Exact Distribution Theory for the Maximum Likelihood Estimators of Local Trend Models

by

Neil Graeme Shephard

Department of Statistical and Mathematical Sciences,

London School of Economics and Political Science

October 1989

Submitted to the University of London for the Degree of Doctor of Philosophy.

Abstract

This thesis has two distinct parts. The second and third chapters concern the theory and practical implementation of computing the value of the (possibly multivariate) distribution function from the known characteristic function. The remaining four chapters consist of a study of the distributional behaviour of the maximum likelihood estimator of some local trend models.

The motivation of the work is the paucity of our knowledge of the behaviour of the maximum likelihood estimator of local trend models. These types of models are being increasingly used in the social sciences, but little is actually known about their theoretical behaviour, especially when we have only small sample sizes.

TABLE OF CONTENTS.

List of Tables

List of Figures

Preface

It is a pleasure to express my gratitude to my supervisor Professor A.C.Harvey for his encouragement throughout the length of this study. I would also like to thank Dr. M.Knott whose continuous criticisms of my ugly proofs were invaluable to me. Without his undying patience this work would be even more difficult to read than it is. The comments and encouragement of Dr. R.W.Farebrother on some earlier drafts of this work have also been useful to me; as was the guidance I received from Dr. Katsuto Tanaka in the very early stages of the research reported here. Finally, I thank Mrs. Dagmar Schmacher for help in typing some of chapter seven of this report, and Ms. Gillian Dunn for showing me how effective the use of commas can be in the construction of sentences!

I also acknowledge the financial support I received from the ESRC in the first year of my study.

N.G.Shephard, North Wold, Norfolk. September 1989.

To My Grandparents

Lydia and Frank

Chapter One

Introduction

Summary The idea of using local, rather than global, trend models is introduced. It is noted that deterministic components can result if various signal–noise ratios are zero. The two major sections of the thesis are summarized. The first, which consists of the second and third chapters, will be concerned with characteristic functions. Chapters four, five, six and seven make up the second section and deal with the probabilities of estimating deterministic components.

Key Words LOCAL TREND MODELS, SIGNAL–NOISE RATIOS, LOCAL LINEAR TREND MODEL, LOCAL LEVEL MODEL, CHARACTERISTIC FUNCTIONS.

(1.1) Motivation

In the classical decomposition of non–seasonal time series data the observations are regarded as the sum of two components; a trend and an irregular. This can be written as

$$y_t = \mu_t + \epsilon_t, \tag{1.1}$$

where y_t is the observed series, μ_t is interpreted as a trend, and ϵ_t is thought of as an additive irregular component.

Traditionally, global trend models have been specified for μ_t by writing, in the linear case,

$$\mu_t = \mu + \beta.t, \tag{1.2}$$

which allows the model to be consistently estimated by ordinary least squares, under weak regularity on the irregular term. Unfortunately, imposing a deterministic trend in this way, can have highly misleading consequences since there are many series for which it is simply not appropriate – see Nelson and Kang(1984), Granger and Newbold(1974) and Harvey(1989).

A more flexible model for the trend allows the level and slope to evolve over time. Such models are called local or stochastic trends. In the linear case μ_t would be given by,

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \tag{1.3a}$$
$$\beta_t = \quad + \beta_{t-1} + \zeta_t, \tag{1.3b}$$

where (ϵ_t), (η_t), and (ζ_t) are assumed to be uncorrelated, zero mean, white noise processes with variances σ^2, $q\sigma^2$ and $p\sigma^2$ respectively. We will call q and p signal–noise ratios. The practical implication of such a formulation is that forecasts depend more on the more recent observations.

When q and p are both zero (1.3) collapses into equation (1.2) and we say that the trend has two (μ and β) deterministic components. We can deduce from this that (1.2) is encompassed within the more general trend model (1.3).

A similar story can be told about the simpler process, the local level model, given by (1.1) and (1.4),

$$\mu_t = \mu_{t-1} + \eta_t. \tag{1.4}$$

This model is a classic model in time series analysis as it provides a rationale for the exponentially weighted moving average forecasting scheme (c.f. Muth(1960)). When q is zero this model collapses to a constant observed with irregular error. Hence, the model given in equations (1.1) and (1.4) has a deterministic component if q is zero.

Little has been written about the possibility that deterministic components are estimated in these types of models. Recently, Newbold(1988) has drawn attention to this problem. Harvey(1989, chapter 4) has attempted to link up this possibility, in the local level model, with that of estimating noninvertible moving average process of order one.

The report given in this thesis is the first full scale analysis of this problem.

(1.2) The Structure of the Thesis

This work has two major sections. The first, which comprises chapters two and three, is concerned with developing algorithms for computing the (possibly multivariate) distribution function from the characteristic function. The second, made up of chapters four, five , six and seven, looks at the probability of estimating deterministic components in various local trend models.

Chapter two provides a unified framework for the theoretical study of inverting known characteristic functions to compute distribution functions. This provides a comparatively easy proof of Gil–Pelaez's(1951) and Gurland's(1948) univariate result. Furthermore, it will give a basis of a simple derivation of a rigorously justified multivariate inversion formula.

The third chapter investigates, in detail, the numerical implementation of the theories discussed in chapter two. An algorithm for a positive variable is given, paying particular attention to a square root problem which frequently arises when dealing with Fredholm determinants. Davies'(1973) algorithm for an unconstrained, univariate random variable is developed in a rigorous fashion. Then a bivariate inversion is given, which is a generalization of a result first suggested by Shively(1986). Finally, a multivariate inversion is derived.

The second section of this thesis starts, in chapter four, with a literature review which brings out the connection between the published results on estimated noninvertible moving averages and the local trend models we analyze here. This chapter also presents a valid Edgeworth expansion for the maximum likelihood estimator of q in the local level model, allowing us to understand the process which occurs as we approach a breakdown in the central limit theorem.

The fifth chapter is the pivotal piece of the work of this thesis. It analyses the possibility of estimating a deterministic component in the local level model. It is shown that this probability is highly sensitive to the way we initialize the nonstationary state, μ_t.

Explanatory variables are introduced into the measurement equation, (1.1), of the local level model in the penultimate chapter. The analysis, which is similar to that given by Sargan and Bhargava(1983b), centres around the use of time trends as explanatory variables. It is shown that the introduction of these types of variables will have a dramatic effect on the possibility of estimating a deterministic component if the regression coefficients are viewed as fixed. On the other hand, if they are deterministic, in the sense of Wold, their effect will be mild.

In the final chapter a local linear trend model is analyzed. This is of particular interest to us, as the occurrence of two deterministic components is equivalent to estimating two unit roots in a moving average process of order two which is a theoretical problem which has previously escaped solution.

Chapter Two

From Characteristic Function to Distribution Function:– The Theory

Summary A unified framework is established for the study of the computation of the distribution function from the characteristic function. A new approach to the proof of Gurland's and Gil–Pelaez's univariate inversion theorem is suggested. A multivariate inversion theorem is then derived using this technique.

Key Words CHARACTERISTIC FUNCTIONS, INVERSION THEOREMS.

(2.1) Introduction.

It is often easier to manipulate characteristic functions than distribution functions. If the characteristic function is known then we can compute the distribution function by using an inversion theorem. This chapter reviews the theoretical basis of inverting characteristic functions, presenting the work within a unified framework based on the well known results of Fourier analysis. Only Theorem 2.5 (and hence Theorem 2.7) of this work is a new result. The proof of Theorem 2.3, however, (and hence Theorem 2.4) is substantially different from the one given in the literature.

Inverting the characteristic function to find the distribution function has a long history, c.f. Lukacs(1970, chapter 2). Lévy's(1925) result is the most famous of these theorems, although in this context its practical use is limited to some special cases unless the random variable of interest is always strictly positive (see Bohmann (1970, pg 238) and Knott(1974, pg 431)). Gurland's(1948) paper gave a more useful inversion theorem, but it is the paper of Gil–Pelaez(1951) which has provided the basis of most of the distributional work completed in this field (c.f. Davies(1973, 1980) and Imhof(1961)). Gurland's and Gil–Pelaez's results are almost identical. Gurland's is based on the principal value of a Lebesgue integral, while Gil–Pelaez removes the need for principal values by using a Riemann integral.

Recently Shively(1986) has generalized Gil–Pelaez's work on Riemann integrals to provide a bivariate inversion theorem, while Shively(1988) used this expression to tabulate critical values of a statistic proposed by Watson and Engle(1985) for testing the stability of the parameters in a regression model. Only Gurland(1948) has attempted to provide a multivariate inversion theorem. Our results are slightly different from those obtained by Gurland. In this chapter we develop a framework for the analysis of univariate inversion theorems which offers an easy multivariate generalization. The result, which is given in Theorem 2.5, is an expression which involves terms which are straightforward to compute. The advent of the wide availability of powerful computers will mean that such multivariate inversions will join univariate inversions, (c.f. Farebrother(1980, 1981) and Davies(1980)), in being a tool which theoretical statisticians can employ routinely when tackling difficult distributional problems.

(2.2) The Univariate Inversion Theorem.

Bohmann(1961) provided an elegant and unified framework for the study of inversion theorems based on the results of Fourier analysis. His work, which relies on the properties of convolutions, is in keeping with the discussion of characteristic functions given by Feller(1971, chapter XV). Although the subject matter of this chapter is rather different from that considered by Bohmann, our work will remain firmly within the framework he suggested.

To establish our notation we introduce some definitions. Let F denote the distribution function of interest. Suppose its corresponding density, f, is integrable in the Lebesgue sense (written $f \in L$) and that its characteristic function is defined as

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx. \tag{2.1}$$

We suppose that φ is known and we wish to compute F directly from it. The basic result we will use to perform this calculation is the Fourier inversion theorem.

<u>Theorem 2.1 (Fourier inversion theorem)</u> Suppose g, $\varphi \in L$, and

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, g(x)\, dx, \text{ then} \tag{2.2}$$

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\, \varphi(t)\, dt, \text{ everywhere.} \tag{2.3}$$

<u>Proof.</u> C.f. Hewitt and Stromberg(1965, pg 409).

The other result which will be central to our development concerns convolutions and is given below.

<u>Theorem 2.2</u> If g, h ϵ L,

$$g^*h(x) = \int_{-\infty}^{\infty} g(x-y)h(y)\, dy, \tag{2.4}$$

and g and h have Fourier transforms ϕ and ψ, then the Fourier transform of $g^*h(x)$ is

$$\xi(t) = \phi(t)\, \psi(t). \tag{2.5}$$

<u>Proof.</u> C.f. Hewitt and Stromberg(1965) theorem (21.31) and theorem (21.41).

g * h is called the convolution of g and h. These two results will be nearly sufficient to enable us to develop all the results we want in this chapter. A simple application of the Fourier inversion theorem gives us the following well known result.

Corollary 2.1 If f, $\varphi \in$ L, then

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)\, dt. \tag{2.6}$$

Proof. This follows trivially from theorem 2.1.

Equally, following, for example, Feller, we can convolute F with the uniform distribution on [–h,h] and then use corollary 2.1 to produce Lévy's important theorem.

Corollary 2.2 (Lévy(1925)) If f, $\varphi \in$ L, then

$$\frac{F(x+h)-F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin ht}{ht} e^{-itx} \varphi(t)\, dt. \tag{2.7}$$

Proof. By corollary 2.1, as left hand side is a density function.

This corollary has been of use to statisticians working in many fields, as it allows the derivation of an algorithm for the calculation of the distribution function from the characteristic function when the random variable of interest is strictly positive, c.f. Bohmann(1970, pg 238) and independently Knott(1974, pg 431). However, as Gil–Pelaez(1951) noted, when it is not positive this expression cannot be used for this purpose. As a result many writers have abandoned the idea of using convolutions. Gil–Pelaez employed the notion of a Riemann integral (see the proof in Kendall, Stuart and Ord(1987, pg 120–21)), while Gurland used a similar idea, but his proof involved the manipulation of principal values of Lebesgue integrals. Theorem 2.3 shows that under weak regularity (which will be relaxed to some extent in theorem 2.4) this diversion was unnecessary and so allows us to get away from a type of derivation which "...detracts from the logical structure of the theory."(Feller(1971, pg 511)).

Theorem 2.3 If f, $\varphi \in L$, then under the assumption of the existence of a mean for the random variable of interest

$$F(x) = \frac{1}{2} - \frac{1}{2\pi} \int_0^\infty \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-itx}}{it}\right] dt \tag{2.8}$$

where $\underset{t}{\Delta}\eta(t) = \eta(t) + \eta(-t)$.

Proof. The function $h(y) = \text{sign}(y)$, $y \in [-h,h]$, has the transform

$$\varphi_h(t) = \int_{-h}^{h} e^{ity}\text{sign}(y)\, dy = \frac{2(\sin ht - 1)}{it} \tag{2.9}$$

which is bounded for all t.

The convolution, written $u_h(x)$, of $h(y)$ with the continuous density $f(x)$ is $2F(x) - F(x+h) - F(x-h)$ which although bounded, is not integrable as $h\to\infty$. The convolution has the transform $2\varphi(t)(\sin ht - 1)/it$, which is bounded by the existence of a mean and is integrable as $\varphi \in L$. Hence, for fixed h we can use the inversion theorem to give the equality

$$\frac{2}{2\pi}\int_{-\infty}^{\infty} (\sin ht - 1)\left[\frac{\varphi(t)e^{-ixt}}{it}\right] dt$$

$$= \frac{2}{2\pi}\int_0^\infty (\sin ht - 1) \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-ixt}}{it}\right] dt = u_h. \tag{2.10}$$

When $h\to\infty$ the left hand side of the (2.10) can be manipulated, using the Riemann–Lebesgue theorem (c.f. Feller(1971), pg 513) and so reduces to

$$\frac{-2}{2\pi}\int_0^\infty \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-itx}}{it}\right] dt = 2F(x) - 1 \qquad \square.$$

The requirement that $\varphi \in L$ can be removed by using an additional convolution to improve the behaviour of the tails of the integrand. In the next theorem we do this by employing Fejér's kernel (there are, of course, many other kernels we could have used, eg. Abel's or Gauss's).

Theorem 2.4 If $f \in L$, then under the assumption of the existence of a mean

$$F(x) = \frac{1}{2} - \frac{1}{2\pi} \lim_{n\to\infty} \int_0^n \left[1 - \frac{t}{n}\right] \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-itx}}{it}\right] dt. \tag{2.11}$$

Proof. Given in the appendix.

(2.3) The Multivariate Inversion Theorem.

Suppose we now become interested in the multivariate generalizations of the above theorems. We suppose $f \in L^p$ and that

$$\varphi(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{it'x} f(x)\, dx \tag{2.12}$$

where $x=(x_1,\dots,x_p)'$ and $t=(t_1,\dots,t_p)'$ where p is some positive integer. It is well known that the Fourier inversion theorem and convolution theory go through to the multivariate case, so allowing us trivial proofs of the following well known corollaries.

Corollary 2.3 (c.f. Cramér(1946, eqn. 10.6.3)) If f, $\varphi \in L^p$, then

$$f(x) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-it'x} \varphi(t)\, dt. \tag{2.13}$$

Proof. The proof follows using a generalization of the proof of corollary 2.1.

Corollary 2.4 (c.f. Cramér(1946, eqn. 10.6.2)) If f, $\varphi \in L^p$, and we define an interval R by the inequalities $x_j - h < X_j < x_j + h$ for $j=1,\dots,p$, then

$$\frac{Pr(R)}{(2h)^p} = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{j=1}^{p} \left[\frac{\sin h.t_j}{h.t_j}\right] e^{-it'x} \varphi(t)\, dt. \tag{2.14}$$

Proof. The proof follows using a generalization of the proof of corollary 2.2.

Little has been written about the theory of computing the distribution function by inverting multivariate characteristic functions. Gurland(1948) extended his univariate procedure to p dimensions using the notion of a principal value of a Lebesgue integral. His result is slightly different from the result we present here. Further, his proof is much more complicated than the one given here. Recently Shively(1986, 1988) has extended Gil–Pelaez's(1951) result to two variables in order to calculate some critical values of a test statistic which arises in econometrics. He used Riemann integrals in his derivation and so his proof does not conform with modern work on integrals, as well as being rather obscure.

As we have seen in section two there is no need to use these techniques. The advantage of using a convolution approach is that the multivariate generalisations now follow using standard results.

__Theorem 2.5__ If f, $\varphi \in L^p$, then under the assumption of the existence of a mean, the following equality holds

$$\frac{(-2)^p}{(2\pi)^p}\int_0^\infty \cdots \int_0^\infty \underset{t_1}{\Delta}\,\underset{t_2}{\Delta}\cdots\underset{t_p}{\Delta}\left[\frac{\varphi(t)e^{-ix't}}{it_1 it_2 \ldots it_p}\right]dt = u^+(x) \tag{2.15}$$

where $u^+(x) = 2^p F(x_1,\ldots,x_p) - 2^{p-1}\Big[F(x_2,x_3,\ldots,x_p)+\ldots+F(x_1,\ldots,x_{p-2},x_{p-1})\Big]$

$+\ 2^{p-2}\Big[F(x_3,x_4,\ldots,x_p)+\ldots+F(x_1,\ldots,x_{p-3},x_{p-2})\Big] + \ldots + (-1)^p.$

Proof. Define

$$h(y_1,\dots,y_p) = \prod_{j=1}^{p} \operatorname{sign}(y_j)\ ,\ \text{where } y_j \epsilon [-h,h] \tag{2.16}$$

so it has the transform

$$\varphi_h(t) = \prod_{j=1}^{p} \frac{2(\sin t_j h_j \ -1)}{i t_j}. \tag{2.17}$$

Consider the convolution of the density function, f, with $h(y_1,\dots,y_p)$, which we will write as u_h

$$u_h(x) = \int_{-h}^{h} \cdots \int_{-h}^{h} f(x_1-y_1,x_2-y_2,\dots,x_p-y_p)\ h(y)\ dy, \tag{2.18}$$

$$= \sum_{a_1,a_2,\dots,a_p} \Delta_{ha_1,ha_2,\dots,ha_p} \tag{2.19}$$

where the summation is taken over all the values $(a_j=\pm 1)$, with

$$\Delta_{h,h,\dots,h} = \sum_{b_1,b_2,\dots,b_p} (-1)^{b_1+\dots+b_p} . F(x_1-hb_1,\dots,x_p-hb_p), \tag{2.20}$$

where the summation is taken over the binary numbers $(b_j=0,1)$ and F denotes a generic distribution function. By the (multivariate) Fourier inversion theorem,

$$\frac{2^p}{(2\pi)^p}\int_{-\infty}^{\infty}\!\!\cdots\int_{-\infty}^{\infty}\varphi(t)e^{-ix't}\prod_{j=1}^{p}\left[\frac{\sin h_jt_j-1}{it_j}\right]dt$$

$$=\frac{2^p}{(2\pi)^p}\int_0^{\infty}\!\!\cdots\int_0^{\infty}\underset{t_1}{\Delta}\,\underset{t_2}{\Delta}\cdots\underset{t_p}{\Delta}\left[\frac{\varphi(t)e^{-ix't}}{it_1it_2\ldots it_p}\right](\sin t_jh_j-1)\,dt=u_h(x).$$

(2.21)

Allowing $h\to\infty$ the left hand side can be simplified by exploiting the Riemann–Lebesgue theorem. The result is that we have

$$\frac{(-2)^p}{(2\pi)^p}\int_0^{\infty}\!\!\cdots\int_0^{\infty}\underset{t_1}{\Delta}\,\underset{t_2}{\Delta}\cdots\underset{t_p}{\Delta}\left[\frac{\varphi(t)e^{-ix't}}{it_1it_2\ldots it_p}\right]dt=u^+(x)\square. \qquad \textbf{(2.22)}$$

<u>**Theorem 2.6**</u> Under the existence of a mean the following equality holds.

$$\underset{t_1}{\Delta}\cdots\underset{t_p}{\Delta}\left[\frac{\varphi(t)e^{-ix't}}{it_1it_2\ldots it_p}\right]=2^{2p-1}i^{p-1}\underset{t_2}{\Delta}\cdots\underset{t_p}{\Delta}\,\mathrm{Im}\left[\frac{\varphi(t)e^{-ix't}}{t_1\,t_2\ldots t_p}\right],\ \text{if p is odd}$$

$$=2^{2p-1}i^{p}\underset{t_2}{\Delta}\cdots\underset{t_p}{\Delta}\,\mathrm{Re}\left[\frac{\varphi(t)e^{-ix't}}{t_1\,t_2\ldots t_p}\right],\ \text{if p is even.}$$

(2.23)

<u>Proof.</u> Trivial.

<u>Examples</u> <u>p=2</u>

$$\frac{2^2}{(2\pi)^2}\int_0^\infty\int_0^\infty \underset{t_1}{\Delta}\underset{t_2}{\Delta}\left[\frac{\varphi(t)e^{-ix't}}{it_1\ it_2}\right] dt_1\, dt_2$$

$$= \frac{2^5}{(2\pi)^2}\int_0^\infty\int_0^\infty \underset{t_2}{\Delta}\ \mathrm{Re}\left[\frac{\varphi(t)e^{-ix't}}{t_1\ \ t_2}\right] dt_1\, dt_2$$

$$= 4F(x_1,x_2) - 2[F(x_1) + F(x_2)] + 1 \tag{2.24}$$

<u>p=3</u>

$$\frac{-2^3}{(2\pi)^3}\int_0^\infty\int_0^\infty\int_0^\infty \underset{t_1}{\Delta}\ \underset{t_2}{\Delta}\ \underset{t_3}{\Delta}\left[\frac{\varphi(t)e^{-ix't}}{it_1\ it_2\ \ it_p}\right] dt_1\, dt_2\, dt_3$$

$$= \frac{2^7}{(2\pi)^3}\int_0^\infty\int_0^\infty\int_0^\infty \underset{t_2}{\Delta}\ \underset{t_3}{\Delta}\ \mathrm{Im}\left[\frac{\varphi(t)e^{-ix't}}{t_1\ \ t_2\ \ \ t_p}\right] dt_1\, dt_2\, dt_3$$

$$= 8F(x_1,x_2,x_3) - 4[F(x_1,x_2)+F(x_1,x_3)+F(x_2,x_3)]$$

$$+\ 2[F(x_1)+F(x_2)+F(x_3)] - 1 \tag{2.25}$$

p=4

$$\frac{2^4}{(2\pi)^4}\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty \underset{t_1}{\Delta}\,\underset{t_2}{\Delta}\,\underset{t_3}{\Delta}\,\underset{t_4}{\Delta}\left[\frac{\varphi(t)e^{-ix't}}{it_1it_2it_3it_4}\right]dt_1\,dt_2\,dt_3\,dt_4$$

$$=\frac{2^9}{(2\pi)^4}\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty \underset{t_2}{\Delta}\,\underset{t_3}{\Delta}\,\underset{t_4}{\Delta}\,\mathrm{Re}\left[\frac{\varphi(t)e^{-ix't}}{t_1\,t_2\,t_3\,t_4}\right]dt_1\,dt_2\,dt_3\,dt_4$$

$$= 16F(x_1,x_2,x_3,x_4) - 8[F(x_1,x_2,x_3) + F(x_1,x_2,x_4) + F(x_1,x_3,x_4) + F(x_2,x_3,x_4)]$$

$$+ 4[F(x_1,x_2) + F(x_1,x_3) + F(x_1,x_4) + F(x_2,x_3) + F(x_2,x_4) + F(x_3,x_4)]$$

$$- 2[F(x_1) + F(x_2) + F(x_3) + F(x_4)] + 1 \qquad (2.26)$$

The assumption of the integrability of φ can be removed by following the approach given in the proof of theorem 2.4. This yields a straightforward, although rather ugly, proof of theorem 2.7.

Theorem 2.7 If $f \in L^p$, then under the assumption of the existence of a mean, the following equality holds

$$\frac{(-2)^p}{(2\pi)^p} \lim_{n\to\infty} \int_0^n \cdots \int_0^n \prod_{j=1}^{p} \left[1 - \frac{t_j}{n}\right] \underset{t_1}{\Delta} \underset{t_2}{\Delta} \cdots \underset{t_p}{\Delta} \left[\frac{\varphi(t)e^{-ix't}}{it_1 it_2 \cdots it_p}\right] dt = u^+(x)$$

where $u^+(x) = 2^p F(x_1,\ldots,x_p) - 2^{p-1}\left[F(x_2,x_3,\ldots,x_p) + \ldots + F(x_1,\ldots,x_{p-2},x_{p-1})\right]$

$$+ 2^{p-2}\left[F(x_3,x_4,\ldots,x_p) + \ldots + F(x_1,\ldots,x_{p-3},x_{p-2})\right] + \ldots + (-1)^p. \tag{2.27}$$

Mathematical Appendix.

Proof of Theorem 2.4 Recall the proof of theorem 2.3. To improve the behaviour of the tails of $u_h(x)$ we convolute it with Fejér's kernel

$$k_n(x) = \frac{1}{2\pi} n. \left[\frac{\sin(nx/2)}{(nx/2)} \right]^2. \tag{A2.1}$$

It is important to note that

$$\text{(i)} \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} k_n(x)\, dx = 1, \tag{A2.2}$$

(ii) the transform of Fejér's kernel is

$$\varphi_n(t) = \frac{1}{2\pi} \left[1 - \frac{|t|}{n} \right] I(|t|<n) \tag{A2.3}$$

where I(.) is an indicator function.

For fixed n the convolution of $u_h(x)$ with $k_n(x)$, and its transform, are integrable, as $\varphi(t)$ is bounded. Hence we can use the inversion theorem to give the equality

$$\frac{2}{2\pi}\int_{-\infty}^{\infty} \left[1 - \frac{|t|}{n}\right] I(|t|<n)(\sin ht - 1)\left[\frac{\varphi(t)e^{-ixt}}{it}\right] dt$$

$$= \frac{2}{2\pi}\int_{0}^{\infty} \left[1 - \frac{|t|}{n}\right] I(t\leq n)(\sin ht - 1) \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-ixt}}{it}\right] dt$$

$$= u_h * k_n(x). \tag{A2.4}$$

When $h \to \infty$ the left hand side of the (A2.4) can be manipulated using the Riemann–Lebesgue theorem (c.f. Feller(1971), pg 513) because for fixed n

$$\left[1-\frac{|t|}{n}\right] I(t \le n) \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-ixt}}{it}\right] \tag{A2.5}$$

is integrable, as $\underset{t}{\Delta}(\varphi(t)e^{-ixt})/it$ is bounded due to the assumption of the existence of a mean for the density f(x). Hence the left hand side of (A2.4) reduces to

$$\frac{-2}{2\pi}\int_0^\infty \left[1-\frac{t}{n}\right] I(t \le n) \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-ixt}}{it}\right] dt. \tag{A2.6}$$

Now think of the right hand side of (A2.4) as $h \to \infty$. Remember it is

$$\int_{-\infty}^\infty \left[2F(x-y)-F(x-y-h)-F(x-y+h)\right] k_n(y)\, dy \tag{A2.7}$$

so using the boundedness of $2F(x-y)-F(x-y-h)-F(x-y+h)$ and the integrability of $k_n(y)$ we can employ the Lebesgue dominated convergence theorem (see Hewitt and Stromberg(1965, section 12.34)) to imply

$$\lim_{h \to \infty} u_h * k_n(x) = u^+ * k_n(x) \tag{A2.8}$$

where $u^+(x) = 2F(x)-1$. Thus,

$$\frac{-2}{2\pi}\int_0^\infty \left[1-\frac{t}{n}\right] I(t \le n) \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-ixt}}{it}\right] dt = u^+ * k_n(x). \tag{A2.9}$$

We now become interested in the behaviour of this equality as $n \to \infty$. Proposition 2.1 deals with the right hand side. This is a well known theory (cf. Hewitt and Stromberg(1965, section 21.42)) although we are using slightly different assumptions from those usually employed and so we include its proof in this appendix. Hence

$$\frac{-2}{2\pi} \lim_{n \to \infty} \int_0^n \left[1-\frac{t}{n}\right] \underset{t}{\Delta} \left[\frac{\varphi(t)e^{-ixt}}{it}\right] dt = 2F(x) - 1 \; \square.$$

<u>Proposition 2.1</u> $$\lim_{n\to\infty} u^+ * k_n(x) = u^+(x)$$

<u>Proof.</u> $$u^+ * k_n(x) - u^+(x) = \int_{\mathbb{R}} (u^+(x-t) - u^+(x))k_n(t)\, dt$$

$$= \int_{\mathbb{R}^+} \beta(t)k_n(t)\, dt \tag{A2.10}$$

where $\beta(t)=u^+(x+t)+u^+(x-t)-2u^+(t)$, which is bounded but not integrable. Choose $\epsilon_1>0$, then using the continuity of the distribution function we can choose $\epsilon>0$ so small that $|\beta(t)|<\epsilon_1$ $\forall t\epsilon[0,\epsilon)$. So write $k(t) = 4/(1+t^2)$ and note that $k_n(t)\leq nu(nt)$, then

$$\left| \int_{\mathbb{R}^+} k_n(t)\beta(t)\, dt \right| \leq \left| \int_0^{\epsilon} k_n(t)\beta(t)\, dt \right| + \left| \int_{\epsilon}^{\infty} k_n(t)\beta(t)\, dt \right|$$

$$\leq \epsilon_1 + \sup_t |\beta(t)| \left| \int_{\epsilon}^{\infty} n\, k(nt)\, dt \right|$$

$$\leq \epsilon_1 + \frac{4}{\epsilon n} \sup_t |\beta(t)| \tag{A2.11}$$

so for $n>4\sup_t |\beta(t)|/\epsilon\epsilon_1$ we have the required result by choosing ϵ_1 sufficiently small □.

Chapter Three

From Characteristic Function to Distribution Function:– The Algorithms

Summary. Algorithms for the implementation of the theorems discussed in chapter two are derived in detail. Attention is concentrated on the errors induced by the use of numerical integration rules. Algorithms for inverting characteristic functions are outlined for the following :– (i) a positive random variable, (ii) an unconstrained variable, (iii) a quadratic form in normal variables, (iv) two unconstrained variables, (v) two quadratic forms in normal variables, (vi) p unconstrained variables, and (vii) p quadratic forms in normal variables.

Key Words. CHARACTERISTIC FUNCTIONS, INVERSION THEOREMS, FREDHOLM DETERMINANTS, DAVIES' ALGORITHM, RIEMANN SUMS, TRAPEZOIDAL RULES, IMHOF PROCEDURE, SERIAL CORRELATION COEFFICIENTS, EIGENVALUE–FREE.

(3.1) Introduction

In chapter two we reviewed the theory of inverting characteristic functions to compute the distribution function. Typically this inversion will be carried out numerically, resulting in the employment of some numerical integration routine.

In this chapter we see how various writers have attempted to implement the theories discussed above. Efficient algorithms have employed the structure of particular problems of interest and so we will have to describe a variety of techniques here. In section (3.2) we look at the case where the variables are strictly positive, using the work of Bohmann and Knott. Section (3.3) looks at the most important of the inversion algorithm, that of Davies, which will be presented in general. In sections (3.4) and (3.5) we see how this algorithm is changed when we work with a variable which is a quadratic form in normal variables.

A bivariate inversion formula will then be discussed at length in section (3.6). This development will then be used in section (3.7) where we discuss two quadratic forms in normal variables. Two examples of the use of the bivariate inversion will also be given in this section to illustrate the feasibility of the method. Finally, in section (3.8) we will generalize Davies' algorithm to cope with p variables.

(3.2) A Positive Random Variable

If a random variable is strictly positive then we can employ Lévy's(1925) theory to invert the characteristic function. An algorithm for performing this inversion has been developed, independently, by Bohmann(1970) and Knott(1974) who used Fourier analysis to show the following result held.

<u>Theorem 3.1</u> If $F(0) = 0$, φ, $f \in L$, then

$$F(h) \leq \frac{h}{\theta} + \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{\sin\ (hk\pi/\theta)}{k} \mathrm{Re}(\varphi(k\pi/\theta)) \leq F(h) + (1 - F(\theta)) \tag{3.1}$$

where Re(.) denotes the real part of a complex function.

<u>Proof.</u> Given in Bohmann(1970).

To implement this algorithm we need to produce a decision rule for the choice of θ and a cutoff point in the infinite sum after M terms. Knott(1974) did not offer a general procedure for achieving this, although his discussion on page 433 is useful in this context. We can produce an automatic criteria for the selection of θ and suggest a sensible way of choosing M. If the moment generating function of the random variable, written X, exists (if φ is analytic then it will exist, c.f. Lukacs(1970, pg 196)) and is written $M_X(t)$, then for small u>0 we have (c.f. Feller(1966, pg 525) and Davies(1973))

$$E\Big[I(X>x) - \exp(u(X-x))\Big] = 1 - F(x) - M_X(u)\exp(-ux) \leq 0 \tag{3.2}$$

where I(.) is an indicator function. Hence,

$$1 - F(x) \leq M_X(u)\exp(-ux). \tag{3.3}$$

Write $\psi(u) = \log M_X(u)$ and choose $x = \psi'(u) = d\psi(u)/du$, then we have that

$$1 - F(\ \psi'(u)) \leq M_X(u).\exp\ (-u\ \psi'(u)) = \exp(-u\varphi'(u) + \varphi(u)) \tag{3.4}$$

and so we can find u such that $1 - F(\psi'(u)\ ) < \epsilon$, where ϵ has to be selected.

It seems difficult to control the error induced by truncating the sum after M terms, unless we know more about the analytic structure of the characteristic function of interest. However, we can introduce sensible decision rules so that it is unlikely that we will make a large mistake. If the characteristic function is expensive to compute then it is worthwhile spending considerable efforts in studying the evolution of the sum as M increases in order to deduce a sensible stopping point. However, usually $\varphi(t)$ is quite cheap to evaluate, in which case a semi–automatic criterion could be used, eg. stop after N occurrences of the terms in the sum being less than η in absolute value. Typically we might take N to be 20 and η as 1.0×10^{-8}.

