Location-based estimation of the autoregressive coefficient in ARX(1) models

by

Timothy Kevin Kuria Kamanu

A thesis

submitted in fulfilment of the requirements

for the degree of

Magister Scientiae

in the Department of Statistics, Faculty of Natural

Sciences,

University of the Western Cape.

NOVEMBER 2006

Supervisor:

Prof. Chris Koen.

To mum, dad, my brothers and sisters.

Keywords

1. Time series analysis
2. Unit root tests
3. Near unit root processes
4. Overdifferencing
5. Unbiased estimation of autocorrelation
6. Probability density of AR(1) coefficient
7. Non-stationary AR(1) processes
8. Mode estimation
9. Testing for stationarity versus testing for nonstationarity
10. Box-Jenkins modelling

JEL Classification: C13, C15, C22, C52

Mathematical Subject Classification: 37M05, 37M10, 62F03, 62F10, 62F25, 62M10

Abstract

In recent years, two estimators have been proposed to correct the bias exhibited by the least-squares (LS) estimator of the lagged dependent variable (LDV) coefficient in dynamic regression models when the sample is finite. They have been termed as 'mean-unbiased' and 'median-unbiased' estimators. Relative to other similar procedures in the literature, the two location-based estimators have the advantage that they offer an exact and uniform methodology for LS estimation of the LDV coefficient in a first order autoregressive model with or without exogenous regressors i.e. ARX(1).

However, no attempt has been made to accurately establish and/or compare the statistical properties among these estimators, or relative to those of the LS estimator when the LDV coefficient is restricted to realistic values. Neither has there been an attempt to compare their performance in terms of their mean squared error (MSE) when various forms of the exogenous regressors are considered. Furthermore, only implicit confidence intervals have been given for the 'median-unbiased' estimator. Explicit confidence bounds that are directly usable for inference are not available for either estimator.

In this study a new estimator of the LDV coefficient is proposed; the 'most-probably-unbiased' estimator. Its performance properties vis-a-vis the existing estimators are determined and compared when the parameter space of the LDV coefficient is restricted. In addition, the following new results are established: (1) an explicit computable form for the density of the LS estimator is derived for the first time and an efficient method for its numerical evaluation is proposed; (2) the exact bias, mean, median and mode of the distribution of the LS estimator are determined in three specifications of the ARX(1) model; (3) the exact variance and MSE of LS estimator is determined; (4) the standard error associated with the determination of same quantities when

simulation rather than numerical integration method is used are established and the methods are compared in terms of computational time and effort; (5) an exact method of evaluating the density of the three estimators is described; (6) their exact bias, mean, variance and MSE are determined and analysed; and finally, (7) a method of obtaining the explicit exact confidence intervals from the distribution functions of the estimators is proposed.

The discussion and results show that the estimators are still biased in the usual sense: 'in expectation'. However the bias is substantially reduced compared to that of the LS estimator. The findings are important in the specification of time-series regression models, point and interval estimation, decision theory, and simulation.

Declaration

I declare that *location-based estimation of the autoregressive coefficient in ARX(1) models* is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

TIMOTHY KEVIN KURIA KAMANU *NOVEMBER 2006*

SIGNED: .....

Acknowledgments

I would like to acknowledge and sincerely thank my sponsors, the University of the Western Cape (UWC) and the African Institute for Mathematical Sciences (AIMS).

I would also like to thank my Supervisor Prof. C. Koen for carefully reading my work. Your insight, attention to detail, ideas, comments and corrections have been an invaluable source of information during my research.

Furthermore, I am grateful to the Dean of the Faculty of Science Prof J. B. Donker, the Head of the Statistics Department Prof. R. Blignaut, Dr. T. Kotze, and my former Co-supervisor Prof. D. Kotze[1], for sponsorship to various activities during the course of my study. I learnt a lot from the data-mining course and conference presentations.

Thank you to: Prof. J. F. Kiviet, Dr. D. Wilcox and Ms. C. J. Cook for taking time to respond to my queries; the inter-library loan staff at UWC for their efficiency in providing some references; my colleagues in, and staff of, the Statistics Department for creating a good study environment; all my friends, Davis, Ian and Sophie to mention just a few; and mum and dad - Mary and Samuel, your constant encouragement, especially when I drifted away from my original research title, is highly appreciated.

Last but not least, I thank the Almighty God for working behind my simulations and for the many uncountable blessings.

This thesis was typeset with LaTeX 2_ε[2] by the author.

[1] Thanks again!!

[2] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The style package *timstyle* was used.

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms

ADF Augmented Dickey-Fuller

AR Autoregressive process

ARMA Autoregressive moving average

ARX Autoregressive process with exogenous regressors ($\mathbf{X}$)

CI Exact confidence interval(s)

CIR Cumulative impulse response

Dev A deviation statistic (Euclidean distance)

DS Difference stationary

EDF Empirical density function

EP Epanechnikov kernel

I Integrated

IID Independently and identically distributed

IQR Standardized sample interquartile range

KPSS Kwiatkowski, Phillips, Schmidt and Shin

LDV Lagged dependent variable

LS Least squares

MGF Moment generating function

MISE Mean-integrated-squared error

MLE Maximum likelihood estimator

MSE Mean squared error

MVUE Minimum variance unbiased estimator

PP Phillips and Perron

RMS Root mean squared

RSS Residual sum of squares

SE Standard error

TR Triangular kernel

TS Trend stationary

Var Variance

Notation

α	Coefficient of lagged dependent variable (true parameter)
$\hat{\alpha}_{\mathsf{UE}}^{\mathsf{A}}$	Mean-based estimator
$\hat{\alpha}_{\mathsf{UE}}$	Estimator resulting from the risk-unbiasedness restrictions
$\hat{\alpha}_{\mathsf{UE}}^{\mathsf{B}}$	Median-based estimator
$\hat{\alpha}_{\mathsf{UE}}^{\mathsf{C}}$	Mode-based estimator
$\hat{\alpha}_{\mathsf{LS}}$	LS estimator: model independent notation
$\hat{\alpha}_j$	LS estimator of α in Model j $(j = 1, 2, 3)$, [see (1.4)-(1.6)]
$\hat{\alpha}^*$	LS estimates
$\hat{\alpha}_{ji}^*$	The i^{th} LS **estimate** of α in Model j: see section 3.2
$\bar{\alpha}_j$	Mean statistic of the LS estimator given α in Model j
$\tilde{\alpha}_j$	Median statistic of the LS estimator given α in Model j
$\ddot{\alpha}_j$	Mode statistic of the LS estimator given α in Model j
β	Column vector of unknown coefficients
$b_T(\alpha)$	Bias function given α and sample size T
B_{trc}	Pre-specified bound on the truncation error
δ	Time trend variable in the series representation
$\mathbb{E}$	Expectation operator
Γ	Full coefficient vector [see model formulation in (1.1)]
H	Parameter: number of columns after inclusion of a redundant regressor

$\mathbf{I}$	Identity matrix
y_1	Initial observation of the series $\{y_t\}$
μ	Constant or intercept variable in the series representation
N	Number of replicates
$\mathcal{N}$	Normally distributed
Ω	Restricted parameter space of the LDV coefficient
$\mathbf{1}$	Column vector of ones
r	Ratio of quadratic forms
σ^2	Variance of the innovations
T	Sample size
$\mathbf{M}$	Transformation matrix in quadratic forms
$\mathbf{t}$	Column vector comprising $1, 2, \ldots, T$
$\mathfrak{t}$	Regression statistic
$\{u_t\}$	Disturbance or innovations assumed Gaussian white noise
$\mathbf{u}$	Disturbances or innovations vector: assumed to be Gaussian white noise
$\mathbf{X}$	Full column-rank matrix of observations on fixed exogenous regressors
$\mathbf{y}$	Vector of observations on the dependent variable
$\mathbf{y}_{-1}$	Vector $\mathbf{y}$ lagged one period
$\mathcal{Z}$	Design matrix [see (1.2)]
$\{y_t\}$	Series observations $t = 1, \ldots, T$
$\mathbf{0}$	Column vector of zeros

Preface

A well-known approach to modelling time series is to consider the series as a sum of a deterministic time trend and a stochastic term. Within this approach, there are two competing models which depend on the nature of the process driving the latter: *trend stationary (TS)* specification and *difference stationary (DS)* specification (Dejong *et al.* 1992; Hamilton 1994, pp. 435f). A TS model is also referred to as an *integrated process of order zero*: I(0). A DS model is variously termed as an *integrated process of order one*: I(1), or a *unit root process* because one of the roots or eigenvalues of the autoregressive polynomial is equal to unity.

Usually, the substantive issue has been to determine whether a series is best characterized as a TS or DS process. The essential difference between the two models is their persistence properties subject to a random shock/innovation or impulse. In a DS model, each shock has some permanent effect on the level of the series: the series has infinite memory. However, a TS model has short memory, the effect of a shock is transitory and is eliminated since the reversion to a deterministic trend eventually dominates in the infinite future (Rudebusch 1993).

The model selection criteria are formalized by *unit root tests* and *stationarity tests*. The distinction as to whether a series is best characterized as TS or DS is important in subsequent application of asymptotic theory to obtain distributions of estimators and test statistics (Hamilton 1994), and in theoretizing of financial cycles for variables such as futures contracts, stock prices, dividends, spot and forward exchange rates and aggregate economic variables such as gross domestic product and real consumption (Phillips 1988).

Typically, one wishes to test for a unit root in the presence of a deterministic trend (Schmidt & Phillips 1992). Nevertheless, **unit root tests have low power against *plausible TS alternatives* that are distinguishable, in behaviour, from a *plausible DS null* model** (Dejong *et*

al. 1992). The same conclusion applies to stationarity tests. Simple power studies are insufficient to answer the question of the specification of time-series regression models (Rudebusch 1992, 1993). The tests must be supplemented by a comparative measure of persistence of the dynamic responses - to random disturbances or shocks - in the TS and DS specifications over a relevant horizon of practical interest. Such a measure is the *cumulative impulse response (CIR)* (Andrews 1993; Patterson 2000). The estimated TS and DS models may have similar or different implications for persistence, depending on the values taken by the model parameters (Rudebusch 1993).

Given the autoregressive structure, the **main problem is estimation of model parameters**. Common estimators, such as the least-squares (LS) estimator[1] may be significantly biased in finite samples, since the presence of the lagged dependent variable (LDV) violates the assumption of non-stochastic regressors in the classical linear regression model. Therefore, the estimated models (TS and DS) are not the 'best' candidates for the TS and DS representations of the data. Better representations would **correct the coefficient estimates of the model parameters for finite-sample bias**.

As shown later, the bias is most severe in the presence of a unit root, where it is known as "Dickey-Fuller bias" (Dickey & Fuller 1979; Rudebusch 1992). Furthermore, it is shown that the LS estimator may be seriously biased in finite samples, over much of the relevant parameter space. As such, for the unit root hypothesis, the estimation bias implies that the LS estimated models probably under-state the amount of persistence in the true TS and DS specifications.

This study examines **estimation rules that aim to improve on the bias exhibited by LS estimators** of the LDV coefficient in various formulations of an autoregressive process of order one; [AR(1)]. The CIR is not examined. In particular, the statistical properties of the estimators proposed by Andrews (1993) and Tanizaki (2000) are established and compared. No restriction is made to the unit root case. The study considers a realistic range of the parameter space over which the LDV coefficient is defined. In addition, the methods discussed are independent of the LS estimated TS or DS specifications, and of the extent of their bias (as implied by the distance of the roots of their associated lag-polynomials, from the unit circle). Furthermore, an assump-

[1] In the classical regression framework, the LS estimator is useful because the *Gauss-Markov theorem* guarantees that within the class of linear estimators, it is consistent and has minimum variance.

tion is made that the AR(1) model specifications have enough support from the data: they are correctly specified.

Chapter 1

Introduction

In many applications, a more general formulation than the pure AR(1) series is preferred. The dependent variable $\mathbf{y}$ can be expressed in a partial regression framework as[1]

$$\boxed{\mathbf{y} = \mathbf{X}\beta + \alpha\mathbf{y}_{-1} + \mathbf{u},} \tag{1.1}$$

where $\mathbf{y} = (y_2, \dots y_T)'$ is a vector of observations on the dependent variable, in a sample of size T, $\mathbf{y}_{-1}$= $(y_1 \dots y_{T-1})'$ is the vector $\mathbf{y}$ lagged one period, α is the coefficient of the LDV, β is a $k \times 1$ vector of unknown coefficients, $\mathbf{X}$ is a full column-rank $(T-1) \times k$ matrix of observations on k fixed exogenous regressors. $\mathbf{u}$ is a $(T-1) \times 1$ disturbances vector, which is considered to be Gaussian white noise, that is $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{T-1})$.

It is well known that the LS estimators of the regression coefficients in (1.1) may be seriously biased in small samples. This is because the assumption of non-stochastic regressors in the classical linear regression model, as required by the *Gauss-Markov theorem*, are not satisfied. The LDV can neither be treated as fixed in repeated sampling nor as distributed independently of the disturbances.

When (1.1) is expressed as

$$\mathbf{y} = \mathcal{Z}\Gamma + \mathbf{u}, \tag{1.2}$$

[1]The linear dynamic regression model (1.1) is sometimes termed ARX(1). It accommodates more explanatory variables which may be endogenous, i.e. effects deriving from within the system. The matrix $\mathbf{X}$ can also be stochastic. In general, $\mathbf{X}$ may comprise constants, linear trend, step-, impulse-, and seasonal- dummy variables or other covariates which do not derive from the dependent variable. It is quite useful in modelling a wide variety of dynamical systems, in practice.

where $\mathcal{Z} = [\,\mathbf{X}\,, \mathbf{y}_{-1}]$, the LS estimator of the coefficient vector Γ, denoted $\hat{\Gamma}$, is given by

$$\hat{\Gamma} = \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = (\mathcal{Z}'\mathcal{Z})^{-1}\mathcal{Z}'\mathbf{y}. \tag{1.3}$$

In (1.3), $\hat{\beta}$ is the LS estimator of the coefficient vector β and $\hat{\alpha}$ is the LS estimator of the LDV coefficient α. In this study, only the bias correction of the latter estimator is considered.

This bias in $\hat{\alpha}$ is usually large near the unit root border and varies with the form of $\mathbf{X}$ in (1.1). The form of the bias is unknown; it is a function of unknown parameters.

Various bias correction mechanisms have been proposed. The performance of the resultant bias-corrected estimators, in terms of variance and mean squared error (MSE), has been studied for example in:

1. Orcutt and Winokur (1969), where the bias-corrected estimators are based on first-order large-T approximations in Marriott and Pope (1954), and the jacknife estimator in Quenouille (1949).

2. MacKinnon and Smith (1998), where the bias-corrected estimators are established using the *bias-function*, which relates the bias of the estimator to the true parameters.

3. Kiviet and Phillips (2003, 2005), where the corrections are determined using explicit expressions obtained from higher-order asymptotic expansions.

Corrections based on asymptotic expansions lead to approximate bias-correction. With the exception of the estimators in Kiviet and Phillips (2003, 2005), the other modified estimators listed above apply corrections in simpler models than (1.1). In addition, the mechanisms are subject to specific conditions on the initial observation for the series (y_1) and are also restricted to specific forms of the exogenous vector $\mathbf{X}$.

However, all the correction mechanisms apply in a specific range on the parameter space of the LDV coefficient and improve bias to different orders of the sample size. For example, the jacknife estimator does not accommodate any exogenous regressor in the model specification. It is unbiased to order T^{-1}. On the other hand, the Kiviet and Phillips (2003) bias-corrected

estimator applies to stationary ($|\alpha| < 1$) specifications in (1.1), whereas the estimator in Kiviet and Phillips (2005) applies to the unit root case ($\alpha = 1$) in the same model. The estimators are unbiased to orders T^{-1} and T^{-2}, respectively.

Over a decade ago, Andrews (1993) proposed a 'median-unbiased' estimator of the LDV coefficient in different formulations of (1.1). Subsequently, Tanizaki (2000) proposed a 'mean-unbiased' estimator for the coefficient. In contrast to the earlier procedures, these correction mechanisms

1. Provide a general methodology to correct for bias in finite samples, irrespective of the form of $\mathbf{X}$ in (1.1).

2. Are impartial to the values of the LDV coefficient α in the space in which it is defined.

3. Are exact in the case of AR(1) series and are approximate for higher-order autoregressive processes.

4. Utilize reasonable assumptions about the initial condition, where a stationary initial condition is allowed if $|\alpha| < 1$ and a fixed initial condition otherwise.

Tanizaki (2000) analyzed the performance of the bias-corrected estimators using a **simulation method**. He considered an unrestricted space for the LDV coefficient. The author analyzed the MSE and root mean squared (RMS), of the 'median-' and 'mean-unbiased' estimators, for various forms of $\mathbf{X}$ in (1.1). However, no attempt has been made to establish their statistical properties in the practically relevant case when the space of the LDV coefficient is restricted. The present work seeks to establish the MSE and RMS of these bias-corrected or *location-based* estimators for this case.

Different *unbiasedness restrictions* are related to the definition of the 'median-' and 'mean-unbiased' estimators in Andrews (1993) and Tanizaki (2000), respectively. A new estimator is proposed, the *'most-probably-unbiased'* estimator of the LDV coefficient in an AR(1) series. Its MSE and RMS properties are studied.

Finally, the problem of inference is considered. Here, practical interest is focussed on the situation where the LDV coefficient is near the unit root border (Rudebusch 1992, 1993). In

this case, unit root tests have low power against plausible trend stationary alternatives whereas stationarity tests also have low power against plausible integrated alternatives (Dejong *et al.* 1992). The power of the tests diminishes as more exogenous regressors are added to the test regression. Unless the tests are supplemented by examining the CIR over a given horizon, the question of existence of a unit root remains uncertain. However, point and interval estimators do provide additional information.

Therefore, the determination of these estimators is incorporated without restriction to the unit root case, but rather over the entire space of the LDV coefficient. In contrast to the bounds given in Andrews (1993) for the LS estimator and the implicit bounds for the 'median-unbiased' estimator, accurate determination of the explicit bounds for all of the bias-corrected estimators is illustrated here for a given T. These are directly usable for inference about the LDV coefficient.

The rest of Chapter 1 is organized as follows. In section 1.1 different forms for the AR(1) model to be considered are defined. Section 1.2 discusses the finite sample properties of the LS estimator. Section 1.3 reviews the literature on bias, bias correction and work related to performance evaluation in terms of MSE.

In Chapters 2 and 3, the computational aspects of the estimators are dealt with in detail. In Chapter 4, the performance properties of the bias-corrected estimators as well as the LS estimator are established and compared, in terms of MSE. Explicit confidence bounds for all the estimators are also determined and analysed. Chapter 5 concludes our quest in the topic.

1.1 The models: AR(1)

Following Andrews (1993), the following forms of an AR(1) series are considered.

$$y_t = \alpha y_{t-1} + u_t, \qquad \alpha \in \Omega \equiv (-1, 1) \quad : \quad \text{Model } 1. \tag{1.4}$$

$$y_t = \mu + \alpha y_{t-1} + u_t, \qquad \alpha \in \Omega \equiv (-1, 1] \quad : \quad \text{Model } 2. \tag{1.5}$$

$$y_t = \mu + \delta t + \alpha y_{t-1} + u_t, \qquad \alpha \in \Omega \equiv (-1, 1] \quad : \quad \text{Model } 3. \tag{1.6}$$

(for $t = 2, \ldots, T$). In (1.4) - (1.6), $\{u_t\}$ is a zero-mean white noise process and Ω is the parameter space of the LDV coefficient. The initial value of the series is either random or fixed,

depending on the value of α (see below).

Following the vector notation in (1.1), the LS estimator of the LDV coefficient α in Model j, denoted $\hat{\alpha}_j$ ($j = 1, 2, 3$) in (1.4) - (1.6) is

$$\hat{\alpha}_j = \frac{\mathbf{y}_{-1}{}' (\mathbf{I} - \mathbf{P}_j) \mathbf{y}}{\mathbf{y}_{-1}{}' (\mathbf{I} - \mathbf{P}_j) \mathbf{y}_{-1}}, \tag{1.7}$$

where $\mathbf{P}_j = \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$ for models $j = 1, 2, 3$. Matrix $\mathbf{X}_j = \mathbf{0}$, $\mathbf{1}$ and $[\mathbf{1}\ \mathbf{t}]$ in models 1, 2 and 3 respectively: see Appendix A for the derivations.

Conditional on y_1 and $\mathbf{X}$, the LS estimator in (1.7) is also the maximum likelihood estimator (MLE) of α when the disturbance u_t is normally distributed (Ali 2002; Kiviet & Phillips 2003).

For Model 1, in (1.4), $\{y_t\}$ is a zero-mean strictly stationary, normal, AR(1) process. The nonstationary case (when $\alpha = 1$) is not considered since the distribution of the LS estimator depends on the initial observation y_1.

In (1.5), μ gives rise to a non-zero mean whereas δ in (1.6) represents a deterministic trend. In this context, $\{y_t\}$ in Model 2 is a strictly stationary, normal AR(1) process with mean $\mu/(1-\alpha)$ when $|\alpha| < 1$ whereas it is a random walk with an arbitrary initial condition when $\alpha = 1$. For Model 3, if $|\alpha| < 1$, $\{y_t\}$ is a strictly stationary, normal, AR(1) process about a deterministic trend. It is a normal random walk with drift δ and arbitrary initial condition when $\alpha = 1$.

The case of $|\alpha| > 1$, where $\{y_t\}$ is non-stationary and the variance of the series grows exponentially as t increases, is not considered. The choice of Ω will follow from the discussion of bias in the next section.

The initial observation plays an important role in the statistical behaviour of the LS estimator in finite samples: MacKinnon and Beach (1978) discuss the effect of y_1 on the MLE of α. In contrast, disregarding or fixing the first observation makes no difference, asymptotically. That is, its effect disappears in large sample theory.

Here, reasonable assumptions about y_1 are utilized by allowing a stationary initial condition when the series is stationary and a fixed start-up otherwise. In particular, $y_1 \sim \mathcal{N}(\mathrm{E}[y_1], \omega^2\sigma^2)$ where $0 \leq w < \infty$. If $\omega = 0$ the series has a fixed initial condition. If $\omega \neq 0$ the start-up is random. In the latter case, $\omega^2 = 1/(1-\alpha)^2$ in order that the variance of y_1 is equal to that of all other y_t.

1.2 Finite sample properties

1.2.1 Bias of the LS estimator

The LS estimator in (1.7) may be significantly biased in small samples. The magnitude of the bias depends on the form of $\mathbf{X}$ in (1.1). In particular, bias increases when a constant (intercept) and/or time-trend is included in the model.

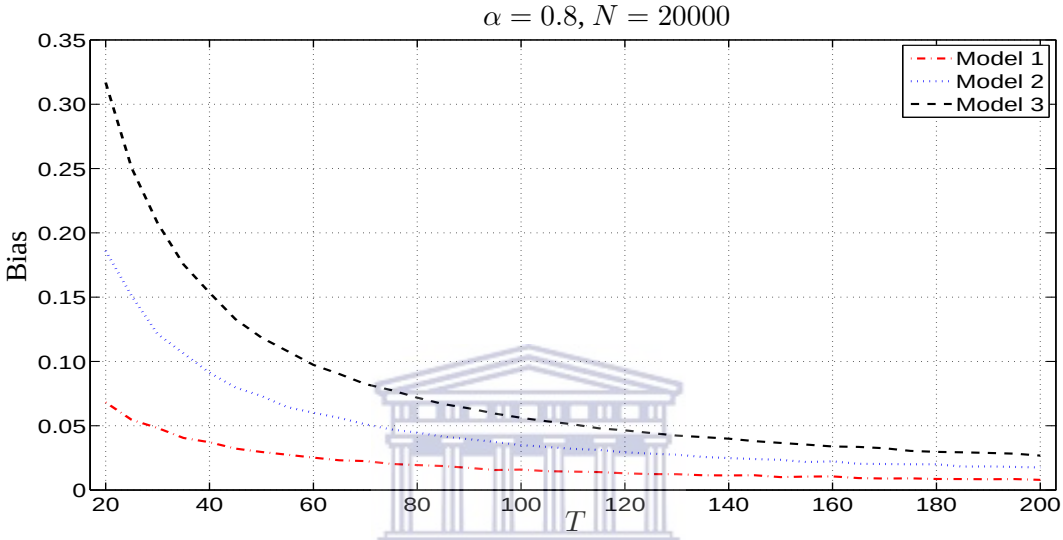


Figure 1.1: Size of the bias of the LS estimator in the models as T increases.

Figure[2] 1.1 is a convenient summary of absolute sizes of bias in each model as T increases. $N = 20000$ replicates for the sampling distribution of the LS estimator are considered when the true LDV coefficient $\alpha = 0.80$. The following features are apparent:

1. In comparison to the other models, the bias in Model 1 (where $\mathbf{X} \equiv \mathbf{0}$) is small. Even for a sample of size $T = 20$ the magnitude of bias is approximately 0.06. However, for the same T in Models 2 and 3, the biases are approximately 3 and 5 as large.

2. Bias decreases as $T \to \infty$: Hurwicz (1950) and Kendall (1954) established that the bias of the LS estimator is $o(T^{-1})$.

Figure 1.2 compares the *bias functions*

$$b_T(\alpha) = \mathbb{E}[\hat{\alpha}_j] - \alpha \quad \text{for} \quad j = 1, 2, 3, \tag{1.8}$$

[2]The computer programs used to obtain most of the figures and tables are given Appendix C.

where $\mathbb{E}$ denotes the usual expectation operator.

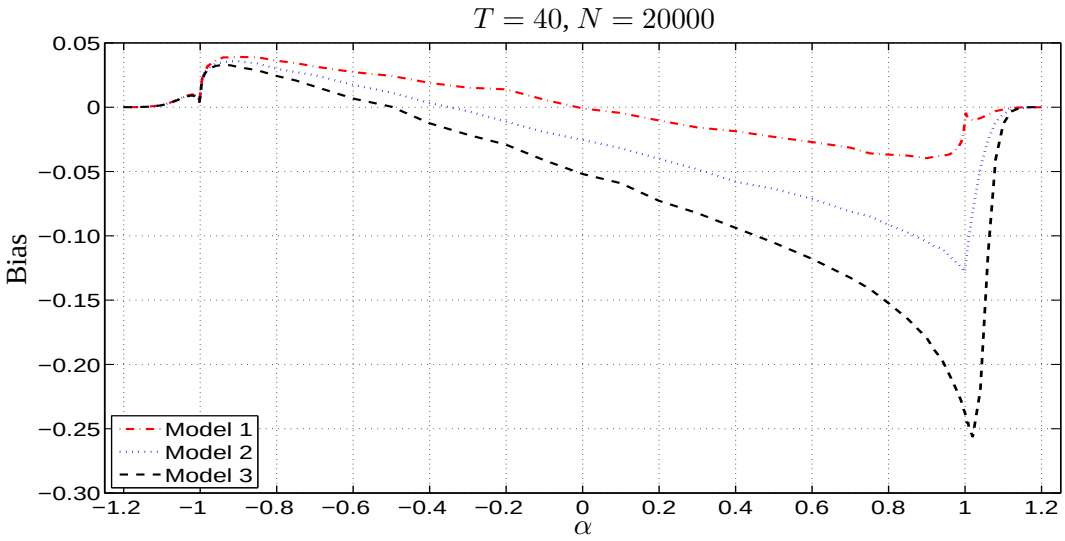


Figure 1.2: Comparison of bias functions for the LS estimator in different models. The domain has been extended for illustrative purposes.

For the sample size considered ($T = 40$), again with 20000 replications, it can be noted that:

1. The LS estimator is positively biased for $\alpha < -0.5$ and negatively biased for $\alpha \geq 0$ in the three models. This shows a slight disagreement with the Model 2 simulation results in Orcutt and Winokur (1969) where negative bias is claimed for all $\alpha \geq -0.5$.

2. It is also evident that bias significantly increases as the LDV coefficient approaches unity. Greater bias is observed for α close to $+1$ than at -1. This is because of increased correlation between the numerator and the denominator in (1.7) for positive α.

For Model 1, the bias is less than 0.05 over the range of α shown. Investigations at moderately larger sample sizes, say $T = 100$, show the LS estimator to be essentially unbiased, with the size of the bias less than 0.01. However, the model is usually inadequate for practical purposes (although it is useful for theoretical considerations as a benchmark model). In fact, the mean of a series is seldom known in practice.

Clearly, for $|\alpha| > 1$, the bias of the LS estimator decreases steeply. Random initial conditions are used over all values shown. Figure 1.2 is used here only for illustrative purposes, on the nature of the LS estimators' bias over an unrestricted space. The results at $|\alpha| = 1$, where the assumption of a fixed initial condition is used, were omitted deliberately.

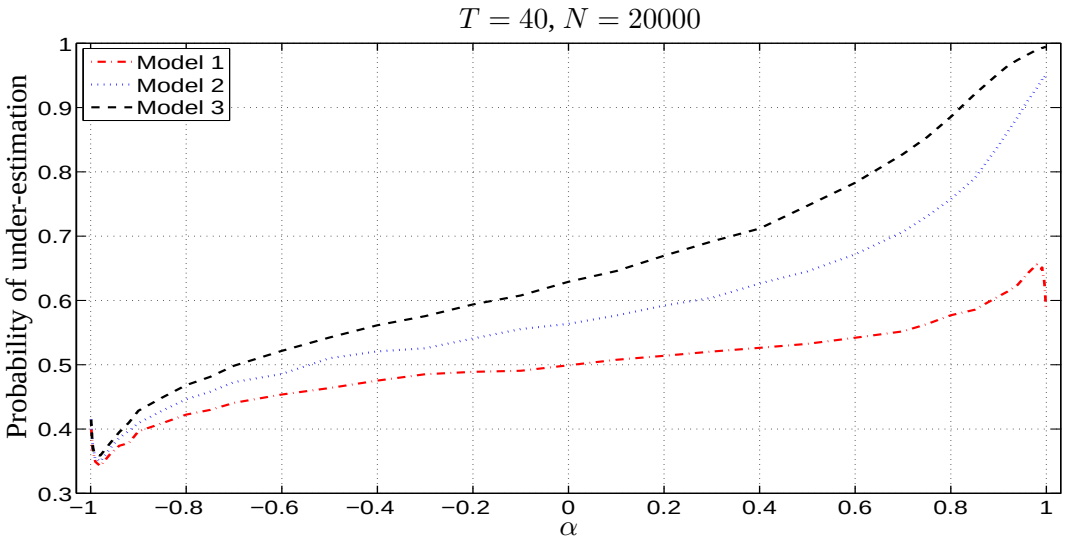


Figure 1.3: Probability that the LS estimator underestimates the true LDV coefficient $\alpha \in \Omega$.

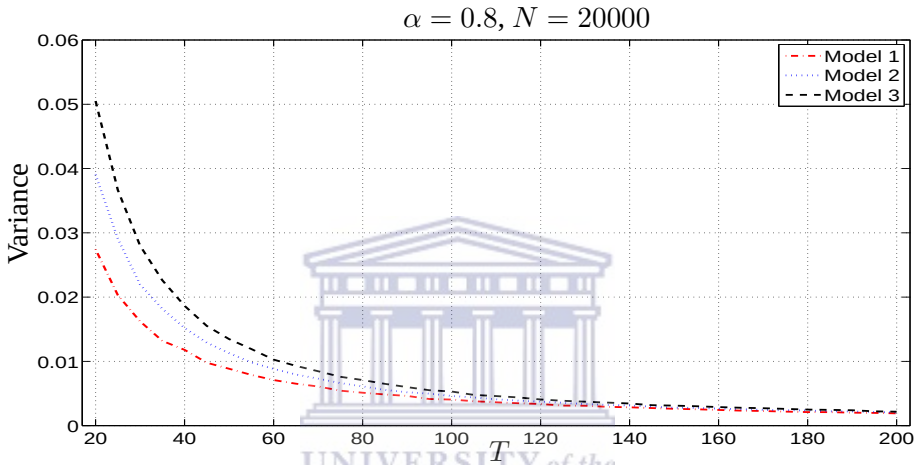
Due to the bias, the LS estimator is not a good estimator of the true parameter. Figure 1.3 depicts the propensity of the estimator to underestimate the true LDV coefficient over Ω. In general, except near $\alpha = -1$, the probability increases as $\alpha \to 1$. When $\alpha > 0$, there is more than a 50% chance that the LDV coefficient is under-estimated. Compared to Model 1, for a given α and T, the probability of under-estimation in Models 2 and 3 are much higher. For example, for Model 3, the LS estimator will almost inevitably under-estimate the true parameter when $\alpha = 1$.

From the above discussion, bias correction is useful especially when $\mathbf{X}$ incorporates a constant or time-trend. Therefore, regard should be placed on the nature and size of bias and their possible consequences for inference(s) based on the LS estimator. Bias is also a factor in the error of prediction (Harvey 1981; Yule & Kendall 1948).

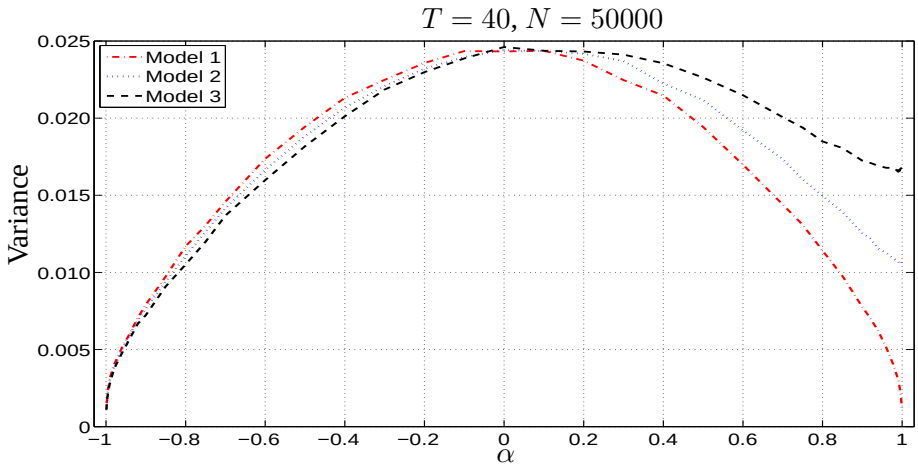
1.2.2 Variance of the LS estimator

In general, estimators with small variance are desirable. As such, a trade-off between unbiasedness and variance is inevitable.

Figure 1.4 shows the variance of the LS estimator. Figure 1.4(a) shows decreasing variance in different models as $T \to \infty$. More replications ($N = 50000$) are used to minimize the influence of the random initial conditions in Figure 1.4(b). As $\alpha \to 1$ the variance of the LS estimator increases as more exogenous regressors are included in the model. It is higher when α is close to zero than at alternative values in the parameter space in all models.



(a) Variance of the LS estimator as T increases.



(b) Variance of the LS estimator over Ω.

Figure 1.4: Variance of the LS estimator in different models. A larger number of replicates N is used in (b) than in (a) to enhance accuracy and ensure smooth curves i.e. eliminate variability due to the random initial condition(s).

1.3 Literature review

In dynamic regression models, the bias property is not peculiar to the LS estimator of α. Other estimators share this behaviour especially near the nonstationary border. The literature on bias, bias-correction and analysis of performance (in terms of MSE) will be reviewed.

Approximate procedures

Apart from the jacknife estimator (Quenouille 1949, 1956), early approaches to bias-correction utilized asymptotic expansions (in stationary models). The methods did not allow for trend and $\alpha = 1$ (i.e. they considered Model 1 only). Emphasis was on establishing the first and second moments of estimators at high accuracy in finite samples.

Hurwicz (1950), Marriott and Pope (1954), Kendall (1954), and White (1961) established large-T approximations of bias to first-order in stationary models. The resultant bias-corrected estimators are unbiased to the same order.

Orcutt and Winokur (1969) used Monte Carlo experiments to investigate bias in Model 2. Based on results in Quenouille (1949) and Marriott and Pope (1954), they proposed two bias-corrected estimators and evaluated their performance in terms of the MSE. The bias-corrected estimators were noted to have larger MSE for α close to zero than the LS estimator, but showed better performance for other values of α in all sample sizes. Furthermore, they showed that the LS estimator and the estimator in Marriott and Pope (1954) had smaller MSE than the jacknife estimator in Quenouille (1949), for all α and T.

Grubb and Symons (1987) used large-T asymptotics and derived an expression for bias to the order T^{-1}. Shaman and Stine (1988), and Stine and Shaman (1989) derived the first-order term of bias of the LS estimator as a function of unknown parameters. They gave expressions for bias for both the LS estimator and the Yule-Walker estimator in stationary models. Using approximations to the Yule-Walker estimator, they showed that the LS estimator has smaller bias. Le Breton and Pham (1989) calculated the exact asymptotic biases of the LS estimator in stationary and unit root models.

Abadir (1993) gave a closed form analytical expression for bias given that Model 1 has a unit root, and developed a simple approximation. Abadir (1995) derived the minimum MSE unbiased estimator for the same model. The latter estimator was expressed in terms of exponential functions in polynomials of T^{-1} and T^{-2}. The author stressed the need for bias-correction through the use of procedures such as those in Andrews (1993).

Near the unit root border, first order approximations to the mean and variance break down meaning that adequate correction cannot be based on such expansions. Alternative techniques have been investigated and are summarized below.

Kiviet and Phillips (2003) gave explicit expressions for bias, variance and MSE of the LS estimator of the LDV coefficient in (1.1) calculated from large-T first-order approximations. In contrast to earlier procedures, the results were obtained using Taylor-type expansions containing quadratic forms in standard normal vectors. They proposed bias-corrected estimators and gave analytical expressions, to first-order, for their variances and MSEs in stationary models. The estimators were analyzed numerically for the case of Model 2. For all α and T, they established that bias-corrected estimators resulting from Marriott and Pope (1954) and Kendall (1954) using large-T, first-order approximations, had uniformly smaller second-order bias [Theorem 4.2 in Kiviet and Phillips (2003)] than the estimators based on similar approximations when expressed in quadratic forms. However, the latter were uniformly more efficient in terms of MSE.

Recently, Kiviet and Phillips (2005) extended the results to the unit root case in (1.1). They determined the bias to order T^{-2}, variance and MSE to order T^{-3}, and provided numerical illustrations of the relevance and accuracy of the analytic expressions in finite samples. The expressions were obtained using higher-order asymptotic expansions. These have the advantage of allowing further theoretical investigations into the limitations of first-order asymptotics, the nature of bias in standard estimators, and thus suggest corrections.

However, even though the second-order approximations are more accurate than those to order T^{-1}, the former are of limited use when α is close to unity. In particular, for near-unit root models, the approximations to order T^{-2} are much more vulnerable to the magnitude of the LDV coefficient than the latter (Kiviet and Phillips 2005).

Exact Procedures

Sawa (1978) evaluated the exact bias and variance of the LS estimator using numerical integration of the moment generating function (MGF) of the estimator in Models 1 and 2 (stationary cases). The LS estimator was represented as a ratio of quadratic forms in normally distributed variables, assuming fixed initial conditions for the series. Nankervis and Savin (1988) corrected and extended the results in Sawa (1978).

Andrews (1993) proposed the 'median-unbiased' estimators of α for AR(1) models.

MacKinnon and Smith (1998) used an alternative technique for bias correction using properties of the bias function and the variance of the parameters, where these can be estimated using computer simulation or otherwise. They showed that bias can be reduced by order T and in some cases eliminated. The authors also analyzed the performance of the resultant estimators in terms of variance and MSE. The MSE is smaller in the neighborhood of $|\alpha| = 1$ but higher for α close to zero.

Tanizaki (2000) introduced the 'mean-unbiased' estimator for AR(1) models based on Andrews (1993).

Chapter 2

CDF and PDF of the LS Estimator

2.1 Introduction

The location statistics (mean, median and mode of the distribution of the LS estimator) can be obtained using:

1. Simulation.
2. Numerical Integration.

The simulation method is discussed in Chapter 3. This Chapter considers the numerical integration method. The method is exact since the actual accuracy can be controlled precisely.

For a given $\alpha \in \Omega$, T and Model j $(j = 1, 2, 3)$, the notation $\bar{\alpha}_j$, $\tilde{\alpha}_j$ and $\ddot{\alpha}_j$ for the mean, median and mode, respectively, is used. The corresponding location functions are determined, over Ω, as $g(\alpha) = \bar{\alpha}_j|\alpha$ for the mean, $g(\alpha) = \tilde{\alpha}_j|\alpha$ for the median, and $g(\alpha) = \ddot{\alpha}_j|\alpha$ for the mode. The functions relate the true value of the coefficient to the statistics.

The present Chapter is organized as follows. Section 2.2 clarifies the derivation of the finite sample distribution of the LS estimator. The LS estimator of α is shown to be a function of the residuals. If the residuals are invariant, interest is computation of the distribution function of the estimators. Implementation of the CDF is discussed with respect to computational efficiency.

In section 2.3, the exact density of the LS estimator is derived: this is a new result. An efficient method for its computation is also established. Finally, numerical evaluation of the location functions is discussed in section 2.4.

2.2 Exact CDF of the LS estimator

The finite sample distribution of the LS estimator can be derived directly from the form of the

- estimator $\hat{\alpha}_j$ in (1.7) or
- regression t-statistic for α [see (NB2) on page 98, for the unit root case].

Both methods exploit the invariance property to restate the LS estimator or the statistic in a form which is invariant with respect to the nuisance parameters in β, through the application of partitioned or partial regression theory, i.e. equation (1.1).

The result of the theory - accredited to the *Frisch-Waugh theorem* (Greene 2003, pp. 27f) - applies in a time-series setting whether the regression is fitted with a time-trend variable or not. The partial regression coefficient of the LDV can be obtained in isolation by regressing $\mathbf{y}$ and $\mathbf{y}_{-1}$ on $\mathbf{X}$, separately. The process, *partialling out* or *netting out*, involves eliminating the effect of $\mathbf{X}$ and using the resultant residuals which are invariant to the coefficients of $\mathbf{X}$ in a simple regression.

The partial regression coefficient, $\hat{\alpha}_j$ $(j = 1, 2, 3)$ in (1.7), can be written as

$$\hat{\alpha}_j = \frac{\mathbf{y}_{-1}{}' \mathbf{M}_{\mathbf{X}} \mathbf{y}}{\mathbf{y}_{-1}{}' \mathbf{M}_{\mathbf{X}} \mathbf{y}_{-1}} = \frac{\hat{\mathbf{u}}_1' \mathbf{M}_{\mathbf{X}} \hat{\mathbf{u}}_2}{\hat{\mathbf{u}}_1' \mathbf{M}_{\mathbf{X}} \hat{\mathbf{u}}_1} \equiv \xi(\hat{\mathbf{u}}), \tag{2.1}$$

where $\hat{\mathbf{u}}_1 = \mathbf{M}_{\mathbf{X}}\mathbf{y}_{-1}$ and $\hat{\mathbf{u}}_2 = \mathbf{M}_{\mathbf{X}}\mathbf{y}$ are the residual vectors from least squares regression of $\mathbf{y}_{-1}$ and $\mathbf{y}$ on $\mathbf{X}$ alone, respectively. $\mathbf{M}_{\mathbf{X}} = \mathbf{I} - \mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ may be termed as the "residual maker" defined on the columns of $\mathbf{X}$ (Greene 2003). The RHS follows from the properties of $\mathbf{M}_{\mathbf{X}}$ i.e. it is symmetric and *idempotent*.

However, as stated in (2.1), the residuals (and hence the distribution of $\hat{\alpha}_j$) depend on the unknown parameter vector β because it is incorporated in $\mathbf{y}_{-1}$, the variance σ^2, and are also

influenced by assumptions about y_1 (Kiviet & Phillips 1992). Therefore, an additional transformation is necessary to render them invariant to these parameters.

Kiviet and Phillips (1992) established statistics for *exact*[1] and/or *similar* inference in an AR(1) multiple regression model when the innovation process is normal white noise. They are generalizations of the statistics (NB1) and (NB2) discussed in Note 1 on page 97 below. The statistics are obtained by decomposing the LDV into parts which depend on the fixed part ($\mathbf{X}\beta$), stochastic part ($\hat{\mathbf{u}}$), the initial observation (y_1), and introducing a redundant regressor. The resultant residuals are invariant to the coefficients in β and the initial variable y_1.

Dejong *et al.* (1992) and Andrews (1993) proved the same distributional invariance. The statistics can be used to test the hypothesis that $\hat{\alpha}_j$ $(j = 1, 2, 3)$ equals any value, say x_0. Dickey and Fuller (1979) cited invariance of test statistics when $x_0 = 1$.

Kiviet and Phillips (1992) gave the LS estimator and the regression t-statistic in terms of residuals derived from those in (2.1) as

$$\hat{\alpha}_j - x_0 = \frac{\hat{\mathbf{u}}'\mathbf{M}_1^0\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\mathbf{M}_2^0\hat{\mathbf{u}}} = \frac{\hat{\mathbf{v}}'\mathbf{M}_1^0\hat{\mathbf{v}}}{\hat{\mathbf{v}}'\mathbf{M}_2^0\hat{\mathbf{v}}} \quad \text{where} \quad \hat{\mathbf{v}} = \frac{\hat{\mathbf{u}}}{\sigma} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad \text{and} \tag{2.2}$$

$$\mathrm{t} = r\left\{\frac{T-H-1}{1-r^2}\right\}^{\frac{1}{2}} \quad \text{with} \quad r = \frac{\hat{\mathbf{v}}'\mathbf{M}_1^0\hat{\mathbf{v}}}{\sqrt{(\hat{\mathbf{v}}'\mathbf{M}_3\hat{\mathbf{v}})(\hat{\mathbf{v}}'\mathbf{M}_2^0\hat{\mathbf{v}})}}, \tag{2.3}$$

where $\hat{\mathbf{u}} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ are the residuals vector from $\mathbf{y}$ or $\mathbf{y}_{-1}$ regressed on $\mathbf{X}$, separately. H satisfies $k \leq H \leq 2k+1$ and $\mathbf{M}_1^0$, $\mathbf{M}_2^0$ and $\mathbf{M}_3^0$ are transformation matrices. The superscript indicates dependence on x_0.

We note that $\hat{\mathbf{u}} = \hat{\mathbf{u}}_1$ or $\hat{\mathbf{u}} = \hat{\mathbf{u}}_2$, depending on the context.

Kiviet and Phillips (1992) showed that the LS estimator satisfies (2.2) - where $\xi(\mathbf{z}) \equiv \hat{\alpha}_j$, $\xi_0(\mathbf{z}) \equiv x_0$ and the RHS, $U(\cdot)$, is a function of zero-mean residuals (Harvey 1981; Krämer & Sonnberger 1986; Butler & Paolella 1999).

It follows from (2.2) that the distribution of $\hat{\alpha}_j$ $(j = 1, 2, 3)$ can be obtained by establishing that of $U(\hat{\mathbf{u}} - x_0)$ for all x_0. This is discussed in the next section. The consequence is that the evaluation of the location statistics can proceed under a scenario where both the parameter and

[1]They are *exact* since actual significance levels can be controlled precisely and *similar* because the null-distribution is invariant to the nuisance parameters (Kiviet & Phillips 1992).

the estimator are subjected to the same one-to-one transformation. That is, using $F_U(0)$ or its corresponding density $f_U(0)$.

The RHSs in (2.2) and (2.3) are ratios of quadratic forms in standard normal variates. They are completely determined by $\mathbf{X}$, x_0 and the sample size T (Kiviet & Phillips 1992, p. 352). Their distribution can be evaluated numerically subject to an acceptable degree of accuracy using Monte-Carlo techniques. However, use of (2.3) is problematic[2] owing to its form. Existing methods of evaluation rely on re-writing the statistics (ratio of quadratic forms) as a single quadratic form in standard variates. The transformation of a ratio of quadratic forms to a single quadratic form is clarified in section 2.2.1, as the notation is introduced.

2.2.1 Derivation of the exact CDF

Andrews (1993) used (2.2) to establish the distribution of the LS estimator of α using the Imhof (1961) technique. As a consequence of invariance, it suffices to consider $\beta = 0$ when determining the distribution of the LS estimator. The T observations of the series denoted by $\mathbf{Y} = (y_1, y_2, \ldots y_T)'$ can be expressed in terms of the underlying errors as

$$\mathbf{Y} = \mathbf{R}_\alpha \hat{\mathbf{v}}, \tag{2.4}$$

where the matrix $\mathbf{R}_\alpha$ is defined for each $\alpha \in \Omega$ by

$$\mathbf{R}_\alpha = \begin{bmatrix} b & 0 & \cdots & \cdots & 0 \\ b\alpha & 1 & \ddots & & 0 \\ b\alpha^2 & \alpha & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ b\alpha^{T-1} & \alpha^{T-2} & \cdots & \alpha & 1 \end{bmatrix} \tag{2.5}$$

and

$$b = \begin{cases} \frac{1}{\sqrt{1-\alpha^2}} & \text{iff} \quad \alpha \in (-1, 1), \\ 0 & \alpha = 1. \end{cases} \tag{2.6}$$

[2]Its distribution is complicated since r^2 is not a ratio of independent χ^2 variates. It can not be established numerically using the Imhof (1961) technique. However, it can be used as a basis for exact inference through simulation experiments (Kiviet & Phillips 1992, p. 354).

(Andrews 1993; Kiviet & Phillips 2003; Broda, Carstensen & Paolella 2004). For a given true value of α, the exact finite sample distribution of $\hat{\alpha}_j$ can be obtained - using (2.2) - as

$$\mathbf{P}_\alpha(\hat{\alpha}_j \leq x) \quad = \quad \mathbf{P}_\alpha(\hat{\mathbf{v}}'\mathbf{W}_{\alpha jx}\hat{\mathbf{v}} \leq 0), \tag{2.7}$$

where $\mathbf{W}_{\alpha jx}$ is a symmetric matrix[3] defined as

$$\mathbf{W}_{\alpha jx} = \mathbf{R}'_\alpha \left[\frac{\mathbf{D}'_0(\mathbf{I}-\mathbf{P}_j)\mathbf{D}_1}{2} + \frac{\mathbf{D}'_1(\mathbf{I}-\mathbf{P}_j)\mathbf{D}_0}{2} - x\,\mathbf{D}'_1(\mathbf{I}-\mathbf{P}_j)\mathbf{D}_1 \right] \mathbf{R}_\alpha, \tag{2.8}$$

where $\mathbf{I}$ is a $(T-1)\times(T-1)$ identity matrix whereas $\mathbf{D}_0 = [\mathbf{0} \vdots \mathbf{I}]$ and $\mathbf{D}_1 = [\mathbf{I} \vdots \mathbf{0}]$ are $(T-1)\times T$ matrices. For brevity, the notation $Q = \hat{\mathbf{v}}'\mathbf{W}_{\alpha jx}\hat{\mathbf{v}}$ is used: Q is a single quadratic form in uncorrelated zero-mean random variates. According to the *principal axis theorem* (Scheffé 1959, p. 396f) there exists an orthogonal transformation such that

$$Q = \hat{\mathbf{v}}'\mathbf{W}_{\alpha jx}\hat{\mathbf{v}} = (\hat{\mathbf{v}}^*)'\mathbf{\Lambda}\,\hat{\mathbf{v}}^* = \sum_{i=1}^{m} \lambda_i \chi_{\eta_i;\kappa_i^2}, \tag{2.9}$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, λ_i are the $m \leq T$ eigenvalues of the matrix $\Sigma_{\hat{\mathbf{v}}^*}\mathbf{W}_{\alpha jx}$ whereas $\chi_{\eta_i;\kappa_i^2}$ denotes independent chi-squared variates with η_i degrees of freedom and non-centrality parameter κ_i^2. $\Sigma_{\hat{\mathbf{v}}^*}$ is the covariance matrix of $\hat{\mathrm{v}}^*$. Standard normal variates are non-central with parameter $\kappa_i = 0$ (Scheffé 1959). It follows that, $\Sigma_{\hat{\mathbf{v}}^*} = \mathbf{I}$ since $\hat{\mathbf{v}}^* \sim \mathcal{N}(0, \mathbf{I}_T)$ and has rank $m = T$.

Furthermore, in numerical implementations one can reasonably expect the eigenvalues to differ by virtue of approximations inherent in the numerical techniques employed in their determination. As such, their multiplicities η_i can be taken to be equal to 1 for all i.

Let the exact distribution function of Q be $F_Q(q)$. Interest here is in

$$\boxed{F_Q(0) \equiv \mathbf{P}(Q \leq 0),} \tag{2.10}$$

which can be obtained by inversion of the characteristic function of Q. Q depends on $\alpha, j, x, T, \hat{\mathbf{v}}^*$ and λ.

Box (1954), Grad and Solomon (1955), and Gurland (1955) gave the characteristic function of Q for standard normal variates as

$$\Psi_Q(u) = \prod_{j=1}^{T} (1 - 2\,\mathbf{i}\,\lambda_j\,u)^{-\frac{1}{2}}. \tag{2.11}$$

[3] See Andrews (1993, p. 163) for methodology and specification in (2.7) and (2.8) respectively.

where $\mathbf{i} \equiv \sqrt{-1}$. Imhof (1961) generalized the result for non-standard normal variates. The exact CDF of the LS estimator, denoted as $F_{\text{LS}}(x) \equiv F_Q(0)$, can be obtained from its corresponding density, denoted as $f_{\text{LS}}(x)$, as a consequence of the *Fourier inversion theorem* (Riley, Hobson & Bence 2002) where

$$f_{\text{LS}}(x) = \frac{1}{2}\int_{-\infty}^{\infty} \Psi_Q(u)\, e^{-\mathbf{i}ux}\, du. \tag{2.12}$$

The results in Imhof (1961), explicitly derived by Gil-Pelaez (1951), are adapted here to the central variates case by expressing the distribution function as

$$\boxed{F_{\text{LS}}(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^{\infty} \frac{\sin\phi(u,x)}{u\rho(u,x)}\, du,} \tag{2.13}$$

where

$$\phi(u,x) = \frac{1}{2}\sum_{i=1}^{T} \tan^{-1}[\lambda_i(x)u] \quad \text{and} \tag{2.14}$$

$$\rho(u,x) = \prod_{i=1}^{T} [1+\lambda_i(x)^2u^2]^{\frac{1}{4}}. \tag{2.15}$$

The integrand at the origin can be obtained by applying l'Hôpital's rule[4] as

$$\lim_{u\to 0} \frac{\sin\phi(u,x)}{u\rho(u,x)} = \frac{1}{2}\sum_{i=1}^{T}\lambda_i. \tag{2.16}$$

The exact distribution of Q can also be obtained by inverting its Laplace transform. However, numerical evaluation of the resultant forms of the distribution function and its corresponding density are difficult to work with, hence approximations are necessary (Grad & Solomon 1955, pp. 427f). Laplace approximations perform very poorly especially for values of α close to unity. In addition, they are exceptionally inaccurate when T is small, but their performance improves as the sample size increases (Paolella 2003, pp. 320f).

