0.09, 0.42 and 0.65, respectively, while the W_{OLS} test has powers of 0.03, 0.18 and 0.38 in this case. The power gain is consistent over all design matrices and all positive ρ values. On the negative side of H_0, however, the powers of the W_{MU} test are slightly lower than those of the W_{OLS} and W_{MLE} tests. This is probably because that the high powers of the W_{OLS} and W_{MLE} tests for negative ρ values are pushed up by the ill-centred power curves.

Compared with the W_{MU} test, the Wald-test based on Andrews' estimator, W_A, is generally more powerful than the W_{OLS} and W_{MLE} tests for positive ρ values, but less powerful than the W_{MU} test. Although it is also able to correct the local biasedness, its power for non-local alternatives are significantly lower than those of W_{MU} except for $X1$. Its power on the left side of H_0, however, is consistently slightly higher than that of W_{MU}. It is often argued that when testing for autocorrelation, good power is mostly needed for positive side of H_0. The W_{MU} test satisfies this criteria best among the tests considered.

Another interesting finding is that for all design matrices except $X7$ and $X8$, the W_{MU} test has powers very similar to those of the $s(0,0.5)$ test, especially on the right side of H_0. As the $s(0,0.5)$ test is tangent and close to the power envelope of this testing problem, this similarity verifies the effectiveness of correcting the Wald test by using the proposed MU estimator.

For design matrix $X7$ and $X8$, the W_{MU} test has a monotonic power curve on both sides of H_0. The W_{OLS} test performs poorly as expected. The power gain by the W_{MU} test is most significant for $T=20$. For example, the powers of the W_{MU} test at $\rho=0.8$ and 1 for $X8$ and $T=20$ are 0.5 and 0.78, respectively, compared with those of W_{OLS}, of 0.11 and 0.6. Interestingly, unlike for other design matrices, the W_{MLE} test has much better power properties than the W_{OLS} test for this design. It has slightly lower powers than the W_{MU} test on the right side of H_0 and better powers on the left side of H_0. The same pattern is depicted for $T=40$.

The power results for $T=60$ are presented in Table 6.3. The problems of local biasedness and non-monotonic power plaguing the tests for smaller sample sizes disappear. However, the power gain from using the W_{MU} test is still quite apparent for all design matrices. The W_{MU} test is significantly more powerful than the other two tests for local alternatives, and slightly more powerful for non-local alternatives. For example, the powers of the W_{OLS}, W_{MLE} and W_{MU} tests at $\rho=0.2$ for $X2$ are 0.25, 0.25 and 0.43, respectively.

From the above analysis, it is clear that the introduction of $\hat{\rho}_{MU}$ as a correction to the Wald-test effectively eliminates the local biasedness and non-monotonic power problem at the same time. The new test has powers greater than their empirical sizes at local alternatives on both sides of H_0 and the power does not drop when ρ departs further from H_0. The power curves are better centred than those of the W_{OLS} and W_{MLE} tests. The only drawback of the new procedure is its slightly lower power on the negative side of H_0, which is usually regarded as less critical than positive autocorrelation.

6.2.6.2 Testing for Random Walk Disturbances

The power comparison of the W_{OLS}, W_{MLE}, W_{MU} and $s(1,0.5)$ tests for testing random walk disturbances against one-sided alternatives are presented in Tables 6.4a – 6.4d and the power curves are plotted in Figures 6.2a – 6.2d for different design matrices and $T=20$, 40. The W_{OLS} and W_{MLE} tests still suffer from local biasedness for all design matrices except $X1$ with 20 observations. The W_{OLS} test is also not immune to this problem for $T=40$ for several design matrices such as $X6$ and $X8$. The point optimal test $s(1,0.5)$ is shown to be more powerful than the W_{OLS} and W_{MLE} tests.

The corrected Wald test W_{MU} effectively eliminates the local biasedness for all design matrices. Its power is above its empirical size for local alternatives. For a sample size of 20, the W_{MU} test is significantly more powerful than the other two tests for all positive ρ values. In particular, for $X6$ and $X8$, the W_{MU} test is more than twice as powerful as the other two tests for all positive ρ values. For a sample size of 40, the performance of the W_{MLE} test improves substantially while the W_{OLS} test is still not performing well. The W_{MU} test is now slightly more powerful than the W_{MLE} test, while still superior to the W_{OLS} test by a big margin, especially for $X2$, $X6$ and $X8$. There is also a similarity between the power curves of the W_{MU} test and the $s(1,0.5)$ test, which verifies the effectiveness of improving the small sample power properties of the Wald test by using MU estimator.

6.2.6.3 Testing the LDV Coefficient

The estimated powers of the Wald-tests based on $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ in the dynamic linear regression model for different design matrices and $T = 20$, 40 are reported in Table 6.5, and the power curves are presented in Figure 6.3. As expected, the Wald test based on the OLS estimator performs poorly in this model. It suffers from local biasedness on the positive side of H_0 for all design matrices and both sample sizes. The powers drop to almost 0 at $\gamma = 0.1$ for $X4$ and $X5$ and $T = 20$. The power curve is poorly centred and seriously asymmetrical. The Wald test based on the proposed MU estimator, however, seems to be immune to this bias problem. Its power curve is centred around H_0 and the powers are above the size on both sides of H_0. This is achieved by a significant increase in power for positive γ values and slightly lower power for negative γ values. The power gain on the positive side of H_0 is particularly large for $T = 20$. For example, for $X5$, the powers of the W_{OLS} test at $\gamma = 0.1$ and 0.2 are 0.04 and 0.05, respectively (both below the size of 10%), while the W_{MU} test has a power of 0.11 and 0.16 in this case. The results show that the modified Wald test has approximately equal powers on the two sides of H_0 provided the departure from H_0 is not too far. For example, the powers of the W_{OLS} test at $\gamma = 0.1$ are lower than those at $\gamma = -0.1$ by more than 0.12 for all design matrices and $T = 20$, while the W_{MU} test has a power difference less than 0.01 for all design matrices at these two points. These results lead us to believe that the bootstrap correction of the local biasedness proposed by Goh and King (1999) is indeed achieved as effectively by the proposed method.

6.3 Prediction Based on MU Estimators

It is commonly agreed that the precision of the forecasts directly depends upon the precision of the parameter estimates. As Phillips (1979) remarked, in autoregressive time series, the serious small sample bias of the parameter estimates will carry over to the conditional distribution of forecasts given the observed values of the endogenous variables used to initiate forecasts. In the linear regression context, Goldberger's (1962) seminal paper on forecasting suggested the importance of efficiently estimating the parameters in the unknown covariance matrix in order to increase forecasting accuracy. Assuming all the parameters are known, Goldberger derived the best linear predictor for a general linear regression model. But in practice, one has to replace the parameters in this optimal predictor by their estimates. The risk of an approximate predictor is therefore affected by the choice of estimator. The exact behaviour of the predictors based on different estimators is usually hard to trace especially in small samples.

The study of prediction accuracy in autoregressive models stems from the early work of Hurwicz (1950b) and Shenton and Johnson (1966) among others. Yamamoto (1976) compared the asymptotic efficiencies of the predictors in a linear regression model with AR(1) disturbances and concluded that the predictor based on the GLS estimator is not necessarily asymptotically more efficient than the one based on the OLS estimator. The asymptotic approach was also taken by Baillie (1979), Fuller and Hasza (1981), Stine (1987) and Kemp (1999) in examining the properties of predictors for autoregressive time series, mainly for the near non-stationary and unit root case in large samples. Spitzer and Baillie (1983), via Monte Carlo simulations, studied the validity of asymptotic approximations to the distributions of the prediction errors in the same model. They concluded that the asymptotic error formula, when used in small samples, do not fully reflect the finite sample risk caused by estimating the parameters.

Phillips (1979), Maekawa (1987), Hoque et al. (1988) and Magnus and Pesaran (1989, 1991) all attempted to, via Edgeworth-type expansions, approximate the finite sample distribution of the forecasting errors in a simple autoregressive model, in order to study the small sample bias and efficiency of the predictors. Monte Carlo studies of forecasting errors in the same model were conducted by Stine (1985), Dielmann (1985), Sampson (1991) and Kemp (1999) among others. It appears that the small sample bias in the LS estimators of the autoregressive parameter usually leads to larger mean squared error of prediction (MSEP) especially when the parameter is near unity.

To improve on the predictors based on the OLS estimators, King and Giles (1984) proposed a pre-test procedure. More recently, Gospodinov (1999) attempted to construct a MU predictor for an AR(p) model by inverting the median function of the probability distribution of the least square predictor. Before we introduce the proposed predictors based on the MU estimators, we start by examining the small sample prediction risk of the linear autoregressive model. Many researchers have suggested improving the prediction accuracy by bootstrap. Attention was mainly given to prediction intervals. These studies include Stine (1985), Masarotto (1990), Thombs and Schucany (1990), Basawa et al. (1991), Kabaila (1993b), Beran (1993), Grigoletto (1998) and Kim (2001).

6.3.1 Prediction Risk and Estimation Bias

For the linear regression model with AR(1) disturbances (the model specified by (4.1) and (4.2)), consider the one period ahead forecast $\hat{y}_{T+1}$, which is given by

$$\begin{aligned}\hat{y}_{T+1} &= x_{T+1}'\hat{\beta}+\hat{u}_{T+1} \\ &= x_{T+1}'\hat{\beta}+\hat{\rho}\hat{u}_T\end{aligned} \tag{6.37}$$

where

$$\hat{\beta}=(X'\Omega^{-1}(\hat{\rho})X)^{-1}X'\Omega^{-1}(\hat{\rho})y, \tag{6.38}$$

$$\hat{u}_t = y_t - x_t'\hat{\beta}, \tag{6.39}$$

and y_T and x_{T+1} are the last available observations of y and x, respectively. $\hat{\rho}$ can be replaced by different estimators such as $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$. If the autoregressive parameter is known, Goldberger (1962) showed that this is the best linear unbiased predictor. But when ρ and β are replaced by their estimates, the efficiency of the predictor (6.37) will depend on the quality of the estimators chosen. Similarly, the h-period ahead forecast $\hat{y}_{T+h}$ is given by,

$$\hat{y}_{T+h} = x_{T+h}'\hat{\beta}+\hat{\rho}^h\hat{u}_T. \tag{6.40}$$

Many researchers have studied the asymptotic efficiency of this predictor in a model with just a constant as the regressor. They concluded that asymptotically, the prediction error is most serious when ρ is in the neighbourhood of 1. Researchers also found that the estimator $\hat{\beta}$ is essentially unbiased for most design matrices and positive autocorrelation, and different estimators of ρ do not have a big impact on the efficiency of $\hat{\beta}$, see for example Rao and Griliches (1969), Magee et al. (1987) and Latif and King (1993). Therefore heuristically speaking, the finite sample MSEP of (6.37) is dominated by the mean squared error of $\hat{\rho}$, namely the terms involving $E(\hat{\rho}-\rho)^2$. This justifies the effort to improve the prediction accuracy by using a less biased $\hat{\rho}$ with small RMSE. Hence we would expect that the proposed predictor

$$\hat{y}_{T+h}^{MU} = x_{T+h}'\hat{\beta}+\hat{\rho}_{MU}^h\hat{u}_T, \tag{6.41}$$

where $\hat{\rho}_{MU}$ is the MU estimator we proposed in Chapter 5, should be more efficient in small samples compared with the ones based on $\hat{\rho}_{OLS}$ and $\hat{\rho}_{MLE}$.

For the dynamic linear regression model, the frequently used h-period-ahead predictor is given by

$$\hat{y}_{T+h} = \hat{\gamma}_{OLS}^h y_T + \hat{\gamma}_{OLS}^{h-1}x_{T+1}'\hat{\beta}+\ldots+\hat{\gamma}_{OLS}x_{T+h-1}'\beta + x_{T+h}'\beta. \tag{6.42}$$

For $h=1$, its finite sample MSEP can be expressed as

$$E[(\hat{y}_{T+1}-y_{T+1})^2|y_T]=E(\hat{\gamma}-\gamma)^2 y_T^2+E(x_{T+1}'\hat{\beta}-x_{T+1}'\beta)^2$$
$$+2E[(\hat{\gamma}-\gamma)(x_{T+1}'\hat{\beta}-x_{T+1}'\beta)y_T]. \tag{6.43}$$

It is also apparent that by reducing the bias and RMSE of the estimator $\hat{\gamma}$, we should be able to reduce the MSEP of the predictor. Therefore we replace $\hat{\gamma}_{OLS}$ in (6.42) by the approximately MU estimator $\hat{\gamma}_{MU}$ we developed in Chapter 4 and suggest using the one-period-ahead predictor

$$\hat{y}_{T+1}^{MU}=\hat{\gamma}_{MU}y_T+x_{T+1}'\hat{\beta}. \tag{6.44}$$

6.3.2 Experimental Design

In the linear regression model with AR(1) disturbances, we compute the one-period-ahead forecasts based on $\hat{\rho}_{OLS}$, $\hat{\rho}_{MLE}$, $\hat{\rho}_A$ and $\hat{\rho}_{MU}$, which are denoted by $\hat{y}_{T+1}^{OLS}$, y_{T+1}^{MLE}, y_{T+1}^{A} and y_{T+1}^{MU}, respectively, for eight different design matrices specified in Chapter 5, and for $\rho=1$, 0.9, …, 0.1, 0 and for $T=20$ and 40. The RMSEPs were calculated based on 2000 forecasts.

In the dynamic linear regression model, we compare the predictors based on $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$. The latter was based on the iterative algorithm proposed in Chapter 4. 1000 forecasts were computed for $\gamma=0.9$, 0.8, …, -0.6, -0.8, $T=20$, 40 and for four design matrices specified in Section 4.3.

6.3.3 Results

6.3.3.1 Linear Regression with AR(1) Disturbances

The one-period-ahead prediction errors of the 4 predictors are reported in Table 6.6a – 6.6c. For most design matrices, the prediction errors based on different estimators all increase with ρ. For design matrices such as $X2$ and $X5$, the prediction error of

$\hat{y}_{T+1}^{OLS}$ at $\rho=1$ is more than twice as much as the error at $\rho=0$. For $T=40$, the prediction errors of $\hat{y}_{T+1}^{OLS}$ and $\hat{y}_{T+1}^{MLE}$ become more evenly distributed across the parameter space. y_{T+1}^{MLE} generally has RMSEPs similar to or slightly higher than those of $\hat{y}_{T+1}^{OLS}$ except for $X5$ and $X8$ with 20 observations. The RMSEPs of $\hat{y}_{T+1}^{MU}$ are consistently smaller than those of $\hat{y}_{T+1}^{OLS}$ and $\hat{y}_{T+1}^{MLE}$ for large positive ρ values (≥ 0.5) and for $T=20$. Especially when ρ is close to 1, the advantage is significant. For example, the RMSEPs of $\hat{y}_{T+1}^{MU}$ at $\rho=0.9$ for $X2$, $X5$, $X6$ and $X8$ for $T=20$ are, 1.23, 2.07, 2.02 and 2.69, respectively, which are much smaller than those of $\hat{y}_{T+1}^{OLS}$: 1.83, 2.85, 4.04 and 2.92, respectively. This advantage is still present for $T=40$ but with a smaller margin. For example, the RMSEPs of $\hat{y}_{T+1}^{MU}$ for $X5$ at $\rho=1$, 0.9, and 0.8 are, 1.06, 1.06 and 1.07, respectively, while $\hat{y}_{T+1}^{OLS}$ has RMSEPs of 2.58, 1.67 and 1.3, respectively. The predictor based on Andrews' estimator, $\hat{y}_{T+1}^{A}$, performs very similarly to $\hat{y}_{T+1}^{MU}$ for $X1$. But for all other design matrices, it always has slightly larger RMSEPs than those of $\hat{y}_{T+1}^{MU}$. These results verify the expected link between the quality of an estimator and the performance of the predictor based on it. The smaller bias and risk of the MU estimators proposed in the previous chapters were translated into smaller RMSEPs in forecasting.

6.3.3.2 Dynamic Linear Regression Model

The estimated one-step-ahead forecast errors for the two predictors based on $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ for the dynamic linear regression model are reported in Table 6.7. The predictor based on the proposed MU estimator generally has smaller RMSEPs than those of the predictor based on the OLS estimator for all positive γ values and for all design matrices considered. The OLS predictor has similar RMSEPs for different γ values, while the MU predictor tends to have larger RMSEPs for γ values close to 1 than those for other γ values. The difference between the two predictors is more significant for $-0.5\leq\gamma\leq0.5$, where the bias in $\hat{\gamma}_{OLS}$ seems to be most serious, as reported in Chapter 4. For example, the difference in RMSEP between the two predictors at $\gamma=0.4$ and -0.4 for $X1$ and $T=20$ are 0.32 and 0.38, respectively, both of which favoured $\hat{y}_{T+1}^{MU}$. Interestingly, for $X1$, the OLS predictor sees little improvement when the sample size increases from 20 to 40, while the MU predictor shows a more apparent drop in the RMSEP when the sample size increases. The advantage of usi·: the new predictor is minimal for design matrix $X4$ and $T=20$. This reminds us that the RMSEP of a predictor is not solely determined by the bias of the estimator the predictor is based on. A similar phenomenon was reported by Fair (1996), who found that the MU estimators do not necessarily lead to better forecasting performance for a dynamic simultaneous equations model.

6.4 Concluding Remarks

This chapter provided some evidence of the effectiveness of improving small sample performance of the Wald test and forecasting accuracy by using MU estimators.

The removal of bias in the estimator leads to the correction of the local biasedness of the Wald test in both the linear regression model with AR(1) disturbances and the dynamic linear regression model. The power curves of the Wald test based on the MU estimator are properly centred at the null point and also tightened on the positive side of H_0. For the non-local alternatives, the modified Wald test is not affected by the problem of non-monotonic power, which usually plagues the tests based on LS estimators. Although the modified Wald-test is asymptotically equivalent to the ones based on the MLE estimators, it provides an effective remedy for the small sample deficiencies of the Wald test that are frequently encountered by researchers.