Finally, before we leave this section, we must introduce a problem which is sometimes encountered when computing the real part of a characteristic function. Frequently, analytic characteristic functions, such as those derived from Fredholm determinants (see chapter six of this thesis), involve the square root of a complex number. For example, the characteristic function of the large sample Cramér–von Mises statistic (c.f. Anderson and Darling(1952)) is

$$\varphi(t) = \sqrt{\frac{\sqrt{2it}}{\sin\sqrt{2it}}}\ . \tag{3.5}$$

This square root is not immediately defined. To see this write a complex number z as $a + b.i$, where $i = \sqrt{-1}$, then we can reexpress z as

$$z = r\{\cos\,\theta + i.\sin\,\theta\}, \qquad \text{where } r = \sqrt{a^2 + b^2} \tag{3.6}$$
$$\text{and } \theta = \text{arc cos } a/r, \text{ where } \theta \in (-\pi,\ \pi),$$

$$\text{so } \sqrt{z} = \sqrt{r}\left[\ \cos\{\theta/2 + k\pi\} + i.\sin\{\theta/2 + k\pi\}\right], \text{ where } k = 0,\ 1. \tag{3.7}$$

Hence $\sqrt{z}$ has two distinct values. Anderson and Darling(1952) overcame the problem of the non–uniqueness of the square root in (3.5) by taking it to be real and positive when the characteristic function is real and positive. Recently, Perron(1989) has proposed an algorithm for the correct selection of k. Our alternative to this, which was developed independently from Perron's, is similar to this technique.

Computers will choose k such that the real part of $\sqrt{z}$ is non–negative, which means the resulting imaginary part may be positive or negative. This automatic choice of k will sometimes give us the wrong root because the real part of the characteristic function can be negative. We will need to use the properties of the characteristic function to ensure that the correct k, and hence the sign of the real part, is selected. This means that in practice we will have to alter the sign of the computed function for some values of t. Although we can suggest an algorithm for achieving this sign change, we will see latter that it will not guarantee that the sign will be allocated correctly.

The algorithm we suggest is motivated by two properties of characteristic functions. Firstly, characteristic functions are continuous (c.f. Grimmett and Stirzaker(1982, pg 101)), which of course means we must choose k so that both the real and imaginary parts of φ do not exhibit discontinuous jumps. Secondly, characteristic functions are always unity at the origin.

Suppose we write $\hat{\varphi}(t_j)$ as the value of $\varphi(t_j)$ returned by the computer at the value t_j, where $t_0 \equiv 0 < t_1 < \ldots < t_M$. Then, by noting $\hat{\varphi}(t_0) = \varphi(t_0) = 1$, we can use the following check for $j = 1,\ldots,M$ to change the sign of $\hat{\varphi}(t_j)$. If we write

$c_0 = 0$ and $s_0 = 1$, then if

$$|\mathrm{Im}\{\hat{\varphi}(t_j)\} + \mathrm{Im}\{\hat{\varphi}(t_{j-1})\}| < |\mathrm{Im}\{\hat{\varphi}(t_j)\} - \mathrm{Im}\{\hat{\varphi}(t_{j-1})\}|, \quad (j = 1, \ldots, M), \tag{3.8}$$

then put

$$c_j = 1, \text{ if } c_{j-1} = 0, \tag{3.9a}$$

$$= 0, \text{ if } c_{j-1} = 1, \tag{3.9b}$$

and

$$s_j = -s_{j-1}, \text{ if } c_j = 1. \tag{3.9c}$$

Otherwise we let $s_j = s_{j-1}$. Then we write

$$\varphi(t_j) = s_j . \hat{\varphi}(t_j). \tag{3.10}$$

(3.8) will be satisfied if there is a sign change in the imaginary component of the characteristic function.

There are three possibilities of serious errors using this technique. The first, and the most important, is that the characteristic function is moving so fast that there are sign changes which we miss completely. This should not happen unless we choose the value of θ to be very small. In any case, if there is an error it will occur when k is small (as $\varphi(t)$ moves quickest when t is small), so it is sensible to inspect a printout of the first few terms of the series. If it looks like there may be problems θ should be increased (this means the procedure will be slightly more expensive).

The second possible error is easily avoidable by using a numerical check on the value of the imaginary part when inequality (3.8) is satisfied. This check is necessary because the imaginary part of $\hat{\varphi}(t_j)$ changes sign when it passes through zero, as well as when it makes discontinuous jumps (which is when we want to change the sign). When it is progressing through zero we should ignore the (3.8) and so not change s_j or c_j.

The final possible error is connected to the one just discussed above. If the real and imaginary parts of $\hat{\varphi}(t)$ are both close to zero then it is quite conceivable that there will be a discontinuous jump in the imaginary part which is difficult to distinguish from a continuous movement through zero. Careful numerical checks will be required to ensure that an error is not induced because of this problem. However, in practice this type of error should not be common.

(3.3) A Real Univariate Random Variable

Suppose the conditions for theorem 2.3 hold, then we can use equation (2.8)

$$F(x) = \tfrac{1}{2} - \frac{1}{2\pi} \int_0^\infty \frac{\Delta}{t} \left[\frac{\varphi(t)e^{-itx}}{it}\right] dt.$$

Following Davies(1973) we will exploit the work of Bohmann(1961) to rewrite (2.8) to give us an expression which can be used to numerically invert characteristic functions. Our proof, however, is somewhat different to that used by Davies(1973).

<u>Theorem 3.2 (Davies(1973, equation 6))</u> If φ, f ϵ L, and the mean exists and

$\delta_{x,t} \Pr(X \leq x) = \Pr(X \leq x-t) - \Pr(X > x+t)$, then

$$F(x) + \sum_{j=1}^{\infty} \cos 2\pi zj \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X \leq x)\right]$$

$$= \frac{1}{2} - \frac{1}{2\pi}\mathrm{Im}\left[\frac{\varphi(\Delta z)\exp[-ix\Delta z]}{z}\right] - \frac{1}{2\pi}\sum_{v=1}^{\infty} \Delta_z \mathrm{Im}\left[\frac{\varphi(\Delta(z+v))\exp[-ix\Delta(z+v)]}{(z+v)}\right] \tag{3.11}$$

where Im(.) denotes the imaginary part of a complex function.

<u>Proof.</u> Given in the appendix.

<u>Corollary 3.1(Trapezoidal Rule)</u> If φ, f ϵ L and the mean exists and

$\delta_{x,t} \Pr(X \leq x) = \Pr(X \leq x-t) - \Pr(X > x+t)$ then,

$$F(x) + \sum_{j=1}^{\infty} \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X \leq x)\right]$$

$$= \frac{1}{2} - \frac{1}{2\pi} \lim_{v \to 0} \mathrm{Im}\left[\frac{\varphi(\Delta v)\exp[-ix\Delta v]}{v}\right] - \frac{2}{2\pi} \sum_{v=1}^{\infty} \mathrm{Im}\left[\frac{\varphi(\Delta v)\exp[-ix\Delta v]}{v}\right]. \tag{3.12}$$

<u>Proof.</u> Put z=0 in theorem 3.2.

The integrating error which results from using a trapezoidal rule is given by $\left|\sum_{j=1}^{\infty} \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X \leq x)\right]\right|$, which is rather difficult to manipulate. Hence, the trapezoidal rule has not been used for this type of inversion. An alternative form of the inversion is given below.

Corollary 3.2(Riemann Sum) If φ, f ϵ L and the mean exists and

$\delta_{x,t}\, \Pr(X\le x) = \Pr(X\le x-t) - \Pr(X>x+t)$ then,

$$F(x) + \sum_{j=1}^{\infty} (-1)^j \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X\le x)\right]$$

$$= \frac{1}{2} - \frac{1}{\pi} \sum_{v=0}^{\infty} \operatorname{Im}\left[\frac{\varphi(\Delta(v+1/2))\exp[\,ix\Delta(v+1/2)]}{(v+1/2)}\right]. \tag{3.13}$$

Proof. Put z=1/2 in theorem 3.2.

The Riemann sum is the most popular of the integration rules as it leads to a simple numerical integration error. The absolute value of the error is less than,

$$\left|\sum_{j=1}^{\infty} (-1)^j \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X\le x)\right]\right|$$

$$\le \max[F(x - \tfrac{2\pi}{\Delta}),\, 1 - F(x + \tfrac{2\pi}{\Delta})]. \tag{3.14}$$

Equation (3.4) showed us that 1–F(x) can be bounded, but using the same type of argument we can see that we can also bound F(x) as

$$F(\psi'(u)) \le M_X(-u)\exp(u\psi'(u))$$

$$= \exp(\psi(-u) + u\psi'(u)) \tag{3.15}$$

Therefore, by appropriate choice of Δ, we can bound this induced error by any small positive real ϵ.

(3.4) Quadratic Form in Normal Variables:– Eigenvalue Based Algorithms

Consider the problem of finding the distribution function of

$$Y = U'AU \tag{3.16}$$

where A is some T×T matrix and $U \sim N(\mu, \Sigma)$. This problem has been addressed by many writers in the last 50 years in connection with, for instance, the Durbin–Watson statistic.

If A is non–negative, then a comparatively cheap algorithm is given by Sheil and O'Muircheartaigh(1977), which has been improved upon by Farebrother(1984a). If μ is a vector of zeros and T is small (less than 70) then Pan's(1964) procedure, highlighted by Durbin and Watson(1971), is usually thought to be efficient, and is programmed in Pascal by Farebrother(1980, 1981) and in FORTRAN by White(1978). An alternative, but very specialized algorithm, has been suggested by Sargan and Bhargava(1983a, appendix A).

A more flexible algorithm has been given by Imhof(1961), which was programmed in FORTRAN by Koerts and Abrahamse(1968). Farebrother(1984c) showed that this subroutine had been improved upon by Davies'(1973) algorithm which has been programmed in Algol by Davies(1980). Imhof's(1961) contribution can now be introduced. He gave a simple expression for $\mathrm{Im}\left[\dfrac{\varphi(\Delta v)\exp[-ix\Delta v]}{v}\right]$.

This allowed the procedure to be carried out in real, rather than complex, arithmetic.

Although Imhof's arguments are well known we go through them here since we will use them in the multivariate case later. For simplicity of exposition, we will assume μ is a vector of zeros and $\Sigma = I$.

Recalling the characteristic function of Y is (c.f. Kendall, Stuart and Ord(1987, pg 489))

$$\varphi(t) = \prod_{j=1}^{T} (1 - 2it\lambda_j)^{-1/2} = |I - 2itA|^{-1/2}, \tag{3.17}$$

where (λ_j) are the eigenvalues of the matrix A and using the following results

$$\mathrm{Im}(z) = \sin\{\arg(z)\}.|z|, \tag{3.18a}$$

$$|wz| = |w|.|z|, \tag{3.18b}$$

$$\arg(wz) = \arg(w) + \arg(z), \tag{3.18c}$$

$$\arg((1 - 2it\lambda_j)^{-1/2}) = \tfrac{1}{2}\arctan 2t\lambda_j, \tag{3.18d}$$

$$\arg(e^{itx}) = tx, \tag{3.18e}$$

$$|(1 - 2it\lambda_j)^{-1/2}| = (1 + 4t^2\lambda_j^2), \tag{3.18f}$$

$$|e^{itx}| = 1, \tag{3.18g}$$

we have the following corollary:

Corollary 3.3 (The Imhof Procedure) If $\delta_{x,t} \Pr(X \leq x) = \Pr(X \leq x-t) - \Pr(X > x+t)$,

then

$$F(x) + \sum_{j=1}^{\infty} (-1)^j \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X \leq x) \right]$$

$$= \frac{1}{2} - \frac{1}{\pi} \sum_{v=0}^{\infty} \frac{\sin\ \epsilon(\Delta(v+1/2))\ \gamma(\Delta(v+1/2))}{(v+1/2)} \tag{3.19}$$

where $\epsilon(t) = \frac{1}{2} \sum_{j=1}^{T} \arctan 2t\lambda_j - tx$ (3.20)

and $\gamma(t) = \prod_{j=1}^{T} (1 + 4t^2\lambda_j^2)^{-1/4}.$ (3.20)

Truncating the infinite sum after M terms implies the introduction of the error,

$$\left| \frac{1}{\pi} \sum_{v=M+1}^{\infty} \frac{\sin\ \epsilon(\Delta(v+1/2))\ \gamma(\Delta(v+1/2))}{(v+1/2)} \right|. \tag{3.21}$$

Davies(1980) gave a useful upper bound on this error which allows M to be chosen so that the error is less than some arbitrarily small number η. We will not reproduce them here as they turn out not to be useful in our later discussion.

(3.5) A Quadratic Form in Normal Variables:– Eigenvalue–Free Algorithms

Frequently the eigenvalues of A are unknown, which means that they have to be computed numerically if (3.20) is to be used to compute the distribution function. It is known (c.f. Golub and van Loan(1983) and Press et al(1986, chapter 11)) that eigenvalue routines require, in general, $O(T^3)$ operations for their completion, which means that the inversion procedure becomes expensive when T is large. In the last decade there has been considerable interest in developing eigenvalue–free inversion algorithms which, it is hoped, will be less expensive when T is large. This section describes some of this work.

Evans and Savin(1984) used a difference equation. Another O(T) algorithm has been recently suggested by Shively, Ansley and Kohn(1988) who employed an extended Kalman filter. This approach is of more interest and so we discuss it here in a little detail.

Suppose we again think of the matrix

$$I - 2it(B - dI) = \kappa_0 I + \kappa_1 B \tag{3.22}$$

where $\kappa_0 = 1 + 2it.d$ and $\kappa_1 = -2it$. Suppose for a moment that instead of κ_0 and κ_1 being complex they were both positive and real. Then we could view $\kappa_0 I + \kappa_1 B$ as the covariance matrix of $(Y_1,\ldots,Y_T)'$, where Y_t is the stochastic process given below, in state space form.

$$Y_t = (1\ 0)\ \alpha_t + \epsilon_t, \tag{3.23a}$$

$$\alpha_t = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \alpha_{t-1} + R_t \eta_t, \text{ and } \alpha_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{3.23b}$$

with $R_t = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ for t=1,...,T−1 and $R_T = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and where (ϵ_t) and (η_t) are Gaussian white noise processes with variances κ_0 and κ_1 respectively. Then, the product of the one step ahead prediction error variances, which are given by the Kalman filter, would give the required determinant. Of course κ_0 and κ_1 are neither real, or in general, positive, but we could use an analytic continuation theorem to show that we can still formally proceed with the Kalman filter to calculate the determinant. (There is a problem with the square root of a complex number, but we shall not be concerned with that here; see Shively et al appendix B).

The Shively et al procedure will also cope with the introduction of explanatory variables into the covariance matrix, by using the modified Kalman filter of Ansley and Kohn(1985), and Kohn and Ansley(1986).

Two other sets of authors have also proposed eigenvalue–free methods for evaluating the characteristic function of this type. Palm and Sneek(1984) and Farebrother(1985) have studied the use of algorithms which produce tridiagonal matrices, which allows an eigenvalue–free calculation of the characteristic function. Farebrother(1989) has implemented his technique in a Pascal algorithm.

(3.6) A Bivariate Generalisation of Davies' Algorithm

Shively(1988a,b) has derived a bivariate generalization to Davies'(1973) Riemann sum algorithm (corollary 3.2). In this section we show that the work we presented in section (3.3) can be easily transferred across to the bivariate framework. We require no new manipulative theorems to do this, and the proof of the theorem follows exactly the same lines as the theorem given above. Our starting point is equation (2.24)

$$\frac{2^2}{(2\pi)^2}\int_0^\infty\int_0^\infty \underset{t_1}{\Delta}\ \underset{t_2}{\Delta}\left[\frac{\varphi(t)\ e^{-ix't}}{it_1\ it_2}\right] dt = 4F(x_1,x_2) - 2[F(x_1) + F(x_2)] + 1$$

$$= u^+(x_1,x_2). \tag{2.24}$$

Theorem 3.3 If φ, f ϵ L^2 and the mean of the relevant random variable exists and

$\delta_{x,t}\Pr(X\leq x) = \Pr(X\leq x-t) - \Pr(X>x+t)$, then

$$2^{-2}u^{+}(x_1,x_2) + \sum_{\substack{j_1=0 \\ j_1\neq j_2=0}}^{\infty} \sum_{j_2=0}^{\infty} (\cos 2\pi z_1 j_1 \cdot \cos 2\pi z_2 j_2)\, \delta_{x_1,\frac{2\pi j_1}{\Delta_1}}\, \delta_{x_2,\frac{2\pi j_2}{\Delta_2}} F(x_1,x_2)$$

$$= \frac{-2}{(2\pi)^2}\Delta_1\Delta_2 \left[\sum_{v_2=-\infty}^{\infty} \Delta_{z_2} \operatorname{Re}\left[\frac{\varphi(z_1,a_2)e^{-ix'(z_1,a_2)'}}{\Delta_1 z_1 \quad a_2}\right] \right.$$

$$\left. + \sum_{v_1=1}^{\infty} \sum_{v_2=-\infty}^{\infty} \Delta_{z_1}\Delta_{z_2} \operatorname{Re}\left[\frac{\varphi(a)e^{-ix'a}}{a_1 \quad a_2}\right] \right] \tag{3.25}$$

where $a = (\Delta_1(z_1 + v_1), \Delta_2(z_2 + v_2))' = (a_1, a_2)'$.

Proof. Given in the appendix.

<u>Corollary 3.4(Trapezium Rule)</u> If φ, $f \in L^2$ and the mean exists for the relevant random variable and $\delta_{x,t} \Pr(X \le x) = \Pr(X \le x-t) - \Pr(X > x+t)$, then

$$2^{-2}u^{+}(x_1,x_2) + \sum_{\substack{j_1=0 \\ j_1 \neq j_2 = 0}}^{\infty} \sum_{j_2=0}^{\infty} \delta_{x_1, \frac{2\pi j_1}{\Delta_1}} \delta_{x_2, \frac{2\pi j_2}{\Delta_2}} F(x_1,x_2)$$

$$= \frac{-1/2}{(2\pi)^2} \left[\lim_{z_1 \to 0} \lim_{z_2 \to 0} \sum_{v_2=-\infty}^{\infty} \Delta_{z_2} \operatorname{Re}\left[\frac{\varphi(z_1, a_2) e^{-ix'(z_1,a_2)'}}{z_1 \quad (v_2+z_2)}\right]\right.$$

$$\left. + \lim_{z_1 \to 0} \lim_{z_2 \to 0} \sum_{v_1=1}^{\infty} \sum_{v_2=-\infty}^{\infty} \Delta_{z_1} \Delta_{z_2} \operatorname{Re}\left[\frac{\varphi(a_1, a_2) e^{-ix'(a_1,a_2)'}}{(v_1+z_1)(v_2+z_2)}\right]\right] \tag{3.26}$$

<u>Proof.</u> Put $z_1 = z_2 = 0$ in theorem 3.3.

Bounding the resulting integration error is difficult. Here it is given by

$$\left| \sum_{\substack{j_1=0 \\ j_1 \neq j_2 = 0}}^{\infty} \sum_{j_2=0}^{\infty} \delta_{x_1, \frac{2\pi j_1}{\Delta_1}} \delta_{x_2, \frac{2\pi j_2}{\Delta_2}} F(x_1,x_2) \right|. \tag{3.27}$$

<u>Corollary 3.5(Riemann Sum)</u> If φ, f ϵ L^2 and the mean exists, and if

$\delta_{x,t} \Pr(X \le x) = \Pr(X \le x-t) - \Pr(X > x+t)$, then

$$2^{-2} u^{+}(x_1,x_2) + \sum_{\substack{j_1=0 \\ j_1 \ne j_2 = 0}}^{\infty} \sum_{j_2=0}^{\infty} (-1)^{j_1+j_2} \delta_{x_1, \frac{2\pi j_1}{\Delta_1}} \delta_{x_2, \frac{2\pi j_2}{\Delta_2}} F(x_1,x_2)$$

$$= \frac{-2}{(2\pi)^2} \left[\sum_{v_1=0}^{\infty} \sum_{v_2=-\infty}^{\infty} \mathrm{Re} \left[\frac{\varphi(\Delta_1 b_1, \Delta_2 b_2) e^{-ix'(\Delta_1 b_1, \Delta_2 b_2)'}}{b_1 \, b_2} \right] \right] \tag{3.28}$$

where $b = (v_1 + 1/2, v_2 + 1/2) = (b_1, b_2)$.

<u>Proof.</u> Put $z_1 = z_2 = 1/2$ in theorem 3.3, and then use the symmetry with respect to z_2, as v_2 is as an integer.

This error bound is much more amenable, as

$$\left| \sum_{j_1=0}^{\infty} \sum_{\substack{j_2=0 \\ j_1 \neq j_2 = 0}}^{\infty} (-1)^{j_1+j_2} \delta_{x_1, \frac{2\pi j_1}{\Delta_1}} \delta_{x_2, \frac{2\pi j_2}{\Delta_2}} F(x_1, x_2) \right|$$

$$\leq \max\left[[1-F(x_1 + \frac{2\pi}{\Delta_1})] + [1-F(x_2 + \frac{2\pi}{\Delta_2})],\ F(x_1 - \frac{2\pi}{\Delta_1}) + F(x_2 - \frac{2\pi}{\Delta_2}) \right]. \tag{3.29}$$

Thus we can use the bounds employed in sections (3.2) and (3.3) to provide a way of selecting the step sizes, Δ_1 and Δ_2.

(3.7) Two Quadratic Forms in Normal Variables

Suppose we are interested in the joint distribution of the two quadratic forms in normal variables

$$Y_1 = U'AU \tag{3.30a}$$

$$Y_2 = U'BU, \tag{3.30b}$$

where A and B are T×T matrices and, for sake of simplicity $U \sim N(0, I)$.

The joint characteristic function of Y_1 and Y_2 is given by

$$\varphi(t_1, t_2) = E \exp\left[iU'(t_1A + t_2B)U\right]$$

$$= \prod_{j=1}^{T} (1 - 2i\delta_j(t_1, t_2))^{-1/2} \tag{3.31}$$

where $\delta_j(t_1,t_2)$ denotes the j–th eigenvalue of the $(t_1A + t_2B)$ matrix. For this case, we have sufficient regularity to allow us to write the equality

$$2^{-2}u^+(x_1,x_2) + \sum_{\substack{j_1=0 \\ j_1 \neq j_2 = 0}}^{\infty} \sum_{j_2=0}^{\infty} (-1)^{j_1+j_2} \delta_{x_1, \frac{2\pi j_1}{\Delta_1}} \delta_{x_2, \frac{2\pi j_2}{\Delta_2}} F(x_1,x_2)$$

$$= \frac{-2}{(2\pi)^2}\left[\sum_{v_1=0}^{\infty} \sum_{v_2=-\infty}^{\infty} \operatorname{Re}\left[\frac{\varphi(\Delta_1 b_1, \Delta_2 b_2)e^{-ix'(\Delta_1 b_1, \Delta_2 b_2)'}}{b_1 \; b_2}\right]\right]. \tag{3.32}$$

Using the arguments of Imhof(1961) we have

$$\arg\left[\varphi(t)e^{-ix't}\right] = \left[\frac{1}{2}\sum_{j=1}^{T} \arctan 2\delta_j(t_1,t_2) - x't\right] \tag{3.33a}$$

$$|\varphi(t)e^{-ix't}| = \prod_{j=1}^{T} \left[1 + 4\delta_j^2(t_1,t_2)\right]^{-1/4} \tag{3.33b}$$

So writing $\epsilon(t) = \arg\left[\varphi(t)e^{-ix't}\right]$ and $\gamma(t) = |\varphi(t)e^{-ix't}|$, we have that

$$\mathrm{Re}\left[\frac{\varphi(\Delta_1 b_1,\Delta_2 b_2)e^{-ix'(\Delta_1 b_1,\Delta_2 b_2)'}}{b_1\ b_2}\right] = \frac{\gamma(\Delta_1 b_1,\ \Delta_2 b_2).\cos\ \epsilon(\Delta_1 b_1,\ \Delta_2 b_2)}{b_1\ b_2} \tag{3.33c}$$

This expression is not very useful for numerical work as it depends on the eigenvalues of the matrix $(t_1A + t_2B)$, which change as t_1 and t_2 vary, thus requiring many numerical eigenvalue calculations if the matrices A and B do not share the same eigenvectors.If A and B do not share the same eigenvectors, then it seems as if this technique could be expensive. This prompts us to investigate the possibility of employing eigenvalue–free techniques. Shively et al(1988) discuss this possibility at some length, when the matrices of interest are band matrices, or slight variants of them.

If the Davies(1973) procedure is abandoned then we can resurrect the eigenvalue technique. Suppose we consider equation (2.32) again. We can then prove the following corollary by thinking about the transform $u = 2t_1$ and $v = t_2/t_1$

Corollary 3.8

$$\int_0^\infty \int_0^\infty \underset{t_1}{\Delta} \underset{t_2}{\Delta} \left[\frac{\varphi(t)e^{-ix't}}{it_1 it_2}\right] dt_1\, dt_2$$

$$= \int_0^\infty \int_0^\infty \frac{\gamma(u,uv)\ \cos\ \epsilon(u,uv)}{u\ v}\, du\, dv \tag{3.34a}$$

where $$\epsilon(u,uv) = \frac{1}{2}\left[\sum_{j=1}^{T} \text{arc tan } u.\delta_j(v) - x't\right] \tag{3.34b}$$

$$\gamma(u,uv) = \prod_{j=1}^{T} \left[1 + u^2\delta_j^2(v)\right]^{-1/4} \tag{3.34c}$$

and where $\delta_j(v)$ is the j–th eigenvalue of $(A + vB)$.

Proof. $$\int_0^\infty \int_0^\infty \underset{t_1}{\Delta} \underset{t_2}{\Delta} \left[\frac{\varphi(t)e^{-ix't}}{it_1 it_2}\right] dt_1\, dt_2$$

$$= \int_0^\infty \int_0^\infty \underset{u}{\Delta} \underset{uv}{\Delta} \left[\frac{\varphi(u,uv)e^{-ix'(u,uv)}}{iu\ \ iv}\right] du\, dv$$

$$= \int_0^\infty \int_0^\infty \underset{u}{\Delta} \underset{v}{\Delta} \left[\frac{\varphi(u,uv)e^{-ix'(u,uv)}}{iu\ \ iv}\right] du\, dv. \tag{3.35}$$

The use of the Imhof result proves the corollary □.

This representation has substantial advantages over the previous one, as now we only need eigenvalue calculations with respect to a single variable.

It does not seem possible to produce an explicit numerical integration routine using this form of the expression. This should cause us little difficulty as we can use standard routines to do this.

<u>Example 3.7.1</u> <u>Independent Chi–Squared Variables</u>

Suppose we choose A and B as diagonal matrices, with diagonal elements being unity and zero. They are given below

$$A = \begin{bmatrix} I_{T/2} & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & I_{T/2} \end{bmatrix} \tag{3.36}$$

when T is an even integer and I_r denotes an order r identity matrix. Y_1 and Y_2 are independent chi–squared variables with T/2 degrees of freedom. Then, of course, their joint distribution function is then the product of the marginal distribution functions and so their distribution function is exactly known. However, this case is an interesting one which will allow us to study the feasibility of the bivariate inversions.

The infinite sums in (3.32) have to be truncated at finite points. Unlike the case of the univariate inversion (c.f. Davies(1980)), no useful bound is possible on the resulting truncation error; sensible ad hoc rules can be invented which should mean that it is unlikely that this error will be large. The rule we employ is to truncate an inner sum if there has been ICOI consecutive terms less than, in absolute value, EPSI. Likewise, the outer sum is truncated using the corresponding numbers ICOO and EPSO. It is important to check the evolution of these sums to ensure that they do not oscillate so slowly that they cause a major truncation error. A simple visual check of the evolution of the outer sum and a small sub–sample of the inner sums should be sufficient to guard against this in practice.

Table 3.1 reports the number of terms used in the summation when we take MI=10, MO=10, EPSA1=1.0×10^{-6}, EPSA2=1.0×10^{-6}. In these tables we allow T, x_1 and x_2 to vary. Throughout, there was no error in the calculation of the distribution function, which exceeded 1×10^{-6}. The tables have elements corresponding to (X_1, X_2).

Table 3.1

Number of Terms in the Summation before the Truncation

T= 16	0.01	0.1	0.5	0.9	0.99	0.999
0.01	1843	1723	1656	1473	1257	1253
0.1	1829	1769	1636	1495	1277	1218
0.5	735	695	711	693	665	621
0.9	1688	1531	1587	1375	1182	997
0.99	1630	1569	1368	1296	1081	969
0.999	611	1484	1355	1265	1068	1043

T= 32	0.01	0.1	0.5	0.9	0.99	0.999
0.01	696	655	586	496	436	479
0.1	654	659	590	470	414	469
0.5	441	435	431	415	381	401
0.9	608	577	534	461	352	384
0.99	526	547	508	453	362	389
0.999	592	561	478	461	393	424

T= 64	0.01	0.1	0.5	0.9	0.99	0.999
0.01	386	362	320	284	257	320
0.1	390	352	324	258	251	306
0.5	357	359	295	263	237	295
0.9	352	301	296	265	230	238
0.99	323	317	292	230	228	259
0.999	340	329	302	242	236	263

T=128	0.01	0.1	0.5	0.9	0.99	0.999
0.01	312	278	240	211	219	252
0.1	277	274	238	206	211	238
0.5	257	255	221	193	195	227
0.9	276	252	197	166	191	222
0.99	288	256	199	168	197	226
0.999	294	238	234	176	203	238

T=256	0.01	0.1	0.5	0.9	0.99	0.999
0.01	242	238	209	174	207	234
0.1	260	234	197	172	205	230
0.5	250	194	165	137	169	192
0.9	222	220	185	160	187	195
0.99	230	228	195	164	197	224
0.999	238	230	195	172	201	232

T=512	0.01	0.1	0.5	0.9	0.99	0.999
0.01	234	204	147	172	205	207
0.1	228	196	172	168	197	205
0.5	212	191	159	159	183	189
0.9	220	190	164	160	189	214
0.99	226	196	166	166	193	218
0.999	230	206	170	172	197	234

T=1024	0.01	0.1	0.5	0.9	0.99	0.999
0.01	232	176	170	170	203	231
0.1	224	197	139	166	197	223
0.5	211	185	155	157	182	213
0.9	193	189	160	164	189	195
0.99	201	195	164	166	197	199
0.999	205	199	172	172	199	209

T=2048	0.01	0.1	0.5	0.9	0.99	0.999
0.01	201	197	170	170	201	207
0.1	195	168	162	168	197	201
0.5	188	158	129	156	184	186
0.9	193	187	160	135	193	218
0.99	203	195	166	170	195	224
0.999	209	203	168	172	176	226

Example 3.7.2 Non–Circular Correlation Coefficients

Consider the joint distribution function of the first two non–circular serial correlation coefficients. We follow Durbin(1980) in defining

$$r_j = \frac{y' A_j y}{y' y}, \qquad (j=1,2) \qquad (3.37)$$

as the serial correlation coefficients, where A_1 and A_2 are given below

$$2.A_1 = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots\ \vdots \\ 0 & 0 & 0 & \cdots & 1\ 1 \end{bmatrix} \text{ and } A_2 = 2A_1^2 - I. \qquad (3.38)$$

For sake of simplicity we allow y to be a T×1 vector containing independent and identically distributed zero mean Gaussian variables.

The eigenvalues of A_j are

$$\mu_{rj} = \cos \left[j\pi(r-1)/T \right] \qquad (r=1,\ldots,T) \qquad (3.39)$$

while they share the same eigenvectors (see Anderson(1971, pg 282–290). This means that we can write

$$\Pr(r_1 \leq d_1, r_2 \leq d_2)$$

$$=\Pr\left[\sum_{t=1}^{T} u_t^2(\mu_{r1} - d_1) \leq 0, \sum_{t=1}^{T} u_t^2(\mu_{r2} - d_2) \leq 0\right] \qquad (3.40)$$

where $u_t \sim NID(0,1)$.

In table 3.2, are the results for various joint probabilities for the serial correlation coefficients. We took $d_1 = d_2 = d$, where $d = \pm 1.96/\sqrt{T}$, which represents the critical values for r_1 and r_2 at a 5% level using the usual asymptotic theory. Bracketed beneath these probabilities are the number of terms in the sum (32) used in their computation. Throughout we took ICOI = ICOO=8, EPSI = 1×10^{-6}, EPSO = 1×10^{-5} and $\epsilon = 6\times10^{-7}$. The marginal distribution functions were evaluated using Davies(1980) algorithm, with the corresponding induced error chosen to be negligible.