In the next section, various approaches to the evaluation of F_Q or its corresponding density are discussed.

[4] See Appendix A for the details. Note: the summations and product in (2.11), (2.14) and (2.15) are over T as opposed to m in (2.9) by virtue of the earlier stated assumption that $\eta_i = 1$ for all i.

2.2.2 Implementation

Many statistics can be expressed as ratios of quadratic forms. The Durbin-Watson test statistic (Sargan & Bhargava 1983) and Von-Neuman ratio are immediate examples. However, computable forms of F_Q such as (2.13), or the exact density f_Q, are often unavailable despite the statistics being simple.

Forchini (2002) gave a closed form solution of F_Q for the MLE of α in a Gaussian AR(1) model, but it is intractable. In general, the approach adopted to reduce Q to a computable form determines the efficacy of implementation. For example, use of *cumulants* and their derivatives leads to expressions involving nested infinite sums. For other cases, unsolved integrals may result which do not lend themselves to efficient numerical computation. More efficient approaches involve converting Q to its diagonal form or tridiagonal form (Farebrother 1990). The latter approach, *diagonalization*, was demonstrated in (2.9).

Various algorithms have been proposed and compared for evaluating the CDF in (2.13) and/or its corresponding PDF $f_{\text{LS}}(x)$. The choice between them depends on the desired level of accuracy and consideration of the computational effort required. A trade-off between accuracy and efficiency is in most cases inevitable in practical implementation. The numerical algorithms may be classified as either exact or approximate.

Exact numerical techniques are referred to as such because they can be implemented subject to a pre-specified tolerance or accuracy level. They include the Imhof (1961) and Davies (1973) algorithms. They are based on inverting the characteristic function obtained by reducing the defining matrices to their eigenvalue or tridiagonal form in (2.9). Shively, Ansley and Kohn (1990) also gave an exact and efficient procedure utilizing a modification of the *Kalman filter*.

Approximate numerical techniques use approximations to the distribution of Q. Lieberman (1994) derived approximations to the distribution based on the saddlepoint method. Such techniques have been shown to be numerically more efficient than the exact procedures especially where interest is in the moments of the distribution. This applies for example to near-unit root processes. The approximations are handy when:

1. $\hat{\mathbf{v}}$ is a more general correlated error process. In this case, expressions for the distribu-

tion of Q that can be realistically computed do not exist (Paolella 2003). Taylor series expansions or the method of Laplace can be used if the mean is assumed to be zero.

2. Accuracy is not the main concern.

However, relative to the exact procedures they have the following limitations:

1. Approximations are not suitable when exact results are required. Exact procedures for the density vary with the mathematical form in which it is expressed.

2. Whilst using saddle point approximations to the PDF, a normalization constant must be computed separately to ensure a proper density function. Although the constant is close to unity, its computation is best done using a separate integration which introduces a computation time constraint.

3. Saddle point approximations of the density are obtained using only the leading term of the moment generating function of Q [the same argument does not apply to saddle point approximations of the CDF (Butler & Paolella 1999)]. Therefore, the resultant density is slightly less accurate than the exact PDF obtained using the Imhof (1961) technique.

   In section 2.3, it is shown that the PDF of the LS estimator can be explicitly determined from the CDF function in (2.13).

Nonetheless, numerical integration utilizing the inversion formula (2.12) is time consuming because of repeated evaluations of the integrand. The numerical determination of the eigenvalues in (2.11) requires $O(T^3)$ operations. In contrast to the eigenvalue or tridiagonalization approach (Farebrother 1990), the method by Shively, Ansley and Kohn (1990) requires only $O(T)$ operations. The latter method requires less computing time and memory.

The method by Shively, Ansley and Kohn (1990) involves obtaining the characteristic function in (2.11) using an eigenvalue-free method, namely by writing it in terms of a determinant. A modified version of the Kalman filter is then used to compute the determinant. However, the tridiagonalization approach is more general. Imhof's algorithm is used here on the basis of ease of implementation and its more general applicability.

A set of *MATLAB* programs for evaluating $F_{\text{LS}}(x)$ is provided in Appendix C. That is, Algorithm 2 - 5. In contrast, Andrews' (1993) implementation employed *Fortran* subroutines due to Koerts and Abrahamse (1971): Farebrother (1990) gave an equivalent *Pascal* implementation.

The *MATLAB* function `quad` is used to evaluate the integral: this uses an extrapolated *Simpson's quadrature* in an adaptive recursive algorithm. For the integral evaluation, there are two sources of error:

1. Numerical error, resulting from the use of quadrature formulae.
2. Truncation error, resulting from replacing $[0, \infty)$ by a finite range $0 \leq u \leq b$ in (2.13).

Since modern quadrature algorithms employ adaptive step sizes, the numerical error can be reasonably contained within a pre-specified tolerance level. Obviously, the choice of b can affect the accuracy and computational efficiency of the method. The methodology leading to efficient choices of b is developed in the next section.

An implementation using the trapezoidal rule - *MATLAB* function `trapz` - is also given in Algorithm 6, but it is not as efficient. As expected, a large number of steps are required to minimize the integration error or match the error rates obtained using Simpson's rule. Davies (1973, p. 417) tabulated the number of terms required to achieve an accuracy to the order of 10^{-5} in numerical integration using the trapezoidal rule. This can be used as a guide during implementation.

2.2.3 Efficiency considerations: integration truncation bounds

It is desirable to determine a suitable replacement for the infinite upper bound b so that the truncation error, denoted as B_{trc}, is minimized. The integration bounds are determined at each x in computing the CDF in (2.13).

Imhof (1961) gave the limit

$$B_{\text{trc}} = \frac{1}{\pi} \int_b^{\infty} \frac{\sin \phi(u,x)}{u\rho(u,x)}\, du < \frac{2}{T\,\pi\, b^{\frac{T}{2}} \prod_{i=1}^{T} |\lambda_i|^{\frac{1}{2}}}, \tag{2.17}$$

where ϕ and ρ are defined in (2.14) and (2.15). The integration limit b can be determined from (2.17) as

$$b_{\text{imh}} = \left[\frac{4}{T^2\,\pi^2\,B_{\text{trc}}^2\,\prod_{i=1}^{T}|\lambda_i|} \right]^{1/T}. \tag{2.18}$$

As given in (2.18), it depends on a pre-specified bound on truncation error B_{trc}, T and the eigenvalues. It is apparent that for $\lambda_i << 1$, excessively large values of b_{imh} would be obtained. These may have dire consequences for program performance.

Proceeding from first principles, useful CDF integration bounds are derived. Given B_{trc}, the *efficient CDF bounds* can be obtained by considering a restriction based on the order statistics $\lambda_{(i)}$ of λ_i. For some suitable value of $m < T$, the product term contained in the expression (2.15) for ρ can be written as

$$\begin{aligned}\prod_{i=1}^{T}(1+\lambda_i^2u^2) &= \underbrace{\prod_{i=1}^{m}(1+\lambda_{(i)}^2u^2)}_{\text{for } \lambda_{(i)}<<1}\ \underbrace{\prod_{i=m+1}^{T}(1+\lambda_{(i)}^2u^2)}_{\text{otherwise}} \\ &> \prod_{i=m+1}^{T}(1+\lambda_{(i)}^2u^2), \end{aligned} \tag{2.19}$$

since $\prod_{i=1}^{m}(1+\lambda_{(i)}^2u^2) > 1$. It follows that

$$\begin{aligned} B_{\text{trc}} = \frac{1}{\pi}\int_b^\infty \frac{\sin\phi(u,x)}{u\rho(u,x)}\,du &< \frac{1}{\pi}\int_b^\infty \frac{1}{u\prod_{i=1}^{T}(1+\lambda_{(i)}^2u^2)^{1/4}}\,du \\ &< \frac{1}{\pi}\int_b^\infty \frac{1}{u\prod_{i=m+1}^{T}(1+\lambda_{(i)}^2u^2)^{1/4}}\,du \\ &< \frac{1}{\pi}\int_b^\infty \frac{1}{u\prod_{i=m+1}^{T}|\lambda_{(i)}^2u^2|^{1/4}}\,du \\ &= \frac{2}{(T-m)\,\pi\,b^{(T-m)/2}\prod_{i=m+1}^{T}|\lambda_{(i)}|^{1/2}}. \end{aligned} \tag{2.20}$$

Therefore,

$$b = \left[\frac{4}{(T-m)^2\,\pi^2\,B_{\text{trc}}^2\,\prod_{i=m+1}^{T}|\lambda_{(i)}|} \right]^{\frac{1}{T-m}}. \tag{2.21}$$

A small value of b is preferable for computational efficiency when evaluating the integral (2.13). Such a bound can be established by choice of an optimal m such that

$$\boxed{b_{\text{eff}} = \min_m \left[\frac{4}{(T-m)^2\,\pi^2\,B_{\text{trc}}^2\,\prod_{j=m+1}^{T}|\lambda_{(i)}|^{1/2}} \right]^{\frac{1}{T-m}}}. \tag{2.22}$$

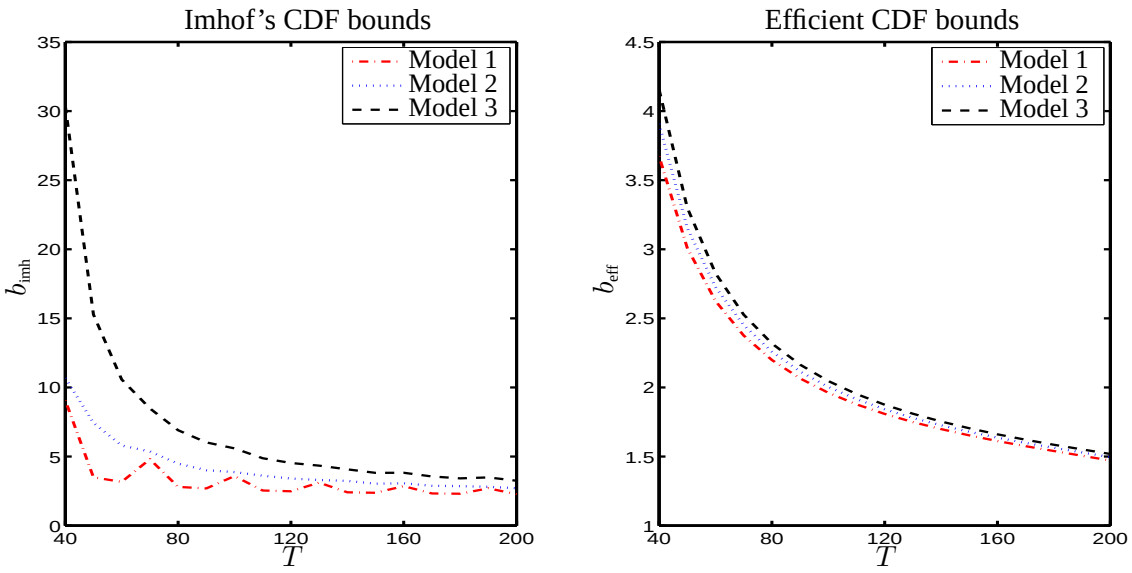


Figure 2.1: CDF integration truncation bounds in different models as T increases ($\alpha = x = 0.5$).

An investigation into the properties of b_{imh} and b_{eff} reveals T as the dominant factor. Figure 2.1 shows that bounds obtained using both methods decrease as T increases. The bound curves for b_{eff} are smoother and the values are significantly smaller than those obtained using Imhof's restriction. The computations are done when the truncation error bound is specified at $B_{\text{trc}} = 10^{-8}$ and $\alpha = x = 0.5$, using Algorithm 7. The Imhof bounds for Model 3 can be quite large for small sample sizes. Furthermore, b_{imh} are larger than b_{eff} in all the cases (α, T and model).

The properties of b_{imh} and b_{eff} also vary with x for a given α. The boxplots in Figure 2.2 depict the distributions of the bounds for fixed values of α in Model 2 when $T = 40$. [The choice $T = 40$ is somewhat arbitrary; any small value could be used for these illustrative cases. Below, other values are sometime used]. The parameter x is allowed to vary in $[-1.5, 1.5]$ at step sizes of 0.001. The upper and lower edges of the boxes indicate the upper and lower quartile values of the bounds whereas the medians are shown by the notches. The whiskers extend from each edge to encompass a factor 1.5 the interquartile range. Outliers are indicated by plus signs in the plots. Notably, the distribution of b_{eff} has no outliers.

In the case of Model 1, a symmetry about $\alpha = 0$ is observed in both panels in Figure 2.2 but that is not so in Model 3 (see Figure 2.3).

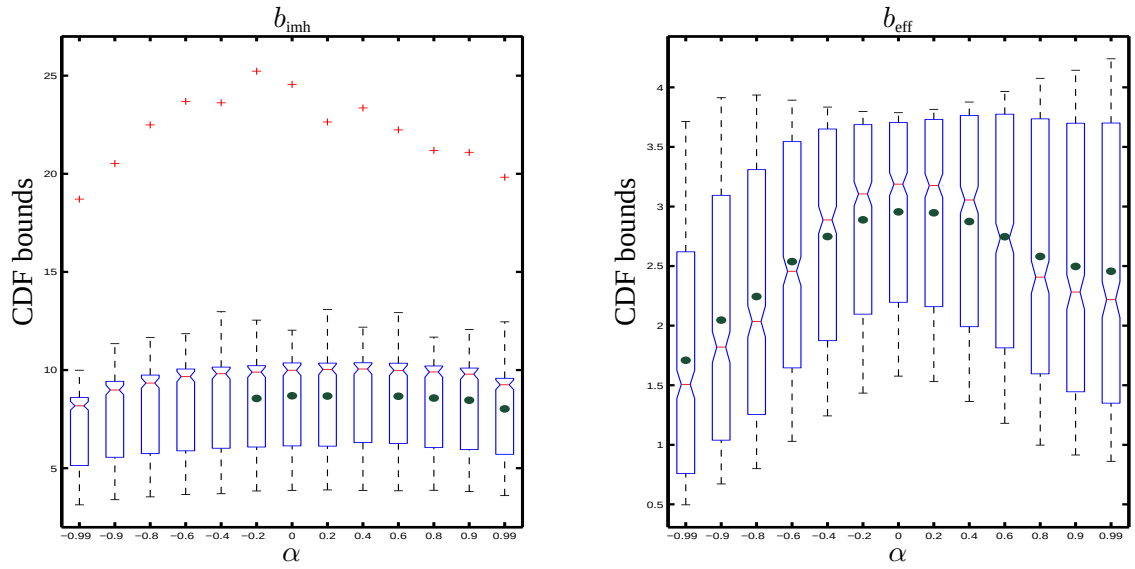


Figure 2.2: Distribution of the CDF integration truncation bounds for a given values of α (Model 2, $T = 40$).

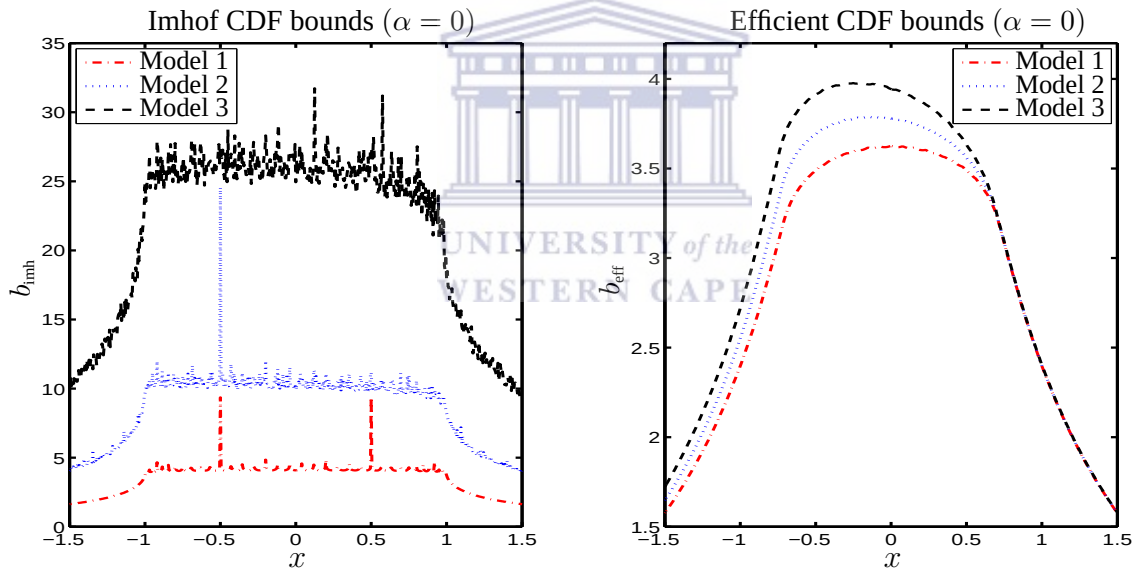


Figure 2.3: CDF integration truncation bounds for $\alpha = 0$, in different models, as x is varied ($T = 40$).

As mentioned earlier, the excessively large values of b_{imh} are occasioned by small values of $|\lambda|$ in the denominator in (2.17). This characteristic is illustrated in the left panel of Figure 2.3 where the properties of the bounds for the three models are compared as x is varied, given that $\alpha = 0$. In fact, infinite Imhof bounds may arise if $\lambda_i = 0$ for some i. As such, the domain of the y-axis for the left panel in Figure 2.2 is $[d_0, \infty)$ for some constant d_0. To illustrate this, for the case of finite bounds for all x, the mean bound is indicated by a dot (•) inside the boxes in Figure 2.2.

For the un-dotted boxes, the mean bound is infinite corresponding to cases where, for some i, $b_{\text{imh}} = \infty$.

However, for all α, x, and suitable B_{trc}, b_{eff} is always finite hence the evaluation of the CDF is faster.

2.3 Exact PDF of the LS estimator

It is non-trivial to determine the density function directly by integration of the inversion formulae in (2.12) except in certain cases (Imhof 1961). It is shown that the exact PDF of the estimator can be obtained by differentiating (2.13) with respect to x.

2.3.1 Derivation of the exact PDF

By reversing the order of integration and differentiation, it follows from (2.13) that

$$f_{\text{LS}}(x) = \frac{\partial\, F_{\text{LS}}(x)}{\partial x} = -\frac{1}{\pi}\int_0^\infty \frac{\partial}{\partial x}\left\{\frac{\sin\phi(u,x)}{u\,\rho(u,x)}\right\} du. \tag{2.23}$$

The following result is obtained[5]

$$\boxed{f_{\text{LS}}(x) = \frac{1}{2\pi}\int_0^\infty \frac{1}{\rho(u,x)}\sum_{i=1}^{T}\left\{\frac{\lambda_i u\sin\phi(u,x) - \cos\phi(u,x)}{(1+\lambda_i(x)^2u^2)}\,\frac{d\,\lambda_i(x)}{d\,x}\right\} du,} \tag{2.24}$$

where $\phi(u,x)$ and $\rho(u,x)$ are as given in (2.14) and (2.15) respectively and the derivative of the i^{th} eigenvalue is given by

$$\begin{aligned}\frac{d\,\lambda_i(x)}{d\,x} &= \mathbf{w}_i'(x)\,\frac{d\,\mathbf{W}_{\alpha jx}}{d\,x}\,\mathbf{w}_i(x)\\ &= -\mathbf{w}_i'(x)\,\mathbf{R}_\alpha'[\mathbf{D}_1'(\mathbf{I}-\mathbf{P}_j)\mathbf{D}_1]\mathbf{R}_\alpha\,\mathbf{w}_i(x) \qquad (j=1,2,3)\end{aligned} \tag{2.25}$$

(Magnus & Neudecker 1999). In (2.25), $\mathbf{W}_{\alpha jx}$ is the matrix defined in (2.8) and $\mathbf{w}_i(x)$ is the eigenvector associated with the eigenvalue $\lambda_i(x)$. Henceforth, the derivative of the latter is denoted by $\lambda_i' \equiv d\lambda_i/dx$, for notational convenience.

[5] See Appendix A for the deduction of the result (2.24) from (2.23).

Implementation

The exact PDF of the LS estimator derived above can be evaluated numerically using the Imhof (1961) technique. The density function is well defined at the origin unlike the integral for the CDF in (2.13). The integral in (2.24) is computed using Simpson's rule, as implemented in the function `quad`.

The set of functions *PDFsimpson*, *PDFintegrand*, *PDFtruncate* and *PDFquadform* (Algorithm 8 - 11 in Appendix C) were used for the task of computing the mean and mode statistics of the LS estimator using (2.24). Explanatory comments are provided within the code. The function *PDFtruncate* facilitates the determination of suitable integration truncation bounds for the PDF, as clarified in the next section.

2.3.2 Efficient truncation bounds: PDF computation

As was done for the CDF in (2.22), plausible replacements of the infinite limit for the integral in (2.24) are desired. A similar approach can be adopted to determine the minimal upper integration limit b which satisfies a pre-specified bound on the truncation error.

In this case, the truncation error bound is

$$B_{\text{trc}} = \frac{1}{2\pi}\int_b^{\infty} \frac{1}{\rho}\sum_{i=1}^{T}\left\{\frac{\lambda_i u\sin\phi - \cos\phi}{1+\lambda_i^2 u^2}\lambda_i'\right\} du. \tag{2.26}$$

A useful restriction based on the order statistics $|\lambda|_{(i)}$ of $|\lambda_i|$ can be established by separately considering the sum and product terms in (2.26). For some n, the summation term can be partitioned as

$$\begin{aligned}\sum_{i=1}^{T}\frac{\lambda_i u\sin\phi-\cos\phi}{(1+\lambda_i^2u^2)}\lambda_i' &< \sum_{i=1}^{T}\frac{|\lambda_i|u+1}{1+\lambda_i^2u^2}|\lambda_i'| \\ &< \underbrace{\sum_{i=1}^{n}|\lambda_i'|\left[|\lambda|_{(i)}u+1\right]}_{\text{for } |\lambda|_{(i)}<<1} + \underbrace{\sum_{i=n+1}^{T}\frac{|\lambda|_{(i)}u+1}{|\lambda|_{(i)}^2u^2}|\lambda_i'|}_{\text{for larger values of } |\lambda|_{(i)}}, \end{aligned} \tag{2.27}$$

where the subscript on λ_i' is understood to be the same as that of the associated order statistic $|\lambda|_{(i)}$. For the product term $\rho(u,x)$, a similar restriction to that in (2.19) applies for some m. An appropriate methodology for the choice of n is proposed below.

From (2.27), it follows that

$$\begin{aligned} B_{\text{trc}} &< \frac{1}{2\pi}\int_b^{\infty} \frac{1}{\prod_{i=m+1}^{T}[|\lambda|_{(i)}u]^{\frac{1}{2}}} \left\{ \sum_{i=1}^{n} |\lambda_i'| \left[|\lambda|_{(i)}u + 1\right] + \sum_{i=n+1}^{T} \frac{|\lambda|_{(i)}u + 1}{|\lambda|_{(i)}^2 u^2} |\lambda_i'| \right\} du \\ &= \frac{S_n(|\lambda|_{(i)})}{(T-m-4)\,\pi \prod_{i=m+1}^{T} |\lambda|_{(i)}^{\frac{1}{2}}\, b^{1/(T-m-4)}}, \end{aligned} \tag{2.28}$$

where

$$S_n(|\lambda|_{(i)}) = \sum_{i=1}^{n} |\lambda_i'| \left[|\lambda|_{(i)} + 1\right] + \sum_{i=n+1}^{T} \frac{|\lambda_i'|}{|\lambda|_{(i)}^2} \left[|\lambda|_{(i)} + 1\right]. \tag{2.29}$$

The calculations leading to (2.28) are given in Appendix A. By inspection, the second term of the sum $S_n(|\lambda|_{(i)})$ in (2.29) will be large if $|\lambda|_{(i)} < 1$ in the denominator. As such, a useful restriction is to choose n such that $|\lambda|_{(n)} < 1$, $|\lambda|_{(n+1)} \geq 1$.

At every iteration in x and given a pre-specified value of B_{trc}, the *efficient PDF bounds* b_{eff} can be established by a choice of an optimum m such that

$$b_{\text{eff}} = \min_{m} \left[\frac{S_n^2}{(T-m-4)^2\,\pi^2\, B_{\text{trc}}^2 \prod_{i=m+1}^{T} |\lambda|_{(i)}} \right]^{\frac{1}{(T-m-4)}}. \tag{2.30}$$

An investigation into the properties of b_{eff} in (2.30) shows similar characteristics as those of the efficient CDF bounds in (2.22). The sample size exerts the most influence: as T increases b_{eff} decreases. In addition, the bounds increase from Model 1 to 3. Table B11 summarizes the results for various combinations of α and T. In larger sample sizes and for a given α, the difference between the bounds for the different models decreases.

For a given α and T, the efficient PDF bounds are greater than the efficient CDF bounds given in Table B10. Section 2.4 discusses the computation of the location statistics, and hence the location functions, using the bounds determined in (2.22) and (2.30).

2.4 Numerical evaluation of the location functions

2.4.1 Median function

The median statistic can be obtain as the p^{th} *quantile* of the distribution of the LS estimator, where[6] $p = 0.5$. For a given $\alpha \in \Omega$ and T, $\tilde{\alpha}_j$ $(j = 1, 2, 3)$ can be obtained using (2.13) as

$$F_{\text{LS}}(\tilde{\alpha}_j) = 0.5. \tag{2.31}$$

Equation (2.31) is solved using the *MATLAB* function `fzero`. The set of functions used are provided in Appendix C. Compared to the other statistics, $\tilde{\alpha}_j$ can be obtained at high accuracy, even at low tolerances for `quad`, without much loss of computational efficiency. The tabulated values in Andrews (1993) can easily be recovered at the default tolerance level `tol` $= 10^{-6}$. However, as shown in Chapter 4, if the derivatives of the median function are required, `tol` must be set at lower levels e.g. 10^{-14} when using either `quad` or the *Lobatto quadrature* algorithm `quadl`.

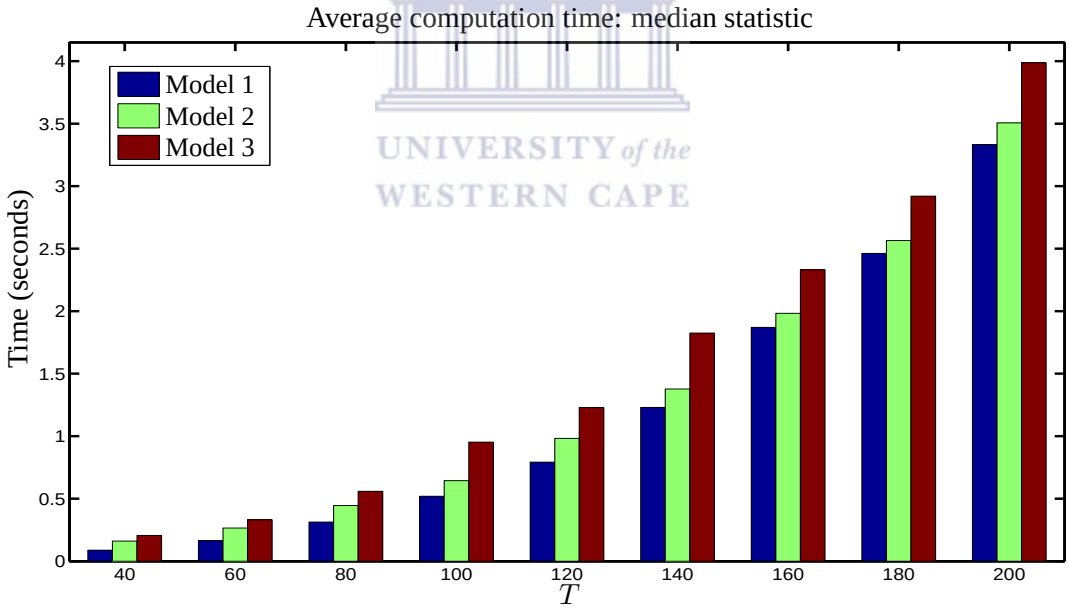


Figure 2.4: The average computation time of the median from the CDF using Imhof's technique (`tol` $= 10^{-6}$).

Tables B2, B5 and B8 in Appendix B summarize the results for the median function obtained using (2.31) in Models 1, 2 and 3 respectively.

[6] At alternative values of p, *central confidence intervals* for α can also be defined (Andrews 1993).

Figure 2.4 shows average times used to compute the median statistic of the LS estimator for various T (see also Table 3.1 on page 35). For each T and model, calculations were repeated 25 times, and the average computation time noted. All computations were carried out using a 1.6 Giga Hertz Centrino processor equivalent to a Pentium IV. The computation time for the median increases with T. For given T, the time increases with the form of the regressor vector $\mathbf{X}$ in (1.1): the average time increases from Model 1 to 3 as additional variables are included in the regression. To arrive at Figure 2.4, a constant upper bound, $b = 10$, was assumed for the integration limit in (2.13). The default tolerance level for `quad` was used.

2.4.2 Mean function

For given α and T, the mean statistic of the LS estimator for Model j $(j = 1, 2, 3)$ is given by

$$\bar{\alpha}_j = \mathbb{E}_{\text{LS}}(\hat{\alpha}_j) = \int_{-\infty}^{\infty} x\, f_{\text{LS}}(x)\, dx. \tag{2.32}$$

The computation of (2.32) proceeds subject to the choice of finite integration limits beyond which the density is effectively zero.

Algorithm 12 in Appendix C is used to obtain $\bar{\alpha}_j$. The integral in (2.32) is evaluated using `trapz`. Step-sizes of length `par` $= 0.0005$ are used and noted to be adequate for accurate results. $f_{\text{LS}}(\cdot)$ is obtained using the functions given in section 2.3.1. As shown in Chapter 3, where the numerical integration and simulation techniques are compared with respect to computation time and effort, evaluation of the mean is more costly than that of the median and mode statistics. However, $\bar{\alpha}_j$ $(j = 1, 2, 3)$ is more 'robust' to the choice of the tolerance parameters of `quad`. Therefore, accurate values can be obtained at the default level, `tol` $= 10^{-6}$.

Tables B1, B4 and B7 show the summarized results for the mean function obtained using (2.32) in Models 1, 2 and 3 respectively.

2.4.3 Mode function

For a given α and T, the mode of the density of the LS estimator for Model j $(j = 1, 2, 3)$ is obtained as

$$\ddot{\alpha}_j = \arg\left\{\max_x \left[f_{\text{LS}}(x)\right]\right\}. \tag{2.33}$$

The *MATLAB* functions `fminsearch` or `fminbnd` can be used to solve (2.33). However, the default options for the two functions are quite undemanding. In particular, their termination tolerance on x (`TolX`) is 1×10^{-4} as compared to that for `fzero`, which is to the order 10^{-16}. It is therefore advisable to use the default optimization structure of `fzero` with the function `fminbnd`.

It is undesirable to search for the mode over the entire domain of x. The bias property of the LS estimator can be relied upon to define an appropriate search range `[L,U]`, and hence improve the computational efficiency of the subroutines. Figures 1.1 and 1.2 can be used (in conjunction with each other) as a guide to defining the search-range in the different models. Algorithm 13 is used for the task of computing the mode function.

Tables B3, B6 and B9 show the summarized results for the mode function obtained using (2.33) in Models 1, 2 and 3 respectively.

Chapter 3

The Location-based Estimators

3.1 Introduction

In this Chapter, the method of obtaining the location-based estimators of α when the parameter space of $\hat{\alpha}_{\text{UE}} \equiv g^{-1}(\cdot)$ is restricted is explained. The *final estimators* can be obtained from the inverse of the appropriate location function g. A new estimator is proposed following from the most-probably-unbiasedness condition discussed in Chapter 1. The estimator is dubbed the *mode-based* estimator since it derives from the mode function.

Prior to establishing the location-based estimators, the technique of obtaining the location functions using simulation is discussed and compared to the numerical integration method - in terms of accuracy and computational efficiency. The simulation method is shown to be adequate in approximating the statistics when a large number of replicates are considered. It is also desirable when T is large.

The work is organized as follows. Section 3.2 deals with the determination of the mean and median function since a common approach can be adopted for both. Computational considerations are also mentioned.

A technique for estimating the mode statistic of the LS estimator is proposed in section 3.3. It uses a popular non-parametric technique for density estimation: *kernel density estimation*. A two-step procedure is given and shown to be efficient and accurate for determining the mode

of the LS estimator. An investigation of its efficiency relative to the numerical technique is conducted. Finally, the mechanism of obtaining the final estimators of the LDV coefficient is introduced.

3.2 Simulation method: mean and median functions

3.2.1 Introduction: notation

In the case of simulation, artificially generated series are used for each model. The technique relies on the sampling distribution of the LS estimates of α for a given T.

Andrews (1993) showed that given the invariance of the LS estimator of α to the actual values of the mean and slope, the LS estimates denoted as $\hat{\alpha}^*$ of α can be obtained from the simple form $y_t = \alpha y_t + u_t$ $(t = 2, \ldots, T)$.

The notation $\hat{\alpha}^*_{ji}$ is used to denote the i^{th} LS estimate of α in Model j $(j = 1, 2, 3)$. For a given α, T and Model j $(j = 1, 2, 3)$, the sampling distribution of the LS estimator can be approximated using simulation as $\{\hat{\alpha}^*_{ji}\}_{i=1}^{N}$. For a given innovation distribution the N simulated LS estimates are independently and identically distributed (IID) with common density $f_{\text{LS}}(\cdot)$ which depends only on α, T and model type.

For fixed α and T, the statistics corresponding to (2.31) and (2.32) for the median and mean of the LS estimator in Model j $(j = 1, 2, 3)$ are denoted by $\tilde{\alpha}^*_j$ and $\bar{\alpha}^*_j$ respectively. They can be obtained as

$$\tilde{\alpha}^*_j = \text{Median of } \left\{\hat{\alpha}^*_{j1}, \ldots, \hat{\alpha}^*_{jN}\right\}, \quad \text{and} \tag{3.1}$$

$$\bar{\alpha}^*_j = \frac{1}{N}\sum_{i=1}^{N} \hat{\alpha}^*_{ji}. \tag{3.2}$$

The mean function can be established as $g(\alpha) = \bar{\alpha}^*_j|\alpha$ while the median function is given by $g(\alpha) = \tilde{\alpha}^*_j|\alpha$.

3.2.2 Accuracy vs efficiency: numerical vs simulation method

As N increases, $\bar{\alpha}_j^*$ and $\tilde{\alpha}_j^*$ will approach their respective exact values $\bar{\alpha}_j$ and $\tilde{\alpha}_j$ obtained using numerical integration. To quantify the accuracy of the simulation method, for fixed α and T, the following statistics are defined

$$\text{Dev}_1 = \bar{\alpha}_j^* - \bar{\alpha}_j \qquad \text{and} \qquad \text{Dev}_2 = \tilde{\alpha}_j^* - \tilde{\alpha}_j. \tag{3.3}$$

Figure 3.1 demonstrates the behaviour of the statistics for a grid of α values when $T = 50$ in Model 1. As expected, when $N = 80000$ the magnitude of the deviations is small compared to when $N = 10000$.

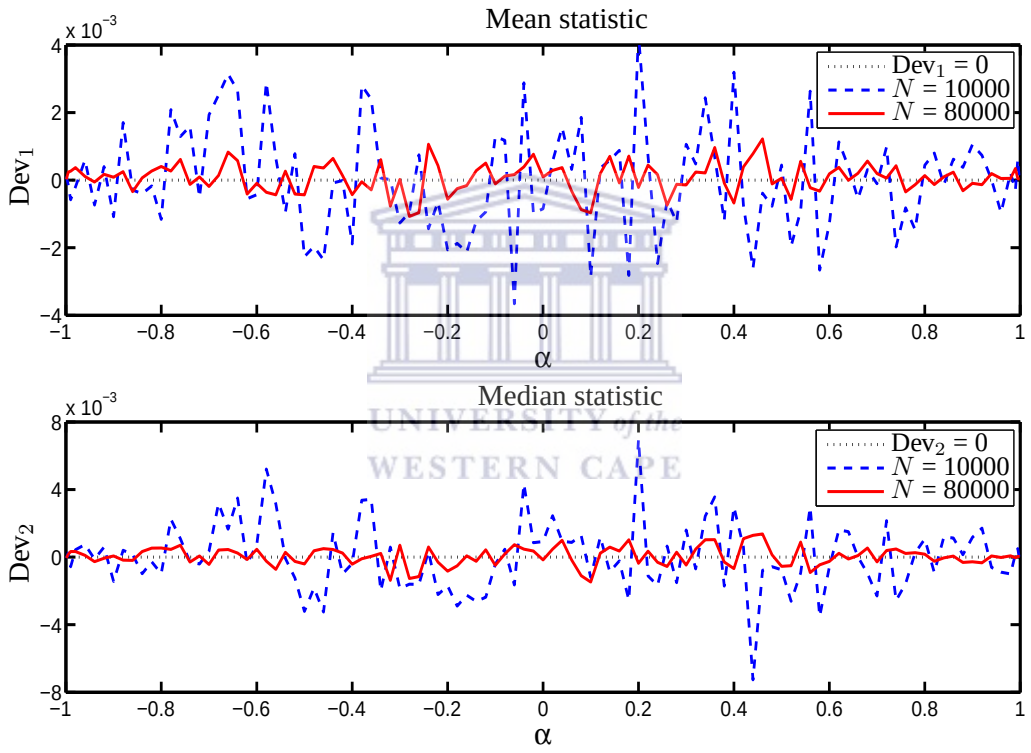
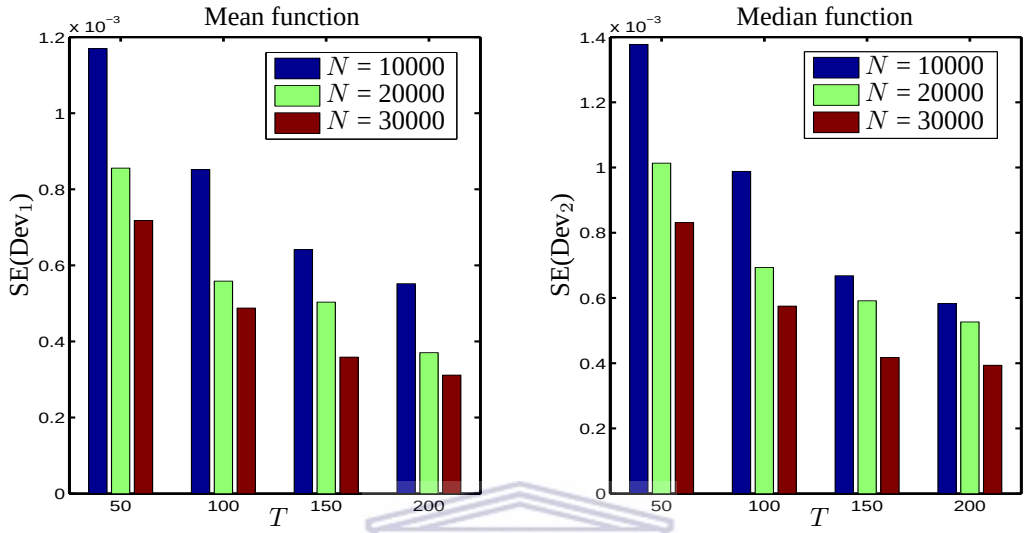


Figure 3.1: Deviations of the mean and median statistics obtained using simulation from exact values determined using numerical integration.

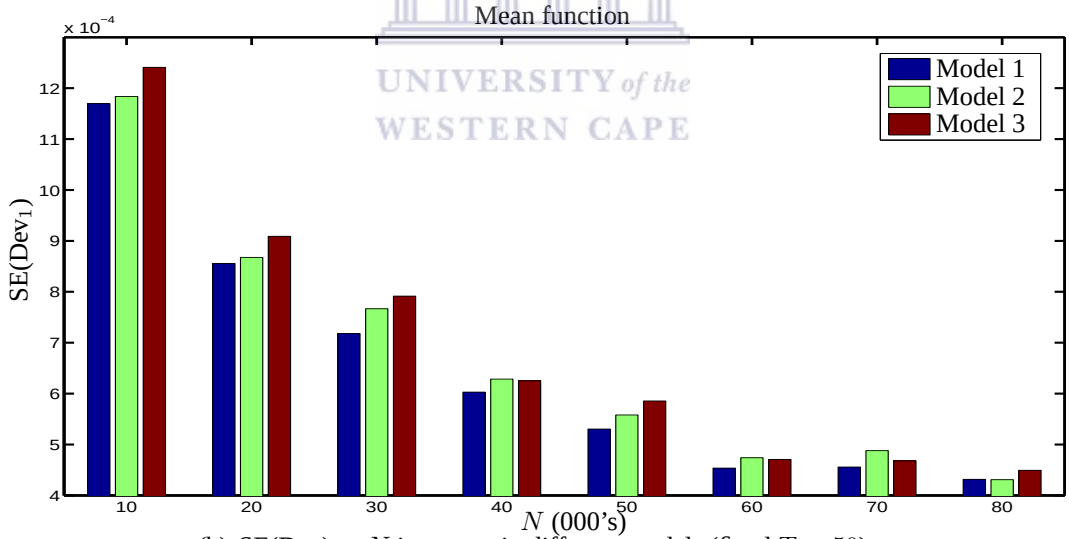
For fixed α and T, it is observed that $|\text{Dev}_1|$ is always smaller than $|\text{Dev}_2|$. Furthermore, the deviations are smaller near $|\alpha| = 1$. The choice of N determines the accuracy in estimation of the mean and median functions. In particular, for fixed T, model and α

$$\text{SE(Dev)} \propto \frac{1}{\sqrt{N}}, \tag{3.4}$$

where SE denotes the standard error. However, N affects the efficiency of the method as an alternative to numerical integration. Since the bias and variance of the LS estimator decrease with increasing T, it follows that the two statistics in (3.3) will decrease - for T large - even for relatively smaller N. Figure 3.2 summarizes the result.



(a) SE(Dev) as T increases (fixed N, Model 1).



(b) SE(Dev) as N increases in different models (fixed $T = 50$).

Figure 3.2: Approximate standard errors in computing the mean function (Dev_1) and median function (Dev_2) in different models.

As T increases the accuracy of the simulation method improves. For example, from the right panel in Figure 3.2(a) an accuracy to the order of 6×10^{-4} can be achieved, for the median

function, when $T = 200$ and $N = 10000$. As such, when T is large, the simulation method may be preferred to the numerical integration technique since it is computationally more efficient. Figure 3.2(b) is an illustration of (3.4) in the different models.

Table 3.1 compares the average computation times of the mean and median statistics under each of the methods for various T. The average is determined by considering the total time used to compute the statistics given 105 values of $\alpha \in (-1, 1]$.

Relative to the simulation method, the computation of the mean by numerical integration[1] is very costly when T is large. The simulation method has the advantage over the numerical integration method that the mean and median both follow almost trivially once the collection of least squares estimates have been generated.

Table 3.1: Comparison of the average computation time (in seconds) of the mean and median statistics; numerical integration vs simulation method.

(a) Numerical integration

	Mean			Median		
$T \setminus$ Model	1	2	3	1	2	3
50	31.4	32.0	32.4	0.1	0.2	0.3
100	93.2	96.2	97.4	0.5	0.6	1.0
150	216.7	223.6	224.1	1.4	1.5	1.9
200	466.2	447.4	443.1	3.3	3.5	4.0

(b) Simulation (mean or median)

T/N	10000	20000	30000	40000	50000
50	1.1	3.2	8.4	17.6	29.4
100	1.1	3.4	8.8	18.4	30.1
150	1.3	3.7	9.0	18.8	30.8
200	1.5	4.9	11.1	20.4	33.0

For the median statistic, preference can be given to the numerical integration technique since it is more accurate and computationally efficient for all T.

Tables B12 and B13 in Appendix B summarize the results obtained using the simulation method ($N = 50000$) for the mean and median functions respectively. Illustrative values of the standard errors when $N = 50000$ in the various models can be ascertained from Figure 3.2. The results obtained using the simulation technique agree very well with the exact values obtained using the numerical integration method.

[1]The default optimization parameters for `quadl` and `fzero` are used. Density values less than 10^{-7} are ignored whereas `par` $= 0.001$, as discussed in section 2.4, is used. A constant bound $b = 10$ is assumed.

3.3 Simulation method: mode function

Histogram maxima are inadequate as estimators of modes (Silverman 1986). Other parametric and non-parametric estimators of the mode (Bickel 2002, 2003; Grenander 1965) of a sampling distribution may be directly extended to determine the mode statistic of the LS estimator. However, most of these methods have high variance especially when outliers are evident in the data (Bickel 2002, 2003).

In a preliminary analysis done by the current author, the effectiveness of the *Fisher z-transformation* in transforming the distribution of LS estimates of the LDV coefficient to approximate normality was investigated, for the case of near-unit-root processes. The transformed data can then be used to estimate[2] the mode. The discussion of the results is deferred to Note 4 on page 99.

A more intuitive natural data-dependent method is to estimate the mode from the *empirical density function* (EDF) determined using a *weighting* or *kernel* function. The method has low bias, high efficiency and is robust to outliers (Wand & Jones 1995).

3.3.1 Kernel density estimation

Given a sample of IID observations[3] $X = \{X_i\}_{i=1}^{N}$ from a continuous univariate distribution with PDF $f_{\text{LS}}(x)$ to be estimated, the kernel estimator $\hat{f}(x,h)$ of the true density is defined as

$$\hat{f}(x;h) = \frac{1}{N\,h}\sum_{i=1}^{N}\mathbb{K}\left(\frac{x-X_i}{h}\right), \tag{3.5}$$

where $\mathbb{K}$ is a kernel function satisfying $\int_x \mathbb{K}(x)dx = 1$ and $h > 0$ is a *smoothing parameter, bandwidth* or *window width* (Hastie, Tibshirani & Friedman 2001).

The kernel technique (*kernel smoothing*) provides a simple way of finding the structure in the data set without knowledge of the functional form for $f_{\text{LS}}(x)$, while remaining mathematically tractable. The estimator $\hat{f}(x,h)$ can be interpreted as a sum of bumps placed at the observations

[2]Under the assumption of normality, the mean, median and mode are equal. Bickel (2003) proposed a power transformation for skew data.

[3]The subscript for the model is dropped for notational convinience. In particular, $\{X_i\} \equiv \{\alpha^*_{ji}\}$ for any given model $j = 1, 2$ or 3 (fixed): see section 3.2.

whereas the kernel $\mathbb{K}$ determines their shape. The bandwidth h determines the width (*spread*) of the bumps.

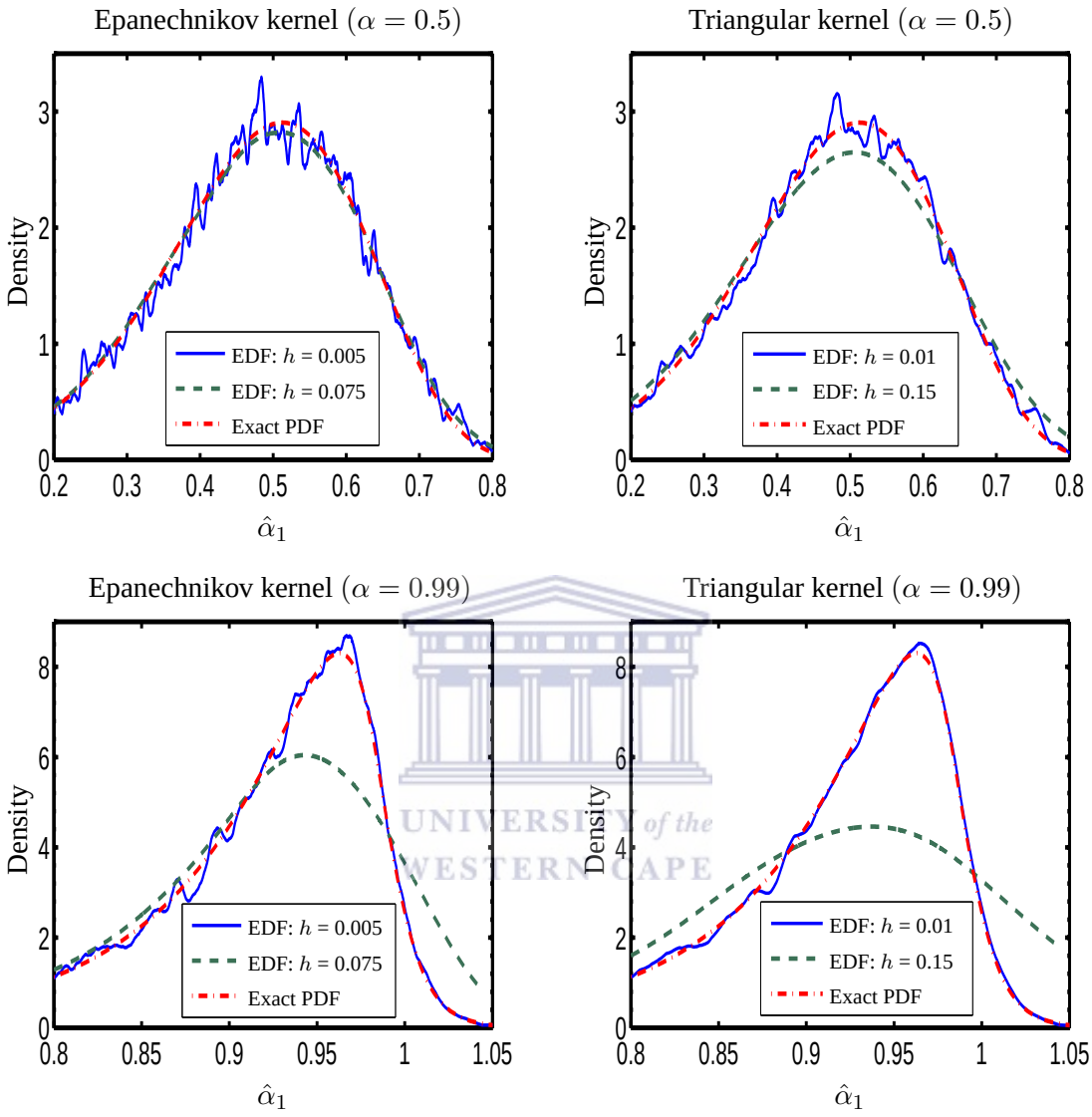


Figure 3.3: Bandwidth selection - effect of h on the EDF of the LS estimator $\hat{\alpha}_1$ (Model 1, $T = 40$, $N = 10000$) i.e. oversmoothing (dashed line) and undersmoothing (continuous line). The exact PDF - obtained using numerical integration - is shown by the dash-dotted line.

The fundamental problem when using (3.5) is the dependence on h. This problem is common to all non-parametric curve estimators (Wand & Jones 1995). The choice of h has a significant influence on the effectiveness of the method and the resulting point estimates of the density. A small bandwidth selection implies that averaging is done over relatively few observations in

the neighbourhood of x, hence rough estimates of $f(x)$ (*undersmoothing*). On the other hand, if a large h is chosen, this may mask essential details in the data set. The 'masking' property is referred to as *oversmoothing*. In particular, as $h \to 0$, the variance of the kernel estimator increases but its systematic error decreases (and vice versa, as h increases). This affects the resultant estimate of the mode.

The effectiveness of two kernels is investigated here with regard to estimating the density of the LS estimator near its maximum. That is:

1. The Epanechnikov kernel (EP)

$$\mathbb{K}(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{if } |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

2. The triangular kernel (TR)

$$\mathbb{K}(x) = \begin{cases} 1-|x| & \text{if } |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

The two kernels have high efficiency in minimizing the error when approximating the true density (Silverman 1986, p. 43). Furthermore, their finite support region $[-1, 1]$ suits the interest in the central part of the distribution.

Figure 3.3 illustrates the effect of the smoothing parameter on the EDF. The left panels show the EDF determined using EP whereas the right, using TR ($N = 10000$). A comparison of the top ($\alpha = 0.50$) and bottom ($\alpha = 0.95$) panels indicates that the same bandwidth cannot be ideal for all $\alpha \in \Omega$. For example, the EDF is undersmoothed when $h = 0.005$ or $h = 0.01$ in the top panels. However, as shown in the bottom panels, the EDF reasonably approximates the PDF when the same values are used. The technique for dealing with this limitation - choice of h - is discussed in the next section.

3.3.2 Numerical implementation

The basic kernel estimator defined in (3.5) applies a single smoothing parameter over the entire set of N observations. Given the skewed nature of the distribution of the LS estimator when

$|\alpha|$ is close to unity, the kernel estimator can be inadequate for determining the overall density over all x. However, interest here is in the central part of the distribution and not necessarily in achieving a good fit over the entire set $\{X_i\}$. The aim is to only ascertain how well (3.5) is suited to determining the mode of the density of the LS estimator.

In practical implementation of the kernel estimators, the choice of smoothing parameter can be done by utilizing the data or subjectively. Methods that rely on the data are called *bandwidth selectors*.

The subjective method involves visual inspection of the smoothing achieved with a range of bandwidth parameters. For example, in Figure 3.3 the behaviour of the EDF in each panel is invariably similar, at the same values of h, for the two kernels. It can be expected that the *ideal* or *'optimal' bandwidth* denoted as h_{opt}, for the kernels in (3.6) and (3.7), falls between $[0.005, 0.075]$. In fact, effective smoothing is achieved when $h_{\text{opt}} \approx 0.05$ for $\alpha = 0.5$, and $h_{\text{opt}} \approx 0.02$ for $\alpha = 0.95$.