A similar conclusion can also be drawn for forecasting accuracy. The predictor based on the MU estimator usually has a smaller error compared with those based on biased estimators. The removal of bias in the estimator leads to improved prediction efficiency.

Table 6.1a

Rejection Probabilities of the W_{OLS}, W_{MLE}, W_A, W_{MU} and $s(0,0.5)$ Tests at 5% Significance Level in the Linear Regression with AR(1) Disturbances; Testing $H_0: \rho=0$ vs. $H_1: \rho\neq 0$, for Design Matrix $X1$ and $X2$

					$X1, T=20$						
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.70	0.57	0.33	0.12	0.03	0.05	0.18	0.45	0.78	0.96	0.99
W_{MLE}	0.71	0.58	0.34	0.13	0.04	0.05	0.18	0.45	0.78	0.96	0.99
W_A	0.82	0.71	0.50	0.24	0.10	0.05	0.10	0.25	0.61	0.90	0.99
W_{MU}	0.83	0.73	0.51	0.26	0.10	0.05	0.10	0.29	0.65	0.90	0.98
$s(0,0.5)$	0.83	0.73	0.51	0.26	0.10	0.05	0.13	0.32	0.61	0.83	0.93
					$X1, T=40$						
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	1.00	0.99	0.88	0.49	0.12	0.05	0.31	0.78	0.98	1.00	1.00
W_{MLE}	1.00	0.99	0.88	0.49	0.12	0.05	0.31	0.78	0.97	1.00	1.00
W_A	1.00	0.99	0.90	0.57	0.16	0.05	0.18	0.63	0.91	0.99	1.00
W_{MU}	1.00	0.99	0.91	0.59	0.18	0.05	0.18	0.62	0.94	1.00	1.00
$s(0,0.5)$	1.00	0.99	0.92	0.60	0.19	0.05	0.23	0.64	0.91	0.99	1.00
					$X2, T=20$						
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.41	0.30	0.14	0.04	0.02	0.05	0.19	0.45	0.77	0.95	0.99
W_{MLE}	0.54	0.39	0.19	0.06	0.03	0.05	0.17	0.42	0.76	0.95	0.99
W_A	0.56	0.46	0.30	0.15	0.06	0.05	0.16	0.40	0.74	0.94	0.99
W_{MU}	0.73	0.61	0.39	0.20	0.09	0.05	0.08	0.21	0.54	0.85	0.97
$s(0,0.5)$	0.67	0.53	0.31	0.13	0.06	0.05	0.13	0.27	0.46	0.58	0.61
					$X2, T=40$						
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.99	0.96	0.80	0.41	0.08	0.05	0.28	0.75	0.97	1.00	1.00
W_{MLE}	0.99	0.97	0.82	0.42	0.09	0.05	0.28	0.75	0.97	1.00	1.00
W_A	0.95	0.86	0.73	0.47	0.16	0.05	0.20	0.66	0.95	1.00	1.00
W_{MU}	1.00	0.98	0.86	0.50	0.13	0.05	0.13	0.53	0.91	0.99	1.00
$s(0,0.5)$	1.00	0.98	0.86	0.50	0.13	0.05	0.20	0.59	0.87	0.98	0.99

Note: 2000 Replications with Simulated Critical Values.

Table 6.1b

Rejection Probabilities of the W_{OLS}, W_{MLE}, W_A, W_{MU} and $s(0,0.5)$ Tests at 5% Significance Level in the Linear Regression with AR(1) Disturbances; Testing $H_0: \rho = 0$ vs. $H_1: \rho \neq 0$, for Design Matrix $X3$ and $X4$

$X3, T=20$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.50	0.35	0.16	0.05	0.02	0.05	0.18	0.43	0.76	0.95	0.99
W_{MLE}	0.59	0.43	0.21	0.07	0.03	0.05	0.17	0.42	0.76	0.95	0.99
W_A	0.63	0.50	0.32	0.15	0.07	0.05	0.15	0.38	0.72	0.93	0.99
W_{MU}	0.77	0.67	0.45	0.23	0.10	0.05	0.09	0.24	0.58	0.87	0.97
$s(0,0.5)$	0.75	0.63	0.4	0.19	0.08	0.05	0.12	0.25	0.45	0.57	0.65
$X3, T=40$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.98	0.94	0.76	0.36	0.06	0.05	0.27	0.76	0.97	1.00	1.00
W_{MLE}	0.98	0.95	0.78	0.38	0.07	0.05	0.28	0.76	0.97	1.00	1.00
W_A	0.87	0.79	0.67	0.46	0.15	0.05	0.21	0.66	0.95	1.00	1.00
W_{MU}	0.99	0.96	0.81	0.44	0.12	0.05	0.11	0.48	0.89	0.99	1.00
$s(0,0.5)$	0.99	0.97	0.82	0.46	0.12	0.05	0.18	0.52	0.81	0.95	0.99
$X4, T=20$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.58	0.38	0.18	0.06	0.03	0.05	0.18	0.44	0.77	0.95	0.99
W_{MLE}	0.67	0.47	0.23	0.08	0.03	0.05	0.18	0.44	0.77	0.95	0.99
W_A	0.65	0.48	0.29	0.15	0.07	0.05	0.15	0.38	0.73	0.93	0.99
W_{MU}	0.81	0.65	0.42	0.20	0.09	0.05	0.09	0.24	0.58	0.87	0.97
$s(0,0.5)$	0.81	0.65	0.42	0.19	0.08	0.05	0.11	0.29	0.55	0.79	0.93
$X4, T=40$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.99	0.97	0.81	0.41	0.09	0.05	0.31	0.77	0.97	1.00	1.00
W_{MLE}	1.00	0.98	0.83	0.42	0.09	0.05	0.30	0.77	0.97	1.00	1.00
W_A	0.90	0.77	0.67	0.50	0.17	0.05	0.20	0.63	0.93	1.00	1.00
W_{MU}	1.00	0.98	0.87	0.49	0.15	0.05	0.14	0.54	0.90	0.99	1.00
$s(0,0.5)$	1.00	0.99	0.88	0.53	0.18	0.05	0.21	0.59	0.85	0.95	0.99

Note: 2000 Replications with Simulated Critical Values.

Table 6.1e

Rejection Probabilities of the W_{OLS}, W_{MLE}, W_A, W_{MU} and $s(0,0.5)$ Tests at 5% Significance Level in the Linear Regression with AR(1) Disturbances; Testing $H_0: \rho = 0$ vs. $H_1: \rho \neq 0$, for Design Matrix $X5$ and $X6$

$X5, T=20$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.51	0.38	0.22	0.09	0.04	0.05	0.17	0.38	0.70	0.91	0.98
W_{MLE}	0.83	0.61	0.33	0.14	0.05	0.05	0.14	0.33	0.67	0.92	0.99
W_A	0.72	0.60	0.43	0.25	0.10	0.05	0.10	0.26	0.58	0.86	0.97
W_{MU}	0.88	0.69	0.40	0.20	0.09	0.05	0.08	0.20	0.50	0.81	0.96
$s(0,0.5)$	0.84	0.62	0.32	0.15	0.06	0.05	0.12	0.29	0.55	0.77	0.84
$X5, T=40$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	1.00	0.98	0.86	0.46	0.12	0.05	0.28	0.74	0.96	1.00	1.00
W_{MLE}	1.00	0.99	0.88	0.49	0.13	0.05	0.27	0.74	0.96	1.00	1.00
W_A	0.95	0.82	0.65	0.50	0.21	0.05	0.19	0.63	0.93	1.00	1.00
W_{MU}	1.00	0.99	0.89	0.52	0.16	0.05	0.15	0.54	0.89	0.99	1.00
$s(0,0.5)$	1.00	0.95	0.70	0.23	0.04	0.05	0.27	0.64	0.85	0.95	0.97
$X6, T=20$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.33	0.28	0.15	0.05	0.03	0.05	0.16	0.38	0.70	0.92	0.99
W_{MLE}	0.60	0.45	0.23	0.09	0.04	0.05	0.14	0.34	0.68	0.92	0.99
W_A	0.60	0.57	0.42	0.25	0.10	0.05	0.09	0.24	0.57	0.86	0.97
W_{MU}	0.40	0.14	0.06	0.05	0.05	0.05	0.06	0.06	0.07	0.27	0.73
$s(0,0.5)$	0.73	0.49	0.25	0.12	0.07	0.05	0.09	0.20	0.38	0.59	0.68
$X6, T=40$											
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.95	0.94	0.75	0.36	0.08	0.05	0.31	0.77	0.97	1.00	1.00
W_{MLE}	0.99	0.96	0.81	0.39	0.09	0.05	0.29	0.76	0.97	1.00	1.00
W_A	0.81	0.77	0.63	0.44	0.16	0.05	0.24	0.68	0.95	1.00	1.00
W_{MU}	0.96	0.76	0.27	0.05	0.01	0.05	0.28	0.73	0.96	1.00	1.00
$s(0,0.5)$	0.99	0.95	0.69	0.27	0.07	0.05	0.18	0.43	0.61	0.75	0.81

Note: 2000 Replications with Simulated Critical Values.

Table 6.2a

Rejection Probabilities of the W_{OLS}, W_{MLE} and W_{MU} Tests at 5% Significance Level in the Linear Regression with AR(1) Disturbances; Testing $H_0\!:\rho=0$ vs. $H_1\!:\rho\neq 0$, for Design Matrix $X7$

					$T=20$				
ρ	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80
W_{OLS}	0.35	0.30	0.20	0.10	0.10	0.21	0.40	0.55	0.53
W_{MLE}	0.74	0.48	0.24	0.10	0.10	0.17	0.40	0.70	0.92
W_{MU}	0.78	0.51	0.26	0.12	0.10	0.18	0.37	0.57	0.67
					$T=40$				
ρ	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80
W_{OLS}	0.93	0.84	0.53	0.16	0.05	0.23	0.61	0.89	0.91
W_{MLE}	0.99	0.88	0.51	0.13	0.05	0.22	0.62	0.94	0.99
W_{MU}	0.98	0.84	0.44	0.11	0.05	0.13	0.44	0.79	0.87

Table 6.2b

Rejection Probabilities of the W_{OLS}, W_{MLE}, W_{MU} and $s(0,0.5)$ Tests at 5% Significance Level in the Linear Regression with AR(1) Disturbances; Testing $H_0\!:\rho=0$ vs. $H_1\!:\rho\neq 0$, for Design Matrix $X8$

						$T=20$					
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.06	0.11	0.12	0.09	0.05	0.05	0.08	0.15	0.21	0.19	0.10
W_{MLE}	0.77	0.54	0.26	0.10	0.05	0.05	0.07	0.17	0.39	0.72	0.91
W_{MU}	0.78	0.50	0.21	0.07	0.04	0.05	0.08	0.17	0.31	0.47	0.64
$s(0,0.5)$	0.69	0.46	0.26	0.15	0.09	0.05	0.04	0.06	0.15	0.33	0.51
						$T=40$					
ρ	1.00	0.80	0.60	0.40	0.20	0.00	-0.20	-0.40	-0.60	-0.80	-0.95
W_{OLS}	0.37	0.68	0.66	0.41	0.13	0.05	0.17	0.47	0.71	0.69	0.34
W_{MLE}	0.99	0.97	0.79	0.37	0.09	0.05	0.16	0.47	0.85	0.99	0.99
W_{MU}	1.00	0.96	0.75	0.33	0.08	0.05	0.13	0.39	0.65	0.73	0.82
$s(0,0.5)$	0.98	0.94	0.80	0.49	0.16	0.05	0.14	0.44	0.75	0.88	0.80

Note: 2000 Replications with Simulated Critical Values.

Figure 6.1

Empirical Power Curves of the W_{OLS}, W_{MLE} and W_{MU} Tests at the 5% Significance Level in the Linear Regression with AR(1) Disturbances; Testing $H_0\!:\rho=0$ against. $H_1\!:\rho\neq 0$

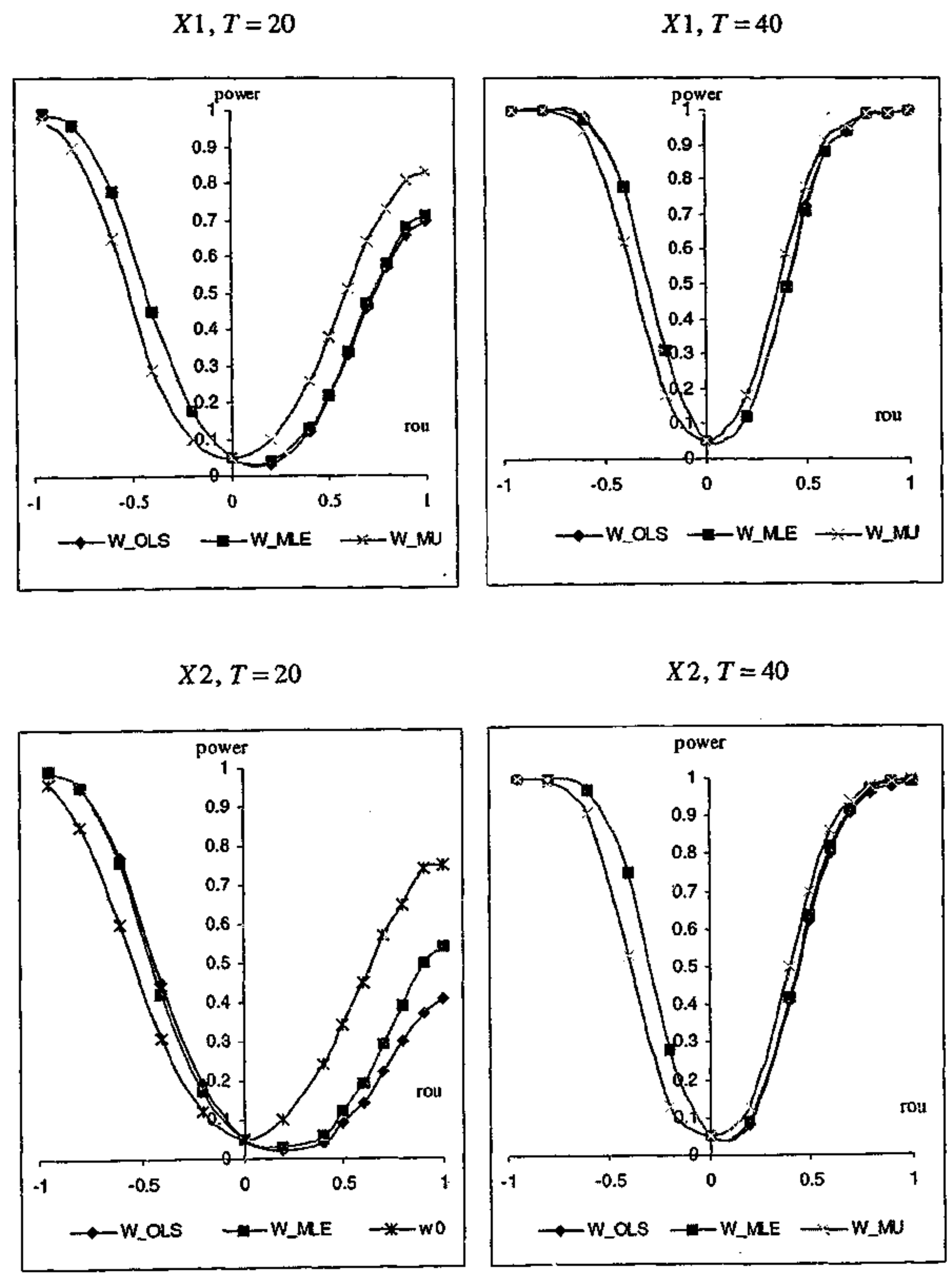
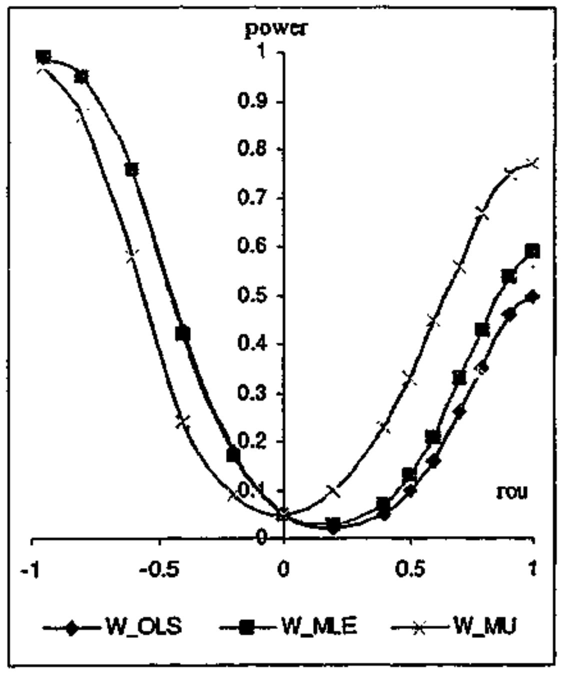
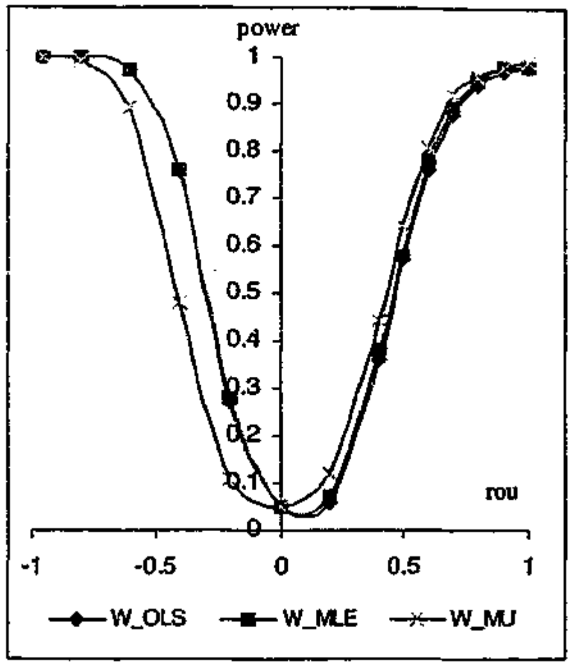