Table 3.2

Exact Joint Probability for the First Two Serial Correlation Coefficients

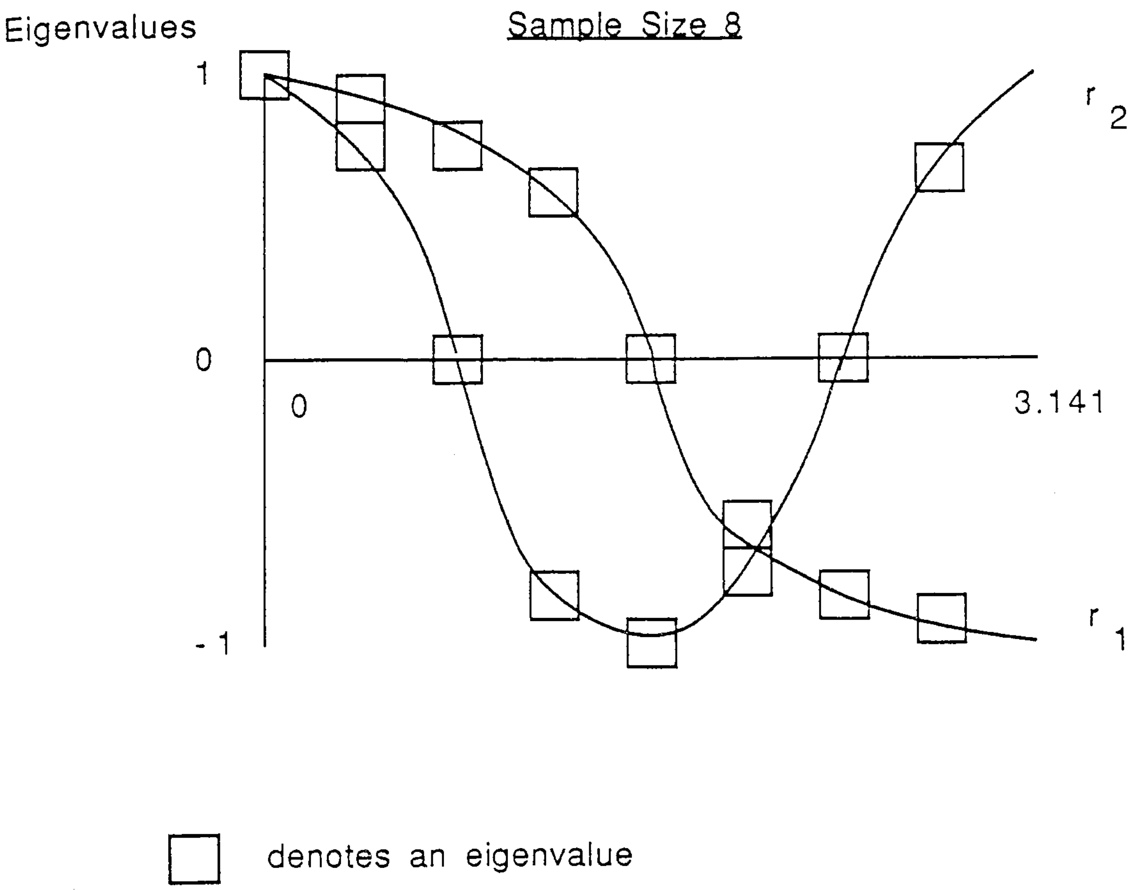
T	$\Pr(X_1>d, X_2\leq -d)$	$\Pr(X_1\leq -d, X_2>d)$	$\Pr(X_1>d, X_2>d)$	$\Pr(X_1\leq -d, X_2\leq -d)$
8	0.00000 (5209)	0.00001 (5829)	0.00446 (5822)	0.00000 (5350)
16	0.00000 (1856)	0.00007 (1338)	0.00389 (1684)	0.00000 (1565)
32	0.00000 (601)	0.00100 (664)	0.00279 (640)	0.00000 (686)
64	0.00007 (373)	0.00107 (356)	0.00204 (374)	0.00004 (358)
128	0.00020 (253)	0.00101 (261)	0.00157 (234)	0.00014 (236)
256	0.00031 (209)	0.00092 (214)	0.00125 (214)	0.00023 (211)
512	0.00040 (201)	0.00085 (181)	0.00105 (199)	0.00033 (203)
1024	0.00047 (248)	0.00079 (246)	0.00092 (246)	0.00041 (248)
2048	0.00054 (434)	0.00077 (432)	0.00085 (440)	0.00049 (434)
∞	0.00062	0.00062	0.00062	0.00062

Note:– The last digit in all these probabilities should be viewed as unreliable. The asymptotic result is calculated using large sample theory.

A Monte Carlo experiment was performed to check the accuracy of these calculations. When T was less than 100, one million replications were used in the experiment, while for larger T we slowly reduced the number of replications. The Gaussian white noise variables were generated using the Numerical Algorithms Group routines G05DDF and G05CBF. Throughout the results were found to be accurate to four decimal figures.

The results are very surprising. They indicate that unless the sample size is large (well over 100), the joint distribution function is significantly asymmetrical. We can explain this rather peculiar behaviour by studying diagram 3.1. This depicts the eigenvalues associated with r_1 and r_2 for T=8. It indicates that there exist values of t for which both μ_{t1} and μ_{t2} are substantially positive, and which will cause $\Pr(X_1>d, X_2>d)$ to be comparatively large. The same type of behaviour should also be expected for $\Pr(X_1\leq -d, X_2>d)$ as there exist some values of t for which μ_{t1} is significantly negative, while μ_{t2} is strongly positive. This probability is small for small T because of the "end effect" on the eigenvalues (ie. they only go up to $\cos\frac{\pi(T-1)}{T}$ and $\cos\frac{2\pi(T-1)}{T}$), although when T rises above 10, this value will grow rapidly until T reaches 100. The other two probabilities will be small as their influential eigenvalues are small, in absolute value. Only when T becomes large will these probabilities start to converge from below, to 0.000625.

This pattern in the eigenvalues generalizes to the whole correlogram. We can see that there is a surprisingly large probability of observing all the serial correlation coefficients to be positive, and a smaller, but non–negligible, probability that they alternate in sign– starting with r_1 being negative.


Eigenvalues
Sample Size 8
1
0
-1
0
3.141
r_2
r_1
denotes an eigenvalue

(3.8) A Multivariate Generalisation of Davies' Algorithm

Recall equation (2.15),

$$\frac{(-2)^p}{(2\pi)^p} \int_0^\infty \cdots \int_0^\infty \underset{t_1}{\Delta} \underset{t_2}{\Delta} \cdots \underset{t_p}{\Delta} \left[\frac{\varphi(t) e^{-x' t}}{i t_1 \cdots i t_p} \right] dt = u^+(x).$$

Introducing the notation that $\sum\limits_{j=0}^{\infty}$ denotes a sum over the variables $j_1, \ldots, j_p$ which vary between zero and infinity, but never all equal zero simultaneously, we have the following theorem.

Theorem 3.4 If φ, f ϵ L^p and the mean of the relevant random variable exists, and if

$\underset{x,t}{\delta}$ Pr(X≤x) = Pr(X≤x–t) – Pr(X>x+t), then

$$2^{-p}u^{+}(x) + \sum_{j=0}^{\infty} \left[\prod_{k=1}^{p} (\cos 2\pi z_k j_k)\ \underset{x_k, \frac{2\pi j_k}{\Delta_k}}{\delta} \right] F(x)$$

$$= \frac{(-i)^p 2^{p-1}}{(2\pi)^p} \Delta_1 \ldots \Delta_p \times \left[\sum_{v_2=-\infty}^{\infty} \sum_{v_3=-\infty}^{\infty} \cdots \sum_{v_p=-\infty}^{\infty} \right.$$

$$\underset{t_2}{\Delta} \underset{t_3}{\Delta} \cdots \underset{t_p}{\Delta} \operatorname{Re}\left[\frac{\varphi(z_1, a_2, \ldots, a_p) e^{-ix'(z_1, a_2, \ldots, a_p)'}}{\Delta_1 z_1 \quad a_1 a_2 a_3 \cdot \cdot \cdot a_p} \right]$$

$$\left. + \sum_{v_1=1}^{\infty} \sum_{v_2=-\infty}^{\infty} \cdots \sum_{v_p=-\infty}^{\infty} \underset{t_1}{\Delta} \underset{t_2}{\Delta} \cdots \underset{t_p}{\Delta} \operatorname{Re}\left[\frac{\varphi(a) e^{-ix'a}}{a_1 a_2 \ldots a_p} \right] \right]$$

if p is even, and

$$= -\frac{(-i)^{p-1}2^{p-1}}{(2\pi)^p}\Delta_1\cdots\Delta_p\times\left[\sum_{v_2=-\infty}^{\infty}\sum_{v_3=-\infty}^{\infty}\cdots\sum_{v_p=-\infty}^{\infty}\right.$$

$$\underset{t_2}{\Delta}\,\underset{t_3}{\Delta}\cdots\underset{t_p}{\Delta}\,\mathrm{Im}\left[\frac{\varphi(z_1,a_2,\ldots,a_p)e^{-ix'(z_1,a_2,\ldots,a_p)'}}{\Delta_1 z_1\quad a_1\,a_2\,a_3\,.\,.\,.\,a_p}\right]$$

$$\left.+\sum_{v_1=1}^{\infty}\sum_{v_2=-\infty}^{\infty}\cdots\sum_{v_p=-\infty}^{\infty}\underset{t_1}{\Delta}\,\underset{t_2}{\Delta}\cdots\underset{t_p}{\Delta}\,\mathrm{Im}\left[\frac{\varphi(a)e^{-ix'a}}{a_1\;a_2\cdots a_p}\right]\right] \tag{3.41}$$

if p is odd, where

$$a = (\Delta_1(z_1 + v_1),\ \Delta_2(z_2 + v_2),\ldots,\ \Delta_p(z_p + v_p))' = (a_1,\ a_2,\ldots,a_p)'.$$

Proof. Trivial generalization of the proof of theorem 3.3. The sole problem is ensuring that the constant on the right hand side of (3.41) is correct. In the case of p being even this constant is, because of proposition 1,

$\frac{(2)^{-2p}.2^{2p-1}i^{p}(-2)^{p}}{(2\pi)^{p}} = \frac{(-i)^{p}2^{p-1}}{(2\pi)^{p}}$. In the case of p being odd, it is

$\frac{(2)^{-2p}.2^{2p-1}i^{p-1}(-2)^{p}}{(2\pi)^{p}} = -\frac{(-i)^{p-1}2^{p-1}}{(2\pi)^{p}}$.

Again we can employ either a Trapezoidal Rule in the integration or a Riemann sum. As in the univariate case, the Trapezoidal Rule does not offer a very convenient form to work with. Thus we only present here the Riemann sum.

Corllary 3.9 (Riemann Sum) If $\delta_{x,t} \Pr(X \leq x) = \Pr(X \leq x-t) - \Pr(X > x+t)$, then,

$$2^{-p}u^{+}(x) + \sum_{j=0}^{\infty} \left[\prod_{k=1}^{p} (-1)^{j_k} \, x_k, \frac{\delta_{2\pi j_k}}{\Delta_k} \right] F(x)$$

$$= \frac{(-i)^p 2^{2p-1}}{(2\pi)^p} \Delta_1 \cdots \Delta_p \left[\sum_{v_1=0}^{\infty} \sum_{v_3=-\infty}^{\infty} \cdots \sum_{v_p=-\infty}^{\infty} \mathrm{Re}\left[\frac{\varphi(b) e^{-ix'b}}{b_1 \, b_2 \cdots b_p} \right] \right]$$

if p is even, and

$$= -\frac{(-i)^{p-1} 2^{2p-1}}{(2\pi)^p} \Delta_1 \cdots \Delta_p \left[\sum_{v_1=0}^{\infty} \sum_{v_3=-\infty}^{\infty} \cdots \sum_{v_p=-\infty}^{\infty} \mathrm{Re}\left[\frac{\varphi(b) e^{-ix'b}}{b_1 \, b_2 \cdots b_p} \right] \right] \tag{3.42}$$

if p is odd, where

$b = (\Delta_1(1/2 + v_1), \Delta_2(1/2 + v_2), \ldots, \Delta_p(1/2 + v_p))' = (b_1, b_2, \ldots, b_p)'$.

Proof. Trivial.

Mathematical Appendix

Proof of Theorem 3.2 Define $\delta_{x,t} \Pr(X\leq x) = \Pr(X\leq x-t) - \Pr(X>x+t)$. Using (2.8) we have

$$\delta_{x,t} F(x) = \frac{-2}{2\pi}\int_0^\infty \cos ut \, \Delta_u \left[\frac{\varphi(u)e^{-iux}}{ui}\right] du. \tag{A1}$$

Hence $\sum_{j=0}^{\infty} \cos 2\pi zj \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X\leq x)\right]$

$$= F(x) + \sum_{j=1}^{\infty} \cos 2\pi zj \left[\delta_{x,\frac{2\pi j}{\Delta}} \Pr(X\leq x)\right]$$

$$= \frac{1}{2} - \frac{\Delta}{2\pi}\sum_{j=-\infty}^{\infty} \exp[-2\pi ijz]\int_0^\infty \cos 2\pi jy \, \Delta_y \left[\frac{\varphi(\Delta y)e^{-ix\Delta y}}{i\Delta y}\right] dy. \tag{A2}$$

The second term on the left hand side of (A2) can be viewed as an error caused by approximating the inversion formula. Using proposition 3.1, which is a slight variant of a well known theorem (Poisson's formula, c.f. Bohmann(1961) or Zygmund(1955, pg 37)) given in the appendix for completeness, we have, for p = 1

$$= \frac{1}{2} - \frac{1}{2}\frac{\Delta}{2\pi}\sum_{v=-\infty}^{\infty} \Delta_z \left[\frac{\varphi(\Delta(z+v))\exp[-ix\Delta(z+v)]}{\Delta(z+v)i}\right]$$

$$= \frac{1}{2} - \frac{1}{2}\frac{\Delta}{2\pi}\frac{\Delta}{z}\left[\frac{\varphi(\Delta z)\exp[-ix\Delta z]}{\Delta z i}\right]$$

$$-\frac{1}{2}\frac{\Delta}{2\pi}\sum_{v=1}^{\infty}\frac{\Delta}{z}\frac{\Delta}{z+v}\left[\frac{\varphi(\Delta(z+v))\exp[-ix\Delta(z+v)]}{\Delta(z+v)i}\right] \tag{A3}$$

$$=\frac{1}{2} - \frac{1}{2\pi}\operatorname{Im}\left[\frac{\varphi(\Delta(z))\exp[-ix\Delta z]}{z}\right] - \frac{1}{2\pi}\sum_{v=1}^{\infty}\frac{\Delta}{z}\operatorname{Im}\left[\frac{\varphi(\Delta(z+v))\exp[-ix\Delta(z+v)]}{(z+v)}\right]$$

□.

Proof of Theorem 3.3

$$\underset{x_1,s_1}{\delta}\ \underset{x_2,s_2}{\delta} F(x_1,x_2) = \frac{(-2)^2}{(2\pi)^2}\int_0^\infty\int_0^\infty [\cos s_1t_1 \cdot \cos s_2t_2]\ \underset{t_1}{\Delta}\ \underset{t_2}{\Delta}\left[\frac{\varphi(t)\ e^{-ix't}}{it_1\ it_2}\right] dt. \tag{A4}$$

Hence

$$2^{-2}u^+(x_1,x_2) + \sum_{\substack{j_1=0 \\ j_1\neq j_2=0}}^{\infty}\ \sum_{j_2=0}^{\infty} (\cos 2\pi z_1 j_1 \cdot \cos 2\pi z_2 j_2)\ \underset{x_1,\frac{2\pi j_1}{\Delta_1}}{\delta}\ \underset{x_2,\frac{2\pi j_2}{\Delta_2}}{\delta} F(x_1,x_2)$$

$$= \frac{1}{(2\pi)^2}\Delta_1\Delta_2 \sum_{j_1=-\infty}^{\infty}\ \sum_{j_2=-\infty}^{\infty} \exp[-2\pi i[j_1z_1 + j_2z_2]] \times$$

$$\int_0^\infty\int_0^\infty [\cos\frac{2\pi j_1}{\Delta_1}t_1 \cdot \cos\frac{2\pi j_2}{\Delta_2}t_2]\ \underset{t_1}{\Delta}\ \underset{t_2}{\Delta}\left[\frac{\varphi(\Delta_1t_1,\ \Delta_2t_2)\ e^{-ix'(\Delta_1t_1,\ \Delta_2t_2)'}}{i\Delta_1t_1 \cdot i\Delta_2t_2}\right] dt \tag{A5}$$

So, using proposition 3.1, we have for $p = 2$

$$= \frac{1}{(2\pi)^2}\cdot\frac{1}{(2)^2}\Delta_1\cdot\Delta_2 \sum_{v_1=-\infty}^{\infty}\ \sum_{v_2=-\infty}^{\infty}\ \underset{z_1}{\Delta}\ \underset{z_2}{\Delta}\left[\frac{\varphi(a)e^{-ix'a}}{ia_1\ ia_2}\right]$$

$$= \frac{1}{(2\pi)^2} \cdot \frac{1}{(2)^2} \Delta_1 . \Delta_2 \left[\sum_{v_2=-\infty}^{\infty} \underset{z_2}{\Delta} \underset{z_1}{\Delta} \left[\frac{\varphi(z_1\,,\,a_2) e^{-ix'(z_1,a_2)'}}{i\Delta_1 z_1 \quad ia_2} \right] \right.$$

$$\left. + \sum_{v_1=1}^{\infty} \sum_{v_2=-\infty}^{\infty} \underset{z_1}{\Delta} \underset{z_2}{\Delta} \underset{a_1}{\Delta} \left[\frac{\varphi(a) e^{-ix'a}}{ia_1 \; ia_2} \right] \right]. \qquad \text{(A6)}$$

Thus

$$2^{-2} u^{+}(x_1,x_2) + \sum_{\substack{j_1=0 \\ j_1 \neq j_2 = 0}}^{\infty} \sum_{j_2=0}^{\infty} (\cos 2\pi z_1 j_1 \, . \, \cos 2\pi z_2 j_2) \; \underset{x_1, \frac{2\pi j_1}{\Delta_1}}{\delta} \; \underset{x_2, \frac{2\pi j_2}{\Delta_2}}{\delta} F(x_1,x_2)$$

$$= \frac{-1/2}{(2\pi)^2} \left[\sum_{v_2=-\infty}^{\infty} \underset{z_2}{\Delta} \operatorname{Re} \left[\frac{\varphi(z_1, a_2) e^{-ix'(z_1,a_2)'}}{z_1 \qquad (v_2+z_2)} \right] + \right.$$

$$\left. \sum_{v_1=1}^{\infty} \sum_{v_2=-\infty}^{\infty} \underset{z_1}{\Delta} \underset{z_2}{\Delta} \operatorname{Re} \left[\frac{\varphi(a)\, e^{-ix'a}}{(v_1+z_1)(v_2+z_2)} \right] \right]$$

□.

Proposition 3.1 is a slight variant of Poisson's formula, c.f. Bohmann(1961, pg 124).

<u>Proposition 3.1</u> Assume the condition

<u>B1</u> $\Delta_{z_1} \cdots \Delta_{z_p} g(z_1,\ldots,z_p) \in L$ and is both continuous and bounded.

Then

$$\sum_{v_1=-\infty}^{\infty} \cdots \sum_{v_p=-\infty}^{\infty} \Delta_{z_1} \cdots \Delta_{z_p} g(v_1+z_1,\ldots,v_p+z_p) =$$

$$2^p \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_p=-\infty}^{\infty} e^{-2\pi i n'z} \int_0^{\infty} \cdots \int_0^{\infty} \Big(\prod_{j=1}^{p} \cos 2\pi n_j y_j\Big) \Delta_{y_1} \cdots \Delta_{y_p} g(y_1,\ldots y_p)\, dy_1 \ldots dy_p \tag{A7}$$

where $n = \begin{bmatrix} n_1 \\ \vdots \\ n_p \end{bmatrix}$, $z = \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix}$.

Proof of Proposition 3.1 Define

$$\underset{z_1}{\Delta} \cdots \underset{z_p}{\Delta} G(z_1,\ldots,z_p) = \sum_{v_1=-\infty}^{\infty} \cdots \sum_{v_p=-\infty}^{\infty} \underset{z_1}{\Delta} \cdots \underset{z_p}{\Delta} g(v_1+z_1,\ldots,v_p+z_p).$$

Then, by B1, $\underset{z_1}{\Delta} \cdots \underset{z_p}{\Delta} G(z_1,\ldots,z_p)$ is integrable, continuous and bounded. It has the Fourier coefficients

$$\int_{-0\cdot 5}^{0\cdot 5} \cdots \int_{-0\cdot 5}^{0\cdot 5} e^{2\pi i n' t} \underset{t_1}{\Delta} \cdots \underset{t_p}{\Delta} G(t_1,\ldots,t_p)\ dt_1 \ldots dt_p$$

$$= 2^p \int_0^{0\cdot 5} \cdots \int_0^{0\cdot 5} \Big(\prod_{j=1}^{p} \cos 2\pi n_j t_j\Big) \underset{t_1}{\Delta} \cdots \underset{t_p}{\Delta} G(t_1,\ldots,t_p)\ dt_1 \ldots dt_p \tag{A8}$$

so using B1 and the Fourier inversion theorem

$$\underset{z_1}{\Delta} \cdots \underset{z_p}{\Delta} G(z_1,\ldots,z_p) = 2^p \sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_p=-\infty}^{\infty} e^{-2\pi i n' z}.$$

$$\int_0^{0\cdot 5} \cdots \int_0^{0\cdot 5} \Big(\prod_{j=1}^{p} \cos 2\pi n_j t_j\Big) \underset{t_1}{\Delta} \cdots \underset{t_p}{\Delta} G(t_1,\ldots,t_p)\ dt_1 \ldots dt_p. \tag{A9}$$

The integral can be reexpressed as

$$\sum_{v_1=-\infty}^{\infty} \cdots \sum_{v_p=-\infty}^{\infty} \int_0^{0\cdot 5} \cdots \int_0^{0\cdot 5} \left(\prod_{j=1}^{p} \cos 2\pi n_j t_j\right) \underset{t}{\Delta} \cdots \underset{t}{\Delta} g(v_1+t_1,\ldots,v_p+t_p) dt_1 \ldots dt_p$$

so using a change of integrating variable, and noting $n_i v_i$ is an integer, it becomes

$$\int_0^{\infty} \cdots \int_0^{\infty} \left(\prod_{j=1}^{p} \cos 2\pi n_j y_j\right) \underset{y_1}{\Delta} \cdots \underset{y_p}{\Delta} g(y_1,\ldots,y_p) \; dy_1 \ldots dy_p$$

and so the proposition is proved □.

Chapter Four

The Connection between the Maximum Likelihood Estimation of Moving Averages and Local Trend Models

Summary This chapter gives a detailed description of the existing simulation and analytic literature on the properties of maximum likelihood (ML) estimators of moving average processes. The connection between moving averages and the local trend models is derived. It is shown that although the existing work on the estimation of noninvertible processes does aid our understanding of the estimation of deterministic components in local trend models, there are substantial gaps in our knowledge which are identifiable.

Finally, a valid Edgeworth expansion for the ML estimator of the local level model is derived. It allows us to describe the process of breakdown in the usefulness of the usual asymptotic results.

Keywords MAXIMUM LIKELIHOOD ESTIMATION, NONINVERTIBILITY, LOCAL TREND MODELS, EDGEWORTH EXPANSIONS.

(4.1) Introduction

Over the last twenty years there has been considerable interest in the properties of maximum likelihood (ML) estimators of moving average processes. In this chapter we focus on the relevance of this substantial literature to the issue of this thesis.

The connection between the local level model and a moving average of order one is easy to show. Recall the local level model is

$$y_t = \mu_t + \epsilon_t, \tag{1.1}$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{1.2}$$

where (ϵ_t), (η_t) are uncorrelated, zero mean, white noise processes with variances σ^2 and $q\sigma^2$ respectively. The differenced version of this is

$$\Delta\, y_t = y_t - y_{t-1} = \eta_t + \Delta\, \epsilon_t, \tag{4.1}$$

and so the autocorrelation function for this, Δy_t, stationary process is

$$\rho(1) = \frac{-1}{\{2 + q\}},\ \rho(s) = 0,\ \forall\ s > 1. \tag{4.2}$$

Thus Δy_t can be written as a first order moving average process

$$\Delta\, y_t = \xi_t + \theta\, \xi_{t-1}, \tag{4.3}$$

where $$\rho(1) = \frac{\theta}{\{1 + \theta^2\}}, \tag{4.4}$$

with

$$q = -\frac{\{1+\theta^2\}}{\theta} - 2, \tag{4.5}$$

and θ is constrained to obey the invertibility convention (which is an identifiability condition) that $|\theta| \leq 1$.

As q is always non–negative, it must be the case that $\theta \in [-1, 0]$. Hence, we see that Δy_t is a special case of a first order moving average process. The occurrence of q being zero is also of interest, as it implies θ must be minus one, ie. the MA(1) process is noninvertible.

Although this result connects the occurrence of deterministic components with noninvertibility, it is not a straightforward matter to deduce the behaviour of q from what we already know about noninvertible processes. The principle reason for this is that the likelihood function for the local level model will be influenced by the assumption we make about the startup procedure we use to initialize the nonstationary state (1.2).

The connection between the local linear trend model and a second order moving average process is derived in chapter seven. In this case, the constraint on the parameter space, imposed by use of the local trend model, is quite severe. As no significantly detailed work has been done on noninvertible second order moving average processes, we do not discuss this model in detail here, but wait until chapter seven to do so.

The rest of this chapter is split into two sections. The first reviews the literature on the estimation of first order noninvertible moving average processes. The second derives a valid Edgeworth expansion for the ML estimator of a local level model.

(4.2) Estimating Noninvertible MA(1)s.

In an unpublished manuscript, Kang(1975) reported that in a simulation experiment she had found that a variety of commonly used estimators of first order moving average processes behaved strangely when $|\theta|$ was close to one, and that estimated boundary cases frequently cropped up. This point was reinforced by a battery of simulation studies that were conducted over the following few years by Cooper and Thompson(1977), Harvey and Phillips(1978) [reported in Harvey(1981, pg 136–9)], Dent and Min(1978), Ansley and Newbold(1980) and Davidson(1981a,b).

The work of Ansley and Newbold(1980) is perhaps the most revealing of these studies. They compared exact ML, exact least squares and conditional least squares estimators of a variety of models, reporting mainly on bias, mean square error and predictive ability. They found that ML estimators were generally more reliable than the other forms of estimators they looked at (see page 181). However, the ML estimator of the parameter in the MA(1) process was shown to have a large probability of occurring near the boundary. Indeed, of the 10,000 simulations they conducted, when the true value of θ was –0.9, 3,278 were between –0.99 and –1.0. This result seemed to confuse these authors as they argued that these values were not exactly minus one but were nearly this value, since "it can be shown that $\hat{\theta}$ (*MLE of* θ) will take the value –1.0 with probability zero." (My italics).

This assertion was shown to be false by Cryer and Ledoulter(1981), who derived the exact sampling distribution of the ML estimator when T = 2. This distribution was seen to be discontinuous at minus one and one, but continuous between these two points. They were also the first to compute the exact probability of observing a boundary case for any value of T.

The first analytic work on noninvertibility was carried out by

Sargan(1977). This was improved and generalized in the published paper of Sargan and Bhargava(1983b). In this they proved T–consistency of the ML estimator of θ when the true process was noninvertible, and derived the limiting probability of observing an estimated noninvertible process. Similar results were obtained in a regression model with MA(1) errors.

Pesaran(1983) produced a slightly simpler analysis of the above setup, but his work did not yield any significantly new results. Anderson and Takemura(1986)'s paper gave a neat and simple analysis of the ML estimation of noninvertible processes. They produced two advances. The first was a proof that the probability that a noninvertible process is estimated, when the true process is not invertible, is $o(T^{-n})$ when n is any integer. The second was to begin to set down a framework for the analysis of estimated noninvertibility in moving averages of a general order. The authors were less successful at the second of these tasks and seemed to have made little headway on this very difficult problem.

More recently, Tanaka and Satchell(1987) attempted to produce an approximation to the whole of the distribution of the ML estimator of θ, when the true process is noninvertible. Although they failed to give a useful approximation, their approach was interesting. They also analyzed a constrained moving average of a general order of the form

$$y_t = \epsilon_t + \theta\,\epsilon_{t-k}, \tag{4.5}$$

which is noninvertible if $|\theta| = 1$ (see figure 7.1 for a detailed look at the case of $k = 2$). They found that the probability of estimating a noninvertible process was sensitive to the choice of k, and tabulated the relevant large sample probabilities on the assumption that the true process was noninvertible. Although their analysis was interesting, it does not provide us with any direct help in our analysis of local trend models.

(4.3) A Valid Edgeworth Expansion for the Local Level Model

Little is known about the sampling distribution of the ML estimator of the parameters of even the simplest of the local trend models. The first work on this topic was carried out by Harvey and Peters(1984),who used the stationary form of the models (ie. they differenced the process until they were stationary) in the frequency domain. This allowed a reasonably simple derivation of the expected information in a sample of size T. Although their form of the information matrix is easy to implement on a computer, it does not significantly aid our understanding as it is a rather complicated trigonometric sum. Although they noted that the limit of these quantities, as the sample size goes to infinity, could be calculated analytically by evaluating some integrals, they were unable to carry out the required manipulations.

In this section we present some very simple results for the local level model. The first is a central limit theorem. This result allows us to study precisely the structure of the information matrix. The second result is an Edgeworth expansion for the ML estimation of the signal noise ratio when σ^2 is known. This enables us to describe the breakdown in the central limit theorem, as the true value of q approaches zero. Throughout, we will assume that (ϵ_t), (η_t) are Gaussian white noise processes, and we will write $T^* = T - 1$; q^* and σ^{*2} as the true values of q and σ^2 respectively and $u_t \sim NID(0, 1)$.

Using the results in chapter five, if $\mu_0 \sim N(0, \kappa)$, where $\kappa \to \infty$, we have that the probability law of the log–likelihood is, apart from the log κ,

$$\ell(1(\sigma^2,q)) = \ell\left[-\frac{T^*}{2} \log \sigma^2 - \frac{1}{2}\sum_{t=1}^{T^*} \log(1 + q\lambda_t) - \frac{1}{2}\sum_{t=1}^{T^*} \frac{\sigma^{*2} u_t^2 (1 + q^*\lambda_t)}{\sigma^2(1 + q\lambda_t)} \right]. \tag{4.6}$$

Then if the ML estimators, written $\hat{\sigma}^2$ and $\hat{q}$, are interior points on the permissible parameter space ($\hat{\sigma}^2$, $\hat{q} > 0$) and q^* is away from zero, we can derive the following result.

Theorem 4.1 If $\hat{\sigma}^2$, $\hat{q} > 0$ and $q^* > 0$, then

$$\sqrt{T^*}\begin{bmatrix} \hat{\sigma}^2 - \sigma^{*2} \\ \hat{q} - q^* \end{bmatrix} \xrightarrow{d} N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{2}{(q+2) - \sqrt{q(q+4)}} \begin{bmatrix} \sigma^4(q+2) & -\sigma^2 q(q+4) \\ -\sigma^2 q(q+4) & (q(q+4))^{3/2} \end{bmatrix}\right] \tag{4.7}$$

Proof. Given in the appendix.

As $q \to 0$, for fixed T^*, the variance of the scaled (by root T^*) ML estimator of q contracts to zero, while the variance of $\hat{\sigma}^2$ converges to $2\sigma^4$. The correlation between $\hat{\sigma}^2$ and $\hat{q}$ is

$$\frac{-q(q+4)}{\sqrt{(q+2)(q(q+4))^{3/2}}} = \frac{-\sqrt[4]{q(q+4)}}{\sqrt{q+2}}. \tag{4.8}$$

(4.8) goes to zero as $q \to 0$, while it converges to minus one as $q \to \infty$.

More information about the sampling behaviour of the ML estimators can be gleaned from performing an Edgeworth expansion. Theorem 4.2 provides this under the assumption that σ^2 is known. This assumption is made in order to simplify the work, and to allow us to concentrate on the estimation of the signal noise ratio, q, which is the major issue of this thesis.

Theorem 4.2 If q, $q^* > 0$ then writing $g(\hat{q};q)$ as the density of $\hat{q}$ for a given true value of q, we have that

$$g(\hat{q};\, q) = \left[\frac{T^* A(q)}{2\pi\,.\,2}\right]^{1/2}.\exp\left[-\frac{T^* A(q)}{2\,.\,2}(\hat{q}-q)^2\right]$$

$$.\left[1+\frac{K(q)}{\sqrt{T^*}}.\left[-\frac{\sqrt{T^*}(\hat{q}-q)}{A(q)}+\frac{1}{6}(\sqrt{T^*}(\hat{q}-q))^3\right]\right](1+O(T^{*-1})).$$

(4.9)

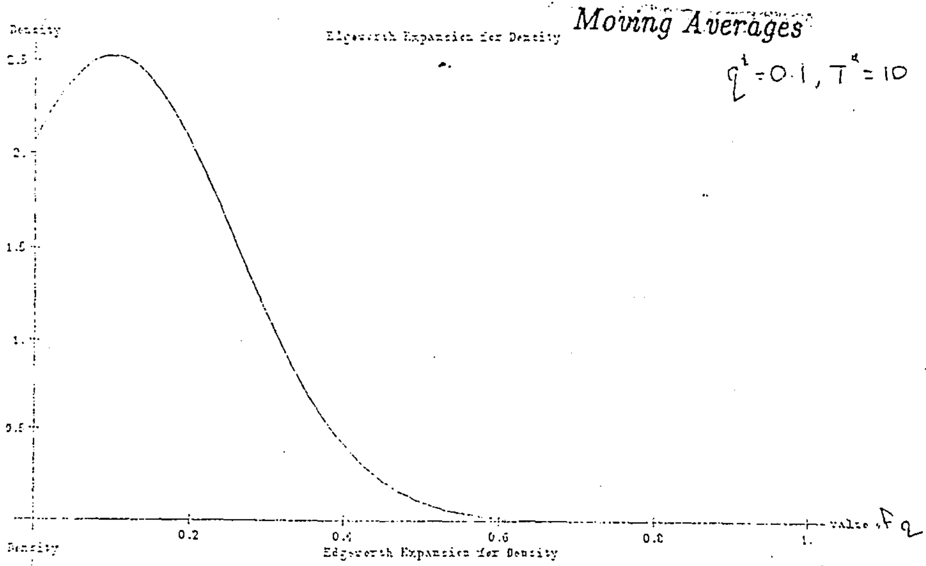
$$\text{where } A(q) = \frac{(q+2)}{(q(q+4))^{3/2}}, \text{ and } K(q) = \frac{6\ +\ q(q+4)}{(q(\,q+4)\,)^{5/2}}.$$

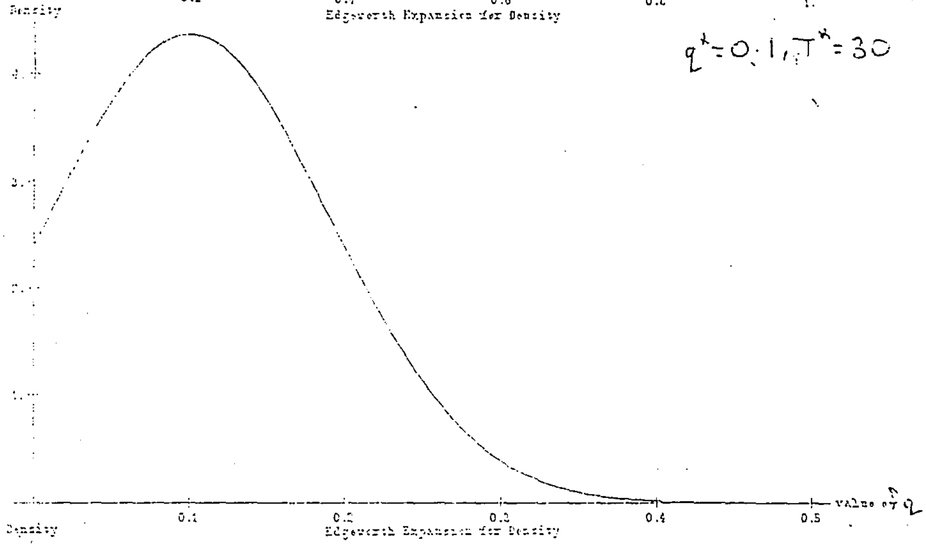
Proof. Given in the appendix.