The subjective method is viable if some structure of the data is known i.e. the approximate location of the mode - implied for instance by the values of mean and median statistics of the LS estimator of α, if known. However, the exercise can be daunting and time consuming. Thus, a more general procedure utilizing automatic bandwidth selectors is preferred because its more attractive than the subjective method.

The theoretical treatment of bandwidth selection, on which the automatic selectors are based, is to peg the choice to the asymptotic global performance of the kernel estimator. For example, using the *mean-integrated-squared-error* (MISE) criterion (Wand & Jones 1995), as a measure of the discrepancy between the EDF and $f_{\text{LS}}(x)$, the optimal bandwidth that minimizes MISE$\{\hat{f}(x;h)\}$ asymptotically is given by

$$\boxed{h_{\text{opt}} = \left[\frac{\mathbb{A}}{N\,\mathbb{B}^2\mathbb{C}}\right]^{\frac{1}{5}},} \tag{3.8}$$

where

$$\mathbb{A} = \int_x \mathbb{K}(x)^2 dx, \qquad \mathbb{B} = \int_x x^2 \mathbb{K}(x) dx \quad \text{and} \quad \mathbb{C} = \int_x f''(x)^2 dx. \tag{3.9}$$

$\mathbb{C}$ is a measure of the total curvature of the underlying true density $f_{\text{LS}}(x)$. It indicates the degree of difficulty in the estimation of the EDF, and is large if the distribution shows high skewness or

kurtosis (Silverman 1986; Wand & Jones 1995). The effect of the size of $\mathbb{C}$ on h_{opt} is apparent by inspection of (3.8).

To avoid mathematical and computational technicalities on which some[4] automatic selectors rely, simpler but generally acceptable selectors are considered here. These bandwidth selectors utilize '*normal-scale rules*' to assign a value to the term $\mathbb{C}$, through reference to some standard family of densities.

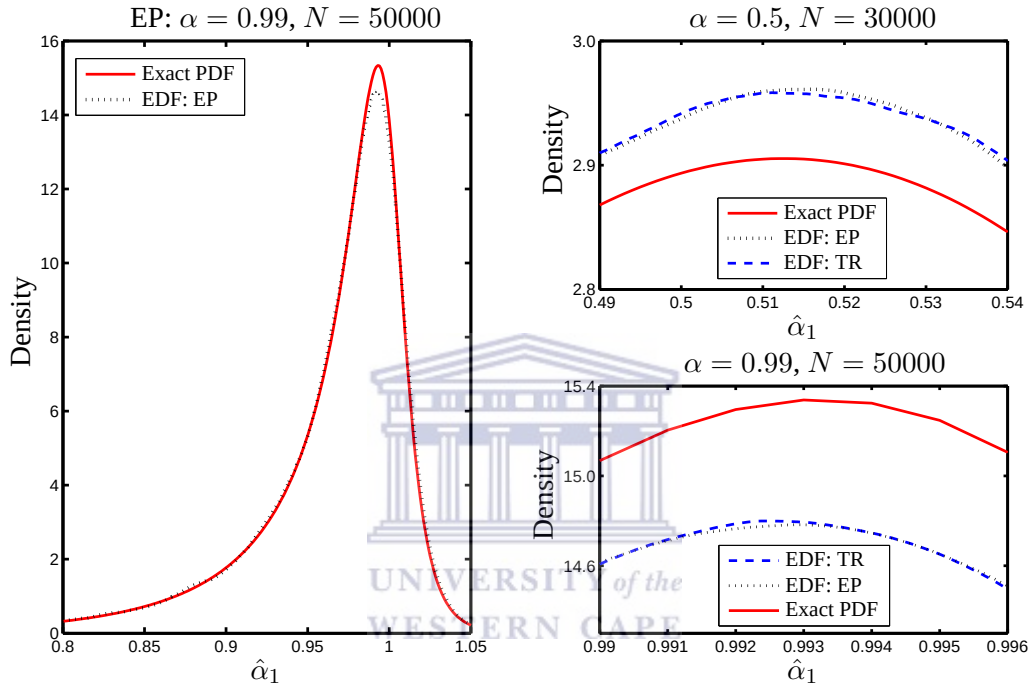


Figure 3.4: Left panel: EDF using EP when $\alpha = 0.99$. Right panels: EDF using EP and TR near the modal value when $\alpha = 0.5$ (top) and $\alpha = 0.99$ (bottom). In all the cases $T = 40$ in Model 1.

Table 3.2: Bandwidth values and the corresponding mode estimates using EP and TR (right panels in Figure 3.4). The exact values of the mode statistic of the LS estimator for $\alpha = 0.5$ and $\alpha = 0.99$ are 0.513 and 0.9933 respectively.

Kernel	$\alpha = 0.50$	
	Bandwidth	Mode
Epanechnikov	0.04081	0.516
Triangular	0.04483	0.511

Kernel	$\alpha = 0.99$	
	Bandwidth	Mode
Epanechnikov	0.0098	0.9922
Triangular	0.0107	0.9929

For the normal density, $\mathbb{C} \approx 3/8\pi^{1/2}\sigma^{-5}$ (Wand & Jones 1995). The parameter σ is estimated from the minimum of the sample standard deviation denoted by s, and a more robust measure

[4]Selectors based on methods such as least-squares cross-validation, likelihood cross-validation, test graph, etc. (Silverman 1986; Wand & Jones 1995) may be superior but are not considered. It is possible that these could yield better estimates of the mode. However, this is beyond our scope.

of scale, namely the standardized sample interquartile range (IQR), which guards against the effect of outliers in the sample set $\{X_i\}_{i=1}^N$. In particular,

$$\hat{\sigma} = \min\{s\,, \text{IQR}\} \quad \text{where} \quad \text{IQR} = \frac{Q_3 - Q_1}{\Phi^{-1}(0.75) - \Phi^{-1}(0.25)} = 0.741[Q_3 - Q_1]. \tag{3.10}$$

In (3.10), Φ^{-1} is the inverse standard normal CDF whereas Q_1 and Q_3 are the first and third quartiles determined from the data.

The quantities $\mathbb{A}$ and $\mathbb{B}$ in (3.9) can be obtained trivially for the kernels. It follows that the 'normal-scale' bandwidths for EP and TR are

$$\boxed{h_{\text{opt}}^{\text{EP}} = 2.345\,\hat{\sigma}\,N^{-\frac{1}{5}} \qquad \text{and} \qquad h_{\text{opt}}^{\text{TR}} = 2.576\,\hat{\sigma}\,N^{-\frac{1}{5}},} \tag{3.11}$$

respectively. Both bandwidths in (3.11) decrease as $|\alpha|$ tends to unity. This property follows from that of the variance of the LS estimator: see Figure 1.4 on page 9. They also decrease as T and N increase.

Table 3.3: Comparison of estimates of the mode of the EDF $\hat{f}(x;h)$ in (3.5) obtained using EP and TR with the exact mode obtained using numerical integration.

(a) $T = 50$.

	$N = 10000$		$N = 50000$		Imhof
α	EP	TR	EP	TR	exact
0.00	-0.003	-0.003	0.000	0.000	**0.000**
0.05	0.048	0.048	0.053	0.053	**0.051**
0.10	0.102	0.102	0.106	0.106	**0.102**
0.15	0.163	0.163	0.155	0.155	**0.153**
0.20	0.208	0.208	0.204	0.204	**0.204**
0.25	0.250	0.251	0.258	0.258	**0.255**
0.30	0.303	0.303	0.308	0.308	**0.306**
0.35	0.355	0.355	0.357	0.357	**0.357**
0.40	0.412	0.412	0.408	0.408	**0.408**
0.45	0.469	0.469	0.457	0.457	**0.459**
0.50	0.515	0.514	0.508	0.508	**0.510**
0.55	0.556	0.556	0.564	0.564	**0.561**
0.60	0.614	0.614	0.607	0.607	**0.612**
0.65	0.662	0.662	0.661	0.661	**0.663**
0.70	0.714	0.714	0.715	0.715	**0.714**
0.75	0.765	0.765	0.765	0.765	**0.764**
0.80	0.812	0.812	0.813	0.813	**0.815**
0.85	0.861	0.861	0.864	0.864	**0.865**
0.90	0.915	0.915	0.912	0.912	**0.915**
0.93	0.940	0.940	0.942	0.942	**0.944**
0.95	0.960	0.960	0.960	0.960	**0.962**
0.97	0.976	0.976	0.976	0.976	**0.978**
0.99	0.991	0.991	0.992	0.992	**0.993**
0.995	0.995	0.995	0.996	0.996	**0.997**
0.999	0.999	0.999	0.999	0.999	**0.999**

(b) $T = 100$.

	$N = 10000$		$N = 50000$		Imhof
α	EP	TR	EP	TR	exact
0.00	-0.002	-0.002	-0.002	-0.002	**0.000**
0.05	0.056	0.056	0.053	0.053	**0.051**
0.10	0.096	0.096	0.106	0.106	**0.101**
0.15	0.149	0.149	0.151	0.151	**0.152**
0.20	0.198	0.198	0.207	0.207	**0.202**
0.25	0.248	0.248	0.254	0.254	**0.253**
0.30	0.295	0.295	0.305	0.305	**0.303**
0.35	0.349	0.349	0.354	0.354	**0.354**
0.40	0.402	0.402	0.406	0.406	**0.404**
0.45	0.451	0.451	0.457	0.457	**0.455**
0.50	0.501	0.501	0.506	0.506	**0.505**
0.55	0.556	0.556	0.560	0.560	**0.556**
0.60	0.607	0.607	0.610	0.610	**0.606**
0.65	0.658	0.658	0.658	0.658	**0.656**
0.70	0.709	0.709	0.708	0.708	**0.707**
0.75	0.757	0.757	0.758	0.758	**0.757**
0.80	0.810	0.810	0.809	0.809	**0.808**
0.85	0.859	0.859	0.859	0.859	**0.858**
0.90	0.908	0.908	0.906	0.906	**0.908**
0.93	0.935	0.935	0.936	0.936	**0.938**
0.95	0.955	0.955	0.957	0.957	**0.958**
0.97	0.976	0.976	0.976	0.976	**0.977**
0.99	0.991	0.991	0.992	0.992	**0.992**
0.995	0.996	0.996	0.996	0.996	**0.996**
0.999	0.999	0.999	0.999	0.999	**0.999**

Figure 3.4 compares the EDF (dotted line) obtained using the optimal bandwidths in (3.11) to the exact density (continuous line) $f_{\text{LS}}(x)$ obtained using Imhof's technique ($T = 40$, Model 1). The left panel shows the effectiveness of the method over the entire domain using EP when $\alpha = 0.99$ and $N = 50000$. The 'normal scale' bandwidths are also noted to be effective for skewed distributions even for relatively small values of N. The top ($\alpha = 0.50$, $N = 30000$) and bottom ($\alpha = 0.99$, $N = 50000$) panels on the right compare $f_{\text{LS}}(x)$ and the EDF near the modal values. Results are summarized in Table 3.2. It is clear from the graphs that although the probability density is systematically over- or underestimated, the mode is nonetheless reasonably well-determined.

To improve the scatter in the values obtained for the mode using either kernel, the following strategy is proposed:

1. Obtain the EDF in the vicinity of the mode using (3.6) or (3.7).

2. Establish the final estimate of the mode by first fitting a low-order polynomial to the kernel estimate (smoothing), and then solving for the point of zero derivative.

To increase the efficiency of the subroutines in the first step above, the EDF is determined at 200 data points within a well-chosen range. The range is set about $\ddot{\alpha}^* \pm 0.5s(X)$ where $\ddot{\alpha}^*$ is a preliminary mode estimate determined from the histogram of $X \equiv \{X_i\}_i^N$ and $s(X)$ is the sample standard deviation. For sufficiently large N, the median of X can also be used for $\ddot{\alpha}^*$, since the median is always close to the mode in the present application. If the histogram is used, caution should be taken when determining the number of bins. In the second step, a third order polynomial is used for $|\alpha| \leq 0.9$ whereas a fourth order polynomial is used for $|\alpha| > 0.9$.

The results are summarized in Table 3.3 for two sample sizes, $T = 50$ and $T = 100$. The findings are given for $N = 10000$ and $N = 50000$. The results agree quite well with the exact mode obtained using Imhof's technique, shown in bold. The differences, which are of the order of $|0.010|$ and $|0.005|$ for $N = 10000$ and $N = 50000$ respectively, are random rather than systematic and decrease with increasing T and N in all the models. Before smoothing is done, the differences are larger than those reported. For example, for the cases of N considered, they are of the order of $|0.021|$ and $|0.019|$ when $T = 50$ and $|0.015|$ for both kernels when $T = 100$.

The question that may arise from the above discussion is; *'why are the bandwidth selectors in (3.11) adequate?'* Intuitively, when $|\alpha|$ is close to unity, the curvature (captured by $\mathbb{C}$ in Equation 3.8) of the underlying density of the LS estimator is large, hence a little smoothing is optimal *viz.*, a small value of h is appropriate. This is ensured by the factor $\hat{\sigma}$ in (3.11) since it decreases as $|\alpha|$ tends to 1. When the underlying function has low curvature, a larger bandwidth is called for. This is so when $|\alpha|$ is close to zero. In this case, $\hat{\sigma}$ is larger and hence the estimation problem is easier.

3.3.3 Accuracy vs efficiency

MATLAB software is array-based and notoriously slow in executing loops. Calculations were therefore vectorized for improved efficiency. The program code for the procedure explained above is provided in Appendix C (Algorithm 14).

Table 3.4 compares the average time taken to compute the mode statistic of the LS estimator when using simulation (given N) to that when using the numerical integration method. It is comparable to Table 3.1 on page 35, for the mean and/or median statistic(s).

Table 3.4: Comparison of the average computation time (in seconds) of the mode statistic; numerical integration vs simulation method.

(a) Numerical integration

T\ Model	1	2	3
50	0.36	0.33	0.31
100	1.66	1.54	1.49
150	4.53	4.51	4.35

(b) Simulation

$T\backslash N$	10000	20000	50000	60000	80000
50	1.8	5.5	29.9	43.5	79.1
100	2.4	6.7	33.1	47.1	83.7
150	3.0	8.0	36.6	53.6	89.0

For the simulation method, the computation time increases as T (more loops) and N increase. Most of the time relates to the generation of LS estimates. Whereas the mean and median are determined almost instantaneously, the kernel implementation to determine the mode takes relatively more time.

It follows from Tables 3.3 and 3.4 that, the numerical integration method may be preferred to simulation on the basis of accuracy and computational efficiency. For the former, in contrast to the case of the mean and median statistics [Table 3.1(a)], the average computation time of the mode statistic decreases slightly from Model 1 to 3.

3.4 The final estimators

3.4.1 The estimators

The conditions leading to the location-based estimators of α in Model j $(j = 1, 2, 3)$ can be written as as follows.

$$\begin{aligned}
\mathbb{E}_\alpha\,[\,[\bar{\alpha}_j - g(\alpha)]^2\,] &\leq \mathbb{E}_\alpha\,[\,[\bar{\alpha}_j - g(\alpha')]^2\,] &&: \text{mean-unbiasedness,} \\
\mathbb{E}_\alpha\,[\,|\tilde{\alpha}_j - g(\alpha)|\,] &\leq \mathbb{E}_\alpha\,[\,|\tilde{\alpha}_j - g(\alpha')|\,] &&: \text{median-unbiasedness,} \\
\text{Pr}_\alpha\,(|\,\ddot{\alpha}_j - g(\alpha)\,| < \epsilon) &\leq \text{Pr}_\alpha\,(|\,\ddot{\alpha}_j - g(\alpha')\,| < \epsilon) &&: \text{mode-unbiasedness,}
\end{aligned} \tag{3.12}$$

for all $\alpha' \neq \alpha; \alpha', \alpha \in \Omega^*$. In (3.12), $\mathbb{E}_\alpha$ denotes expectation given that α is the true parameter, the g's are the location functions whereas $\bar{\alpha}_j$, $\tilde{\alpha}_j$ and $\ddot{\alpha}_j$ respectively are the statistics of interest.

The method of obtaining the location-based estimators utilizes the inverse function g^{-1} of g. If g is unique - i.e. $g^{-1}(g(\alpha)) = \alpha$ - and monotone on Ω^*, given the relations in (3.12), the estimators $\hat{\alpha}_{\text{UE}}$ for α to be used are:

$$\begin{aligned}
\mathbb{E}_\alpha\,[\,[\hat{\alpha}_{\text{UE}} - \alpha]^2\,] &\leq \mathbb{E}_\alpha\,[\,[\hat{\alpha}_{\text{UE}} - \alpha']^2\,] &&: \text{mean-unbiasedness, or} \\
\mathbb{E}_\alpha\,[\,|\hat{\alpha}_{\text{UE}} - \alpha|\,] &\leq \mathbb{E}_\alpha\,[\,|\hat{\alpha}_{\text{UE}} - \alpha'|\,] &&: \text{median-unbiasedness, or} \\
\text{Pr}_\alpha\,(|\,\hat{\alpha}_{\text{UE}} - \alpha\,| < \epsilon) &\leq \text{Pr}_\alpha\,(|\,\hat{\alpha}_{\text{UE}} - \alpha'\,| < \epsilon) &&: \text{mode-unbiasedness,}
\end{aligned} \tag{3.13}$$

for all $\alpha' \neq \alpha; \alpha', \alpha \in \Omega^*$ (Ferguson 1967; Lehmann 1959, 1983; Andrews 1993, p. 144).

3.4.2 The mechanism

An explicit functional form of the location functions (g) in (3.12) does not exist. However, in preceding sections, it was shown that they can be established numerically, given α. For a given α and T, the median, mean and mode functions can be obtained respectively using (2.31), (2.32) and (2.33) (numerical integration method) or (3.1), (3.2) and the EDF defined in (3.5) (simulation method).

The characteristics of g in (3.12) vary with the location function, T and the model. It is interesting that, given the restricted space Ω for the true LDV coefficient α, their ranges (denoted by $\mathbb{D}$) differ. For example, when $T = 50$ in Model 2, $\mathbb{D} = (-0.989, 0.897]$ for the

mean function (Table B4), $\mathbb{D} = (-0.997, 0.914]$ for the median function (Table B5) and $\mathbb{D} = (-0.999, 0.944]$ for the mode function (Table B6). Similarly, the ranges of the functions in Model 3 are $(-0.989, 0.807]$, $(-0.997, 0.824]$ and $(-0.999, 0.861]$ as shown in Tables B7-B9, respectively.

Although the properties of g differ, the method of obtaining the final estimators of α is the same for all the location functions, model and T. By recalling that, in general for $\alpha \in \Omega$, $\alpha \neq g(\alpha) = \bar{\alpha}_j$ (mean), $\alpha \neq g(\alpha) = \tilde{\alpha}_j$ (median) and $\alpha \neq g(\alpha) = \ddot{\alpha}_j$ $(j = 1, 2, 3)$ (mode), if the parameter space, Ω^*, of α is unbounded the estimators $\hat{\alpha}_{\text{UE}}$ of α can be obtained as

$$\boxed{\hat{\alpha}_{\text{UE}} = g^{-1}(\cdot),} \tag{3.14}$$

where $g^{-1}(\cdot)$ satisfies $g^{-1}(g(\alpha)) = \alpha$ for all $\alpha \in \Omega^*$.

3.4.3 Empirical use of $\hat{\alpha}_{\text{UE}}$

The function $g = g(\alpha)$ summarizes the location properties of the LS estimator given that α is the true parameter. Equation (3.14) utilizes the LS estimator of α in Model j $(j = 1, 2, 3)$, denoted by $\hat{\alpha}_{\text{LS}} \equiv \hat{\alpha}_j$ $(j = 1, 2, 3)$.

Given $\alpha \in \Omega^*$, Tanizaki (2000) used (3.14) to obtain[5] the 'mean-unbiased' estimators of α in Models 1 to 3.

Restricted parameter space

When the parameter space of α is bounded - Ω - Andrews (1993) gave a method of restricting the $\hat{\alpha}_{\text{UE}}$ to $[-1, 1]$. The final estimators are given by

$$\boxed{\hat{\alpha}_{\text{UE}} = \begin{cases} 1 & \text{if} \quad \hat{\alpha}_{\text{LS}} > g(1) \\ g^{-1}(\alpha_{\text{ES}}) & \text{if} \quad g(-1) < \hat{\alpha}_{\text{LS}} \leq g(1) \\ -1 & \text{if} \quad \hat{\alpha}_{\text{LS}} \leq g(-1), \end{cases}} \tag{3.15}$$

[5]Intentionally, this terminology is avoided in order to circumvent a possible confusion between 'decision theoretic unbiasedness' and the usual 'unbiasedness in expectation'.

where $g^{-1} : (g(-1),\, g(1)\,] \rightarrow (-1,\, 1]$ such that $g^{-1}(g(\alpha)) = \alpha$ for $\alpha \in \Omega$ and

$$g(-1) \;=\; \lim_{\alpha \rightarrow -1} g(\alpha). \tag{3.16}$$

Since $\Omega = (-1, 1)$ in Model 1, it follows from (3.16) that $g(1) = \lim_{\alpha \rightarrow 1} g(\alpha)$.

The procedures for obtaining $\hat{\alpha}_{\text{UE}}$ are *exact*: they take into account the form of the exogenous regressor $\mathbf{X}$ in (1.1) for any T. They are available if g is uniquely defined and monotonic on Ω. Numerical evaluation of the location functions shows them to be increasing in their domain: see Tables B1 to B9, however an analytic proof of this fact for $\alpha \in (-1, 1]$ and $T \geq 4$ in all the models is not available in literature at present (Andrews 1993, p. 146).

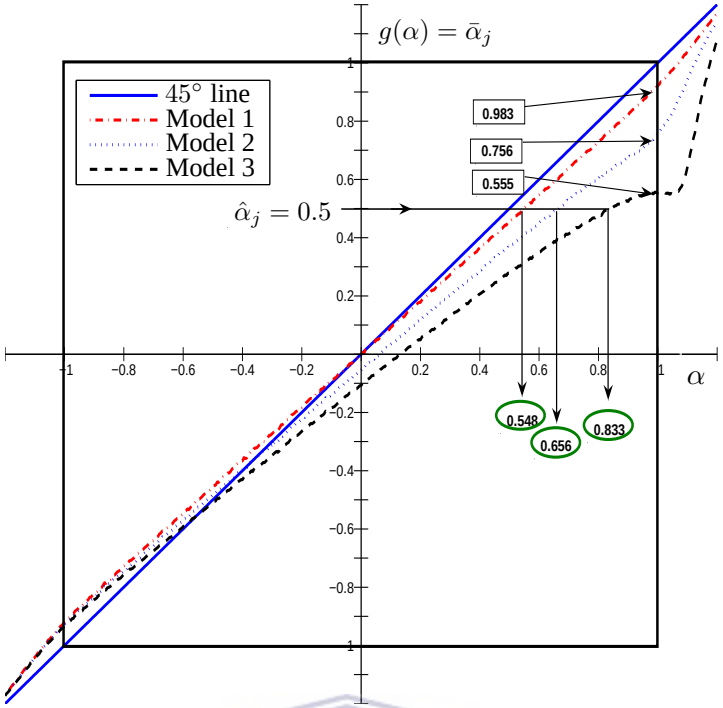
Figure 3.5 illustrates the empirical use of (3.15). For visual clarity, a sample of size $T = 20$ is used to demonstrate the behaviour of the mean and median functions on Ω. Each point on the location functions represents the mean or median of $N = 40000$ values of the LS estimator, determined using simulation. Random initial conditions are used except for $\alpha = 1$ in Model 1, which is conveniently excluded.

The 45° line indicates the case if the LS estimator were indeed unbiased. For a given α and model, the discrepancy between the functions and the line shows the extent of the bias. The left and right edges of the outer-square-box show the extent of the parameter space of α. The intersection of the functions and these edges determines the domain of their inverse functions ($\mathbb{D}$).

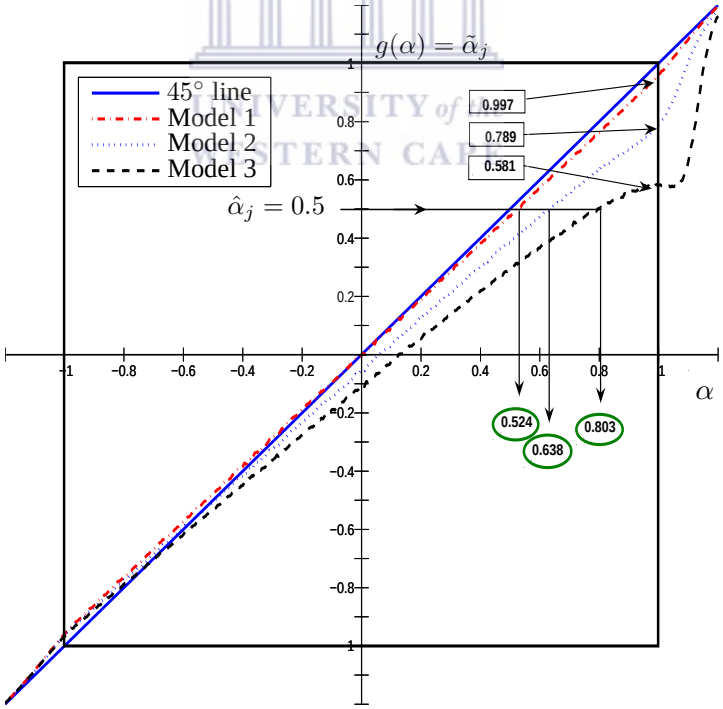
To illustrate: from Figure 3.5(b), given Model 2 where $\mathbb{D} = (-0.997, 0.789]$, the method of obtaining the final estimator is as follows:

1. For any value of $\hat{\alpha}_2 > 0.789$ the median-based estimator of α is $\hat{\alpha}_{\text{UE}} = 1$.
2. For any value of $\hat{\alpha}_2 \leq -0.997$ the corresponding final estimator is $\hat{\alpha}_{\text{UE}} = -1$.
3. For any intermediate value $\hat{\alpha}_2 \in \mathbb{D}$, the median-based estimator of α is obtained by ascertaining the value of α at which $g(\alpha) = \tilde{\alpha}_2$. As an example, if $\hat{\alpha}_2 = 0.5$ then $\hat{\alpha}_{\text{UE}} = 0.627$.

Similarly, from Figure 3.5(a), the mean-based estimator of α given that $\hat{\alpha}_j = 0.5$ $(j = 1, 2, 3)$ can be obtained as 0.548, 0.656 and 0.833 for models 1, 2 and 3, respectively.



(a) Mean function of the LS estimator in different models ($T = 20$).



(b) Median function of the LS estimator in different models ($T = 20$).

Figure 3.5: Illustration of the empirical use of $\hat{\alpha}_{\text{UE}}$ in (3.15). The mean- and median-based estimators of α given that the LS estimator $\hat{\alpha}_j = 0.5$ $(j = 1, 2, 3)$ are given by the values inside the ellipses.

The procedures outlined above can be used with Tables B1 to B9 given in Appendix B. For values of α and T that are not given in the tables, interpolation between the values provided is adequate to give sufficient accuracy for most applications. Detailed tabulations are also supplied in the electronic version of this work.

Intuitively, in (3.13), for the condition of mean-unbiasedness (3.12), the mean-based estimator $\hat{\alpha}_{\text{UE}}$ is the value of α such that the mean of the LS estimator $\bar{\alpha}_j$ $(j = 1, 2, 3)$, corresponds to the true parameter (Tanizaki 2000); similarly for the median-based estimator derived from $\tilde{\alpha}_j$ (Andrews 1993) and for the mode-based estimator derived from $\ddot{\alpha}_j$.

When $T = 20$, the values of the mode-based estimator corresponding to $\hat{\alpha}_j$ in Figure 3.5 can be confirmed to be 0.476, 0.569 and 0.703 in Models $1, 2$ and 3, respectively. It is apparent that the estimator is characterized by smaller factors $|\hat{\alpha}_{\text{UE}} - \hat{\alpha}_{\text{LS}}|$ $(j = 1, 2, 3)$ than the median-based estimator. Furthermore, the factors for the latter are smaller than those resulting from the mean-based estimator.

Tanizaki (2000) gave comparative values for the mean- and median-based estimators given that $\hat{\alpha}_j = 0.5$ while assuming fixed initial conditions for the series. However, such assumptions are rarely justifiable in practice. Furthermore, when using the simulation method, the choice of a small N can lead to a slight inaccuracy in the results[6] when the performance of the estimators is in question. An accurate method of establishing the MSE of the location-based estimators is proposed in Chapter 4.

The consideration of:

1. a restricted parameter space for α,
2. assumptions about the initial conditions,
3. emphasis on the choice of a sufficiently large N when using the simulation method, and
4. a more effective method when analyzing the performance of the location-based estimators,

[6]Tanizaki (2000) used $N = 10000$, whereas $N = 40000$ LS values are used in Figure 3.5. See Figure 3.2 for the approximate standard errors.

forms the main distinction between the present approach and that of Tanizaki (2000).

The statistical properties and the performance of the LS, mean-, median- and mode-based estimators are determined and compared in Chapter 4.

Chapter 4

Comparative Analysis

Estimators are of limited value if one has no idea of their statistical properties. In comparing alternative estimators, statistical theory relies on their small and large sample properties relating to their mean and variance. Our concern is with small sample properties of the estimators derived from the three locational functions.

The essence of unbiasedness is that 'on average' i.e., in expectation, a point estimate is right on target. Although it is a desirable attribute it is not the sole criterion for evaluating the overall performance of an estimator. For instance, an unbiased estimator may not be based on a sufficient statistic. It is arguable that mean-unbiasedness is more desirable than 'optimality' based on other criteria involving e.g. the median or mode. This is because it is based on the mean function, which for given α utilizes all the information in the sampling distribution of the LS estimator, unlike the median and mode functions.

In addition, an important requirement for an estimator is that its sampling variance should be small so that large estimation errors are unlikely to occur. In fact, focusing on unbiasedness may preclude a tolerably biased estimator with a much smaller variance.

It is common practice to require a trade off between unbiasedness and minimum variance. A criterion that recognizes this possible trade-off is the *mean-squared error (MSE)*.

Definition 1 (**Mean-squared error (MSE)).** If $\hat{\theta}$ is an estimator of θ, the MSE of $\hat{\theta}$ is defined by

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta}-\theta)^2] \\ &= \text{Var}(\hat{\theta}) + [\,\mathbb{E}(\hat{\theta})-\theta\,]^2, \end{aligned} \tag{4.1}$$

where $\mathbb{E}$ and Var respectively denote expectation and variance, given the parameter θ. ■

When evaluating the performance of estimators, MSE is preferred to other criteria such as *mean-absolute error* ($\mathbb{E}|\hat{\theta}-\theta|$) because firstly, it is computationally easier to work with, and secondly its simple decomposition into variance and squared-bias in (4.1) allows simpler analysis and interpretation of performance. If $\hat{\theta}$ is unbiased, the squared bias term is equal to zero and $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$. A measure of the typical estimation error associated with $\hat{\theta}$ is given by its *root mean squared error (RMS)*.

Definition 2 (**Root mean squared error (RMS)).** The RMS of $\hat{\theta}$ is given by

$$\text{RMS}(\hat{\theta}) = \sqrt{\text{MSE}(\hat{\theta})}. \quad \blacksquare \tag{4.2}$$

The performance of the mean-, median- and mode-based estimators ($\hat{\alpha}_{\text{UE}}$) can be established and/or compared as follows.

1. Ascertaining their MSE or RMS.
2. Measuring the consequences resulting from their use by the losses resulting from incorrect estimates.
3. Investigating their confidence intervals.
4. Following from the latter two, ascertaining the power gain or losses resulting from the use of $\hat{\alpha}_{\text{UE}}$ for inference.

In comparing the alternative estimators of α, the goal is to answer the following questions:

Q1. What is the nature and extent of the bias exhibited by the LS estimator of α over Ω when various forms of exogenous regressors are included in an AR(1) series?

Q2. Over what region(s) in Ω are the different estimators $\hat{\alpha}_{\text{UE}}$ 'optimal' or adequate? That is

- Which of the estimators has least bias?
- Which estimator has least variance?
- What is the trade-off between bias and variance associated with $\hat{\alpha}_{\text{UE}}$?
- What are the error properties of the final estimators?

Q3. What are the implications of the exact confidence bounds for inference about the LDV coefficient in terms of power gains or losses?

The above questions are pertinent to practical applications particularly in the specification of time series regression models (Rudebusch 1992, 1993). The results give an indication of the behaviour of estimators in higher-order autoregressive processes (Andrews & Chen 1994; Fair 1996; Tanizaki 2000).

Section 4.1 seeks to answer Q1 by establishing the properties of the location functions. In section 4.2, investigations of the MSE and RMS properties of the estimators $\hat{\alpha}_{\text{LS}}$ and $\hat{\alpha}_{\text{UE}}$ when $\alpha \in \Omega$ are done, in view of Q2. A more accurate methodology which is based on the numerical integration technique as opposed to simulation is used. Section 4.3 deals with determination of confidence bounds mentioned in Q3.

4.1 Properties of the location functions

Instead of using the raw data of location functions summarized in Tables B1 to B9, it is more instructive to consider their deviations from the true parameter values. Apart from facilitating visual inspection, the deviations can be used in their own right to establish the rate of change of the location functions with respect to the true coefficient α: the *Jacobian* of the transformation. For fixed α and T, a statistic ∇ is defined by

$$\nabla = \text{Location statistic} - \alpha. \tag{4.3}$$

From (4.3), the following notation is adopted for the different location functions in Model j $(j = 1, 2, 3)$.

$$\begin{aligned} \nabla^{\text{A}} &= \bar{\alpha}_j - \alpha \qquad \text{: mean function } [g(\alpha) = \bar{\alpha}_j]. \\ \nabla^{\text{B}} &= \tilde{\alpha}_j - \alpha \qquad \text{: median function } [g(\alpha) = \tilde{\alpha}_j]. \\ \nabla^{\text{C}} &= \ddot{\alpha}_j - \alpha \qquad \text{: mode function } [g(\alpha) = \ddot{\alpha}_j]. \end{aligned} \tag{4.4}$$

The properties of ∇ follow those of the LS estimator. Figure 4.1 compares the properties of the location functions using the statistics in (4.4) when $T = 50$. Figure 4.1(a) shows the results for Model 1, and Figures 4.1(b) and 4.1(c) show the results for Models 2 and 3 respectively.

In Model 1, the symmetry observed about $\alpha = 0$ follows from the distribution of the LS estimator (Fuller 1976). Andrews (1993) noted the symmetry in the median function. In Models 2 and 3, the behaviour of ∇ is similar but differs from that shown in Model 1 by a characteristic asymmetry. It is clear that the bias of the LS estimator is more pronounced in these models than in Model 1. The following deductions can be made:

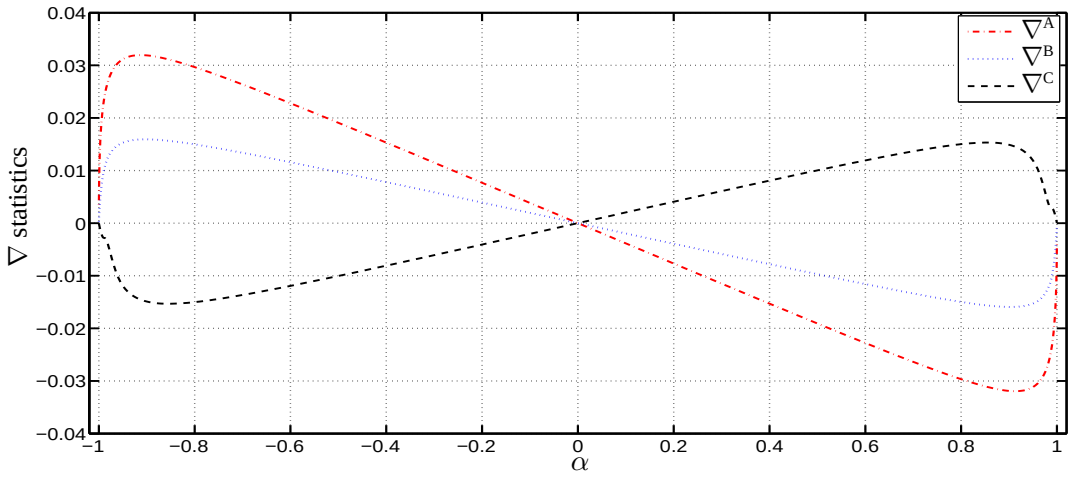
1. In all the models, as T increases $|\nabla|$ decreases: given α, the deviations of location statistic(s) from the true parameter α decrease.

2. The bias of the LS estimator is worst near $|\alpha| = 1$.

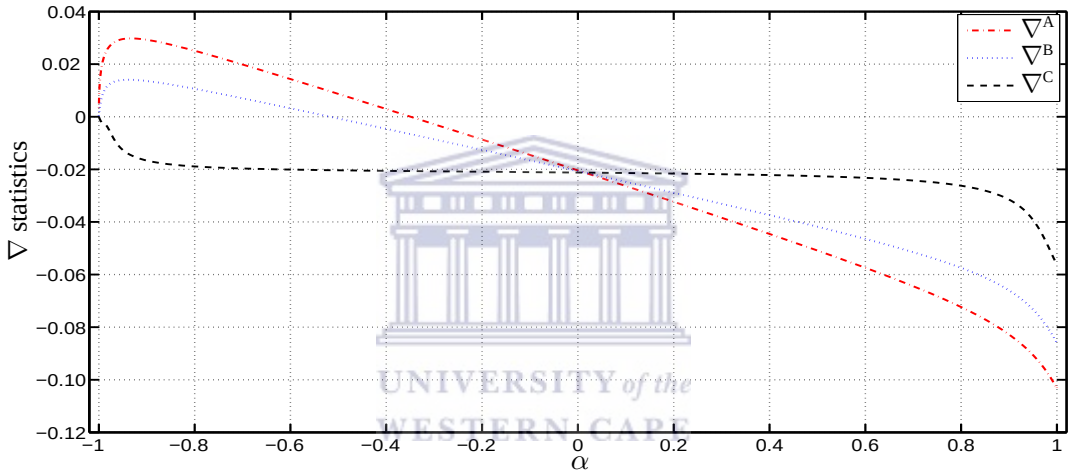
3. As α approaches -1, $|\nabla|$ increases steadily and subsequently decreases to zero in all the models.

4. Except for positive values of α in Model 1, for a given α and T the mode of the LS estimator is always less than the true parameter.

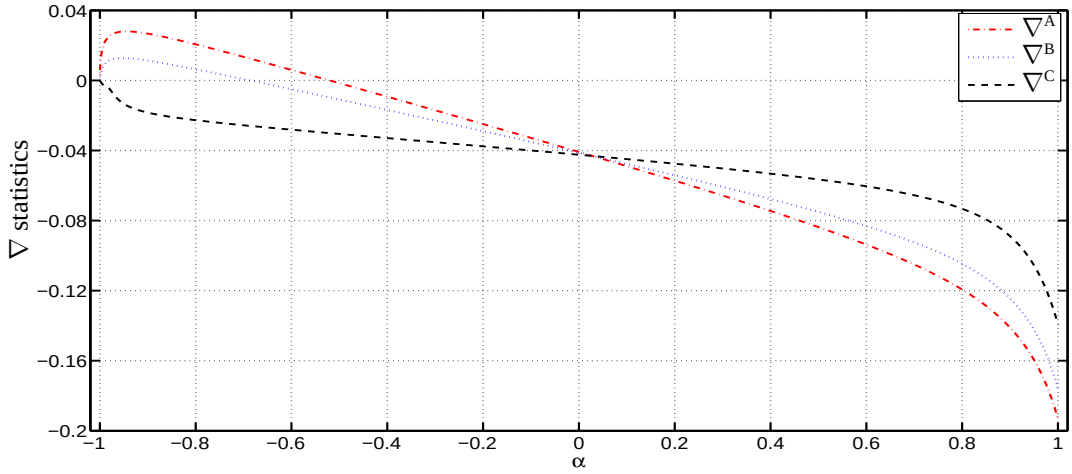
5. In Model 1, when $\alpha = 0$, the mean, median and mode of the distribution of the LS estimator are all equal. In Models 2 and 3, the mean and median functions intersect at $\alpha \approx 0.0141$ and $\alpha \approx 0.0289$ respectively ($T = 50$). The mode function intersects the other functions at very similar values. It follows that the distribution of the LS estimator is almost exactly symmetrical for α close to zero, but becomes skewed as $|\alpha| \to 1$.



(a) Properties of ∇ in Model 1.



(b) Properties of ∇ in Model 2.



(c) Properties of ∇ in Model 3.

Figure 4.1: Properties of the location functions in terms of the ∇ functions defined in (4.3) and (4.4) ($T = 50$).

6. For positive values of α, the absolute difference between the mode of the LS estimator and the true parameter is in general smaller than that of the mean and median. In Model 2, it is constant over much of the parameter space.

4.2 Analysis and computation of MSE

The MSE of the estimators can be obtained numerically or using simulation. Using simulation, a large number of replicates (N) are required for adequate precision. Numerical evaluation is preferable if the explicit form of their density can be evaluated. Since closed form expressions for the densities are not available, an alternative method of evaluation is proposed here. As shown later, the density functions of the location-based estimators may be characterized by both a discrete and continuous part. The MSE and RMS of the estimators are determined at acceptable accuracy levels and analyzed. Firstly, an introduction of the necessary notation and concepts is given.

4.2.1 MSE of the LS estimator $\hat{\alpha}_{\text{LS}}$

In Chapter 2, it was shown that the PDF of the LS estimator $f_{\text{LS}}(x)$ can be evaluated for fixed α and T. It is therefore easy to determine the MSE of the estimator. For consistency in notation, $\hat{\alpha}_{\text{LS}} \in \mathcal{S}$ is used to denote the LS estimator $\hat{\alpha}_j$ in Model j $(j = 1, 2, 3)$, where $\mathcal{S} = (-\infty, \infty)$ is the sample space of the LS estimator. From (2.24) and (4.1), it follows that

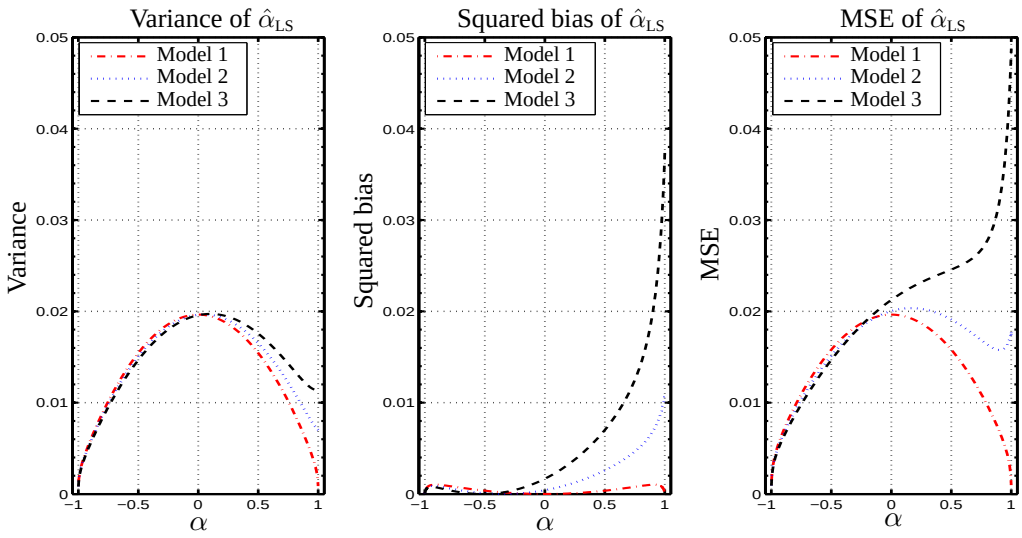
$$\text{MSE}(\hat{\alpha}_{\text{LS}}) = \text{Var}(\hat{\alpha}_{\text{LS}}) + [\mathbb{E}_{\alpha\text{LS}}(\hat{\alpha}_{\text{LS}}) - \alpha]^2, \tag{4.5}$$

where $\mathbb{E}_{\alpha\text{LS}}$ denotes expectation with respect to $\hat{\alpha}_{\text{LS}}$ when α is the true parameter. The variance of the LS estimator is given by

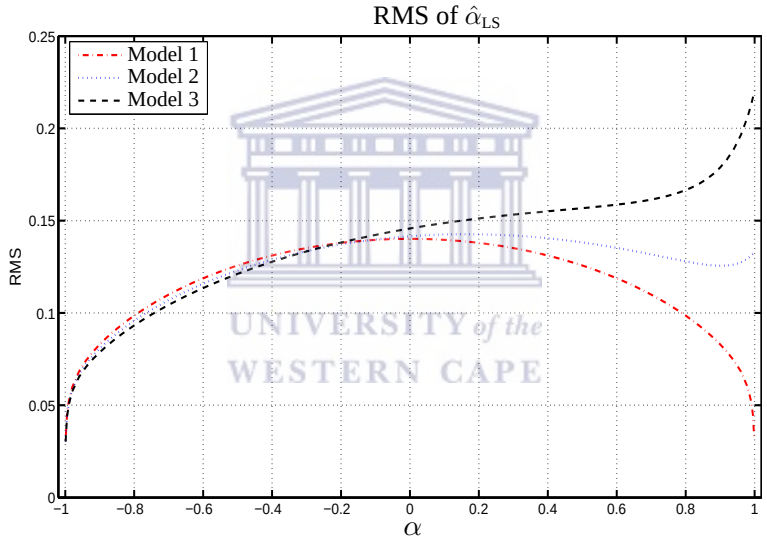
$$\begin{aligned} \text{Var}(\hat{\alpha}_{\text{LS}}) &= \mathbb{E}_{\alpha\text{LS}}[\hat{\alpha}_{\text{LS}}^2] - (\mathbb{E}_{\alpha\text{LS}}[\hat{\alpha}_{\text{LS}}])^2 \\ &= \int_{\mathcal{S}} \hat{\alpha}_{\text{LS}}^2 f_{\text{LS}}(\hat{\alpha}_{\text{LS}}) d\hat{\alpha}_{\text{LS}} - \left\{ \int_{\mathcal{S}} \hat{\alpha}_{\text{LS}} f_{\text{LS}}(\hat{\alpha}_{\text{LS}}) d\hat{\alpha}_{\text{LS}} \right\}^2 . \end{aligned} \tag{4.6}$$

Figure 4.2(a) shows the MSE and its decomposition into variance and squared bias in different models when $T = 50$. Table B15 summarizes the results.

In contrast with Model 1 and 2, the contribution of the squared bias to the MSE of the LS estimator in Model 3 is greater than that of the variance when $\alpha > 0.75$. In addition, in Model 3 the MSE is monotonically increasing on Ω. In all the models, the variance of the LS estimator



(a) Comparison of the variance, squared bias and the MSE of the LS estimator on the same scale ($T = 50$).



(b) RMS of the LS estimator when $T = 50$.

Figure 4.2: (a) Decomposition of the MSE of the LS estimator into variance and squared bias in different models. (b) Properties of the RMS of the LS estimator over Ω in different models.

is high near $\alpha = 0$ but decreases as $|\alpha|$ tends to unity. In general, the estimation error of the LS estimator as quantified by its RMS - shown in Figure 4.2(b) - worsens as α tends to 1 in Models 2 and 3 but improves as T increases. However, in Model 1 it decreases as $|\alpha| \to 1$. For example, near the unit root border the RMS is approximately 3%, 13% and 22% in Models 1, 2 and 3 respectively.

4.2.2 MSE of the estimators $\hat{\alpha}_{\text{UE}}$

To compute the MSE of the estimators $\hat{\alpha}_{\text{UE}}$ their density functions have to be evaluated. Since the location functions are increasing in α, the method of transformation can be used to relate the density of the LS estimator and that of the estimators $\hat{\alpha}_{\text{UE}}$. Given (3.15), it follows that

$$f_{\text{UE}}(\hat{\alpha}_{\text{UE}}) = f_{\text{LS}}(\hat{\alpha}_{\text{LS}}) \left| \frac{d\,\hat{\alpha}_{\text{LS}}}{d\,\hat{\alpha}_{\text{UE}}} \right|, \tag{4.7}$$

where the last term in (4.7) is the *Jacobian* of the transformation from $\hat{\alpha}_{\text{LS}}$ to $\hat{\alpha}_{\text{UE}}$ (to be discussed in section 4.2.3). $f_{\text{LS}}(\hat{\alpha}_{\text{LS}})$ can be determined numerically for a given value of α and T. Owing to the restriction of the estimators $\hat{\alpha}_{\text{UE}}$ to the parameter space $[-1, 1]$ in (3.15), their density functions include probability mass-points when α is near ± 1.

The mean of $\hat{\alpha}_{\text{UE}}$ can be obtained as

$$\begin{aligned} \mathbb{E}_{\alpha \text{UE}}(\hat{\alpha}_{\text{UE}}) \;=\; & \underbrace{\int_{\Omega} \hat{\alpha}_{\text{UE}} f_{\text{LS}}(\hat{\alpha}_{\text{LS}}) \left| \frac{d\,\hat{\alpha}_{\text{LS}}}{d\hat{\alpha}_{\text{UE}}} \right| d\,\hat{\alpha}_{\text{UE}}}_{\text{for} \quad g(-1) < \hat{\alpha}_{\text{LS}} \leq g(1)} + \underbrace{[-1]\ \Pr(\hat{\alpha}_{\text{UE}} = -1)}_{\text{for} \quad \hat{\alpha}_{\text{LS}} \leq g(-1)} \\ & + \underbrace{[+1]\ \Pr(\hat{\alpha}_{\text{UE}} = 1),}_{\text{for} \quad \hat{\alpha}_{\text{LS}} > g(1)} \end{aligned} \tag{4.8}$$

where $\Pr(\hat{\alpha}_{\text{UE}} = -1)$ and $\Pr(\hat{\alpha}_{\text{UE}} = +1)$ denote the probability that $\hat{\alpha}_{\text{UE}} = -1$ and $\hat{\alpha}_{\text{UE}} = 1$ respectively (the mass-point probabilities at ± 1), and $\mathbb{E}_{\alpha \text{UE}}$ denotes expectation with respect to $\hat{\alpha}_{\text{UE}}$ when α is the true parameter.

The MSE of $\hat{\alpha}_{\text{UE}}$ is given by

$$\text{MSE}(\hat{\alpha}_{\text{UE}}) = \text{Var}(\alpha_{\text{UE}}) + [\,\mathbb{E}_{\alpha \text{UE}}(\hat{\alpha}_{\text{UE}}) - \alpha]^2, \tag{4.9}$$

where the variance of $\hat{\alpha}_{\text{UE}}$ is

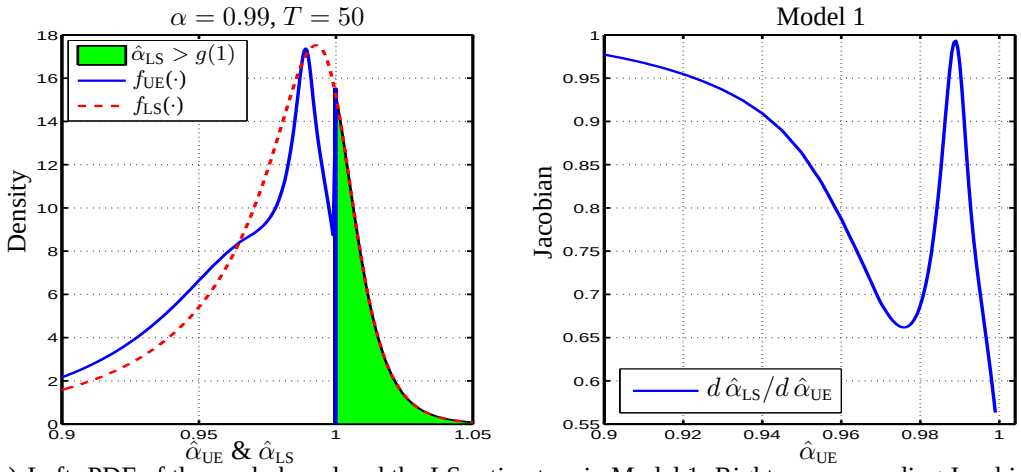
$$\text{Var}(\hat{\alpha}_{\text{UE}}) \;=\; \int_{\Omega} \hat{\alpha}_{\text{UE}}^2 f_{\text{UE}}(\hat{\alpha}_{\text{UE}}) d\hat{\alpha}_{\text{UE}} + \Pr(\hat{\alpha}_{\text{UE}} = \pm 1) - \{\ \mathbb{E}_{\alpha \text{UE}}(\hat{\alpha}_{\text{UE}})\ \}^2 . \tag{4.10}$$

The mass-point probabilities follow from the restriction on $\hat{\alpha}_{\text{UE}}$ in (3.15):

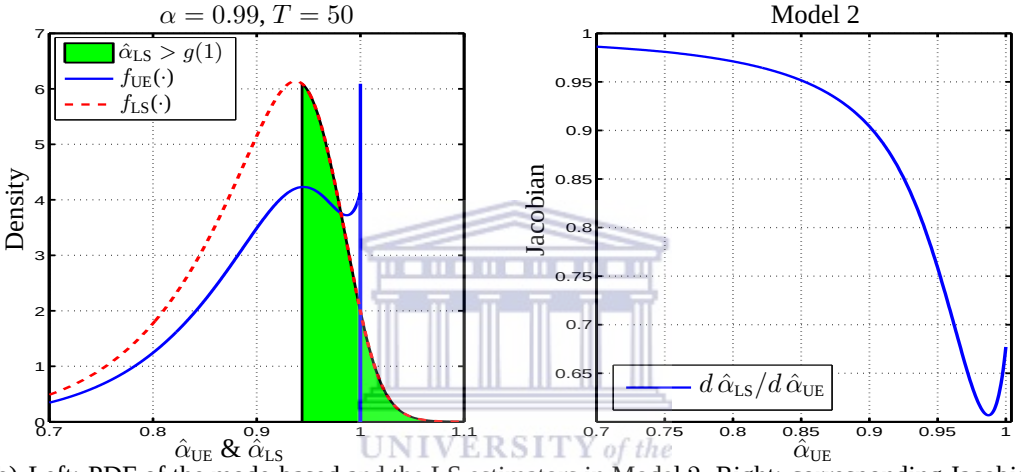
$$\Pr(\hat{\alpha}_{\text{UE}} = +1) \;=\; \int_{g(1)}^{\infty} f_{\text{LS}}(\hat{\alpha}_{\text{LS}})\, d\hat{\alpha}_{\text{LS}} \tag{4.11}$$

and

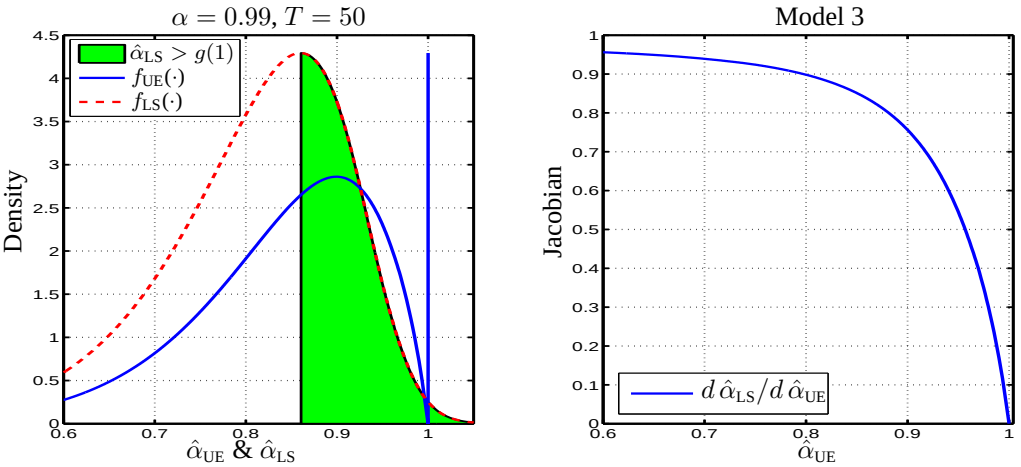
$$\Pr(\hat{\alpha}_{\text{UE}} = -1) \;=\; \int_{-\infty}^{g(-1)} f_{\text{LS}}(\hat{\alpha}_{\text{LS}})\, d\hat{\alpha}_{\text{LS}}. \tag{4.12}$$



(a) Left: PDF of the mode-based and the LS estimators in Model 1. Right: corresponding Jacobian.



