THE UNIVERSITY OF MANITOBA

INFERENCE FOR TIME SERIES MODELS WITH STABLE ERRORS

BY

DANKIT KATASI NASSIUMA

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

WINNIPEG, MANITOBA

MAY, 1992

ISBN 0-315-77825-3

INFERENCE FOR TIME SERIES MODELS WITH STABLE ERRORS

BY

DANKIT KATASI NASSIUMA

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

CONTENTS

ABSTRACT

A recurring theme in this dissertation is the development of practical techniques and statistical inference theory for infinite variance stable random variables. In particular, we aim to study the estimation of parameters, the estimation of missing values as well as properties of nonstationary processes.

In Chapter one, we give an overview of the basic concepts of stable variables as well as the literature related to processes having infinite variance errors. Areas of interest include model identification, parameter estimation, estimation of missing observations, nonstationary processes as well as forecasting.

Chapter two introduces a set of estimators based on estimating functions that are optimal in the sense of minimum scale parameter of the errors. The optimality criteria is established and the strong consistency of the estimators is shown. Limit properties of the estimators are also given. The recursive form of the estimators are derived and similarly shown to be consistent.

Chapter three is concerned with the estimation of missing observations. Estimates based on the optimality criteria developed in Chapter two are derived. Linear estimates as well as estimates of missing observations based on the minimization of the λ^{th} norm ($\lambda \in (0,2)$) are similarly obtained.

In Chapter four we are concerned with nonstationary processes having infinite variance errors. Properties of the estimator based on the optimality criteria developed in Chapter two are studied for the AR(1) model with unit root. Properties of linear processes having time varying scale parameters and those with time varying coefficients are discussed. Of particular interest is the development of conditions under which such nonstationary processes exist.

ACKNOWLEDGMENTS

I first of all thank my Lord Jesus Christ who has consistently made provisions for my deepest needs during my studies and through whom I have come to recognize a form of significance in life which surpasses any statistical significance.

I am very thankful to my advisor, Dr. Thavaneswaran who introduced me to the field of probability theory and who gave me the space and encouragement to venture into infiniteness. My thanks also go to the rest of my supervisory committee members, Professor R. Dhruvarajan, Professor M. Samanta and Professor K. Mount for their valuable suggestions.

I am grateful for the funding availed through the Commonwealth Scholarship and Fellowship Plan which facilitated my undertaking of this program. My thanks also go to the Department of Statistics and my advisor, Dr. Thavaneswaran for the provision of funds for me to attend a workshop at the University of Minnesota in the summer of 1990 which motivated me to work on this topic. I am very grateful to the Head of the Department of Statistics, Professor L.K. Chan for letting me know about the workshop.

My gratitude also go to my parents, Paul and Rebah Nassiuma whose encouragement and trust I cherished throughout my early studies. To my brothers and sisters who believed in me and with whom we have shared so much.

Most of all I thank my wife, Elizabeth who inspite of her own graduate studies, encouraged and supported me throughout my studies and provided the emotional strength that all of us in the family needed. And to my children, Luseno, Manakhe and Rebbea whose cheerfulness provided opportunities for laughter even when weighed down with work and discouraged, and for the sacrifices they made when I missed to be with them because of my studies.

CHAPTER ONE

1. INTRODUCTION

1.1. Overview

A set of observations obtained sequentially in time is known as a time series. When such observations can be represented as a linear function of a sequence of mutually independent and identically distributed (i.i.d) random variables, it is known as a linear process. A purely nondeterministic Gaussian process can be generated by application of a linear filter to an independent and identically distributed (i.i.d) Gaussian sequence of random variables. Commonly applied time series models are based on the assumption that the underlying process is Gaussian. A reason for this is that the large amount of theory that has been developed for the Gaussian processes makes it easy to examine various inferential properties pertaining to parameter estimates and the adequacy of the model of interest. The simplicity implied by these developements point to the strong possibility that the gaussian assumptions could simply be out of convenience rather than the true form of the process. Inferential properties for processes that are not Gaussian on the other hand have not been as extensively studied and this limits their applications.