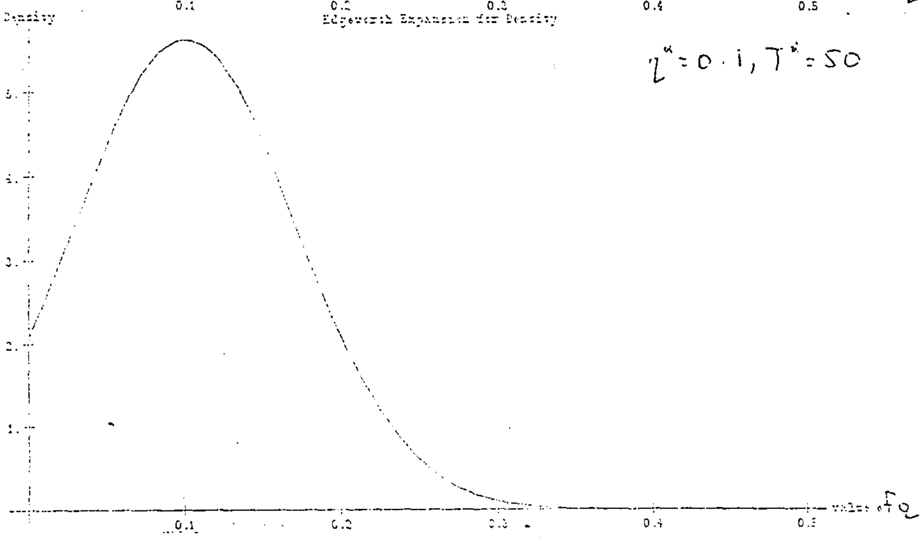
The important feature of this expression is K(q) which is positive for all q. This means that $\hat{q}$ is positively skewed. Further, as $q \to 0$, this skewness explodes, while for $q \to \infty$ it goes to zero.

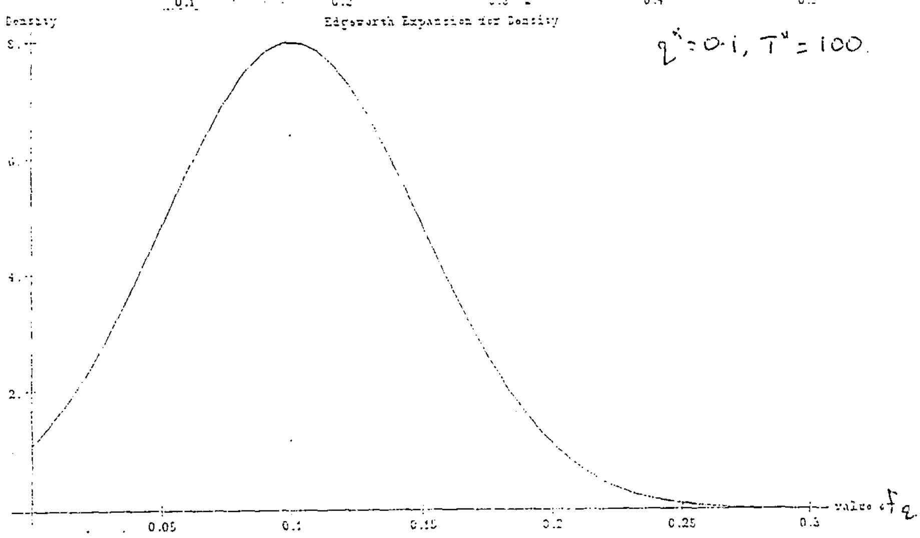
Examples of the density are given in the diagrams 4.1, 4.2 and 4.3, which are displayed below. In the first diagram we keep $q = 0.1$ and allow T^* to increase through the values 10, 30, 50 and 100. We do the same operation in the other two diagrams, with the second diagram having $q = 0.5$, while the final diagram has $q = 1.0$.

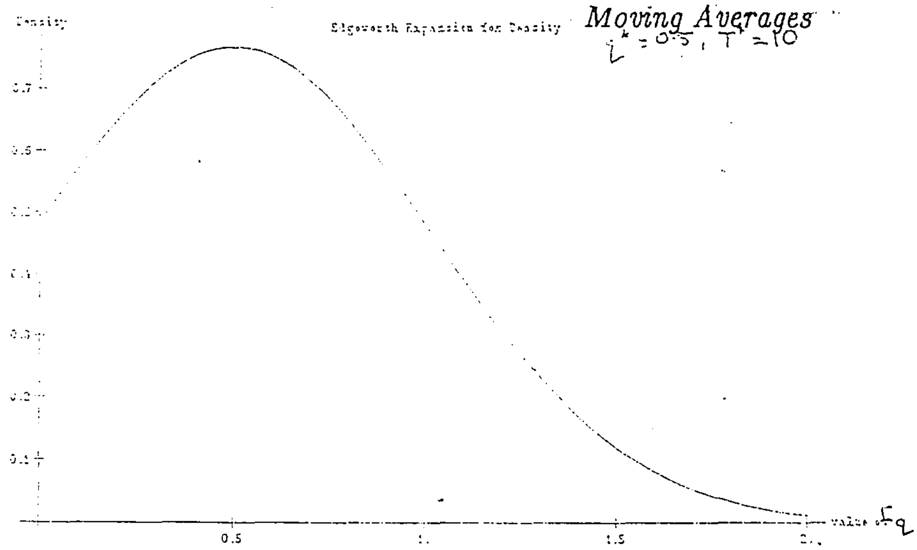
The diagrams indicate that the first order approximation is very accurate if q is significantly away from zero and T^* is 30 or more. Smaller values of q will mean this approximation will be worse, unless the sample size increases significantly.

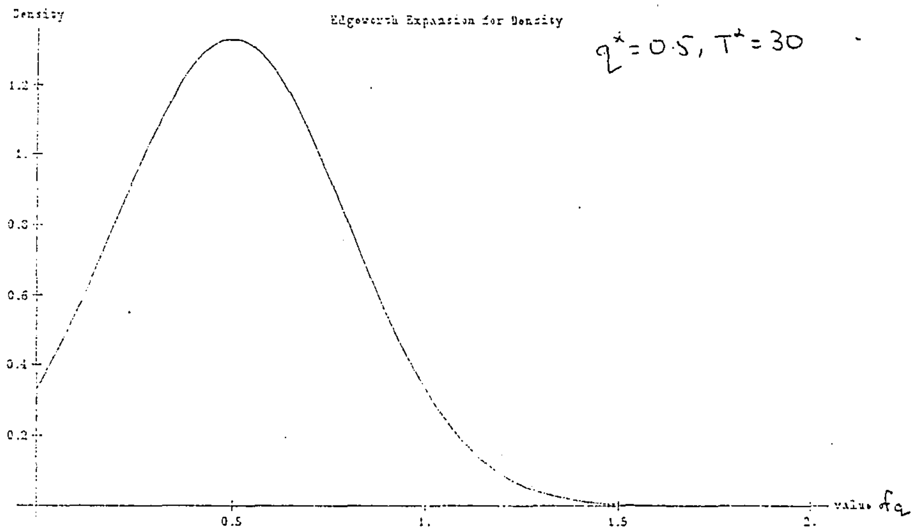

Density
Edgeworth Expansion for Density
$q^* = 0.1,\ T^* = 10$
value of q
0.2
0.4
0.6
0.8
1.

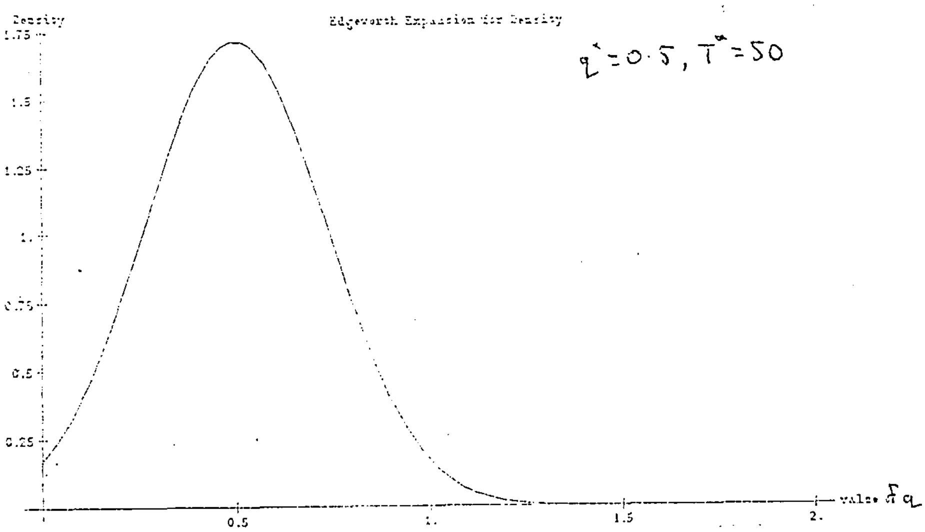

Density
Edgeworth Expansion for Density
$q^* = 0.1,\ T^* = 30$
value of q
0.1
0.2
0.3
0.4
0.5

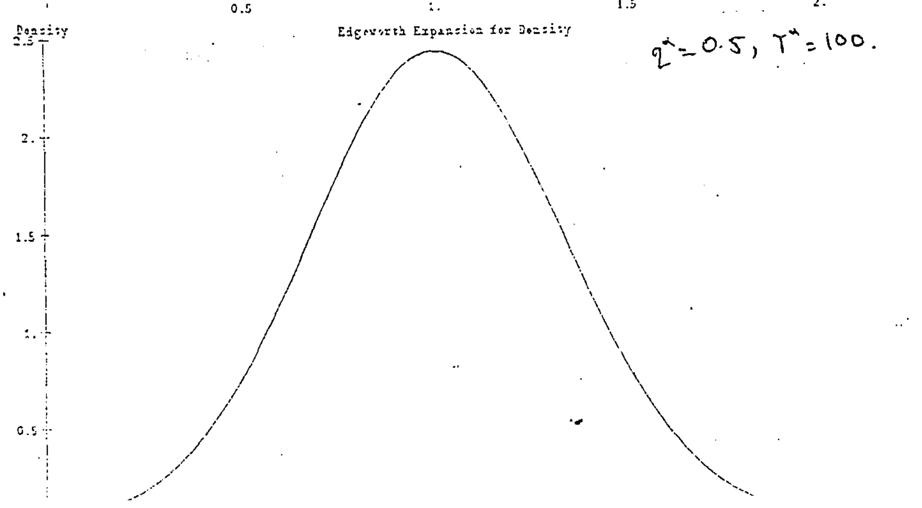

Density
Edgeworth Expansion for Density
$q^* = 0.1,\ T^* = 50$
value of q
0.1
0.2
0.3
0.4
0.5

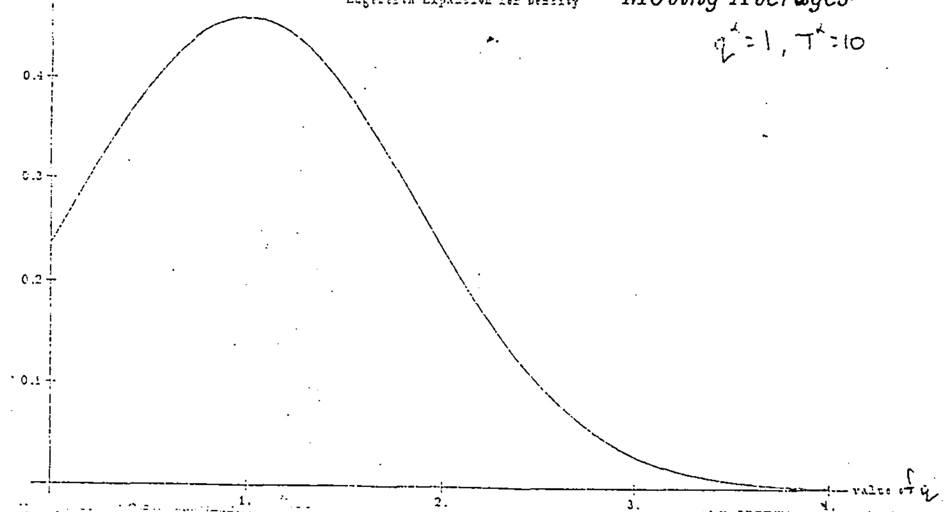

Density
Edgeworth Expansion for Density
$q^* = 0.1,\ T^* = 100$
value of q
0.05
0.1
0.15
0.2
0.25
0.3

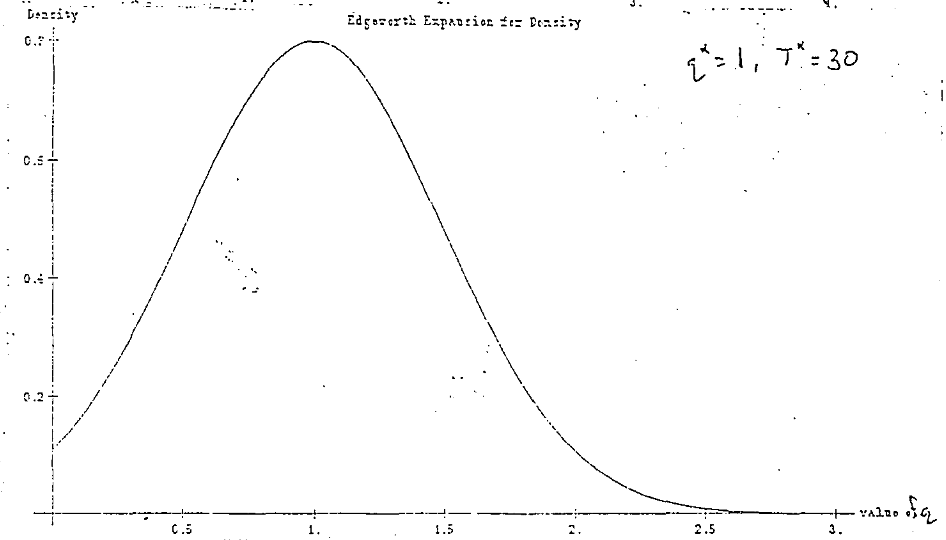

Density
Edgeworth Expansion for Density
$q^* = 0.5,\ T^* = 10$
value of q
0.5
1.
1.5
2.

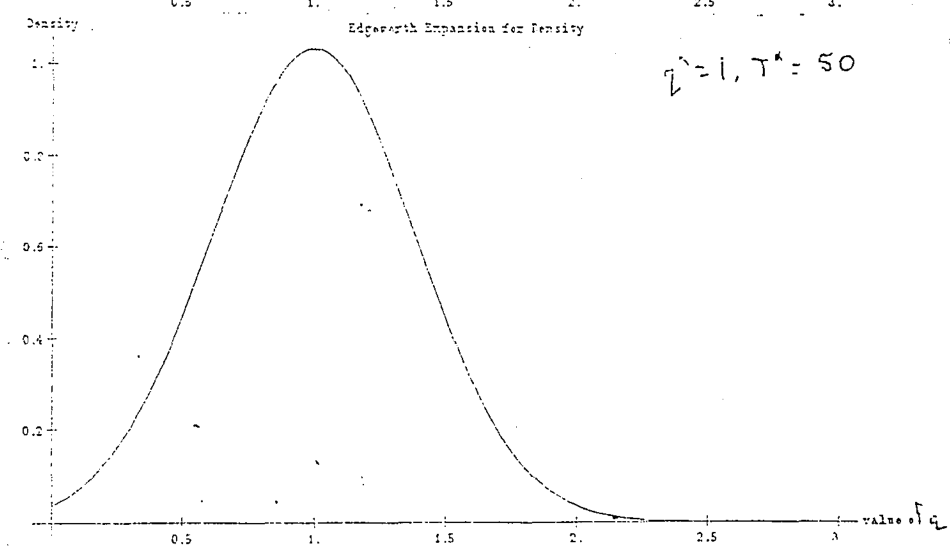

Density
Edgeworth Expansion for Density
$q^* = 0.5,\ T^* = 30$
1.2
1.
0.8
0.6
0.4
0.2
value of q
0.5
1.
1.5
2.

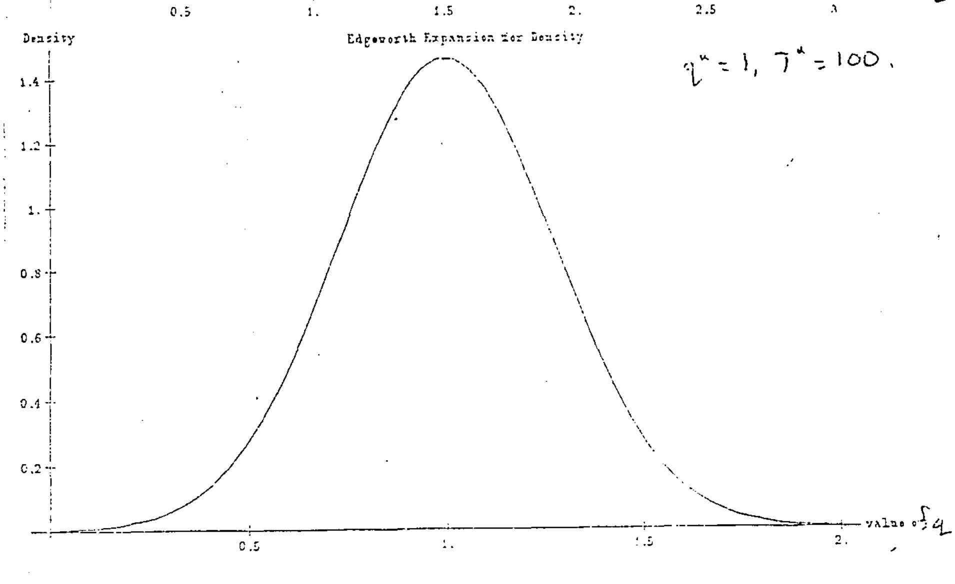

Density
Edgeworth Expansion for Density
$q^* = 0.5,\ T^* = 50$
1.75
1.5
1.25
1.
0.75
0.5
0.25
value of q
0.5
1.
1.5
2.


Density
Edgeworth Expansion for Density
$q^* = 0.5,\ T^* = 100.$
2.5
2.
1.5
1.
0.5


Density
Edgeworth expansion for Density
$q^*=1,\ T^*=10$
0.4
0.3
0.2
0.1
1.
2.
3.
4.
value of q


Density
Edgeworth Expansion for Density
$q^*=1,\ T^*=30$
0.5
0.4
0.2
0.5
1.
1.5
2.
2.5
3.
value of q


Density
Edgeworth Expansion for Density
$q^*=1,\ T^*=50$
1.
0.6
0.4
0.2
0.5
1.
1.5
2.
2.5
value of q


Density
Edgeworth Expansion for Density
$q^*=1,\ T^*=100.$
1.4
1.2
1.
0.8
0.6
0.4
0.2
0.5
1.
1.5
2.
value of q

Mathematical Appendix.

Proof of Theorem 4.1 All the derivatives of the log–likelihood exist (if $q^* > 0$) and so all we need to compute is the matrix of second derivatives (c.f. Cox and Hinkley(1974, pg 297)). But

$$\frac{\partial^2 \ell(\log L)}{\partial \sigma^{2^2}} = \ell\left[\frac{T^*}{2\sigma^4} - \sum_{t=1}^{T^*} \frac{\sigma^{*2} u_t^2 (1 + q^* \lambda_t)}{\sigma^6 (1 + q\lambda_t)}\right] \tag{A4.1}$$

$$\frac{\partial^2 \ell(\log L)}{\partial \sigma^2 \, \partial q} = \ell\left[-\frac{1}{2} \sum_{t=1}^{T^*} \frac{\sigma^{*2} u_t^2 (1 + q^* \lambda_t)\lambda_t}{\sigma^4 (1 + q\lambda_t)^2}\right] \tag{A4.2}$$

$$\frac{\partial^2 \ell(\log L)}{\partial q^2} = \ell\left[-\sum_{t=1}^{T^*} \frac{\sigma^{*2} u_t^2 (1 + q^* \lambda_t)\lambda_t^2}{\sigma^2 (1 + q\lambda_t)^3} + \frac{1}{2} \sum_{t=1}^{T^*} \frac{\lambda_t^2}{(1 + q\lambda_t)^2}\right]. \tag{A4.3}$$

Evaluating q and σ^2 at their true values, we can look at the limits of these quantities using the result that

$$\frac{1}{T^*} \sum_{t=1}^{T^*} \frac{\lambda_t^r}{(1 + q\lambda_t)^r} = \frac{1}{\pi} \int_0^{\pi} \frac{1}{(4\sin^2 \lambda/2 + q)^r} \, d\lambda + O(T^{*-1}) \tag{A4.4}$$

(c.f. Grenander and Szego(1958, pg 221, equation 7) or Hannan(1970, pg 353))

$$= \frac{(-1)^{r-1}}{(r-1)!} \cdot \frac{d^{r-1} (q(q+4))^{-1/2}}{dq^{r-1}} + O(T^{*-1}). \tag{A4.5}$$

(*Computer algebra is useful to evaluate correctly these types of terms– c.f. REDUCE (see Rayna(1987) or Hearn(1985)) and MATHEMATICA (see Wolfram(1988)))*. Hence, the asymptotic covariance matrix is

$$\frac{2\sigma^4}{T^*}\begin{bmatrix} \sigma^{-4} & \dfrac{1}{\sigma^2\sqrt{q(q+4)}} \\ \dfrac{1}{\sigma^2\sqrt{q(q+4)}} & \dfrac{(q+2)}{(q(q+4))^{3/2}} \end{bmatrix}^{-1} \tag{A4.6}$$

$$= \frac{2\sigma^4}{T^*}\cdot\frac{(q(q+4)^{3/2}}{(q+2)-\sqrt{q(q+4)}}\cdot\begin{bmatrix} \dfrac{(q+2)}{(q(q+4))^{3/2}} & -\dfrac{1}{\sigma^2\sqrt{q(q+4)}} \\ -\dfrac{1}{\sigma^2\sqrt{q(q+4)}} & \sigma^{-4} \end{bmatrix} \tag{A4.7}$$

$$= \frac{2\sigma^4}{T^*}\frac{1}{(q+2)-\sqrt{q(q+4)}}\begin{bmatrix} (q+2) & \dfrac{-q(q+4)}{\sigma^2} \\ \dfrac{-q(q+4)}{\sigma^2} & \dfrac{(q(q+4))^{3/2}}{\sigma^4} \end{bmatrix}\Box. \tag{A4.8}$$

Proof of Theorem 4.2

Write $\phi_{T^*}(z,q) = E \exp(izT^*\hat{q})$, and assume

(A4.9)

A1. Where they exists, the j–th cumulant $K_{T^*j}(q)$ of $T^*\hat{q}$, satisfies the relationship

$$i^j K_{T^*j}(q) = d^j \log \phi_{T^*j}(0,q) / dq^j,\ j = 1, \ldots, 3.$$

(A4.10)

A2. For all integers T^* greater than T_1, $T^*\hat{q}$ has cumulants of order up to and including three, and they are continuous functions of q in an open subset of R_+ containing q^*.

Abril(1985) demonstrated that A1 is stronger than Durbin's(1980) assumption 3, while A2 implies assumption 4. Durbin's third assumption is difficult to check directly, but Abril shows that A1 and A2 imply that it holds. This means Durbin's valid Edgeworth expansion for the ML estimator can be used (his theorem one). Abril also showed that A1 and A2 are stronger than the conditions used by Taniguchi(1984) to prove the validity of the expansion for the distribution function.

If $q^* > 0$, then for $q > 0$ both A1 and A2 hold by using the arguments of Shenton and Bowman(1977) or Peers(1978) or Peers and Iqbal(1985), who showed that the cumulants of $T^*\hat{q}$ are related to the expected value of various derivative of the log–likelihood function. We require q to be strictly positive so that we satisfy the open subset requirement in A2.

Having satisfied Durbin's(1980) assumptions 2–4, we can use his theorem one, which states that for a scalar case

$$g(\hat{q};\, q) = \left[\frac{T^{*}I(q)}{2\pi}\right]^{1/2} . \exp\left[-\frac{T^{*}I(q)}{2}(\hat{q}-E\hat{q})^2\right] . \left[1 + \frac{I^{3/2}(q) . K_{T^{*}3}(q)}{6.T^{*3/2}} .\right.$$

$$\left.\left[I^{3/2}(q)T^{*3/2}(\hat{q}-E\hat{q})^3 - 3I^{1/2}(q)T^{*1/2}(\hat{q}-E\hat{q})\right]\right](1 + O(T^{*-1})).$$

(A4.11)

where $I(q)$ is the limiting mean information in an observation, while $K_{T^{*}3}(q) = ET^{*3}(\hat{q}-E\hat{q})^3$. Note that $K_{T^{*}3}(q)$ measure the skewness of $\hat{q}$.

Following Abril we will use Peers(1978) to produce a tractable expression for I and K. He derives the equalities

$$E\,\hat{q} = q + \left[\frac{1}{2}E\left[\frac{d^3 \log L}{dq^3}\right] + E\left[\frac{d\log L}{dq}\frac{d^2\log L}{dq^2}\right]\right]\left[E\,\frac{d^2 \log L}{dq^2}\right]^{-2} + O(T^{*-2}),$$

(A4.12)

$$I_{T^{*}}(q^{*}) = -\,T^{*-1}\left[E\,\frac{d^2\log L}{dq^2}\right] + O(T^{*-1}), \qquad \textbf{(A4.13)}$$

$$K_{T^{*}}(q^{*}) = -T^{*3}\left[2.E\left[\frac{d^3\log L}{dq^3}\right],\right.$$

$$\left. + \, 3.E\left[\frac{d\log L}{dq}\frac{d^2\log L}{dq^2}\right]\right]\left[E\,\frac{d^2\log L}{dq^2}\right]^{-3} + O(1)$$

(A4.14)

where $I_{T^{*}}(q)$ is the mean information in an observation.

Although we are conditioning on q > 0, we do not have to adjust the expectations to take this into account because, following Anderson and Takemura's(1986) work on MA(1)s, we show in chapter five that

$$p(\hat{q} = 0 | q^{*} > 0) = o(T^{*-n}) \text{ for n being any integer.} \tag{A4.15}$$

Hence the validity of the Edgeworth expansion is not affected by our ignoring this conditioning.

Noting that

$$E\left[\frac{d^3 \log L}{dq^3}\right] = 2\sum_{t=1}^{T^*} \frac{\lambda_t^3}{(1+q\lambda_t)^3}, \text{ and } E\left[\frac{d\log L}{dq}\frac{d^2\log L}{dq^2}\right] = -\sum_{t=1}^{T^*} \frac{\lambda_t^3}{(1+q\lambda_t)^3}, \tag{A4.16}$$

and using the result given in equation (A4.4), we have that

$$E\hat{q} = q + O(T^{*-2}), \tag{A4.17}$$

$$I_{T^*}(q) = \frac{A}{2} + O(T^{*-1}), \tag{A4.18}$$

$$K_{T^*3}(q) = -T^{*3}\left[4\,T^*K - 3\,T^*K\right]\left[-\frac{T^*}{2}A\right]^{-3} + O(1). \tag{A4.19}$$

$$= \frac{8T^*K}{A^3} + O(1) \tag{A4.20}$$

Noting the higher order unbiasedness of $\hat{q}$ gives the result immediately.□

Chapter Five

On the Probability of Estimating a Deterministic Component in the Local Level Model

Summary. A local level model has a deterministic level when the signal noise ratio, written q, is zero. This chapter investigates the properties of the maximum likelihood estimator of q, paying particular attention to the case where the true value of q is zero. These properties are shown to be crucially dependent on the initial conditions employed.

Key Words. BOUNDARY ESTIMATES, NONINVERTIBILITY, CHARACTERISTIC FUNCTION, INITIAL CONDITIONS, STRUCTURAL TIME SERIES MODELS.

(5.1) Introduction and Summary.

In studies of least–squares estimates of autoregressive models many writers, for example Phillips(1987), have found their results to be very sensitive to the initial conditions, when the models are nearly non–stationary. As a result of this, it is very natural to enquire about the effect the initial conditions have on the maximum likelihood (ML) estimator of the Gaussian local level model

$$y_t = \mu_t + \epsilon_t \ , \ \epsilon_t \sim \text{NID}(0,\sigma^2) \qquad ,t=1,2,\dots,T \qquad \textbf{(5.1a)}$$
$$\mu_t = \mu_{t-1} + \eta_t \ , \ \eta_t \sim \text{NID}(0,q\sigma^2) \qquad \textbf{(5.1b)}$$

where $(\epsilon_t),(\eta_t)$ are serially and mutually independent.

Muth(1960) showed that this model provides a rationale for exponentially weighted moving averages. Subsequently it has proved to be the basis of more complicated systems, such as the structural time series models suggested by Harvey and Durbin(1986), and the parameter variation models of Cooley and Prescott(1973,1976).

The reduced form of the local level model is an ARIMA(0,1,1) process. When q, the signal noise ratio, goes to zero, the model has a deterministic component, while its reduced form becomes strictly noninvertible. Hence, following the work of Kang(1975), Sargan and Bhargava(1983b), Anderson and Takemura(1986) and Tanaka and Satchell(1987) on MA processes, we may expect the ML estimator to behave in an unusual way when q is near to zero, and to be very sensitive to the specification of the initial conditions for the level (μ_t).

There are three important startup procedures which have received substantial attention in the literature:–

A1. The initial level μ_0 is a fixed, known constant.

A2. $\mu_0 \sim N(0, \kappa)$ $\kappa \to \infty$ where μ_0 is distributed independently from $(\epsilon_t),(\eta_t)$. This diffuse prior has been commonly used in non–stationary state space models, c.f. de Jong(1988).

A3. μ_0 is a fixed, unknown constant. This particular assumption has been used extensively, c.f. Cooley and Prescott(1973,1976), Rosenberg(1973), Nyblom(1986) and Shively(1988b).

In this chapter we study the probability that a (local) ML estimator of q is exactly zero for various true values of q, written q^*, for the local level model under these three startup procedures. We also derive consistency rates for the mle when $q^*=0$. Sections two and three show that the results for the fixed and known startup and the diffuse prior are not too different. However, in section four, we demonstrate that the sampling distribution of the ML estimator will change dramatically when we specify a fixed but unknown startup procedure. Finally, in section five, we use a simulation study to show that these results carry over to the case of the (global) ML estimator.

(5.2) A Fixed and Known Starting Value.

Back substituting into the system, we can write (y_t) as

$$y_t = \mu_0 + \eta_1 + \eta_2 + \ldots + \eta_t + \epsilon_t \tag{5.2}$$

Writing i as a T×1 vector of ones, $y=(y_1,\ldots,y_T)'$, C as a T×T matrix with typical element $((\delta_{i,j}))$, where $\delta_{i,j}$ takes the value one if $i \geq j$ and zero elsewhere, and d is a generic constant, we can express the log–likelihood as

$$l(\sigma^2,q)=d-\frac{T}{2}\log\sigma^2-\frac{1}{2}\log|I+qA|-\frac{(y-\mu_0 i)'(I+qA)^{-1}(y-\mu_0 i)}{2\sigma^2}$$

(5.3)

where $A=CC'$. So concentrating l at

$$\hat{\sigma}^2(q)=\frac{(y-\mu_0.i)'(I+qA)^{-1}(y-\mu_0.i)}{T}, \qquad \textbf{(5.4)}$$

which is positive with probability one (wp1), yields the profile log–likelihood, written M(q)

$$M(q)=d-\frac{T}{2}\log(y-\mu_0.i)'(I+qA)^{-1}(y-\mu_0.i)-\frac{1}{2}\log|I+qA|.$$

(5.5)

Although M(q) only has statistical meaning for $q\geq 0$, it is continuous for small q approaching zero from below, so that the score function, which is the derivative of the profile log–likelihood, exists over the whole non–negative real line, wp1, and is given by

$$s(q)=\frac{T}{2}\frac{(y-\mu_0.i)'(I+qA)^{-1}A(I+qA)^{-1}(y-\mu_0.i)}{(y-\mu_0.i)'(I+qA)^{-1}(y-\mu_0.i)}-\frac{1}{2}\mathrm{tr}((I+qA)^{-1}A)$$

(5.6)

But what does the score function actually tell us? Think of some of the typical shapes M(q) might take when mapped against q. Perhaps the three most important are given below.