Figure 6.1 Continued

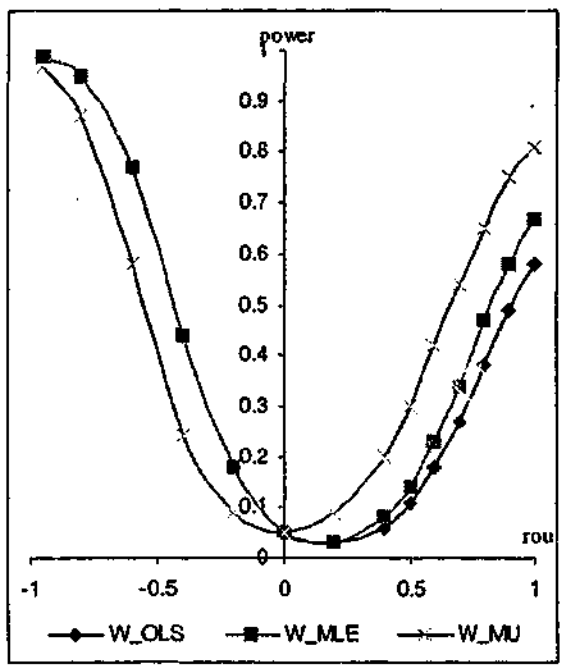
$X3, T=20$



$X3, T=40$



$X4, T=20$



$X4, T=40$

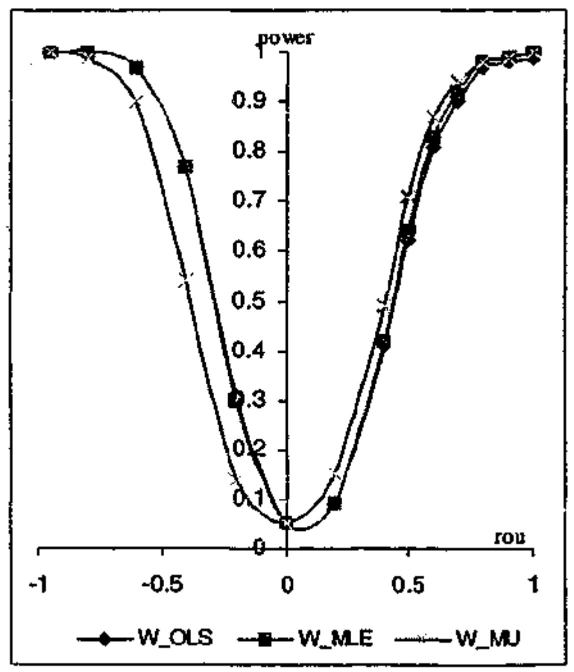
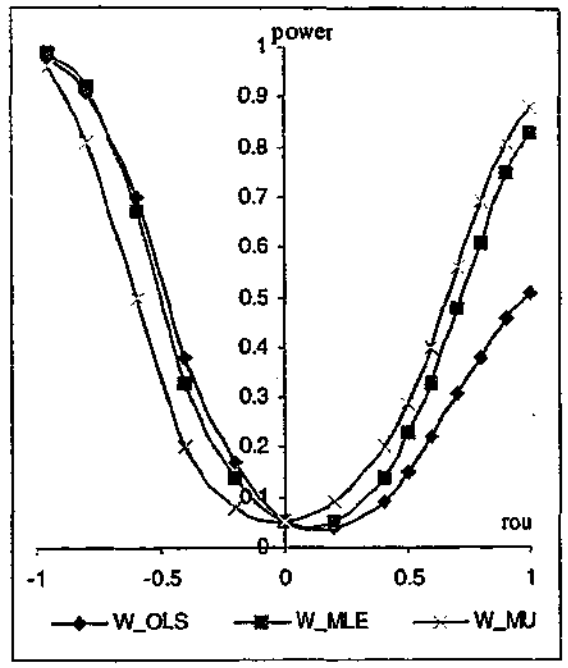
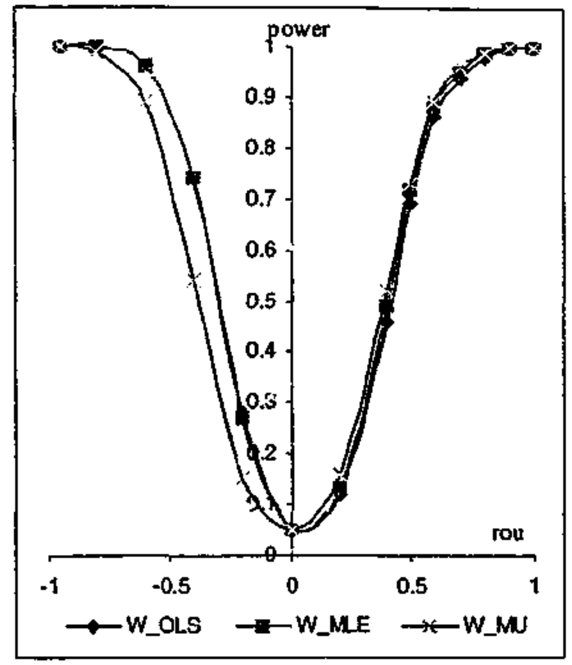


Figure 6.1 Continued

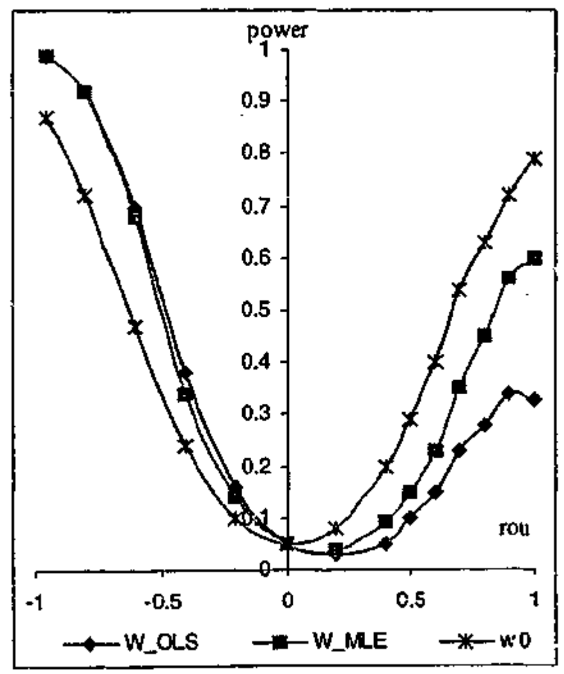
$X5, T=20$



$X5, T=40$



$X6, T=20$



$X6, T=40$

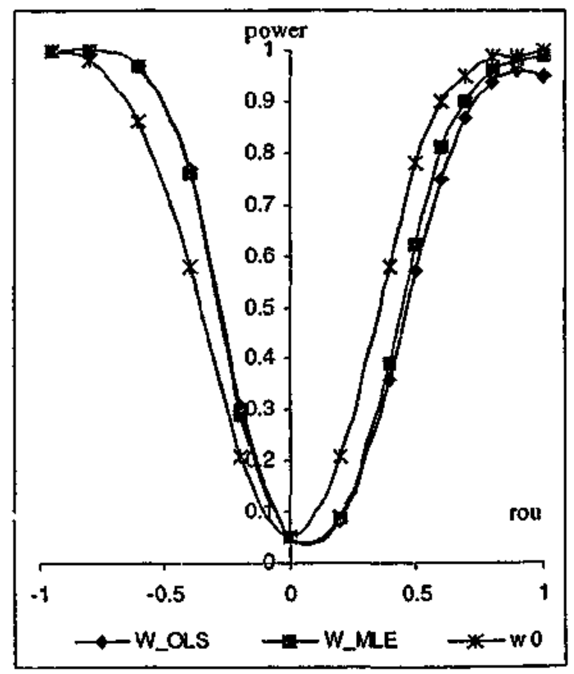
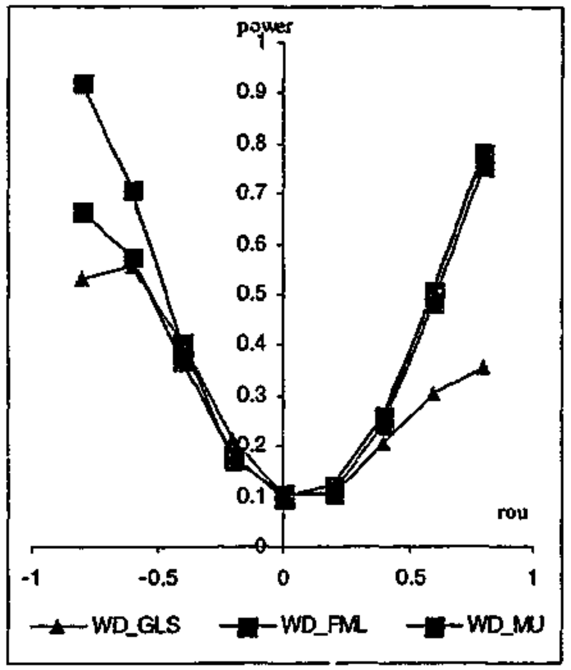
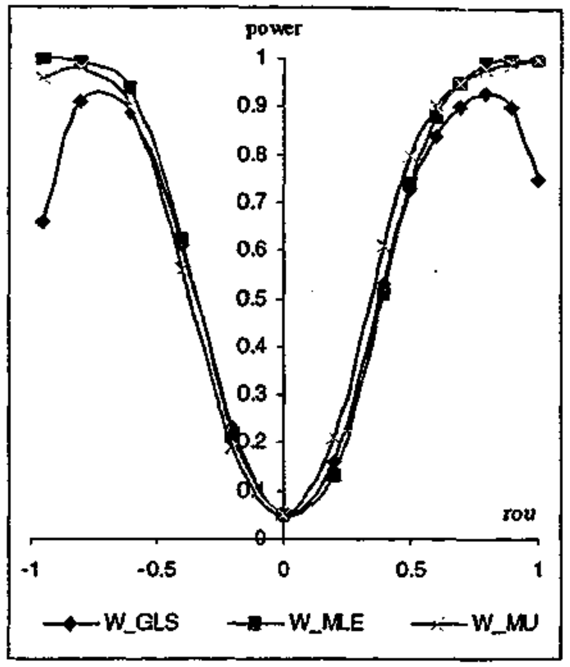


Figure 6.1 Continued

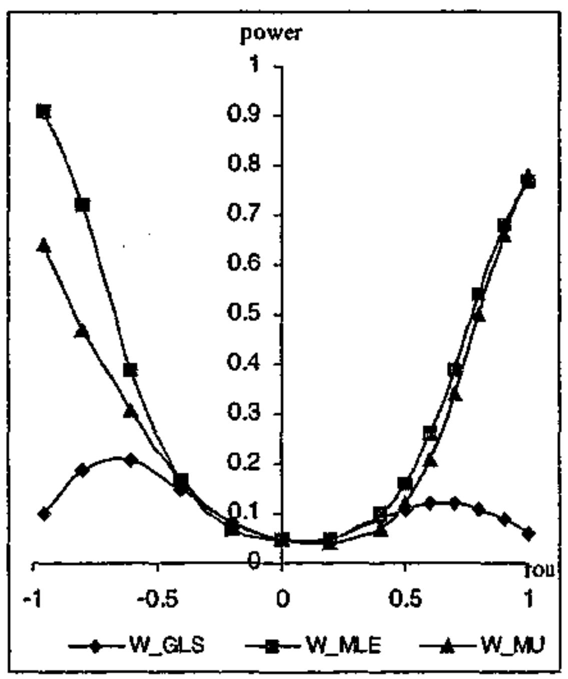
$X7, T=20$



$X7, T=40$



$X8, T=20$



$X8, T=40$

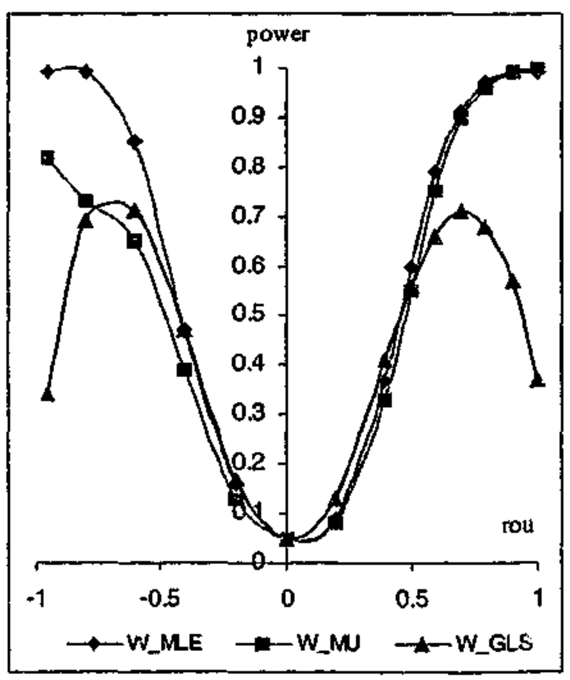


Table 6.3
Rejection Probabilities of the W_{OLS}, W_{MLE} and W_{MU} Tests at the 10% Significance Level for Positive ρ Values in the Linear Regression with AR(1) Disturbances; Testing $H_0: \rho = 0$ against $H_1: \rho \neq 0$; $T = 60$

	$X1$				$X2$		
ρ	W_{OLS}	W_{MLE}	W_{MU}	ρ	W_{OLS}	W_{MLE}	W_{MU}
0.800	0.999	0.999	1.000	0.800	0.999	0.999	1.000
0.600	0.987	0.987	0.992	0.600	0.973	0.976	0.989
0.400	0.830	0.826	0.880	0.400	0.764	0.768	0.881
0.200	0.304	0.299	0.389	0.200	0.247	0.246	0.427
0.000	0.100	0.100	0.100	0.000	0.100	0.100	0.100
	$X3$				$X5$		
ρ	W_{OLS}	W_{MLE}	W_{MU}	ρ	W_{OLS}	W_{MLE}	W_{MU}
0.800	0.998	0.998	1.000	0.800	0.999	0.999	1.000
0.600	0.985	0.984	0.995	0.600	0.986	0.989	0.994
0.400	0.744	0.750	0.854	0.400	0.822	0.830	0.851
0.200	0.222	0.222	0.392	0.200	0.312	0.317	0.377
0.000	0.100	0.100	0.100	0.000	0.100	0.100	0.100
	$X7$				$X8$		
ρ	W_{OLS}	W_{MLE}	W_{MU}	ρ	W_{OLS}	W_{MLE}	W_{MU}
0.800	0.993	1.000	1.000	0.800	0.952	0.999	1.000
0.600	0.978	0.992	0.995	0.600	0.929	0.980	0.984
0.400	0.820	0.849	0.870	0.400	0.751	0.787	0.803
0.200	0.350	0.354	0.378	0.200	0.309	0.287	0.322
0.000	0.100	0.100	0.100	0.000	0.100	0.100	0.100

Note: 2000 Replications with Simulated Critical Values.

Table 6.4a
Rejection Probabilities of the W_{OLS}, W_{MLE}, W_{MU} and $s(1,0)$ Tests at the 5% Significance Level in the Linear Regression with AR(1) or Random Walk Disturbances; Testing $H_0: \rho = 1$ against $H_1: \rho < 1$, for Design Matrix $X1$

ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
				$T=20$					
W_{OLS}	0.05	0.06	0.08	0.12	0.18	0.25	0.37	0.61	0.85
W_{MLE}	0.05	0.06	0.08	0.12	0.19	0.27	0.39	0.65	0.87
W_{MU}	0.05	0.06	0.09	0.14	0.21	0.31	0.43	0.68	0.88
$s(1,0)$	0.05	0.07	0.10	0.14	0.21	0.30	0.43	0.69	0.89
				$T=40$					
ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
W_{OLS}	0.05	0.07	0.13	0.28	0.48	0.71	0.88	0.99	1.00
W_{MLE}	0.05	0.08	0.15	0.34	0.56	0.80	0.92	0.99	1.00
W_{MU}	0.05	0.07	0.15	0.34	0.57	0.79	0.91	0.99	1.00
$s(1,0)$	0.05	0.07	0.15	0.34	0.56	0.79	0.92	1.00	1.00

Note: 2000 Replications with Simulated Critical Values.

Figure 6.2a
Simulated Power Curves of the W_{OLS}, W_{MLE} and W_{MU} Tests at the 5% Significance Level for Design Matrix $X1$

$T=20$ $\qquad$ $T=40$

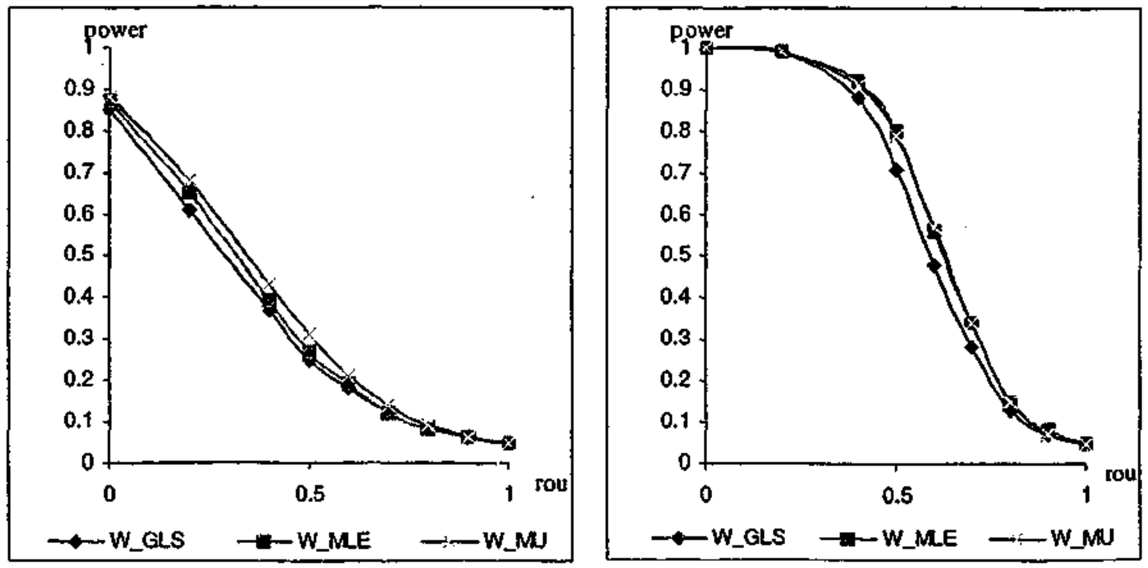


Table 6.4b
Rejection Probabilities of the W_{OLS}, W_{MLE}, W_{MU} and $s(1,0)$ Tests at the 5% Significance Level in the Linear Regression with AR(1) or Random Walk Disturbances; Testing $H_0: \rho = 1$ against $H_1: \rho < 1$, for Design Matrix $X2$

ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
				$T=20$					
W_{OLS}	0.05	0.04	0.06	0.08	0.11	0.16	0.24	0.42	0.68
W_{MLE}	0.05	0.04	0.06	0.09	0.12	0.17	0.24	0.43	0.69
W_{MU}	0.05	0.05	0.08	0.12	0.17	0.25	0.35	0.57	0.78
$s(1,0)$	0.05	0.05	0.08	0.11	0.16	0.23	0.33	0.56	0.79
				$T=40$					
ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
W_{OLS}	0.05	0.07	0.14	0.26	0.44	0.65	0.83	0.98	1.00
W_{MLE}	0.05	0.10	0.21	0.40	0.63	0.84	0.94	1.00	1.00
W_{MU}	0.05	0.10	0.20	0.39	0.61	0.82	0.93	0.99	1.00
$s(1,0)$	0.05	0.09	0.20	0.38	0.60	0.82	0.94	1.00	1.00

Note: 2000 Replications with Simulated Critical Values.