Processes encountered in many practical situations in time series analysis depict properties that deviate from those of the Gaussian process and thus the resulting inferences based on the Gaussian assumption may not always be justified. This has inevitably led to a search for alternative models to the Gaussian models. One possible direction has been to introduce nonlinearity while retaining the Gaussian assumption on the innovation sequence. This has resulted in various nonlinear models being developed which include the Bilinear models by Granger and Andersen (1978), State dependent models by Priestley (1980), Exponential models by Haggan and Ozaki (1981), Random coefficient autoregressive models by Nicolls and Quinn (1982), Threshold models by Tong (1983), and Doubly stochastic models by Tjosteim (1986). Properties of these models have been extensively studied in the literature.

The second alternative has aimed at retaining the linearity property but ascribing non-Gaussianess to the innovation sequence. This approach takes two distinct forms: the finite variance non-Gaussian set of models and the infinite variance case. Research pertaining to finite variance non-Gaussian processes has been vigorously pursued by for example Gaver and Lewis (1980), Jacobs and Lewis (1977) and Lawrence and Lewis (1980) who discuss the exponential and gamma autoregressive models. Some inferences for processes with non-Gaussian stable innovations having infinite variance have similarly been discussed by for example Stuck (1978) and Cline and Brockwell (1985). Models that consist of a blend of non-Gaussianess and nonlinearity have been discussed by Liu (1989) where he considers the existence of a bilinear model with infinite variance innovations. The major interest in this dissertation centres on linear and nonlinear processes having infinite variance stable errors.

Observations of data which show characteristics that do not conform to the Gaussian assumptions have been widely identified but interest has been focused on series that follow the infinite variance stable distribution. The reason for this inclination is twofold: First, the infinite variance stable distributions possess closure properties under convolution which is a vital property for inferential purposes. Secondly, these distributions can be obtained as limiting distributions of sums of i.i.d random variables which is a property that is also possesed by the Gaussian random variables and facillitates the study of limiting properties.

A simple characterization of the stable random variable η is based on its characteristic function. Thus η has a stable distribution if it has a log characteristic function of the form

$$\log(\Phi(w)) = iw\beta - \delta|w|^{\alpha}\left(1 - i\theta\frac{w}{|w|}\tan(\pi\alpha/2)\right) \quad \text{if } \alpha \neq 1$$

$$= iw\beta - \delta|w|\left(1 - 2i\frac{\theta}{\pi}\frac{w}{|w|}\ln(|w|)\right) \quad \text{if } \alpha = 1$$

where $\alpha \in (0,2]$, $\beta \in \Re$, $\delta \in [0,\infty)$, and $\theta \in [-1,1]$ are the exponent, location, scale and symmetry parameters (See Brockwell and Davis, 1987).

To illustrate the basic distinctions which exist between Gaussian processes and infinite variance stable processes, simple plots of simulated data are given in the following figures: Fig. 1.1, Fig 1.2 and Fig. 1.4 depict plots of simulated data from the AR(1), MA(1) and ARMA(1,1) process ξ_t respectively when the process is driven by symmetric stable innovations. The characteristic function of the process innovations η_t is given by $\Phi(w) = \exp(-\delta|w|^{\alpha})$ where α is the characteristic exponent and δ is the scale parameter. In Fig. 1.3, a plot of data from an AR(1) Gaussian processes is given. For simplicity, a common scale parameter $\delta = 1$ was used in the generation of the data depicted in each of the figures but the characteristic exponent was different for each model. The Fortran subroutine GGSTA was used to generate the observations.