Figure 5.1

Stylised Shapes of M(q) Against q

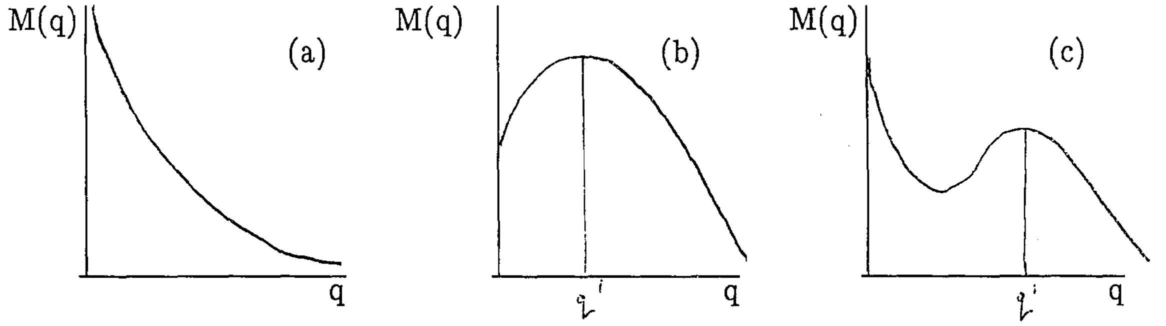


Figure 5.1(a) and 5.1(c) implies a local maximum occurs at q=0 iff s(0)<0, while 5.1(b) and 5.1(c) tell us that maxima occur at $q'>0$ iff, for small $\epsilon>0$, $s(q'+\epsilon)<0$. These ideas will be central in our development of the sampling distribution of the ML estimator of q. Figure 5.1(c) will be of particular importance when we come to study global ML estimator in section 5.5.

Rutherford(1946) showed that the eigenvalues of A are given by

$$\delta_t = \frac{1}{4\sin^2\frac{\pi(t-0.5)}{2T+1}} \quad (t=1,2,\ldots,T), \tag{5.7}$$

so noting $y-\mu_0.i \sim N(0,\sigma^2(I+q^*A))$, we can see that the probability law of the score is

$$\ell\left[\frac{\frac{T}{2}\sum_{t=1}^{T}\frac{u_t^2\,\delta_t\,(1+q^*\delta_t)}{(1+q\,\delta_t)^2}}{\sum_{t=1}^{T}\frac{u_t^2(1+q^*\delta_t)}{(1+q\delta_t)}}-\frac{1}{2}\sum_{t=1}^{T}\frac{\delta_t}{(1+q\delta_t)}\right] \tag{5.8}$$

where $u_t \sim NID(0,1)$. If we write the probability of observing a local maximum at zero, for a given q^*, as $p(q^*)$, then using figure 5.1(a)

$$p(q^*) = \Pr\left[\sum_{t=1}^{T} u_t^2\,(1+q^*\delta_t)(\delta_t-\frac{(T+1)}{2}) < 0\right] \tag{5.9}$$

as $\Sigma\,\delta_t = \text{tr}\,A = (T+1)T/2$.

This probability can be evaluated for any T by the Imhof(1961) procedure. The results of these calculations are reported in table 5.1 given below.

Table 5.1

$p(q^*)$ for the Local Level Model, using a Known, Fixed Start-up.

True Value of q

T	0	0.01	0.1	1	10
10	.654	.572	.329	.146	.099
20	.666	.447	.180	.058	.034
30	.670	.343	.109	.028	.016
40	.672	.270	.070	.015	.008
50	.673	.216	.047	.009	.005

A special case is when $q^*=0$. Allowing $T\to\infty$ we can enforce some simplifications on the above expression as $\delta_t^{-1} = \frac{[t-0.5]^2\pi^2}{T^2} + O(T^{-4})$. This yields

$$\lim_{T\to\infty} p(0) = \Pr\left[\sum_{t=1}^{\infty} \frac{u_t^2}{(t-1/2)^2\pi^2} < 1/2\right]. \tag{5.10}$$

The random variables' characteristic function can be shown to be $1/\sqrt{\cos\sqrt{2i\theta}}$. By analytically inverting this function Nabeya and Tanaka(1988) demonstrated that its distribution function given by

$$F(x) = 2\sqrt{2}\sum_{j=0}^{\infty} \binom{-1/2}{j} \Phi\left[\frac{-2j-1/2}{\sqrt{x}}\right] \tag{5.11}$$

where $\Phi(.)$ is the standard normal distribution. We deduced the required limiting probability by using this expression with $x=1/2$. It was found to be 0.6778.

Finally, in this section, we derive the rate of convergence of the ML estimate to zero. Recall the figure 5.1. These diagrams imply the likelihood will be maximized, (perhaps a local maximum though), in the region [0,r) if the gradient of the likelihood is negative at the point r. It thus becomes important to look at the score function in deriving the consistency of the ML estimate. This is done in the proof of the following theorem. Our method of proof is similar in style to that used by Tanaka and Satchell(1987, proposition one) in their analysis of strictly non–invertible MA(1) processes.

<u>Theorem 5.1.</u> The ML estimator of q is $O_p(T^{-2})$ when $q^*=0$, ie. for any $\epsilon>0$, $\exists\ T_0$ and $c>0$ such that

$$\Pr(s(c/T^2;0) \geq 0\ ) < \epsilon,\ \forall\ T > T_0.$$

<u>Proof.</u> Remember, on $q^*=0$ we are concerned with the probability law of the score which we write as

$$\ell\left[\frac{1}{T^2}\, s\left(\frac{c}{T^2};0\right)\right]$$

$$= \ell\left[\frac{\frac{1}{2}\sum_{t=1}^{T} \frac{u_t^2\ T^2/\delta_t}{(c+T^2/\delta_t)^2}}{\frac{1}{T}\sum_{t=1}^{T} \frac{u_t^2\ T^2/\ \delta_t}{(c+T^2/\delta_t)}} - \frac{1}{2}\sum_{t=1}^{T}\frac{1}{(c+T^2/\delta_t)}\right]. \tag{5.12}$$

By employing a weak law of large numbers on the denominator of the first term we have the limiting form of the scaled score

$$\lim_{T\to\infty} \ell\left[\frac{1}{T^2}\, s\left(\frac{c}{T^2};0\right)\right] \to \ell\left[\frac{1}{2}\sum_{t=1}^{\infty} \frac{u_t^2\ [t-0.5]^2\pi^2}{(c+[t-0.5]^2\pi^2)^2} - \frac{1}{2}\sum_{t=1}^{\infty} \frac{1}{(c+[t-0.5]^2\pi^2)}\right]. \tag{5.13}$$

Writing X(c) to denote this limiting random variable, and using the moments of chi–squared random variables, we can see that

$$\mathrm{EX(c)} = \frac{1}{2}\sum_{t=1}^{\infty} \frac{-c}{(c+[t-0.5]^2\pi^2)^2} < 0, \tag{5.14}$$

$$\mathrm{VX(c)} = \frac{1}{2}\sum_{t=1}^{\infty} \frac{[t-0.5]^4\pi^4}{(c+[t-0.5]^2\pi^2)^4} \tag{5.15}$$

Employing Chebyshev's inequality we have

$$\lim_{T\to\infty} \Pr\left[\frac{1}{T^2}\, s(\frac{c}{T^2};0) \geq 0\right] < \frac{\mathrm{VX(c)}}{[\mathrm{EX(c)}]^2}$$

$$= \frac{2}{c^2}\, \frac{\sum_{t=1}^{T} \frac{[t-0.5]^4\,\pi^4}{(c+[t-0.5]^2\pi^2)^4}}{\left[\sum_{t=1}^{T} \frac{1}{(c+[t-0.5]^2\pi^2)^2}\right]^2} = \frac{g(c)}{c^2}. \tag{5.16}$$

The function g must be bounded for all values of c, which means $\exists$ b such that $\forall$ c $g(c) \leq b$, so setting $c = \sqrt{b/\epsilon}$ gives the desired result immediately.□

This result should not surprise us as we know that the ML estimator of the MA(1) coefficient, θ, is $\pm 1 + O_p(T^{-1})$, (see Sargan and Bhargava(1983b)), when the true value of $\theta = \pm 1$ and that q is related to θ quadratically.

(5.3) The Diffuse Prior.

Although the likelihood for this type of model is usually constructed via the prediction error decomposition, (c.f. Harvey(1984)), our analysis requires it to be explicitly expressed in terms of the parameters of interest. To this end we write H_t as the information set available at time t, thus allowing the diffuse prior to be written as $\mu_0 | H_0 \sim N(0,\kappa)$ so that when $\kappa \to \infty$, $\mu_1 | H_1 \to N(y_1,\sigma^2)$. Ignoring the log κ term in the likelihood (see de Jong(1988)), the likelihood simplifies to $L(\sigma^2,q;y_2,\ldots,y_T | H_1)$ as $\kappa \to \infty$, and so the likelihood for $(y_1,\ldots,y_T)$ is the same as that for the vector random variable v, defined by $v = y - y_1.i$, where i and y are $(T-1)\times 1$ vectors such that $i=(1,\ldots,1)'$ and $y=(y_2,\ldots,y_T)'$.

Back substituting into the local level yields

$$y_1 = \mu_1 + \epsilon_1 \tag{5.17a}$$

$$y_t = \mu_1 + \epsilon_t + \eta_2 + \ldots + \eta_t \tag{5.17b}$$

Thus $v \sim N_{T-1}(0,\sigma^2(I+ii'+qA))$. The log–likelihood is

$$l(\sigma^2,q) = d - \frac{(T-1)}{2}\log\sigma^2 - \frac{1}{2}\log|I+ii'+qA| - \frac{v'(I+ii'+qA)^{-1}v}{2\sigma^2}. \tag{5.18}$$

Concentrating σ^2 out of the likelihood function yields the following profile log–likelihood

$$M(q) = d - \frac{(T-1)}{2}\log v'(I+ii'+qA)^{-1}v - \frac{1}{2}\log|I+ii'+qA|, \tag{5.19}$$

with the corresponding score

$$s(q) = \frac{(T-1)}{2}\frac{v'(I+ii'+qA)^{-1}A(I+ii'+qA)^{-1}v}{v'(I+ii'+qA)^{-1}v} - \frac{1}{2}\mathrm{tr}\left[(I+ii'+qA)^{-1}A\right]. \tag{5.20}$$

Noting that $v \sim N(0, \sigma^2(I+ii'+q^* A))$, theorem 5.2 of the appendix , and the properties of the eigenvalue solution, the probability law of the score can be written as

$$\ell\left[\frac{\frac{(T-1)}{2}\sum_{t=1}^{T-1}\frac{u_t^2\lambda_t\,(1+q^*\lambda_t)}{(1+q\lambda_t)^2}}{\sum_{t=1}^{T-1}\frac{u_t^2(1+q^*\lambda_t)}{(1+q\lambda_t)}} - \frac{1}{2}\sum_{t=1}^{T-1}\frac{\lambda_t}{(1+q\lambda_t)}\right], \tag{5.21}$$

where $u_t \sim NID(0,1)$ and $\lambda_t = \frac{1}{4\sin^2\left[\frac{t\pi}{2T}\right]}$, $t=1,2,\ldots,T-1$. The proof of this is straightforward but messy. The easiest part is showing that $tr\left[(I+ii'+qA)^{-1}A\right] = tr\left[(A^{-1}(I+ii')+qI)^{-1}\right]$ to give the second term. The denominator of the first term follows using the same kind of argument, while the numerator is simply its derivative with respect to q.

Expression (5.21) implies

$$p(q^*) = Pr(s(0;q^*) < 0) = Pr\left[\sum_{t=1}^{T-1} u_t^2\left(\lambda_t - \frac{(T+1)}{6}\right)(1+q^*\lambda_t) < 0\right] \tag{5.22}$$

as $tr\left[(I+ii')^{-1}A\right] = \frac{(T-1)(T+1)}{6}$. The Imhof procedure was used to calculate this probability and the results are presented in table 5.2.

Table 5.2

Tabulation of the Probability of a Local Maximum Occuring at q=0 for a Diffuse Prior.

True Value of q

T–1	0	0.01	0.1	1	10
10	.637	.612	.468	.215	.116
20	.647	.559	.287	.075	.032
30	.650	.487	.176	.030	.011
40	.652	.414	.110	.013	.004
50	.653	.349	.070	.006	.002

If we compare these results with those given in table 5.1 and diagram 5.2, we see that they are slightly smaller for $q^*=0$ and tend to be larger for $q^*>0$.

In the case where q^* is exactly zero, the outcome is particularly interesting. Then p(0) is exactly analogous to the probability of obtaining a strictly non–invertible MA parameter in an overdifferenced MA(1) model (see Sargan and Bhargava(1983b) corollary one, when corrected). This is not surprising, since the likelihood function for the local level model corresponds to the exact likelihood function for an MA(1), applied to first differences.

As $T\to\infty$ there are some simplifications which can be enforced on p(0). Since $\lambda_t^{-1}=\dfrac{t^2\pi^2}{T^2}+O(T^{-4})$ we can use the weak law of large numbers to see that

$$\lim_{T\to\infty} p(0) = \Pr\left[\sum_{t=1}^{\infty}\frac{u_t^2}{t^2\pi^2}<\frac{1}{6}\right]. \tag{5.23}$$

The distribution function of the random variable $\Sigma u_t^2/t^2\pi^2$ has been tabulated by

Anderson and Darling(1952). Using their tables, we deduced that the required limiting probability is 0.6574, which is not very different from the case considered in section two.

We can also prove that the mle of q is T^2–consistent when $q^*=0$. The proof of this proposition follows exactly the same lines as for theorem 5.1 when one recalls that $\lambda_t^{-1} = \frac{t^2\pi^2}{T^2} + O(T^{-4})$, so giving the limiting score

$$\lim_{T\to\infty} \ell\left[\frac{1}{T^2} s(\frac{c}{T^2}\,;\,0)\right] \to \ell\left[\frac{1}{2}\sum_{t=1}^{\infty} \frac{u_t^2 \; t^2\pi^2}{(c+t^2\pi^2)^2} - \frac{1}{2}\sum_{t=1}^{\infty} \frac{1}{(c+t^2\pi^2)}\right]. \tag{5.24}$$

Another interesting property of the ML estimator is the rate at which $p(q^*)$ goes to zero when $q^* = 0$. This problem was first tackled for the MA(1) case by Anderson and Takemura(1986) and their work is carried over to our case in the proof of the following theorem.

Theorem 5.3 Define $\epsilon>0$ and n as any fixed constants. Then $\exists\, T_o$ such that $\forall\, T>T_o$

$$T^n\, p(q^*) < \epsilon, \text{ when } q^* > 0.$$

Proof. Given in the appendix.

Finally, we can see that a multivariate version of the local level model is discussed in the appendix of this chapter.

(5.4) A Fixed Unknown Starting Value.

Using assumption A3 we can back substitute to obtain

$$y_t = \mu_0 + \eta_1 + \eta_2 + \ldots + \eta_t + \epsilon_t \tag{5.25}$$

Writing $v = y - \mu_0 . i$, where $y = (y_1, \ldots, y_T)'$ and $i = (1, 1, \ldots, 1)'$, the log–likelihood is given by

$$\ell(\sigma^2, q, \mu_0) = d - \tfrac{1}{2} \log |\sigma^2(I+qA)| - \tfrac{T}{2} v'(I+qA)^{-1} v/\sigma^2. \tag{5.26}$$

Concentrating σ^2 and μ from ℓ yields the following profile log–likelihood

$$M(q) = d - \tfrac{T}{2} \log \hat{v}'(I+qA)^{-1}\hat{v} - \tfrac{1}{2}\log|I+qA| \tag{5.27}$$

where $\hat{v} = y - \hat{\mu}(q).i$. We will use $P(q)$ to denote $I - i(i'(I+qA)^{-1}i)^{-1}i'(I+qA)^{-1}$ thus allowing the profile log–likelihood to be written as

$$M(q) = d - \tfrac{T}{2}\log y'P(q)'(I+qA)^{-1}P(q)y - \tfrac{1}{2}\log|I+qA| \tag{5.28}$$

Now as $\frac{dP(q)}{dq} = (I-P(q))A(I+qA)^{-1}(I-P(q))$, and $(I+qA)^{-1}P(q) = P(q)'(I+qA)^{-1}$, we can see that the score is

$$s(q) = \frac{T}{2}\frac{y'P(q)'(I+qA)^{-1}A(I+qA)^{-1}P(q)y}{y'P(q)'(I+qA)^{-1}P(q)y} - \frac{1}{2}\,\mathrm{tr}\left[(I+qA)^{-1}A\right] \tag{5.29}$$

by observing the equality $P(q)(I-P(q))=0$.

However, $P(q)(y-\mu_0.i) = P(q)y \sim N(0,\sigma^2P(q)(I+q^*A)P(q)')$, so using the symmetry and idempotency of $P(0)$ we can see that

$$p(q^*) = \Pr\left[\frac{T}{2}\frac{u'P(0)AP(0)[P(0)(I+q^*A)P(0)]u}{u'[P(0)(1+q^*A)P(0)]u} - \frac{1}{2}\frac{(T+1)T}{2} < 0\right] \tag{5.30}$$

since the trace of A is $\frac{T(T+1)}{2}$ and where u is a $T\times 1$ vector of independent standard normals. So if we write the non–zero eigenvalues of $P(0)AP(0)$ as (ω_t), and note that the eigenvectors of this matrix also diagonalise $P(0)$, then we can see that this probability can be written as

$$p(q^*) = \Pr\left[\sum_{t=1}^{T-1} u_t^2\left(\omega_t - \frac{(T+1)}{2}\right)(1+q^*\omega_t) < 0\right] \tag{5.31}$$

Theorem 5.4 of the appendix states that the non–zero eigenvalues of the matrix $P(0)AP(0)$ are $\omega_t = \dfrac{1}{4\sin^2\left[\frac{t\pi}{2T}\right]}$, $t=1,2,\ldots,T-1$. Hence the Imhof procedure can be employed to calculate this probability. The results of these calculations are given below in table 5.3.

Table 5.3

$p(q^*)$ for the Local Level Model using a Fixed, But Unknown, Starting Value.

True Value of q

T–1	0	0.01	0.1	1	10
10	.964	.955	.878	.597	.445
20	.961	.921	.665	.333	.237
30	.961	.866	.494	.199	.129
40	.961	.798	.373	.124	.077
50	.960	.725	.285	.081	.049

When $q^*=0$ and $T\to\infty$ we can enforce some simplifications on the above expression to give

$$\lim_{T\to\infty} p(0) = \Pr\left[\sum_{t=1}^{\infty} \frac{u_t^2}{t^2\pi^2} < \frac{1}{2}\right]. \tag{5.32}$$

Again using the Anderson and Darling(1952) tables it was found that this probability was 0.9602. Note that this value is dramatically bigger than for the first two cases.

Finally we prove the following theorem for the mle of q.

Theorem 5.5 The ML estimator of the signal–noise ratio q is $O_p(T^{-2})$ when $q^*=0$, ie. for any $\epsilon>0$, $\exists\ T_0$ and $c>0$ such that

$$\Pr(s(c/T^2;0) \geq 0\ ) < \epsilon,\ \forall\ T > T_0 \tag{5.33}$$

Proof. Remember at $q^*=0$ we are concerned with the probability law of the score, which we write as

$$\ell\left[\frac{1}{T^2}\, s(\frac{1}{T^2}\,;0) > 0\right] =$$

$$\ell\left[\frac{T}{2}\,\frac{u'P(q)'(I+qA)^{-1}A(I+qA)^{-1}P(q)u}{u'P(q)'(I+qA)^{-1}P(q)u} - \tfrac{1}{2}\mathrm{tr}[(I+qA)^{-1}A]\right]. \tag{5.34}$$

Then if we write $Q(q)=(I+qA)^{-0.5}P(q)(I+qA)^{0.5}$, and note $Q(q)$ is symmetric and idempotent for all q, and write the eigenvalues of $Q(q)AQ(q)$ as (ζ_t) we have that

$$\lim_{T\to\infty} \Pr\left[\frac{1}{T^2}\, s(\frac{c}{T^2}\,;0) > 0\right]$$

$$= \lim_{T\to\infty} \Pr\left[\frac{\dfrac{1}{2}\displaystyle\sum_{t=1}^{T-1} \frac{u_t^2\ T^2/\zeta_t}{(c+T^2/\zeta_t)^2}}{\dfrac{1}{T}\displaystyle\sum_{t=1}^{T-1} \frac{u_t^2\ T^2/\ \zeta_t}{(c+T^2/\zeta\ )}} - \frac{1}{2}\sum_{t=1}^{T} \frac{1}{(c+T^2/\delta_t)} > 0\right]. \tag{5.35}$$

The denominator of the first term can be bounded from above for every c by using Poincare's theorem (see Magnus and Neudecker(1988), pg 209–210)) and by ignoring the largest eigenvalue of the matrix A. This sum then converges to unity as $T\to\infty$. The numerator can be bounded from below by dropping the smallest eigenvalue of A. Thus, as $T\to\infty$, we have the result that the required probability is less than

$$\Pr\left[\frac{1}{2}\sum_{t=1}^{\infty}\frac{u_t^2\ [t-0.5]^2\pi^2}{(c+[t-0.5]^2\pi^2)^2}-\frac{1}{2}\sum_{t=1}^{\infty}\frac{1}{(c+[t-0.5]^2\pi^2)}>0\right].$$

(5.36)

Hence, theorem 5.5 follows by the method given in the proof of theorem 5.1.□

(5.5) The (Global) Maximum of the Profile Likelihood.

In section two we saw that a local maximum of the profile likelihood occurs at zero iff $s(0)<0$, but that this event did not imply the (global) ML estimate is zero. Instead we can see that the probability of a (global) ML estimate being zero has an upper bound $p(q^*)$. In this section we use a simulation experiment to show that this upper bound is a good approximation to the desired probability. Our study is based on 10,000 replications for the cases where T=10,50(10) and q=0,0.01,0.1,1. The gaussian white noise variables were generated using the NAG routine G05DDF. The results using a known, but fixed, startup are almost identical to those using a diffuse prior and so are not reported here. Table IV shows the proportion of (global) ML estimates which are exactly at zero. Bracketed beneath these results are the corresponding proportion of local maximums at zero. The results in the table indicate that the occurrences of boundary cases for the (global) maximum likelihood estimator is almost as common as for the local case discussed at length above. This is consistent with the observation made by Cryer and Ledolter(1981) in connection with the estimation of non–invertible moving average processes. Only when the true value of q becomes quite large (in the region of one) does this result break down. Nevertheless, even for these values, the fixed, but unknown, startup procedure gives many more boundary cases than does the diffuse prior.

Table 5.4
Simulation Results. Percentage of Global and Local Zeros

	Diffuse Prior				Fixed, Unknown Startup			
T–1 q^*	0.0	0.01	0.1	1.0	0.0	0.01	0.1	1.0
10	60 (63)	59 (62)	44 (47)	18 (21)	94 (97)	92 (97)	83 (91)	48 (64)
20	62 (64)	55 (57)	27 (29)	6 (8)	93 (97)	88 (93)	58 (68)	16 (35)
30	64 (65)	48 (49)	16 (18)	1 (3)	93 (96)	81 (88)	37 (50)	5 (21)
40	63 (64)	40 (41)	9 (11)	1 (1)	93 (96)	74 (80)	24 (38)	2 (12)
50	64 (65)	33 (34)	6 (7)	0 (0)	93 (96)	65 (73)	15 (29)	1 (8)

(5.6) Conclusion.

ML estimation of the signal–noise ratio, q, generally requires that a diffuse prior be assumed for the initial level, or that it be treated as a fixed but unknown constant. This chapter has shown that the small sample properties of the two estimators may be quite different when the true value of q is close or equal to zero. This result holds for local as well as global maxima in the likelihood function.

Treating the initial level as an unknown constant to be estimated results in a much higher probability of estimating q to be zero. When the true q is zero, this probability is 0.96 against 0.66 for the diffuse prior. While this is clearly good if q really is zero, the fact that the fixed initial level estimator has, for example, a 0.24 probability of being zero when q=0.1 and T=40, while the corresponding probability for the diffuse prior estimator is only 0.09, is much less attractive. Overall, the assumption of a fixed initial level will lead to a deterministic level being found far too often. This is undesirable from the point of view of forecasting since then there is no discounting of past observations.

Mathematical Appendix.

Theorem 5.2 The eigenvalues of the $(T-1)\times(T-1)$ matrix $C'(I+ii')^{-1}C$ are given by

$$\lambda_t = \frac{1}{4\sin^2\left[\frac{t\pi}{2T}\right]}, \quad t=1,2,\ldots,T-1.$$

Proof. Recall

$$C'(I+ii')^{-1}C = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} I - \frac{ii'}{T} \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$= \frac{1}{T} \begin{bmatrix} (T-1) & (T-2) & \ldots 2 & 1 \\ (T-2) & 2(T-2) & \ldots 3 & 2 \\ \vdots & \vdots & \ddots & \\ 1 & 2 & \ldots & (T-1) \end{bmatrix} = F \tag{5.A1}$$

say. It is well known that

$$F^{-1} = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix}. \tag{5.A2}$$

A simple application of Anderson's(1971) theorem 6.5.5 proves the theorem by the properties of the eigenvalue solution.□

<u>**Proof of Theorem 5.3**</u> Recall the expression for $p(q^*)$ and note that (λ_t) is monotonically decreasing in t. Then we can write

$$\Pr\left[\sum_{t=1}^{T-1} u_t^2\{\lambda_t - (T+1)/6\}(q^*\lambda_t + 1) < 0\right]$$

$$=\Pr\left[\sum_{\lambda_t > \frac{(T+1)}{6}} u_t^{\ 2}\{\lambda_t-(T+1)/6\}(q^*\lambda_t+1) < \sum_{\lambda_t \le \frac{(T+1)}{6}} u_t^{\ 2}\{(T+1)/6 - \lambda_t\}(q^*\lambda_t+1)\right] \tag{5.A3}$$

$$\le\Pr\left[\sum_{\lambda_t > \frac{(T+1}{6}} u_t^{\ 2}\{\lambda_t-(T+1)/6\}(q^*\lambda_t+1) < \frac{(T+1)}{6}\{1+q^*\frac{(T+1)}{6}\} \sum_{\lambda_t \le \frac{(T+1)}{6}} u_t^{\ 2}\right]. \tag{5.A4}$$

To improve tractability, it is useful to discover how many $\lambda_t > (T+1)/6$, for large T. But the event $1/4\sin^2 t\pi/2T > (T+1)/6 \Leftrightarrow$ the event $t < (2T/\pi)\arcsin\{\surd(1.5/[T+1])\}$. The RHS of this inequality is approximately $(2T/\pi)\surd(1.5/[T+1])$, ie $O(T^{1/2})$. If we define m as any fixed positive integer, then $\exists\, T_0$ such that $\forall\, T > T_0$

$$\lambda_t > (T+1)/6 \quad , t=1,2,\ldots,m.$$

Thus $\forall\, T > T_0$

$$p(q^*) \le \Pr\left[\sum_{t=1}^{m} u_t^{\ 2}\{\lambda_t-(T+1)/6\}(q^*\lambda_t+1) < ((T+1)/6)\{1+q^*(T+1)/6\} \sum_{t=m+1}^{T-1} u_t^{\ 2}\right] \tag{5.A5}$$

$$\leq \Pr\left[\{\lambda_m-(T+1)/6\}(q^*\lambda_m+1)\sum_{t=1}^{m} u_t^2 < ((T+1)/6)\ \{1+q^*(T+1)/6\}\sum_{t=m+1}^{T-1} u_t^2\right] \tag{5.A6}$$

Writing the first sum as χ_r^2 we note the independence of the two chi–squared variables and that $r = O(T)$, as m is fixed. Hence, using an expectations operator defined on the measures of the χ_r^2 variable, and employing the density function of a χ_m^2 variable yields

$$= a\ E\int_0^{b\chi_r^2} x^{(m/2-1)}e^{-x/2}dx, \tag{5.A6}$$

where $a = 1/(2^{m/2}.\Gamma(1/2))$ and $b = \dfrac{[(T+1)/6]\ \{1+q^*[(T+1)/6]\}}{\{\lambda_m-[(T+1)/6]\}\ \{q^*\lambda_m+1\}}$.

(5.A7)

But it is easy to see that $b=O(T^{-2})$ if $q^*>0$ (and $O(T^{-1})$ if $q^*=0$), which means that

$$\forall T>T_0,\ p(q^*) < a\ E\int_0^{b\chi_r^2} x^{(m/2-1)}\ dx \tag{5.A8}$$

$$= (2a/m)E(b\chi_r^2)^{m/2} \tag{5.A9}$$

so for even m

$$= (2a/m)\ (b)^{m.2}\ .\ r(r+2)(r+4)\ldots(r+m-2) \tag{5.A10}$$

$$= O(T^{-m})O(T^{m/2})$$

if $q^* > 0$ (and $O(1)$ of $q^* = 0$)

$$= O(T^{-m/2}). \tag{5.A11}$$

Hence, it is not possible to scale $p(q^*)$ by a factor T^n, where n is any fixed constant, to stop $p(q^*)$ from going to zero as $T \to \infty$.□

<u>Theorem 5.4</u> The (T−1) non–zero eigenvalues of the T×T matrix $C'(I-\frac{ii'}{i'i})C$ are given by

$$\omega_t = \frac{1}{4\sin^2\left[\frac{t\pi}{2T}\right]},\ t=1,2,\dots,T-1.$$

<u>Proof.</u> $C'(I-\frac{ii'}{i'i})C = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}\begin{bmatrix} I-\frac{ii'}{i'i} \end{bmatrix}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$

$$= \frac{1}{T}\begin{bmatrix} 0 & & 0^{*\prime} & & \\ & (T-1) & (T-2) & \dots 2 & 1 \\ 0^{*} & (T-2) & 2(T-2) & \dots 3 & 2 \\ & \vdots & \vdots & \ddots & \\ & 1 & 2 & \dots & (T-1) \end{bmatrix} \qquad \textbf{(5.A12)}$$

where 0^* is a (T−1)×1 vector of zeros. Hence using the properties of the eigenvalue solution and theorem 5.2, the theorem has been proved.

A Multivariate Appendix

A multivariate local level model has recently received some attention in the literature; see Enns et al(1982), and Harvey(1986). The general form of this model for a P dimensional series is

$$y_t = \mu_t + \epsilon_t \,,\ \epsilon_t \sim \mathrm{NID}(0,\Sigma_\epsilon) \tag{5.B1a}$$

$$\mu_t = \mu_{t-1} + \eta_t \,,\ \eta_t \sim \mathrm{NID}(0,\Sigma_\eta) \tag{5.B1b}$$

$$\mu_0 \sim \mathrm{N}(0,\kappa I), \text{ and } \kappa\to\infty \tag{5.B1c}$$

where (ϵ_t), (η_t), μ_0 are totally independent and the "true value" of Σ_ϵ is positive definite, while Σ_η is positive semi–definite.

Such a group of equations are called "seemingly unrelated time series equations", or SUTSE, after the so called seemingly unrelated equation model in econometrics.

A special case of this model has been studied by Fernandez–Macho and Harvey(1989). This model, which is called a "homogeneous" system, is given below.

$$y_t = \mu_t + \epsilon_t \,,\ \epsilon_t \sim \mathrm{NID}(0,\Sigma) \tag{5.B2a}$$

$$\mu_t = \mu_{t-1} + \eta_t \,,\ \eta_t \sim \mathrm{NID}(0,q\Sigma) \tag{5.B2b}$$

$$\mu_0 \sim \mathrm{N}(0,\kappa I), \text{ and } \kappa\to\infty \tag{5.B2c}$$

where q is a non–negative scalar.

In this section we will study the probability that the maximum likelihood estimator of q will be exactly zero.

Using precisely the projection technique employed earlier in this chapter we have the result that the likelihood for $y=(y_1{}',\dots,y_T{})'$ is the same, apart from a log κ term, as that given for $v = y - i \otimes y_1$, where i is a (T–1)x1 vector of ones. Now, $v \sim N_{P(T-1)}(0,(I+ii'+qA)\otimes\Sigma)$ so the desired log–likelihood is

$$\ell(q,\Sigma;y) = -\frac{1}{2}\log|(I+ii'+qA)\otimes\Sigma| - \frac{v'((I+ii'+qA)\otimes\Sigma)^{-1}v}{2}. \tag{5.B3}$$

Writing vec(V′)=v, where V′ is Px(T–1), allows us to see that (5.B3) is equal to

$$= -\frac{P}{2}\log|I+ii'+qA| - \frac{T-1}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\{V\Sigma^{-1}V'(I+ii'+qA)^{-1}\}. \tag{5.B4}$$

So using Magnus and Neudecker(1988, pg 314–317) we can concentrate at

$$\hat{\Sigma}(q) = \frac{V'(I+ii'+qA)^{-1}V}{T-1}. \tag{5.B5}$$

Noting that $\hat{\Sigma}(q)$ is positive definite wp1, the concentrated support is

$$M(q) = \ell(q,\hat{\Sigma}(q);y) = -\frac{P}{2}.\log|I+ii'+qA| - \frac{T-1}{2}\log|V'(I+ii'+qA)V|, \tag{5.B6}$$

which leads to

$$s(q) = \frac{dM(q)}{dq} = -\frac{P}{2}\mathrm{tr}((I+ii'+qA)^{-1}A)$$

$$+ \frac{T-1}{2}\mathrm{tr}\left[(V'(I+ii'+qA)^{-1}V)^{-1}V'(I+ii'+qA)^{-1}A(I+ii'+qA)^{-1}V\right]. \tag{5.B7}$$

Evaluating the score at $q = 0$ and its resulting probability law under $q^* = 0$ we have

$$p(0) = \Pr(s(0) < 0)$$

$$= \Pr\left[\mathrm{tr}\left[(U'U)^{-1}U'(I+ii')^{-1/2}A(I+ii')^{-1/2}U\right] \leq P.\frac{T+1}{6}\right]. \tag{5.B8}$$

Using a weak law of large numbers, $(U'U)/T \xrightarrow{P} I_P$, so

$$\lim_{T\to\infty} p(0) = \Pr\left[\sum_{t=1}^{\infty}\sum_{i=1}^{P}\frac{u_{it}^2}{t^2\pi^2} < \frac{P}{6}\right] \tag{5.B9}$$

where $u_{it} \sim NID(0,1)\ \forall\ i = 1,\cdots,P$ and $t=1,2,\ldots$. We become interested in the characteristic function of the random variable on the left hand side of this inequality. It is easy to show that this is given by

$$\varphi(s) = \prod_{t=1}^{\infty}\{1-(2is/t^2\pi^2)\}^{-P/2} = \{\sqrt{(2is)}/\sin\sqrt{(2is)}\}^{-P/2} \tag{5.B10}$$

which allows the limiting probability to be evaluated.