(b) Left: PDF of the mode-based and the LS estimators in Model 2. Right: corresponding Jacobian.



(c) Left: PDF of the mode-based and the LS estimators in Model 3. Right: corresponding Jacobian.

Figure 4.3: Left panels: superimposed PDFs of the mode-based estimator $f_{\text{UE}}(\cdot)$ and the LS estimator $f_{\text{LS}}(\cdot)$ given that the true parameter $\alpha = 0.99$ and $T = 50$ in various models. The probability of the mass-point at $+1$ is shown by the shaded region. Right panels: corresponding Jacobians $|d\hat{\alpha}_{\text{LS}}/d\hat{\alpha}_{\text{UE}}|$ used to obtain the density of $\hat{\alpha}_{\text{UE}}$.

The PDF of $\hat{\alpha}_{\text{UE}}$ depends on the properties of the Jacobian. The latter depends on α, T, the model considered and the location function. For example, Figure 4.3 shows the superimposed densities of $\hat{\alpha}_{\text{UE}}$ (continuous line) for the mode-based estimator and that of the LS estimator $f_{\text{LS}}(\cdot)$ (dashed line) given that the true parameter $\alpha = 0.99$ and $T = 50$. The right panels show the varied features of the corresponding Jacobians used to obtain the density $f_{\text{UE}}(\cdot)$. The probability of the mass-point at $+1$ given in (4.11) is equal to the area of the shaded region. This gives rise to the peculiar spike in $f_{\text{UE}}(\cdot)$ at $\alpha = 1$. Figures 4.3(a)-(c) show the results for Models 1, 2 and 3, respectively. Figure 4.4 is the analogue of Figure 4.3(a) for the median-based estimator.

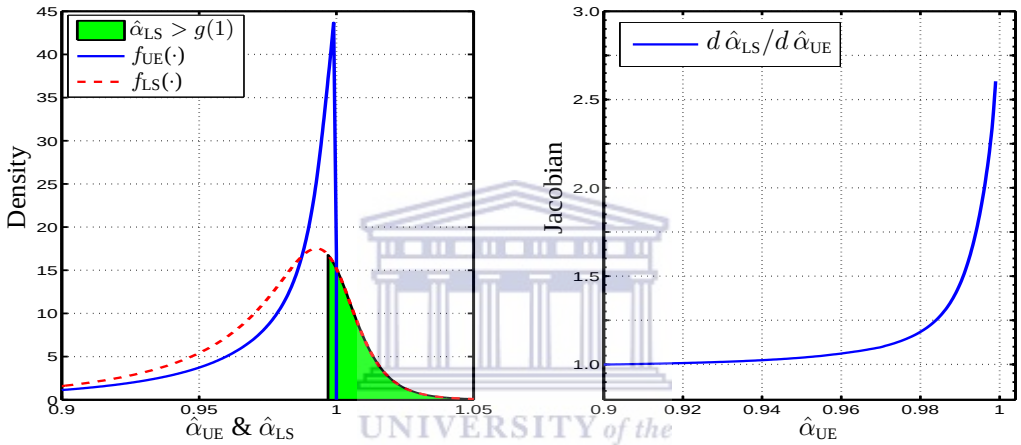


Figure 4.4: Left: PDFs of the median-based estimator $f_{\text{UE}}(\cdot)$ and the LS estimator $f_{\text{LS}}(\cdot)$ given that the true parameter $\alpha = 0.99$ and $T = 50$ in Model 1. Right: corresponding Jacobian.

Given the nature of the density of the LS estimator $f_{\text{LS}}(\hat{\alpha}_{\text{LS}})$, at least one of the point-mass probabilities is negligible when $|\alpha|$ is near 1. However, when $|\alpha|$ is very close to 0 both point-mass probabilities are $\leq 10^{-8}$ when[1] $T = 50$.

[1]However, for smaller sample sizes (say $T \leq 20$) both point masses may be non-negligible irrespective of the value of α.

4.2.3 Calculation of the Jacobians

Since the explicit form of the location functions g is not known, the Jacobian estimator can only be evaluated numerically.

Because the location functions show asymptotic behaviour[2] when α is near ± 1, it would seem plausible to model $g(\hat{\alpha}_{\text{UE}})$ by a logarithmic form, say

$$g(\hat{\alpha}_{\text{UE}}) = \log(\texttt{Const}) + \log(1 - \hat{\alpha}_{\text{UE}}), \tag{4.13}$$

for some constant `Const`. The Jacobian can then be obtained by analytic differentiation. However, investigations here showed that $g(\hat{\alpha}_{\text{UE}})$ is non-linear in the variable $\log(1 - \hat{\alpha}_{\text{UE}})$, with clearly different functional behaviour for different models/parameters.

Another option is the finite-difference approximation

$$\frac{dg(\hat{\alpha}_{\text{UE}})}{d\hat{\alpha}_{\text{UE}}} \approx \frac{\Delta g(\hat{\alpha}_{\text{UE}})}{\Delta \hat{\alpha}_{\text{UE}}}.$$

The approximation improves as $\Delta\hat{\alpha}_{\text{UE}} \to 0$. However, such an approximation is inadequate and inefficient when high accuracy is required. Better results are obtained from piecewise-polynomial approximations over a localized neighbourhood of points in the domain of $g(\cdot)$:

$$\frac{d\hat{\alpha}_{\text{LS}}}{d\hat{\alpha}_{\text{UE}}} \approx \left.\frac{dg_{\text{np}}(\cdot)}{d\hat{\alpha}_{\text{UE}}}\right|_{\alpha^0}, \tag{4.14}$$

where $g_{\text{np}}(\cdot)$ is a low-order polynomial fitted to `np` neighbouring points and α^0 is the midpoint of the neighbourhood.

The effectiveness of the method depends on the choice of `np`, the degree of the polynomial used, and the stepsizes over which the location functions are available. The fitting procedure described above is implemented using the *MATLAB* function `polyfit`[3]. The functions `polyder` and `polyval` are used to respectively obtain analytical derivatives of g_{np}, and to evaluate the derivatives numerically.

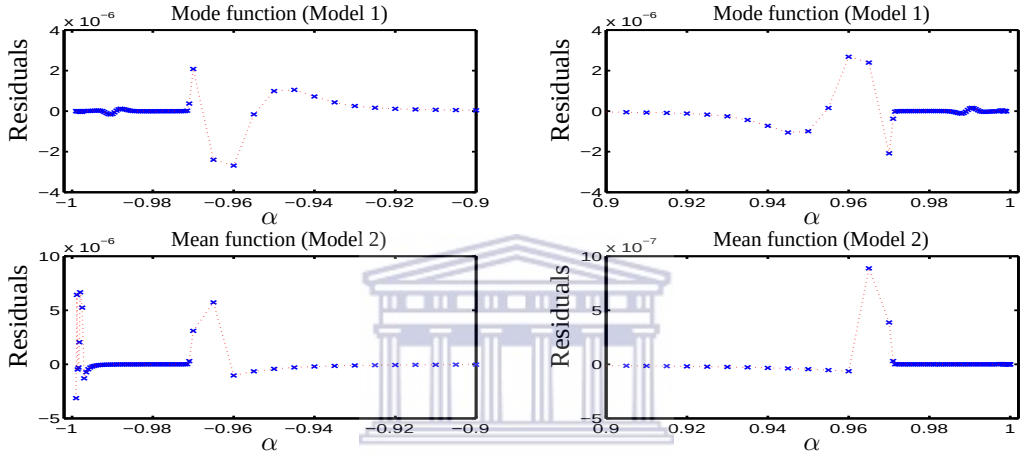
The choice of `np` and the degree of the polynomial relies heavily on the local behaviour of the location function about α^0. For $\alpha \in [-0.75, 0.75]$, a visual inspection shows the functions to be

[2] See the local behaviour of the ∇ functions in section 4.1.

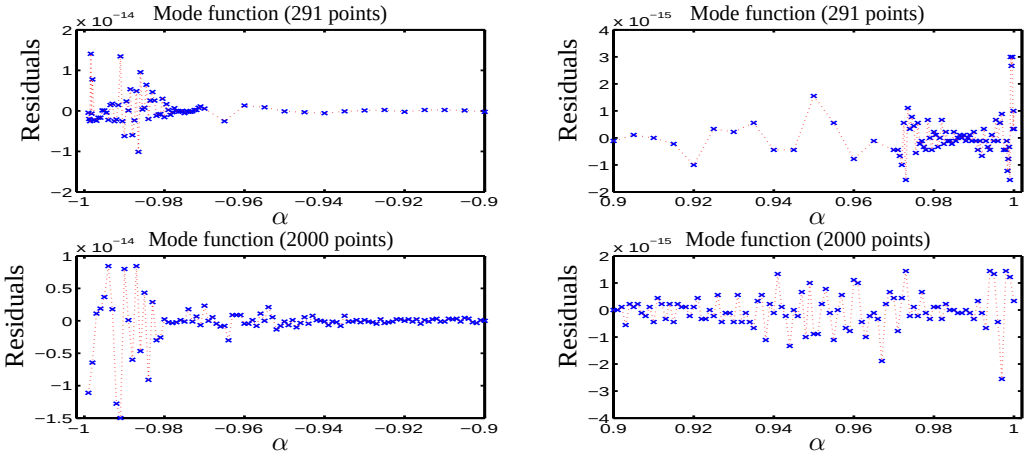
[3] *polyfit* generates the 'best fit' polynomial in the least square sense for a given data set.

almost linear. Therefore, a first- or second-order polynomial can be used. For other values in Ω, increasing np as well as the degree improves the numerical accuracy of the fitting procedure and the resultant derivatives.

At the end-points of Ω, polynomial approximations using the first or last np points over which the location functions are defined, can be used. However, the spread - the mesh - of points used should be adequate to capture the true behaviour of the underlying functions and hence guarantee the accuracy of the fitting algorithm. As such, np should be increased for $|\alpha| > 0.75$ to ensure a close fit especially near ± 1.



(a) Top panels: Residuals from a 5-point-polynomial fit for the mode function in Model 1 (285 points). Bottom panels: Same for the mean function in Model 2 (291 points).



(b) Residuals from a 3-point-polynomial fit for the mode function in Model 2. Top: residuals over 291 points. Bottom: effect of a cubic spline interpolation (2000 points) in the same domain

Figure 4.5: Residuals profile from the procedure described in the text when a second-order polynomial is used. Left panels: characteristics near the -1 boundary point. Right panels: characteristics near $+1$.

The ideal np and the degree of g_{np} can be determined either by visual inspection of the fit or by analyzing the properties of the resultant residuals. The top panels in Figure 4.5(a) show the residuals resulting from a second-order polynomial fit as described above, for the mode function in Model 1 [$\mathsf{np} = 5$ neighbouring point are used to establish $g_{np}(\alpha^0)$]. Similarly, the bottom panels show the results for the mean function in Model 2. The panels on the left and right illustrate the characteristics of the residuals near $\alpha = -1$ and $+1$ respectively. For both functions, the mesh consisted of 291 non-equidistant points in Ω.

It is evident from the character of the residuals that such a combination of np and polynomial degree is not appropriate for all α. Since the curvature of the location functions change over the domain of α, the nature of the residuals also varies. Irrespective of accuracy of the fit ($\approx 10^{-6}$), the patterns present in all panels indicate systematics that result from the choice of the parameters.

Near $\alpha = -1$, the rate of change of the mean function is greater than that of the mode and median functions in all models (see Figure 4.1 on page 54). A higher degree polynomial may therefore be suitable. However, use of e.g. a third degree polynomial over a 5-point-neighborhood is dubious. Furthermore, if a large np is chosen, this may result in poor representation of local behaviour. Although centering and scaling may improve the numerical properties of the resultant polynomials, the derivatives may be adversely affected.

The peculiar behaviour near $|\alpha| = 0.96$ in Figure 4.5(a) is due to the effect of unequal stepsizes (used on the domain of the location functions) on the fitting procedure. To eliminate such effects, as well as decrease the dependence of the fitting algorithm on the choice of np, the degree of polynomial used, and the value of α, a three-step procedure - which was found to ensure sufficient accuracy - is proposed. That is:

1. Establishing the location functions over a well-chosen[4] mesh of points in Ω.

2. Refining the data set obtained above to a large number of uniformly-spaced-points via a cubic spline interpolation.

[4]An appropriate mesh (unequally spaced) can be determined by studying the behaviour of the statistic ∇ over Ω (see Chapters 2 and 3).

3. Using $\texttt{np} = 3$ with a second degree polynomial to obtain piecewise polynomial approximations and their derivatives.

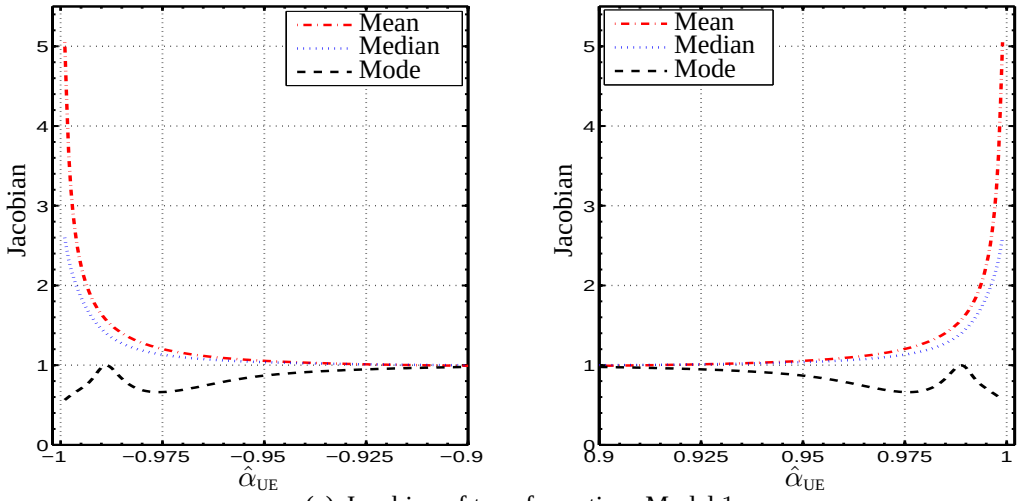
The refinement is effected using the *MATLAB* function `interp1` with `v5cubic` or `spline` options to cater for the unequally spaced points.

Figure 4.5(b) shows the residuals after using the procedure outlined above for the mode function in Model 2. In contrast to Figure 4.5(a), the top panels show that the choice of $\texttt{np} = 3$ reduces the effect arising from unequal step-sizes. The residuals mimic white noise. Their order of magnitude ($\approx 10^{-14}$) indicates the accuracy of the fit and hence the Jacobians. In fact, higher accuracy ($\approx 10^{-20}$) is noted for intermediate values of $|\alpha|$ near 0. The bottom panels show an improved profile of the residual as well as accuracy[5] after interpolation is effected (2000 points): see the scales of the vertical axes.

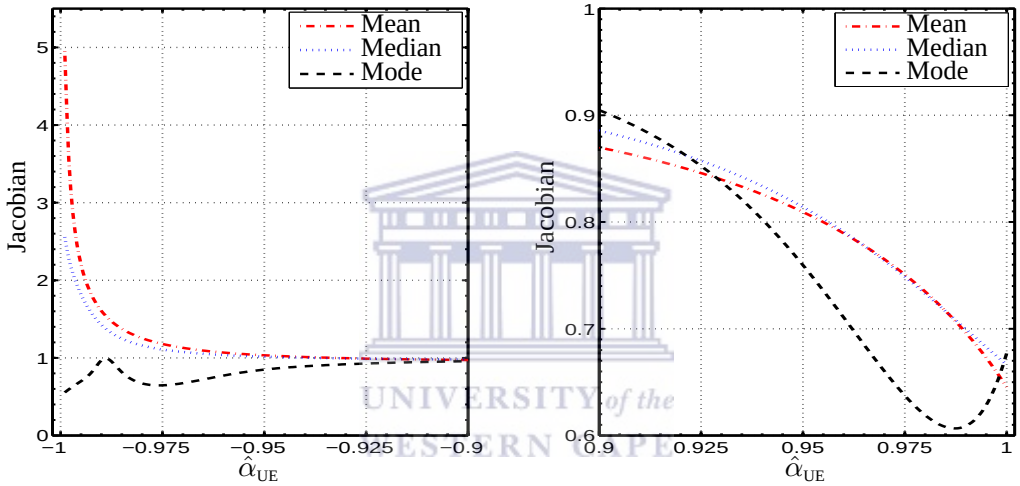
In order to avoid the noted numerical effects in the Jacobian computations, the location statistics must be computed at higher precision for the median and mode functions. The median and mode statistics of the LS estimator are affected more than the mean, to the choice of the optimization parameters used by `quad` or `quadl` when evaluating the integrals for the CDF and PDF in (2.13) and (2.24), respectively. A tolerance level of 1×10^{-14} is used for the Lobatto quadrature algorithm utilized by `quadl`. However, for the mean, a tolerance of 1×10^{-9} is sufficient to obtain accurate results. Of course, this has further-reaching implications for the computational expense of the mean of the LS estimator than for the median and mode. The bound on the truncation error is pre-specified at $B_{\text{trc}} = 10^{-15}$ in all cases.

The location functions are tabulated to high accuracy to ensure smoothness and eliminate spurious effects. In the case of the mean statistic, a point discontinuity in the density was noted when both the LDV coefficient and the LS estimator are equal to 1. Thorough investigations tracked this to the matrix $\mathbf{R}_\alpha$ given on page 16. In particular, by its definition, when the LDV coefficient equals unity, its columns and/or rows are not independent. Therefore, the eigenvalues have multiplicities ($\lambda = 0.500$). The eigenvectors are not independent as is assumed throughout. The result affects the mean, and normalization calculations in Models 2 and 3: the mean of the LS estimator and the area under its PDF are over- and under-stated respectively.

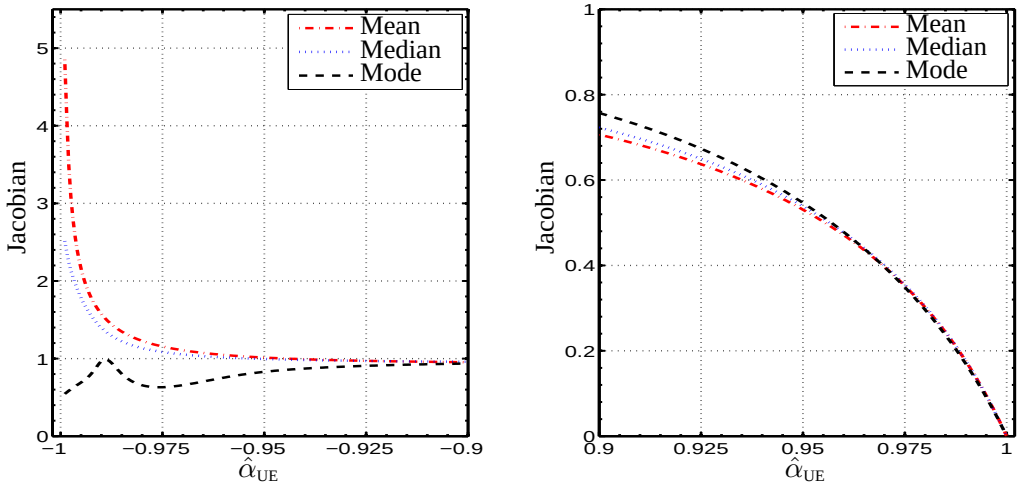
[5] Although the gain in accuracy may seem to be minimal, interpolation provides a **general procedure** for all the location functions. It also enables **lesser computational effort** whilst establishing the same.



(a) Jacobian of transformation: Model 1



(b) Jacobian of transformation: Model 2



(c) Jacobian of transformation: Model 3

Figure 4.6: Comparison of the Jacobian functions for the three models. In Figure 4.6(a), for Model 1, the Jacobian is shown over $\Omega \equiv (-0.999, 0.999)$ whereas in Figures 4.6(b) and 4.6(c) it is computed over $\Omega \equiv (-0.999, 1]$, for Models 2 and 3 respectively.

Figure 4.6 shows the Jacobians of the three location functions when $T = 50$. Figures 4.6(a)-(c) illustrate the behaviour in Models 1, 2 and 3 respectively. The panels on the left show the characteristics near $\alpha = -1$ and on the right near $\alpha = 1$. The Jacobian is close to 1 for intermediate values of α. For fixed T in Model 1, when $\alpha \in [-0.75, 0.75]$ the values of the Jacobian for the mean and median functions are less than 1 whereas those for the mode function are greater than 1.

In Model 3, when $\alpha > -0.5$ the Jacobian functions decrease monotonically. Near $\alpha = 1$, the Jacobian is approximately equal to zero for all the location functions. Near $\alpha = -1$, larger values are obtained for the mean function (due to its rapid change at the point) than for the median and mode functions. Furthermore, for all the location functions, the properties of the Jacobian near $\alpha = -1$ are model-independent. The Jacobian for the mode function shows some peculiar features (e.g. a non-monotonic behaviour near $\alpha = -1$ in all the models). The numerical procedure and algorithm used to obtain the Jacobian are given Appendix C - Algorithm 15.

4.2.4 Calculation of the MSE of $\hat{\alpha}_{\text{UE}}$

A check on the accuracy in computing the Jacobians and hence the MSE of $\hat{\alpha}_{\text{UE}}$ in (4.9) is the accuracy with which the PDF of the estimators, in (4.7), is evaluated. In particular,

$$\int_{\Omega} f_{\text{UE}}(\hat{\alpha}_{\text{UE}})\, d\hat{\alpha}_{\text{UE}} + \Pr(\hat{\alpha}_{\text{UE}} = \pm 1) = 1 \tag{4.15}$$

should hold.

The parameter space of $\hat{\alpha}_{\text{UE}}$ is $\{(-1, 1)\} \cup \{1\} \cup \{-1\}$ whereas that of the true LDV coefficient is $\Omega \equiv (-1, 1)$ for Model 1 and $\Omega \equiv (-1, 1]$ for Models 2 or 3. The assumptions about the practical limits of the domain of g affects the properties of $f_{\text{UE}}(\hat{\alpha}_{\text{UE}})$ and hence the MSE calculations.

For example, Figure 4.7 compares the Jacobian for the different location functions near $\alpha = -1$ when $T = 50$. Given that

$$\begin{aligned} &\Omega_1 = (-0.999, 0.999) \quad \text{or} \quad \Omega_2 \equiv (-0.9999, 0.9999) \quad \text{in Model 1, and} \\ &\Omega_1 = (-0.999, 1] \quad\quad\ \text{or} \quad \Omega_2 \equiv (-0.9999, 1] \quad\quad\ \text{in Models 2 or 3,} \end{aligned} \tag{4.16}$$

the continuous lines show the case when $\Omega \equiv \Omega_1$ (extracts from Figure 4.6) and the dots when $\Omega \equiv \Omega_2$.

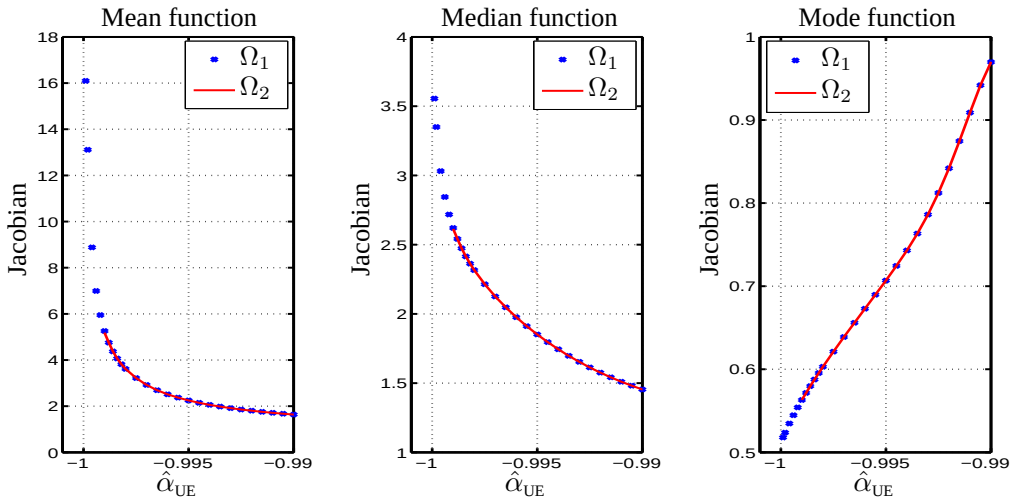


Figure 4.7: Details of Jacobian functions given in Figure 4.6 near $\alpha = -1$ when $\Omega = \Omega_1$ (continuous lines) in all the models. The dots show the case when $\Omega \equiv \Omega_2$. Left to right panels: Jacobian function for the mean, median and mode functions when $T = 50$.

Given Ω_2, when α is close to -1, the Jacobian for the mean and median functions is relatively higher than when $\Omega \equiv \Omega_1$. Similarly, that for the mode function is lower. For example, when $\alpha = -0.999$ in all the models the Jacobian for the mean function is approximately equal to 5 whereas at $\alpha = -0.9999$ it is three times as large, at 16. For the median and mode functions, the Jacobian is approximately 2.5 and 0.57 when $\alpha = -0.999$ and approximately 3.5 and 0.52 when $\alpha = -0.9999$, respectively.

When $\Omega = \Omega_2$, the area (4.15) is slightly overstated - when α is very close to 1 - for the mean- and median-based estimators due to the relatively large values of the Jacobian (e.g. for the latter the area is overstated by about 0.05%). For the mode-based estimator the choice of Ω has little effect. For uniformity in the subroutines, $\Omega \equiv \Omega_1$ was considered.

In all the models, an accuracy of the order 10^{-4} was obtained for the area in (4.15) when $\alpha \approx 1$. Improved accuracy of the order of 10^{-5} to 10^{-9} was noted for other values in Ω. A suitable step-size (0.0005) was used when evaluating $f_{\text{UE}}(\cdot)$ to enable efficient computation near $g(-1)$ and $g(1)$ and hence the probability mass at ± 1. An additional check on the implementation was

done by ensuring equality of the variance in (4.10) against the identity

$$\begin{aligned}\text{Var}(\hat{\alpha}_{\text{UE}}) &= \int_{\Omega} [\,\hat{\alpha}_{\text{UE}} - \mathbb{E}_{\hat{\alpha}_{\text{UE}}}(\hat{\alpha}_{\text{UE}})\,]^2 f_{\text{UE}}(\hat{\alpha}_{\text{UE}}) d\hat{\alpha}_{\text{UE}} + [1 - \mathbb{E}_{\hat{\alpha}_{\text{UE}}}(\hat{\alpha}_{\text{UE}})]^2 \Pr(\hat{\alpha}_{\text{UE}} = 1) \\ &\quad + [-1 - \mathbb{E}_{\hat{\alpha}_{\text{UE}}}(\hat{\alpha}_{\text{UE}})]^2 \Pr(\hat{\alpha}_{\text{UE}} = -1), \quad (4.17)\end{aligned}$$

where $\mathbb{E}_{\hat{\alpha}_{\text{UE}}}(\hat{\alpha}_{\text{UE}})$ is given in (4.8). The two agreed very well. The results are discussed in the following section. The subroutines used are provided in Appendix C.

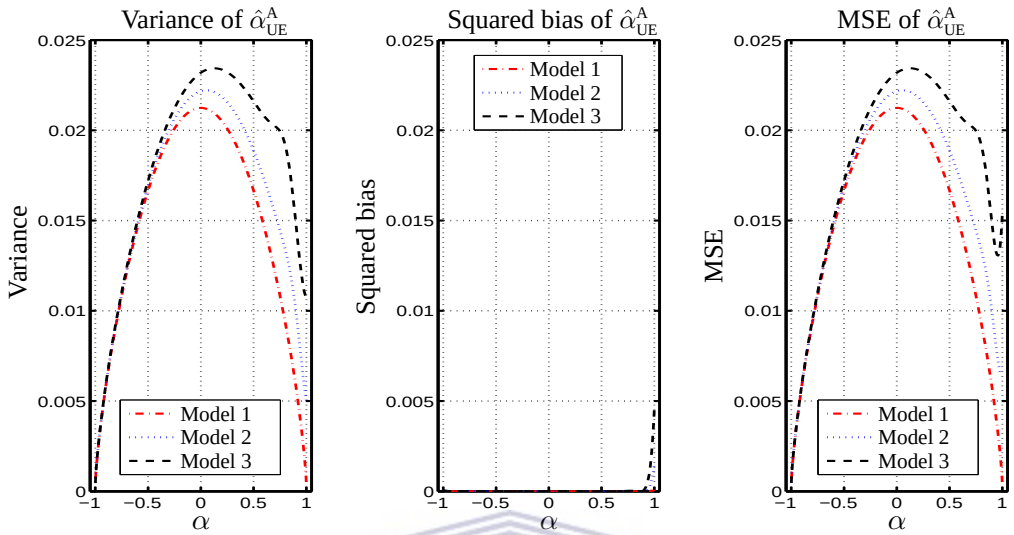
4.2.5 Results : the MSE of $\hat{\alpha}_{\text{UE}}$

The following notation is here introduced to compare the performance of the location-based estimators.

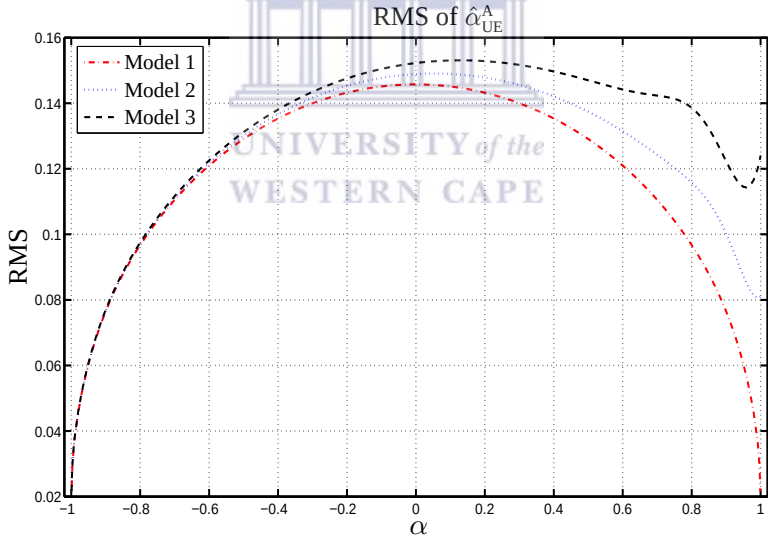
$$\begin{aligned}\hat{\alpha}_{\text{UE}}^{\text{A}} &: \text{mean-based.} \\ \hat{\alpha}_{\text{UE}}^{\text{B}} &: \text{median-based.} \\ \hat{\alpha}_{\text{UE}}^{\text{C}} &: \text{mode-based.}\end{aligned} \quad (4.18)$$

Figure 4.8 shows the properties of the MSE and RMS of $\hat{\alpha}_{\text{UE}}^{\text{A}}$ in different models when $T = 50$. It is the analogue of Figure 4.2 (on page 57) for the LS estimator. The decomposition of the MSE of $\hat{\alpha}_{\text{UE}}^{\text{B}}$ and $\hat{\alpha}_{\text{UE}}^{\text{C}}$ shows similar characteristics to that of $\hat{\alpha}_{\text{UE}}^{\text{A}}$ (see Figure 4.11 on page 75).

Figure 4.9 compares the properties of the variance and squared bias of the location-based estimators $(\hat{\alpha}_{\text{UE}})$ with the variance and squared bias of the LS estimator $(\hat{\alpha}_{\text{LS}})$ for the same sample size. Similarly, in Figure 4.10, the MSE and RMS of $\hat{\alpha}_{\text{UE}}$ is compared with the MSE and RMS of $\hat{\alpha}_{\text{LS}}$. Note that the scales on the vertical axes are not the same in all the diagrams.

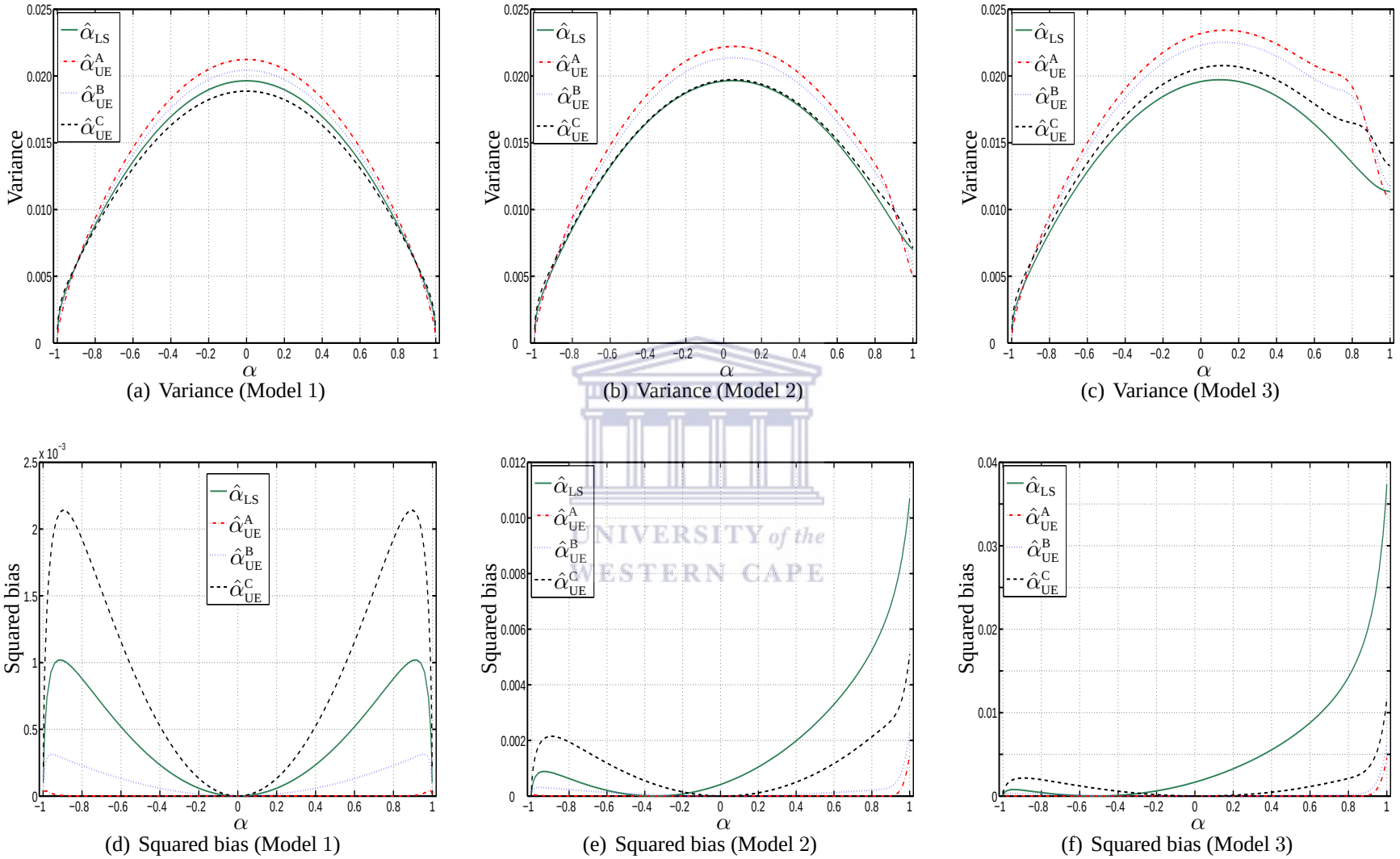


(a) Variance, squared bias and MSE of $\hat{\alpha}^{A}_{UE}$ when $T = 50$ in different models (equally scaled axes).



(b) RMS of $\hat{\alpha}^{A}_{UE}$ in the different models when $T = 50$.

Figure 4.8: (a) Decomposition of the MSE of $\hat{\alpha}^{A}_{UE}$ into variance and squared bias. (b) RMS of $\hat{\alpha}^{A}_{UE}$ over Ω.



(a) Variance (Model 1) (b) Variance (Model 2) (c) Variance (Model 3)

(d) Squared bias (Model 1) (e) Squared bias (Model 2) (f) Squared bias (Model 3)

Figure 4.9: Comparison of the variance and squared bias of the estimators $\hat{\alpha}_{\text{UE}}$ with the variance and squared bias of $\hat{\alpha}_{\text{LS}}$ in different models ($T = 50$).

The exact values of the variance, squared bias, MSE and the RMS are summarized in Tables B16, B17 and B18 for $\hat{\alpha}_{\text{UE}}^{\text{A}}$, $\hat{\alpha}_{\text{UE}}^{\text{B}}$ and $\hat{\alpha}_{\text{UE}}^{\text{C}}$ respectively. All four quantities decrease with T. As compared to the results for the LS estimator, the following deductions can be made

1. In Models 2 and 3, the estimators $\hat{\alpha}_{\text{UE}}$ have significantly smaller bias than $\hat{\alpha}_{\text{LS}}$ over much of Ω. This is especially so for the near unit-root cases. In addition, the bias of $\hat{\alpha}_{\text{UE}}$ is more uniform on Ω than that of $\hat{\alpha}_{\text{LS}}$, in all the models.

2. For all the estimators, as $\alpha \to -1$ the bias decreases to zero. When α is close to zero in Models 2 and 3, the bias of $\hat{\alpha}_{\text{UE}}$ is approximately zero in contrast to that of $\hat{\alpha}_{\text{LS}}$.

3. In contrast to the LS estimator, the contribution of the squared bias to the MSE of the estimators $\hat{\alpha}_{\text{UE}}$ is less than that of the variance in all the models.

4. The bias of $\hat{\alpha}_{\text{UE}}^{\text{A}}$ is approximately zero over much of Ω: its squared bias is less than that of $\hat{\alpha}_{\text{LS}}$ for all combinations of α, T and/or model.

5. Except for $\alpha \in [-0.71, -0.41]$ in Model 2 and $\alpha \in [-0.53, -0.25]$ in Model 3 ($T = 50$), the squared bias of $\hat{\alpha}_{\text{UE}}^{\text{B}}$ is less than that of $\hat{\alpha}_{\text{LS}}$.

6. Except for $\alpha = 0$ in Model 1 (where the bias of the estimators $\hat{\alpha}_{\text{UE}}$ are almost equal), $\alpha \in [-1, -0.17]$ in Model 2, and $\alpha \in [-1, -0.3]$ in Model 3 ($T = 50$), the squared bias of $\hat{\alpha}_{\text{UE}}^{\text{C}}$ is less than that of $\hat{\alpha}_{\text{LS}}$.

7. For the variance of $\hat{\alpha}_{\text{UE}}$ vis-a-vis the variance of the LS estimator when $T = 50$:

   (a) $\text{Var}(\hat{\alpha}_{\text{UE}}^{\text{A}}) \geq \text{Var}(\hat{\alpha}_{\text{LS}})$ except when $\alpha > 0.90$ in Model 1; $\alpha > 0.93$ in Model 2; and $\alpha > 0.97$ in Model 3.

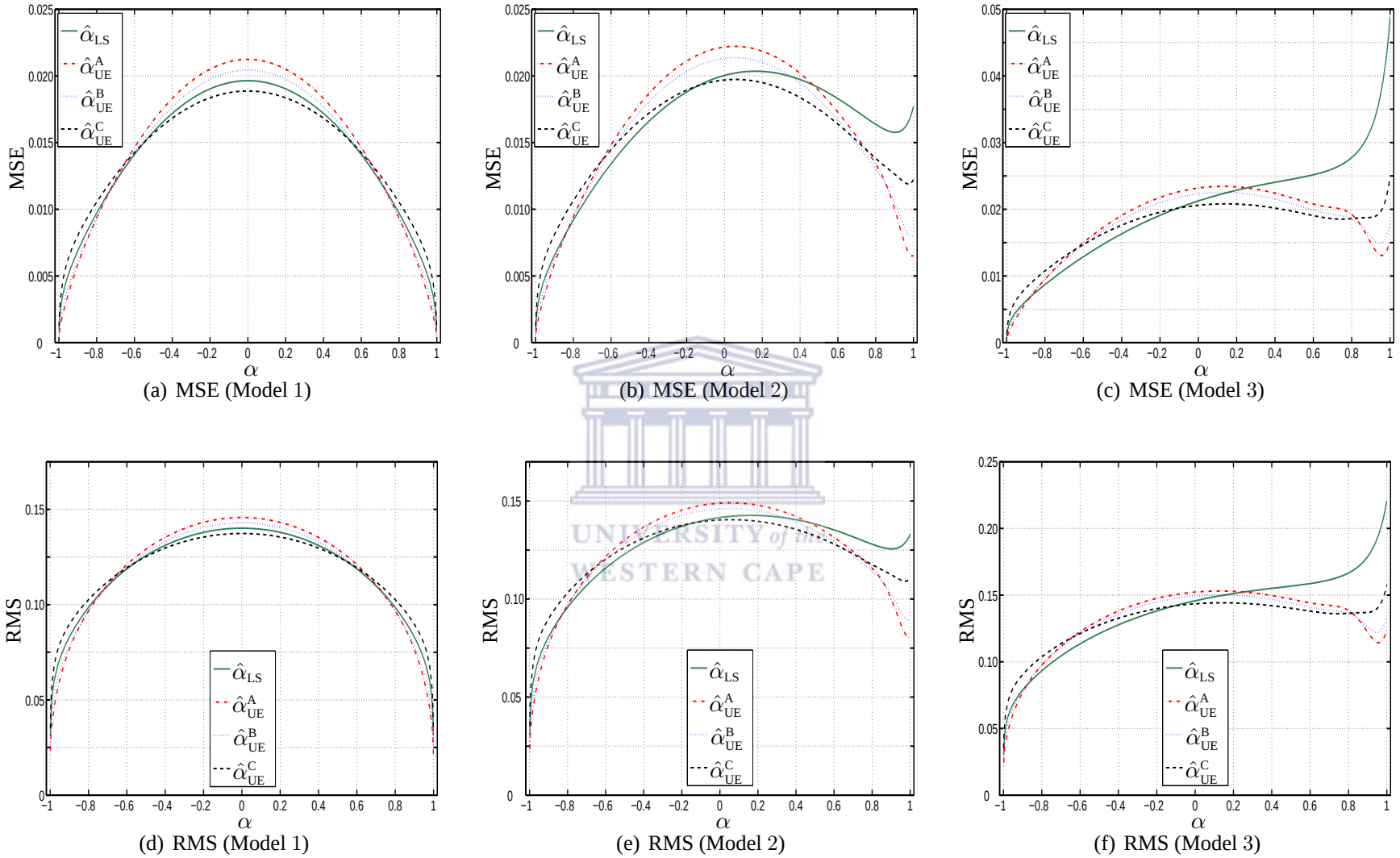
   (b) $\text{Var}(\hat{\alpha}_{\text{UE}}^{\text{B}}) \geq \text{Var}(\hat{\alpha}_{\text{LS}})$ except when $\alpha > 0.90$ in Model 1 and $\alpha > 0.95$ in Model 2.

Table 4.1: Summary of ranges on Ω where the MSE of $\hat{\alpha}_{\text{UE}}$ is less than or equal to that of $\hat{\alpha}_{\text{LS}}$ in different models ($T = 50$).

Estimator \ Model	1	2	3
$\hat{\alpha}_{\text{UE}}^{\text{A}}$	$\lvert\alpha\rvert \geq 0.55$	$\alpha \in [-1, -0.85] \cup [0.47, 1]$	$\alpha \in [-1, -0.87] \cup [0.25, 1]$
$\hat{\alpha}_{\text{UE}}^{\text{B}}$	$\lvert\alpha\rvert \geq 0.65$	$\alpha \in [-1, -0.85] \cup [0.35, 1]$	$\alpha \in [-1, -0.90] \cup [0.15, 1]$
$\hat{\alpha}_{\text{UE}}^{\text{C}}$	$\lvert\alpha\rvert \leq 0.55$	$\alpha \in (-0.15, 1]$	$\alpha \in (-0.10, 1]$

(c) $\text{Var}(\hat{\alpha}_{\text{UE}}^{\text{C}}) \leq \text{Var}(\hat{\alpha}_{\text{LS}})$ in Model 1 except when $|\alpha| > 0.90$. However, in Models 2 and 3, $\text{Var}(\hat{\alpha}_{\text{UE}}^{\text{C}}) \geq \text{Var}(\hat{\alpha}_{\text{LS}})$ for all α.

8. With the exception of Model 3, where the MSE of $\hat{\alpha}_{\text{UE}}^{\text{C}}$ is largest near $\alpha = 1$ (because of substantial bias), the MSE of $\hat{\alpha}_{\text{UE}}$ is largest near $\alpha = 0$. Furthermore, for the sample size considered, It was that $\text{MSE}(\hat{\alpha}_{\text{UE}}) \leq \text{MSE}(\hat{\alpha}_{\text{LS}})$ in the ranges in Table 4.1.



(a) MSE (Model 1) (b) MSE (Model 2) (c) MSE (Model 3)

(d) RMS (Model 1) (e) RMS (Model 2) (f) RMS (Model 3)

Figure 4.10: Comparison of the MSE and RMS of the estimators $\hat{\alpha}_{\text{UE}}$ with the MSE and RMS of $\hat{\alpha}_{\text{LS}}$ in different models ($T = 50$)

4.2.6 Efficiency of the estimators $\hat{\alpha}_{\text{UE}}$ against $\hat{\alpha}_{\text{LS}}$

Figure 4.11 shows a comparison of the variance, bias and MSE between the estimators $\hat{\alpha}_{\text{UE}}$. The results for $\hat{\alpha}^{\text{A}}_{\text{UE}}$, $\hat{\alpha}^{\text{B}}_{\text{UE}}$ and $\hat{\alpha}^{\text{C}}_{\text{UE}}$ are shown in the left, middle and right panels respectively. The relative contributions of the variance and squared bias to the MSE in all the models is illustrated.

With regard to the question Q2 on page 52, Figure 4.12 shows the biases of $\hat{\alpha}_{\text{UE}}$ and $\hat{\alpha}_{\text{LS}}$ in each model. Figures 4.13 and 4.14 summarize the ratios of the variance, MSE and RMS of $\hat{\alpha}_{\text{UE}}$ to quantities calculated for $\hat{\alpha}_{\text{LS}}$. Values less than one therefore indicate improvement of $\hat{\alpha}_{\text{UE}}$ over $\hat{\alpha}_{\text{LS}}$.

The following deductions can be made

1. With the exception of $\hat{\alpha}^{\text{C}}_{\text{UE}}$ in Model 1, the bias of $\hat{\alpha}_{\text{UE}}$ is significantly improved over much of Ω. The bias of the estimators $\hat{\alpha}_{\text{UE}}$ changes signs near $\alpha = 0$ in all the models. In Model 1 the bias, variance, MSE and RMS of $\hat{\alpha}_{\text{UE}}$ show a symmetry about $\alpha = 0$.

2. Interestingly, for the sample size considered, when $\alpha \in [0.04, 0.865]$ for $\hat{\alpha}^{\text{B}}_{\text{UE}}$ and $\alpha \in [0.03, 0.55]$ for $\hat{\alpha}^{\text{C}}_{\text{UE}}$, the absolute bias of the estimators in Models 1 is greater than that in Models 2 whereas that in Model 2 is greater than that in Model 3. However, as $|\alpha| \to 1$ the bias of $\hat{\alpha}_{\text{UE}}$ increases as more exogenous regressors are added to the simple AR(1) specification i.e. from Models 1 to 3.

3. For all α and T in a given model

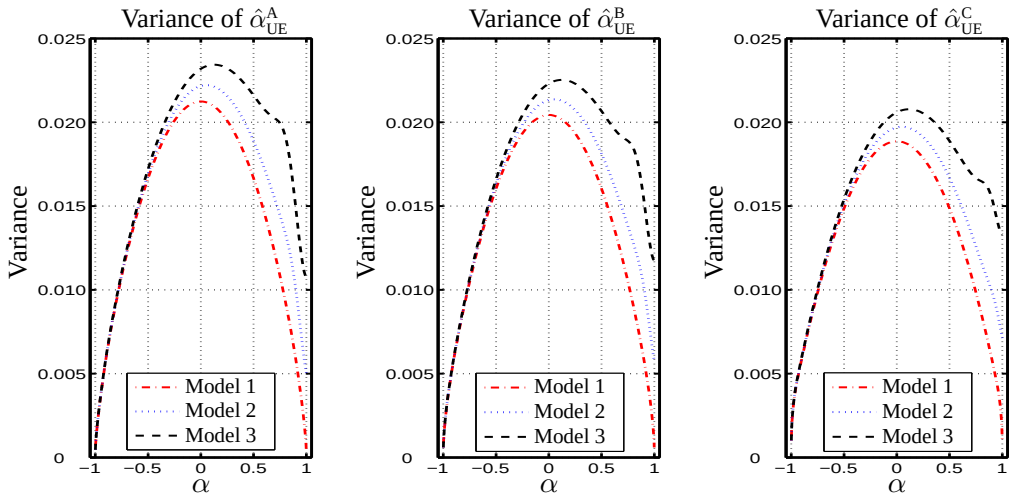
$$|\text{Bias}(\hat{\alpha}^{\text{C}}_{\text{UE}})| \geq |\text{Bias}(\hat{\alpha}^{\text{B}}_{\text{UE}})| \geq |\text{Bias}(\hat{\alpha}^{\text{A}}_{\text{UE}})|, \tag{4.19}$$

   Table B19 contains the exact values of the bias of $\hat{\alpha}_{\text{UE}}$. It reaches a maximum at or near $\alpha = 1$. Table 4.2 summarizes the ratios (in %) of the bias, for $\hat{\alpha}^{\text{A}}_{\text{UE}}$, $\hat{\alpha}^{\text{B}}_{\text{UE}}$ and $\hat{\alpha}^{\text{C}}_{\text{UE}}$ relative to the LS estimator.

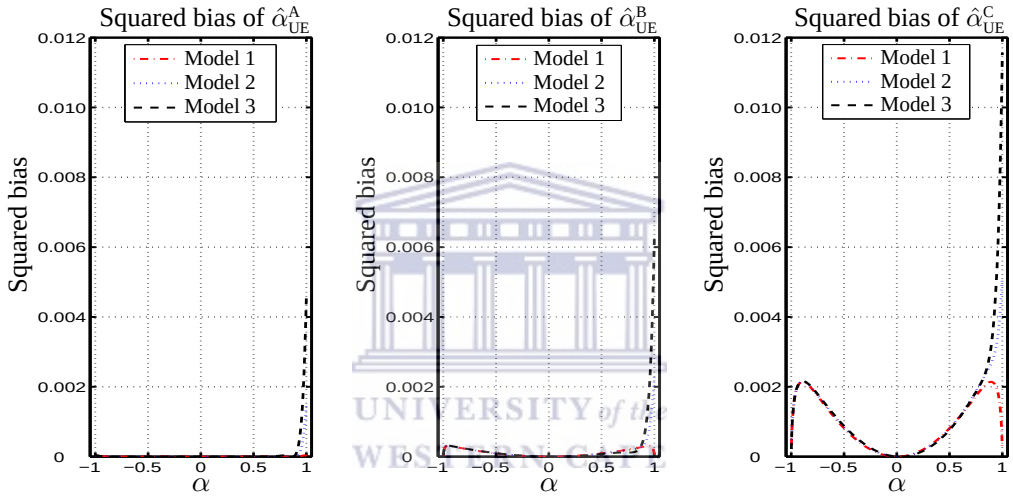
4. Given α and T

$$\text{Var}(\hat{\alpha}_{\text{UE1}}) \leq \text{Var}(\hat{\alpha}_{\text{UE2}}) \leq \text{Var}(\hat{\alpha}_{\text{UE3}}), \tag{4.20}$$

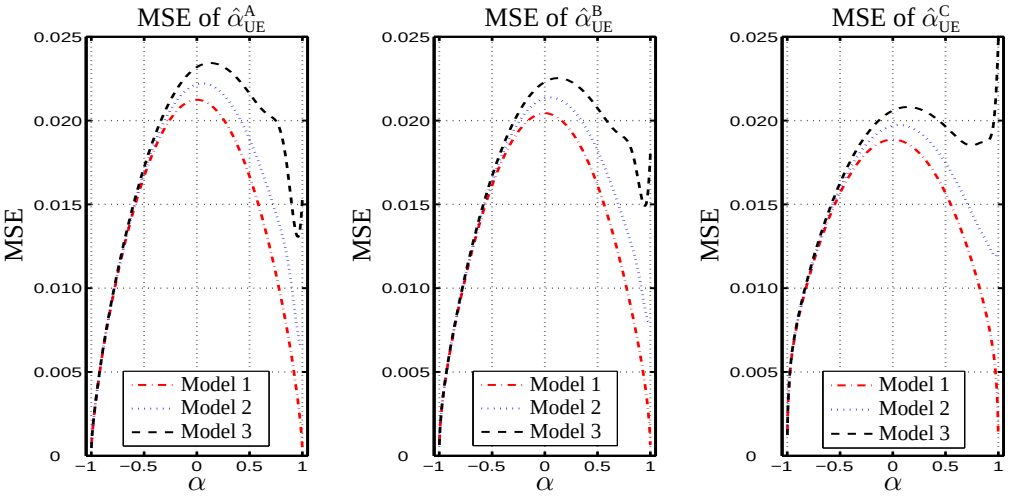
   where the numerical subscripts denote the model. Inequalities similar to (4.20) also hold for the MSE and the RMS of $\hat{\alpha}^{\text{A}}_{\text{UE}}$, $\hat{\alpha}^{\text{B}}_{\text{UE}}$ and $\hat{\alpha}^{\text{C}}_{\text{UE}}$.