Fig. 1.1: AR(1) model $\xi_t = 0.4\xi_{t-1} + \eta_t$, $\alpha = 1.2$

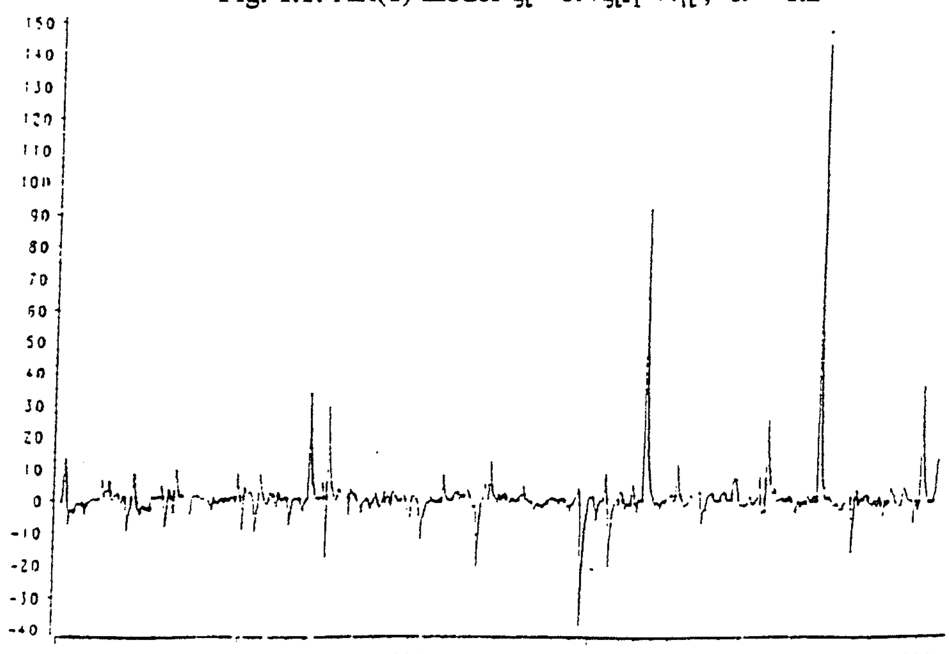


Fig. 1.2: $\xi_t = 0.6\eta_{t-1} + \eta_t$, $\alpha = 0.5$