Using the results of Tanaka and Satchell(1987) we can tabulate the limiting probability, as $T\to\infty$, of the estimated parameter being on the boundary for various values of P. The results are given in table 5.5.

Table 5.5

P	$\lim_{T \to \infty} p(0;P)$
1	0.65744
2	0.61673
3	0.59659
4	0.58411
5	0.57545
6	0.56899
7	0.56894
8	0.55986
9	0.55646
10	0.55359
11	0.55111
12	0.54895
∞	0.50000

This table indicates that the high probability of observing an estimate of q which is exactly zero diminishes as P, the dimension of the system, increases.

Chapter Six

On the Probability of Estimating a Deterministic Component in the Dynamic Regression Model

Summary A dynamic regression model has a fixed intercept when the signal noise ratio, written q, is zero. This chapter investigates the properties of the maximum likelihood estimator of q, paying particular attention to the case where the true value of q is zero. These properties are shown to be crucially dependent on whether the regressors' coefficients are viewed as fixed or deterministic. If they are treated as fixed, then their presence will tend to have a dramatically distorting effect on the mle of q. If, on the other hand, they are initialized using a diffuse prior, then their effect will be mild.

Key Words DYNAMIC REGRESSION, STARTUP PROCEDURES, FREDHOLM DETERMINANTS, TIME TRENDS.

(6.1) Introduction

In chapter five we studied some aspects of the local level model, which is given by

$$y_t = \mu_t \quad + \epsilon_t, \qquad \epsilon_t \sim NID(0,\sigma^2) \tag{6.1a}$$
$$\mu_t = \mu_{t-1} \quad + \eta_t, \qquad \eta_t \sim NID(0,q\sigma^2). \tag{6.1b}$$

When the signal noise ratio q is zero, the local level model has a fixed level. In the last chapter we looked at the probability that the maximum likelihood (ML) estimator of q is zero. In practice, we are likely to initialize μ_0 using a diffuse prior, or by regarding it as a fixed, but unknown, constant which has to be estimated. In that work we demonstrated that this probability is sensitive to the assumption made about the initial value of the level component, μ_0. We showed that when the true value of q is zero, the respective probabilities are 0.66 and 0.96 for these two start–up procedures. While this is advantageous for the fixed starting value method, if q is really zero, the fact that it has, for example, a 0.24 probability of being zero when q=0.1 and T=40, is much less attractive compared to the corresponding probability of 0.09 for the diffuse prior estimator. Overall the assumption of a fixed initial level will lead to a deterministic level being found too often. This is undesirable from the point of view of forecasting since a deterministic level implies there is no discounting of past observations.

In this chapter we show that the same issue arises in the dynamic regression model (also called the intercept variation model)

$$y_t = \mu_t + x_t'\beta + \epsilon_t \quad , \epsilon_t \sim \text{NID}(0,\sigma^2) \tag{6.2a}$$

$$\mu_t = \mu_{t-1} + \eta_t \quad , \eta_t \sim \text{NID}(0,q\sigma^2) \tag{6.2b}$$

$$\mu_0 \sim \text{NID}(0,\kappa),\ \kappa \to \infty \tag{6.2c}$$

where (x_t) is a $(N \times 1)$ vector of fixed regressors, with μ_0 assumed to follow a diffuse prior, independent of (ϵ_t) and (η_t). This model has received considerable attention in econometrics; c.f. Cooley and Prescott(1973), Cooley(1975), Brown, Kleidon and Marsh(1983), Fama and Gibbons(1982), Trzcinka(1982), Hendrick(1973), Raj and Ullah(1981), Nabeya and Tanaka(1988), Leybourne and McCabe(1989) and Rosenberg(1973).

In this chapter we focus our attention on the assumption we make about β, and the effect this has on the probability of estimating a fixed intercept. We study two possible start–up regimes for β. The first is that β is a fixed, unknown constant, which we will have to estimate. This corresponds to the "classical" regression model and has been briefly studied by Sargan and Bhargava(1983b) in the context of MA(1) errors; that is a differenced version of the above model. The alternative is to regard β as deterministic in the sense of Wold (c.f. Chatfield(1984, pg 54–55)), by specifying the model

$$y_t = \mu_t + x_t'\beta + \epsilon_t \quad , \epsilon_t \sim \text{NID}(0,\sigma^2) \tag{6.3a}$$

$$\mu_t = \mu_{t-1} + \eta_t \quad , \eta_t \sim \text{NID}(0,q\sigma^2) \tag{6.3b}$$

$$\beta_t = \beta_{t-1} \tag{6.3c}$$

$$\begin{bmatrix}\mu_0 \\ \beta_0\end{bmatrix} \sim N_{N+1}(0,\kappa I),\ \kappa \to \infty. \tag{6.3d}$$

Thus β here has been given a diffuse prior. We will see that the assumption we place on β will be extremely influential in determining the probability of estimating a fixed intercept.

In the rest of this paper we study the effect these two assumptions have on the probability that a (local) ML estimator of q is exactly zero for various true values of q for the dynamic regression model. In all cases, however, we will assume that μ_0 has a diffuse prior.

Section two looks at the case of the fixed, but unknown constant, which has to be estimated. Section three investigates the diffuse case. Finally, in section four we use a simulation study to show that these results carry over to the case of the (global) ML estimator.

(6.2) The Fixed, Dynamic Regression Model

(6.2.1) The General Fixed, Dynamic Regression Model

Following the development given in chapter five, we can show that if we ignore a log κ term in the log likelihood, the likelihood for $(y_1,\ldots,y_T)'$ is the same as that for the vector random variable v, defined by $v = y - (y_1 - x_1'\beta)i$, where i and y are $(T-1)\times 1$ vectors such that $i=(1,\ldots,1)'$ and $y=(y_2,\ldots,y_T)'$. Back substitution gives us

$$y_1 - x_1'\beta = \mu_1 + \epsilon_1 \qquad \textbf{(6.4a)}$$

$$y_t - x_t'\beta = \mu_1 + \epsilon_t + \eta_2 + \ldots + \eta_t. \qquad \textbf{(6.4b)}$$

Thus $v \sim N(X^*\beta, \sigma^2(I+ii'+qA))$, where $A = CC'$, with $C=((I(i\geq j)))$ and $I(.)$ being the usual indicator function. Writing $X=(x_2,\ldots,x_T)' - i\otimes x_1'$ and $z = y - y_1 i$, then the log–likelihood is

$$l(q,\sigma^2,\beta) = c - \frac{(T-1)}{2}\log\sigma^2 - \tfrac{1}{2}\log|I+ii'+qA| - \tfrac{1}{2}\frac{(z-X\beta)'(I+ii'+qA)^{-1}(z-X\beta)}{\sigma^2}. \tag{6.5}$$

Assuming X is of full rank, we can concentrate the nuisance parameters σ^2 and β out of the likelihood function at

$$\hat{\sigma}^2(q,\beta) = \frac{1}{(T-1)}(z-X\beta)'(I+ii'+qA)^{-1}(z-X\beta) \tag{6.6}$$

and $\hat{\beta}(q) = (X'(I+ii'+qA)^{-1}X)^{-1}X'(I+ii'+qA)^{-1}z = [I - Q(q)]\, z$, say,

allowing us to write the profile log–likelihood function compactly as

$$M(q) = C - \frac{(T-1)}{2}\log z'Q(q)'(I+ii'+qA)^{-1}Q(q)z - \tfrac{1}{2}\log|I+ii'+qA|. \tag{6.7}$$

Then, writing the probability of observing a (local) maximum at zero, for a given true value of q, written q^*, as Pr (q^*) we have (see chapter five, figure 5.1) that

$$\Pr(q^*) = \Pr\left[\left.\frac{dM(q)}{dq}\right|_{q=0} < 0\right]. \tag{6.8}$$

This probability is given in the following theorem.

<u>Theorem 6.1</u> Writing $P(0) = (I+ii')^{-1/2}Q(0)(I+ii')^{1/2}$ and the (T−1−N) non–zero eigenvalues of $P(0)(I+ii')^{-0.5}A(I+ii')^{-0.5}P(0)$ as (ψ_t) we have that

$$\Pr(q^*) = \Pr\left[\sum_{t=1}^{T-1-N} u_t^2 \left(\psi_t - \frac{(T+1)}{6}\right)(1+q^*\psi_t) < 0\right] \tag{6.9}$$

where $u_t \sim NID(0,\sigma^2)$.

<u>Proof.</u> Given in the appendix.

Of course, (ψ_t) depends on the particular design in use and so this probability cannot be evaluated explicitly once and for all. However, we can find bounds on it using the idempotency of P(0) and Poincare's theory (c.f. Magnus and Neudecker(1988, pg 209–211)). Notice that these bounds are not necessarily achievable.

The eigenvalues of $(I+ii')^{-1/2}A(I+ii')^{-1/2}$ are the same as those of the matrix $C'(I+ii')^{-1}C$, and are given by $\lambda_t = \dfrac{1}{4\sin^2(t\pi/2T)}$ (see chapter five, theorem 5.2)). Writing $\delta_t(q^*,T) = (\lambda_t - \frac{(T+1)}{6})(1+q^*\lambda_t)$ we have an upper bound on this probability of

$$\Pr(q^*) \le \Pr\left[\sum_{t=1}^{T-1-N} u_t^2\ \delta_{(t)}(q^*,T) < 0\right] = \Pr_u(q^*), \tag{6.10}$$

whilst the lower bound is given by

$$\Pr(q^*) \geq \Pr\left[\sum_{t=1+N}^{T-1} u_t^2 \; \delta_{(t)}(q^*,T) < 0\right] = \Pr_l(q^*) \tag{6.11}$$

where $\delta_{(1)}(q^*,T) \leq \delta_{(2)}(q^*,T) \leq \ldots \leq \delta_{(T-1)}(q^*,T)$ are the ordered $\delta_t(q^*,T)$. These bounds are tabulated below in table 6.1 for $q^* = 0$ using Davies'(1973) procedure.

Table 6.1
Upper and Lower Bounds on Pr(0) for the Fixed, Dynamic Regression Model

N	1		2		3		4	
T	$Pr_l(0)$	$Pr_u(0)$	$Pr_l(0)$	$Pr_u(0)$	$Pr_l(0)$	$Pr_u(0)$	$Pr_l(0)$	$Pr_u(0)$
10	.580	.988	.513	1.00	.433	1.00	.335	1.00
20	.622	.977	.596	1.00	.567	1.00	.536	1.00
30	.635	.975	.618	1.00	.601	1.00	.583	1.00
40	.641	.973	.629	1.00	.616	1.00	.603	1.00
50	.644	.972	.635	1.00	.625	1.00	.615	1.00

These are precisely the bounds found by Sargan and Bhargava(1983b) in their study of regression models with non–invertible MA(1) disturbances.

We must also be interested in $Pr_l(q^*)$ and $Pr_u(q^*)$ for $q^* > 0$. To illustrate the behaviour of these probabilities we take N=1 and calculate the following table for the upper bound.

Table 6.2

Upper Bound on $\Pr(q^*)$ for the Fixed, Dynamic Regression Model with N=1

	True Value of q				
T	0	0.01	0.1	1	10
10	.988	.988	.982	.932	.830
20	.977	.971	.912	.621	.423
30	.975	.958	.800	.379	.212
40	.973	.940	.671	.226	.110
50	.973	.917	.547	.137	.057

The lower bound behaves so like that for the local level that it is not reproduced here (see table 5.2).

Finally, in this section we can derive the consistency rate of the mle of q, when the true value of the signal noise ratio is zero. This is done in the proof of the following theorem.

Theorem 6.2 The ML estimator of q is $O_p(T^{-2})$ when $q^*=0$, ie for any $\epsilon>0$, $\exists$ $T_0, c>0$ such that

$$\Pr\left[\left. \frac{dM(q)}{dq} \right|_{q=c/T^2} > 0 \right] < \epsilon \,, \forall\, T>T_0. \qquad (6.13).$$

Proof. Given in the appendix.

(6.2.2) A Univariate Model with A Partially Deterministic Linear Trend

An important special case of the above dynamic regression model is the partially deterministic linear time trend model. This type of model has frequently been used as part of some more complicated model; see Watson(1986), Nelson and Plosser(1982) and Clark(1987). We will write this model as

$$y_t = \mu_t + \beta.t + \epsilon_t \qquad (6.14a)$$

$$\mu_t = \mu_{t-1} + \eta_t \qquad (6.14b)$$

$$\mu_0 \sim N(0,\kappa) \ \kappa\to\infty \qquad (6.14c)$$

Writing $d=(1,2,...,T-1)'$, we have that $P(0)$ is

$$P(0) = I - \frac{(I+ii')^{-1/2}dd'(I+ii')^{-1/2}}{d'(I+ii')^{-1}d} \qquad (6.15)$$

so that we become interested in the eigenvalues of the matrix

$$C'(I+ii')^{-1}C - \frac{C'(I+ii')^{-1}dd'(I+ii)^{-1}C}{d'(I+ii')^{-1}d}. \qquad (6.16)$$

Writing the (T–2) non–zero eigenvalues of this matrix as (ξ_t), then the exact probability of observing the mle of being zero is given by

$$\Pr(q^*) = \Pr\left[\sum_{t=1}^{T-2} u_t^2(\xi_t - \frac{(T+1)}{6})(1+q^*\xi_t) < 0\right] \qquad (6.17)$$

where (ξ_t) are given in the following theorem.

Theorem 6.3 (ξ_t) are given by

$$\xi_{2t+1} = \frac{1}{4\sin^2(t\pi/T)}, \qquad \begin{array}{l} t=1,2,\ldots,\frac{1}{2}(T-1)\text{ , if T is odd} \\ t=1,2,\ldots,\frac{1}{2}(T-2)\text{ , if T is even} \end{array} \tag{6.18a}$$

and

$$\xi_{2t} = \frac{1}{4\sin^2(\theta_t/2)}, \qquad \begin{array}{l} t=1,2,\ldots,\frac{1}{2}(T-3)\text{ , if T is odd} \\ t=1,2,\ldots,\frac{1}{2}(T-2)\text{ , if T is even.} \end{array} \tag{6.18b}$$

where θ_t is the smallest solution to the equation $\frac{\tan(T\theta_t/2)}{\tan(\theta_t/2)} = T$, which is in the interval $T\theta_t \epsilon\ (2\pi t, 2\pi(t+0.5))$.

Proof. Given in the appendix.

It is unfortunate that this does not yield a closed form solution for (ξ_{2t}), but its values can be found comparatively cheaply. The resultant eigenvalues can be used in the Davies procedure to produce the following table.

Table 6.3

Tabulation of $p(q^*)$ for the Partially Deterministic Trend Model

	True Value of q				
T	0.0	0.01	0.1	1.0	10.0
10	.988	.987	.981	.930	.827
20	.976	.970	.908	.603	.394
30	.973	.956	.790	.354	.192
40	.971	.937	.655	.208	.099
50	.971	.912	.526	.124	.053

Note how extremely close this is to the upper bound for the dynamic regression model.

In the case where $q^* = 0$, and $T \to \infty$, the outcome is particularly interesting. As we do not have a close form solution for all the eigenvalues, we resort to using the Fredholm approach (see Kac, Keifer and Wolfowitz(1955) and Nabeya and Tanaka(1988)) to derive the characteristic function of interest. We first find its limiting kernel

$$\ell\left[u' \frac{C'(I+ii')^{-1}C}{T}u - u' \frac{C'(I+ii')^{-1}dd'(I+ii')^{-1}C}{Td'(I+ii')^{-1}d} u\right] \tag{6.19}$$

$$= \ell\left[\frac{1}{T}\sum_{j,k=1}^{T-2}\sum K^*\left(\frac{j}{T},\frac{k}{T}\right) u_j u_k\right] \tag{6.20}$$

$$\to \ell\left[\int_0^1\int_0^1 K(s,t)\, dw(s)\, dw(t)\right] \text{ as } T \to \infty, \tag{6.21}$$

where $K(s,t)=\min(s,t)-st-3st(1-s)(1-t)$ and $w(t)$ is Brownian motion with $Ew(t)=0$

and Ew(s)w(t)=min(s,t). This implies that the limiting random variable has a characteristic function given is $(D(2it))^{-1/2}$, where $D(\lambda)$ is the Fredholm determinant associated with the integral equation

$$f(t) = \int_0^1 K(s,t)\, f(s)\, ds. \tag{6.22}$$

This Fredholm determinant was derived in theorem 6 (the case of m=1) of Nabeya and Tanaka(1988) and is given below

$$D(\lambda) = \frac{12}{\lambda^2}\left[2 - \sqrt{\lambda}.\ \sin\sqrt{\lambda} - 2\cos\sqrt{\lambda}\right]. \tag{6.23}$$

This allows the limiting probability to be evaluated. Using Knott(1974)'s numerical inversion theorem we calculate that

$$\lim_{T\to\infty} \Pr(0) = 0.96787, \tag{6.24}$$

which is near the upper bound for the dynamic regression model when N=1.

A similar manipulation can be employed on the dynamic regression model when $x_t=(t,t^2)'$ using the second Fredholm determinant given in theorem 7 of Nabeya and Tanaka(1988). It is

$$D(\lambda) = \frac{8640}{\lambda^4}\left[2 + \frac{\lambda}{3} + \sqrt{\lambda}(-2+\frac{\lambda}{12})\sin\sqrt{\lambda} + (-2+\frac{2\lambda}{3})\cos\sqrt{\lambda}\right]. \tag{6.25}$$

This gives a limiting result of 0.99874, which is very close to the upper bound associated with the dynamic regression model when N=2.

(6.2.3) Regression Models Containing Partially Deterministic Trends

Suppose we now go back to the general fixed, dynamic regression model, but with $x_t=(t;x_t^{*\prime})'$ where x_t^* is a $(N^*\times 1)$ vector of regressors.

The bounds given in section one can be tightened using the explicit time trend regressor incorporated in the design. The effect of this is studied in King(1981) in connection with the Durbin–Watson statistic.

Supposing $X=[X_1,X_2]$ is of full rank. Then, if we write $P_Y = I - Y(Y'Y)^{-1}Y'$, it is well known that

$$P_X = P_{X_1}P_{P_{X_1}.X_2}. \tag{6.26}$$

We apply this to our model by writing $P_1(0) = I - \dfrac{(I+ii')^{-1/2}dd'(I+ii')^{-1/2}}{d'(I+ii')^{-1}d}$. Then there exists a symmetric matrix $P_2(0)$ such that $P(0)=P_2(0)P_1(0)$. Thus

$$P(0)(I+ii')^{-1/2}A(I+ii')^{-1/2}P(0)$$

$$= P_2(0)P_1(0)(I+ii')^{-1/2}A(I+ii')^{-1/2}P_1(0)P_2(0), \tag{6.27}$$

Poincare's theorem can be used again but this time on $P_1(0)(I+ii')^{-1/2}A(I+ii')^{-1/2}P_1(0)$. We have already denoted the non–zero eigenvalues of this matrix (ξ_t). So writing $\Delta_t(q^*,T) = (\xi_t - \frac{(T+1)}{6})(1+q^*\xi_t)$ we have

$$\Pr(q^*) \leq \Pr\left[\sum_{t=1}^{T-2-N^*} u_t^2 \, \Delta_{(t)}(q^*,T) < 0\right] = \Pr_u(q^*) \tag{6.28}$$

$$\Pr(q^*) \geq \Pr\left[\sum_{t=1+N^*}^{T-2} u_t^2 \, \Delta_{(t)}(q^*,T) < 0\right] = \Pr_l(q^*), \tag{6.29}$$

where $\Delta_{(1)}(q^*,T) \leq \Delta_{(2)}(q^*,T) \leq \ldots \leq \Delta_{(T-2)}(q^*,T)$ are the ordered $\Delta_t(q^*,T)$. These probabilities are of some importance. Table 6.4 gives them for $q^*=0$ for various N^* and T.

Table 6.4

Upper and Lower Bounds on Pr(0) for the Fixed, Dynamic Regression Model

N^*	1		2		3		4	
T	$Pr_l(0)$	$Pr_u(0)$	$Pr_l(0)$	$Pr_u(0)$	$Pr_l(0)$	$Pr_u(0)$	$Pr_l(0)$	$Pr_u(0)$
10	.977	1.00	.957	1.00	.919	1.00	.846	1.00
20	.970	1.00	.962	1.00	.951	1.00	.939	1.00
30	.969	1.00	.964	1.00	.958	1.00	.952	1.00
40	.968	1.00	.965	1.00	.961	1.00	.957	1.00
50	.968	1.00	.965	1.00	.962	1.00	.959	1.00

If we compare this table with table 6.1 we can see that the presence of the deterministic drift component dramatically changes $Pr_l(0)$ and $Pr_u(0)$. In effect, for $N^* = 1$, the $Pr_l(0)$ and $Pr_u(0)$ now correspond to the upper bounds for N=1 and N=2 respectively. A similar pattern holds for $N^* > 1$.

Likewise, when we consider $q^*>0$ and $N^*=1$, $Pr_l(q^*)$ behaves in a similar way to that already reported in table 6.2 for the upper bound in the dynamic regression model. $Pr_u(q^*)$ is reported in the table given below. The table is also instructive since it is a good approximation to $Pr_u(q^*)$ for N=2 in the intercept variation model.

Table 6.5

Tabulation of the Upper Bound for Pr(q*) for Various q^*

T	True Value of q 0.0	0.01	0.1	1.0	10.0
10	1.00	1.00	1.00	1.00	1.00
20	1.00	1.00	1.00	.986	.926
30	1.00	1.00	.994	.861	.693
40	1.00	.999	.973	.695	.489
50	.999	.998	.932	.535	.334

Table 6.5 implies that if $q^*=0$ the probability of observing a local maximum on the boundary is almost one. Equally, when $q^*>0$ $Pr(q^*)$ may be sizable for large T and moderate q^*. Thus, observing an estimated value of q to be exactly zero is little evidence for the hypothesis that the true value of q is in fact zero.

(6.3) The Diffuse, Dynamic Regression Model

(6.3.1) The General Diffuse, Dynamic Linear Regression Model

The ML estimator of q is sensitive to the design matrix because of the inability of the determinant term in the log–likelihood, equation (6.5), to reflect our uncertainty about β. This is a familiar problem which can be solved by using a conditioning argument. To illustrate this approach we will first think about a simple case. Suppose we have observed $(y_1,\ldots,y_T)'$ and that

$$y_t = \mu + \epsilon_t, \text{ where } \epsilon_t \sim \text{NID}(0, \sigma^2). \tag{6.30}$$

Then the ML estimator of μ, $\bar{Y}$ is unbiased, but the corresponding estimator of σ^2 is biased. To overcome the effect of the presence of the unknown μ on the estimator of σ^2 we could condition on the first observation. This "approximation" to the likelihood leads to the ML estimators of both μ and σ^2 being unbiased.

The above "approximation" is exact if we consider a slightly different model,

$$y_t = \mu_t + \epsilon_t, \text{ where } \epsilon_t \sim \text{NID}(0, \sigma^2) \tag{6.31a}$$

$$\mu_t = \mu_{t-1} \tag{6.31b}$$

$$\mu_0 \sim N(0, \kappa),\ \kappa \to \infty, \text{ where } \mu_0 \text{ is independent from the } (\epsilon_t). \tag{6.31c}$$

Here the diffuse prior annihilates the marginal distribution of y_1, making the approximation become exact.

A more complicated example is

$$y_t = x_t'\beta + \epsilon_t, \text{ where } \epsilon_t \sim \text{NID}(0, \sigma^2). \tag{6.32}$$

We can condition on the first N observations to enable us to construct an unbiased estimator of σ^2. Alternatively, we can set down the following model which performs this approximation directly

$$y_t = x_t'\beta_t + \epsilon_t, \text{ where } \epsilon_t \sim \text{NID}(0, \sigma^2) \tag{6.33a}$$
$$\beta_t = \beta_{t-1} \tag{6.33b}$$
$$\beta_0 \sim \text{N}(0, \kappa),\ \kappa \to \infty, \text{ where } \beta_0 \text{ is independent from the } (\epsilon_t). \tag{6.33c}$$

It is the dynamic generalization of this model which will form the basis of our discussion in this section. Remember, the diffuse dynamic regression model is given by

$$y_t = \mu_t + x_t'\beta_t + \epsilon_t \tag{6.34a}$$
$$\mu_t = \mu_{t-1} + \eta_t \tag{6.34b}$$
$$\beta_t = \beta_{t-1} \tag{6.34c}$$

where we initialize the state equation by employing a (N+1)×1 dimensional diffuse prior

$$\begin{bmatrix} \mu_0 \\ \beta_0 \end{bmatrix} \sim \text{N}(0, \kappa I),\ \kappa \to \infty. \tag{6.34d}$$

If we use the projection technique we employed in section two, then it is not difficult to show that

Theorem 6.4

$$\ell(\sigma^2,q) = -\tfrac{1}{2}\log|F+qG| - \tfrac{1}{2}\frac{v'(F+qG)^{-1}v}{\sigma^2} - \frac{(T-N-1)}{2}\log\sigma^2 \tag{6.35}$$

where $F = I + ii' + X^*(X^+)^{-1}D(X^+)^{-1\prime}X^{*\prime} + X^*(X^+)^{-1}E + E'(X^+)^{-1\prime}X^{*\prime}$,

$$G = A + X^*(X^+)^{-1}(X^+)^{-1\prime}X^{*\prime},\ \text{where } X^+ = \begin{pmatrix} \Delta x_2' \\ \vdots \\ \Delta x_{N+1}' \end{pmatrix},$$

$$X^* = \begin{pmatrix} x_{N+2} - x_{N+1} \\ \vdots \\ x_T - x_{N+1} \end{pmatrix}$$

and $E = \begin{bmatrix} 1\,1\,1 \ \ldots\, 1 \\ \\ \square \\ \\ \end{bmatrix}$ as a $N\times(N-n)$ matrix and $D = \begin{bmatrix} 2 & -1 & 0 & \cdots & & 0 \\ -1 & 2 & -1 & \cdots & & 0 \\ 0 & -1 & 2 & \cdots & & 0 \\ \vdots & \vdots & \vdots & & & \\ & & & & 2 & -1 \\ 0 & 0 & 0 & & -1 & 2 \end{bmatrix}$

Proof. Given in the appendix.

Although this is a reasonably convenient form for the numerical evaluation of $\Pr(q^*)$, it does not allow us to compare its analytic form with the expressions given in section two. However, we can use de Jong's(1988) result to rewrite the probability law of the log–likelihood. This new expression has the same probability law as (6.35) when it is evaluated under the assumption that the observations are generated under the fixed startup.

$$\ell(\sigma^2,q) = -\tfrac{1}{2}\log|F+qG| - \frac{1}{2}\frac{z'Q(q)'(I+i\,i'+qA)^{-1}Q(q)z}{\sigma^2} - \frac{(T-N-1)}{2}\log\sigma^2 \tag{6.36}$$

$$= +\frac{1}{2}\sum_{t=1}^{T-N-1} \log \lambda_t[Q(q)'(I+ii'+qA)^{-1}Q(q)] - \frac{1}{2}\frac{z'Q(q)'(I+i\,i'+qA)^{-1}Q(q)z}{\sigma^2}$$

$$-\frac{(T-N-1)}{2}\log\sigma^2 \tag{6.37}$$

where $\lambda_t[Q(q)'(I+ii'+qA)^{-1}Q(q)]$ denotes the non–zero eigenvalues of $Q(q)'(I+ii'+qA)^{-1}Q(q)$. If we compare this with 6.5 we see that we have merely altered a log–determinant term.

It is then not difficult to show that

$$\Pr(q^*) = \Pr\left[\sum_{t=1}^{T-N-1} u_t^2(\psi_t - \overline{\psi})(1+q^*\psi_t) < 0\right] \tag{6.38}$$

where $\overline{\psi} = \sum_{t=1}^{T-N-1} \psi_t/(T-N-1)$.

This probability can be evaluated routinely by using the Davies algorithm for any given set of regressors. However, it does not seem clear how one might bound this probability in the way we did for the fixed startup parameterization, since the regressors enter the likelihood function in quite a complicated way.

Before we look at some leading examples of this result, the following theorem is proved.

Theorem 6.5 The mle of q is $O_p(T^{-2})$ when $q^*=0$, ie for $\epsilon>0$, $\exists$ T_0,c>0 such that

$$\Pr\left(s(c/T^2;0) < 0\right) < \epsilon \ \forall \ T>T_0. \qquad (6.39).$$

Proof. Given in the appendix.

(6.3.2) A Univariate Model With A Partially Deterministic Trend

One of the models considered in section 6.2.2 was the partially deterministic trend model

$$y_t = \mu_t \quad +\beta.t \quad + \epsilon_t \qquad (6.40a)$$

$$\mu_t = \mu_{t-1} \quad + \eta_t. \qquad (6.40b)$$

The relevant eigenvalues are given by theorem 6.3 and are used in the Davies algorithm to produce the following table.

Table 6.6

Tabulation of $\Pr(q^*)$ for the Diffuse,
Partially Deterministic Linear Trend Model

	True Value of q				
T	0.0	0.01	0.1	1.0	10.0
10	.595	.590	.548	.367	.214
20	.610	.585	.433	.141	.049
30	.614	.558	.305	.052	.013
40	.617	.521	.205	.020	.004
50	.618	.477	.136	.008	.001

A number of features of this table are remarkable. Firstly, if we compare it to results from the fixed, dynamic regression model we can see that Pr(0) is considerably lower in this case; 0.62 compared to 0.96. While it is good to have such a high value if q really is zero, the fact that the fixed startup for the dynamic regression model results in $\Pr(q^*)$ being, for example, 0.655 when $q = 0.1$ and $T = 40$, compared to 0.205, is much less attractive. Secondly, the results are very similar to those found in chapter five for the local level model initialized by a diffuse prior. For example, for the local level model Pr(0) was 0.66 while Pr(0.1), when $T = 40$, was 0.110.

The limiting probability as $T \to \infty$ when $q^* = 0$ is of some interest. In this case, we can use the results which we derived in the second part of section 6.3.1. Remember the appropriate Fredholm determinant was

$$D(\lambda) = \frac{12}{\lambda^2}\left[2 - \sqrt{\lambda}.\ \sin\sqrt{\lambda} - 2\cos\sqrt{\lambda}\right] \tag{6.23}$$

and note that

$$\mathrm{tr}\Big[P(0)(I+ii')^{-0\cdot5}A(I+ii')^{-0\cdot5}P(0)\Big] = \mathrm{tr}\Big[A(I+ii')^{-1} - \frac{(I+ii')^{-1}dd'(I+ii')^{-1}A}{d'(I+ii')^{-1}d}\Big].$$

But $\mathrm{tr}\Big[(I+ii')^{-1}A\Big] = \frac{(T-1)(T+1)}{2}$, $d'(I+ii')^{-1}d = \frac{T(T-1)(T+1)}{12}$, $C'(I+ii')^{-1}d$ has the typical element $\frac{t(T-t)}{2}$, so $\mathrm{tr}\Big[C'(I+ii')^{-1}dd'(I+ii')^{-1}C\Big] = \frac{1}{4}\sum_{t=1}^{T-1} t^2(T-t)^2$ $= \frac{T(T-1)}{120}\Big[T^3+T^2+T+1\Big]$ using the equalities in Anderson(1971, pg 83).

So $\lim\limits_{T\to\infty} T^{-2}\,\mathrm{tr}\Big[P(0)(I+ii')^{-1/2}A(I+ii')^{-1/2}P(0)\Big] = \frac{1}{6} - \frac{1}{10} = \frac{1}{15}.$

We can use Knott(1974)'s inversion formulation to work out the required limiting probability, which turns out to be $\lim\limits_{T\to\infty} \Pr(0) = 0.62257$.

A similar extension can be employed on the dynamic regression model when $x_t' = (t,t^2)$. We give here only the limiting result for $q^*=0$, but the exact probability for $q^*>0$ can be evaluated. To find the $\lim\limits_{T\to\infty} \Pr(0)$ we can use the Fredholm approach. The required Fredholm determinant is

$$D(\lambda) = \frac{8640}{\lambda^4}\Big[2 + \frac{\lambda}{3} + \sqrt{\lambda}(-2+\frac{\lambda}{12})\sin\sqrt{\lambda} + (-2+\frac{2\lambda}{3})\cos\sqrt{\lambda}\Big] \qquad \textbf{(6.25)}$$

while we must also look at $\lim\limits_{T\to\infty} \frac{1}{T^2}\,\mathrm{tr}\Big[P(0)(I+ii')^{-0\cdot5}A(I+ii')^{-0\cdot5}P(0)\Big]$. This is a rather intractable expression; but by noting that it is the expectation of the kernel which generated the Fredholm determinant given above we can find the value of the limit of the trace by calculating the first cumulant associated with the corresponding characteristic function. It turns out, after a great deal of simple but tedious algebra, that

$$\left.\frac{d\log\ D(\lambda)^{-0.5}}{d\lambda}\right|_{\lambda=0} = \frac{36\times 8640\times 2}{10!\times 4} = \frac{3}{70}. \tag{6.41}$$

Using Knott's formulation we find that $\lim_{T\to\infty} \Pr(0) = 0.60699$. This value is slightly lower than in the case of a linear trend. Note also how dramatically lower this value is compared to that recorded for the fixed, dynamic regression model.

(6.4) The (Global) Maximum of the Profile Likelihood

In this section we use a simulation experiment to show that $\Pr(q^*)$ is a good approximation of the probability of a global maximum being observed at $q = 0$ for the case of a partially deterministic linear trend. The study is based on 10,000 replications for the cases where T=10,50(10) and q=0,0.01,0.1,1. The gaussian white noise variables were generated using the NAG routine G05DDF. Table 6.7 shows the proportion of (global) ML estimates which are exactly at zero. Bracketed beneath these results are the corresponding proportion of local maximums at zero. The results in the table indicate that the occurrences of boundary cases for the (global) maximum likelihood estimator is almost as common for this model as for the local maximum analyzed in the proceeding section. This is consistent with the observation made by Cryer and Ledolter(1981) in connection with the estimation of non–invertible moving average processes. Only when the true value of q becomes quite large, (in the region of one), does this result break down.