(a) Variance of the estimators $\hat{\alpha}_{UE}$ in different models (equally scaled axes).



(b) Squared bias of the estimators $\hat{\alpha}_{UE}$ in different models (equally scaled axes).



(c) MSE of the estimators $\hat{\alpha}_{UE}$ in different models (equally scaled axes).

Figure 4.11: Comparison of the (a) variance, (b) squared bias, and (c) MSE, between the estimators $\hat{\alpha}^{A}_{UE}$, $\hat{\alpha}^{B}_{UE}$ and $\hat{\alpha}^{C}_{UE}$ ($T = 50$).

Table 4.2: Ratios (in %) of the bias for $\hat{\alpha}_{\text{UE}}^{\text{A}}$, $\hat{\alpha}_{\text{UE}}^{\text{B}}$ and $\hat{\alpha}_{\text{UE}}^{\text{C}}$, relative to the bias of $\hat{\alpha}_{\text{LS}}$ in different models ($T = 50$).

	$\hat{\alpha}_{\text{UE}}^{\text{A}}$			$\hat{\alpha}_{\text{UE}}^{\text{B}}$			$\hat{\alpha}_{\text{UE}}^{\text{C}}$		
α/Model	1	2	3	1	2	3	1	2	3
-0.999	38.17	39.40	40.70	74.02	76.20	78.45	156.00	159.98	164.11
-0.800	1.63	1.97	2.46	50.99	60.63	74.30	147.89	175.80	215.37
0.200	0.16	(0.27)	(0.50)	50.10	10.88	5.38	149.82	33.15	17.14
0.400	0.23	(0.39)	(0.83)	50.17	16.52	9.14	149.65	50.22	29.00
0.800	1.63	(2.16)	(4.87)	50.99	20.09	9.35	147.89	63.80	38.54
0.990	31.93	32.98	31.66	68.31	42.79	37.88	139.30	67.44	53.38
0.999	38.17	37.34	34.68	74.02	46.26	40.59	156.00	68.94	55.36
1.000	-	37.83	35.02	-	46.65	40.90	-	69.12	55.60

5. In Model 3 the variance of the estimators $\hat{\alpha}_{\text{UE}}$ is characterized by a shoulder near $\alpha = 0.80$ [see Figure 4.9(c)]. This is due to the restriction on the parameter space of $\hat{\alpha}_{\text{UE}}$ which gives rise to the mass-point probabilities [see (3.15) and (4.17)].

6. From Figures 4.13(a)-(c), when α is close to ± 1 in each of the models

$$\text{Var}(\hat{\alpha}_{\text{UE}}^{\text{A}}) \leq \text{Var}(\hat{\alpha}_{\text{UE}}^{\text{B}}) \leq \text{Var}(\hat{\alpha}_{\text{UE}}^{\text{C}}). \tag{4.21}$$

Otherwise, i.e. for intermediate values of α,

$$\text{Var}(\hat{\alpha}_{\text{UE}}^{\text{C}}) \leq \text{Var}(\hat{\alpha}_{\text{UE}}^{\text{B}}) \leq \text{Var}(\hat{\alpha}_{\text{UE}}^{\text{A}}). \tag{4.22}$$

For example, when $T = 50$, (4.21) holds except when $|\alpha| > 0.90$ in Model 1. The variance of $\hat{\alpha}_{\text{UE}}$ is largest when α is close to zero but decreases as $|\alpha| \to 1$. Table 4.3 compares the variance of $\hat{\alpha}_{\text{UE}}$ with respect to that of $\hat{\alpha}_{\text{LS}}$ in percentage terms.

Table 4.3: Ratio of the variance of $\hat{\alpha}_{\text{UE}}$ expressed as a percentage to the variance of $\hat{\alpha}_{\text{LS}}$ ($T = 50$).

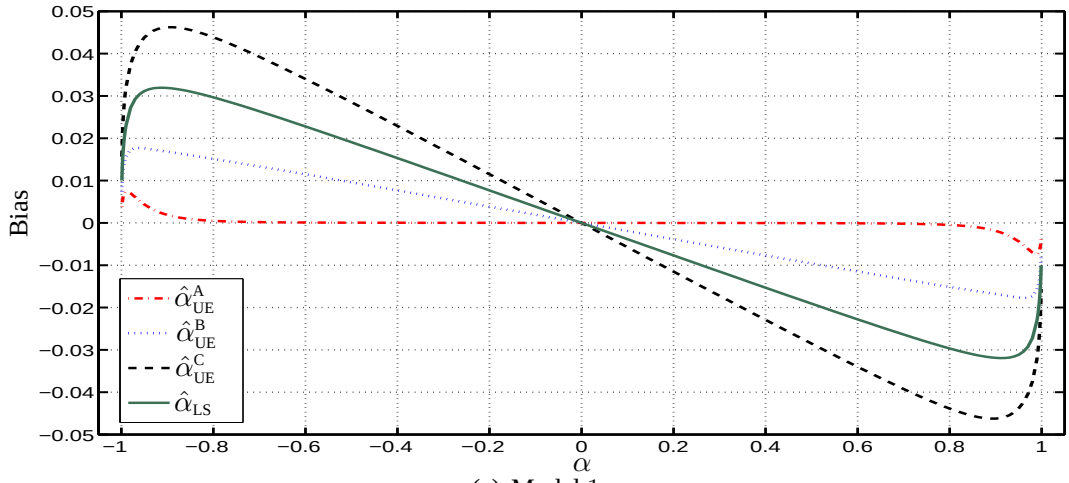
	$\hat{\alpha}_{\text{UE}}^{\text{A}}$			$\hat{\alpha}_{\text{UE}}^{\text{B}}$			$\hat{\alpha}_{\text{UE}}^{\text{C}}$		
α/Model	1	2	3	1	2	3	1	2	3
-0.999	51.99	53.99	56.08	68.80	71.45	74.20	116.12	120.56	125.15
-0.800	105.92	110.28	114.87	102.94	107.27	111.82	97.47	101.73	106.22
-0.400	108.01	112.59	117.49	103.96	108.41	113.19	96.16	100.38	104.91
0.000	108.14	112.99	118.41	104.03	108.70	113.91	96.06	100.36	105.17
0.400	108.01	113.81	121.12	103.96	109.33	116.13	96.16	100.74	106.57
0.800	105.92	120.34	142.16	102.94	115.12	136.90	97.47	104.96	122.13
0.950	87.87	94.50	105.70	94.28	102.19	113.28	106.19	111.58	122.96
0.990	64.00	75.83	95.65	77.92	85.96	104.18	112.41	104.22	117.47
0.999	51.99	71.96	95.02	68.80	82.26	103.59	116.12	101.85	117.07
1.000	-	71.55	95.01	-	81.87	103.58	-	101.58	117.06

7. From Figures 4.13(d)-(f), when α is close to ± 1 in each model

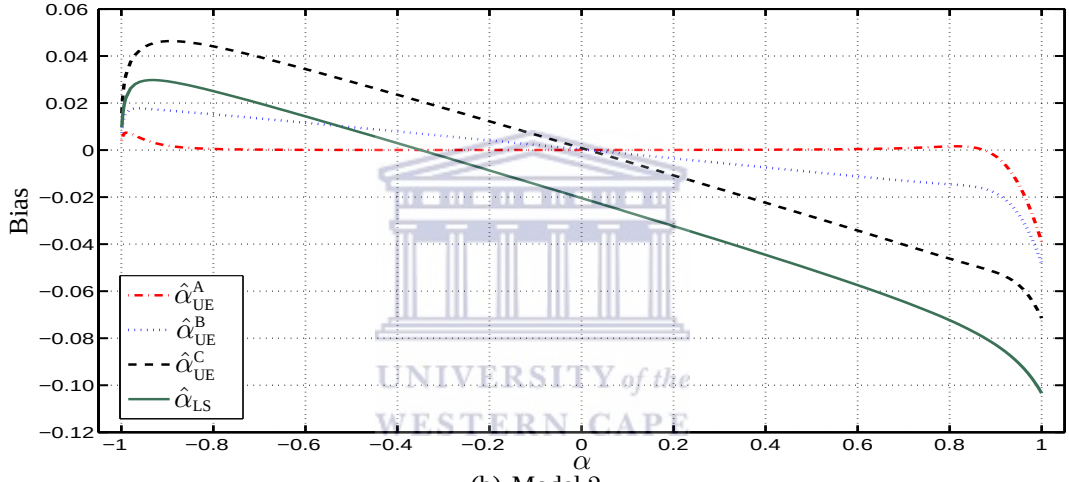
$$\text{MSE}(\hat{\alpha}_{\text{UE}}^{\text{A}}) \leq \text{MSE}(\hat{\alpha}_{\text{UE}}^{\text{B}}) \leq \text{MSE}(\hat{\alpha}_{\text{UE}}^{\text{C}}). \tag{4.23}$$

For intermediate values of α

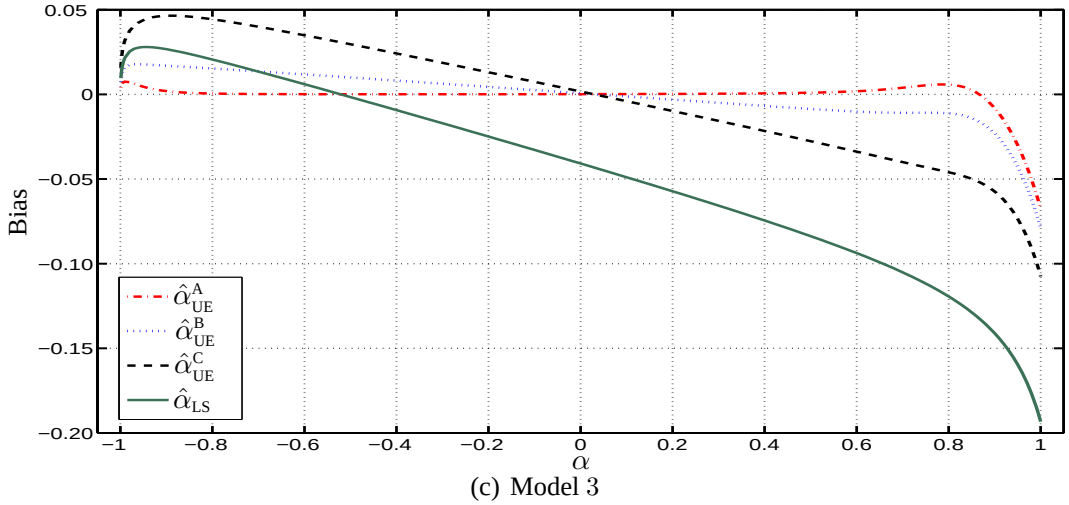
$$\text{MSE}(\hat{\alpha}_{\text{UE}}^{\text{C}}) \leq \text{MSE}(\hat{\alpha}_{\text{UE}}^{\text{B}}) \leq \text{MSE}(\hat{\alpha}_{\text{UE}}^{\text{A}}). \tag{4.24}$$



(a) Model 1

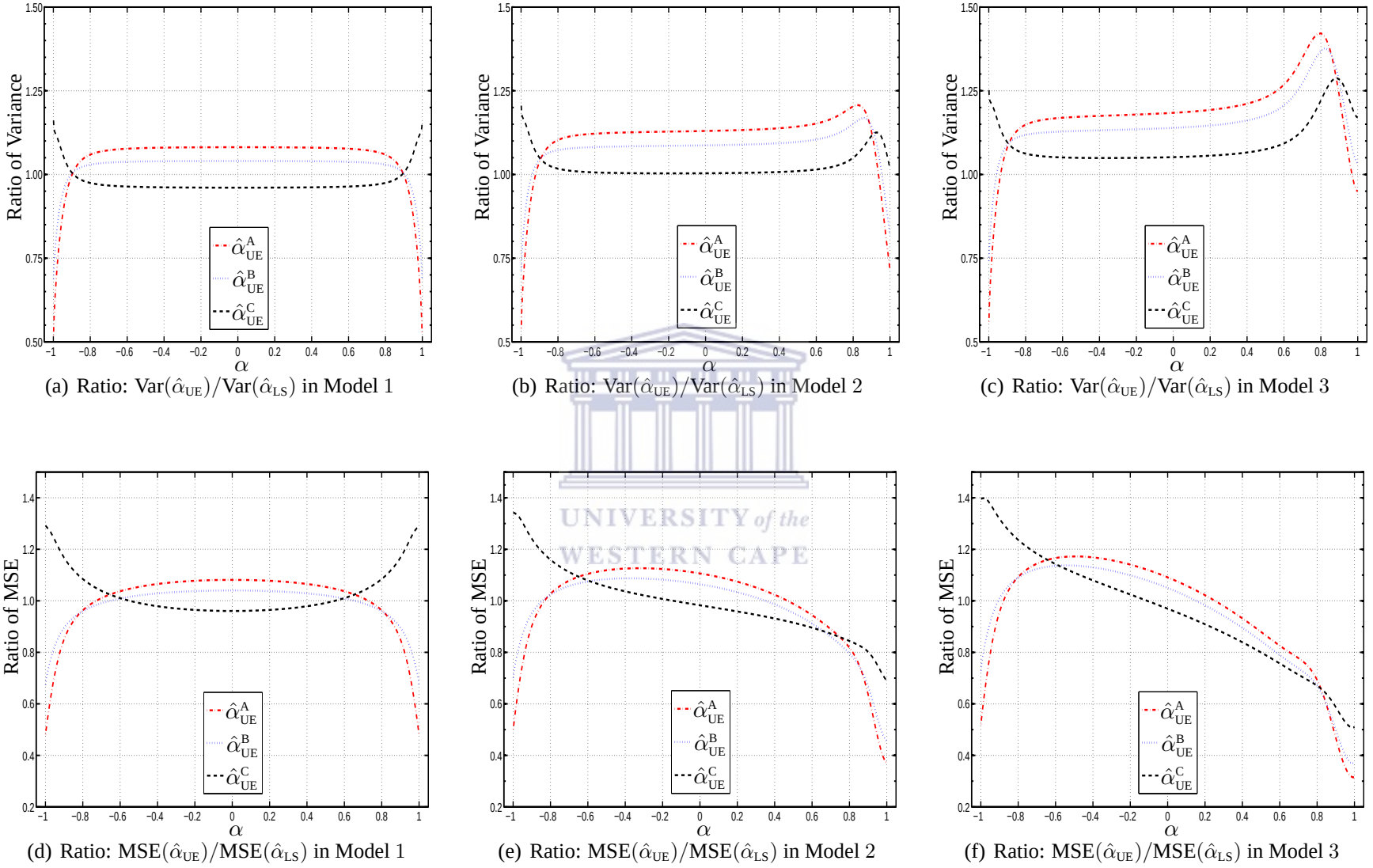


(b) Model 2



(c) Model 3

Figure 4.12: Comparison of the bias of $\hat{\alpha}_{\text{UE}}$ to that of $\hat{\alpha}_{\text{LS}}$ in different models ($T = 50$).



(a) Ratio: $\text{Var}(\hat{\alpha}_{\text{UE}})/\text{Var}(\hat{\alpha}_{\text{LS}})$ in Model 1

(b) Ratio: $\text{Var}(\hat{\alpha}_{\text{UE}})/\text{Var}(\hat{\alpha}_{\text{LS}})$ in Model 2

(c) Ratio: $\text{Var}(\hat{\alpha}_{\text{UE}})/\text{Var}(\hat{\alpha}_{\text{LS}})$ in Model 3

(d) Ratio: $\text{MSE}(\hat{\alpha}_{\text{UE}})/\text{MSE}(\hat{\alpha}_{\text{LS}})$ in Model 1

(e) Ratio: $\text{MSE}(\hat{\alpha}_{\text{UE}})/\text{MSE}(\hat{\alpha}_{\text{LS}})$ in Model 2

(f) Ratio: $\text{MSE}(\hat{\alpha}_{\text{UE}})/\text{MSE}(\hat{\alpha}_{\text{LS}})$ in Model 3

Figure 4.13: Ratios of variance and MSE of the estimators $\hat{\alpha}_{\text{UE}}$ relative to $\hat{\alpha}_{\text{LS}}$ in a sample of size $T = 50$.

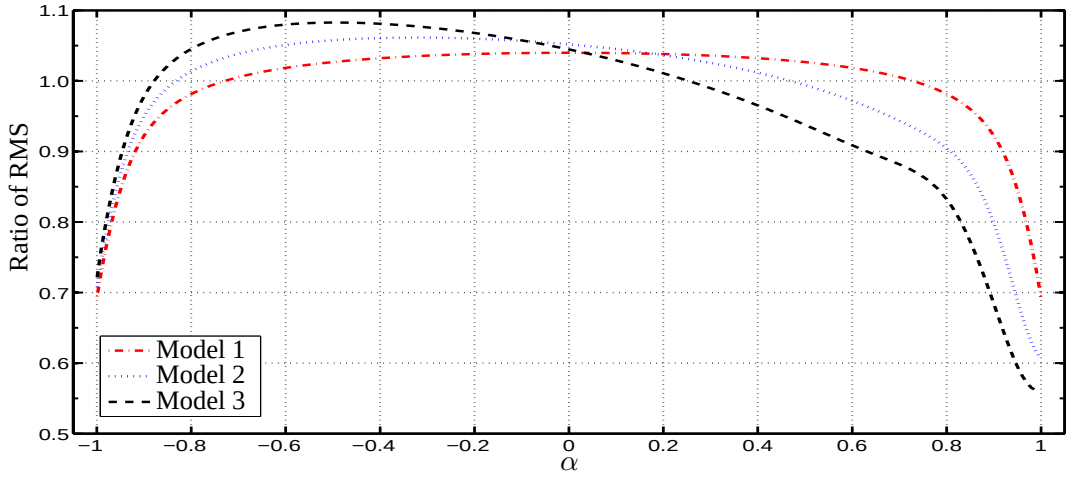
From Figures 4.2 through to 4.11

1. There is an obvious trade-off between the bias and variance of the estimators: estimators with small variance typically have larger bias, and vice versa (Tables 4.3 and 4.2).

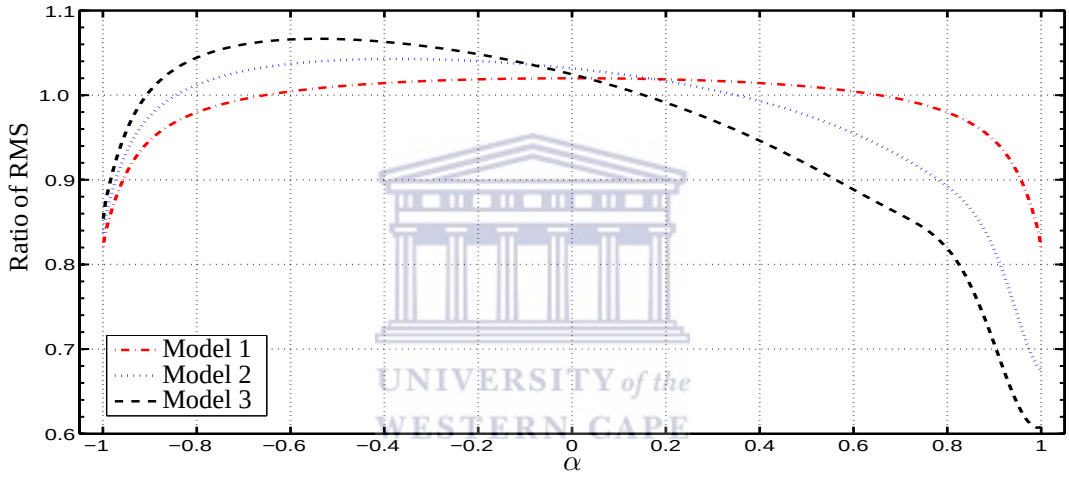
2. Relative to the LS estimator, the mean- and median-based estimators have smaller MSE when $|\alpha|$ is close to unity (all models). The mode-based estimator shows similar properties only in Models 2 and 3: its RMS is non-increasing over Ω when an intercept or a time-trend is incorporated in the vector $\mathbf{X}$ in (1.1) (Figure 4.14).

3. From Figure 4.13, when the LDV coefficient is close to -1, the MSE of the location-based estimators increases as more exogenous regressors are added. The opposite applies when $\alpha \to 1$. Therefore, substantial improvement in estimation of the LDV coefficient is achieved in those models which contain exogenous regressors: the new estimators are particularly useful in Models 2 and 3 where the bias of $\hat{\alpha}_{\text{LS}}$ is large.

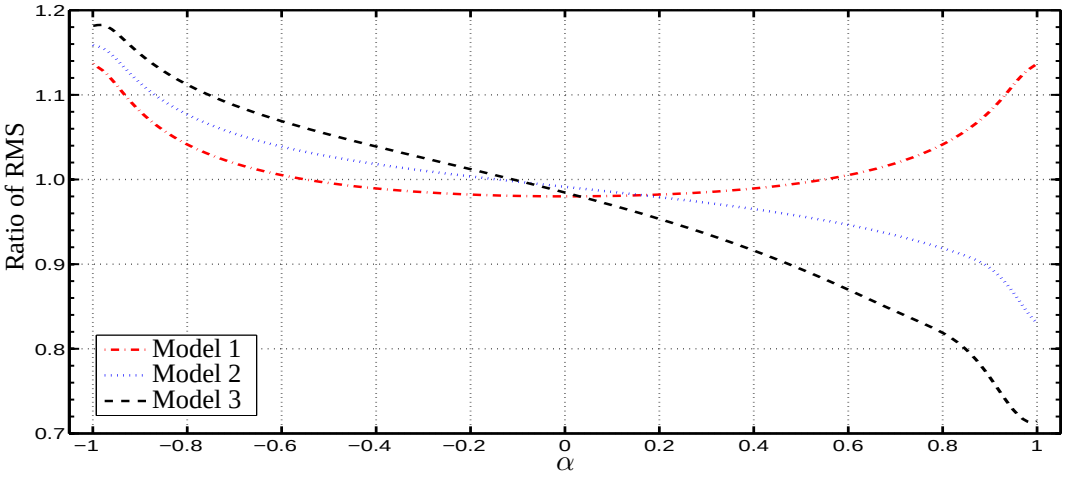
4. The mean-based estimator is of course much more useful for improving the bias of the LS estimator than the other two estimators.



(a) Ratio: $\text{RMS}(\hat{\alpha}_{\text{UE}})/\text{RMS}(\hat{\alpha}_{\text{LS}})$ for $\hat{\alpha}_{\text{UE}}^{\text{A}}$.



(b) Ratio: $\text{RMS}(\hat{\alpha}_{\text{UE}})/\text{RMS}(\hat{\alpha}_{\text{LS}})$ for $\hat{\alpha}_{\text{UE}}^{\text{B}}$.



(c) Ratio: $\text{RMS}(\hat{\alpha}_{\text{UE}})/\text{RMS}(\hat{\alpha}_{\text{LS}})$ for $\hat{\alpha}_{\text{UE}}^{\text{C}}$.

Figure 4.14: Ratios of the RMS in different models ($T = 50$).

4.3 Confidence intervals

In frequentist theory, confidence intervals (CI) provide a useful tool for inference (Williams 2001). In this section question Q3 (on page 52) is considered. Interval estimators of $\hat{\alpha}_{\text{UE}}$ are determined and compared. The CI provide a measure of their accuracy. The CI can be used directly to construct one- and two-sided tests of $\mathbf{H}_0 : \alpha = \alpha_0$ for arbitrary $\alpha_0 \in (-1, 1]$ (Andrews 1993).

4.3.1 Exact confidence intervals: location-based estimators

The CI take into account both the finite sample size and the extraneous regressor matrix $\mathbf{X}$ in (1.1). As such, they are more useful than asymptotic bounds (Dufour 1990). Unlike Andrews (1993, pp. 147, 152), where the confidence bounds for the median-based estimator were obtained from the inverse quantile function of $\hat{\alpha}_{\text{LS}}$ defined on Ω, the CI here are obtained directly from CDF $F_{\text{UE}}(\cdot)$ of the estimators $\hat{\alpha}_{\text{UE}}$. The CDF is established from the corresponding PDF $f_{\text{UE}}(\cdot)$ defined in (4.7). The approach is more direct, theoretically appealing and accurate than the implicit iterative scheme proposed by Andrews (1993).

For illustration, the 90% confidence bounds for the LDV coefficient α can be obtained by establishing the 0.05 and 0.95 quantiles of the CDF or CMF. The notation $\text{CI}_j^{0.05}$ is used to denote the 0.05 quantile of the estimators in Model j $(j = 1, 2, 3)$. For the LS estimator, the bounds can be derived from the distribution function defined in (2.13) using Algorithm 2. In this case, p is equal to 0.05 and 0.95.

The accuracy of the confidence bounds depends on the accuracy with which the PDF of $\hat{\alpha}_{\text{UE}}$ in (4.7) is determined. The quantiles of the LS estimator (determined using the same procedure as those of $\hat{\alpha}_{\text{UE}}$) were compared to the values in Andrews (1993) as a check on the accuracy of the proposed Algorithm. The subroutine (Algorithm 17) utilizes *MATLAB* functions `cumsum` and `cumtrapz` and is invoked whilst computing the MSE of $\hat{\alpha}_{\text{UE}}$ discussed in section 4.2.4.

4.3.2 Results: confidence width

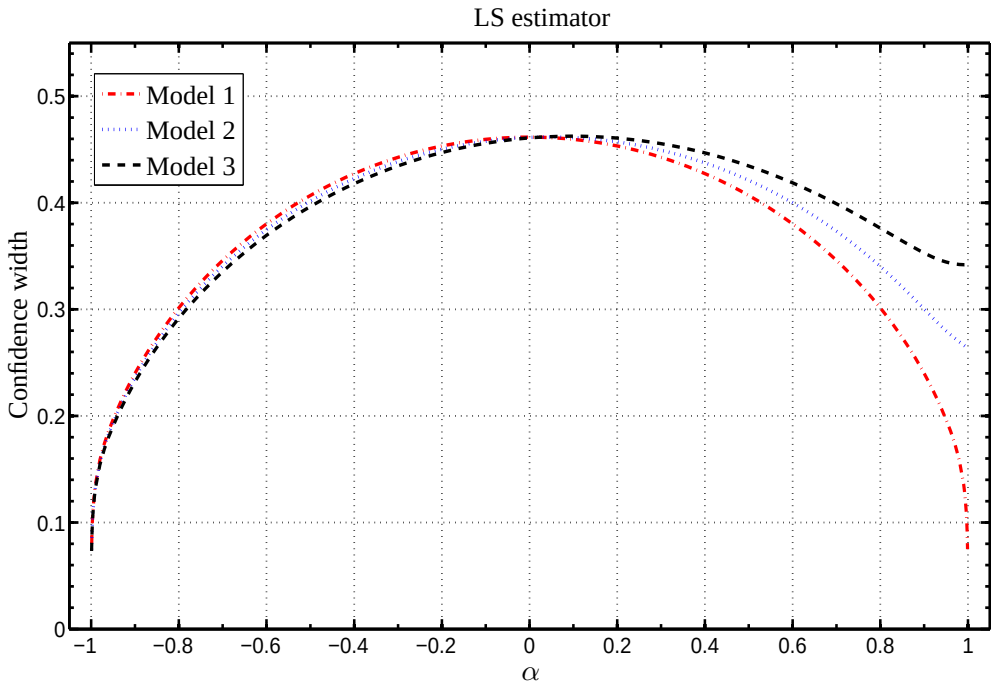
Table 4.4 gives summary results for the CI of $\hat{\alpha}^{\text{A}}_{\text{UE}}$, $\hat{\alpha}^{\text{B}}_{\text{UE}}$, $\hat{\alpha}^{\text{C}}_{\text{UE}}$ and $\hat{\alpha}_{\text{LS}}$ in Model 1 when $T = 50$. The CI of the LS estimator are shown in bold. In Model 1, the quantile function - which relates the quantiles to $\alpha \in \Omega$ - shows a symmetry about $\alpha = 0$ (Andrews 1993). Tables B20 and B21 contain the results for Models 2 and 3 respectively.

Table 4.4: The 0.05 and 0.95 quantiles of the estimators $\hat{\alpha}_{\text{UE}}$ for Model 1 when $T = 50$. Quantiles for the LS estimator are shown in bold.

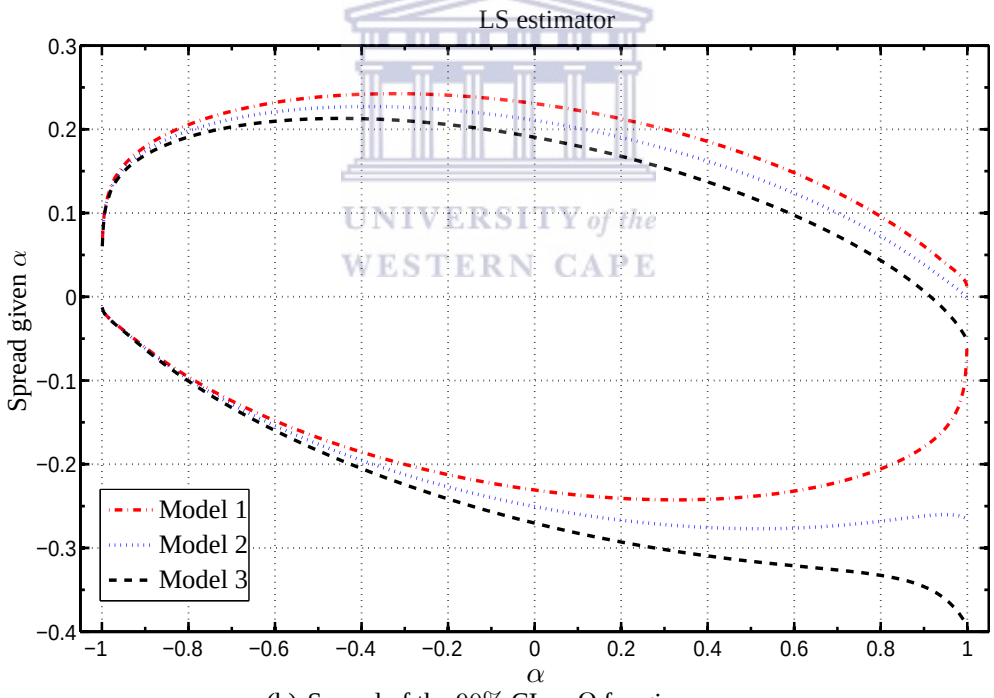
Estimator	$\hat{\alpha}^{\text{A}}_{\text{UE}}$		$\hat{\alpha}^{\text{B}}_{\text{UE}}$		$\hat{\alpha}^{\text{C}}_{\text{UE}}$		$\hat{\alpha}_{\text{LS}}$	
α/Quantile	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95
0.00	-0.240	0.240	-0.235	0.235	-0.226	0.226	**-0.231**	**0.231**
0.05	-0.191	0.288	-0.188	0.283	-0.180	0.272	**-0.184**	**0.277**
0.10	-0.142	0.336	-0.140	0.329	-0.134	0.316	**-0.137**	**0.323**
0.15	-0.093	0.383	-0.091	0.375	-0.087	0.361	**-0.089**	**0.368**
0.20	-0.042	0.429	-0.042	0.421	-0.040	0.404	**-0.041**	**0.413**
0.25	0.008	0.475	0.008	0.466	0.008	0.448	**0.008**	**0.457**
0.30	0.060	0.520	0.059	0.510	0.056	0.490	**0.057**	**0.500**
0.35	0.112	0.565	0.110	0.554	0.105	0.532	**0.108**	**0.543**
0.40	0.164	0.609	0.161	0.597	0.155	0.574	**0.158**	**0.586**
0.45	0.218	0.652	0.214	0.640	0.205	0.615	**0.209**	**0.627**
0.50	0.272	0.694	0.267	0.681	0.256	0.655	**0.261**	**0.668**
0.55	0.327	0.736	0.321	0.722	0.308	0.695	**0.314**	**0.709**
0.60	0.383	0.777	0.375	0.763	0.361	0.734	**0.368**	**0.748**
0.65	0.439	0.817	0.431	0.802	0.414	0.772	**0.423**	**0.787**
0.70	0.497	0.855	0.488	0.840	0.469	0.809	**0.478**	**0.824**
0.75	0.557	0.893	0.546	0.877	0.525	0.845	**0.536**	**0.861**
0.80	0.618	0.928	0.606	0.912	0.583	0.881	**0.594**	**0.896**
0.85	0.681	0.960	0.669	0.945	0.643	0.915	**0.656**	**0.929**
0.90	0.749	0.986	0.735	0.974	0.707	0.949	**0.721**	**0.961**
0.93	0.792	0.996	0.777	0.989	0.748	0.971	**0.763**	**0.978**
0.95	0.824	1.000	0.808	0.996	0.778	0.987	**0.793**	**0.990**
0.97	0.859	1.000	0.843	1.000	0.813	1.000	**0.828**	**1.002**
0.99	0.908	1.000	0.892	1.000	0.861	1.000	**0.876**	**1.012**
0.995	0.929	1.000	0.914	1.000	0.882	1.000	**0.898**	**1.014**
0.999	0.967	1.000	0.952	1.000	0.923	1.000	**0.937**	**1.012**

Figure 4.15 shows the properties of the CI for the LS estimator $(T = 50)$. Their width, obtained as the difference between the 0.95 and 0.05 quantile values, is shown in Figure 4.15(a).

Instead of plotting the actual CI against α, for comparative purposes, it is more instructive to consider the spread of the bounds over Ω for given α. Figure 4.15(b) shows the results. The spread values are obtained by deducting α from the corresponding quantile values. For example, in the case of Model 3 at $\alpha = 1$, the actual bounds are $\text{CI}^{0.05}_3 = 1 - 0.394 = 0.606$ whereas $\text{CI}^{0.95}_3 = 1 - 0.052 = 0.948$ where -0.394 and -0.052 are the spread values shown on the y-axis. As noted in Andrews (1993, p. 146), the 0.95 quantile function in Model 1 dips slightly near $\alpha = 1$ when $10 \leq T \leq 60$: see Table 4.4 above.



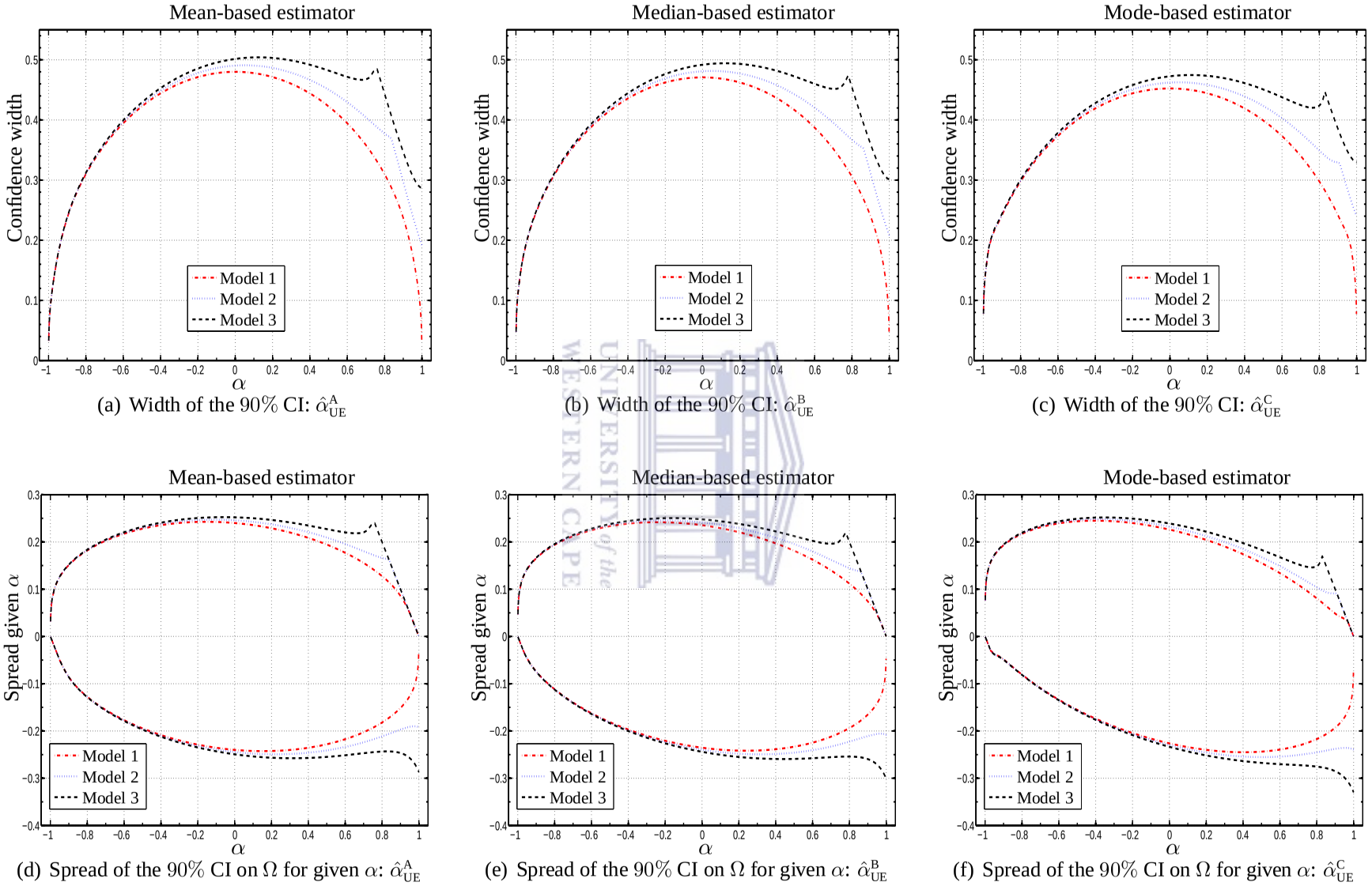
(a) Width of the 90% CI: $\hat{\alpha}_{\text{LS}}$



(b) Spread of the 90% CI on Ω for given α.

Figure 4.15: 90% confidence intervals for the LS estimator in different models when $T = 50$. The actual bounds in (b) can be obtained by adding α to the corresponding spread value on the y-axis.

In Figure 4.16, the CI of each of the estimators $\hat{\alpha}_{\text{UE}}$ in different models are compared whereas in Figure 4.17 the properties of the CI of $\hat{\alpha}_{\text{LS}}$, $\hat{\alpha}^{\text{A}}_{\text{UE}}$, $\hat{\alpha}^{\text{B}}_{\text{UE}}$ and $\hat{\alpha}^{\text{C}}_{\text{UE}}$ in each model are illustrated.



(a) Width of the 90% CI: $\hat{\alpha}^{A}_{UE}$ (b) Width of the 90% CI: $\hat{\alpha}^{B}_{UE}$ (c) Width of the 90% CI: $\hat{\alpha}^{C}_{UE}$

(d) Spread of the 90% CI on Ω for given α: $\hat{\alpha}^{A}_{UE}$ (e) Spread of the 90% CI on Ω for given α: $\hat{\alpha}^{B}_{UE}$ (f) Spread of the 90% CI on Ω for given α: $\hat{\alpha}^{C}_{UE}$

Figure 4.16: Comparison of the 90% confidence intervals of location-based estimators in different models ($T = 50$).

The following deductions can be made:

1. For all the estimators $\hat{\alpha}_{\text{UE}}$, the width of the CI is smallest near $\alpha = -1$. When α is close to -1, inference based on $\hat{\alpha}^{\text{A}}_{\text{UE}}$ is more precise than that of $\hat{\alpha}^{\text{B}}_{\text{UE}}$ or $\hat{\alpha}^{\text{C}}_{\text{UE}}$ since the bounds are tighter. From Figure 4.16, the minimum width associated with $\hat{\alpha}^{\text{A}}_{\text{UE}}$ is approximately 0.032 for all the models. For the case of $\hat{\alpha}^{\text{B}}_{\text{UE}}$ and $\hat{\alpha}^{\text{C}}_{\text{UE}}$ the minimum widths are approximately 0.047 and 0.077 respectively. Near $\alpha = -1$

$$\text{Width}(\hat{\alpha}^{\text{A}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}^{\text{B}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}^{\text{C}}_{\text{UE}}). \tag{4.25}$$

   For $\alpha < 0$, the widths of the 90% CI given α are similar for all the location-based estimators. They are maximum near $\alpha = 0$.

2. For all α and T, the width of the CI varies with the form of $\mathbf{X}$ in (1.1). It increases from Model 1 to Model 3 but decreases with increasing T. In particular

$$\text{Width}(\text{CI}_1) \leq \text{Width}(\text{CI}_2) \leq \text{Width}(\text{CI}_3), \tag{4.26}$$

   where $\text{Width}(\text{CI}_j)$ denotes the width of the CI in Model j $(j = 1, 2, 3)$.

3. The restriction on the parameter space of $\hat{\alpha}_{\text{UE}}$ leads to an abrupt but uniform decrease in the width of the CI as the LDV coefficient $\alpha \to 1$. As apparent in Figures 4.16(d)-(f), the upper bound of the 90% confidence band is equal to 1 when α is close to 1.

4. From Figures 4.17 - (d) and (f) - when $\alpha > 0$ in Models 2 and 3, $\text{CI}^{0.05}(\hat{\alpha}_{\text{UE}})$ are characterized by an upward shift, relative to $\text{CI}^{0.05}(\hat{\alpha}_{\text{LS}})$. The corresponding 0.95 quantiles also shift upwards.

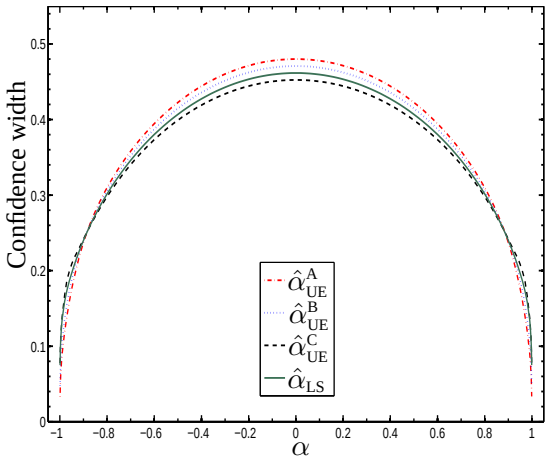
5. Given $T = 50$, the width of the 90% CI over Ω may be summarized as follows

   (a) In Model 1, for intermediate values i.e. $\alpha \in [-0.85, 0.85]$,

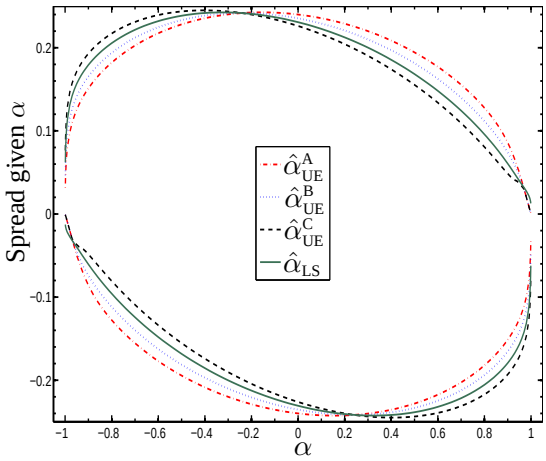
$$\text{Width}(\hat{\alpha}^{\text{C}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}_{\text{LS}}) \leq \text{Width}(\hat{\alpha}^{\text{B}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}^{\text{A}}_{\text{UE}}). \tag{4.27}$$

   (b) In Models 2 and 3, when $\alpha < -0.90$ the converse of (4.27) holds. However, for $\alpha \in [-0.90, 0.85]$

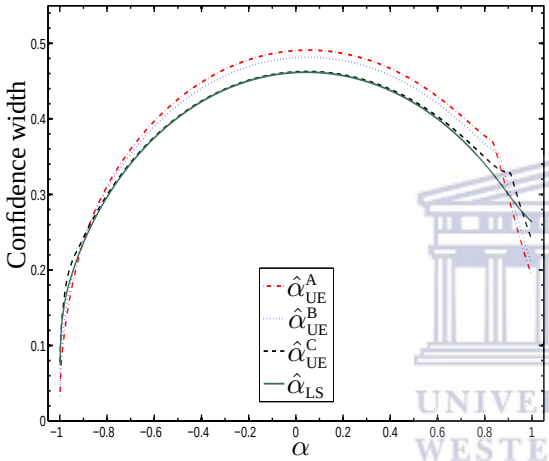
$$\text{Width}(\hat{\alpha}_{\text{LS}}) \lessapprox \text{Width}(\hat{\alpha}^{\text{C}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}^{\text{B}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}^{\text{A}}_{\text{UE}}), \tag{4.28}$$



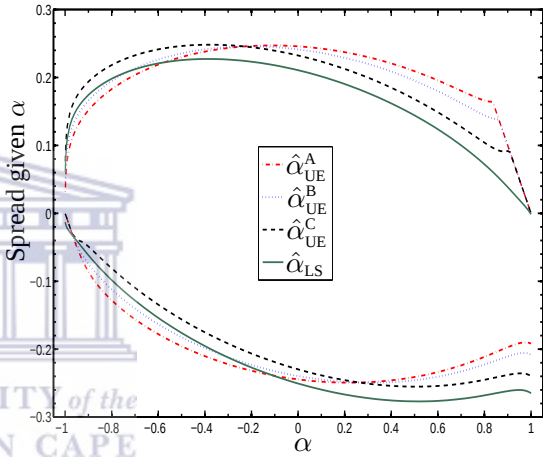
(a) Width: Model 1



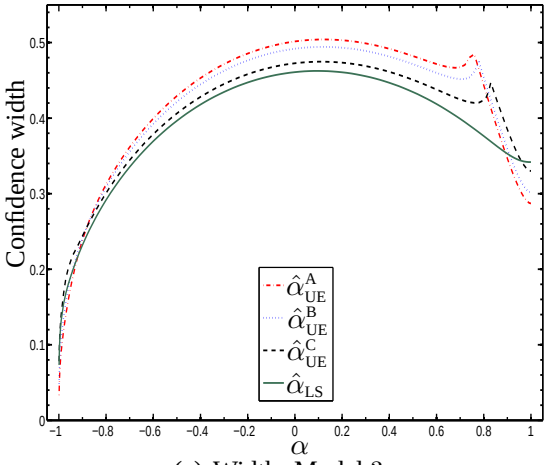
(b) Spread of the CI on Ω for given α: Model 1



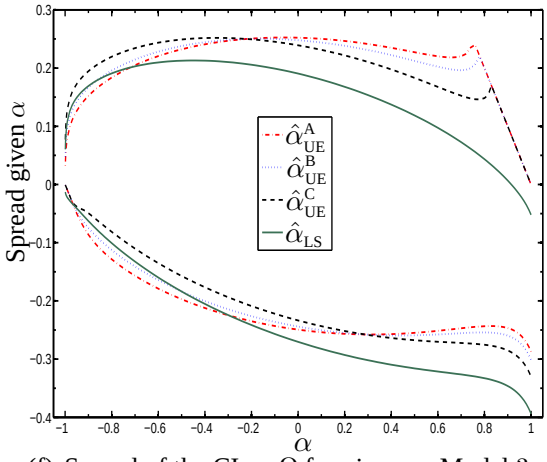
(c) Width: Model 2



(d) Spread of the CI on Ω for given α: Model 2



(e) Width: Model 3



(f) Spread of the CI on Ω for given α: Model 3

Figure 4.17: Comparison of CI of the estimators $\hat{\alpha}_{\text{UE}}$ in different models when $T = 50$ (see also Tables B20 and B21).

while for the near unit-root case i.e. $\alpha > 0.95$

$$\text{Width}(\hat{\alpha}^{\text{A}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}^{\text{B}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}^{\text{C}}_{\text{UE}}) \leq \text{Width}(\hat{\alpha}_{\text{LS}}). \tag{4.29}$$

4.3.3 Ratio of CI

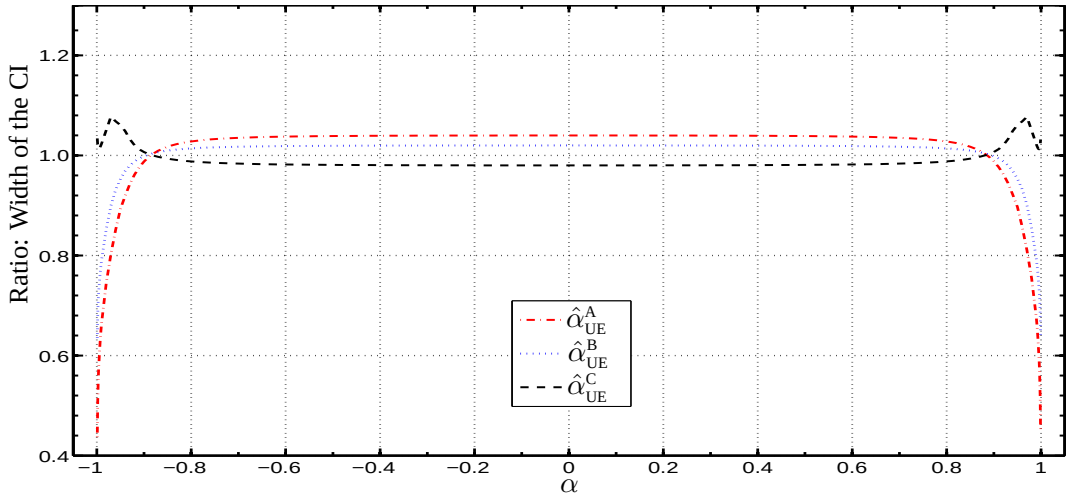
The CI of the estimators $\hat{\alpha}_{\text{UE}}$ at alternative level(s) and T follow the characteristics discussed above. As such, the stipulated regions in Ω can be used as a guide to inference about the LDV coefficient. The width of the CI indicate the *power* properties of the CI-based tests: although the coverage probability of the estimators are the same, the empirical power of the tests may differ.

For comparative purposes, the ratio $\frac{\text{Width}(\hat{\alpha}_{\text{UE}})}{\text{Width}(\hat{\alpha}_{\text{LS}})}$ was studied. Figure 4.18 summarizes the results in Equations (4.25) to (4.29).

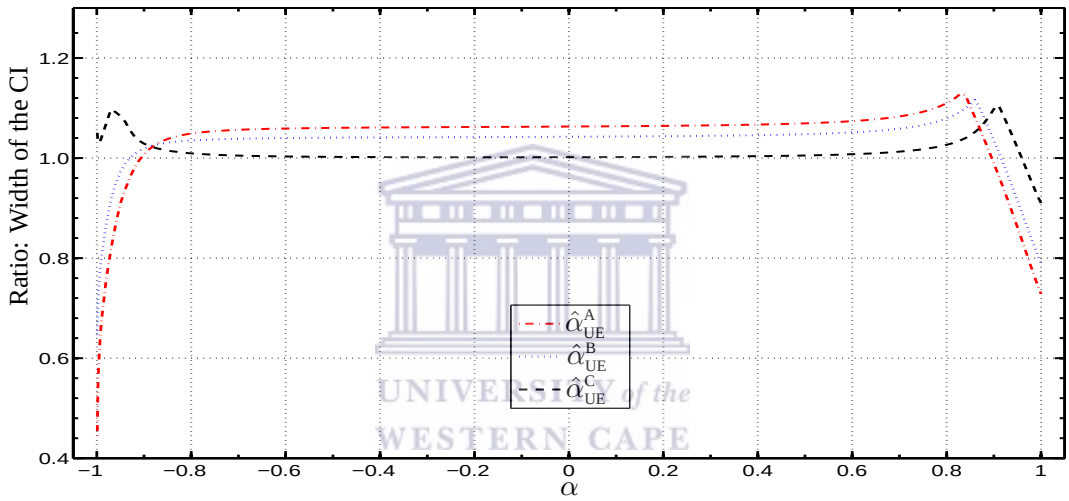
In general, when $|\alpha|$ is close to unity the width of the CI of the mean-based estimator $\hat{\alpha}^{\text{A}}_{\text{UE}}$ is smaller than that of the other estimators in all the models. In particular, for the near unit-root case Equations (4.25) to (4.29) imply that more exact (precise) inference can be achieved based on the CI of $\hat{\alpha}^{\text{A}}_{\text{UE}}$.

In Model 1, since the CI of $\hat{\alpha}^{\text{A}}_{\text{UE}}$ are narrower near $\alpha = 1$ than the CI of $\hat{\alpha}^{\text{B}}_{\text{UE}}$ and $\hat{\alpha}^{\text{C}}_{\text{UE}}$, tests based on the latter may show lower power on the unit-root hypotheses.