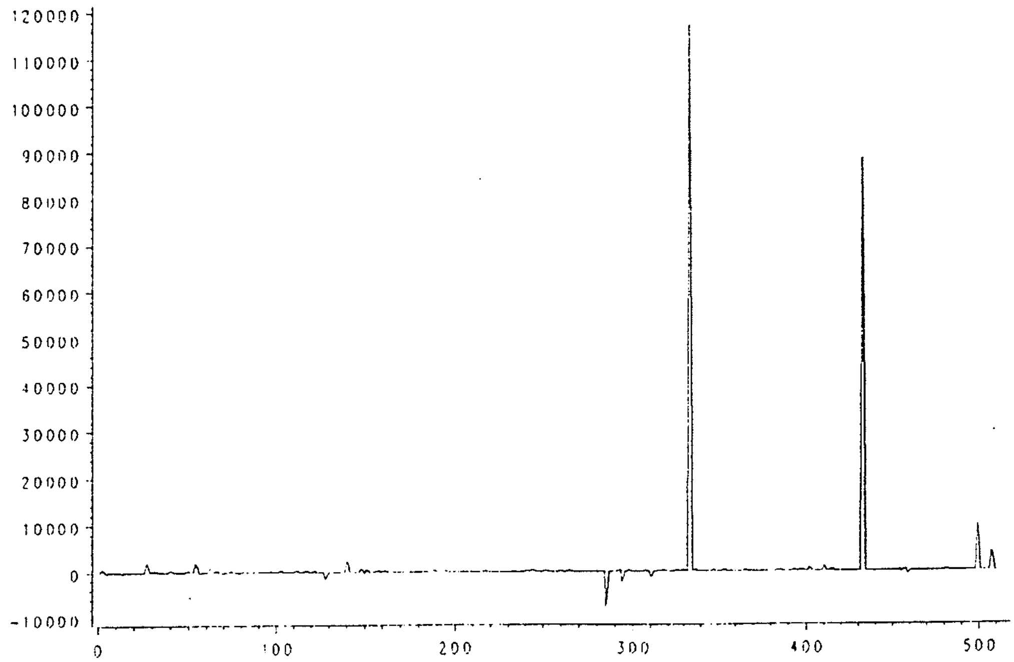


Fig. 1.3: $\xi_t = 0.4\xi_{t-1} + \eta_t$, $\alpha = 2$

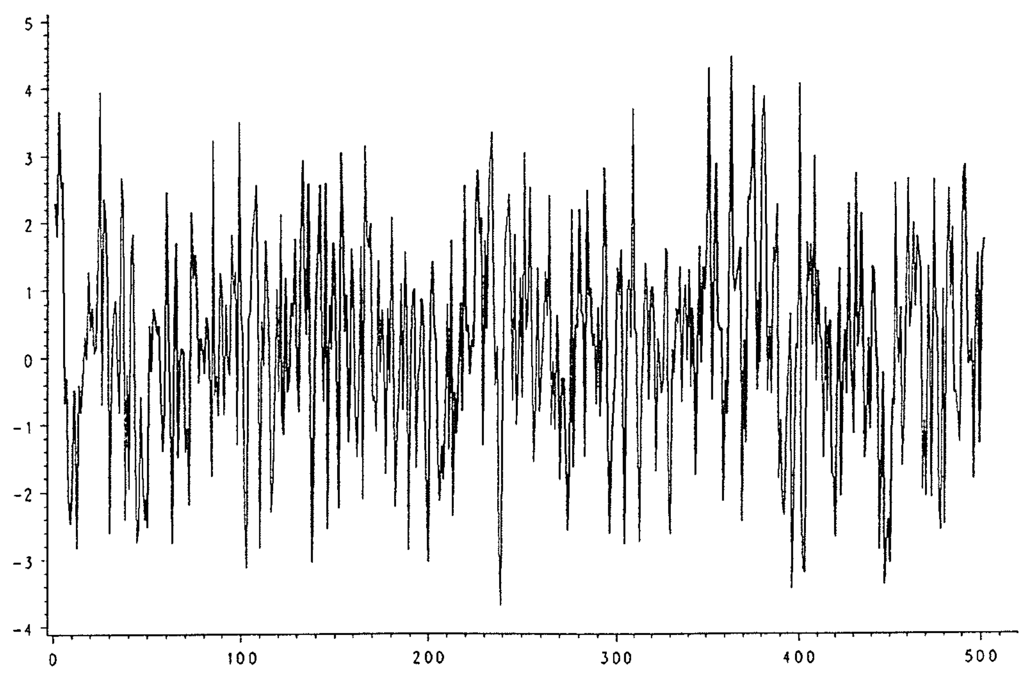
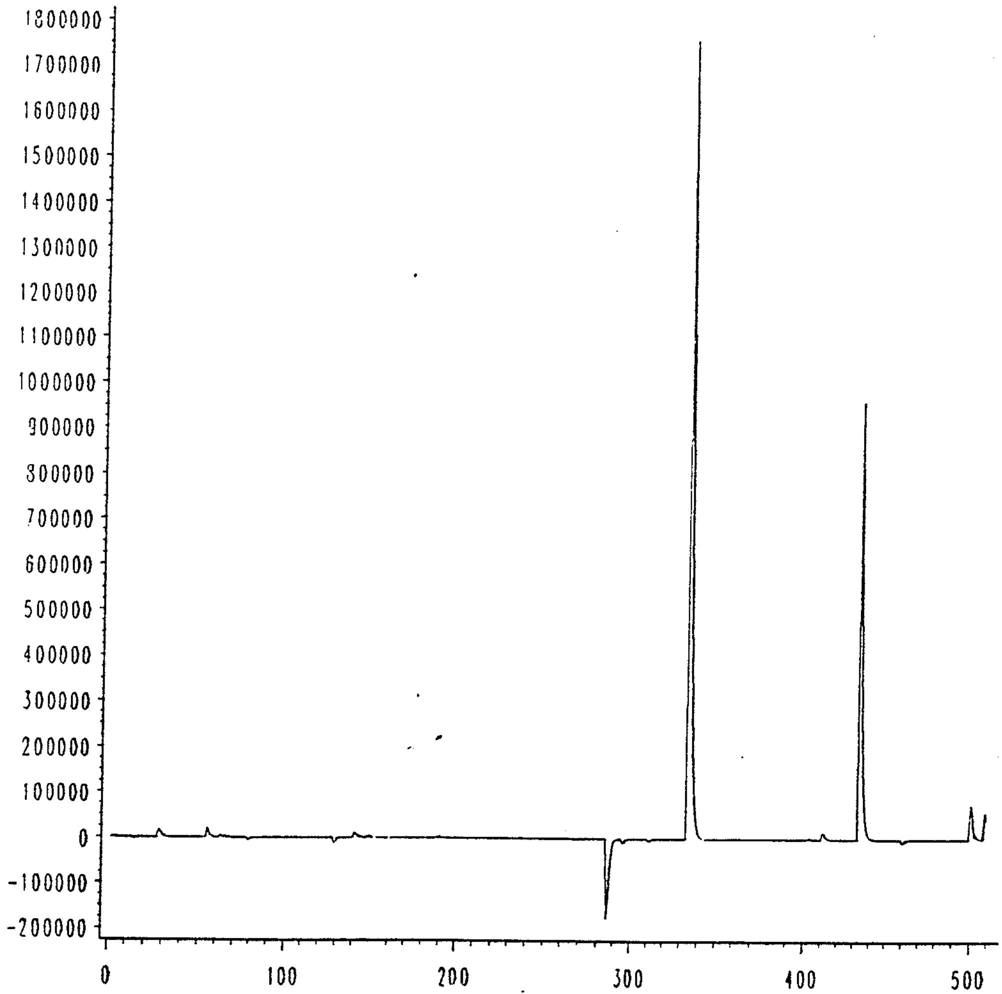