Table 6.7

Simulation Results. Percentage of Global and Local Zeros.

	Diffuse Prior				Fixed, Unknown Startup			
T–1 q^*	0.0	0.01	0.1	1.0	0.0	0.01	0.1	1.0
10	59 (61)	58 (60)	52 (54)	30 (31)	96 (98)	96 (98)	94 (97)	78 (87)
20	61 (62)	57 (58)	40 (41)	10 (11)	95 (97)	93 (96)	83 (88)	40 (55)
30	61 (62)	55 (55)	28 (29)	4 (4)	95 (97)	92 (95)	68 (77)	16 (32)
40	61 (61)	50 (51)	18 (18)	1 (2)	95 (97)	90 (93)	53 (63)	7 (18)
50	61 (62)	45 (46)	11 (12)	0 (1)	95 (97)	86 (91)	39 (49)	2 (10)

(6.5) Conclusion

We saw in chapter five that the probability that the signal noise ratio q is estimated to be exactly zero is very sensitive to the choice of the initial conditions we place on μ_t. In this chapter we have developed this idea, and have shown that the probability is also sensitive to the selection of regressing coefficients.

This finding is rather surprising. The sensitivity of the distributional behaviour of q to the regression design matrix is a very disturbing characteristic of the fixed, unknown startup procedure. It can be removed by the use of a diffuse prior on the coefficients. This acts like a conditioning argument, with the first N observations being used to provide initial values for the coefficients.

Mathematical Appendix

Proof of Theorem 6.1 As $\frac{dQ(q)}{dq} = (I-Q(q)')A(I+ii'+qA)^{-1}(I-Q(q))$,

$$ \tag{6.A1} $$

we can use the fact that $(I-Q(q))Q(q) = 0$ to show that

$$s(q) = \frac{dM(q)}{dq}$$

$$= \frac{(T-1)}{2}\frac{z'Q(q)'(I+ii'+qA)^{-1}A(I+ii'+qA)^{-1}Q(q)z}{z'Q(q)'(I+ii'+qA)^{-1}Q(q)z} - \frac{1}{2}\mathrm{tr}\left[(I+ii'+qA)^{-1}A\right] \tag{6.A2}$$

However, $Q(q)(z-X\beta) = Q(q)z \sim N(0,\sigma^2 Q(q)(I+ii'+q^*A)Q(q)')$. So writing $P(0)=(I+ii')^{-1/2}Q(0)(I+ii')^{1/2}$, which is an idempotent matrix, leads us to the observation that

$$\Pr(q^*) = \Pr\left[\frac{(T-1)}{2}\frac{u'P(0)(I+q^*D)P(0)\ \ P(0)DP(0)u}{u'P(0)\cancel{D}P(0)u} - \frac{1}{2}\frac{(T-1)(T+1)}{6} < 0\right]$$

(I+q*D)

$$ \tag{6.A3} $$

where $D = (I+ii')^{-1/2}A(I+ii')^{-1/2}$ and $u \sim N(0, I)$. If we write the eigenvalues of $P(0)DP(0)$ as (ψ_t) and note that the eigenvectors of this matrix also diagonalise the idempotent, symmetric matrix $P(0)$, then this probability can be written as

$$\Pr(q^*) = \Pr\left[\sum_{t=1}^{T-1-N} u_t^2\left(\psi_t - \frac{(T-1)}{6}\right)(1+q^*\psi_t) < 0\right]\square. \tag{6.A4}$$

<u>Proof of Theorem 6.2</u> On $q^* = 0$, we have the probability law of the score

$$\ell(s(q)) = \ell\left[\frac{(T-1)}{2}\frac{u'R(q)P(q)D(q)P(q)R(q)'u}{u'R(q)P(q)R(q)'u} - \frac{1}{2}\,\mathrm{tr}\left[(I+ii'+qA)^{-1}A\right]\right], \tag{6.A5}$$

where $P(q) = (I+ii'+qA)^{-1/2}Q(q)(I+ii'+qA)^{1/2}$,

$D(q) = (I+ii'+qA)^{-1/2}A(I+ii'+qA)^{-1/2}$, and

$R(q) = (I+ii')^{1/2}(I+ii'+qA)^{-1/2}$.

Proposition 6.1 states that $R(q)R(q)'$ and $D(q)$ have the same eigenvectors, which implies that we can use Poincare's theorem to prove that

$$\ell\left[u'R(q)P(q)D(q)P(q)R(q)'u\right] \leq \ell\left[\sum_{t=1}^{T-1-N} u_t^2\left[\frac{\lambda_t}{(1+q\lambda_t)^2}\right]_{(t)}\right], \tag{6.A6}$$

where the bracketed subscript denotes the ordered index.

It is easier to see that

$$\ell\left[\frac{u'R(q)P(q)R(q)'u}{T}\right] \geq \frac{1}{T}\,\ell\left[\sum_{t=N+1}^{T-1}\frac{u_t^2}{(1+q\lambda_t)}\right] \xrightarrow{p} 1, \tag{6.A7}$$

while $\mathrm{tr}\left[(I+ii'+qA)^{-1}A\right] = \displaystyle\sum_{t=1}^{T-1}\frac{\lambda_t}{(1+q\lambda_t)}$. $\qquad$ (6.A8)

Thus,

$$\lim_{T\to\infty}\frac{1}{T^2}\ell\left[\sum_{t=1}^{T-1-N} u_t^2\left[\frac{\lambda_t}{(1+q\lambda_t)^2}\right]_{(t)}\right]\to \qquad \ell\left[\sum_{t=1}^{\infty} u_t^2\left[\frac{t^2\pi^2}{(c+t^2\pi^2)^2}\right]_{(t)}\right] \tag{6.A9}$$

Clearly for all finite c the smallest values are occurring when t is infinite, so

$$=\ell\left[\sum_{t=1}^{\infty} u_t^2\left[\frac{t^2\pi^2}{(c+t^2\pi^2)^2}\right]\right]. \tag{6.A10}$$

Hence we have that $\lim_{T\to\infty}\ell(1/T^2\ s(c/T^2;0))$

$$\geq \ell\left[\frac{1}{2}\sum_{t=1}^{\infty}\frac{u_t^2}{(c+t^2\pi^2)^2}-\frac{1}{2}\sum_{t=1}^{\infty}\frac{1}{(c+t^2\pi^2)}\right]. \tag{6.A11}$$

Then, using the arguments deployed in the proof of theorem 5.1, we get the required consistency result immediately □.

Proof of Theorem 6.3 We need to solve the following equation

$$\left[C'(I+ii')^{-1}C - \frac{C'(I+ii')^{-1}dd'(I+ii')^{-1}C}{d'(I+ii')^{-1}d}\right] y = \lambda y. \tag{6.A12}$$

Writing $z=C'(I+ii')^{-1}Cy$, $F=C'(I+ii')^{-1}C$ and noting the equality

$$(CC')^{-1} = \begin{bmatrix} 2 & -1 & 0 & \cdots & & 0 \\ -1 & 2 & -1 & \cdots & & 0 \\ 0 & -1 & 2 & \cdots & & 0 \\ \vdots & \vdots & \vdots & & & \\ & & & & 2 & -1 \\ 0 & 0 & 0 & & -1 & 1 \end{bmatrix} \tag{6.A13}$$

so that we see that $d'(CC')^{-1}C = i'$, we look at the problem of solving

$$\left[I - \frac{C'(I+ii')^{-1}di'}{d'(I+ii')^{-1}d}\right] z = \lambda F^{-1} z. \tag{6.A14}$$

The left hand side of this equality is simply $z-e\nu$, where $\nu=i'z$ and

$$e=\frac{C'(I+ii')^{-1}d}{d'(I+ii')^{-1}d}, \tag{6.A15}$$

which gives the typical term $e_t=\frac{6t(T-t)}{T(T-1)(T+1)}$. This allows us to write out the above eigenvalue equation term by term, to give

$$\gamma z_t = \gamma\nu e_t + 2z_t - z_{t-1} - z_{t+1}, \text{ where } \gamma=\lambda^{-1}, \tag{6.A16}$$

with the three boundary conditions that

$$(1)\ z_0 = 0 \tag{6.A17a}$$

$$(2)\ z_T = 0 \tag{6.A17b}$$

$$(3)\ i'(I-ei')z = \lambda i'F^{-1}z = z_1 + z_{T-1} = 0. \tag{6.A17c}$$

Thus to find the required eigenvalues, we need to solve a second order difference equation subject to three constraints. Goldberg(1968, pg 184–8) has given the general solution to the above equation as

$$z_t = \sin(t\theta + B'), \tag{6.A18}$$

where B' is some constant and θ is such that $\gamma = 4\sin^2(\theta/2)$.

To achieve a particular solution to the equation we use the method of undermined coefficients. This yields $z_t^* = \frac{\nu}{c}\left[\frac{2}{\gamma} + Tt - t^2\right]$ so giving the solution $z_t = \sin(t\theta + B') + \frac{\nu}{c}\left[\frac{2}{\gamma} + Tt - t^2\right]$.

Imposing the constraint $z_0 = 0$ implies $0 = \sin B' + \frac{2\nu}{c\gamma}$. Hence we can write z_t as

$$z_t = 2\cos\left[\frac{t\theta}{2} + B'\right]\sin\frac{t\theta}{2} + (Tt - t^2)\sin B'(\cos\theta - 1). \tag{6.A19}$$

Imposing $z_T = 0$ implies $0 = \cos\left[\frac{t\theta}{2} + B'\right]\sin\frac{t\theta}{2}$ which implies that either

$$\theta = \frac{2k\pi}{T}\text{, k being some integer,} \tag{6.A20a}$$

or $$\theta = \frac{2r\pi + \pi - B'}{T}\text{, r being some integer.} \tag{6.A20b}$$

The final constraint implies that $\sin\frac{T\theta}{2}\cos\left[\frac{(T-2)\theta + B'}{2}\right] =$ $T(1 - \cos\theta)\sin B'$, so if we impose the first of the conditions implied by $z_T = 0$, we obtain sin B' =0, so giving the first part of the theorem using the definition of the eigenvalue solution. The imposition of the the second condition is more complicated. It implies, as

$$\cos(r\pi + 0.5\pi - \theta) = \sin(r\pi + 0.5\pi)\sin\,\theta \qquad (6.A21a)$$

$$\sin(r\pi + 0.5\pi - \frac{T\theta}{2}) = \sin(r\pi + 0.5\pi)\cos\frac{T\theta}{2}, \qquad (6.A21b)$$

that $\sin\frac{T\theta}{2}\sin\theta\sin(r\pi+\frac{\pi}{2}) = T\sin(r\pi+\frac{\pi}{2})2\sin^2\frac{\theta}{2}\cos\frac{T\theta}{2}$ **(6.A21c)**

we need to solve the equation $\frac{\tan(T\theta/w)}{\tan(\theta/2)} = T$, which, of course, has no closed form solution. However, it is easy to see that the rth smallest solution to this equation has to be in the interval $T\theta\epsilon(2\pi r, 2\pi(r+\frac{1}{2}))$, which proves the theorem □.

Proof of Theorem 6.4 It is not difficult to see that we will have to use (N+1) observations in the construction of a proper prior. Given the information set available at time N+1, we can write

$$\mu_{N+1} = y_{N+1} - x'_{N+1}\beta_{N+1} - \epsilon_{N+1}, \tag{6.A22a}$$
$$X^{+}\beta = y^{+} - y^{+}_{-1} + \epsilon^{+}_{-1} - \epsilon^{+} - \eta^{+}, \tag{6.A22b}$$

where $X^{+} = \begin{pmatrix} \Delta x'_2 \\ \vdots \\ \Delta x'_{N+1} \end{pmatrix}$, $y^{+} = \begin{pmatrix} y_2 \\ \vdots \\ y_{N+1} \end{pmatrix}$, $y^{+}_{-1} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$

and $\epsilon^{+} = \begin{pmatrix} \epsilon_2 \\ \vdots \\ \epsilon_{N+1} \end{pmatrix}$, $\epsilon^{+}_{-1} = \begin{pmatrix} \epsilon_1 \\ \\ \epsilon_N \end{pmatrix}$, $\eta^{+} = \begin{pmatrix} \eta_2 \\ \vdots \\ \eta_{N+1} \end{pmatrix}$.

Thus we have the result that

$$\begin{aligned} v_s &= y_s - \hat{y}_{s|N+1} \\ &= (x_s - x_{N+1})'(X^{+})^{-1}(\epsilon^{+}_{-1} - \epsilon^{+} - \eta^{+}) + \epsilon_s - \epsilon_{N+1} + \eta_{N+1} + \ldots + \eta_s \,. \end{aligned}$$

If we write $E = \begin{bmatrix} 1\ 1\ 1\ \ldots\ 1 \\ \\ \square \\ \\ \end{bmatrix}$ as a (N×N–n) matrix, and $D = \begin{bmatrix} 2 & -1 & 0 & \cdots & & 0 \\ -1 & 2 & -1 & \cdots & & 0 \\ 0 & -1 & 2 & \cdots & & 0 \\ \vdots & \vdots & \vdots & & 2 & -1 \\ 0 & 0 & 0 & & -1 & 1 \end{bmatrix}$,

then

$$v \sim N(0, \sigma^2(I + qA + ii' + X^{*}(X^{+})^{-1}(D+qI)(X^{+})^{-1\prime}X^{*\prime}$$

$$+X^{*}(X^{+})^{-1}E + E'(X^{+})^{-1\prime}X^{*\prime})$$

where $v=(v_{N+2},\ldots,v_T)'$ and $X^{*}= (x_{N+2}- x_{N+1},\ldots,x_T-x_{N+1})'$. Hence the log–likelihood function for $(y_1,y_2,\ldots,y_T)'$ is simply

$$l(\sigma^2,q)=-\frac{(T-N-1)}{2}\log\sigma^2-\frac{1}{2}\log|F+qG|-\frac{v'(F+qG)^{-1}v}{2}, \qquad \textbf{(6.A23)}$$

where $F = I + ii' + X^{*}(X^{+})^{-1}D(X^{+})^{-1\prime}X^{*\prime}+X^{*}(X^{+})^{-1}E + E'(X^{+})^{-1\prime}X^{*\prime}$ and
$G = A + X^{*}(X^{+})^{-1}(X^{+})^{-1\prime}X^{*\prime}$ □.

Proof of Theorem 6.5 Now

$$s(q) = -\tfrac{1}{2} \sum_{t=1}^{T-N-1} \lambda_t(P(q)(I+ii'+qA)^{-1/2}A(I+ii'+qA)^{-1/2}P(q))$$

$$+ \frac{(T-N-1)}{2} \frac{z'Q(q)'(I+ii'+qA)^{-1}A(I+ii'+qA)^{-1}Q(q)z}{z'Q(q)'(I+ii'+qA)^{-1}Q(q)z}. \tag{6.A24}$$

Using the same methodology as in the proof of theorem 6.2, we can manipulate the ratio of quadratic forms under the assumption that $q^* = 0$. The sum of eigenvalues can be tackled in the same way using a lower, rather than upper, bound on the eigenvalues. This means that this sum is greater than

$$\sum_{t=N+1}^{T-1} \left[\frac{\lambda_t}{(1+q\lambda_t)}\right]$$

$$\to \sum_{t=N+1}^{T-1} \frac{1}{(c+t^2\pi^2)}, \text{ as } T\to\infty, \text{ for } q = c/T^2.$$

So,

$$\lim_{T\to\infty} \ell(1/T^2\, s(c/T^2;0))$$

$$\geq \ell\left[\frac{1}{2}\sum_{t=1}^{\infty} u_t^2 \frac{t^2\pi^2}{(c+t^2\pi^2)^2} - \frac{1}{2}\sum_{t=N+1}^{\infty} \frac{1}{c+t^2\pi^2}\right] \tag{6.A25}$$

$$= \ell(X(c)).$$

$$\text{Now } EX(c) = c \sum_{t=N+1}^{\infty} \frac{-1}{(c+t^2\pi^2)^2} - \frac{1}{2} \sum_{t=1}^{N} \frac{t^2\pi^2}{(c + t^2\pi^2)^2} \tag{6.A26}$$

and both $\sum_{t=N+1}^{\infty} \frac{1}{(c+t^2\pi^2)^2}$ and $-\frac{1}{2} \sum_{t=1}^{N} \frac{t^2\pi^2}{(c + t^2\pi^2)^2}$ are bounded for all c. Hence the expectation is of order c^2. However, the variance of X(c) is simply

$$\sum_{t=1}^{\infty} \frac{t^4\pi^4}{(c+t^2\pi^2)^4}, \text{ which is bounded for all c.}$$

Thus, following the proof of theorem 5.1, the theorem is proved using Chebyshev's theorem by allowing c to be chosen arbitrarily large.□

Proposition 6.1 $(I+ii'+qA)^{-1/2}A(I+ii'+qA)^{-1/2}$ and $(I+ii'+qA)^{-1/2}(I+ii')(I+ii'+qA)^{-1/2}$ have the same eigenvectors.

Proof. $(I+ii'+qA)^{-1/2}A(I+ii'+qA)^{-1/2}$ is of full rank. So $\exists$ P such that $P'P = I_{T-1}$, and Δ is a $(T-1)\times(T-1)$ diagonal matrix, such that

$$(I+ii'+qA)^{-1/2}A(I+ii'+qA)^{-1/2}\,P = P\,\Delta. \tag{6.A27}$$

Writing $P^* = (I+ii'+qA)^{1/2}P$, we have that

$$(I+ii')A^{-1}P^* + qP^* = P^*\Delta^{-1}. \tag{6.A28}$$

Equally, $(I+ii'+qA)^{-1/2}(I+ii')(I+ii'+qA)^{-1/2}$ is of full rank. So $\exists$ R such that $P'P = I_{T-1}$, and Λ is $(T-1)\times(T-1)$ diagonal matrix such that

$$(I+ii'+qA)^{-1/2}(I+ii')(I+ii'+qA)^{-1/2}\,R = R\,\Lambda. \tag{6.A29}$$

So writing $R^* = (I+ii'+qA)^{1/2}R$, we have that

$$(I+ii'+qA)(I+ii')^{-1}\,R^* = R^*\,\Lambda^{-1}. \tag{6.A30}$$

Thus by comparing (6.A30) with (6.A28) we have proved the theorem by using trivial eigenvalue manipulations.□

Chapter Seven

On the Probability of Estimating Deterministic Components in the Local Linear Trend Model

Summary The link between the local linear trend and an unconstrained second order moving average process is established. This suggests there maybe three cases which generate irregular asymptotic results.

Kitagawa's trend model is analyzed in some detail. An approximation to the log–likelihood is suggested for this model. The result is that we can prove T^4–consistency on the ML estimator of the signal–noise ratio on the slope component (β_t).

The local linear trend is much harder to analyze as the score vector has two components. Consistency results are established for the ML estimator of the two signal–noise ratios under various assumptions. Finally, the probability of estimating deterministic components in this model is found.

Key Words CHARACTERISTIC FUNCTIONS, NUMERICAL INVERSIONS, SIGNAL NOISE RATIOS, NONINVERTIBILITY OF MOVING AVERAGES.

(7.1) Introduction.

The local linear trend model has provided the framework for much of the work in the structural modelling of time series which has occurred in the last decade. The model is given by

$$y_t = \mu_t \qquad\qquad + \epsilon_t,\ \epsilon_t \sim \mathrm{NID}(0, \sigma^2) \tag{7.1a}$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t,\ \eta_t \sim \mathrm{NID}(0, q\sigma^2) \tag{7.1b}$$
$$\beta_t = \qquad \beta_{t-1} + \zeta_t,\ \zeta_t \sim \mathrm{NID}(0, p\sigma^2) \tag{7.1c}$$

where (ϵ_t), (η_t) and (ζ_t) are assumed to be independent. (μ_t) can be viewed as a stochastic level, while (β_t) is a stochastic increment term.

The local linear trend can be viewed as a rationalization of Holt's(1957) (see Chatfield(1984)) forecasting scheme. Its natural interpretability has meant that it has found applications as a part of a wider class of models used in the work of, for instance, Harrison and Stevens(1976), Harvey(1989), Clark(1987) and Watson(1986).

This paper deals with the important case where the initialization of the model takes the form of a bivariate diffuse prior, where

$$\begin{bmatrix} \mu_0 \\ \beta_0 \end{bmatrix} \sim \mathrm{N}(0, \kappa I),\ \kappa \to \infty \tag{7.2}$$

with μ_0 and β_0 assumed to be totally independent of the noise in (7.1).

Of course, when twice differenced, this model has a restricted moving average representation. To understand this we look at the autocovariance generating function (AGF) of an unrestricted MA(2) model and the twice differenced local linear trend. The parameters of the MA(2) process will be taken to follow the convention that the associated characteristic equation will have roots which are on or inside the unit circle. This is merely an identification condition. $\gamma(t)$ will denote the t–th autocovariance of the random variable of interest.

The AGF of the MA(2) $z_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$ (where u_t is defined as white noise, with variance σ^{+2}) is

$$g(L) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\, L^{\tau} = \sigma^{+2}[\{1 + \theta_1 L + \theta_2 L^2\}\{1 + \theta_1 L^{-1} + \theta_2 L^{-2}\}]$$

$$= \sigma^{+2}[\{1 + \theta_1^2 + \theta_2^2\} + L\{\theta_1 + \theta_1\theta_2\} + L^2\theta_2 + L^{-1}\{\theta_1 + \theta_1\theta_2\} + L^{-2}\theta_2], \tag{7.3}$$

while for the twice differenced local linear trend, we have

$$g(L) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\, L^{\tau} = \sigma^2[\{(1-L)(1-L^{-1})q\} + p + \{(1-L)^2(1-L^{-1})^2\}]$$

$$= \sigma^2[\{2q+p+6\} + L\{-q-4\} + L^{-1}\{-q-4\} + L^2 + L^{-2}]. \tag{7.4}$$

Equating the first two autocorrelations achieves the following relationships:

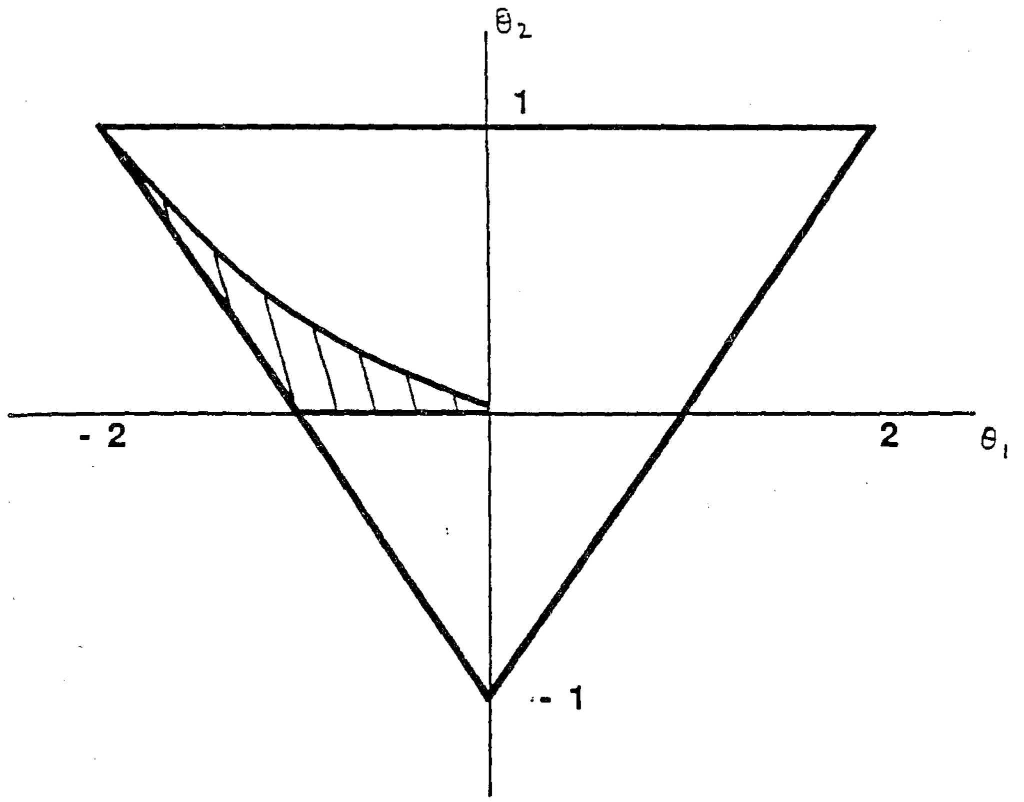
Theorem 7.1

$$q = -4 - \frac{\{\theta_1(1+\theta_2)\}}{\theta_2}, \quad p = \frac{\{1 + \theta_1^2 + \theta_2^2\}}{\theta_2} - 2q - 6 \tag{7.5}$$

Proof Given in the appendix.

These relationships will allow us to display the precise relationship between the parameters of the unrestricted MA(2) process and the differenced local linear trend. This is best done in diagram 7.1, which displays the region of invertibility of the unrestricted MA(2) (see theorem 4.2, pg 172, of Goldberg(1958)), and the corresponding permissible parameter space for the local linear trend. This is similar to the one given by Godolphin and Stone(1980). Note their results are concerned with a slightly different local linear trend.

Diagram 7.1



Shaded area is given by the invertibility convention, $\theta_1 \le 0$, $\theta_2 \ge 0$ and the equation

$$0 = -4 - \frac{\{\theta_1(1+\theta_2)\}}{\theta_2} .$$

The diagram 7.1 allows us to pick out the main features of the various stochastic trend models proposed in the literature. In particular, we can see that the local linear trend model considerably restricts the permissible parameter space. Further, the imposition of q = 0, as suggested by Kitagawa(1981) and Kitagawa and Gersch(1984), has an enormous effect on this space.

The diagram suggests that there will be three non–standard distribution problems associated with the ML estimation of this model:–

(i) under the restriction q = 0, if the true value of p is zero,

(ii) if the true value of q is positive, but the true value of p is zero,

(iii) if the true value of q and p are both zero.

It turns out that each of these problems will give rise to significantly different results.

The third case is of particular importance for it corresponds to the estimation of two unit roots for the restricted MA(2) model. This is of interest as no researchers have been able to make any progress with the unrestricted model.

The distributional work associated with the estimation of noninvertible MA(2) is very difficult and as a result, little progress has been made in this area. Only Anderson and Takemura(1986) and Tanaka and Satchell(1987) have written in this area and even they have made limited progress. Tanaka and Satchell's work is the most advanced of the two. In their paper they analyzed the following model

$$y_t = \epsilon_t + \theta\, \epsilon_{t-q}, \tag{7.6}$$

where (ϵ_t) is a gaussian white noise process. Thus (7.6) is a constrained MA(q) process. If we look at diagram 7.1 we can see that their analysis for the case q = 2 is a very specialized one. We can also see that their work gives us few clues for the analysis we want to pursue because it does not conform to the shaded area which is of interest to us.

(7.2) Kitagawa's Local Trend Model

Suppose we think of the local linear trend model, given by equation (7.1) and (7.2). The log–likelihood for $(y_1,...,y_T)'$ is the same, apart from a log k term (see de Jong(1988)), as that for $v=(v_3,...,v_T)'$, where

$$\begin{aligned} v_t &= \Delta^2 y_t \\ &= y_t - 2y_{t-1} + y_{t-2}, \qquad (t=3,...,T) \end{aligned} \qquad (7.7)$$

It is therefore not difficult to show that

$$v \sim N(0, \sigma^2(B+qA+pI)), \text{ where} \qquad (7.8)$$

$$B = \begin{bmatrix} 6 & -4 & 1 & 0 & . & . & . & 0 & 0 & 0 \\ -4 & 6 & -4 & 1 & . & . & . & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & . & . & . & 0 & 0 & 0 \\ 0 & 1 & -4 & 6 & . & . & . & 0 & 0 & 0 \\ . & . & . & . & & & & . & . & . \\ . & . & . & . & & & & . & . & . \\ . & . & . & . & & & & . & . & . \\ 0 & 0 & 0 & 0 & . & . & . & 6 & -4 & 1 \\ 0 & 0 & 0 & 0 & . & . & . & -4 & 6 & -4 \\ 0 & 0 & 0 & 0 & . & . & . & 1 & -4 & 6 \end{bmatrix}, \qquad (7.9)$$

$$\text{and} \qquad A = \begin{bmatrix} 2 & -1 & 0 & . & . & . & 0 & 0 \\ -1 & 2 & -1 & . & . & . & 0 & 0 \\ 0 & -1 & 2 & . & . & . & 0 & 0 \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ 0 & 0 & 0 & . & . & . & 2 & -1 \\ 0 & 0 & 0 & . & . & . & -1 & 2 \end{bmatrix}. \tag{7.10}$$

When q is constrained to be zero the model corresponds to the Kitagawa(1981) and Kitagawa and Gersch(1984) trend model. Young(1984, pg 74) calls this model a double integrated random walk. It can also be shown to be a cubic spline (see Wecker and Ansley(1983)) for equally spaced observations. Imposing this constraint tends to result in a smoother estimated trend (see Harvey(1989, ch. 6.1)).

The log–likelihood function is

$$l(\sigma^2,p) = -\frac{(T-2)}{2}\log\sigma^2 - \frac{1}{2}\log|B+pI| - \frac{v'(B+pI)^{-1}v}{2\sigma^2} \tag{7.11}$$

and this yields the following profile log–likelihood by concentrating σ^2 out of the log–likelihood,

$$M(p) = -\frac{(T-2)}{2}\log v'(B+pI)^{-1}v - \frac{1}{2}\log|B+pI|. \tag{7.12}$$

The probability law of the score is

$$\ell\left[\frac{dM(p)}{dp}\right] = \ell\left[\frac{(T-2)}{2}\,\frac{v'(B+pI)^{-1}(B+pI)^{-1}v}{v'(B+pI)^{-1}v} - \frac{1}{2}\,\mathrm{tr}\left[(B+pI)^{-1}\right]\right]$$

$$= \ell\left[\frac{(T-2)}{2}\frac{\sum_{t=1}^{T-2}\frac{u_t^2(p^*+\delta_t^+)}{(p+\delta_t^+)^2}}{\sum_{t=1}^{T-2}\frac{u_t^2(p^*+\delta_t^+)}{(p+\delta_t^+)}} - \frac{1}{2}\sum_{t=1}^{T-2}\frac{1}{(p+\delta_t^+)}\right], \tag{7.13}$$

where δ_t^+ denotes the t–th eigenvalue of B, $u_t \sim NID(0,1)$, and p^* denotes the true value of p.

The probability of having an estimated value of p being zero is

$$Pr(p^*) = Pr\left[\frac{dM(p)}{dp}\bigg|_{p=0} < 0\right]. \tag{7.14}$$

Some convenient simplifications can be achieved if we analyze the structure of the matrix B. Marshall(1989) noted that if we write

$$F = \begin{bmatrix} 0 & 1 & 0 & . & . & . & 0 & 0 \\ 1 & 0 & 1 & . & . & . & 0 & 0 \\ 0 & 1 & 0 & . & . & . & 0 & 0 \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ 0 & 0 & 0 & . & . & . & 0 & 1 \\ 0 & 0 & 0 & . & . & . & 1 & 0 \end{bmatrix} \tag{7.15}$$

then $B = 4I + F^2 - 4F + G$, where

$$G = \begin{bmatrix} 1 & 0 & 0 & . & . & . & 0 & 0 \\ 0 & 0 & 0 & . & . & . & 0 & 0 \\ 0 & 0 & 0 & . & . & . & 0 & 0 \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ . & . & . & & & & . & . \\ 0 & 0 & 0 & . & . & . & 0 & 0 \\ 0 & 0 & 0 & . & . & . & 0 & 1 \end{bmatrix}. \tag{7.16}$$

Hence, if we were to make the mild assumptions that $\epsilon_1=0$ and $\epsilon_T=0$, we can use the eigenvalues of F (c.f. Anderson(1971, theorem 6.5.5)) to deduce that the exact eigenvalues of $4I + F^2 - 4F$ are

$$\delta_t = 4 + w_t^2 - 4w_t, \text{ where } w_t = 2\cos\frac{t\pi}{T-1}$$

$$= (2 - w_t)^2 \tag{7.17}$$

$$= \left[4\sin^2\frac{\pi t}{2(T-1)}\right]^2, \quad (t=1,\ldots,T-2) \tag{7.18}$$

We can either calculate δ_t^+ numerically, or use the approximate values δ_t, when we come to compute $\Pr(p^*)$. We showed in chapter five that this probability is not sensitive to these types of assumptions and so we use δ_t here. It is easy to show that

$$\Pr(p^*) = \Pr\left[\sum_{t=1}^{T-2} u_t^2 \frac{(p^*+\delta_t)}{\delta_t}[\delta_t^{-1} - \overline{\delta^{-1}}] < 0\right] \tag{7.19}$$

where $\overline{\delta^{-1}} = \dfrac{1}{(T-2)}\displaystyle\sum_{t=1}^{T-2} \delta_t^{-1}$.

Using Davies'(1980) algorithm we can compute the following table.