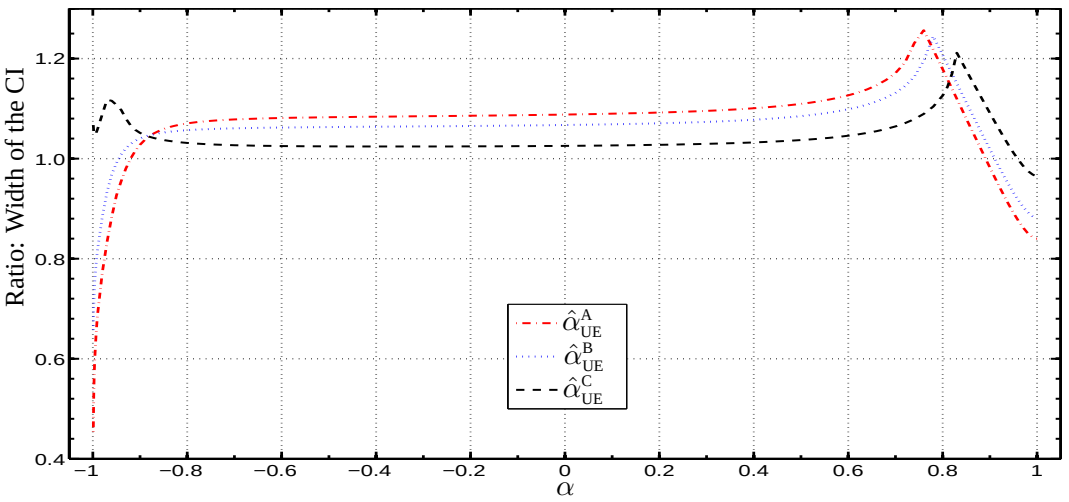
For intermediate values $\alpha \in [-0.85, 0.85]$ in all the models, the CI of $\hat{\alpha}^{\text{C}}_{\text{UE}}$ are tighter than the CI of $\hat{\alpha}^{\text{B}}_{\text{UE}}$ whereas the latter are narrower than the CI of $\hat{\alpha}^{\text{A}}_{\text{UE}}$. This may be a reason for preferring $\hat{\alpha}^{\text{C}}_{\text{UE}}$ in some circumstances.



(a) Ratio: $\frac{\text{Width}(\hat{\alpha}_{\text{UE}})}{\text{Width}(\hat{\alpha}_{\text{LS}})}$ in Model 1.



(b) Ratio: $\frac{\text{Width}(\hat{\alpha}_{\text{UE}})}{\text{Width}(\hat{\alpha}_{\text{LS}})}$ in Model 2.



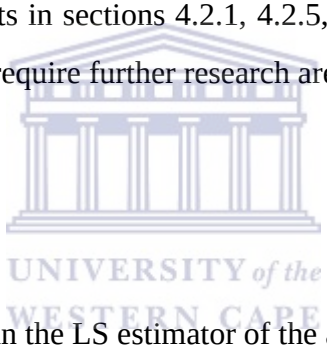
(c) Ratio: $\frac{\text{Width}(\hat{\alpha}_{\text{UE}})}{\text{Width}(\hat{\alpha}_{\text{LS}})}$ in Model 3.

Figure 4.18: Comparison of the ratio of the widths of the 90% CI for the estimators $\hat{\alpha}_{\text{UE}}$.

Chapter 5

Conclusions

This Chapter is organized as follows. Section 5.1 summarizes the work done in the preceding chapters. In section 5.2, the results in sections 4.2.1, 4.2.5, 4.2.6 and 4.3.2 are discussed in a broader sense. Some aspects that require further research are listed in section 5.3.



5.1 Review

In Chapter 1, the problem of bias in the LS estimator of the autoregressive coefficient in AR(1) models with or without exogenous regressors was discussed. A review of procedures aimed at reducing this bias was carried out and their performance (in terms of MSE) was mentioned. Particular attention was given to estimators proposed by Andrews (1993) and Tanizaki (2000).

In Chapter 2, the evaluation of the CDF of the LS estimator via numerical integration - using the Imhof (1961) algorithm - was illustrated. Transformation of a ratio of quadratic forms to a single quadratic form was also clarified. Computational aspects were expounded and useful restrictions leading to efficient truncation bounds to replace the infinite upper limits of integration, in the expression of the CDF, were proposed. The properties of these bounds were studied and compared with those given by Imhof (1961).

Using the form of the CDF of the LS estimator, an explicit computable form for the PDF of the LS estimator was derived for the first time. Its implementation was also discussed with regard

to accuracy and computation time or effort. A method of obtaining efficient truncation bounds for the integral of the PDF was proposed and their properties analysed.

The point estimators (median, mean and mode statistics of the LS estimator) of the LDV coefficient were evaluated from the expressions of the exact CDF and PDF functions. Determination of the exact confidence intervals for the LS estimator of α was also mentioned.

In Chapter 3, computation of the mean and median statistics using simulation was illustrated. Accuracy versus efficiency studies vis-a-vis the numerical integration method were conducted. Approximate standard errors in determining the mean and median functions (over Ω) using the simulation technique were given. Simulation was shown to be preferable to the numerical integration method when T is large.

A technique based on applying kernel smoothing to simulation results was described as an alternative to establishing the mode of the PDF of the LS estimator. Kernel results for the mode function of the LS estimator were compared with those based on the Imhof (1961) technique.

The method of obtaining the mean-, median- and mode-based estimators of the LDV coefficient from the respective location functions was given.

In Chapter 4, the properties of the location functions were described. The MSE and RMS of the LS estimator were determined numerically using its PDF. Their properties on the parameter space Ω were illustrated for a sample size $T = 50$.

The location functions were also used to establish the Jacobian of the transformation from $\hat{\alpha}_{\text{LS}}$ to $\hat{\alpha}_{\text{UE}}$. This information was used to construct a numerical scheme to evaluate the PDF of the location-based estimators. Due to their restriction on Ω, the density of the final estimators was shown to be characterized by a continuous and discrete part - in distinct contrast to that of the LS estimator. The first and second moments, MSE and RMS of the location-based estimators were determined. Their bias, variance and MSE over Ω were compared with those of the LS estimator.

A method of accurately establishing explicit rather than implicit (Andrews 1993) interval estimators of the LDV coefficient, using the CDF of the mean-, median- and mode-based estimators, was proposed. The properties of these exact bounds were explored by studying the respective 90% confidence intervals for a sample size $T = 50$.

5.2 Discussion

The bias in the LS estimator of the LDV coefficient in AR(1) models is quite significant in small samples especially when the model incorporates exogenous regressors and/or when the LDV coefficient is close to unity. Therefore, in agreement with the results in Abadir (1995), the results here indicate that estimation techniques that lead to smaller biases may be useful prior to the use of LS estimates for inference or otherwise.

A similarity was noted in the properties of the bias, variance and MSE of the LS and location-based estimators to other findings in the literature (Orcutt & Winokur 1969; Shaman & Stine 1988; Kiviet & Phillips 2003, 2005; etc.). In general, their variance and MSE are large for α close to zero. Furthermore, these quantities, as well as bias decreases with T.

The terminology used in Andrews (1993) and Tanizaki (2000), i.e. 'mean-unbiased' and 'median-unbiased', can be potentially confusing especially to a non-statistician. However, the discussion and results have shown that the estimators are biased in the usual sense: in expectation. The same applies to the estimator established from the mode function.

In fact, it was shown that the median- and mode-based estimators may be more biased than the LS estimator in a restricted subspace of Ω. However, all the location-based estimators show substantial improvements in terms of bias when the model incorporates exogenous variables and the LDV coefficient is close to 1.

In contrast to the other two location-based estimators, the mean-based estimator shows outstanding performance for all α and T, in terms of its bias. For the moderate sample size considered, bias is less than 40% that of the LS estimator, when α is close to unity. It is practically unbiased for other values of α in all the models: about 0% to 5% of the bias in the LS estimator. Of course, its good bias performance is due to the mean-unbiasedness criterion in (3.12).

Since the procedures leading to the location-based estimators and the corresponding confidence intervals incorporate the form of the exogenous regressor vector $\mathbf{X}$ in Equation (1.1) as well as the finite sample size, they offer an improvement over asymptotic results in terms of accuracy. In particular, methods based on asymptotic expansions have been shown to be unsatisfactory when the parameters of the exogenous regressor vector $\mathbf{X}$ and T are small (Kiviet & Phillips

2005). The approximate methods do not yield exact inferences (Dufour 1990): asymptotic critical values can be very unreliable, especially near the non-stationary border.

Therefore, the numerical results given provide a benchmark for analysis of the results and the performance of approximate methods for the moments, MSE or RMS. However, the significance of approximate methods cannot be overlooked. In fact, they complement the numerical results. In particular, moment approximation methods that utilize asymptotic expansions (Kendall 1954; Grubb & Symons 1987; Kiviet & Phillips 1993; Kiviet, Phillips & Schipp 1999) or exponential forms (Abadir 1993, 1995) provide analytic expressions that allow further theoretical investigation. Such expressions have been exploited to suggest improvement of inference and estimation methods in small and moderate samples.

The MSE criterion avails a small sample method to evaluate the performance of the various estimators and also provides a basis for selection of one over the other. It offers a common approach of balancing-off the desire for unbiasedness against that of minimum variance over the parameter space of the LDV coefficient α. Although the variance of the location-based estimators is generally greater than that of the LS estimator, the former show an improvement in the RMS especially when α is large. This is because of the bias improvement. Near the unit root boundary, the mean-based estimator shows better results than the median- or mode-based estimators. For example, its MSE ($T = 50$) is about 48%, 37% and 31% that of the LS estimator in Models 1, 2 and 3 respectively: this implies that the RMS is approximately 70%, 60% and 55%.

However, it is common practice to trade off some increase in bias for a decrease in variance or vice versa. This depends on the use to which the estimators are put. The numerical results indicate that the mode-based estimator may be preferred to the median- and mean-based estimators since it has minimum variance. None of the location-based estimators is a *minimum variance unbiased estimator (MVUE)* except when the LDV coefficient is near $\alpha = 0$.

The properties of the density need to be studied further. It has been shown (Forchini 2002) that the closed form expression of the exact CDF of a ratio of quadratic forms may fail to be analytic at certain points of its domain.

The results here also illuminate the factors which are important in determining the bias, variance

and the MSE of estimators in dynamic regression models and the need for correct specification of the models. As mentioned in the Preface, given that the models have enough support from the data, most interest is in the specification of plausible TS and DS models: unit root and stationarity tests. A majority of the tests are asymptotic based, but have been modified to improve their small-sample performance (Broda, Carstensen & Paolella 2004).

However, although the empirical evidence (Perron 1988; Phillips & Perron 1988; Dejong & Whiteman 1991; Rudebusch 1992; etc.) obtained using such tests justifies an assumption of a unit root in the autoregressive representation of many time series, it is extremely robust across various data samples and test procedures.

The discriminatory power of standard tests has been examined and/or challenged by, amongst others: Perron (1988), incorporating structural trend breaks; Phillips (1991), Schotman and van Dijk (1991), and Stock (1994), using Bayesian techniques; Phillips (1987, 1988), Ng and Perron (1995), and Perron and Ng (1996), allowing more general innovation processes; Shin and So (2001), using recursive mean adjustment; Xiao (2001b), for stationarity tests; Xiao (2001a), using near-unit-root tests; Elliot, Rothenberg and Stock (1996), and Thompson (2002, 2003a), using efficient unit root tests; Juhl and Xiao (2003), using point optimal tests; Hasan and Koenker (1997), Rothenberg and Thompson (2003), and Thompson (2003b), using rank based tests; and Broda, Carstensen and Paolella (2004), using nearly efficient unit root tests in small samples. The general result is that the power of unit root tests and their counterpart stationarity tests can be quite low.

However, the indictment of low power does not compromise the results from these tests. In fact, all statistical tests have low power against alternatives that are 'local' to the null. Whilst using large-sample first-order approximate bias-correction, Marriott and Pope (1954), and Rudebusch (1993) noted that when the LDV coefficient of the estimated TS model is close to unity, the consequences of the use of bias-corrected estimators on the unit-root hypothesis are that:

1. The estimated LDV coefficient of the bias-corrected TS model is much closer to 1.
2. The bias-corrected TS model exhibits greater persistence to random shocks than the un-corrected model. The model reverts to trend slightly less rapidly than the LS estimated model and will be harder to distinguish from the DS model.

3. The power of unit root tests is much lower against the bias-corrected TS model than against the uncorrected TS model.

4. The impulse response of the bias-corrected TS model is indistinguishable from that of the LS estimated DS model (when bias is ignored) or the bias-corrected DS model.

It follows that, for a particular data sample, the question on existence of a unit root is even more uncertain.

5.3 Subsequent research

1. The eigenvalue approach to the calculation of the PDF and CDF was used. However, it has been shown (Farebrother 1990) that the diagonalization or tridiagonalization methods are more computationally intensive than eigenvalue-free methods (Farebrother 1990; Shively, Ansley & Kohn 1990; Kohn, Shively & Ansley 1993). These techniques could be tried and accuracy checks conducted.

2. The point and interval estimators developed apply to AR(1) models. It would be interesting to establish whether the properties of the location-based estimators of the LDV coefficient for the AR(1) series carries over to AR(p) models. In fact, the procedures in Andrews and Chen (1994) entails transforming the higher order processes to an AR(1).

3. Other stochastic simulation methods in Fair (1996) and Tanizaki (2000) may also be investigated and extended to the new estimator (mode-based estimator).

4. The empirical use of the mean- and mode-based estimators in typical data sets such as the Nelson and Plosser (1982) series of macroeconomic models vis-a-vis other estimators is still to be tested. The median-based estimator has been considered for instance in Rudebusch (1992, 1993) and Lopez, Murray and Papell (2004).

5. Investigation as to whether the procedures carry over to a Bayesian framework would be interesting.

6. Given the results in Forchini (2002), it is worthwhile to ascertain whether the density of the LS estimator is analytic over its entire domain. For example, for Model 1, it is well known that the LS estimator is superconsistent when $\alpha = 1$; the distribution has different functional forms over different intervals separated by this point of nonanalyticity.

7. The exact numerical results obtained here assumed normal IID innovations. Their robustness under different disturbance structures may be considered.

8. Variation of the forms of the exogenous regressor(s) other than the constant and/or time-trend, is also a feasible venture.

Notes

Note 1 : Unit root tests and stationarity tests

The literature on testing for unit roots and stationarity is enormous. Phillips and Xiao (1998) provide a chronology of the development.

Asymptotic properties of the LS estimator

As shown in section 1.2, the behavior of the LS estimator $\hat{\alpha}_j$ $(j = 1, 2, 3)$ varies over Ω. In the asymptotic framework, the same follows. If $|\alpha| < 1$, the rate of convergence to the limiting distribution is $\sqrt{T}$. If $\alpha = 1$, the estimator converges to the true value at a super-consistent rate T. For $|\alpha| < 1$, the limiting distributions are standard normal.

When $\alpha = 1$, Dickey and Fuller (1979, 1981) derived a representation for the distribution of the LS estimator in Model 1. Phillips (1987) characterized the limiting distribution as the ratio of a χ^2-variate and a non-standard distribution expressed in terms of functionals of a *Wiener process*. Phillips (1988) extended the results in Phillips (1987) to incorporate a drift and time-trend in the model specification (Models 2 and 3, respectively).

For nearly-integrated series, the distributions converge weakly to functionals of a diffusion process rather than standard Brownian motion. In both cases, the distributions depend on whether a constant or time-trend is included in the regression to be estimated. Critical values can only be obtained using Monte Carlo experiments. The leading cases are tabulated in Fuller (1996).

Unit root tests

The classical unit root tests are due to Dickey and Fuller (1979). They are commonly referred to as *Dickey-Fuller tests*. They were developed to detect the presence of a unit root in a general I(1) process: they are based on the regressions (1.4) to (1.6). The tests are based on the LS estimator $\hat{\alpha}_{\text{LS}} \equiv \hat{\alpha}_j$ $(j = 1, 2, 3)$ in finite samples using the test statistics

$$T(\hat{\alpha}_{\text{LS}} - 1) \quad \text{and} \tag{NB1}$$

$$\mathfrak{t} = \frac{\hat{\alpha}_{\text{LS}} - 1}{\hat{\sigma}^2_{\hat{\alpha}_{\text{LS}}}}, \tag{NB2}$$

where $\hat{\sigma}^2_{\hat{\alpha}_{\text{LS}}}$ is the estimated standard error of $\hat{\alpha}_{\text{LS}}$. The *coefficient tests* in (NB1) use the super-consistency property of the LS estimator $\hat{\alpha}_{\text{LS}}$ near $\alpha = 1$, whereas the $\mathfrak{t}$*-ratio tests* in (NB2) are based on the usual regression $\mathfrak{t}$-statistic. The latter does not have the usual Student-$\mathfrak{t}$ distribution, but is skewed toward negative values. The critical values are determined from the asymptotic distributions discussed earlier. Common derived procedures include the *Phillips and Perron (PP) tests* (Phillips 1987; Phillips & Perron 1988) and the *Augmented Dickey-Fuller (ADF) tests* (Said & Dickey 1984) which only applies to (NB2).

The null and alternative hypothesis are

$$\begin{aligned} \text{H}_0 \quad &: \quad \alpha = 1 \qquad \{y_t\} \text{ is difference stationary,} \\ \text{H}_1 \quad &: \quad |\alpha| < 1 \qquad \{y_t\} \text{ is trend stationary.} \end{aligned} \tag{NB3}$$

Stationarity tests

Stationarity tests are based on testing for the moving average root in Δy_t where Δ is the difference operator defined by $\Delta y_t = y_t - y_{t-1}$. Alternatively, these can be considered as tests that the *autoregressive moving average (ARMA)* representation $\Delta y_t = \Phi(\mathfrak{B})u_t$ is *non-invertible* where $\Phi(\mathfrak{B})$ is an operator in the usual back-shift operator $\mathfrak{B}$ defined by $\mathfrak{B}y_t = y_{t-1}$. The statistic for a common test due to Kwiatkowski *et al.* (1992), is given as

$$\text{KPSS} = \left(\frac{1}{T^2} \sum_{t=1}^{T} \hat{S}_t \right) / \Upsilon^2, \tag{NB4}$$

where $\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j$, and $\hat{u}_j$ are the estimated residuals from regressions in (1.4) to (1.6) and $\hat{\Upsilon}^2$ is the long run variance of u_t estimated using $\hat{u}_t$.

The null and alternative hypothesis are

$$\begin{aligned} \mathrm{H}_0 &: \Phi(1) = 0 \qquad \text{characteristic polynomial } \Phi(z) = 0 \text{ has a unit root,} \\ \mathrm{H}_1 &: \Phi(1) > 0 \qquad \Phi(z) \text{ has root outside the unit circle.} \end{aligned} \tag{NB5}$$

King (1980), Nabeya and Tanaka (1988), Tanaka (1990), and Leybourne and McCabe (1994) also consider stationarity tests for the models in (1.4) to (1.6).

Note 2 : Transformation method: mode estimation

In section 3.3, it was mentioned that the mode can be estimated by first transforming the data to approximate normality where the mean, median and mode are the same estimand. Bickel (2003) proposed using a power transformation. This Note summarizes the results obtained after applying the *Fisher* $\mathfrak{Z}$*-transformation* to the case where the distribution of the LS estimator is skewed: near-unit root.

Given a set of sample data $\{X_i\}_{i=1}^{N}$, a new set $\{Z_i\}_{i=1}^{N}$ can be obtained using the transformation

$$Z_i = \text{arctanh}(X_i) = \frac{1}{2} \ln \left(\frac{1 + X_i}{1 - X_i} \right) \qquad i = 1, 2, \ldots, N. \tag{NB6}$$

Figure 5.1 (left panels) show histograms of $N = 40000$ replicates of the LS estimator given $\alpha = 0.995$ and $\alpha = 0.999$ for Model 1 $(T = 40)$. The panels on the right show transformed data obtained using (NB6) and the estimates of α using the mean and median. It is evident that the transformation alleviates the skewness. However, from the exact results - for the mean and median functions - given in Tables B1 and B2, it is clear that both statistics are biased upward when using this method. Furthermore, the two statistics are not equal.

By implication, the mode can not be estimated adequately after tarnsformation to approximate symmetry (Bickel 2002, 2003).

The results should not be surprising: the limitation of the method can be seen for the case[1] if

[1]The occurrence $\hat{\alpha}_i^* = 1$ for $i = 1, \ldots, N$ (see section 3.2) is observed to be rare - numerically. In fact, **it does not occur in the reported results**; the mean statistic is not affected when analysing the effectiveness of method.

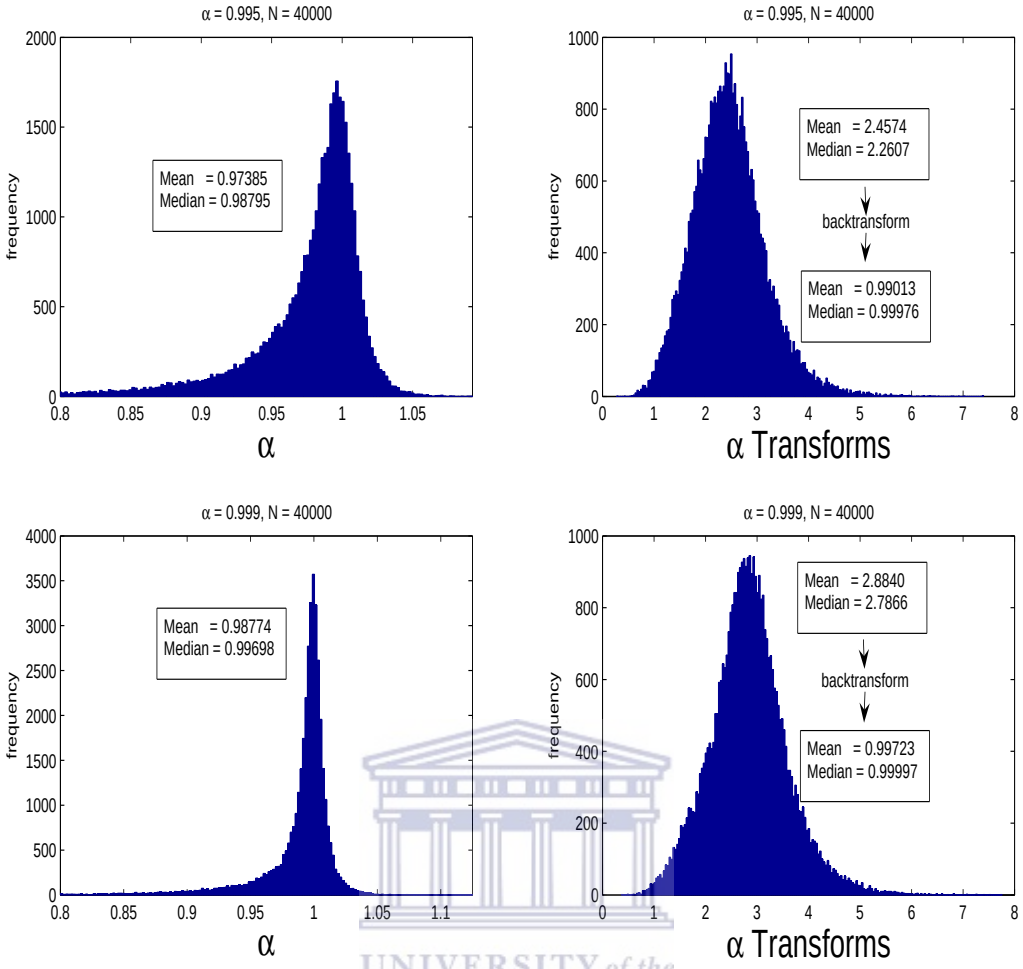


Figure 5.1: Histograms showing the effect of Fisher 3 transformation on LS estimated data $(T = 40)$.

$X_i = 1$ in (NB6). However, other transformations can be tried. An immediate example is the Box and Cox (1964) family of transformations. Variants of the latter are sometimes applied to transform data exhibiting non-normality to approximate normality (Wand & Jones 1995). In fact, the *Transformation kernel density estimator* uses suitable transforms. For example, for heavily skewed data a possible two-parameter family of convex transformations is the *shifted power family* defined by

$$T(X_i; \lambda_1, \lambda_2) = \begin{cases} (X_i - \lambda_1)^{\lambda_2} \text{sign}(\lambda_2), & \text{if} \quad \lambda_2 \neq 0, \\ \ln(X_i + \lambda_1) & \text{if} \quad \lambda_2 = 0, \end{cases} \tag{NB7}$$

where λ_1, λ_2 are parameters. (NB7) is an extension of the Box-Cox family of transformations. However, this method is not pursued further.

Note 3 : Symmetry, Eigenvalues and Eigenvectors

This Note highlights the importance of the symmetry of $\mathbf{W}_{\alpha jx}$ in (2.8) - mentioned when deriving the CDF expression in (2.13) - to the evaluation of the PDF in (2.24), determination and numerical evaluation of the derivatives of the eigenvalues in (2.25).

A real symmetric matrix has only real eigenvalues (Magnus & Neudecker 1999, p. 14). In particular, if $\mathbf{A}$ is a real and symmetric $n \times n$ matrix and $\mathbf{w}$ is an eigenvector associated with a simple eigenvalue λ of $\mathbf{A}$, the derivative of $\lambda = \lambda(\mathbf{A})$ is defined by the differential

$$d\lambda = \frac{\mathbf{w}'\,(d\mathbf{A})\,\mathbf{w}}{\mathbf{w}'\mathbf{w}} = \mathbf{w}'\,(d\mathbf{A})\,\mathbf{w} \tag{NB8}$$

(Magnus & Neudecker 1999, pp. 157f).

Equation (NB8) holds only for symmetric matrices where the normalization constant $\mathbf{w}'\mathbf{w} = 1$. Since the eigenvalues are real, the corresponding eigenvectors are also real. (NB8) is implemented as (2.25).

However, if symmetry is not required (Magnus & Neudecker 1999), the eigenvalues may be complex. Consequently, the eigenvectors are also complex. The analogous result to (NB8) is

$$d\lambda = \frac{\mathbf{w}_1^*\,(d\mathbf{A})\,\mathbf{w}_2}{\mathbf{w}_1^*\mathbf{w}_2}, \tag{NB9}$$

where $\mathbf{w}_1$ and $\mathbf{w}_2$ are the left and right eigenvectors associated with λ. $\mathbf{w}_1^*$ is the complex conjugate of $\mathbf{w}_1$. Normalization using $\mathbf{w}_2^*\mathbf{w}_2 = 1$ "leads to trouble" (Magnus & Neudecker 1999, p. 162).

Implementation of (NB9) is more complex than (NB8), since it requires more computations. That is:

1. A check on whether λ is real or complex.
2. Determination of $\mathbf{w}_1$, $\mathbf{w}_2$ and their conjugates.
3. Evaluation of the normalization factor.

It follows that symmetry of $\mathbf{W}_{\alpha jx}$ in (2.8) - which guarantees real eigenvalues - is even more useful for the case of evaluation of the PDF.

Furthermore, (NB8) applies only to simple eigenvalues. The occurrence of multiplicities is problematic when differentiating eigenvalues and eigenvectors. This is because the conditions of the *implicit function theorem* (Magnus & Neudecker 1999, pp. 142f) are not satisfied: λ is not differentiable everywhere. However, for lengthy series it is somewhat improbable for multiplicities to occur in practice.

Appendices

Appendix A : Mathematical

Notation: The LS estimator of α

The derivation of the LS estimator of α, $\hat{\alpha}_j$ $(j = 1, 2, 3)$ in (1.7), is provided below. The aim is to minimize the residual sum of squares, RSS $= \mathbf{u}'\mathbf{u} = (\mathbf{y} - \alpha\mathbf{y}_{-1} - \mathbf{X}\beta)'(\mathbf{y} - \alpha\mathbf{y}_{-1} - \mathbf{X}\beta)$ with $\mathbf{y} = (y_2, y_3, \ldots, y_T)'$ and $\mathbf{y}_{-1} = (y_1, y_2, \ldots, y_{T-1})'$. The k *normal equations* are

$$\frac{\partial \text{RSS}}{\partial \beta} = -2\mathbf{X}'(\mathbf{y} - \alpha\mathbf{y}_{-1} - \mathbf{X}\beta) = 0 \quad \text{and} \tag{A1}$$

$$\frac{\partial \text{RSS}}{\partial \alpha} = -2\mathbf{y}_{-1}{}'(\mathbf{y} - \alpha\mathbf{y}_{-1} - \mathbf{X}\beta) = 0. \tag{A2}$$

From (A1) it is trivial to show that $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \alpha\mathbf{y}_{-1})$. Substituting for β in (A2) gives

$$\begin{aligned}
&-\mathbf{y}_{-1}{}'\mathbf{y} + \alpha\mathbf{y}_{-1}{}'\mathbf{y}_{-1} + \mathbf{y}_{-1}{}'\mathbf{X}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \alpha\mathbf{y}_{-1})\right] = 0, \\
&-\mathbf{y}_{-1}{}'\mathbf{y} + \alpha\mathbf{y}_{-1}{}'\mathbf{y}_{-1} + \mathbf{y}_{-1}{}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \alpha\mathbf{y}_{-1}{}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{-1} = 0, \\
&\alpha\mathbf{y}_{-1}{}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y}_{-1} - \mathbf{y}_{-1}{}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y} = 0,
\end{aligned}$$

where $\mathbf{I}$ is a $(T-1) \times (T-1)$ identity matrix. It follows that

$$\hat{\alpha}_j = \frac{\mathbf{y}_{-1}{}'\left(\mathbf{I} - \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\right)\mathbf{y}}{\mathbf{y}_{-1}\left(\mathbf{I} - \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\right)\mathbf{y}_{-1}} = \frac{\mathbf{y}_{-1}{}'\left(\mathbf{I} - \mathbf{P}_j\right)\mathbf{y}}{\mathbf{y}_{-1}{}'\left(\mathbf{I} - \mathbf{P}_j\right)\mathbf{y}_{-1}}, \tag{A3}$$

where $\mathbf{P}_j = \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$. The matrices $\mathbf{X}_j$ are defined by

$$\mathbf{X}_1 = \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \mathbf{X}_2 = \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad \mathbf{X}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T-1 \end{bmatrix}.$$

Alternative derivation

The LS estimator of α can also be obtained using the partitioned inverse matrix of $\mathcal{Z}'\mathcal{Z}$ in (1.3). The LS estimator of $\hat{\Gamma}$ is

$$\hat{\Gamma} = \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{y}_{-1} \\ \mathbf{X}'\mathbf{y}_{-1} & \mathbf{y}_{-1}'\mathbf{y}_{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{y}_{-1}'\mathbf{y} \end{bmatrix}. \tag{A4}$$

The normal equations $\mathcal{Z}'\mathcal{Z}\Gamma = \mathcal{Z}'y$ can be written as

$$\mathbf{X}'\mathbf{X}\beta + \mathbf{X}'\mathbf{y}_{-1}\alpha = \mathbf{X}'\mathbf{y} \quad \text{and} \tag{A5}$$

$$\mathbf{y}_{-1}'\mathbf{X}\beta + \mathbf{y}_{-1}'\mathbf{y}_{-1}\alpha = \mathbf{y}_{-1}'\mathbf{y}. \tag{A6}$$

From (A5), it follows that

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{y}_{-1}\alpha). \tag{A7}$$

If (A5) is pre-multiplied by $\mathbf{y}_{-1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, then

$$\mathbf{y}_{-1}'\mathbf{X}\beta + \mathbf{y}_{-1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{-1}\alpha = \mathbf{y}_{-1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{A8}$$

Substituting for the first term in (A8) from (A6), then

$$\begin{aligned} \left[\mathbf{y}_{-1}'\mathbf{y}_{-1} - \mathbf{y}_{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{-1}\right]\alpha &= \mathbf{y}_{-1}'\mathbf{y} - \mathbf{y}_{-1}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \\ \left[\mathbf{y}_{-1}'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{y}_{-1}\right]\alpha &= \mathbf{y}_{-1}'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{y}. \end{aligned} \tag{A9}$$

The result (A3) follows from (A9).

Integrand at the origin: CDF function

Using l'Hôpital's rule, from (2.13) and differentiating with respect to u, it follows that

$$\lim_{u\to 0}\frac{\sin\phi}{u\rho} = \lim_{u\to 0}\left[\frac{\cos\phi\,\frac{\partial\phi}{\partial u}}{\rho + u\frac{\partial\rho}{\partial u}}\right], \tag{A10}$$

where ϕ and ρ are functions of u and x as defined in (2.14) and (2.15) respectively. Since $\partial\lambda(x)/\partial u = 0$, then

$$\frac{\partial\phi}{\partial u} = \frac{1}{2}\sum_{i=1}^{T}\frac{\lambda_i}{1+\lambda_i^2u^2} \quad \text{and} \tag{A11}$$

$$\begin{aligned}\frac{\partial\rho}{\partial u} &= \frac{1}{2}\left[\frac{\lambda_1^2u}{(1+\lambda_1^2u^2)}\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}+\cdots+\frac{\lambda_T^2u}{(1+\lambda_T^2u^2)}\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}\right]\\ &= \frac{1}{2}\sum_{i=1}^{T}\frac{\lambda_i^2u}{1+\lambda_i^2u^2}\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}.\end{aligned} \tag{A12}$$

Since $\lim_{u\to 0}\phi(u,x)=0$, substituting (A11) and (A12) into (A10), it follows that

$$\begin{aligned}\lim_{u\to 0}\frac{\sin\phi(u,x)}{u\,\rho(u,x)} &= \lim_{u\to 0}\left[\frac{\cos\phi\,\frac{1}{2}\sum_i\frac{\lambda_i}{1+\lambda_i^2u^2}}{\prod_k(1+\lambda_i^2u^2)^{1/4}\left\{1+\frac{1}{2}\sum_i\frac{\lambda_i^2u^2}{1+\lambda_i^2u^2}\right\}}\right]\\ &= \frac{1}{2}\sum_i\lambda_i,\end{aligned} \tag{A13}$$

where i and k varies from 1 to T for the sums and product in (A13).

Differentiation of the CDF to obtain the PDF of the LS estimator

Given $F_{\text{LS}}(u,x)$ in (2.13), it follows from (2.23) that

$$f_{\text{LS}}(x) = -\frac{1}{\pi}\int_0^\infty \frac{u\rho\left\{\cos\phi\,\frac{\partial\phi}{\partial x}\right\}-\sin\phi\left\{\frac{d[u\,\rho]}{\partial x}\right\}}{u^2\,\rho^2}\,du, \tag{A14}$$

where ϕ and ρ are given in (2.14) and (2.15) respectively.

The derivatives are given by:

$$\frac{\partial\phi}{\partial x} = \frac{1}{2}\sum_{i=1}^{T}\frac{u}{(1+\lambda_i^2u^2)}\frac{\partial\lambda_i}{\partial x} \equiv \frac{1}{2}\sum_{i=1}^{T}\frac{\lambda_i'u}{(1+\lambda_i^2u^2)}. \tag{A15}$$

$$\begin{aligned}\frac{\partial\rho}{\partial x} &= \frac{1}{2}\left[\frac{u^2\lambda_1\lambda_1'}{(1+\lambda_1^2u^2)}\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}+\cdots+\frac{u^2\lambda_T\lambda_T'}{(1+\lambda_T^2u^2)}\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}\right]\\ &= \frac{1}{2}\sum_{i=1}^{T}\left[\frac{u^2\lambda_i\lambda_i'}{(1+\lambda_i^2u^2)}\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}\right].\end{aligned} \tag{A16}$$

Substituting the result in (A15) and (A16) into (A14) gives

$$
\begin{aligned}
f_{\text{LS}}(x) &= -\frac{1}{\pi}\int_0^\infty \frac{u^2\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}\left\{\frac{1}{2}\sum_{i=1}^{T}\left[\frac{\lambda_i'}{(1+\lambda_i^2u^2)}\left(\cos\phi-\lambda_i u\sin\phi\right)\right]\right\}}{u^2\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{2}}}\,du \\
&= \frac{1}{2\pi}\int_0^\infty \frac{\sum_{i=1}^{T}\left[\frac{\lambda_i'}{(1+\lambda_i^2u^2)}\left(\lambda_i u\sin\phi-\cos\phi\right)\right]}{\prod_{k=1}^{T}(1+\lambda_i^2u^2)^{\frac{1}{4}}}\,du \\
&= \frac{1}{2\pi}\int_0^\infty \frac{1}{\rho}\sum_{i=1}^{T}\left\{\frac{\cos\phi-\lambda_i u\sin\phi}{(1+\lambda_i^2u^2)}\,\mathbf{w}_i'\,\mathbf{R}_\alpha'[\mathbf{D}_1'(\mathbf{I}-\mathbf{P}_j)\mathbf{D}_1]\mathbf{R}_\alpha\,\mathbf{w}_i\right\}du, \quad \text{(A17)}
\end{aligned}
$$

where $\mathbf{w}_i$ is the eigenvector associated to λ_i. The matrices $\mathbf{R}_\alpha$, $\mathbf{D}_1$ and $\mathbf{P}_j$ $(j=1,2,3)$ follow from (2.8) given (2.25).

Efficient truncation bound: PDF evaluation

From (2.19), for the product term $\rho(u,x)$, then

$$
\frac{1}{\prod_{i=m+1}^{T}(1+|\lambda|_{(i)}^2u^2)^{\frac{1}{4}}} < \frac{1}{\prod_{i=m+1}^{T}[|\lambda|_{(i)}u]^{\frac{1}{2}}}. \quad \text{(A18)}
$$

From the expression for the truncation error B_{trc} in (2.26) and the restrictions following the use of order statistic $|\lambda|_{(i)}$ in (2.27) with (A18) above, it follows that

$$
\begin{aligned}
B_{\text{trc}} &= \frac{1}{2\pi}\int_b^\infty \frac{1}{\rho}\sum_{i=1}^{T}\frac{\lambda_i u\sin\phi-\cos\phi}{(1+\lambda_i^2u^2)}\,\lambda_i' du \\
&< \frac{1}{2\pi}\int_b^\infty \frac{1}{\prod_{i=m+1}^{T}[|\lambda|_{(i)}u]^{\frac{1}{2}}}\left\{\sum_{i=1}^{n}|\lambda_i'|\left[|\lambda|_{(i)}u+1\right]+\sum_{i=n+1}^{T}\frac{|\lambda|_{(i)}u+1}{|\lambda|_{(i)}^2u^2}\,|\lambda_i'|\right\}du \\
&= \frac{1}{2\pi\prod_{i=m+1}^{T}|\lambda|_{(i)}^{\frac{1}{2}}}\int_0^\infty\left\{\sum_{i=1}^{n}\frac{|\lambda|_{(i)}|\lambda_i'|}{u^{\frac{T-m-2}{2}}}+\sum_{i=1}^{n}\frac{|\lambda_i'|}{u^{\frac{T-m}{2}}}+\cdots\right. \\
&\qquad\left.\cdots+\sum_{i=n+1}^{T}\frac{|\lambda|_{(i)}|\lambda_i'|}{|\lambda|_{(i)}^2u^{\frac{T-m+2}{2}}}+\sum_{i=n+1}^{T}\frac{|\lambda_i'|}{|\lambda|_{(i)}^2u^{\frac{T-m+4}{2}}}\right\}du \\
&= \frac{-1}{\pi\prod_{i=m+1}^{T}|\lambda|_{(i)}^{\frac{1}{2}}}\left[\sum_{i=1}^{n}\frac{|\lambda|_{(i)}|\lambda_i'|}{(T-m-4)u^{\frac{T-m-4}{2}}}+\sum_{i=1}^{n}\frac{|\lambda_i'|}{(T-m-2)u^{\frac{T-m-2}{2}}}+\cdots\right. \\
&\qquad\left.\cdots+\sum_{i=n+1}^{T}\frac{|\lambda|_{(i)}|\lambda_i'|}{(T-m)u^{\frac{T-m}{2}}|\lambda|_{(i)}^2}+\sum_{i=n+1}^{T}\frac{|\lambda_i'|}{(T-m+2)u^{\frac{T-m+2}{2}}|\lambda|_{(i)}^2}\right]_b^\infty \\
&< \frac{-1}{(T-m-4)\pi\prod_{i=m+1}^{T}|\lambda|_{(i)}^{\frac{1}{2}}}\left[\sum_{i=1}^{n}|\lambda|_{(i)}|\lambda_i'|+\sum_{i=1}^{n}|\lambda_i'|+\cdots\right. \\
&\qquad\left.\cdots+\sum_{i=n+1}^{T}\frac{|\lambda|_{(i)}|\lambda_i'|}{|\lambda|_{(i)}^2}+\sum_{i=n+1}^{T}\frac{|\lambda_i'|}{|\lambda|_{(i)}^2}\right]\left[\frac{1}{u^{\frac{T-m-4}{2}}}\right]_b^\infty \\
&= \frac{1}{(T-m-4)\,\pi\,b^{\frac{T-m-4}{2}}\prod_{i=m+1}^{T}|\lambda|_{(i)}^{\frac{1}{2}}}\left[\sum_{i=1}^{n}|\lambda_i'|\left[|\lambda|_{(i)}+1\right]+\sum_{i=n+1}^{T}\frac{|\lambda_i'|}{|\lambda|_{(i)}^2}\left[|\lambda|_{(i)}+1\right]\right] \\
&< \frac{S_n(|\lambda|_{(i)})}{(T-m-4)\,\pi\,b^{\frac{T-m-4}{2}}\prod_{i=m+1}^{T}|\lambda|_{(i)}^{\frac{1}{2}}}, \qquad \text{(A19)}
\end{aligned}
$$

where $S_n(|\lambda|_{(i)})$ is given in (2.29).

Appendix B : Tabular results

Tables B1-B9 give summarized exact results for the location functions obtained from the CDF and PDF in (2.13) and (2.24), respectively. For Model 1 the functions are exactly uneven, hence only values for $\alpha \geq 0$ are reported.

Table B1: Mean function of the LS estimator of α for Model 1.

α / T	40	50	60	70	80	90	100	125	150	200
0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	0.048	0.048	0.048	0.049	0.049	0.049	0.049	0.049	0.049	0.050
0.10	0.095	0.096	0.097	0.097	0.098	0.098	0.098	0.098	0.099	0.099
0.15	0.143	0.144	0.145	0.146	0.146	0.147	0.147	0.148	0.148	0.149
0.20	0.191	0.192	0.194	0.194	0.195	0.196	0.196	0.197	0.197	0.198
0.25	0.238	0.240	0.242	0.243	0.244	0.245	0.245	0.246	0.247	0.248
0.30	0.286	0.288	0.290	0.292	0.293	0.293	0.294	0.295	0.296	0.297
0.35	0.333	0.337	0.339	0.340	0.341	0.342	0.343	0.344	0.345	0.347
0.40	0.381	0.385	0.387	0.389	0.390	0.391	0.392	0.394	0.395	0.396
0.45	0.429	0.433	0.436	0.438	0.439	0.440	0.441	0.443	0.444	0.446
0.50	0.476	0.481	0.484	0.486	0.488	0.489	0.490	0.492	0.493	0.495
0.55	0.524	0.529	0.532	0.535	0.537	0.538	0.539	0.541	0.543	0.545
0.60	0.572	0.577	0.581	0.583	0.585	0.587	0.588	0.591	0.592	0.594
0.65	0.620	0.625	0.629	0.632	0.634	0.636	0.637	0.640	0.641	0.644
0.70	0.668	0.674	0.678	0.681	0.683	0.685	0.686	0.689	0.691	0.693
0.75	0.715	0.722	0.726	0.730	0.732	0.734	0.735	0.738	0.740	0.743
0.80	0.764	0.770	0.775	0.778	0.781	0.783	0.785	0.788	0.790	0.792
0.85	0.812	0.819	0.824	0.827	0.830	0.832	0.834	0.837	0.839	0.842
0.90	0.861	0.868	0.873	0.876	0.879	0.881	0.883	0.886	0.888	0.891
0.93	0.892	0.898	0.903	0.906	0.909	0.911	0.913	0.916	0.918	0.921
0.95	0.913	0.919	0.923	0.927	0.929	0.931	0.933	0.936	0.938	0.941
0.97	0.935	0.941	0.945	0.948	0.950	0.952	0.953	0.956	0.958	0.961
0.99	0.963	0.967	0.970	0.972	0.974	0.975	0.976	0.978	0.980	0.982
0.995	0.974	0.976	0.978	0.980	0.981	0.982	0.983	0.985	0.986	0.988
0.999	0.988	0.989	0.990	0.991	0.991	0.992	0.992	0.993	0.994	0.995

Table B2: Median function of the LS estimator of α for Model 1.

α / T	40	50	60	70	80	90	100	125	150	200
0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	0.049	0.049	0.049	0.049	0.049	0.049	0.050	0.050	0.050	0.050
0.10	0.098	0.098	0.098	0.099	0.099	0.099	0.099	0.099	0.099	0.100
0.15	0.146	0.147	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149
0.20	0.195	0.196	0.197	0.197	0.198	0.198	0.198	0.198	0.199	0.199
0.25	0.244	0.245	0.246	0.246	0.247	0.247	0.248	0.248	0.248	0.249
0.30	0.293	0.294	0.295	0.296	0.296	0.297	0.297	0.298	0.298	0.299
0.35	0.342	0.343	0.344	0.345	0.346	0.346	0.347	0.347	0.348	0.348
0.40	0.390	0.392	0.393	0.394	0.395	0.396	0.396	0.397	0.397	0.398
0.45	0.439	0.441	0.443	0.444	0.444	0.445	0.446	0.446	0.447	0.448
0.50	0.488	0.490	0.492	0.493	0.494	0.495	0.495	0.496	0.497	0.498
0.55	0.537	0.539	0.541	0.542	0.543	0.544	0.545	0.546	0.546	0.547
0.60	0.586	0.588	0.590	0.592	0.593	0.593	0.594	0.595	0.596	0.597
0.65	0.635	0.638	0.640	0.641	0.642	0.643	0.644	0.645	0.646	0.647
0.70	0.683	0.687	0.689	0.690	0.691	0.692	0.693	0.695	0.695	0.697
0.75	0.732	0.736	0.738	0.740	0.741	0.742	0.743	0.744	0.745	0.746
0.80	0.782	0.785	0.787	0.789	0.790	0.791	0.792	0.794	0.795	0.796
0.85	0.831	0.834	0.837	0.839	0.840	0.841	0.842	0.843	0.844	0.846
0.90	0.881	0.884	0.886	0.888	0.890	0.891	0.892	0.893	0.894	0.896
0.93	0.911	0.914	0.917	0.918	0.920	0.921	0.921	0.923	0.924	0.926
0.95	0.932	0.935	0.937	0.938	0.940	0.941	0.942	0.943	0.944	0.946
0.97	0.953	0.956	0.958	0.959	0.960	0.961	0.962	0.963	0.964	0.966
0.99	0.979	0.980	0.981	0.982	0.983	0.983	0.984	0.985	0.985	0.986
0.995	0.988	0.988	0.989	0.989	0.989	0.990	0.990	0.991	0.991	0.992
0.999	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.998

Table B3: Mode function of the LS estimator of α for Model 1.

α / T	40	50	60	70	80	90	100	125	150	200
0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.050	0.050	0.050
0.10	0.103	0.102	0.102	0.101	0.101	0.101	0.101	0.101	0.101	0.101
0.15	0.154	0.153	0.153	0.152	0.152	0.152	0.152	0.151	0.151	0.151
0.20	0.205	0.204	0.203	0.203	0.203	0.202	0.202	0.202	0.201	0.201
0.25	0.256	0.255	0.254	0.254	0.253	0.253	0.253	0.252	0.252	0.251
0.30	0.308	0.306	0.305	0.304	0.304	0.303	0.303	0.302	0.302	0.302
0.35	0.359	0.357	0.356	0.355	0.354	0.354	0.354	0.353	0.352	0.352
0.40	0.410	0.408	0.407	0.406	0.405	0.404	0.404	0.403	0.403	0.402
0.45	0.461	0.459	0.458	0.456	0.456	0.455	0.455	0.454	0.453	0.452
0.50	0.513	0.510	0.508	0.507	0.506	0.506	0.505	0.504	0.503	0.503
0.55	0.564	0.561	0.559	0.558	0.557	0.556	0.556	0.554	0.554	0.553
0.60	0.615	0.612	0.610	0.609	0.607	0.607	0.606	0.605	0.604	0.603
0.65	0.666	0.663	0.661	0.659	0.658	0.657	0.656	0.655	0.654	0.653
0.70	0.717	0.714	0.711	0.710	0.709	0.708	0.707	0.706	0.705	0.703
0.75	0.768	0.764	0.762	0.760	0.759	0.758	0.757	0.756	0.755	0.754
0.80	0.818	0.815	0.813	0.811	0.810	0.809	0.808	0.806	0.805	0.804
0.85	0.869	0.865	0.863	0.861	0.860	0.859	0.858	0.857	0.855	0.854
0.90	0.918	0.915	0.913	0.911	0.910	0.909	0.908	0.907	0.906	0.904
0.93	0.946	0.944	0.942	0.941	0.940	0.939	0.938	0.937	0.936	0.934
0.95	0.963	0.962	0.961	0.960	0.959	0.958	0.958	0.956	0.956	0.954
0.97	0.977	0.978	0.978	0.978	0.977	0.977	0.977	0.976	0.975	0.974
0.99	0.993	0.993	0.992	0.992	0.992	0.992	0.992	0.992	0.993	0.993
0.995	0.997	0.997	0.997	0.997	0.997	0.997	0.996	0.996	0.996	0.996
0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999

Table B4: Mean function of the LS estimator of α for Model 2.

α / T	40	50	60	70	80	90	100	125	150	200
-0.999	-0.988	-0.989	-0.990	-0.991	-0.991	-0.992	-0.992	-0.993	-0.994	-0.995
-0.80	-0.770	-0.775	-0.779	-0.782	-0.784	-0.785	-0.787	-0.789	-0.791	-0.793
-0.60	-0.583	-0.586	-0.588	-0.589	-0.591	-0.592	-0.592	-0.594	-0.595	-0.596
-0.40	-0.397	-0.397	-0.397	-0.398	-0.398	-0.398	-0.398	-0.399	-0.399	-0.399
-0.20	-0.211	-0.209	-0.207	-0.206	-0.205	-0.205	-0.204	-0.203	-0.203	-0.202
0.00	-0.026	-0.020	-0.017	-0.014	-0.013	-0.011	-0.010	-0.008	-0.007	-0.005
0.10	0.067	0.074	0.078	0.081	0.084	0.085	0.087	0.090	0.091	0.093
0.20	0.159	0.168	0.173	0.177	0.180	0.182	0.184	0.187	0.189	0.192
0.30	0.252	0.262	0.268	0.273	0.276	0.279	0.281	0.285	0.287	0.290
0.40	0.344	0.355	0.363	0.368	0.372	0.375	0.378	0.382	0.385	0.389
0.50	0.436	0.449	0.458	0.464	0.468	0.472	0.475	0.480	0.483	0.487
0.60	0.528	0.543	0.552	0.559	0.564	0.568	0.572	0.577	0.581	0.586
0.70	0.619	0.636	0.647	0.654	0.660	0.665	0.668	0.675	0.679	0.684
0.80	0.709	0.728	0.740	0.749	0.756	0.761	0.765	0.772	0.777	0.783
0.90	0.796	0.817	0.832	0.842	0.849	0.855	0.860	0.868	0.874	0.881
0.93	0.820	0.843	0.858	0.869	0.877	0.883	0.888	0.897	0.903	0.910
0.95	0.836	0.859	0.875	0.886	0.895	0.901	0.906	0.916	0.922	0.929
0.97	0.851	0.875	0.891	0.903	0.912	0.918	0.924	0.934	0.940	0.948
0.99	0.865	0.890	0.906	0.918	0.927	0.934	0.940	0.950	0.957	0.966
1.00	0.872	0.897	0.913	0.925	0.934	0.942	0.947	0.958	0.965	0.973

Table B5: Median function of the LS estimator of α for Model 2.

α / T	40	50	60	70	80	90	100	125	150	200
-0.999	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.998
-0.80	-0.787	-0.789	-0.791	-0.792	-0.793	-0.794	-0.794	-0.795	-0.796	-0.797
-0.60	-0.596	-0.597	-0.597	-0.598	-0.598	-0.598	-0.598	-0.599	-0.599	-0.599
-0.40	-0.406	-0.405	-0.404	-0.403	-0.403	-0.402	-0.402	-0.402	-0.401	-0.401
-0.20	-0.216	-0.213	-0.210	-0.209	-0.208	-0.207	-0.206	-0.205	-0.204	-0.203
0.00	-0.026	-0.021	-0.017	-0.015	-0.013	-0.011	-0.010	-0.008	-0.007	-0.005
0.10	0.069	0.075	0.079	0.082	0.085	0.086	0.088	0.090	0.092	0.094
0.20	0.164	0.171	0.176	0.180	0.182	0.184	0.186	0.189	0.191	0.193
0.30	0.258	0.267	0.273	0.277	0.280	0.282	0.284	0.287	0.289	0.292
0.40	0.353	0.363	0.369	0.374	0.377	0.380	0.382	0.385	0.388	0.391
0.50	0.447	0.458	0.465	0.470	0.474	0.477	0.480	0.484	0.486	0.490
0.60	0.541	0.554	0.562	0.567	0.572	0.575	0.577	0.582	0.585	0.589
0.70	0.635	0.649	0.658	0.664	0.669	0.672	0.675	0.680	0.684	0.688
0.80	0.727	0.743	0.753	0.760	0.765	0.769	0.773	0.778	0.782	0.787
0.90	0.816	0.834	0.846	0.854	0.861	0.865	0.869	0.876	0.880	0.885
0.93	0.841	0.860	0.873	0.882	0.888	0.893	0.897	0.904	0.909	0.915
0.95	0.857	0.877	0.890	0.899	0.906	0.912	0.916	0.923	0.928	0.934
0.97	0.872	0.893	0.906	0.916	0.923	0.929	0.933	0.941	0.947	0.953
0.99	0.886	0.907	0.921	0.931	0.939	0.945	0.950	0.958	0.964	0.971
1.00	0.893	0.914	0.928	0.938	0.946	0.952	0.957	0.965	0.971	0.978

Table B6: Mode function of the LS estimator of α for Model 2.