Fig. 1.4: $\xi_t = 0.4\xi_{t-1} + 0.3\eta_{t-1} + \eta_t$, $\alpha = 0.4$



The sharp bursts of observations in the infinite variance stable processes distinguish them from the Gaussian process. Such outlying observations in infinite variance stable processes lead to high values of the tail probabilities.

Some earlier work based on utilizing infinite variance stable distributions was in Economics by Mandelbrot (1963) when considering the models for stock market price changes. Fama (1965) when similarly modeling stock market prices attributed their discrepancies to the possibility of the process having stable innovations and thus fitted an adequate model on this basis. Granger and Orr (1972) made some observations on available alternatives when the Gaussian assumptions are not appropriate. Stuck and Kleiner (1974) on examining telephone signals concluded that models with stable innovations would more adequately represent the process. Stable processes have similarly been found appropriate in various practical situations in Geophysics, Economics, Engineering, Electronics and Biology (Zolotarev, 1986).

Inferential problems for infinite variance processes have also been discussed by Cline and Brockwell (1985) who established some conditions for the existence of a solution for a stationary linear processes with stable innovations and this made it possible to develop linear predictors for such processes. Properties of least squares estimates for infinite variance processes were also discussed by Kanter and Steiger (1974) and Yohai and Maronna (1977). Recently, the problem of identification for such processes has been studied by for example Bhansali (1988) and Knight (1989). We aim to extend this study of inference for the infinite variance processes to a wider range of problems with the view to developing practical techniques and relevant theory.

The increasing evidence in regard to infinite variance stable processes thus provides an impetus to further examine inferential problems for such processes in this dissertation.

1.2. The L^λ space, regular variation and the stable law

The set of random variables in this study are those defined in the L^λ ($\lambda \in (0,2)$) space and we thus start by recalling some properties which pertain to them.