Table 7.1

Tabulation of $\Pr(p^*)$ for Kitagawa's Trend Model

True Value of p

T	0	0.001	0.01	0.1	1.0
10	.649	.634	.534	.276	.115
20	.667	.477	.207	.068	.028
30	.673	.268	.087	.023	.016
40	.675	.155	.035	.015	.015
50	.677	.094	.019	.012	.012

Of particular interest is the case where $p^*=0$. For this case we can use the fact that

$$\delta_t = \frac{\pi^4 t^4}{T^4} + 0(T^{-6}) \tag{7.20}$$

to see that

$$\lim_{T\to\infty} \Pr(0) = \Pr\left[\sum_{t=1}^{\infty} \frac{u_t^2}{t^4\pi^4} < \sum_{t=1}^{\infty} \frac{1}{t^4\pi^4}\right]. \tag{7.21}$$

It is known (c.f. Abramowitz and Stegun(1965, pg 807)) that

$$\sum_{t=1}^{\infty} t^{-4} = \frac{(2\pi)^4}{2\,.\,4\,!}\ |B_4|, \tag{7.22}$$

where B_4 is the 4–th Bernoulli number. This is $\frac{-1}{30}$, so

$$\lim_{T\to\infty} \Pr(0) = \Pr\left[\sum_{t=1}^{\infty} \frac{u_t^2}{t^4\pi^4} < \frac{1}{90}\right]. \tag{7.23}$$

The characteristic function of $\sum_{t=1}^{\infty} \frac{u_t^2}{t^4\pi^4}$ is known to be (c.f. Tanaka and Satchell(1987))

$$\varphi(\theta) = \left[\frac{\sin(2i\theta)^{1/4}}{(2i\theta)^{1/4}} \cdot \frac{\sin h(2i\theta)^{1/4}}{(2i\theta)^{1/4}}\right]^{-1/2}. \tag{7.24}$$

Using Knott's(1974) numerical inversion this value is 0.6815.

Finally, in this section, the following theorem will be proved.

<u>**Theorem 7.2**</u> When $p^*=0$, the ML estimator of p in $0_p(T^{-4})$, i.e. for any $\epsilon>0$, $\exists\, T_o$ and c>0 such that

$$p\left[\frac{dM(p)}{dp}\bigg|_{p=c/T^4} > 0\right] < \epsilon, \forall\, T > T_o. \tag{7.25}$$

<u>Proof.</u> Given in the appendix.

(7.3) The Local Linear Trend Model

When both p and q are unknown, the log–likelihood is

$$l(p,q,\sigma^2) = -\frac{(T-2)}{2}\log\sigma^2 - \frac{1}{2}\log|B+qA+pI)| - \frac{v'(B+qA+pI)^{-1}v}{2\sigma^2} \tag{7.26}$$

giving the profile log–likelihood as

$$M(p,q) = -\frac{(T-2)}{2}\log v'(B+qA+pI)^{-1}v - \frac{1}{2}\log|B+qA+pI|. \tag{7.27}$$

The score vector has two elements; the first is

$$\frac{dM(p,q)}{dq} = \frac{(T-2)}{2}.\frac{v'(B+qA+pI)^{-1}A(B+qA+pI)^{-1}v}{v'(B+qA+pI)^{-1}v} - \frac{1}{2}\,\mathrm{tr}\left[(B+qA+pI)^{-1}A\right], \tag{7.28}$$

while the second is

$$\frac{dM(p,q)}{dp} = \frac{(T-2)}{2}.\frac{v'(B+qA+pI)^{-1}(B+qA+pI)^{-1}v}{v'(B+qA+pI)^{-1}v} - \frac{1}{2}\,\mathrm{tr}\left[(B+qA+pI)^{-1}\right]. \tag{7.29}$$

As we can write

$$A = 2I - F, \tag{7.30}$$

if we were to assume that $\epsilon_1=0$ and $\epsilon_T=0$, the probability law of the first element of the score would be

$$\ell\left[\frac{(T-2)}{2}\cdot\frac{\displaystyle\sum_{t=1}^{T-2}\frac{u_t^2\lambda_t(\lambda_t^2+q^*\lambda_t+p^*)}{(\lambda_t^2+q\lambda_t+p)^2}}{\displaystyle\sum_{t=1}^{T-2}\frac{u_t^2(\lambda_t^2+q^*\lambda_t+p^*)}{(\lambda_t^2+q\lambda_t+p)}}-\frac{1}{2}\sum_{t=1}^{T-2}\frac{\lambda_t}{(\lambda_t^2+q\lambda_t+p)}\right], \tag{7.31}$$

where λ_t are the eigenvalues of A and are such that $\delta_t=\lambda_t^2$. The second element of the score is

$$\ell\left[\frac{(T-2)}{2}\cdot\frac{\displaystyle\sum_{t=1}^{T-2}\frac{u_t^2(\lambda_t^2+q^*\lambda_t+p^*)}{(\lambda_t^2+q\lambda_t+p)^2}}{\displaystyle\sum_{t=1}^{T-2}\frac{u_t^2(\lambda_t^2+q^*\lambda_t+p^*)}{(\lambda_t^2+q\lambda_t+p)}}-\frac{1}{2}\sum_{t=1}^{T-2}\frac{1}{(\lambda_t^2+q\lambda_t+p)}\right]. \tag{7.32}$$

Hence, the probability that both q and p are estimated to be zero is

$$\Pr\left[\begin{array}{l} \sum_{t=1}^{T-2} \dfrac{u_t^2 \ (\lambda_t^{\ 2}+q^* \lambda_t+p^*)}{\lambda_t^{\ 2}} \ (\lambda_t^{-1} - \overline{\lambda^{-1}}) < 0 \\ \sum_{t=1}^{T-2} \dfrac{u_t^2(\lambda_t^{\ 2}+q^* \lambda_t+p^*)}{\lambda_t^{\ 2}} \ (\lambda_t^{-2} - \overline{\lambda^{-2}}) < 0 \end{array}\right]. \qquad \textbf{(7.33)}$$

This probability can be evaluated exactly using the routines discussed in chapter 3. The results of these calculations are given below. The marginal probabilities are computed using Davies'(1980) algorithm for evaluating the distribution function of a univariate random variable.

The joint probability reported here is of some interest as it represents the probability that a deterministic trend is estimated from the data. The marginal probabilities are unfortunately less informative. They are upper bounds on the probability that the parameter is estimated to be exactly zero.

Tabulation of the Probability of Boundary Cases

Sample Size of 10

Parameter Value		Probability of a zero Estimate*		
q	p	p	q	p and q
.00	.000	0.649	0.632	0.615
.00	.001	0.634	0.618	0.601
.00	.010	0.534	0.520	0.503
.01	.000	0.632	0.618	0.598
.01	.001	0.619	0.602	0.585
.01	.010	0.526	0.510	0.494
.10	.000	0.529	0.509	0.493
.10	.001	0.521	0.501	0.485
.10	.010	0.464	0.445	0.430

the marginal probabilities are upper bounds. See the text.

Sample Size of 20

Parameter Value		Probability of a zero Estimate		
q	p	p	q	p and q
.00	.000	0.668	0.647	0.625
.00	.001	0.478	0.461	0.442
.00	.010	0.207	0.180	0.172
.01	.000	0.597	0.573	0.553
.01	.001	0.451	0.429	0.412
.01	.010	0.205	0.176	0.169
.10	.000	0.365	0.318	0.307
.10	.001	0.325	0.280	0.270
.10	.010	0.192	0.150	0.146

Sample Size of 30

Parameter Value		Probability of a zero Estimate		
q	p	p	q	p and q
.00	.000	0.672	0.650	0.628
.00	.001	0.270	0.248	0.236
.00	.010	0.080	0.052	0.039
.01	.000	0.531	0.502	0.482
.01	.001	0.259	0.230	0.220
.01	.010	0.080	0.051	0.038
.10	.000	0.260	0.194	0.187
.10	.001	0.197	0.139	0.136
.10	.010	0.077	0.045	0.032

Sample Size of 40

Parameter Value		Probability of a zero Estimate		
q	p	p	q	p and q
.00	.000	0.675	0.652	0.629
.00	.001	0.156	0.124	0.121
.00	.010	0.038	0.022	0.009
.01	.000	0.464	0.428	0.411
.01	.001	0.150	0.114	0.112
.01	.010	0.038	0.022	0.009
.10	.000	0.194	0.119	0.112
.10	.001	0.117	0.066	0.063
.10	.010	0.038	0.021	0.009

Sample Size of 50

Parameter Value		Probability of a zero Estimate		
q	p	p	q	p and q
.00	.000	0.677	0.653	0.630
.00	.001	0.087	0.059	0.050
.00	.010	0.027	0.016	0.004
.01	.000	0.404	0.361	0.347
.01	.001	0.091	0.058	0.057
.01	.010	0.018	0.005	0.004
.10	.000	0.147	0.077	0.075
.10	.001	0.074	0.034	0.033
.10	.010	0.019	0.005	0.004

These tables indicate that the true value of p is very important in determining the probability that p and/or q are estimated to be zero. Although this is not particularly surprising, it is interesting that the value of p does change the probability of a boundary value for q so dramatically. It would seem that the estimation procedure does sometimes mix up the presence of variation due to noise on the μ_t term and that due to noise on the β_t term.

Finally, we prove the following consistency theorems.

<u>Theorem 7.3</u> When $q^*>0$ is known, and if $p^*=0$ then the ML estimator of p is $O_p(T^{-2})$, i.e. for any $\epsilon>0$, $\exists$ T_o and $c>0$ such that

$$\Pr\left[\frac{dM(q,p)}{dp}\bigg|_{q=q^*,p=c/T^2} > 0\right] < \epsilon, \forall\, T > T_o. \tag{7.34}$$

<u>Proof</u> Given in the appendix.

<u>Corollary 7.1</u> When $q^*>0$ but is unknown, and if $p^*=0$ then the ML estimator of p is $O_p(T^{-2})$, i.e. for any $\epsilon>0$, $\exists$ T_o and $c>0$

$$\Pr\left[\frac{dM(q,p)}{dp}\bigg|_{q=\hat{q},\, p=c/T^2} > 0\right] = \epsilon, \forall T > T_o \tag{7.35}$$

where $\hat{q}$ is the ML estimator of q.

<u>Proof.</u> **Follows precisely the arguments of the proof of theorem 3.1 as $\hat{q}=q^2+0_p(1)$ even when p^2 is possibly zero; see Dunsmuir and Hannan(1976) □.**

<u>Theorem 7.4</u> If $p^{*}=q^{*}=0$, then the ML estimator of q is $O_p(T^{-2})$ and the ML estimator of p is $O_p(T^{-4})$, i.e. for any η, $\epsilon>0$, $\exists$ T_o and d, c>0 such that

$$P_1 = \Pr\left[\left.\frac{dM(p,q)}{dq}\right|_{\substack{q=c/T^2 \\ p=d/T^4}} > 0\right] < \eta, \text{ and} \tag{7.36}$$

$$P_2 = \Pr\left[\left.\frac{dM(p,q)}{dp}\right|_{\substack{q=c/T^2 \\ p=d/T^4}} > 0\right] < \epsilon, \ \forall\ T>T_0. \tag{7.37}$$

<u>Proof.</u> Given in the appendix.

These theorems indicate that the rate at which the estimator of p approaches zero is very rapid. This speed is accelerated when the true value of q is also zero.

Mathematical Appendix

Proof of Theorem 7.1 Equating the first two autocorrelations we have

$$\frac{1}{2q+p+6} = \frac{\theta_2}{1+\theta_1^2+\theta_2^2} \text{ and } \frac{-(q+4)}{2q+p+6} = \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} \tag{A7.1}$$

So $-q-4 = \dfrac{\theta_1(1+\theta_2)}{\theta_2}$, which gives the first result, while

$p+6 = -2p + \dfrac{1+\theta_1^2+\theta_2^2}{\theta_2}$, allows the second result □.

Proof of Theorem 7.2

The probability law of the score, when $p^*=0$, is

$$\ell\left[T^{-4}\, s(p;0)\right] = \ell\left[\frac{(T-2)}{2}\frac{\sum_{t=1}^{T-2}\frac{u_t^2\,\delta_t^+}{(p+\delta_t^+)^2}}{\sum_{t=1}^{T-2}\frac{u_t^2\,\delta_t^+}{(p+\delta_t^+)}} - \frac{1}{2}\sum_{t=1}^{T-2}\frac{1}{(p+\delta_t^+)}\right]$$

(A7.2)

Noting $\frac{1}{(T-2)}\sum_{t=1}^{T-2}\frac{u_t^2\,\delta_t}{(p_t+\delta_t)}\bigg|_{p=c/T^4} \overset{p}{\to} 1,$ **(A7.3)**

we have that $\lim_{T\to\infty}\ell\left[T^{-4}\, s(p);0\right] = \frac{1}{2}\,\ell\left[\sum_{t=1}^{T-2}\frac{u_t^2\,\pi^4 t^4}{(c+\pi^4 t^4)^2} - \sum_{t=1}^{T-2}\frac{1}{c+\pi^4 t^4}\right].$

(A7.4)

Following theorem 5.1, the proof is completed □.

Proof of Theorem 7.3

Recall $\ell\left[\frac{dM(q^*,p)}{dp}\right]$

$$= \ell\left[\frac{(T-2)}{2}\frac{\sum_{t=1}^{T-2}\frac{u_t^2(\lambda_t^2+q^*\lambda_t)}{(\lambda_t^2+q^*\lambda_t+p)^2}}{\sum_{t=1}^{T-2}\frac{u_t^2(\lambda_t^2+q^*\lambda_t)}{(\lambda_t^2+q^*\lambda_t+p)}} - \frac{1}{2}\sum_{t=1}^{T-2}\frac{1}{(\lambda_t^2+q^*\lambda_t+p)}\right] \tag{A7.5}$$

But $\frac{1}{T-2}\sum_{t=1}^{T-2}\frac{u_t^2(\lambda_t^2+q^*\lambda_t)}{(\lambda_t^2+q^*\lambda_t+(c/T^2))} \overset{p}{\to} 1,$

as $\frac{\lambda_t^2}{T^2} = \lambda_t\pi^2t^2 + 0(T^{-4}) = O(T^{-2})$, so

$$\lim_{T\to\infty}\ell\left[T^{-2}\frac{dM(q^*,p)}{dp}\right] = \ell\left[\frac{1}{2}\sum_{t=1}^{\infty}\frac{u_t^2q^*t^2\pi^2}{(q^*t^2\pi^2+c)^2} - \frac{1}{2}\sum_{t=1}^{\infty}\frac{1}{c+q^2t^2\pi^2}\right]. \tag{A7.6}$$

Using the usual argument, the theorem follows □.

<u>Proof of Theorem 7.4</u>

It is easy to see that $\lim\limits_{T\to\infty} \ell\left[T^{-4} \frac{dM(p,q)}{dq} \Bigg|_{\substack{\epsilon=c/T^2\\ p=d/T^4}} \right]$

$$= \ell\left[\frac{1}{2} \sum_{t=1}^{\infty} \frac{u_t^2 t^6 \pi^6}{(t^4\pi^4+ct^2\pi^2+d)^2} - \frac{1}{2} \sum_{t=1}^{\infty} \frac{t^2\pi^2}{(t^4\pi^4+ct^2\pi^2+d\ )} \right] = \ell(X(c,d)), \tag{A7.7}$$

and $\lim\limits_{T\to\infty} \ell\left[\frac{1}{T^4} \frac{dM(p,q)}{dp} \Bigg|_{\substack{q=c/T^2\\ p=d/T^4}} \right]$

$$= \ell\left[\frac{1}{2} \sum_{t=1}^{\infty} \frac{u_t^2 t^4 \pi^4}{(t^4\pi^4+ct^2\pi^2+d)^2} - \frac{1}{2} \sum_{t=1}^{\infty} \frac{1}{(t^4\pi^4+ct^2\pi^2+d\ )} \right] = \ell(Y(c,d)). \tag{A7.8}$$

So $EX(c,d) = -\frac{1}{2} \sum_{t=1}^{\infty} \frac{ct^4\pi^4+dt^2\pi^2}{(t^4\pi^4+ct^2\pi^2+d)^2}$,

$$< -\frac{c}{2} \sum_{t=1}^{\infty} \frac{t^4\pi^4}{(t^4\pi^4+ct^2\pi^2+d)^2} \tag{A7.9}$$

and $$VX(c,d) = \frac{1}{2} \sum_{t=1}^{\infty} \frac{t^{12}\pi^{12}}{(t^4\pi^4+ct^2\pi^2+d)^4}. \tag{A7.10}$$

So $\lim\limits_{T\to\infty} P_1 < \frac{VX(c,d)}{[E(c,d)]^2}$

$$< \frac{\dfrac{2}{c^2} \displaystyle\sum_{t=1}^{\infty} \frac{t^{12}\pi^{12}}{(t^4\pi^4+ct^2\pi^2+d)^4}}{\displaystyle\sum_{t=1}^{\infty} \frac{t^4\pi^4}{(t^4\pi^4+ct^2\pi^2+d)^2}} = \frac{2}{c^2} g(c,d). \qquad \textbf{(A7.11)}$$

But g(c,d) is bounded from above for all c and d, so $\lim\limits_{T\to\infty} P_1$ can be made arbitrarily small by choosing a large enough c.

Having chosen c, we can now work on the value of d.

$$EY(c,d) = -\frac{1}{2}\sum_{t=1}^{\infty} \frac{ct^2\pi^2+d}{(t^4\pi^4+ct^2\pi^2+d)^2}$$

$$-\frac{d}{2}\sum_{t=1}^{\infty} \frac{1}{(t^4\pi^4+ct^2\pi^2+d)^2}, \qquad \textbf{(A7.12)}$$

and $$VY(c,d) = \frac{1}{2}\sum_{t=1}^{\infty} \frac{t^8\pi^8}{(t^4\pi^4+ct^2\pi^2+d)^4} \qquad \textbf{(A7.13)}$$

So $\lim\limits_{t\to\infty} P_2$ can be made arbitrarily small by choosing d appropriately large □.

References.

Abramowitz,M. and Stegun,I.A.(1965) "Handbook of mathematical functions," Dover Publications, New York.

Abril,J.C.(1985) "Asymptotic expansions for time series problems with applications to moving average models," Ph.D. thesis, University of London.

Anderson,T.W.(1971) "The Statistical Analysis of Time Series," John Wiley and Sons, New York.

Anderson,T.W. and Darling,D.A.(1952) "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," Annals of Mathematical Statistics 23 pg. 193–212.

Anderson,T.W. and Takemura,A.(1986) "Why do noninvertible estimated moving averages occur?" Journal of Time Series Analysis. 7 pg. 235–254.

Ansley,C.F. and Newbold,P.(1980) "Finite Sample Properties of Estimators for autoregressive moving average models," Journal of Econometrics 13 pg. 159–183.

Ansley,C.F. and Kohn,R.(1985) "Estimation, filtering and smoothing in state space models with incompletely specified initial conditions," Annals of Statistics 13 pg. 1286–1316.

Bohmann,H.(1961) "Approximate Fourier analysis of distribution functions," Arkiv för Matematik 4 pg. 99–1157.

Bohmann,H.(1970) "A method to calculate the distribution function when the characteristic function is known," Nordisk Tidskr. Informationsbehandling (BIT) 10 pg. 237–42.

Bohmann,H.(1972) "From characteristic function to distribution function via Fourier analysis," Nordisk Tidskr. Informationsbehandling (BIT) 12 pg. 237–42.

Brown,P.,Kleidon,A.W.,Marsh,T.A.(1983) "New Evidence on the Nature of Size–Related Anomalies in Stock Prices" Journal of Financial Economics 12 pg. 33–56.

Chatfield,C.(1984) "The analysis of time series, an introduction," Chapman and Hall.

Clark,P.K.(1987) "The Cyclical Component of U.S. Economic Activity," The Quarterly Journal of Economics 52 pg. 797–814.

Cooley,T.F. and Prescott,E.C.(1973) "Systematic (non–random) variation models varying parameter regression: a theory and some applications" Annals of Economic and Social Measurement 2 pg. 463–473.

Cooley,T.F. and Prescott,E.C.(1976) "Estimation in the presence of stochastic parameter variation," Econometrica 44 pg. 167–84.

Cooley,T.F.(1975) "A Comparison of Robust and Varying Parameter Estimates of a Macro–Econometric Model," Annals of Economic and Social Measurement 4 pg. 373–388.

Cooper,D.M. and Thompson,R.(1977) "A note on the estimation of the parameters of the autoregressive–moving average process," Biometrika 64 pg. 625–8.

Cox,D.R. and Hinkley,D.V.(1974) "Theoretical Statistics," Chapman and Hall.

Cramér,H.(1946) "Mathematical methods of statistics," Princeton University Press, Princeton.

Cryer,J.D. and Ledolter,J.(1981) "Small–sample properties of the maximum likelihood estimator in the first–order moving average model," Biometrika 68 pg. 691–694.

Davidson,J.(1981a) "Small sample properties of estimators of the moving average process," in E.G.Charatsis(1979), ed. "Proceedings of the Econometric Society European Meeting 1979: Selected econometric papers in memory of Stephan Valananis," North–Holland Amsterdam.

Davidson,J.(1981b) "Problems with the estimation of moving average processes," Journal of Econometrics 16 pg. 295–310.

Davies,R.B.(1973) "Numerical inversion of a characteristic function," Biometrika 60 pg. 415–417.

Davies,R.B.(1980) "AS 155: The distribution of a linear combination of χ^2 random variables," Applied Statistics 29 pg. 323–333.

de Jong,P.(1988) "The likelihood for a state space model," Biometrkia 75 pg. 165–9.

Dent,W. and Min,A–S.(1978) "A Monte Carlo study of autoregressive integrated moving average processes," Journal of Econometrics 7 pg. 23–55.

Durbin,J.(1980) "Approximations for densities of sufficient estimators," Biometrika 67 pg. 311–33.

Durbin,J. and Watson,G.S.(1971) "Testing for Serial Correlation in Least Squares Regression. III," Biometrika 58 pg. 1–19.

Dunsmuir,W. and Hannan,E.J.(1976) "Vector linear time series models," Advances in Applied Probability 8 pg. 339–364.

Enns,P.G., Mackak,J.A., Spivey,W.A. and Wrobleski,W.J.(1982) "Forecasting applications of an adaptive multiple exponential smoothing model," Management Science 28 pg. 1035–44.

Evans,G.B.A. and Savin,N.E.(1984) "Testing for unit roots 2." Econometrica 52 pg. 1241–1269.

Fama,E.F. and Gibbons,M.R.(1982) "Inflation, Relative Returns and Capital Investment," Journal of Monetary Economics 9 pg. 297–323

Farebrother,R.W.(1980) "AS 153: Pan's procedure for the tail probabilities of the Durbin–Watson statistic," Applied Statistics 29 pg. 224–227.

Farebrother,R.W.(1981) "Correction to AS 153, Pan's procedure for the tail probabilities of the Durbin–Watson statistic," Applied Statistics 30 pg. 189.

Farebrother,R.W.(1984a) "AS 204: The Distribution of a Positive Linear Combination of χ^2 Random Variables," Applied Statistics 33 pg. 332–339.

Farebrother,R.W.(1984b) "The distribution of a linear combination of central χ^2 random variables. A remark on AS 153: Pan's procedure for the tail probabilities of the Durbin–Watson statistic," Applied Statistics 33 pg. 363–366.

Farebrother,R.W.(1984c) "A remark on Algorithms AS106, AS 153 and AS 155. The distribution of a linear combination of χ^2 random variables, Applied Statistics 33 pg. 366–369.

Farebrother,R.W.(1985) "Eigenvalue–free methods for computing the distribution of a quadratic form in normal variables," Statistische Hefte 26 pg. 287–302.

Farebrother,R.W.(1989) "The Distribution of a Quadratic Form in Normal Variables," Discussion Papers in Computing CP101, Department of Econometrics and Social Statistics, University of Manchester.

Feller,W.(1966) "Introduction to Probability Theory and its Applications, Volume Two" 1st edition, John Wiley and Sons.

Feller,W.(1971) "Introduction to Probability Theory and its Applications, Volume Two" 2nd edition, John Wiley and Sons.

Fernandez–Macho,F.J. and Harvey,A.C.(1989) "Seemingly unrelated time series equations and a test for homogeneity," Journal of Business and Economic Statistics (to appear)

Gil–Pelaez,J.(1951) "Note on the Inversion Theorem," Biometrika 37 pg. 481–2

Godolphin,E.J. and Stone,J.M.(1980) "On the structural representation for polynomial–projecting predictor models base on the Kalman filter," Journal of the Royal Statistical Society, Series B 42 pg. 35–45.

Goldberg,S.(1958) "Difference Equations," John Wiley, New York.

Golub,G.H. and Van Load,C.F.(1983) "Matrix Computations," John Hopkins University Press, Baltimore.

Granger,C.W.J. and Newbold,P.(1974) "Spurious regressions in economics," Journal of Econometrics, 2 pg. 111–120.

Grenander,U. and Szego,G.(1958) "Toeplitz Forms and their Applications," University of California Press, Berkeley, California.

Grimmett,G.R. and Stirzaker(1982) "Probability and Random Processes," Oxford University Press, Oxford.

Gurland,J.(1948) "Inversion Formulae for the Distribution of Ratios," Annals of Mathematical Statistics 19 pg. 228–37

Hannan,E.J.(1970) "Multiple Time Series," Wiley, New York.

Harvey,A.C.(1981) "Time Series Models," Philip Allen Publishers Limited, Oxford.

Harvey,A.C.(1984) "Dynamic Models, the Prediction Error Decomposition and State Space," in "Econometrics and Quantitative Economics," ed. by Hendry,D.F. and Wallis,K.F.

Harvey,A.C.(1986) "Analysis and generalisation of a multivariate exponential smoothing model," Management Science 32 pg 374–380.

Harvey,A.C.(1989) "Forecasting, Structural Time Series Models and the Kalman Filter," Cambridge University Press (to appear).

Harvey,A.C. and Peters,S.(1984) "Estimation procedures for structural time series models," ESRC/LSE Econometrics Project Discussion Papers A44 (forthcoming in Journal of Forecasting).

Harvey,A.C.(1977) and Phillips,G.D.A. "A comparison of estimators in the ARMA(1,1) model," Unpublished paper, University of Kent.

Harvey,A.C. and Durbin,J.(1986) "The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modeling," Journal of the Royal Statistical Society, Series A. 149 pg. 187–227.

Harvey,A.C.,Henry,S.G.B.,Peters,S. and Wren–Lewis,S.(1986) "Stochastic trends in dynamic regression models: an application to the employment–output equation," Economic Journal 96 pg. 975–985.

Hearn,A.C.(1985) "Reduce User's Manual," The Rand Corporation, Santa Monica.

Hendrick,C.L.(1973) "Expectations and the Labour Supply," American Economic Review 63 pg. 968–974.

Hewitt,E. and Stromberg,K.R.(1965) "Real and Abstract Analysis," Springer–Verlag, New York.

Holt,C.C.(1957) "Forecasting seasonal and trends by exponentially weighted moving averages," Carnegie Institute of Technology, Pittsburgh, Pennsylvania, ONR Research Memorandum No. 52.

Imhof,J.P.(1961) "Computing the distribution of quadratic forms in normal variables," Biometrika 48 pg. 419–426.

Kac,M.,Kiefer,J. and Wolfowitz,J.(1955) "On tests of normality and other tests of goodness of fit based on distance measures," Annals of Mathematical Statistics 26 pg. 189–211.

Kang,K.M.(1975) "A Comparison of estimators for Moving average Processes," unpublished, from the Australian Bureau of the Census.

Kendall,M.G., Stuart,A. and Ord,J.K.(1987) "Advanced theory of statistics, vol 1," 5th ed. Griffin, London.

King,M.L. (1981) "The Durbin–Watson Test for Serial Correlation: Bounds for Regressions with Trend and/or seasonal Dummy Variables," Econometrica 49 pg. 1571–1581.

Kitagawa,G.(1981) "A nonstationary time series model and its fitting by a recursive filter," Journal of Time Series Analysis 2 pg. 103–116.

Kitagawa,G. and Gersch,W.(1984) "A smoothness priors – state space modelling of time series with trend and seasonality," Journal of the American Statistical Association 79 pg. 378–389.

Knott,M.(1974) "The Distribution of the Cramér–von Mises Statistic for Small Sample Sizes," Journal of the Royal Statistical Society, Series B 36 pg. 430–438.

Koerts,J. and Abrahamse,A.P.J.(1968) "On the theory and Application of the general linear model" Rotterdam University Press.

Kohn,R. and Ansley,C.F.(1986) "Estimation, prediction and interpolation for ARIMA models with missing data," Journal of the American Statistical Association 81 pg. 751–61.

Leybourne,S.J. and McCabe,B.P. (1989) "On the distribution of some test statistics for coefficient constancy," Biometrika 76 pg. 169–178.

Lévy,P.(1925) "Calcul des probabilités," Gauthier–Villars, Paris.

Lukacs,E.(1970) "Characteristic Functions," Griffin, London.

Magnus,J.R. and Neudecker,H.(1988) "Matrix Differential Calculus" Wiley, Chichester.

Marshall,P.(1989) "An Alternative expression for the likelihood of a structural time series model," unpublished paper, Department of Statistics, LSE.

Muth,J.F.(1960) "Optimal Properties of Exponentially Weighted Forecasts", Journal of the American Statistical Association 55 pg. 299–305.

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 16 pg. 218–235.

Nelson,C.R. and Plosser,C.I.(1982) "Trends and Random Walks in Macroeconomic Time Series, Some Evidence and Implications," Journal of Monetary Economics 10 pg. 139–162.

Nelson,C.R. and Kang,H.(1984) "Pitfalls in the use of Time as an Explanatory Variable in Regression," Journal of Business and Economic Statistics 2 pg. 73–82.

Newbold,P.(1988) "Some Recent Developments in Time Series Analysis–III," International Statistical Review 56 pg. 17–29.

Nyblom,J.(1986) "Testing for Deterministic Linear Trend in Time Series," Journal of the American Statistical Association 81 pg. 545–549.

Palm,F.C. and Sneek,J.M.(1984) "Significance tests and spurious correlation in regression models with autocorrelated errors," Statistische Hefte 25 pg. 87–105.

Pan, Jie–Jian(1964) "Distribution of the noncircular serial correlation coefficients," (in Chinese) Shuxue Jinzhan 7 pg. 328–337. Translated by N.N.Chan for Selected Translations in Mathematical Statistics and Probability 7 1968, pg. 281–292.

Peers,H.W.(1978) "Second–order sufficiency and statistical invariance," Biometrika 65 pg. 489–96.

Peers,W.H. and Iqbal,M.(1985) "Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications," Journal of the Royal Statistical Society,Series B. 47 pg. 547–554.

Perron,P.(1989) "The Calculation of the Limiting Distribution of the Least Squares Estimator in a Near–Integrated Model," Econometric Theory 5 pg. 241–55.

Pesaran,M.H.(1983) "A note on the maximum likelihood estimation of regression models with first order moving average errors with roots on the unit circle," Australian Journal of Statistics 25 pg. 442–448.

Phillips, P.C.B.(1987) "Towards a unified asymptotic theory foe autoregression," Biometrika 74 pg. 535–47.

Press,W.H., Flannery,B.P., Teukolsky,S.A. and Vetterling,W.T.(1986) "Numerical Recipes: The Art of Scientific Computing," Cambridge University Press, Cambridge.

Raj,B. and Ullah,A. (1981) "Econometrics, A Varying Coefficients Approach," Croom Helm, London.

Rayna,G.(1987) "Reduced: Software for Algebraic Computation," Springer–Verlag, New York.

Rosenburg,B.(1973) "A survey of Stochastic Parameter Regression," Annals of Economic and Social Measurement 2 pg. 381–397.

Rutherford,D.E.(1946) "Some continuant determinants arising in Physics and Chemistry," Proceedings of the Royal Society of Edinburgh section A62 pg. 229–36.

Sargan,J.D.(1977) "Estimating the first order moving average when the true coefficient is one," working paper, London School of Economics.

Sargan,J.D. and Bhargava,A.(1983a) "Testing residuals from least squares regression for being generated by the Gaussian random walk," Econometrica 51 pg. 153–174.

Sargan,J.D. and Bhargava,A.(1983b) "Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle," Econometrica 51 pg. 799–820.

Sheil,J. and I.O'Muircheartaigh(1977) "AS106: The distribution of non–negative quadratic forms in normal variables," Applied Statistics 26 pg. 92–98.

Shenton,L.R. and Bowman,K.O.(1977) "Maximum likelihood estimmation in small samples," Griffin Statistical Monographs No. 38, London.

Shively,T.S.(1986) "Issues in the analysis of stochastic coefficient regression models," Ph.D. Dissertation, The University of Chicago.

Shively,T.S.(1988) "Numerical Inversion of a Bivariate Characteristic Function," Working paper, University of Texas at Austin.

Shively,T.S.(1988a) "An exact test for a stochastic coefficient in a time series regression model," Journal of Time Series Analysis 9 pg. 81–88.

Shively,T.S.(1988b) "An analysis of tests for regression coefficient stability," Journal of Econometrics 39 pg. 367–386.

Shively,T.S.,Ansley,C.F. and Kohn,R.(1989) "Fast evaluation of the distribution of the Durbin–Watson and other invariant test statistics in regression," unpublished paper.

Tanaka,K and Satchell,S.E.(1987) "Asymptotic Properties of the Maximum Likelihood and Non–Linear Least Squares Estimator for noninvertible moving average models" presented to the Econometric Study Group, Autumn Term, London School of Economics.

Taniguchi,M.(1984) "Validity of Edgeworth expansions for Statistics of Time Series," Journal of Time Series Analysis, 5 pg. 37–51.

Trzcinka,C.(1982) "The Pricing of Tax–Exempt Bonds and the Miller Hypothesis," Journal of Finance 37 pg. 907–923

Watson,M.W. and Engle,R.F.(1985) "Testing for regression coefficient stability with a stationary AR(1) alternative," Review of Economics and Statistics 67 pg. 341–46.

Watson,M.W.,(1986) "Univariate Detrending Methods with Stochastic Trends," Journal of Monetary Economics 18 pg. 49–75.

Wecker,W.E. and Ansley,C.F.(1983) "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing," Journal of the American Statistical Association 78 pg 81–89.

White,K.J.(1978) "A general computer program for econometric models– SHAZAM" Econometrica 46 pg. 239–240.

Wolfram,S.(1988) "Mathematica: A System for doing Mathematics by Computer," Addison–Wesley, Redwood City, California.

Young,P.C.(1984) "Recursive Estimation and Time–Series Analysis, An Introduction" Springer–Verlag, Berlin.

Zygmund,A.(1955) "Trigonometrical Series," Dover Publications,New York.