α / T	40	50	60	70	80	90	100	125	150	200
-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999
-0.80	-0.823	-0.819	-0.816	-0.814	-0.812	-0.811	-0.810	-0.808	-0.807	-0.805
-0.60	-0.625	-0.620	-0.617	-0.614	-0.613	-0.611	-0.610	-0.608	-0.607	-0.605
-0.40	-0.426	-0.421	-0.417	-0.415	-0.413	-0.411	-0.410	-0.408	-0.407	-0.405
-0.20	-0.226	-0.221	-0.217	-0.215	-0.213	-0.211	-0.210	-0.208	-0.207	-0.205
0.00	-0.027	-0.021	-0.018	-0.015	-0.013	-0.011	-0.010	-0.008	-0.007	-0.005
0.10	0.073	0.079	0.082	0.085	0.087	0.088	0.090	0.092	0.093	0.095
0.20	0.172	0.178	0.182	0.185	0.187	0.188	0.190	0.192	0.193	0.195
0.30	0.272	0.278	0.282	0.285	0.287	0.288	0.290	0.292	0.293	0.295
0.40	0.372	0.378	0.382	0.385	0.387	0.388	0.389	0.392	0.393	0.395
0.50	0.471	0.477	0.482	0.484	0.486	0.488	0.489	0.492	0.493	0.495
0.60	0.570	0.577	0.581	0.584	0.586	0.588	0.589	0.591	0.593	0.595
0.70	0.668	0.676	0.680	0.684	0.686	0.688	0.689	0.691	0.693	0.695
0.80	0.765	0.774	0.779	0.783	0.785	0.787	0.788	0.791	0.793	0.795
0.90	0.858	0.869	0.875	0.880	0.883	0.885	0.887	0.890	0.892	0.894
0.93	0.882	0.895	0.902	0.908	0.911	0.914	0.916	0.919	0.921	0.924
0.95	0.897	0.911	0.919	0.925	0.929	0.932	0.934	0.938	0.941	0.944
0.97	0.911	0.925	0.934	0.941	0.945	0.949	0.952	0.956	0.959	0.963
0.99	0.923	0.937	0.947	0.954	0.959	0.963	0.966	0.972	0.975	0.980
1.00	0.929	0.944	0.953	0.960	0.965	0.969	0.972	0.978	0.982	0.986

Table B7: Mean function of the LS estimator of α for Model 3.

α / T	40	50	60	70	80	90	100	125	150	200
-0.999	-0.988	-0.989	-0.990	-0.991	-0.992	-0.992	-0.992	-0.993	-0.994	-0.995
-0.80	-0.775	-0.779	-0.782	-0.785	-0.786	-0.788	-0.789	-0.791	-0.792	-0.794
-0.60	-0.593	-0.594	-0.595	-0.595	-0.596	-0.596	-0.597	-0.597	-0.598	-0.598
-0.40	-0.412	-0.409	-0.408	-0.406	-0.405	-0.405	-0.404	-0.403	-0.403	-0.402
-0.20	-0.231	-0.225	-0.221	-0.218	-0.215	-0.214	-0.212	-0.210	-0.208	-0.206
0.00	-0.051	-0.041	-0.034	-0.029	-0.025	-0.022	-0.020	-0.016	-0.013	-0.010
0.10	0.039	0.051	0.059	0.065	0.070	0.073	0.076	0.081	0.084	0.088
0.20	0.128	0.143	0.152	0.159	0.165	0.169	0.172	0.177	0.181	0.186
0.30	0.217	0.234	0.245	0.253	0.259	0.264	0.268	0.274	0.278	0.284
0.40	0.306	0.326	0.338	0.347	0.354	0.359	0.363	0.371	0.376	0.382
0.50	0.394	0.416	0.431	0.441	0.448	0.454	0.459	0.467	0.473	0.480
0.60	0.481	0.506	0.523	0.534	0.543	0.549	0.554	0.564	0.570	0.578
0.70	0.567	0.595	0.613	0.626	0.636	0.644	0.649	0.660	0.667	0.675
0.80	0.648	0.681	0.702	0.717	0.728	0.737	0.743	0.755	0.763	0.773
0.90	0.720	0.759	0.785	0.803	0.816	0.826	0.834	0.849	0.858	0.869
0.93	0.738	0.779	0.806	0.825	0.840	0.851	0.860	0.875	0.885	0.897
0.95	0.748	0.790	0.819	0.839	0.854	0.866	0.875	0.892	0.903	0.916
0.97	0.756	0.800	0.829	0.851	0.867	0.879	0.889	0.907	0.918	0.933
0.99	0.761	0.806	0.836	0.858	0.875	0.888	0.899	0.918	0.931	0.946
1.00	0.762	0.807	0.837	0.860	0.877	0.890	0.901	0.920	0.933	0.950

Table B8: Median function of the LS estimator of α for Model 3.

α / T	40	50	60	70	80	90	100	125	150	200
-0.999	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.997	-0.998
-0.80	-0.793	-0.794	-0.794	-0.795	-0.796	-0.796	-0.796	-0.797	-0.798	-0.798
-0.60	-0.607	-0.605	-0.604	-0.603	-0.603	-0.603	-0.602	-0.602	-0.601	-0.601
-0.40	-0.421	-0.417	-0.414	-0.412	-0.410	-0.409	-0.408	-0.407	-0.405	-0.404
-0.20	-0.237	-0.229	-0.224	-0.220	-0.218	-0.216	-0.214	-0.211	-0.209	-0.207
0.00	-0.052	-0.041	-0.034	-0.029	-0.026	-0.023	-0.020	-0.016	-0.013	-0.010
0.10	0.040	0.052	0.061	0.066	0.071	0.074	0.077	0.081	0.084	0.088
0.20	0.132	0.146	0.155	0.162	0.167	0.170	0.173	0.179	0.182	0.187
0.30	0.223	0.239	0.250	0.257	0.263	0.267	0.270	0.276	0.280	0.285
0.40	0.314	0.332	0.344	0.352	0.359	0.363	0.367	0.374	0.378	0.384
0.50	0.405	0.425	0.438	0.447	0.454	0.460	0.464	0.471	0.476	0.482
0.60	0.494	0.517	0.532	0.542	0.550	0.556	0.560	0.568	0.574	0.581
0.70	0.582	0.608	0.624	0.636	0.644	0.651	0.656	0.665	0.671	0.679
0.80	0.665	0.695	0.715	0.728	0.738	0.745	0.751	0.762	0.769	0.777
0.90	0.740	0.776	0.799	0.815	0.827	0.836	0.844	0.856	0.864	0.874
0.93	0.758	0.796	0.821	0.839	0.852	0.861	0.869	0.883	0.892	0.902
0.95	0.768	0.808	0.834	0.852	0.866	0.877	0.885	0.900	0.909	0.921
0.97	0.776	0.817	0.845	0.864	0.879	0.890	0.899	0.915	0.925	0.938
0.99	0.781	0.823	0.851	0.872	0.887	0.899	0.909	0.926	0.938	0.952
1.00	0.782	0.824	0.853	0.873	0.889	0.901	0.911	0.928	0.940	0.955

Table B9: Mode function of the LS estimator of α for Model 3.

α / T	40	50	60	70	80	90	100	125	150	200
-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999	-0.999
-0.80	-0.828	-0.823	-0.819	-0.817	-0.815	-0.813	-0.812	-0.809	-0.808	-0.806
-0.60	-0.635	-0.628	-0.623	-0.620	-0.618	-0.616	-0.614	-0.611	-0.609	-0.607
-0.40	-0.441	-0.433	-0.427	-0.423	-0.420	-0.418	-0.416	-0.413	-0.411	-0.408
-0.20	-0.248	-0.238	-0.231	-0.227	-0.223	-0.220	-0.218	-0.215	-0.212	-0.209
0.00	-0.054	-0.042	-0.035	-0.030	-0.026	-0.023	-0.021	-0.016	-0.014	-0.010
0.10	0.043	0.055	0.063	0.069	0.073	0.076	0.078	0.083	0.086	0.089
0.20	0.139	0.152	0.161	0.167	0.171	0.174	0.177	0.182	0.185	0.189
0.30	0.236	0.250	0.259	0.265	0.270	0.273	0.276	0.281	0.284	0.288
0.40	0.332	0.347	0.356	0.363	0.368	0.372	0.375	0.380	0.383	0.388
0.50	0.427	0.443	0.454	0.461	0.466	0.470	0.473	0.479	0.483	0.487
0.60	0.522	0.540	0.551	0.559	0.564	0.569	0.572	0.578	0.582	0.586
0.70	0.615	0.634	0.647	0.656	0.662	0.666	0.670	0.677	0.681	0.686
0.80	0.704	0.727	0.741	0.751	0.758	0.763	0.768	0.775	0.779	0.785
0.90	0.782	0.811	0.829	0.842	0.851	0.857	0.862	0.871	0.877	0.883
0.93	0.801	0.832	0.852	0.866	0.876	0.883	0.889	0.899	0.905	0.912
0.95	0.812	0.844	0.866	0.880	0.891	0.899	0.905	0.916	0.923	0.931
0.97	0.820	0.854	0.876	0.892	0.903	0.912	0.919	0.931	0.939	0.948
0.99	0.825	0.860	0.883	0.900	0.912	0.921	0.929	0.943	0.951	0.962
1.00	0.825	0.861	0.884	0.901	0.913	0.923	0.931	0.945	0.954	0.965

Table B10: CDF integration truncation bounds as given by the Imhof's restriction in (2.18) and the method derived in (2.22). $B_{\text{trc}} = 10^{-8}$ and $\alpha = x$.

		Imhof CDF bounds			Efficient CDF bounds		
T	α	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
40	-0.99	3.56	8.77	22.27	3.60	3.67	3.74
	-0.80	4.27	10.52	26.28	2.88	2.90	2.92
	-0.50	9.06	23.68	27.49	2.56	2.57	2.58
	0.00	4.08	9.90	24.94	2.39	2.40	2.41
	0.50	9.06	10.68	30.33	2.29	2.30	2.30
	0.80	4.27	10.50	26.15	2.22	2.22	2.22
	0.99	3.56	8.14	19.17	2.17	2.17	2.17
50	-0.99	3.15	6.53	13.00	3.14	3.18	3.21
	-0.80	3.55	7.35	15.70	3.13	3.17	3.22
	-0.50	3.49	7.53	14.82	3.01	3.07	3.13
	0.00	3.51	7.29	15.52	2.94	3.03	3.12
	0.50	3.49	7.44	15.27	3.01	3.15	3.30
	0.80	3.55	7.06	15.49	3.13	3.33	3.53
	0.99	3.15	6.10	13.17	3.14	3.44	3.55
100	-0.99	2.46	3.51	5.01	2.39	2.40	2.41
	-0.80	2.58	3.70	5.36	2.13	2.14	2.14
	-0.50	3.58	5.21	5.51	1.96	1.97	1.98
	0.00	2.61	3.73	5.41	1.89	1.91	1.93
	0.50	3.58	3.87	5.59	1.96	2.00	2.05
	0.80	2.58	3.68	5.61	2.13	2.20	2.27
	0.99	2.46	3.40	4.88	2.39	2.51	2.59
150	-0.99	2.29	2.88	3.68	2.19	2.19	2.20
	-0.80	2.35	3.00	3.78	1.81	1.81	1.81
	-0.50	2.37	3.00	3.84	1.65	1.66	1.66
	0.00	2.38	3.04	3.84	1.60	1.61	1.62
	0.50	2.37	3.03	3.81	1.65	1.68	1.70
	0.80	2.35	2.99	3.80	1.81	1.87	1.92
	0.99	2.29	2.88	3.68	2.19	2.27	2.32
200	-0.99	2.23	2.65	3.17	2.09	2.10	2.10
	-0.80	2.26	2.71	3.24	1.37	1.37	1.37
	-0.50	2.27	2.75	3.27	1.47	1.47	1.47
	0.00	2.27	2.72	3.27	1.47	1.47	1.48
	0.50	2.27	2.71	3.25	1.47	1.50	1.52
	0.80	2.26	2.72	3.26	1.37	1.48	1.60
	0.99	2.23	2.64	3.20	2.09	2.15	2.20

Table B11: PDF integration truncation bounds as given in (2.30). $B_{\text{trc}} = 10^{-8}$ and $\alpha = x$.

		Efficient PDF bounds		
T	α	Model 1	Model 2	Model 3
40	-0.99	5.84	6.05	6.27
	-0.80	5.70	5.92	6.17
	-0.50	5.56	5.81	6.09
	0.00	5.50	5.82	6.17
	0.50	5.56	6.00	6.47
	0.80	5.70	6.28	6.86
	0.99	5.84	6.25	6.39
50	-0.99	4.52	4.62	4.71
	-0.80	4.33	4.43	4.54
	-0.50	4.17	4.29	4.42
	0.00	4.10	4.26	4.43
	0.50	4.17	4.41	4.65
	0.80	4.33	4.65	4.97
	0.99	4.52	4.88	4.96
100	-0.99	2.86	2.87	2.88
	-0.80	2.54	2.56	2.57
	-0.50	2.35	2.37	2.38
	0.00	2.26	2.29	2.32
	0.50	2.35	2.40	2.46
	0.80	2.54	2.63	2.71
	0.99	2.86	3.00	3.08
150	-0.99	2.48	2.48	2.49
	-0.80	2.11	2.11	2.11
	-0.50	1.91	1.91	1.92
	0.00	1.83	1.84	1.86
	0.50	1.91	1.94	1.96
	0.80	2.11	2.15	2.20
	0.99	2.48	2.56	2.62
200	-0.99	2.31	2.31	2.31
	-0.80	1.88	1.88	1.89
	-0.50	1.69	1.70	1.70
	0.00	1.64	1.64	1.65
	0.50	1.69	1.71	1.73
	0.80	1.88	1.92	1.96
	0.99	2.31	2.37	2.41

Table B12: Simulation results: mean function of the LS estimator for given T ($N = 50000$). Comparative values (numerical integration) are given in Tables B1, B4 and B7. See Figure 3.2 for the approximate standard errors given N.

	Model 1				Model 2				Model 3			
$\alpha\backslash T$	50	100	150	200	50	100	150	200	50	100	150	200
-0.999	-0.989	-0.992	-0.994	-0.995	-0.989	-0.992	-0.994	-0.995	-0.989	-0.992	-0.994	-0.995
-0.80	-0.770	-0.785	-0.790	-0.792	-0.775	-0.787	-0.791	-0.793	-0.779	-0.789	-0.793	-0.794
-0.60	-0.578	-0.588	-0.593	-0.594	-0.586	-0.593	-0.595	-0.596	-0.595	-0.597	-0.598	-0.598
-0.40	-0.384	-0.392	-0.395	-0.396	-0.397	-0.398	-0.399	-0.399	-0.409	-0.404	-0.403	-0.402
-0.20	-0.192	-0.196	-0.198	-0.198	-0.208	-0.204	-0.203	-0.202	-0.225	-0.212	-0.208	-0.206
0.00	0.001	-0.001	0.000	0.000	-0.020	-0.011	-0.007	-0.005	-0.040	-0.021	-0.014	-0.010
0.10	0.097	0.098	0.099	0.099	0.074	0.087	0.091	0.093	0.052	0.076	0.084	0.088
0.20	0.191	0.196	0.197	0.197	0.167	0.184	0.190	0.191	0.142	0.172	0.181	0.185
0.30	0.289	0.294	0.295	0.297	0.262	0.280	0.287	0.291	0.235	0.267	0.278	0.284
0.40	0.386	0.393	0.395	0.396	0.356	0.378	0.386	0.389	0.326	0.364	0.376	0.382
0.50	0.482	0.490	0.494	0.495	0.450	0.475	0.484	0.487	0.417	0.459	0.473	0.480
0.60	0.577	0.588	0.592	0.594	0.542	0.571	0.581	0.586	0.506	0.554	0.570	0.577
0.70	0.673	0.687	0.691	0.693	0.635	0.669	0.679	0.684	0.594	0.650	0.667	0.675
0.80	0.770	0.785	0.790	0.792	0.727	0.765	0.777	0.783	0.680	0.744	0.763	0.773
0.90	0.868	0.883	0.889	0.891	0.817	0.860	0.874	0.880	0.759	0.834	0.858	0.869
0.93	0.898	0.912	0.918	0.921	0.843	0.888	0.903	0.910	0.778	0.859	0.885	0.897
0.95	0.919	0.933	0.938	0.941	0.860	0.906	0.921	0.929	0.791	0.875	0.903	0.915
0.97	0.941	0.954	0.958	0.961	0.875	0.924	0.940	0.948	0.799	0.889	0.919	0.933
0.99	0.967	0.976	0.980	0.982	0.890	0.940	0.957	0.966	0.805	0.899	0.930	0.946
1.00	-	-	-	-	0.896	0.948	0.965	0.973	0.806	0.901	0.933	0.950

Table B13: Simulation results: median function of the LS estimator for given T ($N = 50000$). Comparative values (numerical integration) are given in Tables B2, B5 and B8.

	Model 1				Model 2				Model 3			
$\alpha\backslash T$	50	100	150	200	50	100	150	200	50	100	150	200
-0.999	-0.997	-0.997	-0.997	-0.998	-0.997	-0.997	-0.997	-0.998	-0.997	-0.997	-0.997	-0.998
-0.80	-0.785	-0.792	-0.795	-0.796	-0.789	-0.794	-0.796	-0.797	-0.793	-0.797	-0.798	-0.798
-0.60	-0.589	-0.595	-0.597	-0.597	-0.597	-0.598	-0.600	-0.599	-0.606	-0.602	-0.603	-0.601
-0.40	-0.391	-0.396	-0.398	-0.398	-0.404	-0.402	-0.402	-0.401	-0.416	-0.408	-0.406	-0.404
-0.20	-0.195	-0.197	-0.199	-0.198	-0.212	-0.206	-0.204	-0.203	-0.228	-0.213	-0.209	-0.207
0.00	0.001	0.000	0.000	0.000	-0.020	-0.010	-0.007	-0.005	-0.040	-0.020	-0.013	-0.010
0.10	0.099	0.100	0.100	0.099	0.076	0.088	0.092	0.094	0.052	0.077	0.084	0.088
0.20	0.195	0.199	0.199	0.198	0.171	0.186	0.191	0.192	0.145	0.174	0.183	0.186
0.30	0.294	0.297	0.297	0.299	0.268	0.283	0.288	0.292	0.240	0.270	0.279	0.286
0.40	0.394	0.397	0.398	0.398	0.364	0.382	0.388	0.391	0.334	0.368	0.379	0.384
0.50	0.491	0.495	0.497	0.497	0.459	0.479	0.487	0.490	0.425	0.464	0.477	0.482
0.60	0.588	0.594	0.596	0.597	0.554	0.577	0.585	0.588	0.517	0.559	0.574	0.580
0.70	0.685	0.694	0.696	0.696	0.648	0.675	0.684	0.687	0.607	0.657	0.672	0.678
0.80	0.785	0.792	0.795	0.796	0.742	0.773	0.782	0.787	0.695	0.752	0.769	0.777
0.90	0.884	0.892	0.894	0.895	0.834	0.869	0.880	0.885	0.775	0.843	0.864	0.874
0.93	0.913	0.921	0.924	0.925	0.860	0.897	0.909	0.914	0.794	0.869	0.892	0.902
0.95	0.935	0.942	0.944	0.945	0.878	0.915	0.928	0.934	0.807	0.885	0.909	0.921
0.97	0.956	0.962	0.964	0.966	0.893	0.933	0.947	0.953	0.817	0.899	0.925	0.938
0.99	0.980	0.984	0.985	0.986	0.907	0.950	0.964	0.971	0.823	0.909	0.937	0.952
1.00	-	-	-	-	0.913	0.957	0.971	0.978	0.823	0.911	0.940	0.955

Table B14: Simulation results: 90% confidence intervals for the LS estimator of α ($T = 60$, $N = 40000$).

	Model 1		Model 2		Model 3	
α	0.05	0.95	0.05	0.95	0.05	0.95
-0.999	-1.010	-0.943	-1.010	-0.944	-1.011	-0.945
-0.80	-0.890	-0.620	-0.892	-0.626	-0.894	-0.632
-0.60	-0.738	-0.392	-0.743	-0.402	-0.748	-0.410
-0.40	-0.571	-0.182	-0.580	-0.194	-0.588	-0.206
-0.20	-0.395	0.020	-0.406	0.007	-0.419	-0.008
0.00	-0.211	0.211	-0.228	0.195	-0.244	0.178
0.10	-0.117	0.306	-0.136	0.288	-0.153	0.270
0.20	-0.023	0.394	-0.043	0.376	-0.065	0.358
0.30	0.079	0.484	0.054	0.465	0.030	0.445
0.40	0.181	0.571	0.155	0.551	0.127	0.530
0.50	0.284	0.656	0.254	0.636	0.222	0.615
0.60	0.391	0.739	0.354	0.718	0.319	0.695
0.70	0.503	0.815	0.462	0.795	0.416	0.773
0.80	0.617	0.890	0.565	0.871	0.514	0.846
0.90	0.742	0.957	0.674	0.938	0.608	0.912
0.93	0.784	0.975	0.707	0.957	0.631	0.929
0.95	0.813	0.987	0.730	0.970	0.647	0.940
0.97	0.847	0.998	0.748	0.981	0.659	0.949
0.99	0.891	1.009	0.768	0.994	0.665	0.955
0.995	0.910	1.011	0.771	0.997	0.666	0.956
0.999	0.943	1.010	0.777	0.998	0.666	0.956
1	-	-	0.776	0.999	0.665	0.957

Table B15: Exact values for the decomposition of the MSE of the LS estimator in different models. The values for the 'variance', 'squared bias' and 'MSE' have been multiplied by a factor of 10^3 for reporting purposes. See also Figure 4.2(a).

	Variance			Squared bias			MSE			RMS		
α/ Models	1	2	3	1	2	3	1	2	3	1	2	3
-0.999	0.872	0.847	0.823	0.101	0.096	0.092	0.973	0.943	0.915	0.031	0.031	0.030
-0.90	5.865	5.671	5.485	1.017	0.856	0.713	6.882	6.527	6.198	0.083	0.081	0.079
-0.80	8.832	8.547	8.273	0.880	0.629	0.424	9.712	9.176	8.698	0.099	0.096	0.093
-0.70	11.388	11.041	10.708	0.695	0.394	0.181	12.084	11.436	10.889	0.110	0.107	0.104
-0.60	13.587	13.206	12.838	0.519	0.205	0.036	14.106	13.411	12.874	0.119	0.116	0.113
-0.50	15.441	15.052	14.675	0.364	0.075	0.003	15.805	15.128	14.678	0.126	0.123	0.121
-0.40	16.955	16.586	16.224	0.234	0.009	0.086	17.189	16.594	16.310	0.131	0.129	0.128
-0.30	18.131	17.808	17.488	0.132	0.008	0.291	18.263	17.816	17.778	0.135	0.133	0.133
-0.20	18.971	18.720	18.467	0.059	0.075	0.619	19.029	18.795	19.087	0.138	0.137	0.138
-0.10	19.474	19.324	19.165	0.015	0.211	1.076	19.489	19.535	20.241	0.140	0.140	0.142
0.00	19.642	19.620	19.582	0.000	0.416	1.666	19.642	20.037	21.248	0.140	0.142	0.146
0.10	19.474	19.609	19.720	0.015	0.694	2.395	19.489	20.303	22.115	0.140	0.142	0.149
0.20	18.971	19.291	19.582	0.059	1.046	3.274	19.029	20.337	22.856	0.138	0.143	0.151
0.30	18.131	18.669	19.172	0.132	1.475	4.317	18.263	20.143	23.489	0.135	0.142	0.153
0.40	16.955	17.743	18.496	0.234	1.986	5.548	17.189	19.729	24.044	0.131	0.140	0.155
0.50	15.441	16.518	17.562	0.364	2.588	7.010	15.805	19.106	24.573	0.126	0.138	0.157
0.60	13.587	14.998	16.389	0.519	3.298	8.784	14.106	18.296	25.174	0.119	0.135	0.159
0.70	11.388	13.195	15.010	0.695	4.151	11.045	12.084	17.346	26.054	0.110	0.132	0.161
0.80	8.832	11.136	13.498	0.880	5.236	14.242	9.712	16.371	27.740	0.099	0.128	0.167
0.90	5.865	8.910	12.060	1.017	6.853	19.908	6.882	15.763	31.968	0.083	0.126	0.179
0.99	2.175	7.144	11.372	0.529	10.030	33.946	2.704	17.174	45.317	0.052	0.131	0.213
0.999	0.872	7.028	11.366	0.101	10.625	36.989	0.973	17.653	48.355	0.031	0.133	0.220
1	-	7.016	11.366	-	10.698	37.370	-	17.714	48.736	-	0.133	0.221

Table B16: Exact values for the decomposition of the MSE of the estimator $\hat{\alpha}_{\text{UE}}^{\text{A}}$ in different models. The values for the 'variance', 'squared bias' and 'MSE' have been multiplied by a factor of 10^3 for reporting purposes.

	Variance			Squared bias			MSE			RMS		
α/Model	1	2	3	1	2	3	1	2	3	1	2	3
-0.999	0.456	0.460	0.464	0.008	0.008	0.008	0.464	0.468	0.473	0.022	0.022	0.022
-0.90	5.850	5.885	5.924	0.004	0.004	0.004	5.855	5.889	5.928	0.077	0.077	0.077
-0.80	9.355	9.425	9.503	0.000	0.000	0.000	9.355	9.425	9.503	0.097	0.097	0.097
-0.70	12.212	12.331	12.461	0.000	0.000	0.000	12.212	12.331	12.461	0.111	0.111	0.112
-0.60	14.632	14.815	15.013	0.000	0.000	0.000	14.632	14.815	15.013	0.121	0.122	0.123
-0.50	16.661	16.924	17.208	0.000	0.000	0.000	16.661	16.924	17.208	0.129	0.130	0.131
-0.40	18.313	18.673	19.061	0.000	0.000	0.000	18.313	18.673	19.061	0.135	0.137	0.138
-0.30	19.595	20.070	20.581	0.000	0.000	0.000	19.595	20.070	20.581	0.140	0.142	0.143
-0.20	20.510	21.116	21.772	0.000	0.000	0.000	20.510	21.116	21.772	0.143	0.145	0.148
-0.10	21.058	21.815	22.638	0.000	0.000	0.000	21.058	21.815	22.638	0.145	0.148	0.150
0.00	21.240	22.169	23.187	0.000	0.000	0.000	21.240	22.169	23.187	0.146	0.149	0.152
0.10	21.058	22.180	23.423	0.000	0.000	0.000	21.058	22.180	23.423	0.145	0.149	0.153
0.20	20.510	21.851	23.358	0.000	0.000	0.000	20.510	21.851	23.358	0.143	0.148	0.153
0.30	19.595	21.187	23.008	0.000	0.000	0.000	19.595	21.187	23.008	0.140	0.146	0.152
0.40	18.313	20.193	22.403	0.000	0.000	0.000	18.313	20.193	22.404	0.135	0.142	0.150
0.50	16.661	18.882	21.608	0.000	0.000	0.001	16.661	18.882	21.609	0.129	0.137	0.147
0.60	14.632	17.279	20.786	0.000	0.000	0.004	14.632	17.279	20.789	0.121	0.131	0.144
0.70	12.212	15.442	20.242	0.000	0.001	0.015	12.212	15.443	20.258	0.111	0.124	0.142
0.80	9.355	13.401	19.186	0.000	0.002	0.033	9.355	13.404	19.220	0.097	0.116	0.139
0.90	5.850	10.037	14.889	0.004	0.007	0.067	5.854	10.044	14.955	0.077	0.100	0.122
0.99	1.396	5.417	10.876	0.041	1.091	3.412	1.437	6.509	14.288	0.038	0.081	0.120
0.999	0.456	5.057	10.799	0.009	1.482	4.457	0.465	6.539	15.256	0.022	0.081	0.124
1	-	5.020	10.798	-	1.531	4.594	-	6.552	15.392	-	0.081	0.124

Table B17: Exact values for the decomposition of the MSE of the estimator $\hat{\alpha}_{\text{UE}}^{\text{B}}$ in different models. The values for the 'variance', 'squared bias' and 'MSE' have been multiplied by a factor 10^3 for reporting purposes.

	Variance			Squared bias			MSE			RMS		
α/Model	1	2	3	1	2	3	1	2	3	1	2	3
-0.999	0.602	0.607	0.612	0.042	0.043	0.043	0.644	0.650	0.656	0.025	0.025	0.026
-0.90	5.882	5.924	5.970	0.286	0.287	0.289	6.168	6.211	6.260	0.079	0.079	0.079
-0.80	9.092	9.168	9.252	0.229	0.231	0.234	9.321	9.399	9.486	0.097	0.097	0.097
-0.70	11.792	11.916	12.050	0.178	0.181	0.185	11.970	12.097	12.234	0.109	0.110	0.111
-0.60	14.101	14.286	14.486	0.132	0.135	0.139	14.232	14.421	14.625	0.119	0.120	0.121
-0.50	16.042	16.305	16.587	0.092	0.096	0.100	16.134	16.400	16.687	0.127	0.128	0.129
-0.40	17.625	17.981	18.363	0.059	0.062	0.066	17.684	18.043	18.430	0.133	0.134	0.136
-0.30	18.855	19.319	19.819	0.033	0.036	0.040	18.888	19.355	19.859	0.137	0.139	0.141
-0.20	19.732	20.321	20.959	0.015	0.017	0.020	19.747	20.338	20.978	0.141	0.143	0.145
-0.10	20.258	20.990	21.786	0.004	0.005	0.007	20.262	20.995	21.793	0.142	0.145	0.148
0.00	20.433	21.326	22.305	0.000	0.000	0.000	20.433	21.326	22.306	0.143	0.146	0.149
0.10	20.258	21.332	22.522	0.004	0.003	0.001	20.262	21.335	22.523	0.142	0.146	0.150
0.20	19.732	21.010	22.445	0.015	0.012	0.009	19.747	21.023	22.455	0.141	0.145	0.150
0.30	18.855	20.364	22.090	0.033	0.030	0.025	18.888	20.394	22.114	0.137	0.143	0.149
0.40	17.625	19.399	21.479	0.059	0.054	0.046	17.684	19.453	21.525	0.133	0.139	0.147
0.50	16.042	18.125	20.664	0.092	0.086	0.074	16.134	18.211	20.738	0.127	0.135	0.144
0.60	14.101	16.562	19.770	0.132	0.125	0.102	14.232	16.688	19.872	0.119	0.129	0.141
0.70	11.792	14.761	19.103	0.178	0.169	0.119	11.970	14.930	19.222	0.109	0.122	0.139
0.80	9.092	12.820	18.477	0.229	0.211	0.125	9.321	13.031	18.602	0.097	0.114	0.136
0.90	5.882	10.195	15.407	0.286	0.334	0.531	6.168	10.530	15.938	0.079	0.103	0.126
0.99	1.696	6.140	11.846	0.229	1.837	4.882	1.925	7.977	16.727	0.044	0.089	0.129
0.999	0.602	5.781	11.773	0.043	2.274	6.106	0.645	8.055	17.878	0.025	0.090	0.134
1	-	5.744	11.771	-	2.329	6.266	-	8.073	18.037	-	0.090	0.134

Table B18: Exact value for the decomposition of the MSE of the estimator $\hat{\alpha}_{\text{UE}}^{\text{C}}$ in different models. The values for the 'variance', 'squared bias' and 'MSE' have been multiplied by a factor 10^3 for reporting purposes.

	Variance			Squared bias			MSE			RMS		
α/Model	1	2	3	1	2	3	1	2	3	1	2	3
-0.999	1.013	1.021	1.030	0.221	0.219	0.218	1.234	1.240	1.248	0.035	0.035	0.035
-0.90	5.898	5.953	6.013	2.140	2.150	2.161	8.038	8.102	8.174	0.090	0.090	0.090
-0.80	8.609	8.695	8.788	1.924	1.945	1.967	10.533	10.639	10.755	0.103	0.103	0.104
-0.70	11.013	11.144	11.285	1.539	1.567	1.596	12.552	12.711	12.881	0.112	0.113	0.113
-0.60	13.097	13.287	13.491	1.156	1.187	1.219	14.253	14.473	14.710	0.119	0.120	0.121
-0.50	14.860	15.122	15.402	0.813	0.844	0.878	15.673	15.966	16.280	0.125	0.126	0.128
-0.40	16.303	16.649	17.021	0.525	0.554	0.585	16.827	17.203	17.606	0.130	0.131	0.133
-0.30	17.425	17.869	18.348	0.297	0.321	0.348	17.721	18.190	18.696	0.133	0.135	0.137
-0.20	18.226	18.782	19.384	0.132	0.150	0.171	18.358	18.932	19.555	0.135	0.138	0.140
-0.10	18.707	19.389	20.132	0.033	0.043	0.055	18.740	19.432	20.187	0.137	0.139	0.142
0.00	18.867	19.691	20.594	0.000	0.001	0.003	18.867	19.691	20.597	0.137	0.140	0.144
0.10	18.707	19.688	20.775	0.033	0.024	0.016	18.740	19.712	20.791	0.137	0.140	0.144
0.20	18.226	19.382	20.680	0.132	0.115	0.096	18.358	19.497	20.776	0.135	0.140	0.144
0.30	17.425	18.776	20.320	0.297	0.273	0.245	17.721	19.050	20.565	0.133	0.138	0.143
0.40	16.303	17.874	19.710	0.525	0.501	0.467	16.827	18.375	20.177	0.130	0.136	0.142
0.50	14.860	16.684	18.883	0.813	0.799	0.764	15.673	17.483	19.647	0.125	0.132	0.140
0.60	13.097	15.220	17.908	1.156	1.169	1.141	14.253	16.389	19.049	0.119	0.128	0.138
0.70	11.013	13.517	16.980	1.539	1.616	1.599	12.552	15.132	18.579	0.112	0.123	0.136
0.80	8.609	11.689	16.484	1.924	2.131	2.117	10.533	13.820	18.600	0.103	0.118	0.136
0.90	5.898	9.938	15.483	2.140	2.702	3.292	8.038	12.641	18.776	0.090	0.112	0.137
0.99	2.445	7.445	13.357	1.020	4.562	9.687	3.465	12.007	23.044	0.059	0.110	0.152
0.999	1.013	7.158	13.305	0.221	5.050	11.353	1.234	12.207	24.658	0.035	0.110	0.157
1	-	7.127	13.304	-	5.110	11.571	-	12.238	24.874	-	0.111	0.158

Table B19: Exact bias of $\hat{\alpha}_{\text{UE}}^{\text{A}}$, $\hat{\alpha}_{\text{UE}}^{\text{B}}$, $\hat{\alpha}_{\text{UE}}^{\text{C}}$ and $\hat{\alpha}_{\text{LS}}$ in different models. The reported values have been multiplied by 10^3.

	$\hat{\alpha}_{\text{UE}}^{\text{A}}$			$\hat{\alpha}_{\text{UE}}^{\text{B}}$			$\hat{\alpha}_{\text{UE}}^{\text{C}}$			$\hat{\alpha}_{\text{LS}}$		
α/Model	1	2	3	1	2	3	1	2	3	1	2	3
-0.999	3.83	3.86	3.90	7.42	7.47	7.52	15.65	15.69	15.74	10.03	9.81	9.59
-0.90	1.89	1.91	1.94	16.81	16.86	16.92	46.23	46.34	46.46	31.89	29.26	26.69
-0.80	0.48	0.49	0.51	15.12	15.21	15.30	43.86	44.10	44.35	29.66	25.08	20.59
-0.70	0.19	0.20	0.21	13.33	13.45	13.58	39.24	39.59	39.95	26.37	19.86	13.46
-0.60	0.10	0.11	0.13	11.47	11.63	11.80	34.00	34.45	34.92	22.79	14.33	6.01
-0.50	0.06	0.07	0.09	9.58	9.78	9.98	28.52	29.06	29.63	19.08	8.68	-1.59
-0.40	0.03	0.05	0.08	7.68	7.90	8.14	22.91	23.53	24.19	15.31	2.95	-9.28
-0.30	0.02	0.04	0.09	5.77	6.02	6.29	17.22	17.92	18.66	11.50	-2.84	-17.05
-0.20	0.01	0.04	0.10	3.85	4.12	4.43	11.50	12.26	13.06	7.68	-8.66	-24.89
-0.10	0.01	0.05	0.12	1.92	2.22	2.56	5.76	6.56	7.42	3.84	-14.51	-32.81
0.00	0.00	0.05	0.15	0.00	0.31	0.68	0.00	0.82	1.72	0.00	-20.41	-40.82
0.10	-0.01	0.07	0.21	-1.92	-1.60	-1.20	-5.76	-4.93	-4.02	-3.84	-26.35	-48.94
0.20	-0.01	0.09	0.28	-3.85	-3.52	-3.08	-11.50	-10.72	-9.81	-7.68	-32.34	-57.22
0.30	-0.02	0.12	0.41	-5.77	-5.44	-4.95	-17.22	-16.53	-15.67	-11.50	-38.40	-65.70
0.40	-0.03	0.17	0.62	-7.68	-7.36	-6.81	-22.91	-22.38	-21.60	-15.31	-44.56	-74.49
0.50	-0.06	0.26	1.01	-9.58	-9.29	-8.59	-28.52	-28.26	-27.64	-19.08	-50.87	-83.73
0.60	-0.10	0.43	1.86	-11.47	-11.19	-10.14	-34.00	-34.20	-33.80	-22.79	-57.43	-93.72
0.70	-0.19	0.82	3.92	-13.33	-13.01	-10.90	-39.24	-40.20	-39.99	-26.37	-64.43	-105.09
0.80	-0.48	1.56	5.82	-15.12	-14.53	-11.16	-43.86	-46.16	-45.99	-29.66	-72.36	-119.34
0.90	-1.89	-2.56	-8.11	-16.81	-18.28	-22.99	-46.23	-51.99	-57.34	-31.89	-82.78	-141.10
0.95	-4.37	-14.50	-29.96	-17.62	-27.29	-42.64	-43.81	-57.36	-73.48	-31.16	-90.59	-159.55
0.99	-7.34	-33.03	-58.34	-15.71	-42.85	-69.79	-32.03	-67.55	-98.34	-22.99	-100.15	-184.24
0.999	-3.83	-38.48	-66.69	-7.42	-47.68	-78.07	-15.65	-71.07	-106.48	-10.03	-103.08	-192.32
1	-	-39.12	-67.69	-	-48.25	-79.07	-	-71.49	-107.48	-	-103.43	-193.31

Table B20: The 0.05 and 0.95 quantiles of the mean-, median- and mode-based estimators of α for Model 2 when $T = 50$. Similar quantiles for the LS estimator are shown in bold.

Estimator	$\hat{\alpha}^{A}_{UE}$		$\hat{\alpha}^{B}_{UE}$		$\hat{\alpha}^{C}_{UE}$		$\hat{\alpha}_{LS}$	
α/Quantile	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95
-0.999	-1.000	-0.967	-1.000	-0.952	-1.000	-0.922	**-1.012**	**-0.938**
-0.80	-0.928	-0.617	-0.912	-0.605	-0.881	-0.582	**-0.898**	**-0.602**
-0.60	-0.778	-0.381	-0.763	-0.374	-0.735	-0.359	**-0.754**	**-0.379**
-0.40	-0.610	-0.162	-0.599	-0.159	-0.575	-0.152	**-0.595**	**-0.173**
-0.20	-0.432	0.046	-0.423	0.046	-0.407	0.044	**-0.427**	**0.023**
0.00	-0.245	0.246	-0.240	0.241	-0.230	0.232	**-0.251**	**0.211**
0.10	-0.148	0.343	-0.145	0.336	-0.139	0.323	**-0.160**	**0.302**
0.20	-0.049	0.437	-0.048	0.429	-0.046	0.413	**-0.067**	**0.390**
0.30	0.051	0.530	0.051	0.520	0.049	0.500	**0.028**	**0.477**
0.40	0.154	0.621	0.151	0.609	0.146	0.585	**0.124**	**0.562**
0.50	0.259	0.709	0.254	0.695	0.245	0.668	**0.223**	**0.644**
0.60	0.366	0.796	0.359	0.780	0.346	0.749	**0.324**	**0.724**
0.70	0.476	0.880	0.467	0.862	0.449	0.827	**0.427**	**0.800**
0.80	0.589	0.966	0.577	0.944	0.555	0.904	**0.532**	**0.872**
0.85	0.646	1.000	0.633	0.989	0.608	0.944	**0.585**	**0.906**
0.90	0.703	1.000	0.689	1.000	0.662	0.992	**0.638**	**0.938**
0.93	0.737	1.000	0.722	1.000	0.694	1.000	**0.670**	**0.957**
0.95	0.759	1.000	0.744	1.000	0.714	1.000	**0.690**	**0.970**
0.97	0.780	1.000	0.764	1.000	0.734	1.000	**0.709**	**0.982**
0.99	0.799	1.000	0.783	1.000	0.752	1.000	**0.727**	**0.994**
1.00	0.808	1.000	0.792	1.000	0.760	1.000	**0.735**	**0.999**

Table B21: The 0.05 and 0.95 quantiles of the mean-, median- and mode-based estimators of α for Model 3 when $T = 50$. Similar quantiles for the LS estimator are shown in bold.

Estimator	$\hat{\alpha}^{A}_{UE}$		$\hat{\alpha}^{B}_{UE}$		$\hat{\alpha}^{C}_{UE}$		$\hat{\alpha}_{LS}$	
α/Quantile	0.05	0.95	0.05	0.95	0.05	0.95	0.05	0.95
-0.999	-1.000	-0.967	-1.000	-0.952	-1.000	-0.922	**-1.012**	**-0.939**
-0.80	-0.929	-0.616	-0.913	-0.604	-0.881	-0.581	**-0.901**	**-0.609**
-0.60	-0.779	-0.379	-0.764	-0.372	-0.735	-0.356	**-0.760**	**-0.390**
-0.40	-0.612	-0.159	-0.600	-0.156	-0.577	-0.148	**-0.605**	**-0.187**
-0.20	-0.435	0.051	-0.426	0.050	-0.409	0.049	**-0.441**	**0.006**
0.00	-0.249	0.252	-0.244	0.248	-0.234	0.239	**-0.270**	**0.191**
0.10	-0.154	0.350	-0.150	0.344	-0.143	0.331	**-0.182**	**0.280**
0.20	-0.057	0.447	-0.055	0.438	-0.052	0.422	**-0.093**	**0.368**
0.30	0.042	0.541	0.042	0.531	0.042	0.510	**-0.002**	**0.454**
0.40	0.143	0.635	0.141	0.622	0.136	0.598	**0.090**	**0.537**
0.50	0.245	0.727	0.241	0.712	0.232	0.683	**0.184**	**0.619**
0.60	0.349	0.820	0.342	0.802	0.330	0.768	**0.279**	**0.697**
0.70	0.453	0.920	0.445	0.896	0.428	0.853	**0.374**	**0.773**
0.80	0.557	1.000	0.546	1.000	0.525	0.948	**0.467**	**0.843**
0.85	0.607	1.000	0.595	1.000	0.571	1.000	**0.512**	**0.876**
0.90	0.653	1.000	0.640	1.000	0.615	1.000	**0.554**	**0.907**
0.93	0.678	1.000	0.665	1.000	0.638	1.000	**0.576**	**0.923**
0.95	0.693	1.000	0.679	1.000	0.651	1.000	**0.589**	**0.933**
0.97	0.704	1.000	0.690	1.000	0.662	1.000	**0.599**	**0.941**
0.99	0.712	1.000	0.697	1.000	0.669	1.000	**0.605**	**0.947**
1.00	0.713	1.000	0.698	1.000	0.670	1.000	**0.606**	**0.948**

Appendix C : *MATLAB* programs

Algorithm 1: LS estimator of the LDV coefficient α

1. For a given true value of α and T, generate the series $y_t = \alpha y_{t-1} + u_t$, $t = 2, \ldots, T$ for each model with $u_t \sim \mathcal{N}(0, 1)$. That is, given invariance discussed by Andrews (1993), it suffices to consider $\mu = \delta = 0$ and $\sigma^2 = 1$. The initial condition on the series, y_1, is stationary if $\alpha \in (-1, 1)$ and fixed ($y_1 = 0$) otherwise.

2. Compute the LS estimate of α for Model 1, $[\,\mu\,\alpha\,]'$ for Model 2 and $[\,\mu\,\delta\,\alpha\,]'$ for Model 3.

3. Repeat steps 1 and 2, N times.

Algorithm 1 gives the *MATLAB* code for Steps 1 to 3.


```
function [Z1 Z2 Z3] = LsEst(alpha,T,N);
%.....................................
for j = 1:N
    %....................
    %Generation of series
    if abs(alpha) == 1;
        y(1) = 0;
    else
        y(1) = randn/(1-alpha*alpha);
    end
    for i=2:T
        y(i) = alpha * y(i-1) + randn;
    end
    y = y';
    %....................
    Y = y(2:T);
    Q = ones(T-1,1);
    R = [1:1:T-1]';
    S = y(1:T-1);
    %....................
    X1 = S;         %Model 1
    Z1(j) = X1\Y;
    X2 = [Q S];     %Model 2
    U = X2\Y;
    Z2(j) = U(2);
    X3 = [Q R S];   %Model 3
    V = X3\Y;
```

```
    Z3(j) = V(3);
    %.....................
    A = []; B = []; U = []; V = []; X1 = []; X2 = []; X3 = []; y = [];
end
```

The outputs `Z1`, `Z2` and `Z3` are $N \times 1$ vectors of the LS estimator of α for Models 1, 2 and 3 respectively. The inputs variables `alpha` and `T` denote the true value of the coefficient being estimated and the sample size.

Figure 1.1 is obtained from data results obtained as follows:

- Repeat steps 1 to 3 for fixed $\alpha = 0.80$ and varying $T \in [20, 200]$ at intervals of size 10.
- Determine the mean of the N LS estimates $\hat{\alpha}$ of α in step 3.
- Obtain the usual bias of the LS estimator as $\mathbb{E}[\hat{\alpha}] - \alpha$ i.e. the difference between mean obtained above and $\alpha = 0.8$.

Figure 1.2 is obtained in a similar manner: for a constant sample size $T = 40$ while varying $\alpha \in [-1.2, 1.2]$. Figure 1.3 can be obtained by ascertaining the proportion of estimates in `Z1`, `Z2` and `Z3` less than the true coefficient at each iteration in α. Figure 1.4 summarizes the variance of the vector of LS estimates given α and T.

Numerical evaluation of the location functions

The set of *MATLAB* functions provided below are used to determine the location function using the numerical integration method.

For a given $\alpha \in \Omega$, T and model, the median of the $F_{\text{LS}}(x)$ in (2.13) can be established from (2.31) using Algorithm 2 as

$$\tilde{\alpha}_j = \texttt{fzero(@}(x)\textit{CDFsimpson}(T, \alpha, x, p, j), \alpha\texttt{)} \qquad (j = 1, 2, 3),$$

with $p = 0.5$. The function *LSmean* (Algorithm 12) is used to obtain $\bar{\alpha}_j$ in (2.32). The mode statistic for Model j $(j = 1, 2, 3)$ can be obtained using either of the following arguments to the function *PDFmin* in Algorithm 13

$$\ddot{\alpha}_j = \begin{cases} \texttt{fminbnd(@}(x)\textit{PDFmin}(T, \alpha, x, j), \texttt{L,U,opt)} \\ \texttt{fminsearch(@}(x)\textit{PDFmin}(T, \alpha, x, j), \alpha, \texttt{opt)} \end{cases} \quad \text{or}$$

where `opt` are the default optimization parameters of `fzero`.

The following correspondence is made between the notation used here-above and that used within the code. Outputs: `mdn` $\equiv \tilde{\alpha}_j$, `mn` $\equiv \bar{\alpha}_j$ and `mode` $\equiv \ddot{\alpha}_j$. Inputs: `T` $\equiv T$, `alpha` $\equiv \alpha$, `x` $\equiv x$, `p` $= p$, `model` $= j$ $(j = , 1, 2, 3)$. `L` and `U` are explained in section 2.4.3. Other explanatory statements are given under each subheading and comments (in `%`) are made within the code for clarification.