Suppose a set of iid random variables $\{\eta_i , i = 1, 2,...\}$ is defined on the probability space (Ω,S,P) such that on this space, the λ^{th} moments for these random variables exist. Such a space which we denote as L^λ ($\lambda > 0$) is a metric space. In showing that the L^λ space is a metric space, we distinguish the two cases that arise, i.e., $\lambda < 1$ and $\lambda \geq 1$ and this is facilitated by first of all considering the following inequality:

Minkowski's inequality

Suppose the two random variables ξ and η are defined on the L^λ space such that their λ^{th} moments exist. Then the following inequality holds:

$$[E(|\eta + \xi|^\lambda)]^{1/\lambda} \leq [E(|\xi|^\lambda)]^{1/\lambda} + [E(|\eta|^\lambda)]^{1/\lambda} \quad \text{for } \lambda \geq 1$$

Proof:

This inequality is trivial when $\lambda = 1$ while for the case when $\lambda > 1$, we note that

$$E(|\eta + \xi|^\lambda) \leq E(|\xi||\eta + \xi|^{\lambda-1}) + E(|\eta||\eta + \xi|^{\lambda-1})$$

and the result follows by applying Holder's inequality.

When $\lambda = 2$, this leads to the Schwarz inequality. The case where $\lambda \geq 1$ constitutes the space of random variables which is more commonly known as the banach space. Similarly for the two random variables defined on the space L^λ, their norm exists and is defined for example for ξ as

$$||\xi|| = [E(|\xi|^{\lambda})]^{1/\lambda}.$$

This norm satisfies the following conditions.

(i) $||\xi|| \geq 0$ and $||\xi|| = 0$ iff $\xi \to 0$ a.s.

(ii) $||\beta\xi|| = |\beta|\ ||\xi||$ where $\xi, \beta \in \Re$.

(iii) $||\eta + \xi|| \leq ||\xi|| + ||\eta||$.

A metric can be introduced in every normed linear space by simply defining the distance between ξ and η as $d(\xi,\eta) = ||\xi - \eta||^{\lambda}$. Requirements (i) and (ii) need no proof while requirement (iii) follows from Minkowski's inequality. It is thus easy to see that L^{λ} is a metric space when $\lambda \geq 1$.

In the case when $\lambda < 1$, the distance measure $d(\xi,\eta)$ defined as $d(\xi,\eta) = E(|\xi - \eta|^{\lambda})$ similarly exists and defines a metric. It is easy to see that $d(\xi,\eta)$ satisfies the identity and symmetry axioms for a metric space, i.e., $d(\xi,\eta) \geq 0$ and $d(\xi,\eta) = d(\eta,\xi)$ respectively. The triangular inequality is similarly satisfied based on the following Theorem.

Theorem 1.2.1

Suppose $b, c, d \in \Re$ are any numbers. Then for $q \geq 0$ such that $b_q = 1$ if $q \leq 1$ and $b_q = 2^{q-1}$ if $q > 1$,

$$|c + d|^q \leq b_q\,(|c|^q + |d|^q)$$

A proof of this is given in Shiryayev (1984, pp. 192).

This theorem basically ensures that the following inequality holds:

$$E(|\eta + \xi|^{\lambda}) \leq E(|\xi|^{\lambda}) + E(|\eta|^{\lambda})$$

It thus becomes clear that L^{λ} is a metric space when $\lambda < 1$. Properties of metric spaces for which $\lambda \in (0,2)$ are very vital when considering stable random variables since their moments in the above range of λ exist.

The results discussed in this Section will be assumed when studying the concepts of projections in the L^{λ} space both for the stationary and nonstationary processes.