Figure 6.2b
Simulated Power Curves of the W_{OLS}, W_{MLE} and W_{MU} Tests at the 5% Significance Level for Design Matrix $X2$

$T=20$ $\qquad$ $T=40$

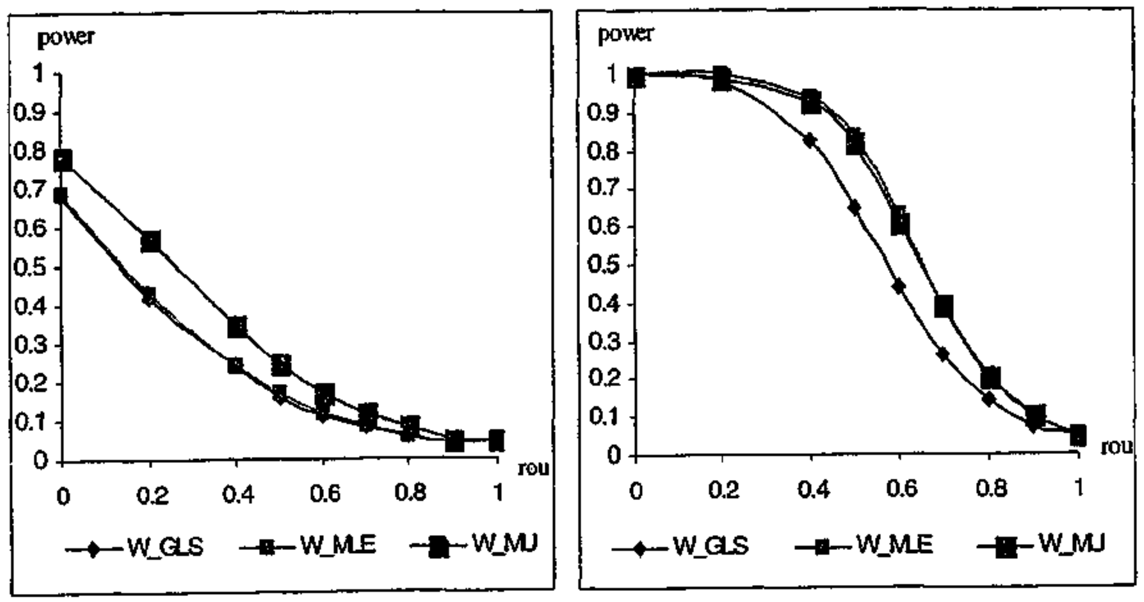


Table 6.4c
Rejection Probabilities of the W_{OLS}, W_{MLE}, W_{MU} and $S(1,0)$ Tests at the 5% Significance Level in the Linear Regression with AR(1) or Random Walk Disturbances; Testing $H_0\!:\rho=1$ against $H_1\!:\rho<1$, for Design Matrix $X6$

				$T=20$					
ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
W_{OLS}	0.05	0.03	0.04	0.05	0.07	0.09	0.14	0.27	0.51
W_{MLE}	0.05	0.03	0.04	0.05	0.07	0.10	0.14	0.27	0.50
W_{MU}	0.05	0.07	0.12	0.17	0.26	0.34	0.45	0.63	0.76
$S(1,0)$	0.05	0.07	0.10	0.15	0.22	0.30	0.43	0.63	0.83
				$T=40$					
ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
W_{OLS}	0.05	0.02	0.03	0.07	0.13	0.25	0.44	0.80	0.98
W_{MLE}	0.05	0.11	0.23	0.47	0.71	0.88	0.96	1.00	1.00
W_{MU}	0.05	0.13	0.30	0.56	0.76	0.88	0.94	0.98	0.98
$S(1,0)$	0.05	0.11	0.23	0.47	0.70	0.87	0.97	1.00	1.00

Note: 2000 Replications with Simulated Critical Values.

Figure 6.2c
Simulated Power Curves of the W_{OLS}, W_{MLE} and W_{MU} Tests at the 5% Significance Level for Design Matrix $X6$

$T=20$ $T=40$

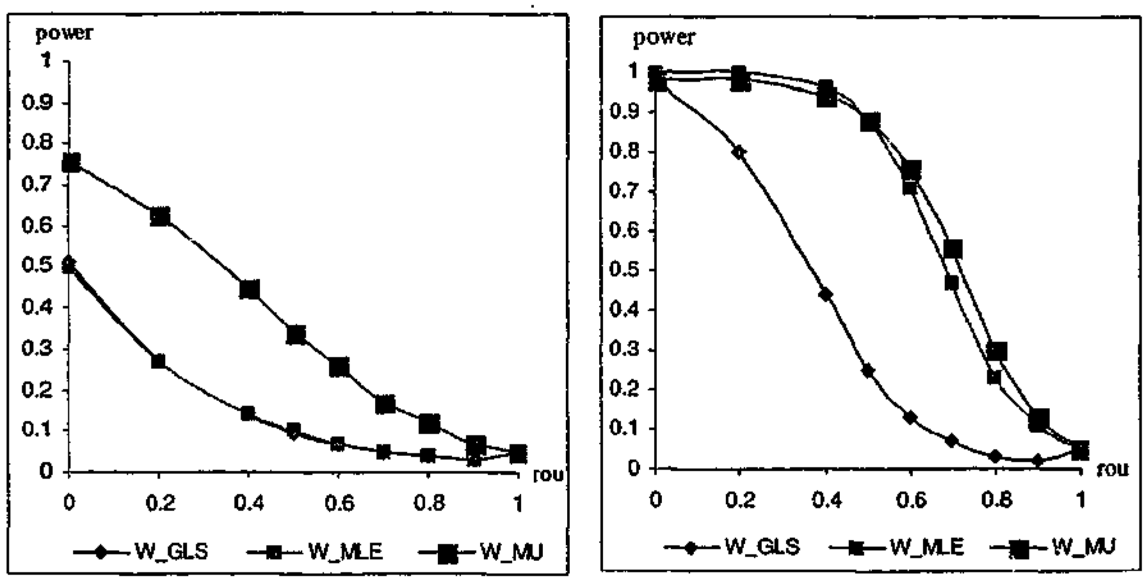


Table 6.4d
Rejection Probabilities of the W_{OLS}, W_{MLE}, W_{MU} and $S(1,0)$ Tests at the 5% Significance Level in the Linear Regression with AR(1) or Random Walk Disturbances; Testing $H_0\!:\rho=1$ against $H_1\!:\rho<1$, for Design Matrix $X8$

				$T=20$					
ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
W_{OLS}	0.05	0.05	0.05	0.06	0.07	0.10	0.12	0.22	0.37
W_{MLE}	0.05	0.05	0.07	0.10	0.12	0.20	0.28	0.42	0.51
W_{MU}	0.05	0.09	0.15	0.24	0.36	0.48	0.59	0.77	0.85
$S(1,0)$	0.05	0.09	0.14	0.23	0.34	0.46	0.58	0.77	0.87
				$T=40$					
ρ	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.20	0.00
W_{OLS}	0.05	0.02	0.01	0.01	0.02	0.03	0.07	0.25	0.56
W_{MLE}	0.05	0.12	0.33	0.61	0.83	0.95	0.98	1.00	1.00
W_{MU}	0.05	0.13	0.35	0.61	0.82	0.94	0.97	1.00	1.00
$S(1,0)$	0.05	0.14	0.35	0.63	0.84	0.96	0.99	1.00	1.00

Note: 2000 Replications with Simulated Critical Values.

Figure 6.2d
Simulated Power Curves of the W_{OLS}, W_{MLE} and W_{MU} Tests at the 5% Significance Level for Design Matrix $X8$

$T=20$ $T=40$

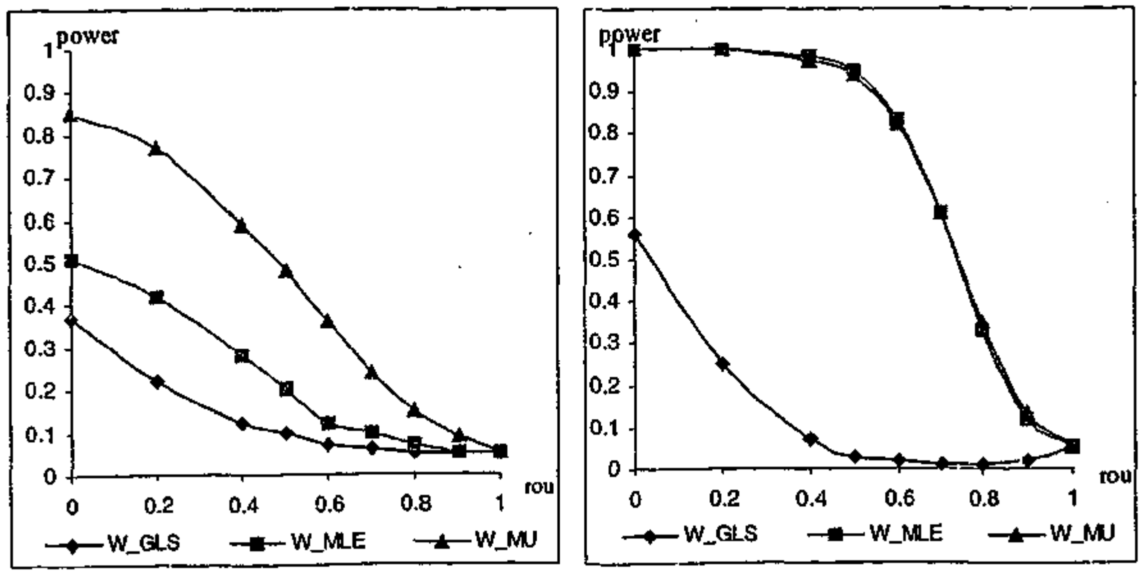


Table 6.5
Rejection Probabilities of the Wald Tests Based on $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ at the 10% Significance Level in the Dynamic Linear Regression Model for Design Matrices $X1$, $X4$, and $X5$;

γ	0.90	0.80	0.60	0.40	0.20	0.10	0.00	-0.10	-0.20	-0.40	-0.60	-0.80
						$X1$, $T=20$						
W_{OLS}	1.00	1.00	0.64	0.22	0.07	0.06	0.10	0.19	0.28	0.66	0.91	0.98
W_{MU}	1.00	1.00	0.97	0.56	0.20	0.13	0.10	0.14	0.20	0.534	0.86	0.98
						$X1$, $T=40$						
W_{OLS}	1.00	1.00	0.96	0.62	0.17	0.08	0.10	0.22	0.42	0.86	0.99	1.00
W_{MU}	1.00	1.00	0.99	0.82	0.32	0.16	0.10	0.16	0.31	0.80	0.99	1.00
						$X4$, $T=20$						
W_{OLS}	1.00	1.00	0.95	0.14	0.05	0.06	0.10	0.18	0.27	0.66	0.90	0.99
W_{MU}	1.00	1.00	1.00	0.77	0.21	0.10	0.10	0.11	0.17	0.50	0.78	0.97
						$X5$, $T=20$						
W_{OLS}	0.56	0.40	0.24	0.09	0.04	0.03	0.10	0.17	0.29	0.68	0.90	0.98
W_{MU}	0.82	0.75	0.55	0.34	0.16	0.11	0.10	0.11	0.20	0.51	0.80	0.98

Notes: All experiments are based on 1000 replications.
Simulated critical values are used.

Figure 6.3
Simulated Power Curves of the Wald Tests Based on $\hat{\gamma}_{OLS}$ and $\hat{\gamma}_{MU}$ at the 10% Significance Level in the Dynamic Linear Regression Model

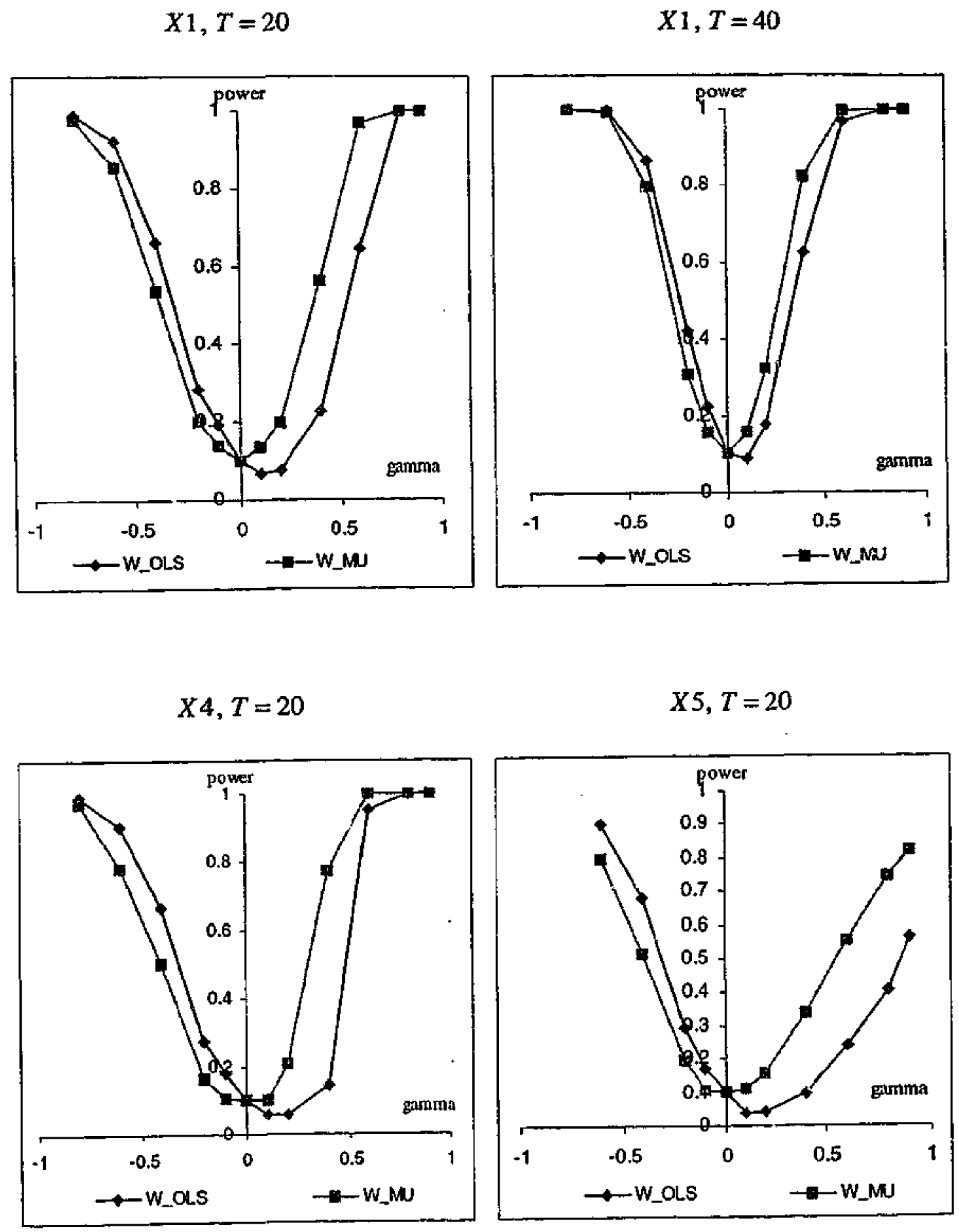


Table 6.6a

RMSEPs of $\hat{y}^{OLS}_{T+1}$, $\hat{y}^{MLE}_{T+1}$, $\hat{y}^{A}_{T+1}$ and $\hat{y}^{MU}_{T+1}$ for Positive ρ Values in the Linear Regression with AR(1) Disturbances: RMSEP of 2000 Forecasts for Design Matrix $X1$, $X2$ and $X3$

	$X1, T=20$				$X1, T=40$			
ρ	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$
1.00	1.13	1.12	1.05	1.04	1.09	1.09	1.05	1.05
0.90	1.18	1.17	1.11	1.10	1.09	1.08	1.06	1.05
0.80	1.17	1.16	1.12	1.11	1.07	1.07	1.06	1.06
0.70	1.15	1.15	1.13	1.12	1.06	1.05	1.06	1.05
0.60	1.17	1.17	1.16	1.15	1.05	1.05	1.05	1.05
0.50	1.14	1.14	1.14	1.13	1.06	1.06	1.06	1.06
0.40	1.14	1.14	1.14	1.14	1.07	1.07	1.06	1.06
0.20	1.15	1.15	1.15	1.16	1.08	1.08	1.08	1.08
0.00	1.12	1.12	1.12	1.14	1.05	1.05	1.05	1.06