Algorithm 2: Numerical evaluation of the CDF of the LS estimator

The function determines the quantiles of the CDF in (2.13) using `quad` at a pre-specified tolerance say `tol`$= 10^{-8}$. The output `F` is given by $F_{\text{LS}}(x) - p$ i.e. using (2.31). It calls other subroutines `CDFquadform`, `CDFtruncate` and `CDFintegrand` in Algorithms (5), (4) and (3) below. `E` is a $T \times 1$ vector of eigenvalues.

```
function F = CDFsimpson(T,alpha,x,p,model);
%.....................................
tol = 1.e-8; %tolerance parameter

E = CDFquadform(T,alpha,x,model);
E = real(E);
a = 0;                                  %Lower limit
b = CDFtruncate(E);                     %Most efficient upper bound

y = quad(@CDFintegrand,a,b,tol,[],E);
F = 0.5 - y - p;
```


Algorithm 3: Integrand evaluation: CDF

The function `CDFintegrand` accepts vector arguments `E` [the eigenvalues of $\mathbf{W}_{\alpha jx}$ in (2.8)] and `P`. The latter is an adaptive-stepsize vector of points determined by the quadrature algorithms. For example, `P` is a 5×1 vector when utilizing `quad`. `y` is a vector of the integrand evaluated at each element of `P`.

```
function y = CDFintegrand(P,E);
%.....................................
n = length(P);
for j = 1:n;
    u = P(j);
    if u == 0;
        x(j) = 0;
        g = 0.5 * sum(E);
        y(j) = g;
    else
        theta = 0.5 * sum( atan(E*u) );
        rho =  prod((1 + (E.^2).* u^2).^(0.25));
        A = sin(theta);
        B = u*rho;
        x(j) = u;
        y(j) = (1/pi) * (A/B);
    end
end
```

Algorithm 4: Most efficient truncation bounds: CDF computation

The function `CDFtruncate` determines the most efficient truncation bound when evaluating the integral at each iterative step in x using the formula (2.22).

```
function b = CDFtruncate(E);
%.....................................
T = length(E);
te = 1.e-15;         %Truncation error
A = abs(E);
B = sort(A);         %Order statistics of 'E'
%.......................
%Constant terms
x1 = pi^2;
x2 = te^2;
%.......................
%Restrictions
%(1) E << 1 (product term)
for m = 1:T-1;
    M = B(m:T);
    x3 = 1/(T-m);
    x4 = (T-m)^2;
    x5 = prod(M);                %product from 'j=m+1' to 'T'
    %....................
    %MAXIMIZATION problem
    p = ((x1*x4*x2*x5)/4)^x3;
    q(m) = p;
    r(m) = m;
    M = []; x3 = []; x4 = []; []; p = [];
end
%.......................
%SOLUTION: Most efficient bound
[c j] = max(q);                  %where 'c' = max of function
                                 %      'j' = index of 'm'
s = r(j);                        %maximizing value of 'm'
Ms = B(s:T);
s3 = 1/(T-s);
s4 = (T-s)^2;
s5 = prod(Ms);
b = (4/(x1*s4*x2*s5))^s3;       %upper limit (min of function)
```


Algorithm 5: Quadratic form in residual vector

Function expresses the LS estimator in terms of invariant residuals and the defining quadratic form in (2.9). It ascertains the eigenvalues of $\mathbf{W}_{\alpha jx}$ for each model.

```
function E = CDFquadform(T,alpha,x,model);
%.....................................
%Generation of series
if abs(alpha) == 1;
    y(1) = 0;
else
    y(1) = randn/sqrt( 1 - alpha*alpha );
end
for i=2:T
    y(i) = alpha * y(i-1) + randn;
end
y1 = y';
%....................
%Vector and Matrices definition
```

```
yy = y1(2:T);
ym = y1(1:T-1);
I = eye(T-1);
D1 = [zeros(T-1,1) I];
D2 = [I zeros(T-1,1)];
%.....................
%Definition of R_alpha Matrix
if abs(alpha) == 1;
    b = 0;
else
    b = 1/sqrt( 1 - alpha*alpha );
end
B = ones(T,T);
B(:,1) = b.*B(:,1);
for i = 1:T
    for j = 1:T
        k = i-j;
        if k>=0
            R(i,j) = B(i,j).*(alpha.^(abs(k)));
        else
            R(i,j) = 0;
        end
    end
end
%.....................
Model = model;
switch Model;
    case 1       %Model 1
        P1 = zeros(T-1,T-1);
        W_1 = R'*((D1'*(I - P1)*D2)/2 + (D2'*(I - P1)*D1)/2 - c*(D2'*(I - P1)*D2))*R;
        E = eig(W_1);
    case 2       %Model 2
        X2 = [ones(T-1,1)];
        P2 = X2*inv(X2'*X2)*X2';
        W_2 = R'*((D1'*(I - P2)*D2)/2 + (D2'*(I - P2)*D1)/2 - c*(D2'*(I - P2)*D2))*R;
        E = eig(W_2);
    case 3    %Model 3
        X3 = [ones(T-1,1) [1:1:T-1]'];
        P3 = X3*inv(X3'*X3)*X3';
        W_3 = R'*((D1'*(I - P3)*D2)/2 + (D2'*(I - P3)*D1)/2 - c*(D2'*(I - P3)*D2))*R;
        E = eig(W_3);
    otherwise
        E = [];
        fprintf('input valid parameters \n');
end
```


Algorithm 6: Trapezoidal implementation: CDF computation

Function `CDFtrapezoidal` computes the integral (2.13) using the trapezoidal rule. The main difference from the Simpson's rule implementation is that `P` in Algorithm 3 is user defined by the choice of a reasonably large number of subintervals say `M` $= 2000$. Better accuracy is achieved with smaller stepsizes. The *MATLAB* function `trapz` is used to obtain the result(s).

```
function F = CDFtrapezoidal(T,alpha,c,p,model);
%.....................................
E = CDFquadform(T,alpha,c,model);
E = real(E);
a = 0;                      %Lower limit
b = CDFtruncate(E);         %Upper limit
```

```
M = 2000;                   %Subintervals
P = linspace(a,b,M);        %Points of integrand evaluation
%.....................
for i = 1:M;
    u = P(i);
    if u == 0;
        x(i) = 0;
        g = 0.5 * sum(E);
        y(i) = g;
    else
        theta = 0.5 * sum( atan(E*u) );
        rho =  prod((1 + (E.^2).* u^2).^(0.25));
        A = sin(theta);
        B = u*rho;
        x(i) = u;
        y(i) = (1/pi) * (A/B);
    end
end
%.....................
[x' y'];                    %Points and integrand value
h = (b-a)/M;                %Subinterval length
z = trapz(x,y);             %Trapezoidal result: unit spacing
J = z * h;                  %Actual result after rescaling
%.....................
F = 0.5 - J - p;
```

Algorithm 7: Imhof's CDF truncation bounds

Function `b = CDFIMFtruncate(E)` ascertains the truncation bound at a pre-specified error of truncation using Equation (2.18). The function may be called separately given the vector of eigenvalues `E` or by `CDFsimpson` in Algorithm 2.


```
function b = CDFIMFtruncate(E);
%.....................................
%Definitions
T = length(E);
te = 1.e-8;         %Truncation error
A = abs(E);
%.....................
%constant terms
M = A;
x1 = pi^2;
x2 = te^2;
x3 = 1/T;
x4 = T^2;
x5 = prod(M);
b = (4/(x1*x4*x2*x5))^x3;
```

Algorithm 8: Numerical evaluation of the PDF

Function `f = PDFsimpson(T,alpha,x,model)` evaluates the density function f for given values of T, α, x and model parameters j, $j = 1, 2, 3$ using either Simpson's quadrature implemented by `quad` or the Lobatto quadrature utilized by `quadl`.

```
function f = PDFsimpson(T,alpha,x,model);
%.....................................
[E V DW] = PDFquadform(T,alpha,x,model);

a = 0;                    %Lower limit
b = PDFtruncate(E,V,DW);  %Upper limit
tolr = []; %Default tolerance adequate for computation of mean of LS estimator

f = quadl(@PDFintegrand,a,b,tolr,[],E,V,DW); %Lobatto quadrature
```

Algorithm 9: Integrand evaluation: PDF

Function `y = PDFintegrand(P,E,V,DW)` accepts the eigenvalue, eigenvectors and derivatives of the eigenvalues and returns the integrand evaluated at different points contained in `P`. `V` and `DW` symbolically denote the eigenvectors and derivatives of the eigenvalues [see (2.25) on page 25] in `E`.


```
function y = PDFintegrand(P,E,V,DW)
%.....................................
n = length(P);
for j = 1:n;
    u = P(j);
    theta = 0.5 * sum( atan(E*u) );
    rho =  prod((1 + (E.^2).* u^2).^(0.25));
    A = ((E.*u).*sin(theta) - cos(theta))./(1 + (E.^2).* u^2);
    B = sum((A .* DW)/rho);
    x(j) = u;
    y(j) = 1/(2*pi)*B;
end
```

Algorithm 10: Minimal / efficient truncation bounds: PDF

Function `b = PDFtruncate(E,V,DW)` evaluates the most efficient truncation bounds for a pre-specified level on the truncation error, say 10^{-20}. The formulation (2.30) is implemented as a maximization problem for the choice of m.

```
function b = PDFtruncate(E,V,DW);
%.....................................
te = 1.e-20;         %Truncation error
A = abs(E);
B = abs(DW);
C = [A B];
D = sortrows(C,1);  %Order statistics of 'E' and 'DW'
%.....................
```

```
%Restrictions
%(1) E <= 1 (summation term)
T = length(E);
n = length(find(D(:,1) <=1 ));
A1 = D(1:n,1);                      %E <= 1
B1 = D(1:n,2);                      %DW for E <= 1
A2 = D(n+1:T,1);                    %E > 1
B2 = D(n+1:T,2);                    %DW for E > 1
Sn1 = sum(B1.*(A1+1));              %Sum from 'j=1' to 'n'
Sn2 = sum((B2./A2.^2).*(A2+1));     %Sum from 'j=n+1' to 'T'
Sn = Sn1 + Sn2;
%.....................
%Constants
x1 = Sn^2;
x2 = pi^2;
x3 = te^2;
%.....................
%(2) E << 1 (product term)
MT = D(:,1);                        %All eigenvalues
for m = 1:T-5;
    M = MT(m:T);
    x4 = 1/(T-m-4);
    x5 = (T-m-4)^2;
    x6 = prod(M);                   %product from 'j=m+1' to 'T'
    %..................
    %maximization
    p = ((x2*x5*x3*x6)/x1)^x4;
    q(m) = p;
    r(m) = m;
    %..................
    M = [];  x4 = []; x5 = []; x6 = []; p = [];
end
```

Algorithm 11: Quadratic form: PDF computation

Function `[E V DW] = PDFquadform(T,alpha,c,model)` adopts a similar structure as `E = CDFquadform(T,alpha,c,model)` except for the additional outputs. The startup code is identical except for the following section. Duplication of code is avoided owing to space restrictions.


```
function [E V DW] = PDFquadform(T,alpha,c,model)
%.......................................


%SEE THE FUNCTION CDF QUADFORM


%.......................................
Model = model;
switch Model;
  case 1      %MODEL ONE
      P1 = zeros(T-1,T-1);
      W_1 = R'*((D1'*(I - P1)*D2)/2 + (D2'*(I - P1)*D1)/2 - c*(D2'*(I - P1)*D2))*R;
      [V,E1] = eig(W_1);
      E = diag(E1);
      DWC = - R'*(D2'*(I-P1)*D2)*R;
      DW = diag(V' * DWC * V);
  case 2      %MODEL TWO
      X2 = [ones(T-1,1)];
      P2 = X2*inv(X2'*X2)*X2';
      W_2 = R'*((D1'*(I - P2)*D2)/2 + (D2'*(I - P2)*D1)/2 - c*(D2'*(I - P2)*D2))*R;
```

```
      [V,E1] = eig(W_2);
      E = diag(E1);
      DWC = - R'*(D2'*(I-P2)*D2)*R;
      DW = diag(V' * DWC * V);
  case 3   %MODEL THREE
      X3 = [ones(T-1,1) [1:1:T-1]'];
      P3 = X3*inv(X3'*X3)*X3';
      W_3 = R'*((D1'*(I - P3)*D2)/2 + (D2'*(I - P3)*D1)/2 - c*(D2'*(I - P3)*D2))*R;
      [V,E1] = eig(W_3);
      E = diag(E1);
      DWC = - R'*(D2'*(I-P3)*D2)*R;
      DW = diag(V' * DWC * V);
  otherwise
      [E V DW] = [];
      fprintf('function aborted; Input valid model parameter; 1 , 2 or 3 \n');
end
```

Algorithm 12: Mean of the LS estimator

The function `mn = LSmean(T,alpha,model,par)` determines the mean of the LS estimator defined in (2.32) via the numerical integration method. Density values $\leq$ `tol` $= 10^{-8}$ are ignored. `par` is a step-size parameter over which the density calculations proceed. The computation of the variance, MSE and RMS of the LS estimator is done using a similar subroutine.


```
function mn = LSmean(T,alpha,model,par);
%-------------------------------------------------------------------------
tol = 1.e-8;
c = alpha;
f = PDFsimpson(T,alpha,c,model);
%-----------------------
x = c;
y = f;
m = x * f;
%-----------------------
%PDF right
j = 1;
f1 = f;
c1 = c;
while abs(f1) > tol;
    c1 = c1 + par;
    f1 = PDFsimpson(T,alpha,c1,model);
    x1(j) = c1;
    y1(j) = f1;
    m1(j) = c1 * f1;
    j = j + 1;
end
%-----------------------
%PDF left
k = 1;
f2 = f;
c2 = c;
while abs(f2) > tol;
    c2 = c2 - par;
    f2 = PDFsimpson(T,alpha,c2,model);
    x2(k) = c2;
    y2(k) = f2;
    m2(k) = c2 * f2;
    k = k + 1;
end
```

```
nx2 = length(x2);
x2 = x2(nx2:-1:1);
y2 = y2(nx2:-1:1);
m2 = m2(nx2:-1:1);
%----------------------
X = [x2 x x1];
Y = [y2 y y1];
M = [m2 m m1];
mn   = trapz(X,M);
```

Algorithm 13: Maximum of density

The function `mode = PDFmin(T,alpha,c,model)` is a minor modification of `f = PDFsimpson(T,alpha,x,model)`. It facilitates use of `fminsearch` or `fminbnd` in determination of the mode of the LS estimator for given α and T, as discussed in section 2.4.3.

```
function mod = PDFmin(T,alpha,x,model);
%.......................................
[E V DW] = PDFquadform(T,alpha,c,model);

a = 0;                              %Lower limit
b = PDFtruncate(E,V,DW);            %Higher bound on truncation error advisable e.g 1.e-15
tolr = 1.e-14; %High tolerance value is necessary for the mode

f = quad(@PDFintegrand,a,b,tolr,[],E,V,DW);   %Simpsons quadrature
mode = -f;
```

Algorithm 14: Mode using the simulation method: implementation of the Epanechnikov and Triangular kernels.

The function `A = Kernel(alpha,T,N,model)` determines the mode of the EDF of the LS estimator in (3.5). It uses the two kernels in (3.6) and (3.7). The output vector `A` consists of α, the mode estimates before (`EP` and `TR`) and after (`EPS` and `TRS`) smoothing.

```
function A = Kernel(alpha,T,N,model);
%.......................................
z = LsEst(alpha,T,N,model);
z = sort(z);
%.....................
%Bandwidths
stdz = std(z);
iqr = 0.741 *( z(0.75*N)- z(0.25*N) );
sig = min(stdz,iqr);
he = 2.345 * iqr * N^(-0.2);            %Epanechnikov kernel
ht = 2.576 * iqr * N^(-0.2);            %Triangular kernel
%.....................
%Using the Histogram
R = 0.5;
[a b] = hist(z,400);
c = [a' b'];
[d1 d2] = max(c);
n1 = d1(1);
n2 = d2(1);
```

```
x  = b(n2);                              %OR x = median(z); %(using the median)
xmin = x -  R * stdz;
xmax = x +  R * stdz;
m1 = length(find(z <= xmin));
m2 = length(find(z <= xmax));
h = floor((m2-m1)/2000);
%.....................
%Kernel functions
i = 1;
for j = m1:h:m2;
    x0 = z(j);
    X(i) = x0;
    XE = abs( (x0 - z)./he );
    XT = abs( (x0 - z)./ht );
    x0 = [];
    %.................
    m = find(XE > 1);
    XE(m) = [];
    m = [];
    SEP = sum( 0.75 * (1 - XE.^2) ); %Epanechnikov
    %.................
    m = find(XT > 1);
    XT(m) = [];
    m = [];
    STR = sum( 1 - abs(XT) );          %Triangular
    %.................
    %EDF point estimates
    F(i) = ( 1 /(N*he) ) * SEP;        %EDF EP
    G(i) = ( 1 /(N*ht) ) * STR;        %EDF TR
    i = i + 1;
    x0 = [];  SEP = []; STR = [];
end
%.....................
%Mode BEFORE smoothing
[Fmax u] = max(F);                     %EP
EP = X(u);
Fmax = []; u = [];
%............
[Gmax u] = max(G);                     %TR
TR = X(u);
Gmax = []; u = [];
%.....................
%Mode AFTER smoothing
if abs(alpha) <= 0.9;
    dp = 3;
else
    dp = 4;
end
%............
P1       = polyfit(X,F,dp);        %Polynomial fit
Q1       = polyval(P1,X);
A1       = polyder(P1);            %Derivative of polynomial
B1       = roots(A1);
[Bmin v] = min(abs(real(B1 - alpha)));
EPS      = real(B1(v));             %Solution
P1 = []; Bmin = []; v = [];
%............
P2       = polyfit(X,G,dp);
Q2       = polyval(P2,X);
A2       = polyder(P2);
B2       = roots(A2);
[Bmin v] = min(abs(real(B2 - alpha)));
TRS = real(B2(v));
P1 = []; Bmin = []; v = [];
%.....................
A = [alpha EP EPS TR TRS];
```

Algorithm 15: Jacobian of transformation: $d\alpha_{\text{ES}}/d\alpha_{\text{UE}}$

The function `J = Jacobian(model,a,fnct,n,dp,Polate)` determines the Jacobian of the transformation from α_{LS} to α_{UE} given the location functions in a particular model. The input parameters are: `model`, the model; '`a`', a matrix on which the location functions are defined over an adequate mesh of points (first column) in Ω, as discussed in the text; `fnct`, the location functions - the columns of `a` i.e. 2, 3 and 4 are used for the mean, median and mode functions, respectively; `n`, the number of points over which the polynomials $g_n(\cdot)$ defined in (4.14) are calculated; `dp`, the degree of the polynomial; and `Polate`, an interpolation option where the value 1 indicates interpolation with $M = 2000$ points and 2 represents no interpolation. The $M \times 2$ output vector `J` consists of the values of α in the first column against the corresponding Jacobian values. Each row in `a` should be in the order $[\,\alpha\,\bar{\alpha}_j\,\tilde{\alpha}_j\,\ddot{\alpha}_j\,]$ for a given model and T.


```
function J = Jacobian(model,a,fnct,n,dp,Polate);
%.....................................
x = a(:,1);             % Alpha values
y = a(:,fnct);          % (2)Mean (3)Median (4)Mode
N = length(x);
%.....................
%Interpolation option
if Polate == 1;
    u = x(1);
    v = x(N);
    M = 2000;
    U = linspace(u,v,M);
    V = interp1(x,y,U,'spline');
    x = U'; y = V';
    N = M;
end
%.....................
if fnct == 2 | fnct == 3 |fnct == 4;
    i = 1;
    while i < N-n+1;
        b = x(i);
        X = x(i:i+n-1);
        Y = y(i:i+n-1);
        %.....................
        %Midpoint
        m1 = mod(n,2);
        m2 = floor(n/2);
        m  = m1 + m2;
        %.....................
        if m1 == 0;
            xm(i) = (X(m2)+ X(m2+1))/2;
            ym(i) = (Y(m2)+ Y(m2+1))/2;
        else
            xm(i) = X(m2+1);
            ym(i) = Y(m2+1);
        end
        %.....................
        %Polynomial fit and Derivative
        P = polyfit(X,Y,dp);
```

```
        D = polyder(P);
        Pm(i) = polyval(P,xm(i));
        Dm(i) = polyval(D,xm(i));
        %.....................
        if i == 1;
            xs = X;
            ys = Y;
            Ps = polyval(P,X);
            Ds = polyval(D,X);
        elseif i == N-n;
            X = x(N-n+1:N);
            Y = y(N-n+1:N);
            P = polyfit(X,Y,dp);
            D = polyder(P);
            xe = x(N-n+2:N);
            Pe = polyval(P,xe);
            De = polyval(D,xe);
        else
        end
        i = i + 1;
    end
end
%.....................
A = [xm' Dm'];       %Intermediate values
B = [xs  Ds ];        %Endpoints near -1
C = [xe  De ];        %Endpoints near +1
delS = find( B(:,1) >= xm(1) );
B(delS,:) = [];
%.....................
J = [B; A; C]; %Jacobian values
%.....................
%Dumping implementation parameter values
a = []; b = []; x = []; y = []; N = []; fnct = []; X = [];
Y = []; m = []; m1 = []; m2 = []; xm = []; ym = []; Pm = [];
Dm = []; P = []; D = []; xs = []; ys = []; Ps = []; Ds = [];
Pe = []; De = []; A = []; B = []; C =[];
```


Algorithm 16: MSE computation

The function `[X Y Z] = MSE(model,fnct,np,dp,Data,Polate,tol,T,a)` determines the MSE of the location-based estimators for given α, T in a particular model; MSE of the LS estimator (for checking purposes); and exact confidence intervals. Following the MSE computations, it determines the mean, bias and variance. A check procedure, using variance as discussed in the text, is provided. `Polate` is (1) yes and (2) no. `par` is a stepsize parameter used in computation of the density of the location-based estimators. The other parameters are as given in evaluation of the Jacobian in the previous algorithm. Explanatory statements are included within the code.

```
function [X Y Z] = MSE(model,fnct,np,dp,Data,Polate,T,a);
%.....................................
par = 0.0005;
tol = 1.e-8; %   Tolerance parameter (PDF)
N = length(a(:,1));
L = a(1,fnct);                                  %   g(-1) value
U = a(N,fnct);                                  %   g(1) value
J = Jacobian(model,fnct,np,dp,Data,Interpolate);    %   (1)Interpolate,(2) Dont
```

```
M = length( J(:,1) );
%.....................................
%EQUIVALENT LS ESTIMATOR
r = length( find(a(:,1) <= alpha) );
if r == 0;
    c = a(1,fnct);
elseif r == N;
    c = a(N,fnct);
else
    A1 = a(r,1);        L1 = a(r,fnct);
    A2 = a(r+1,1);      L2 = a(r+1,fnct);
    c = L1 + ( (alpha - A1)/(A2 - A1)) * (L2 - L1);
    A1 = []; A2 = []; L1 = []; L2 = []; r = [];
end
g = PDFsimpson(T,alpha,c,model);                    %PDF LS
%FINAL ESTIMATOR (CHECK: EQUAL TO 'alpha')
r = length(find( a(:,fnct) <= c ) );
if r == 0;
    z = a(1,1);
elseif r == N;
    z = a(N,1);
else
    A1 = a(r,1);        L1 = a(r,fnct);
    A2 = a(r+1,1);      L2 = a(r+1,fnct);
    z = A1 + ( (c - L1)/(L2 - L1)) * (A2 - A1);
    A1 = []; A2 = []; L1 = []; L2 = []; r = [];
end
%JACOBIAN OF TRANSFORMATION
r = length( find(J(:,1) <= z) );
if r == 0;
    jac = J(1,2);
elseif r == M;
    jac = J(M,2);
else
    A1 = J(r,1);        J1 = J(r,2);
    A2 = J(r+1,1);      J2 = J(r+1,2);
    jac = J1 + ((z - A1)/(A2 - A1)) * (J2 - J1);
    A1 = []; A2 = []; J1 = []; J2 = []; r = [];
end
h = g * jac;                %PDF OF FINAL ESTIMATOR
Ax  = c;                    %LS value
Ay  = g;                    %PDF of LS estimator
Az  = z;                    %Final estimator
Af  = h;                    %PDF of Final estimator
AJ = jac;                   %Jacobian check
%.....................................
%PDF TO THE RIGHT OF 'c': INITIALIZATION
c1     = c;                 %Initialize LS
PDFtol = Ay;                %Initialize PDF of AF
i      = 1;                 %Counter
j      = 1;                 %Counter for LS > g(1)
while abs(PDFtol) > tol;
    c1 = c1 + par;
    %.....................
    if c1 > U;              %LS > g(1) in Andrews,1993
        %Initialize: Discrete calculation & Area
        if j == 1;
            z   = a(N,1);
            Dx0 = U;
            Dz0 = 1;
            Dz0a = z;       %Within (-1,1)
            Dy0 = PDFsimpson(T,alpha,Dx0,model);
            %.............
            %Jacobian
            r = length( find(J(:,1) <= z) );
            if r == 0;
                jac = J(1,2);
            elseif r == M;
```

```
            jac = J(M,2);
        else
            A1 = J(r,1);        A2 = J(r+1,1);
            J1 = J(r,2);        J2 = J(r+1,2);
            jac = J1 + ((z - A1)/(A2 - A1)) * (J2 - J1);
            A1 = []; A2 = []; J1 = []; J2 = []; r = [];
        end
        %..............
        Dj0 = jac;
        Df0 = Dy0 * jac;                %PDF of Final Estimator
        Dmp1 = Dy0;
        MP = 0;                         %Probability Mass Positive (@ +1)
    end
    g = PDFsimpson(T,alpha,c1,model);   %PDF LS estimator
    z = 1;                              %Equivalent Final estimator
    %.................
    %DISCRETIZATION: Masspoint probability (trapezoidal rule)
    Dmp2 = g;
    MP = MP + ( (0.5 *(Dmp1 + Dmp2))* par );
    Dmp1 = Dmp2;
    Dmp2 = [];
    Dx1(j) = c1;                        %LS values
    PMP(j) = MP;                        %Enforcing indexing
    %NUMERICAL METHOD: Masspoint probs
    Dx(j) = c1;                         %LS value
    Dy(j) = g;                          %PDF of LS estimator
    Dz(j) = z;                          %Final estimator
    PDFtol = g;                         %For tolerance check
    j      = j + 1;                     %Counter when LS > g(1)
else
    %FINAL ESTIMATOR
    r = length(find( a(:,fnct) <= c1 ) );
    if r == 0;
        z = a(1,1);
    elseif r == N;
        z = a(N,1);
    else
        A1 = a(r,1);        L1 = a(r,fnct);
        A2 = a(r+1,1);      L2 = a(r+1,fnct);
        z = A1 + ( (c1 - L1)/(L2 - L1)) * (A2 - A1);
        A1 = []; A2 = []; L1 = []; L2 = []; r = [];
    end
    g = PDFsimpson(T,alpha,c1,model);   %PDF LS estimator
    %.................
    %Jacobian of transformation
    r = length( find(J(:,1) <= z) );
    if r == 0;
        jac = J(1,2);
    elseif r == M;
        jac = J(M,2);
    else
        A1 = J(r,1);        J1 = J(r,2);
        A2 = J(r+1,1);      J2 = J(r+1,2);
        jac = J1 + ((z - A1)/(A2 - A1)) * (J2 - J1);
        A1 = []; A2 = []; J1 = []; J2 = []; r = [];
    end
    h1 = g * jac;                   %PDF of Final estimator
    PDFtol = g;
    %.................
    Bx(i)  = c1;                        %LS value
    By(i)  = g;                         %PDF of LS estimator
    Bz(i)  = z;                         %Final estimator
    Bf(i)  = h1;                        %PDF of Final estimator
    BJ(i)  = jac;                       %Jacobian check
    i = i + 1;
end%...end IF loop 'c1'
end%.......end WHILE loop 'h1'
%.........................................
```

```
%PDF TO THE LEFT OF 'c' : INITIALIZATION
c2     = c;                                  %Initialize LS
PDFtol = Ay;                                 %Initialize PDF of AF
k      = 1;                                  %counter for LS within [L U]
p      = 1;                                  %Counter for LS < g(-1)
par    = par0;
while abs(PDFtol) > tol;
    c2 = c2 - par;
    %.....................
    if c2 < L;
        %Initialize: Discrete calculation & Area
        if p == 1;
            z   = a(1,1);
            Ex0 = L;                         %LS values
            Ez0 = -1;
            Ez0a = z;                       %Final Estimator within [-1,1]
            Ey0 = PDFsimpson(T,alpha,Ex0,model); %PDF of LS
            %.............
            %Jacobian
            r = length( find(J(:,1) <= z) );
            if r == 0;
                jac = J(1,2);
            elseif r == M;
                jac = J(M,2);
            else
                A1 = J(r,1);        A2 = J(r+1,1);
                J1 = J(r,2);        J2 = J(r+1,2);
                jac = J1 + ((z - A1)/(A2 - A1)) * (J2 - J1);
                A1 = []; A2 = []; J1 = []; J2 = []; r = [];
            end
            %.............
            Ej0 = jac;
            Ef0 = Ey0 * jac;             %PDF of Final Estimator
            Dmp1 = Ey0;
            MN = 0;                       %Probability Mass Negative (@ -1)
        end
        g = PDFsimpson(T,alpha,c2,model); %PDF LS estimator
        z = -1;                           %Equivalent Final estimator
        %.................
        %DISCRETIZATION: Masspoint probability
        Dmp2 = g;
        MN = MN + ( (0.5 *(Dmp1 + Dmp2))* par );
        Dmp1 = Dmp2;
        Dmp2 = [];
        Dx2(p) = c2;
        PMN(p) = MN;
        %.................
        Ex(p)  = c2;                    %LS value
        Ey(p)  = g;                        %PDF of LS estimator
        Ez(p)  = z;                        %Final estimator
        PDFtol = g;                     %For tolerance check
        p      = p + 1;                   %Counter when LS < g(-1)
    else
        %FINAL ESTIMATOR
        r = length(find( a(:,fnct) <= c2 ) );
        if r == 0;
            z = a(1,1);
        elseif r == N;
            z = a(N,1);
        else
            A1 = a(r,1);        L1 = a(r,fnct);
            A2 = a(r+1,1);      L2 = a(r+1,fnct);
            z = A1 + ( (c2 - L1)/(L2 - L1)) * (A2 - A1);
            A1 = []; A2 = []; L1 = []; L2 = []; r = [];
        end
        g = PDFsimpson(T,alpha,c2,model);   %PDF LS estimator
        %.................
        %Jacobian of transformation
```

```
        r = length( find(J(:,1) <= z) );
        if r == 0;
            jac = J(1,2);
        elseif r == M;
            jac = J(M,2);
        else
            A1 = J(r,1);        A2 = J(r+1,1);
            J1 = J(r,2);        J2 = J(r+1,2);
            jac = J1 + ((z - A1)/(A2 - A1)) * (J2 - J1);
            A1 = []; A2 = []; J1 = []; J2 = []; r = [];
        end
        %..................
        h2 = g * jac;                    %PDF of Final estimator
        Cx(k)  = c2;                    %LS value
        Cy(k)  = g;                      %PDF of LS estimator
        Cz(k)  = z;                       %Final Estimator
        Cf(k)  = h2;                      %PDF of Final estimator
        CJ(k)  = jac;                    %Jacobian check
        PDFtol = g;
        k = k + 1;
    end%...end IF loop 'c2'
end%.......end WHILE loop 'h2'
%.........................................
%MATRIX CONCATENATION
A = [Ax' Az' Ay' Af' AJ'];                 %[LS UE PDFLS PDPUE JAC]
%..................
if  i > 1;   %Equivalent to ( if B ~= [] )
    B = [Bx' Bz' By' Bf' BJ'];             %[LS UE PDFLS PDPUE JAC]
else
    B = [];
end
%..................
if k > 1;    %Equivalent to ( if C ~= [] )
    C = [Cx' Cz' Cy' Cf' CJ'];             %[LS UE PDFLS PDPUE JAC]
    nc = length(Cx);
    C = C(nc:-1:1,:);                      %Reversed indices
else
    C = [];
end
%..................
if j > 1;
    %Numerical calculation
    DX0 = [Dx0 Dz0a Dy0 Df0 Dj0];          %[LS UE PDFLS PDPUE JAC]
    B   = [B; DX0];                          %Including the endpoint at U
    DX1 = [Dx0 Dz0 Dy0 Dy0];                 %Including LSPDF at U
    D   = [Dx' Dz' Dy' Dy'];                 %[LS UE PDFLS PDPUE]
    D   = [DX1; D];                          %To cater for continuity
    %Discret calculation
    Ds0 = [Dx0 Dz0 Dy0 0];
    Dsc = [Ds0; Dx1' Dz' Dy' PMP'];          %[LS UE PDFLS CUMMAREA]
else
    B    = B;
    DX0  = [];
    DX1  = [];
    D    = [];
    Ds0  = [];
    Dsc  = [];
end
%..................
if  p > 1;
    %numerical calculation
    EX0 = [Ex0 Ez0a Ey0 Ef0 Ej0];            %[LS UE PDFLS PDPUE JAC]
    C   = [EX0; C];                          %Including endpoint at L
    EX1 = [Ex0 Ez0 Ey0 Ey0];                 %Including LSPDF at L
    E   = [Ex' Ez' Ey' Ey'];                 %[LS UE PDFLS PDPUE]
    E   = [EX1; E];                          %To cater for continuity
    ne  = length(Ex);
    E = E(ne:-1:1,:);                        %Reversed indices
```

```
    %Discrete calculation
    Es0 = [Ex0 Ez0 Ey0 0];
    Esc = [Es0; Ex' Ez' Ey' PMN'];          %[LS UE PDFLS CUMMAREA]
    nesc = length(Esc(:,1));
    Esc  = Esc(nesc:-1:1,:);
else
    C   = C;
    EX0 = [];
    EX1 = [];
    E   = [];
    Es0 = [];
    Esc = [];
end
%.........................................
%MASSPOINTS COMPUTATIONS ( +1 )
if j > 1;
    nd1 = length(D(:,1));
    if nd1 <= 2;
        %...................
        %DISCRETE METHOD: Trapezoidal method
        nd2 = length(Dsc(:,1));
        DscProb  = Dsc(nd2,4);              %Area: (trapezoidal rule)
        PMPoint1 = DscProb;                 %Assignment
        %DISCRETE METHOD: Summation method
        DscProb2 = sum( Dsc(:,3) ) * par0
        E1MP1     = sum( Dsc(:,1) .* Dsc(:,3) );            %Sum  x times P(x)
        E2MP1     = sum( Dsc(:,1).^2 .* Dsc(:,3) );         %Sum  x^2 times P(x)
    else
        %...................
        %NUMERICAL METHOD
        NumProb  = trapz( D(:,1), D(:,3) );                 %Area
        E1MP1     = trapz( D(:,1), D(:,1) .* D(:,3) );      %E[x]
        E2MP1     = trapz( D(:,1), D(:,1).^2 .* D(:,3) );   %E[x^2]
        PMPoint1 = NumProb;                                 %Assignment
    end
else
    PMPoint1 = 0;
    E1MP1     = 0;
    E2MP1     = 0;
    DscProb  = 0;
end
%.........................................
%MASSPOINTS COMPUTATIONS ( -1 )
if p > 1;
    ne = length( E(:,1) );
    if ne <= 2;
        %...................
        %DISCRETE METHOD: Trapezoidal method
        DscProb  = Esc(1,4);                                %Area
        PMPoint2 = DscProb;                                 %Assignment
        %DISCRETE METHOD: Summation method
        DscProb2 = sum( Esc(:,3) ) * par0;
        E1MP2     = sum( Esc(:,1)    .* Esc(:,3) ) * par0; %Sum  x times P(x)
        E2MP2     = sum( Esc(:,1).^2 .* Esc(:,3) ) * par0; %Sum  x^2 times P(x)
    else
        %...................
        %NUMERICAL METHOD
        NumProb  = trapz( E(:,1), E(:,3) );                 %Area
        E1MP2     = trapz( E(:,1), E(:,1) .* E(:,3) );       %E[x]
        E2MP2     = trapz( E(:,1), E(:,1).^2 .* E(:,3) ); %E[x^2]
        PMPoint2 = NumProb;                                 %Assignment
    end
else
    PMPoint2 = 0;
    E1MP2     = 0;
    E2MP2     = 0;
    DscProb  = 0;
end
```

```
%.....................................
E1MP = E1MP1 + E1MP2;
E2MP = E2MP1 + E2MP2;
%.....................................
%MERGED RESULTS AND DETERMINATION OF CONFIDENCE BOUNDS
X = [C; A; B];                      %[LS UE PDFLS PDPUE JACOBIAN]
%....................
%Incoporating MP@[-1]
if p > 1;
    Y = [ E; X(:,1:4) ];            %[LS UE PDFLS PDPUE]
    Z = E;
    choice = 2;
    x = confidence(X,Z,par0,choice);%CONFIDENCE INTERVALS (CI)
end;
%....................
%Incoporating MP@[+1]
if j > 1;
    Y = [ X(:,1:4); D ];                %[LS UE PDFLS PDPUE]
    Z = D;
    choice = 1;
    x = confidence(X,Z,par0,choice);    %CONFIDENCE INTERVALS (CI)
end;
%....................
%No mass point
if j <= 1 & p <= 1;
    Y = X(:,1:4);                       %No mass point
    Z = [];
    choice = 0;
    x = confidence(X,Z,par0,choice);    %CONFIDENCE INTERVALS (CI)
end
LLS = x(1,2);           LUE = x(1,3);
RLS = x(2,2);           RUE = x(2,3);
%.....................................
%AREA, MEAN, VARIANCE, SQUARE BIAS and MSE COMPUTATION
%.....................................
%LS ESTIMATOR : Direct computation (Mean, Variance,Square bias )
%....................
LS     = Y(:,1);
PLS    = Y(:,3);
mnls  = trapz( LS, LS    .* PLS);
vrls1  = trapz( LS, LS.^2 .* PLS ) - mnls^2;
vrls2  = trapz( LS, (LS - mnls).^2 .* PLS );
sbls   = (mnls - alpha)^2;
msels1 = vrls1 + sbls;
msels2 = vrls2 + sbls;
%....................
%AREA CALCULATION
%Mass Point
MParea = PMPoint1 + PMPoint2;
%....................
%UNDER PDF of LS
XLS  = X(:,1);
XPLS = X(:,3);
ALS  = trapz( XLS,XPLS );                   %Within [L,U]
AreaLS = ALS + MParea;                      %Total
%....................
UNDER PDF of UE
XUE  = X(:,2);
XPUE = X(:,4);
AUE  = trapz( X(:,2),X(:,4) );              %Within [-1,1]
AreaUE = AUE + MParea;                      %Total
%.....................................
%OPTION TWO: UNBIASED ESTIMATOR: Indirect computation of Mean & Variance
%.....................................
if j > 1 | p > 1;
    ZLS  = Z(:,1);
    ZUE  = Z(:,2);
    ZPLS = Z(:,3);
```

```
    ZPUE = Z(:,4);
    nz = length(ZLS);
    %.....................
    if nz <= 2;
        %DISCRETE METHOD
        MPR   = sum( ZUE .* ZPLS ) * par0;                %Pr(UE = 1) or Pr(UE = -1)
        E1ZUE = sum( ZUE .* ZPLS ) * par0;                %Sum x times P(x)
        MNUE  = E1XUE + E1ZUE;                            %Mean of UE
        ZVUE  = sum( (ZUE - MNUE).^2 * MPR   ) * par0;    %Sum x^2 times P(x)
        ZVUE1 = sum( (ZUE - MNUE).^2 .* ZPLS ) * par0;    %Sum x^2 times P(x): Direct
    else
        %NUMERICAL METHOD
        MPR   = trapz( ZLS, ZUE .* ZPLS );                %Pr(UE = 1) or Pr(UE = -1)
        E1ZUE = trapz( ZLS, ZUE .* ZPLS );                %E[x] in Z
        MNUE  = E1XUE + E1ZUE;                            %Mean of UE
        ZVUE  = trapz( ZLS, ( ZUE - MNUE ).^2 * MPR  );   %E[(x - E{x})^2]
        ZVUE1 = trapz( ZLS, ( ZUE - MNUE ).^2 .* ZPLS );  %E[(x - E{x})^2]: Direct
    end
    %.....................
else
    MPR  = 0;
    E1ZUE = 0;
    MNUE  = E1XUE;
    ZVUE = 0;
    ZVUE1 = 0;
end
%.......................................
VRUE1  = XVUE + ZVUE1;
VRUE2  = XVUE + ZVUE;
SBUE2  = (MNUE - alpha)^2;
MSEUE1 = VRUE1 + SBUE2;
MSEUE2 = VRUE2 + SBUE2;
%.......................................
```

Algorithm 17: Exact confidence intervals

The function `x = confidence(X,Z,par0,choice)` determines 5% and 95% quantiles of the CDF of the location-based estimators. It is effected as a subcall routine during MSE computation. It is also used to establish exact bounds for LS estimator for check purposes on accuracy. Explanatory text is included in the subroutines.

```
function x = confidence(X,Z,par0,choice);
%.......................................
XLS  = X(:,1);      %LS estimator
XPLS = X(:,3);      %PDF of LS
XUE  = X(:,2);      %Unbiased estimator
XPUE = X(:,4);      %PDF of UE
X1  = cumtrapz(XUE,XPUE);
X2  = cumtrapz(XLS,XPLS);
%.......................................
switch choice;
    %.........................
    case 0; %No mass point j = 1 & p = 1;
        %UNBIASED ESTIMATOR
        UE = XUE;
        CPUE = X1;
        i = 1;
        for ci = [0.05 0.95];
            r = length( find( CPUE <= ci ) );
            A = UE(r);      C = CPUE(r);
            B = UE(r+1);    D = CPUE(r+1);
```

```
        cue(i) = A + ((ci - C)/(D - C))*(B-A);
        A = []; B = []; C = []; D = []; r = [];
        i = i + 1;
    end
    %.....................
    LS = XLS;
    CPLS = X2;
    i = 1;
    for ci = [0.05 0.95];
        r = length( find( CPLS <= ci ) );
        A = LS(r);      C = CPLS(r);
        B = LS(r+1);    D = CPLS(r+1);
        cls(i) = A + ((ci - C)/(D - C))*(B-A);
        A = []; B = []; C = []; D = []; r = [];
        i = i + 1;
    end
    %.........................
case 2; %Masspoint probability at ( -1 )
    ZLS  = Z(:,1);      %LS estimator
    ZPLS = Z(:,3);      %PDF of LS
    ZUE  = Z(:,2);      %Unbiased estimator
    ZPUE = Z(:,4);      %PDF of UE
    nz = length(ZLS);
    if nz <= 2;
        Z1 = cumsum( ZPUE ) * par0;         %Discrete method
        Z2 = cumsum( ZPLS ) * par0;
    else
        Z1 = cumtrapz( ZLS, ZPUE );      %Numerical method
        Z2 = cumtrapz( ZLS, ZPLS );
    end
    %.....................
    %UNBIASED ESTIMATOR
    pzue = Z1( length(Z1) );
    X1   = X1 + pzue;
    UE   = [ZUE; XUE];
    CPUE = [Z1; X1];
    i = 1;
    for ci = [ 0.05 0.95 ];
        if ci < pzue;
            cue(i) = -1;
        else
            r = length( find( CPUE <= ci ) );
            A = UE(r);      C = CPUE(r);
            B = UE(r+1);    D = CPUE(r+1);
            cue(i) = A + ((ci - C)/(D - C))*(B-A);
            A = []; B = []; C = []; D = []; r = [];
        end
        i = i + 1;
    end
    %.....................
    %LS ESTIMATOR
    pzls = Z2( length(Z2) );
    X2   = X2 + pzls;
    LS   = [ZLS; XLS];
    CPLS = [Z2; X2 ];
    i = 1;
    for ci = [0.05 0.95];
        r = length( find( CPLS <= ci ) );
        A = LS(r);      C = CPLS(r);
        B = LS(r+1);    D = CPLS(r+1);
        cls(i) = A + ((ci - C)/(D - C))*(B-A);
        A = []; B = []; C = []; D = []; r = [];
        i = i + 1;
    end
    %.........................
case 1; %Masspoint probability at ( +1 )
    ZLS  = Z(:,1);      %LS estimator
    ZPLS = Z(:,3);      %PDF of LS
```

```
        ZUE  = Z(:,2);      %Unbiased estimator
        ZPUE = Z(:,4);      %PDF of UE
        nz = length(ZLS);
        if nz <= 2;
            Z1 = cumsum( ZPUE ) * par0;         %Discrete method
            Z2 = cumsum( ZPLS ) * par0;
        else
            Z1 = cumtrapz( ZLS, ZPUE );      %Numerical method
            Z2 = cumtrapz( ZLS, ZPLS );
        end
        %.....................
        %UNBIASED ESTIMATOR
        pxue  = X1( length(X1) );       %Cummulative area in [-1,1]
        Z1  = Z1 + pxue;
        UE  = [XUE; ZUE];
        CPUE = [X1; Z1];
        i = 1;
        for ci = [ 0.05 0.95 ];
            if ci > pxue;
                cue(i) = 1;
            else
                r = length( find( CPUE <= ci ) );
                A = UE(r);      C = CPUE(r);
                B = UE(r+1);    D = CPUE(r+1);
                cue(i) = A + ((ci - C)/(D - C))*(B-A);
                A = []; B = []; C = []; D = []; r = [];
            end
            i = i + 1;
        end
        %.....................
        %LS ESTIMATOR
        pxls = X2( length(X2) );
        Z2 = Z2 + pxls;
        LS   = [XLS; ZLS];
        CPLS = [X2; Z2];
        i = 1;
        for ci = [0.05 0.95];
            r = length( find( CPLS <= ci ) );
            A = LS(r);      C = CPLS(r);
            B = LS(r+1);    D = CPLS(r+1);
            cls(i) = A + ((ci - C)/(D - C))*(B-A);
            A = []; B = []; C = []; D = []; r = [];
            i = i + 1;
        end
    otherwise
end
x = [[0.05 0.95]' cls' cue'];
%.......................................
```

Bibliography

[1] Abadir, K. M. (1993). OLS bias in a nonstationary regression. *Econometric Theory,* Vol. 9, No. 1, 81-93.

[2] Abadir, K. M. (1995). Unbiased estimation as a solution to testing for random walks. *Economics Letters,* Vol. 47, No. 3, 263-268.

[3] Ali, M. M. (2002). Exact distribution of the least squares estimator in a first order autoregressive model. *Econometrics review,* Vol. 21, No. 1, 89-119.

[4] Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica,* Vol. 61, No. 1, 139-165.

[5] Andrews, D. W. K. and Chen, H. (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business and Economic Statistics,* Vol. 12, No. 2, 187-204.

[6] Beach, C. M. and MacKinnon, J. G. (1978). A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica,* Vol. 46, No. 1, 51-58.

[7] Bickel, D. R. (2002). Robust estimators of the mode and skewness of continuous data. *Computational Statistics and Data Analysis,* Vol. 39, No. 2, 153-163.

[8] Bickel, D. R. (2003). Robust and efficient estimation of the mode of continuous data: the mode as a viable measure of central tendency. *Journal of Statistical Computation and Simulation,* Vol. 73, No. 12, 899-912.

[9] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance and of the correlation between errors in the one-way classification. *The Annals of Mathematical Statistics,* Vol. 25, No. 2, 290-302.

[10] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 26, No. 2, 211-252.

[11] Broda, S., Carstensen, K. and Paolella, M. S. (2004). *Assessing and improving the performance of nearly efficient unit root tests in small samples.* Swiss Banking Institute, University of Zurich, November. Unpublished paper.

[12] Butler, R. W. and Paolella, M. S. (1999). *Saddlepoint approximations to the density and distribution of ratios of quadratic forms in normal variables with application to the sample autocorrelation function.* Department of Statistics, Colorado State University, WP 98-16. Unpublished paper.

[13] Davies, R. B. (1973). Numerical inversion of a characteristic function. *Biometrika,* Vol. 60, No. 2, 415-417.

[14] DeJong, D. N. and Whiteman, C. H. (1991). Reconsidering 'trends and random walks in macroeconomic time series'. *Journal of Monetary Economics ,* Vol. 28, No. 2, 221-254.

[15] Dejong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992). Integration versus trend stationarity in time series. *Econometrica,* Vol. 60, No. 2, 423-433.

[16] Dickey, D. A. and Fuller, W. A. (1979). Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association,* Vol. 74, No. 366, 427-431.

[17] Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica,* Vol. 49, No. 4, 1057-1072.

[18] Dufour, J. M. (1990). Exact tests and confidence sets in linear regressions with autocorrelated errors. *Econometrica,* Vol. 58, No. 2, 475-494.

[19] Elliot, G., Rothenberg, T. J. and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica,* Vol. 64, No. 4, 813-836.

[20] Fair, R. C. (1996). Computing median unbiased estimates in macroeconomic models. *Journal of Applied Econometrics*, Vol. 11, No. 4, 431-435.

[21] Farebrother, R. W. (1989). The distribution of a quadratic form in normal variables. *Journal of Applied Statistics,* Vol. 39, No. 2, 294-309.

[22] Ferguson, T. S. (1967). *Mathematical statistics: a decision theoretic approach.* California: Academic Press.

[23] Forchini, G. (2002). The exact cumulative distribution of a ratio of quadratic forms in normal variables, with application to the AR(1) model. *Econometric Theory,* Vol. 18, No. 4, 823-852.

[24] Fuller, W. A. (1996). *Introduction to statistical time series, second edition.* New York: John Wiley & Sons.

[25] Gil-Pelaez, J. (1951). Note on inversion theorem. *Biometrika,* Vol. 38, No. 3/4, 481-482.

[26] Grad, A. and Solomon, H. (1954). Distribution of quadratic forms and some applications. *Annals of Mathematical Statistics,* Vol. 26, No. 3, 464-477.

[27] Greene, W. H. (2003). *Econometric analysis, fifth edition.* New York: Prentice Hall.

[28] Grenander, U. (1965). Some direct estimates of the mode. *The Annals of Mathematical Statistics*, Vol. 36, No. 1, 131-138.

[29] Grubb, D. and Symons, J. (1987). Bias in regressions with a lagged-dependent variable. *Econometrics Theory,* Vol. 3, No. 3, 371-386.

[30] Gurland, J. (1948). Inversion formulae for the distribution of ratios. *Annals of Mathematical Statistics*, Vol. 19, No. 2, 228-237.

[31] Hamilton, J. D. (1994). *Time series analysis.* Princeton, N.J: Princeton University Press.

[32] Harvey, A. C. (1981). *The econometric analysis of time series.* Oxford: Phillip Allan

[33] Hasan, M. N. and Koenker, R. W. (1997). Robust rank tests of the unit root hypothesis. *Econometrica,* Vol. 65, No. 1, 133-161.

[34] Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *The elements of statistical learning.* New York: Springer.

[35] Hurwicz, L. (1950). *"Least-squares bias in time series", in Koopmans, T. C., statistical inference in dynamic economic models (365-383).* New York: John Wiley & Sons.

[36] Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika,* Vol. 48, No. 3/4, 419-426.

[37] Juhl, T. and Xiao, Z. (2003). Power functions and envelopes for unit root tests. *Econometric Theory,* Vol. 19, No. 2, 240-253.

[38] Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika,* Vol. 41, No. 3, 403-404.

[39] King, M. L. (1980). Robust tests for spherical symmetry and their application to least squares regression. *Annals of Statistics,* Vol. 8, No. 6, 1265-1271.

[40] Kiviet, J. F. and Phillips, G. D. A. (1992). Exact similar tests for unit roots and cointegration. *Oxford Bulletin of Economics and Statistics,* Vol. 54, No. 3, 349-367.

[41] Kiviet, J. F. and Phillips, G. D. A. (1993). Alternative bias approximations in regressions with a lagged dependent variable. *Econometric theory,* Vol. 9, No. 1, 62-80.

[42] Kiviet, J. F. and Phillips, G. D. A. (2003). *Improved coefficient and variance estimation in stable first order dynamic regression models.* Department of Quantitative Economics, University of Amsterdam, 30 June. Unpublished paper.

[43] Kiviet, J. F. and Phillips, G. D. A. (2005). Moment approximation for least-squares estimators in dynamic regression models with a unit root. *Econometrics Journal,* Vol. 8, No. 2, 115-142.

[44] Kiviet, J. F., Phillips G. D. A. and Schipp, B. (1999). Alternative bias approximations in first-order dynamic reduced form models. *Journal of Economic Dynamics and Control,* Vol. 23, No. 7, 909-928

[45] Kohn, R., Shively, T. S. and Ansley, C. F. (1993). Computing p-values for the generalized Durbin-Watson statistic and residual autocorrelations in regression. *Applied Statistics,* Vol. 42, No. 1, 249-269.

[46] Krämer, W. and Sonnberger, H. (1986). *The linear regression model under test.* Heidelberg: Physica Verlag

[47] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics,* Vol. 54, No. 1-3, 159-178.

[48] Le Breton, A. and Pham, D. T. (1989). On the bias of the least squares estimator for the first order autoregressive process. *Annals of the Institute of Statistical Mathematics,* Vol. 41, No. 2, 555-563.

[49] Lehmann, E. L. (1959). *Testing statistical hypotheses.* New York: John Wiley & Sons.

[50] Lehmann, E. L. (1983). *Theory of point estimation.* New York: John Wiley & Sons.

[51] Leybourne, S. J. and McCabe, B. P. M. (1994). A consistent test for a unit root. *Journal of Business and Economic Statistics,* Vol. 12, No. 2, 157-166.

[52] Lieberman, O. (1994). Saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables. *Journal of the American Statistical Association,* Vol. 89, No. 427, 924-928.

[53] Lopez, C., Murray, C. J. and Papell, D. (2004). *More powerful unit root tests and the PPP puzzle.* Department of Economics, University of Cincinnati, June. Unpublished paper.

[54] MacKinnon, J. G. and Smith, A. A. (1988). Approximate bias correction in econometrics. *Journal of Econometrics,* Vol. 85, No. 2, 205-230.

[55] Magnus, J. R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics.* New York: John Wiley & Sons.

[56] Marriott, F. H. C. and Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika,* Vol. 41, No. 3, 390-402.

[57] Nabeya, S. and Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Annals of Statistics,* Vol. 16, No. 1, 218-235.

[58] Nankervis, J. C. and Savin, N. E. (1988). The exact moments of the least-squares estimator for the autoregressive model: corrections and extensions. *Journal of Econometrics,* Vol. 37, No. 3, 381-388.

[59] Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics* , Vol. 10, No. 2, 139-162.

[60] Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association,* Vol. 90, No. 429, 268-281.

[61] Ng, S. and Perron, P. (2001). Lag length selection and construction of unit root tests with good size and power. *Econometrica,* Vol. 69, No. 6, 1519-1554.

[62] Orcutt, G. H. and Winokur, H. S. (1969). First order autoregression: inference, estimation, and prediction. *Econometrica,* Vol. 37, No. 1, 1-14.

[63] Paolella, M. S. (2003). Computing moments of ratios of quadratic forms in normal variables. *Computational Statistics and Data Analysis,* Vol. 42, No. 3, 313-331.

[64] Patterson, K. (2000). Finite sample bias of the least squres estimator in an AR(p) model: estimation, inference, simulation and examples. *Applied Economics,* Vol. 32, No. 15, 1993-2005.

[65] Perron, P. (1988). Trends and random walks in macroeconomic time series: further evidence from a new approach. *Journal of Economic Dynamics and Control,* Vol. 12, No. 2/3, 297-332.

[66] Perron, P. and Ng, S. (1996). Useful modification to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies,* Vol. 63, No. 3, 435-463.

[67] Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica,* Vol. 55, No. 2, 277-302.

[68] Phillips, P. C. B. (1988). Regression theory for nearly integrated time series. *Econometrica,* Vol. 56, No. 5, 1021-1043.

[69] Phillips, P. C. B. (1991). To criticize the critics: an objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics,* Vol. 6, No. 4, 333-364.

[70] Phillips, P. C. B. and Perron, P. (1988). Testing for unit roots in time series regression. *Biometrika,* Vol. 75, No. 2, 335-346.

[71] Phillips, P. C. B. and Xiao, Z. (1998). A primer on unit root testing. *Journal of Economic Surveys,* Vol. 12, No. 5, 423-469.

[72] Phillips, P. C. B., Moon, H. R. and Xiao, Z. (2001). How to estimate autoregressive roots near unity. *Journal of Time Series Analysis,* Vol. 22, No. 5, 595-612.

[73] Quenouille, M. H. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society. Series B (Methodological),* Vol. 11, No. 1, 68-84.

[74] Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika,* Vol. 43, No. 3, 353-360.

[75] Riley, K. F., Hobson, M. P. and Bence, S. J. (2002). *Mathematical methods for physics and engineering, second edition.* Cambridge: Cambridge University Press.

[76] Rothenberg, T. and Thompson, S. B. (2003). *Robust tests in the linear model.* Department of Economics, Harvard University, 2 April. Unpublished paper.

[77] Rudebusch, G. D. (1992). Trends and random walks in macroeconomic time series: a re-examination. *International Economic Review,* Vol. 33, No. 3, 661-680.

[78] Rudebusch, G. D. (1993). The uncertain unit root in real GNP. *The American Economic Review,* Vol. 83, No. 1, 264-272.

[79] Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika,* Vol. 71, No. 3, 599-607.

[80] Sargan, J. D. and Bhargava, A. (1983). Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica,* Vol. 51, No. 1, 153-174.

[81] Sawa, T. (1978). The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics,* Vol. 8, No. 2, 159-172.

[82] Scheffé, H. (1959). *The analysis of variance.* New York: John Wiley & Sons.

[83] Schmidt, P. and Phillips, P. C. B. (1992). LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics,* Vol. 54, No. 3, 257-287.

[84] Schotman, P. and van Dijk, H. K. (1991). A Bayesian analysis of the unit root in real exchange rates. *Journal of Econometrics,* Vol. 49, No. 1/2, 195-238.

[85] Shaman, P. and Stine, R. A. (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association,* Vol. 83, No. 403, 842-848.

[86] Shin, D. W. and So, B. S. (2001). Recursive mean adjustment for unit root tests. *Journal of Time Series Analysis,* Vol. 22, No. 5, 595-612.

[87] Shively, T. S., Ansley, C. F. and Kohn, R. (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,* Vol. 85, No. 411, 676-685.

[88] Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* London: Chapman & Hall.

[89] Stine, R. A. and Shaman, P. (1989). A fixed point characterization for bias of autoregressive estimators. *The Annals of Statistics,* Vol. 17, No. 3, 1275-1284.

[90] Stock, J. H. (1994). Deciding between I(1) and I(0). *Journal of Econometrics,* Vol. 63, No. 1, 105-131.

[91] Tanaka, K. (1990). Testing for a moving average unit root. *Econometric Theory,* Vol. 6, No. 4, 433-444.

[92] Tanizaki, H. (2000). Bias correction of OLSE in regression models with lagged dependent variables. *Computational Statistics and Data Analysis,* Vol. 34, No. 4, 495-511.

[93] Thompson, S. B. (2002). *Robust confidence intervals for autoregressive coefficients near one.* Department of Economics, Harvard University, 15 May. Unpublished paper.

[94] Thompson, S. B. (2003a). *Optimal versus robust inference in nearly integrated non-Gaussian models.* Department of Economics, Harvard University, 10 January. Unpublished paper.

[95] Thompson, S. B. (2003b). *Robust tests of the unit root hypothesis should not be 'modified'.* Department of Economics, Harvard University, 20 March. Unpublished paper.

[96] Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing.* London: Chapman & Hall.

[97] White, J. S. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika,* Vol. 48, No. 1/2, 85-94.

[98] Williams, D. (2001). *Weighing the odds: a course in probability and statistics.* London: Cambridge University Press.

[99] Xiao, Z. (2001a). Likelihood-based inference in trending time series with a root near unity. *Econometric Theory,* Vol. 17, No. 6, 1082-1112.

[100] Xiao, Z. (2001b). Testing the null hypothesis of stationarity against an autoregressive unit root alternative. *Journal of Time Series Analysis,* Vol. 22, No. 1, 87-105.

[101] Yule, G. U. and Kendall, M. G. (1948). *An introduction to the theory of statistics*. London: Charles Griffin & Co.