Regular Variation and domains of attraction

When dealing with monotone functions ϑ which are obtained from a probability distribution F by evaluating the integral of $\eta^{\lambda}F(d\eta)$ over $(0,\xi)$ or (ξ,∞), the change of variables in the integration lead to a family of functions of the form $a_t\vartheta(t\xi)$. Interest is then usually centered on investigating the asymptotic behavior of $a_t\vartheta(t\xi)$ as t tends to infinity. Suppose that the limit $G(t\xi)$ of $a_t\vartheta(t\xi)$ exists as $\xi \to \infty$. It then suffices to consider the norming factors of the above function, i.e., $G(1)/\vartheta(t)$ for positive and real $G(1)$. Based on this, a regularly varying function can be defined as follows:

Definition 1.2.1

Suppose ϑ be a positive monotone function defined on $(0, \infty)$ such that the following limit

$$\lim_{t\to\infty} \frac{\vartheta(t\xi)}{\vartheta(t)} \to G(\xi) \leq \infty$$

is bounded on a given set of points. Let $G(\xi) = \xi^{\alpha}$ where $\alpha \in (-\infty,\infty)$. The function ϑ is then said to be regularly varying at infinity and α is known as its characteristic exponent.

Definition 1.2.2

Let the function ϑ be as in definition 1.2.1 and suppose that it further satisfies the relation $\vartheta(\xi) = \xi^{\alpha}\Lambda(\xi)$. The function $\Lambda(\xi)$ is then said to be slowly varying at infinity iff for all positive ξ, the following limit exists:

$$\lim_{t\to\infty} \frac{\Lambda(t\xi)}{\Lambda(t)} \to 1.$$

This definition then leads to an alternative characterization of a regularly varying function which we state in the next definition.

Definition 1.2.3

A function $\vartheta(\xi)$ defined on $[0,\infty)$ is said to vary regularly at infinity with characteristic exponent α if for all $\xi > 0$ and $\Lambda(\xi)$ slowly varying at infinity, $\vartheta(\xi) = \xi^{\alpha}\Lambda(\xi)$.

In the following Theorem, a necessary and sufficient condition for a random variable to be in the domain of attraction of the stable law is given.

Definition:

A sequence of random variables $\{\xi_i, i = 1, 2, \ldots\}$ having distribution function F is said to be in the domain of attraction of the stable law with exponent α is that there exists a slowly varying function $\Lambda(\xi)$ such that as $\xi \to \infty$,

$$1 - F(\xi) \approx p\xi^{-\alpha}\Lambda(\xi)$$

where $$p = \lim_{\xi\to\infty} \frac{[1 - F(\xi)]}{1 - F(\xi) + F(-\xi)},$$

i.e., $p \in [0,1]$ (See Feller, 1971).

By considering Λ as a distribution function, the fact that a random variable ξ with distribution F has regularly varying tails implies that the tail behavior of ξ is regularly varying. The aspect of stable distributions being regularly varying has been used by Feller (1971) in the study of various characteristics pertaining to stochastic processes with stable innovations while De Haan (1970) utilized it to study extreme value problems. The applications developed by Feller (1971) are of particular interest in time series analysis and some of these are reviewed in the following Section.

Stable Random Variables

Characterization of stable distributions is usually based on the sum of the random variables and on the characteristic function and this is given in the following definition:

Definition 1.2.4

A random variable η is said to have a stable distribution if for every positive integer n, there exist constants $a_n > 0$ and b_n such that $S_n = \eta_1 + \eta_2 + ... + \eta_n$ has the same distribution as $a_n\eta + b_n$ for all independent and identically distributed random variables $(\eta_1, \eta_2, ..., \eta_n)$ generated from the same distribution as η. The random variable η has a log characteristic function given as

$$\log(\Phi(w)) = iw\beta - \delta|w|^{\alpha}(1 - i\theta\frac{w}{|w|}\tan(\pi\alpha/2)) \qquad \text{if } \alpha \neq 1$$

$$= iw\beta - \delta|w|(1 - 2i\frac{\theta}{\pi}\frac{w}{|w|}\ln(|w|)) \qquad \text{if } \alpha = 1 \tag{1.2.1}$$

where $\alpha \in (0,2]$, $\beta \in \Re$, $\delta \in [0,\infty)$, and $\theta \in [-1,1]$ are the exponent, location, scale and symmetry parameters respectively (See Feller, 1971). If $\theta = 0$ then η is symmetric stable about β and if $\beta = 0$ then the random variable is symmetric stable about zero with log characteristic function

$$\log(\Phi(w)) = -\delta|w|^{\alpha}. \qquad (1.2.2)$$

Similarly if $\theta = 0$ and $\alpha = 1$ then η has the Cauchy distribution. The case when $\alpha = 2$ leads to the Gaussian distribution.