	$X2, T=20$				$X2, T=40$			
ρ	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$
1.00	1.71	1.69	1.67	1.64	1.62	1.60	1.56	1.55
0.90	1.68	1.66	1.65	1.62	1.52	1.51	1.52	1.51
0.80	1.60	1.58	1.59	1.56	1.40	1.40	1.42	1.41
0.70	1.51	1.50	1.52	1.49	1.30	1.30	1.33	1.32
0.60	1.44	1.43	1.45	1.43	1.21	1.21	1.22	1.22
0.50	1.37	1.37	1.39	1.37	1.16	1.16	1.17	1.17
0.40	1.32	1.32	1.32	1.33	1.11	1.11	1.11	1.11
0.20	1.23	1.24	1.24	1.26	1.06	1.06	1.06	1.06
0.00	1.16	1.16	1.17	1.19	1.03	1.03	1.03	1.04

	$X3, T=20$				$X3, T=40$			
ρ	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$
1.00	1.19	1.14	1.13	1.04	1.09	1.06	1.07	1.02
0.90	1.24	1.20	1.21	1.11	1.09	1.07	1.11	1.04
0.80	1.22	1.19	1.21	1.12	1.10	1.09	1.12	1.07
0.70	1.19	1.18	1.19	1.13	1.10	1.10	1.14	1.09
0.60	1.20	1.20	1.21	1.16	1.11	1.10	1.12	1.10
0.50	1.17	1.17	1.19	1.14	1.08	1.08	1.10	1.08
0.40	1.17	1.17	1.18	1.16	1.11	1.11	1.13	1.11
0.20	1.17	1.17	1.18	1.18	1.10	1.10	1.10	1.11
0.00	1.13	1.14	1.14	1.16	1.12	1.12	1.13	1.14

Table 6.6b

RMSEPs of $\hat{y}^{OLS}_{T+1}$, $\hat{y}^{MLE}_{T+1}$, $\hat{y}^{A}_{T+1}$ and $\hat{y}^{MU}_{T+1}$ for Positive ρ Values in the Linear Regression with AR(1) Disturbances: RMSEP of 2000 Forecasts for Design Matrix $X4$, $X5$ and $X6$

	$X4, T=20$				$X4, T=40$			
ρ	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$
1.00	1.29	1.23	1.22	1.11	1.14	1.10	1.14	1.07
0.90	1.30	1.25	1.28	1.16	1.12	1.10	1.16	1.07
0.80	1.29	1.26	1.28	1.18	1.10	1.09	1.14	1.08
0.70	1.24	1.22	1.25	1.18	1.09	1.08	1.14	1.07
0.60	1.24	1.23	1.26	1.20	1.08	1.07	1.10	1.07
0.50	1.19	1.19	1.21	1.18	1.09	1.08	1.09	1.08
0.40	1.17	1.17	1.18	1.17	1.09	1.09	1.10	1.09
0.20	1.14	1.14	1.15	1.16	1.09	1.10	1.10	1.11
0.00	1.10	1.11	1.11	1.13	1.07	1.07	1.08	1.08

	$X5, T=20$				$X5, T=40$			
ρ	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$
1.00	2.01	1.49	1.71	1.42	1.34	1.23	1.28	1.22
0.90	1.81	1.50	1.64	1.44	1.26	1.24	1.29	1.24
0.80	1.66	1.50	1.58	1.45	1.25	1.24	1.28	1.24
0.70	1.56	1.50	1.55	1.47	1.29	1.28	1.31	1.28
0.60	1.50	1.48	1.51	1.47	1.28	1.28	1.30	1.28
0.50	1.46	1.47	1.49	1.47	1.27	1.27	1.28	1.27
0.40	1.43	1.45	1.46	1.47	1.35	1.35	1.35	1.35
0.20	1.40	1.42	1.42	1.46	1.39	1.39	1.40	1.40
0.00	1.41	1.43	1.42	1.46	1.38	1.39	1.39	1.40

	$X6, T=20$				$X6, T=40$			
ρ	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$	$\hat{y}^{OLS}_{T+1}$	$\hat{y}^{MLE}_{T+1}$	$\hat{y}^{A}_{T+1}$	$\hat{y}^{MU}_{T+1}$
1.00	2.85	2.47	2.54	2.07	1.27	1.19	1.27	1.18
0.90	2.56	2.31	2.34	2.07	1.26	1.21	1.28	1.20
0.80	2.31	2.16	2.24	2.06	1.25	1.22	1.28	1.22
0.70	2.21	2.14	2.23	2.15	1.30	1.27	1.33	1.26
0.60	2.15	2.11	2.20	2.18	1.30	1.28	1.32	1.28
0.50	2.16	2.17	2.24	2.28	1.29	1.29	1.31	1.28
0.40	2.16	2.18	2.22	2.35	1.37	1.37	1.38	1.36
0.20	2.21	2.25	2.24	2.52	1.40	1.40	1.40	1.43
0.00	2.25	2.31	2.27	2.70	1.37	1.38	1.38	1.50

Table 6.6c

RMSEPs of $\hat{y}_{T+1}^{OLS}$, $\hat{y}_{T+1}^{MLE}$, and $\hat{y}_{T+1}^{MU}$ for Positive ρ Values in the Linear Regression with AR(1) Disturbances: RMSEP of 2000 Forecasts for Design Matrix $X8$

	$T=20$			$T=40$		
ρ	$\hat{y}_{T+1}^{OLS}$	$\hat{y}_{T+1}^{MLE}$	$\hat{y}_{T+1}^{MU}$	$\hat{y}_{T+1}^{OLS}$	$\hat{y}_{T+1}^{MLE}$	$\hat{y}_{T+1}^{MU}$
1.00	1.83	1.37	1.23	2.58	1.09	1.06
0.90	1.60	1.33	1.26	1.67	1.08	1.06
0.80	1.46	1.32	1.26	1.30	1.07	1.07
0.70	1.39	1.30	1.27	1.17	1.07	1.07
0.60	1.32	1.29	1.26	1.12	1.06	1.06
0.50	1.29	1.28	1.27	1.05	1.04	1.05
0.40	1.26	1.27	1.27	1.03	1.04	1.05
0.20	1.20	1.24	1.25	1.06	1.07	1.07
0.00	1.15	1.29	1.23	1.02	1.03	1.03

Table 6.7

RMSEPs of $\hat{y}_{T+1}^{OLS}$ and $\hat{y}_{T+1}^{MU}$ of 1000 Forecasts in the Dynamic Linear Regression Model

γ	0.90	0.80	0.60	0.40	0.20	0.10	0.00	-0.10	-0.20	-0.40	-0.60	-0.80
						$X1, T=20$						
$\hat{y}_{T+1}^{OLS}$	1.51	1.43	1.49	1.50	1.51	1.55	1.57	1.51	1.50	1.54	1.59	1.51
$\hat{y}_{T+1}^{MU}$	1.38	1.46	1.22	1.18	1.14	1.20	1.22	1.11	1.14	1.16	1.20	1.16
						$X1, T=40$						
$\hat{y}_{T+1}^{OLS}$	1.49	1.41	1.44	1.40	1.49	1.49	1.48	1.43	1.47	1.51	1.48	1.53
$\hat{y}_{T+1}^{MU}$	1.63	1.24	1.07	1.07	1.09	1.13	1.08	1.04	1.05	1.08	1.07	1.11
						$X4, T=20$						
$\hat{y}_{T+1}^{OLS}$	1.13	1.10	1.08	1.13	1.15	1.12	1.16	1.18	1.11	1.13	1.11	1.10
$\hat{y}_{T+1}^{MU}$	1.12	1.09	1.08	1.10	1.09	1.07	1.10	1.15	1.07	1.11	1.11	1.11
						$X5, T=20$						
$\hat{y}_{T+1}^{OLS}$	1.31	1.33	1.35	1.27	1.38	1.30	1.36	1.38	1.34	1.37	1.35	1.35
$\hat{y}_{T+1}^{MU}$	1.16	1.21	1.23	1.29	1.28	1.21	1.25	1.29	1.27	1.27	1.27	1.29

Chapter 7

Conclusion

7.1 Introduction

Econometric modelling plays a central role in empirical economic studies. The availability of rapidly expanding economic data bases has created a demand for more complex models capable of capturing the state of dynamic economic systems. Within the large enterprise of econometric modelling, point estimation remains a fundamental device for analysing and disclosing the relationship between economic variables. It also serves as a building block and provides inputs for other inference procedures such as hypothesis testing and forecasting. A prominent request from empirical economic research is that inference procedures should be relevant and efficient for the very situation under study, not just for the ideal but unrealistic case of indefinitely large samples, on which most of the classic estimation and hypothesis testing procedures are based. The estimators or tests are deemed efficient only if they are able to explore fully the information contained in the data at hand. Such demand has driven econometricians to search for modelling methodology that is based on exact sample results, instead of relying on the ones only supported by asymptotic theory. The ever-increasing computing capability has facilitated this shift of focus. In terms of estimation, researchers are not satisfied to just have an estimator which is asymptotically normal but with unknown small sample performance. Research work in pursuit of exact finite sample inference procedures has flourished in recent years. The main thrust of this thesis is consistent with this trend. Its main aim is to search for some small sample bias-correction techniques that are capable of constructing estimators that are (approximately) median-unbiased. In the course of pursuing this aim, two general applicable methods were developed, with illustrations given in estimation of the linear autoregressive models.

In the section that follows, we discuss the core ideas of the methods of constructing MU estimators, and summarise the major findings. Some recommendations are drawn for how these methods should be used in practice.

Aspects related to, but not covered in, this thesis will be deliberated on in the final section of the chapter as potential topics for future research.

7.2 Summary of Findings

Median-unbiasedness enjoys some attractive advantages over mean-unbiasedness as a measure of impartiality of a point estimator in some circumstances. Efforts were directed to survey these circumstances in the literature review chapter. A comparison of the definition of the two unbiasedness criteria revealed a host of problems with mean-unbiasedness, such as lack of robustness, not being invariant to one-to-one transformations and being ill-defined when the parameter space has a closed boundary. The use of MU estimators provides a remedy to these problems associated with the mean-unbiased estimators. Apart from the early examples, MU estimators are shown to be used most frequently in estimating linear autoregressive models with high persistence. They provide an alternative inference device that is complimentary to the unit root tests which may suffer from poor power. However, these examples only considered models without exogenous variables. The extension to models with explanatory variables is important for practical research.

Despite its importance, there is a lack of a systematic approach towards constructing MU estimators in the literature. There is however, a vast literature on bias reduction techniques in the context of mean-unbiased estimation. Bias can either be corrected after the initial estimator has been computed or be prevented beforehand. This can be achieved by either evaluating the bias function analytically or via resampling schemes such as jackknife or bootstrap. It is natural for us to modify some of these techniques and apply them to correcting the median bias of an estimator. Therefore the origin of this research lies in the idea of borrowing techniques from the mean bias correction literature in order to develop methods for constructing MU estimators.

Chapter 3 developed two general applicable methods of constructing MU estimators. One is based on adjusting the estimating equations and the second is

based on inverting the median function of a significance test statistic. Both methods can be regarded as extensions of Lehmann's (1959) result, which links the existence of an optimal MU estimator to the conditional distribution of the sufficient statistics. When an estimating equation generates a biased estimator, one can effectively adjust these equations to reduce the estimation bias. The condition for a MU estimating equation to deliver a MU estimator was shown to be more general than those for a mean-unbiased estimating equation to produce a mean-unbiased estimator. So for a given estimating equation, we suggest subtracting its median function from the original estimating function and if the difference is monotonic, we will get a MU estimator by solving the adjusted estimating equations. The advantage of this approach is that no analytical or simulated bias function is required, while the disadvantage is the difficulty of verifying the monotonicity of the new estimating function. The proposed adjusted estimating equation was shown to be equivalent to the modified versions of two existing bias reduction techniques. An iterative algorithm was developed to solve the adjusted estimating equation.

Chapter 4 provides two examples of applying the method of constructing MU estimators by adjusting the estimating equations. In both examples, we chose to adjust the marginal likelihood score equations due to their better small sample performance compared with the profile likelihood counterpart. The adjustment to the marginal likelihood score in the linear regression model with AR(1) disturbances can be computed exactly using Imhof's (1961) algorithm, and the new estimator was shown to be almost free of bias in most cases. While in the dynamic linear regression model, the median function of the marginal likelihood score is not invariant to nuisance parameters, and we have to substitute these nuisance parameters by their consistent estimators and adjust the estimating equation recursively. As a result, the new estimator is approximately MU. It was found that the remaining bias in the new estimator is minimal compared with that of the OLS estimator. In both models, the RMSE of the new estimator is generally smaller than that of the OLS estimator especially for positive parameter values. For the lagged dependent variable coefficient, the confidence intervals based on the new estimator were shown to have better coverage probabilities than those based on the OLS estimator. These two examples lead us to believe that correcting the median bias in an estimator by adjusting the estimating equations towards median-unbiasedness can be effective.

The proposed method does not require knowledge of the form of the bias function. Our results show that the bias correction can be quite accurate and the overall risk of the new estimator tends to be smaller than that of the biased estimators. The drawback of this approach is that the likelihood function and the scores are non-standard when the autoregressive coefficient goes to unity. The limiting distribution of the marginal likelihood score in this case is not clear. Therefore it is not easy to extend this method to cover the interesting case of unit roots and random walk disturbances.

In case the proposed adjustment to the estimating equations is hard to compute or the monotonicity of the adjusted estimating function does not hold, a MU estimator can be constructed by inverting some 'well-chosen' test statistics at the 50% significance level. This was the second method of computing MU estimators developed in Chapter 3. Depending on whether the median function of the test statistic is monotonic, two different methods were proposed: fixed point inversion of the median function of a single test statistic and grid inversion of a median envelope. The latter is theoretically more reliable despite the extra computational costs. We addressed the issue of choosing a good test to invert, which was largely ignored by most of previous studies. Effort was directed to disclose the relationship between the power properties of a test and the effectiveness of inverting its median function for a MU estimator. In many cases when a UMP test does not exist, we recommend two classes of optimal tests, point optimal tests and locally best tests to be considered as good candidates when choosing a test statistic to invert, mainly because of their sound small sample power properties. This approach is suitable for estimating a large class of linear models with autoregressive errors, with heteroscedastic errors and with time-varying coefficients.

We illustrated the method of inverting the median function of a significance test to construct MU estimators in a practical example in Chapter 5. Most popular tests, such as the DW test, the LM test and the t test were shown to have non-monotonic median functions for some design matrices. In particular, we pointed out Andrews' (1993) estimator based on the OLS estimator could not be extended to models with exogenous regressors due to the same problem. The method based on the POI test statistics provides a remedy. It was shown that the POI test has a strictly

monotonic median envelope for all design matrices considered. Monte Carlo evidence showed that when the median function of a single POI test is monotonic for a design matrix, the MU estimator based on fixed point inversion method is almost exactly median unbiased for all parameter values and has a smaller RMSE compared with other estimators. In particular, it generally performs better than Andrews' (1993) MU estimator except in the model with only an intercept and a time trend as the regressors. For the design matrices where a single POI test fails to deliver a monotonic median function, inverting the median envelope of a series of POI tests can be used to construct a MU estimator. It was shown that the proposed estimator almost eliminates the bias present in the OLS and MLE estimators. The bias correction is strikingly substantial for these designs, as they represent the extreme cases where the small sample bias of the conventional estimators is most serious. Finally, we examined the robustness of the proposed estimator to non-normal errors and error structure misspecifications. It was found that the estimator based on inverting median envelope is more robust to non-normal errors than Andrews' estimator. It performs well under all error structures examined, except for errors with an MA component or with the autoregressive order misspecified.

Chapter 6 is concerned with hypothesis testing and forecasting based on the MU estimators developed in the previous chapters. We provided some convincing evidence of the effectiveness of improving the small sample performance of the Wald test by using MU estimators. The removal of the bias in the estimator leads to the correction of the local biasedness of the Wald test in both the linear regression model with AR(1) disturbances and the dynamic linear regression model. The power curve of the Wald test based on the MU estimator is properly centred at the null hypothesis and also tightened on the positive side of H_0. For non-local alternatives, the modified Wald test is not affected by the problem of non-monotonic power, which usually plagues Wald tests based on the OLS estimators. Although the modified Wald-test is asymptotically equivalent to those based on the ML estimator, it provides an effective remedy for the small sample deficiencies of the Wald test that are frequently encountered by researchers. A similar conclusion can also be drawn for forecasting accuracy. The predictor based on the MU estimator usually has smaller average prediction errors compared with those based on biased estimators.

Therefore it is useful to correct the small sample bias of the estimator in order to increase prediction accuracy.

To summarise the major findings, this thesis has established two small sample estimation procedures based on exploring the estimating equations and the exact distributions of the significance test statistics as alternatives to the existing bias-correction techniques for situations where mean-unbiasedness is too restrictive or robustness is highly desirable. It was shown that in many cases, estimation procedures based on asymptotic theory are usually more general and easier to use, but may suffer from small sample deficiencies. The procedures that are able to explore the information contained in given data sets may therefore be preferable. However, the proposed procedures, just like any other single estimation procedure, cannot be consistently superior to the other procedures. Hence, the choice of which procedure to use clearly depends on the model of interest, data set given and inference procedures under examination, which dictate not only the utility function used to assess the quality of an inference procedure, but also the cost-effectiveness of using each procedure. Bearing this in mind, the improvement in the small sample performance achieved by applying the proposed estimation procedures should be enough for researchers to use them in empirical econometric modelling.

7.3 Limitations and Future Research

The research reported in this thesis is by no means totally comprehensive in nature, and possibilities for extension and further development exist. The quest to find exact sample solutions for any econometric model has always been a hard task. The research reported in this thesis is no exception. Our effort to resolve some given problems may itself bring about more new open questions. Below we outline some limitations to our approach and suggest scope for future research.

1. The thesis is mainly concerned with estimating a single scalar parameter of interest. Some attempts were included to study estimating multiple parameters simultaneously in the dynamic linear regression model in Chapter 4. Methods were also proposed to extend our algorithms to multi-variate

parameter cases in Chapter 3. But like most other applications existing in the literature, MU estimation has not been used efficiently for multi-dimensional parameter estimation problems. More work has to be done to better define an impartiality measure for a multi-variate estimator. Some recent developments on constructing simultaneous confidence sets (bands) for multi-parameter models, which include Hall (1987), Beran and Miller (1986), Beran (1993), Grigoletto (1998), Chan et al. (1999) and Wright (2000b), may provide new revelations for calibrating the point estimators of multi-dimensional parameters. In the interest of maintaining a good focus for this thesis, we did not pursue this issue further. These problems, and methods to deal with them, in particular, warrant further research.

2. Constructing MU estimators by inverting the median envelope of POI test statistics was proposed in the context of a general class of linear models with various specifications of the error covariance matrix. Although the method was only illustrated in estimating first order autoregressive disturbances, it is readily extended to other models. It would be interesting to apply the proposed method to models with time-varying coefficients, MA(1) errors or heteroscedasticity errors, where POI test procedures have also been developed. For example, a promising application is to apply the proposed method to the POI tests developed by Shively (1988) and Brooks (1992) in the random walk coefficient model to improve the small sample efficiency of estimation and other inference procedures in this context.

3. The error terms in this study were assumed to be well-behaved and normally distributed. This is mainly to facilitate the exact evaluation of the median functions of test statistics via algorithms such as Imhof (1961). However, without the normality assumption, the methods proposed can still be applied with the median functions approximated by simulation. As it is impossible to simulate a continuous median function for indefinite number of points, non-parametric curve fitting techniques may be required to extrapolate the median functions outside the grid of simulated points. This approach was discussed in Hansen (1999). The efficiency of the proposed procedure based on simulated median functions, however, remains to be seen.

4. Although in Chapter 5, we examined the robustness of the proposed MU estimators under non-normal errors, it remains to be seen if these estimators are also robust to outliers or contaminated data, which ought to be a property possessed by the concept of median-unbiasedness. How to treat outliers and contaminated data in time series models has always been a difficult problem. Median-unbiased estimators may provide more robustness than conventional estimators. Therefore it would be useful to examine, for example, the breakdown properties of the proposed MU estimators.

5. The comparisons of estimators in this study were usually restricted to the proposed MU estimators and the conventional LS or ML estimators. In Chapter 5, we compared the proposed MU estimator with Andrews' MU estimator and also examined several other MU estimators based on inverting different test statistics. A possible research direction is to develop optimality results for various MU estimators in autoregressive models. Andrews (1993) admitted that it was not clear if his MU estimator was optimal in any sense. The same can be said about the MU estimators proposed in this thesis. More work needs to be done to assess the proposed MU estimators in terms of optimality measures such as the concentration measure or the closeness to the Cramer-Rao efficiency lower bound (see review in Chapter 2).

6. A possible topic for exploration is to extend the proposed estimation procedure to more complicated time series models, in which estimating autoregressive type of models is a component of the inference procedures. Examples of such models include panel data AR(1)/unit root model, autoregressive conditional heteroscedasticity models, vector autoregressive models and cointegrated systems. Linear autoregressive models with exogenous variables are essential for understanding these models. The small sample MU estimation procedures developed in this thesis may serve as a building block in developing exact finite sample inference procedures for these more complicated time series models.

References

Ahtola, J. and G.C. Tiao (1984): "Parameter Inference for a Nearly Nonstationary First-Order Autoregressive Model," *Biometrika*, 71, 263-272.

Altonji, J. and O. Ashenfelter (1980): "Wage Movement and the Labour Market Equilibrium Hypothesis," *Economica*, 47, 217-245.

Amemiya, T. (1982): "Two Stage Least Absolute Deviation Estimates," *Econometrica*, 50, 689-711.

Amemiya, T. (1985): *Advanced Econometrics*, Harvard, Cambridge, MA.

Andrews, D.W.K. (1986): "A Note on the Unbiasedness of Feasible GLS, Quasi-Maximum Likelihood, Robust, Adaptive, and Spectral Estimators of the Linear Model," *Econometrica*, 54, 687-698.

Andrews, D.W.K. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive Unit Root Models," *Econometrica*, 61, 139-165.

Andrews, D.W.K. and H.Y. Chen (1994): "Approximately Median Unbiased Estimation of Autoregressive Models," *Journal of Business & Economic Statistics*, 12, 187-204.

Andrews, D.W.K., and P.C.B. Phillips (1987): "Best Median Unbiased Estimation in Linear Regression with Bounded Asymmetric Loss Functions," *Journal of American Statistical Association*, 82, 886-893.

Andrews, D.W.K. and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383-1414.

Angelis, D.D., P. Hall and G.A. Young (1993): "Analytical and Bootstrap Approximations to Estimator Distribution in L1 Regression," *Journal of the American Statistical Association*, 88, 1310-1316.

Ara, I. (1995): "Marginal Likelihood Based Tests of Regression Disturbances," Unpublished Ph.D Thesis, Department of Econometrics and Business Statistics, Monash University.

Atukorala R. (2000): "The Use of an Information Criterion for Assessing Asymptotic Approximations in Econometrics," Unpublished Ph.D Thesis, Department of Econometrics and Business Statistics, Monash University.

Babu, G.J. and C.R. Rao (1988): "Joint Asymptotic Distribution of Marginal Quantile Function at a Point," *Journal of Multivariate Analysis*, 27, 15-23.

Bai, J.S. (1995): "Least Absolute Deviation Estimation of a Shift," *Econometric Theory*, 11, 403-436.

Bai, Z.D., X.R. Chen, B.Q. Mao and C.R. Rao (1988): "Asymptotic Theory of Least Distance Estimates in Multivariate Analysis," Technical Report 88-09, Centre for Multivariate Analysis, University of Pittsburgh.

Baillie, R.T. (1979): "The Asymptotic Mean-Squared Error of Multistep Prediction from the Regression Model with Autoregressive Errors," *Journal of the American Statistical Association*, 74, 175-184.

Barnard, G.A. (1963): "Comment on 'The Spectral Analysis of Point Processes' by M.S. Bartlett," *Journal of the Royal Statistical Society B*, 25, 294.

Barndorff-Nielsen, O. (1983): "On a Formula for the Distribution of the Maximum Likelihood Estimator," *Biometrika*, 70, 343-365.

Barndorff-Nielsen, O. (1986): "Inference on Full or Partial Parameters Based on the Standardised Likelihood Ratio," *Biometrika*, 73, 307-322.

Barndorff-Nielsen, O. (1994): "Adjusted Versions of Profile Likelihood, Directed Likelihood and Extended Likelihood," *Journal of the Royal Statistical Society B*, 56, 125-140.

Barndorff-Nielsen, O. and D.R. Cox (1984): "Bartlett Adjustments to the Likelihood Ratio Statistic and the Distribution of the Maximum Likelihood Estimator," *Journal of the Royal Statistical Society B*, 483-495.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991): "Bootstrapping Unstable First-order Autoregressive Processes," *Annals of Statistics*, 19, 1098-1101.

Bartels R. (1992): "On the Power Function of the Durbin-Watson Test," *Journal of Econometrics*, 51, 101-112.

Bartlett, M.S. (1955): "Approximate Confidence Intervals, III: A Bias Correction," *Biometrika*, 42, 201-204.

Bassett, G. and R. Koenker (1978): "Asymptotic Theory of Least Absolute Error Regression," *Journal of the American Statistical Association*, 73, 618-622.

Beach, C.M., and J.G. MacKinnon (1978): "A Maximum Likelihood Procedure for Regression with Autrocorrelated Errors", *Econometrica*, 46, 51-58.

Beale, E.M.L. (1960): "Confidence Regions in Nonlinear Estimation," *Journal of the Royal Statistical Society B*, 22, 41-88.

Bellhouse, D.R. (1978): "Marginal Likelihood Methods for Distributed Lag Models," *Statistiche Hefte*, 19, 2-14.

Beran, R. (1987): "Prepivoting to Reduce Level Error of Confidence Sets," *Biometrika*, 74, 457-468.

Beran, R. (1988): "Prepivoting Tests Statistics, A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687-697.

Beran, R. (1993): "Probability-centred Prediction Regions," *Annals of Statistics*, 21, 1967-1981.

Beran, R. and P.W. Millar (1986): "Confidence Sets for a Multivariate Distributions," *Annals of Statistics*, 14, 431-443.

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.

Bhargava, A. (1986): "On the Theory of Testing for Unit Root in Observed Time Series," *Review of Economic Studies*, 53, 369-384.

Birnbaum, A. (1961): "A Unified Theory of Estimation I," *Annals of Mathematical Statistics*, 32, 112-135.

Birnbaum, A. (1964): "Median Unbiased Estimators," *Bulletin of Mathematical Statistics*, 11, 25-34.

Bjerkholt, O. (1995): *Foundations of Modern Econometrics: The Selected Essays of Raynar Frisch, Vol. 1*, Edward Elgar, Aldershot.

Bobkoski, M.J. (1983): "Hypothesis Testing in Nonstationary Time Series," Unpublished Ph.D Thesis, University of Wisconsin.

Boldin, M.V. (1994): "On Median Estimates and Tests in Autoregressive Models," *Mathematical Methods of Statistics*, 3, 114-129.

Bose, A. (1995): "Estimating the Asymptotic Dispersion of the L1 Median," *Annals of the Institute of Statistical Mathematics*, 47, 267-271.

Box, M.J. (1971): "Bias in Nonlinear Estimation," *Journal of the Royal Statistical Society*, 33, 171-190.

Brooks, R.D. (1993): "Alternative Point Optimal Tests for Regression Coefficient Stability," *Journal of Econometrics*, 57, 365-376.

Brooks, R.D. and M.L. King (1994): "Hypothesis Testing of Varying Coefficient Regression Models: Procedures and Applications," *Pakistan Journal of Statistics*, 10, 301-357.

Brown, B.M. (1983): "Statistical Uses of Spatial Median," *Journal of the Royal Statistical Society B*, 45, 25-30.

Brown, G.W. (1947): "On Small-sample Estimation," *Annals of Mathematical Statistics*, 18, 582-585.

Brown, L.D., A. Cohen and W.E. Strawderman (1976): "A Complete Class Theorem for Strict Monotone Likelihood Ratio With Applications," *Annals of Statistics*, 4, 712-722.

Burguete, J.F., A.R. Gallant and G. Souza (1982): "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Review*, 1, 151-190.

Cambell, J.R. and B.E. Honore (1993): "Median Unbiasedness of Estimators of Panel Data Censored Regression Models," *Journal of Applied Econometrics*, 9, 246-256.

Cambell J.Y and N.G. Mankiw (1987): "Permanent and Transitory Components in Macroeconomic Fluctuations," *American Economic Review*, 87, 111-117.

Carpenter, J. (1999): "Test Inversion Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society B*, 61, 159-172.

Cavanagh, C.L. (1985): "Roots Local to Unity," mimeo, Harvard University.

Chan, N.H. (1988): "The Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association*, 83, 857-862.

Chan, N.H. and C.Z. Wei (1987): "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367-401.

Chan, W.S., S.H. Cheung, and K.H. Wu (1999): "Exact Joint Forecast Regions for Vector Autoregressive Models," *Journal of Applied Statistics*, 26, 35-44.

Chen, D. and M.L. King (1998): "Calibrating Estimating Equations for Median-unbiased Estimators," in *Conference Proceedings of the Econometric Society Australasian Meeting: Econometric Theory*, Sydney, 457-483.

Chen, D. and M.L. King (1999): "Inverting POI Tests for Median-Unbiased Estimators," in: L. Nailon, ed., *Conference Proceedings of the Fifth Annual Doctoral Research Conference*, Mt. Eliza, 31-41.

Chen, D. and M.L. King (2000): "Approximately Median-unbiased Estimation of the Dynamic Linear Regression Model," in: J. Nairn, ed., *Conference Proceedings of the Sixth Doctoral Research Conference*, Ballarat, 68-86.

Chen, D. and M.L. King (2001): "Some General Methods for Constructing Median-unbiased Estimators," Paper presented at the departmental seminar, Department of Econometrics and Business Statistics, Monash University, Melbourne.

Chipman, J.S., L. Hurwitz, M.K. Richter and H.F. Sonnenschein (1971): *Preferences, Utility and Demand: A Minnesota Symposium*, Harcourt Brace Jovanovich, New York.

Chu, J.T. (1956): "On the Distribution of the Sample Median," *Biometrika*, 43, 112-116.

Cochrane, D., and G.H. Orcutt (1949):"Applications of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", *Journal of American Statistical Association*, 44, 32-61.

Cochrane, J.H. (1988): "How Big Is the Random Walk in GNP?" *Journal of Political Economy*, 96, 893-920.

Cook, R.D., C.L. Tsai and B.C. Wei (1986): "Bias in Nonlinear Regression," *Biometrika*, 73, 615-623.

Cordeiro, G.M. and R. Klein (1994): "Bias Correction in ARMA Models," *Statistics and Probability Letters*, 19, 169-176.

Cordeiro, G.M., and P. McCullagh (1991): "Bias Correction in the Generalized Linear Models," *Journal of the Royal Statistical Society B*, 53, 629-643.

Cox, D.R. and D.V. Hinkley (1974): *Theoretical Statistics*, Chapman and Hall, London.

Cox, D.R. and N. Reid (1987): "Parameter Orthogonality and Approximate Conditional Inference," *Journal of the Royal Statistical Society B*, 49, 1-39.

Cox, D.R. and N. Reid (1993): "A Note on the Calculation of Adjusted Profile Likelihood," *Journal of the Royal Statistical Society B*, 55, 467-471.

Cribari-Neto, F. and G.M. Cordeiro (1996): "On Bartlett and Bartlett-type Corrections," *Econometric Reviews*, 15, 339-367.

Crowder, M. (1987): "On Linear and Quadratic Estimating Functions," *Biometrika*, 74, 591-597.

Cruddas, A.M., N. Reid and D.R. Cox (1989): "A Time Series Illustration of Approximate Conditional Inference," *Biometrika*, 76, 231-237.

Darnell, A. (1984): "Economic Statistics and Econometrics," in: J. Creedy and D.P. O'Brien, eds., *Economic Analysis in Historical Perspective*, Butterworth, London, 152-185.

Darnell, A. (1994): *The History of Econometrics, Vol. 1*, Edward Elgar, Aldershot.

Davidson A.C. and D.V. Hinkley (1996): *Bootstrap Methods and Their Applications*, Cambridge University Press, New York.

DeJong, D.N., J.C. Nankervis, N.E. Savin, and C.H. Whiteman (1992a): "Integration versus Trend-stationarity in Time Series," *Econometrica*, 60, 423-433.

DeJong, D.N., J.C. Nankervis, N.E. Savin, and C.H. Whiteman (1992b): "The Power Problems of Unit Root Statistics for Autoregressive Time Series with a Unit Root," *Journal of Econometrics*, 5[illegible], 323-343.

Dempster, A.P., N.M. Laird and D.B. Rubin (1977): "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society B*, 39, 1-38.

DiCiccio, T.J. and S.E. Stern (1993): "An Adjustment to Profile Likelihood Based on Observed Information," Technical Report, Department of Statistics, Stanford University.

Dickey, D.A. and W.A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

Dickey, D.A. and W.A. Fuller (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057-1072.

Dielbold, F.X. (2001): "Econometrics: Retrospect and prospect," *Journal Of Econometrics*, 100, 73-75.

Dielman, T.E. (1985): "Regression Forecasts When Disturbances are Autocorrelated," *Journal of Forecasting*, 4, 263-271.

Donoho, D.L. (1982): "Breakdown Properties of Multivariate Location Estimators," unpublished Ph.D thesis, Harvard University.

Dufour, J.M. (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475-494.

Dufour, J.M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365-1387.

Dufour, J.M and M.L. King (1991): "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regression with Stationary or Nonstationary AR(1) Errors," *Journal of Econometrics*, 47, 115-143.

Dufour, J.M. and J.F. Kiviet (1998), "Exact Inference Methods for First Order Autoregressive Distributed Lag Models', *Econometrica*, 66, 79-104.

Durbin, J. (1960): "The Fitting of Time-Series Models", *Review of the International Statistical Institute*, 28, 233-243.

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, 159-177.

Dwass, M. (1957): "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28, 181-187.

Efron, B. (1975): "Defining the Curvature of a Statistical Problem," *Annals of Statistics*, 3, 1189-1242.

Efron, B. (1979): "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

Efron, B. (1982): "Maximum Likelihood and Decision Theory," *Annals of Statistics*, 10, 340-356.

Efron, B. (1985): "Bootstrap Confidence Intervals for a Class of Parametric Problems," *Biometrika*, 72, 45-58.

Efron, B. (1987): "Better Bootstrap Confidence Intervals," *Journal of the American Statistical Association*, 82, 171-200.

Efron, B. and R.J. Tibshirani (1986): "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science*, 1, 54-96.

Efron, B. and R.J. Tibshirani (1993): *An Introduction to the Bootstrap*, New York, Chapman and Hall.

Einsenhart, C. (1949): "Probability Centre Lines for Standard Deviation and Range Charts," *Industrial Quality Control*, 6, 24-26.

Einsenhart, C. and C.S. Martin (1948): "The Relative Frequencies with which Certain Estimators of the Standard Deviation of a Normal Population Tend to Underestimate its Value," *Annals of Mathematical Statistics*, 19, 600.

Elliot, G. (1999): "Efficient Tests for a Unit Root When the Initial Observation is Drawn from its Unconditional Distribution," *International Economic Review*, 40, 767-783.

Elliot, G., T.J. Rothenberg and J.H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813-836.

Enders, W. and B. Falk (1998): "Threshold-autoregressive, Median-Unbiased, and Cointegration Tests of Purchasing Power Parity," *International Journal of Forecasting*, 14, 171-186.

Evans, G.B.A. and N.E. Savin (1981): "Testing for Unit Roots: 1," *Econometrica*, 49, 753-779.

Evans, G.B.A. and N.E. Savin (1984): "Testing for Unit Roots: 2," *Econometrica*, 52, 1241-1269.

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.

Fair, R.C. (1996): "Computing Median Unbiased Estimates in Macroeconometric Models," *Journal of Applied Econometrics*, 11, 431-435.

Ferrari, S.L.P., D.A. Botter, G.M. Cordeiro and F. Cribari-Neto (1996): "Second and Third Order Bias Reduction for One-parameter Family Models," *Economics Letters*, 30, 339-345.

Ferrari, S.L.P. and F. Cribari-Neto (1993): "On the Corrections to the Wald Test of Nonlinear Restrictions," *Economics Letters*, 42, 321-326.

Ferrari, S.L.P. and F. Cribari-Neto (1998): "On Bootstrap and Analytical Bias Corrections," *Economics Letters*, 58, 7-15.

Ferreira, P.E. (1982): "Multiparametric Estimating Equations, " *Annals of Institute of Mathematical Statistics*, 34, 423-431.

Fiebig, D.G. (1985): "Evaluating Estimators without Moments," *Review of Economics and Statistics*, 529-534.

Fiebig, D.G., M. McAleer and R. Bartels (1992): "Properties of Ordinary Least Squares Estimators in Regression Models with Nonspherical Disturbances," *Journal of Econometrics*, 54, 321-334

Firth, D. (1993): "Bias Reduction of Maximum Likelihood Estimates," *Biometrika*, 80, 27-38.

Fisher, N. (1985): "Spherical Medians," *Journal of the Royal Statistical Society B*, 47, 342-348.

Fisher, R.A. (1925): "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, 22, 700-725.

Flavin, M.A. (1981): "The Adjustment of Consumption to Changing Expectations about Future Income," *Journal of Political Economy*, 89, 974-1009.

Franzini, L. and A.C. Harvey (1983): "Testing for Deterministic Trend and Seasonal Components in Time Series Models," *Biometrika*, 70, 673-682.

Fraser, D.A.S. (1956): "Sufficient Statistics with Nuisance Parameters," *Annals of Mathematical Statistics*, 27, 838-842.

Frisch, R. (1933): "Editorial Note," *Econometrica*, 1, 1-4.

Froyen, R.T. and R.N. Waud (1984): "The Changing Relationship Between Aggregate Price and Output: The British Experience," *Economica*, 51, 53-67.

Fuller, W.A. (1996): *Introduction to Statistical Time Series*, John Wiley and Sons, New York.

Fuller, W.A. and D.P. Hasza (1981): "Properties of Predictors or Autoregressive Time Series," *Journal of the American Statistical Association*, 76, 155-161.

Geary, P.T. and J. Kennan (1982): "The Employment Real Wage Relationship: An International Study," *Journal of Political Economy*, 90, 854-871.

Ghosh, J.K. and R. Mukerjee (1994): "Adjusted Versus Conditional Likelihood: Power Properties and Bartlett-type Adjustment," *Journal of Royal Statistical Society B*, 56, 185-188.

Ghosh, M. and P.K. Sen (1989): "Median Unbiasedness and Pitman Closeness," *Journal of the American Statistical Association*, 84, 1089-1091.

Girshick, M.A., F. Mosteller and L.J. Savage (1946): "Unbiased Estimates for Certain Binomial Sampling Problems with Applications," *Annals of Mathematical Statistics*, 17, 13-23.

Glahe, F.R. and J.G. Hunt (1970): "Small Sample Properties of Simultaneous Equation Least Absolute Estimators Vis-à-vis Least Squares Estimators," *Econometrica*, 38, 742-753.

Godambe, V.P. (1960): "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31, 1208-1211.

Godambe, V.P. (1976): "Conditional Likelihood and Unconditional Optimum Estimating Equations," *Biometrika*, 63, 277-284.

Godambe, V.P. (1980): "On Sufficiency and Ancillary in the Presence of a Nuisance Parameter," *Biometrika*, 67, 155-162.

Godambe, V.P. (1984): "On Ancillarity and Fisher Information in the Presence of a Nuisance Parameter," *Biometrika*, 71, 626-629.

Godambe, V.P. (1985): "The Foundations of Finite Sample Estimation in Stochastic Processes," *Biometrika*, 72, 419-428.

Godambe, V.P. (1997): "Estimating Functions: A Synthesis of Least Squares and Maximum Likelihood Methods," in: V.P. Godambe, ed., *Selected Proceedings of the Symposium on Estimating Equations*, Institute of Mathematical Statistics, Hayward, 5-15.

Godambe, V.P. and M.E. Thompson (1974): "Estimating Equations in the Presence of Nuisance Parameters," *The Annals of Statistics*, 2, 568-571.

Godambe, V.P. and M.E. Thompson (1984): "Robust Estimation through Estimating Equations," *Biometrika*, 71, 115-125.

Godfrey, L.G. (1988): *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches.* Cambridge University Press, Cambridge.

Goh, K.L. (1998): "Some Solutions to Small-Sample Problems of Wald Tests in Econometrics," Unpublished Ph.D Thesis, Monash University.

Goh, K.L. and M.L. King (1999): "A Correction for Local Biasedness of the Wald and Null Wald Tests," *Oxford Bulletin of Economics and Statistics*, 61, 435-451.

Goh, K.L. and M.L. King (2000): "A Solution to Non-monotonic Power of the Wald Test in Non-linear Models', *Pakistan Journal of Statistics*, 16, 195-205.

Goldberger, A.S. (1962): "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association*, 57, 369-375.

Gonsalez, M., J.M. Sanchez and J. Romo (1994): "The Bootstrap: A Review," *Computational Statistics*, 9, 165-205.

Gospodinov, N. (1999): "Median Unbiased Forecasts for Highly Persistent Autoregressive Processes," Unpublished monograph, Boston College, Chestnut Hill.

Granger, C.W.J. (2001): "Macroeconometrics -- Past and Future," *Journal of Econometrics*, 100, 17-19.

Green, P.J. (1981): "Peeling Bivariate Data," in: V. Barnett, ed., *Interpreting Multivariate Data*, Wiley, New York, 3-20.

Grigoletto, M. (1998): "Bootstrap Prediction Intervals for Autoregressions: Some Alternatives," *International Journal of Forecasting*, 14, 447-456.

Griliches, Z. (1961): "A Note on the Serial Correlation Bias in Estimates of Distributed Lags," *Econometrica*, 29, 65-73.

Grose, S. (1998), "Marginal Likelihood Methods in Econometrics," Unpublished Ph.D. Thesis, Monash University.

Grubb, D. and J. Symons (1987): "Bias in Regression with a Lagged Dependent Variable," *Econometric Theory*, 3, 371-386.

Haavelmo, T. (1944): "The Probability Approach in Econometrics," *Econometrica*, 12, 1-118.

Haldane, J.B.S. (1948): "Note on the Median of a Multivariate Distribution," *Biometrika*, 25, 414-415.

Hall, P. (1987): "On the Bootstrap and Likelihood-based Confidence Regions," *Biometrika*, 74, 481-493.

Hall, P. (1988): "Theoretical Comparisons of Bootstrap Confidence Intervals," *Annals of Statistics*, 16, 927-985.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansions*. New York, Springer-Verlag.

Hall, P. (1994): "Methodology and Theory of the Bootstrap," in: R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Vol. IV*, Elsevier, Amsterdam, 2341-2381.

Hall, P. and M.A. Martin (1988): "On Bootstrap Resampling and Iteration," *Biometrika*, 75, 661-671.

Hall, R.E. (1978): "Stochastic Implications of the Life-Cycle Permanent Income Hypothesis," *Journal of Political Economy*, 86, 971-1007.

Halmos, P.R. (1946): "The Theory of Unbiased Estimator," *Annals of Mathematical Statistics*, 17, 34-43.

Hansen, B. (1999): "Grid Bootstrap," *Reviews of Economics and Statistics*, 81, 594-607.

Harvey, A.C. (1985): "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 3, 216-227.

Hatanaka, M. (1974): "An Efficient Two-step Estimator for the Dynamic Adjustment Model with Autoregressive Errors," *Journal of Econometrics*, 2, 199-220.

Hauck, Jr., W.W. and A. Donner (1977): "Wald Test as Applied to Hypotheses in Logit Analysis," *Journal of the American Statistical Association*, 72, 851-853.

Hayakawa, T. (1975): "The Likelihood Ratio Criterion for a Composite Hypothesis Under a Local Alternative," *Biometrika*, 62, 451-460.

Hayakawa, T. and M.L. Puri (1985): "Asymptotic Expansions of the Distributions of Some Test Statistics," *Annals of the Institute of Statistical Mathematics (Part A)*, 29, 95-108.

Heckman, J.J. (2001): "Econometrics and Empirical Economics," *Journal of Econometrics*, 100, 3-5.

Hendry, D.F. (1995): *Dynamic Econometrics*, Oxford University Press, New York.

Hendry, D.F. and G. Mizon (1978): "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Bank of England Demand-for-Money Function," *Economic Journal*, 88, 549-563.

Henshaw, R.C. (1968): "Testing Single-equation Least Squares Regression Models for Autocorrelated Disturbances," *Econometrica*, 34, 646-660.

Hillier, G.H. (1987): "Classes of Similar Regions and Their Power Properties for Some Econometric Testing Problems," *Econometric Theory*, 3, 1-44.

Hirji, K.F., A.A. Tsiantis and C.R. Mehta (1989): "Median Unbiased Estimation for Binary Data," *The American Statistician*, 43, 7-11.

Honda, Y. (1989): "On the Optimality of Some Tests of the Error Covariance Matrix in the Linear Regression Model," *Journal of the Royal Statistical Society B*, 51, 71-79.

Honore, B.E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-565.

Hoque, A., J.R. Magnus and B. Pesaran (1988): "The Exact Multi-Period Mean-Square Forecast Error for the First-Order Autoregressive Model," *Journal of Econometrics*, 39, 327-346.

Horowitz, J.L. (1994): "Bootstrap-based Critical Values for the Information Matrix Test," *Econometrica*, 61, 395-411.

Horowitz, J.L. and N.E. Savin (2000): "Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach," *Journal of Econometrics*, 95, 375-389.

Hsiao, C. (1997): "Cointegration and Dynamic Simultaneous Equations Model," *Econometrica*, 647-670.

Huber, P.J. (1981): *Robust Statistics*. Wiley, New York.

Huber, P.J. (1987): "The Place of the L1-norm in Robust Estimation," in: Y. Doldge, ed., *Statistical Data Analysis Based on the L1-norm and Related Methods*, Elsevier, Amsterdam, 23-34.

Hurwicz, L. (1950a): "Least-Squares Bias in Time Series," in: T.C. Koopmans, ed., *Statistical Inference in Dynamic Economic Models*, Wiley, New York, 365-383.

Hurwicz, L. (1950): "Prediction and Least Squares," in: T.C. Koopmans, ed., *Statistical Inference in Dynamic Economic Models*, Wiley, New York, 266-300.

Imhof, J.P. (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. (1987): "Bias in the Ordinary Least Squares Estimator in the Dynamic Linear Regression Model with Autocorrelated Errors," Working Paper No. 10/87, Department of Econometrics and Operations Research, Monash University.

Jensen, D.R. (1979): "Linear Models without Moments," *Biometrika*, 66, 611-7.

Joekel, H.K. (1986): "Finite Sample Properties and Asymptotic Efficiency of Monte Carlo Tests," *Annals of Statistics*, 14, 336-347.

Kabaila, P. (1993a): "Some Properties of Profile Bootstrap Confidence Intervals," *Australian Journal of Statistics*, 35, 205-214.

Kabaila, P. (1993b): "On Bootstrap Predictive Inference for Autoregressive Processes," *Journal of Time Series Analysis*, 14, 473-484.

Kalbfleisch, J.D. and D.A. Sprott (1970): "Application of Likelihood Methods to Models Involving Large Numbers of Parameters," *Journal of Royal Statistical Society, B*, 32, 175-208.

Kalbfleisch, J.D. and D.A. Sprott (1973): "Marginal and Conditional Likelihoods," *Sankhya A*, 35, 311-328.

Keating, J.P. and R.C. Gupta (1984): "Simultaneous Comparisons of Scale Parameters," *Sankhya B*, 46, 275-280.

Keating, J.P. and R.L. Mason (1985): "Pitman's Measure of Closeness," *Sankhya B*, 47, 22-30.

Kemp, G.C.R. (1999): "The Behaviour of Forecast Errors from a Nearly Integrated AR(1) Model as both Sample Size and Forecast Horizon Become Large," *Econometric Theory*, 15, 238-256.

Kendall, M.G. (1954): "Note on Bias in the Estimation of Autocorrelation," *Biometrika*, 41, 403-404.

Kendall, M.G. and A. Stuart (1967): *The Advanced Theory of Statistics*, 2, 2nd Ed. Charles Griffin, London.

Kilian, L. (1998): "Small-sample Confidence Intervals for Impulse Responsive Functions," *The Review of Economics and Statistics*, 80, 218-230.

Kim, J.H. (2001): "Bootstrap-after-bootstrap Prediction Intervals for Autoregressive Models," *Journal of Business and Economic Statistics*, 19, 117-128.

King, M.L. (1980): "Robust Tests for Spherical Symmetry and Their Applications to Least Squares Regression," *Annals of Statistics*, 8, 1265-1271.

King, M.L. (1983): "Testing for Autoregressive against Moving Average Errors in the Linear Regression Model," *Journal of Econometrics*, 35-51.

King, M.L. (1984): "A New Test for Fourth-order Autoregressive Disturbances," *Journal of Econometrics*, 24, 269-277.

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*, London: Routledge and Kegan Paul, 19-73.

King, M.L. (1987b): "Towards a Theory of Point Optimal Testing," *Econometric Review*, 6, 169-218.

King. M.L. (1987c): "An Alternative Test for Regression Coefficient Stability," *Review of Economics and Statistics*, 69, 379-381.

King, M.L. (1996): "Hypothesis Testing in the Presence of Nuisance Parameters," *Journal of Statistical Planning and Inference*, 50, 103-120.

King, M.L. and D.E.A. Giles (1984): "Autocorrelation Pre-testing in the Linear Model: Estimation, Testing and Prediction," *Journal of Econometrics*, 25. 35-48.

King, M.L., and G.H. Hillier (1985): "Locally Best Invariant Tests for the Error Covariance Matrix of the Linear Regression Model," *Journal of the Royal Statistical Society B*, 47, 98-102.

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.

Kiviet, J.F. and J.M. Dufour (1997): "Exact Tests in Single Equation Autoregressive Distributed Lag Models," *Journal of Econometrics*, 80, 325-353.

Kiviet, J.F. and G.D.A. Phillips (1990): "Exact Similar Tests for the Root of a First-order Autoregressive Regression Model," University of Amsterdam, AE Report 12/90, paper presented at Econometric Society World Congress 1990, Barcelona, Spain.

Kiviet, J.F. and G.D.A. Phillips (1992): "Exact Similar Tests for Unit Root and Cointegration," *Oxford Bulletin of Economics and Statistics*, 54, 349-367.

Kiviet, J.F. and G.D.A. Phillips (1993): "Alternative Bias Approximations in Regressions with a Lagged Dependent Variable," *Econometric Theory*, 9, 62-80.

Kiviet, J.F. and G.D.A. Phillips (1994): "Bias Assessment and Reduction in the Linear Error-Correction Models," *Journal of Econometrics*, 63, 215-243.

Kiviet, J.F. and G.D.A. Phillips (1996): "Higher-order Asymptotic Expansions of the Least Squares Estimation Bias in the First-order Dynamic Regression Models," University of Amsterdam, AE Report 1/96.

Kobayashi, M. (1985): "Comparison of Efficiencies of Several Estimators for Linear Regressions with Autocorrelated Errors," *Journal of the American Statistical Association*, 80, 951-953.

Koenker, R. and G. Bassett (1982): "Tests of Hypotheses and L1 Estimation," *Econometrica*, 50, 1577-1583.

Kramer, W. (1985): "The Power of the Durbin-Watson Test for Regression without an Intercept," *Journal of Econometrics*, 28, 363-370.

Kramer, W. and H. Zeisel (1990): "Finite Sample Power of Linear Regression Autocorrelation Tests," *Journal of Econometrics*, 43, 363-372.

Krishnakumar, J. (2001): "A Short Comment on the JE Open Forum Essays," *Journal of Econometrics*, 100, 77-78.

Laplace, P.S. (1774): "Memoire sur la Probabilite des Causes par les Evenements," in *Oeuvres Completes de Laplace 8*, Gauthier-Villars, Paris, 27-65.

Laskar, M.R. and M.L. King (1997): "Modified Wald Test for Regression Disturbances," *Economic Letters*, 56, 5-11.

Laskar, M.R., and M.L. King (1998): "Estimation and Testing of Regression Disturbances Based on Modified Likelihood Functions," *Journal of Statistical Planning and Inference*, 71, 75-92.

Latif, A. and M.L. King (1993): "Linear Regression Forecasting in the Presence of AR(1) Disturbances," *Journal of Forecasting*, 12, 513-524.

Le Breton, A., and D.T. Pham (1989): "On the Bias of the Least Squares Estimator for the First Order Autoregressive Process," *Annals of the Institute of Statistical Mathematics*, 41, 555-563.

Lehmann, E.L. (1951): "A General Concept of Unbiasedness," *Annals of Mathematical Statistics*, 22, 587-592.

Lehmann, E.L. (1959): *Testing Statistical Hypotheses*, John Wiley and Sons, New York.

Lehmann, E.L. (1983): *Theory of Point Estimation*, John Wiley and Sons, New York.

Lele, S. (1991): "Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes," *Journal of the Royal Statistical Society B*, 53, 253-267.

Levenbach, H. (1972): "Estimation of Autoregressive Parameters from a Marginal Likelihood Function," *Biometrika*, 59, 61-71.

Levin, B. and F. Kong (1990): "Bartlett's Bias Correction to the Profile Score Function is a Saddlepoint Correction," *Biometrika*, 77, 219-221.

Liang, K.Y. (1987): "Estimating Functions and Conditional Likelihood Function," *Biometrika*, 74, 695-702.

Lieberman, O. (1998): "From Unbiased Linear Estimating Equations to Unbiased Estimators," *Biometrika*, 85, 244-250.

Liu, R.Y. (1988): "On a Notion of Simplicial Depth," *Proceedings of the National Academy of Science of USA*, 85, 1732-1734.

Liu, R.Y. (1990): "On a Notion of Data Depth Based on Random Simplices," *Annals of Statistics*, 18, 405-414.

Lopuhaa, H.P. and P.J. Rousseeuw (1988): "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices," Technical

Report 87-14, Faculty of Mathematics and Informatics, Delft University of Technology.

Maasoumi, E. (2001): "On the Relevance of First-order Asymptotic Theory to Economics," *Journal of Econometrics*, 100, 83-86.

MacKinnon, J.G. and A.A. Smith (1998): "Approximate Bias Correction in Econometrics," *Journal of Econometrics*, 85, 205-230.

Maddala, G.S. and I.M. Kim (1998): *Unit Roots, Cointegration and Structural Changes*, Cambridge University Press, Cambridge.

Maekawa, K. (1983): "An Approximation to the Distribution of the Least Squares Estimator in an Autoregressive Model with Exogenous Variables," *Econometrica*, 51, 229-239.

Maekawa, K. (1987): "Finite Sample Properties of Several Predictors from an Autoregressive Model," *Econometric Theory*, 3, 359-370.

Magdalinos, M.A. (1990): "The Classical Principles of Testing Using Instrumental Variables Estimates," *Journal of Econometrics*, 44, 241-279.

Magee, L. (1989): "An Edgeworth Test Size Correction for the Linear Model with AR(1) Errors," *Econometrica*, 57, 661-674.

Magee, L., A. Ullah and V.K. Srivastava (1987): "Efficiency of Estimators in the Regression Model with First-order Autocorrelated Errors," in: M.L. King and D.E.A. Giles, eds., *Specification Analysis in the Linear Model*, London: Routledge and Kegan Paul, 81-98.

Magnus, J.R. and B. Pesaran (1989): "The Exact Multi-period Mean-square Forecast Error for the First-order Autoregressive Model with an Intercept," *Journal of Econometrics*, 42, 157-179.

Magnus, J.R. and B. Pesaran (1991): "The Bias of Forecasts from a First-order Autoregression," *Econometric Theory*, 7, 222-235.

Mak, T.K. (1993): "Solving Non-linear Estimation Equations," *Journal of the Royal Statistical Society B*, 55, 945-955.

Mantel, N. (1987): "Understanding Wald's Test for Exponential Families," *American Statistician*, 41, 147-148.

Mardia, K.V., H.R. Southworth and C.C. Taylor (1999): "On Bias in Maximum Likelihood Estimators," *Journal of Statistical Planning and Inference*, 76, 31-39.

Mahmood, M. (2000), "Exploring Estimating Equations to Improve Estimation Procedures in the Linear Model," Unpublished Ph.D Thesis, Monash University.

Marriot, F.H.C. and J.A. Pope (1954): "Bias in the Estimation of Autocorrelations," *Biometrika*, 41, 390-402.

Masarotto, G. (1990): "Bootstrap Prediction Intervals for Autoregressions," *International Journal of Forecasting*, 6, 229-239.

McCullagh, P. and R. Tibshirani (1990): "A Simple Method for the Adjustment of Profile Likelihoods," *Journal of Royal Statistical Society, B*, 52, 325-344.

McKean, J.W. and R.M. Shrader (1984): "A Comparison of Methods for Studentizing the Sample Median," *Communications in Statistics: Theory*, B6, 751-773.

McManus, D.A. and J.C. Nankervis and N.E. Savin (1994): "Multiple Optima and Asymptotic Approximations in the Partial Adjustment Model," *Journal of Econometrics*, 62, 91-128.

Michel, R. (1972): "Bounds for the Efficiency of Approximately Median Unbiased Estimates," *Metrika*, 18, 90-96.

Milan, L. (1984): "Testing for the Hildreth-Houck Random Coefficient Model," Unpublished M.Ec. thesis, Monash University.

Nankervis, J.C. and N.E. Savin (1985): "Testing the Autoregressive Parameter with the *t* Statistic," *Journal of Econometrics*, 27, 143-161.

Nankervis, J.C. and N.E. Savin (1987): "Finite Sample Distribution of *t* and *F* Statistics in an AR(1) Model with an Exogenous Variable," *Econometric Theory*, 3, 387-408.

Nankervis, J.C. and N.E. Savin (1988a): "The Exact Moments of the Least Squares Estimator for the Autoregressive Model: Corrections and Extensions," *Journal of Econometrics*, 37, 381-388.

Nankervis, J.C. and N.E. Savin (1988b): "The Student *t* Approximation in a Stationary First Order Autoregressive Model," *Econometrica*, 56, 119-145.

Nankervis, J.C. and N.E. Savin (1996): "The Level and Power of the Bootstrap *t* Test in the AR(1) Model with Trend," *Journal of Business and Economic Statistics*, 14, 161-168.

Narula, S.C. and J.F. Wellington (1982): "The Minimum Sum of Absolute Errors Regression: A State of the Art Survey," *International Statistical Review*, 50, 317-326.

Nelson, C.R., and C.I. Plosser (1982): "Trends and Random Walks in Macroeconomics Time Series," *Journal of Monetary Economics*, 10, 139-162.

Nelson, F.D. and N.E. Savin (1988): "The Non-Monotonicity of the Power Function of the Wald Test in Nonlinear Models," Working Paper Series No. 88-7, Department of Economics, University of Iowa.

Nelson, F.D. and N.E. Savin (1990): "The Danger of Extrapolating Asymptotic Local Power," *Econometrica*, 58, 977-981.

Nyblom, J. (1986): "Testing for Deterministic Linear Trend in Time Series," *Journal of the American Statistical Association*, 81, 545-549.

Nyblom, J. (1989): "Testing for the Constancy of Parameter Over Time," *Journal of the American Statistical Association*, 84, 223-230.

Oja, H. (1983): "Descriptive Statistics for Multivariate Distributions," *Statistics and Probability Letters*, 1, 323-333.

Orcutt, G.H. and H.S. Winokur (1969): "First Order Autoregression: Inference, Estimation, and Prediction," *Econometrica*, 37, 1-14.

Oya, K. (1997): "Wald, LM and LR Test Statistics of Linear Hypotheses in Structural Equation Model," *Econometric Reviews*, 16, 157-178.

Pantula, S.G., G. Gonzalez-Farias and W.A. Fuller (1994): "A Comparison of Unit-root Test Criteria," *Journal of Business and Economics Statistics*, 12, 449-459.

Park, R.E. and B.M. Mitchell (1980): "Estimating the Autocorrelated Error Model with Trended Data," *Journal of Econometrics*, 13, 185-201.

Patterson, K.D. (2000): "Bias Reduction in Autoregressive Models," *Economics Letters*, 68, 135-141.

Peers, H.W. (1971): "Likelihood Ratio and Associated Test Criteria," *Biometrika*, 58, 577-587.

Pfanzagl, J. (1970): "On the Asymptotic Efficiency of Median Unbiased Estimates," *Annals of Mathematical Statistics*, 41, 1500-1509.

Pfanzagl, J. (1971): "On Median Unbiased Estimates," *Metrika*, 17, 82-91.

Pfanzagl, J. (1979): "On Optimal Median Unbiased Estimators in the Presence of Nuisance Parameters," *Annals of Statistics*, 7, 187-193.

Phillips, P.C.B. (1979): "The Sampling Distribution of Forecasts from a First-Order Autoregression," *Journal of Econometrics*, 9, 241-261.

Phillips, P.C.B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277-302.

Phillips, P.C.B. (1988): "The ET Interview: Professor James Durbin," *Econometric Theory*, 4, 125-157.

Phillips, P.C.B. (2001): "Trending Time Series and Macroeconomic Activity: Some Present and Future Challenges," *Journal of Econometrics*, 100, 21-27.

Phillips, P.C.B. and J.Y. Park (1988): "On the Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1065-1083.

Phillips, P.C.B. and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335-346.

Phillips, P.C.B. and Z. Xiao (1998): "A Primer on Unit Root Testing," *Journal of Economic Surveys*, 12, 423-469.

Pitman, E.J.G. (1937): "The 'Closest' Estimates of Statistical Parameters," *Proceedings of the Cambridge Philosophical Society*, 33, 212-222.

Podivinsky, J.M. and M.L. King (2000): "The Exact Power Envelope of Tests for a Unit Root," University of Southampton Discussion Papers in Economics and Econometrics, No. 0026.

Powell, J.L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-325.

Prais, S.J., and C.B. Winsten (1954):" Trend Estimators and Serial Correlation," unpublished Cowles Commission discussion paper: Stat. No. 383, Chicago.

Quandt, R.E. (1960): "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324-330.

Quenouille, M.H. (1949): "Notes on Bias in Estimation," *Biometrika*, 43, 353-360.

Rahman, S. and M.L. King (1997): "Marginal Likelihood Score Based Tests of Regression Disturbances in the Presence of Nuisance Parameters," *Journal of Econometrics*, 82, 63-80.

Rao, C.R. (1988): "Methodology Based on the L1-Norm in Statistical Inference," *Sankhya A*, 50, 289-313.

Rao, C.R. and G.J. Babu (1988): "Joint Asymptotic Distribution of Marginal Quantile Functions in Samples from a Multivariate Population," *Journal of Multivariate Analysis*, 27, 15-23.

Rao, P. and Z. Griliches (1969): "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," *Journal of American Statistical Association*, 64, 253-272.

Rayner, R.K. (1990): "Bootstrapping *p* Values and Power in the First-Order Autoregression: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 8, 251-263.

Read, C.B. (1985): "Median Unbiased Estimator," in: S. Kotz and N.L. Johnson, eds., *Encyclopedia of Statistical Sciences 5*, John Wiley, New York, 424-426.

Ronchetti, E. (1982): "Robust Testing in Linear Models: The Infinitesimal Approach," Ph.D dissertation, Swiss Federal Institute of Technology, Zurich.

Rothenberg, T.J. (1988): "Approximate Power Functions for Some Robust Tests of Regression Coefficients," *Econometrica*, 56, 997-1009.

Rudebusch, G.D. (1992): "Trends and Random Walks in Macroeconomic Time Series: A Re-Examination," *International Economic Review*, 33, 661-680.

Rudebusch, G.D. (1993): "The Uncertain Unit Root in Real GNP," *American Economic Review*, 83, 264-272.

Saikkonen, P. and R. Luukonen (1993): "Point Optimal Test for Testing the Order of Differencing in ARIMA Models," *Econometric Theory*, 9, 343-362.

Sampson, M. (1991): "The Effect of Parameter Uncertainty on Forecast Variance and Confidence Intervals for Unit Root and Trend Stationary Time Series Models," *Journal of Applied Econometrics*, 6, 67-76.

Sargan, J.D. and A. Bhargava (1983): "Testing Residuals from Least Squares Regression of Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153-174.

Savage L.J. (1954): *The Foundations of Statistics*, John Wiley and Sons, New York.

Sawa, T. (1978): "The Exact Moments of the Least Squares Estimator for the Autoregressive Model," *Journal of Econometrics*, 8, 159-172.

Schaefer, R.L. (1983): "Bias Correction in Maximum Likelihood Logistic Regression," *Statistics and Medicine*, 2, 71-78.

Schmidt, P. and P.C.B. Phillips (1992): "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, 257-287.

Seheult, A.H., P.J. Diggle and D.A. Evans (1976): "Discussion of Paper by V. Barnett," *Journal of the Royal Statistical Society A*, 139, 351-352.

Shaman, P., and R.A. Stine (1988): "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association*, 83, 842-848.

Shao, J. and D. Tu (1995): *The Jackknife and Bootstrap*. Springer-Verlag, New York.

Shenton, L.R. and W. L. Johnson (1965): "Moments of a Serial Correlation Coefficient," *Journal of the Royal Statistical Society. Series B*, 27, 308-320.

Shepherd, N.G. (1993): "Maximum Likelihood Estimation of Regression Models with Stochastic Trend Components," *Journal of the American Statistical Association*, 88, 590-595.

Shepherd, N.G. and A.C. Harvey (1990): "On the Probability of Estimating a Deterministic Component in the Local Level Model," *Journal of Time Series Analysis*, 11, 339-347.

Shively, T.S. (1988): "An Exact Test for a Coefficient in a Time Series Regression Model," *Journal of Time Series Analysis*, 9, 81-88.

Shively, T.S., G.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.

Shwert, G.W. (1987): "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," *Journal of Econometrics*, 20, 73-103.

Shwert, G.W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147-160.

Silvapulle, P. and M.L. King (1991): "Testing Moving Average Against Autoregressive Disturbances in the Linear Regression Model," *Journal of Business and Economics Statistics*, 9, 329-335.

Small, C.G. (1990): "A Survey of Multidimensional Median," *International of Statistical Review*, 58, 263-277.

Spitzer, J.J. (1979): "Small-sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors," *Journal of the American Statistical Association*, 74, 41-47.

Spitzer, J.J. and R.T. Baillie (1983): "Small-sample Properties of Predictions from the Regression Model with Autoregressive Errors," *Journal of the American Statistical Association*, 78, 258-263.

So, B.S. and D. W. Shin (1999): "Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals," *Econometric Theory*, 15, 165-176.

So, B.S. and D.W. Shin (2000): "Gaussian Tests for Seasonal Unit Roots Based on Cauchy Estimation and Recursive Mean Adjustments," *Journal of Econometrics*, 99, 107-137.

Stern, S.E. (1997): "A Second-order Adjustment to the Profile Likelihood in the Case of a Multidimensional Parameter of Interest," *Journal of the Royal Statistical Society B*, 59, 653-655.

Stine, R.A. (1985): "Bootstrap Prediction Intervals for Regression," *Journal of the American Statistical Association*, 80, 1026-1031.

Stine, R.A. (1987): "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association*, 82, 1072-1078.

Stock, J.H. (1991):"Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435-459.

Stock, J.H. (1994): "Unit Roots, Structural Breaks and Trends," in: R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics 4*, Elsevier, North Holland, 2739-2481.

Stock, J.H. (2001): "Macro-econometrics," *Journal of Econometrics*, 100, 29-32

Stock, J.H. and M.W. Watson (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of American Statistical Association*, 93, 349-358.

Strasser, H. (1978): "Admissible Representation of Asymptotically Optimal Estimates," *Annals of Statistics*, 6, 867-881.

Stroud, T.W.F. (1971): "On Obtaining Large-sample Tests from Asymptotically Normal Estimators," Annals of Mathematical Statistics, 42, 1412-1424.

Stuart, A. and J.K. Ord (1991): *Kendall's Advanced Theory of Statistics* 2. 5th Ed., Edward Arnold, London.

Sung, N.K. (1988): "Cramer-Rao Analogue for Median-unbiased Estimators," Unpublished Ph. D. dissertation, Iowa State University, Ames, Iowa.

Sung, N.K. (1990): "A Generalized Cramer-Rao Analogue for Median-unbiased Estimators," *Journal of Multivariate Analysis*, 32, 204-212.

Tanaka, K. (1983): "Asymptotic Expansions Associated with the AR(1) Model with Unknown Mean," *Econometrica*, 51, 1221-1231.

Taylor, L.D. and T.A. Wilson (1964): "Three-Pass Least Squares: A Method for Estimating Models with a Lagged Dependent Variable," *The Review of Economics and Statistics*, 46, 329-346.

Thombs, L.A. and W.R. Schucany (1990): "Bootstrap Prediction Intervals for Autoregression," *Journal of the American Statistical Association*, 89, 1303-1313.

Tillman, J.A. (1975), "The Power of the Durbin-Watson Test," *Econometrica*, 43, 959-974.

Tsai, C.L., R.D. Cook and B.C. Wei (1986): "Bias in Nonlinear Regression," *Biometrka*, 73, 615-623.

Tse, Y.K. (1984): "An Empirical Comparison of Small Sample Distributions of Estimators of the First Order Autoregression," *Journal of Statistical Computation and Simulation*, 19, 227-236.

Tukey, J.W. (1975): "Mathematics and the Picturing of Data," in *Proceedings of the International Congress of Mathematicians*, Vancouver, 2, 523-531.

Tunnicliffe-Wilson, G. (1989): "On the Use of Marginal Likelihood in Time Series Model Estimation," *Journal of the Royal Statistical Society B*, 51, 15-27.

Van Der Hart, H.R. (1961): "Some Extensions of the Idea of Bias," *Annals of Mathematical Statistics*, 32, 436-447.

Voinov, V.G. and M.S. Nikulin (1993a): *Unbiased Estimators and Their Applications I*, Kluwer Academic Publishers, Boston

Voinov, V.G. and M.S. Nikulin (1993b): *Unbiased Estimators and Their Applications II*, Kluwer Academic Publishers, Boston

Vinod, H.D. (1993): "Bootstrap Methods: Applications in Econometrics," in: S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics 11*, Elsevier, North Holland, 573-610.

Vinod, H.D. (1997): "Using Godambe-Durbin Estimating Functions in Econometrics," in: V.P. Godambe, ed., *Selected Proceedings of the Symposium on Estimating Equations*, Institute of Mathematical Statistics, Hayward, 215-237.

Weiss, A.A. (1991): "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation," *Econometric Theory*, 7, 46-68.

Wright, J.H. (2000a): "Confidence Intervals for Univariate Impulse Response With a Near Unit Root," *Journal of Business and Economic Statistics*, 18, 368-373.

Wright, J.H. (2000b): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of the American Statistical Association*, 18, 211-223.

Yamamoto, T. (1976): "Asymptotic Mean-Squared Prediction Error for an Autoregressive Model with Estimated Coefficients," *Applied Statistics*, 25, 123-127.

Yamamoto, T. (1979): "On the Prediction Efficiency of the Generalized Least Squares Model with an Estimated Variance Covariance Matrix," *International Economic Review*, 20, 693-705.

Zaman, A. (1981): "Estimators without Moments," *Journal of Econometrics*, 15, 289-298.

Zeisel, H. (1989): "On the Power of the Durbin-Watson Test Under High Autocorrelation," *Communications in Statistics*, 18, 3907-3916.

Zelelinski, Y. J. (1999): "A Median Unbiased Estimator of the Autoregressive Parameter," *Journal of Time Series Analysis*, 21, 362-369.