For the stable distributions, the second moment does not exist if $\alpha \in (1,2)$ while for the case when $\alpha \in (0,1]$, neither the first nor the second moment exists. In both cases the variance can thus not be defined. An alternative measure of the variance as a measure of the spread for stable random variables is the dispersion as defined in Stuck (1978). This alternative measure is equivalent to the tail probability of a regularly varying random variable (Cline and Brockwell, 1985). A basic definition of the dispersion is given as follows.

Definition 1.2.5

Let ξ be given by the summation

$$\xi = \sum_{i=0}^{\infty} \psi_i \eta_i$$

where $\{\eta_i, i = 1, 2,...\}$ is a sequence of independently distributed stable random variables with scale parameter $\delta = 1$ and characteristic exponent α, $0< \alpha \leq 2$ such that

$$\sum_{i=0}^{\infty} |\psi_i|^{\alpha} < \infty .$$

The dispersion, which we denote as disp(ξ)), is

$$\mathrm{disp}(\xi_t) = \sum_{i=0}^{\infty} |\psi_i|^{\alpha} .$$

It is easy to see that this measure reduces to the variance measure when the characteristic exponent is 2.

Parameter estimation for stable distributions

In the literature, various techniques are usually applied in the estimation of parameters of stable distributions and one that has been of greatest interest was proposed by Press (1972a) and is based on the empirical characteristic function. This is usually referred to as the method of moments. The implementation of this method is given in this Section.

Consider the case where the observations $\{\xi_j, j = 1, 2, \ldots, n\}$ are available for the variable ξ. The empirical characteristic function is then obtained as

$$\hat{\Phi}(w) = \frac{1}{n}\sum_{j=1}^{n} e^{iw\xi_j}, \quad w \in (-\infty,\infty).$$

Remark: For each w, $|(\hat{\Phi}(w)|$ is bounded above by 1 and thus $\hat{\Phi}(w)$ has finite moments. Now since $\hat{\Phi}(w)$ represents a stochastic process for $w \in (-\infty,\infty)$, it implies that for each specified w, $\hat{\Phi}(w)$ constitutes the sample mean of iid random variables. Similarly by the strong law of large numbers, $\hat{\Phi}(w)$ is a consistent estimator of $\Phi(w)$.

Paulson et al. (1975) carried out simulation studies to show the effectiveness of this procedure for the estimation of the parameters. Analytical point estimates of the parameters are obtained by noting from equation (1.2.2) that,

$$|\Phi(w)| = e^{-\delta|w|^{\alpha}}$$

and thus

$$\delta|w|^{\alpha} = -\ln|\Phi(w)|.$$

Now for any w_1 and w_2 such that $w_1 \neq w_2$, we have

$$\delta|w_k|^\alpha = -\ln|\Phi(w_k)| \quad, k = 1, 2 \ ; \alpha \neq 1 \ \text{ and } w_1, w_2 \neq 0 \tag{1.2.3}$$

Replacing $\Phi(w_k)$ with its estimate $\hat{\Phi}(w_k)$ and solving the two equations in (1.2.3) simultaneously for α and δ, we obtain

$$\hat{\alpha} = \ln\left|\frac{\ln|\hat{\Phi}(w_1)|}{\ln|\hat{\Phi}(w_2)|}\right|[\ln|w_1/w_2|]^{-1}$$

$$\hat{\delta} = -|w_2|^{-\hat{\alpha}}\,[\ln|\hat{\Phi}(w_2)|]$$

Estimation of θ and β is attained by letting

$$H(w) = I_M(\ln|\Phi(w)|$$

where $I_M(\vartheta(\xi))$ represents the imaginary part of the function $\vartheta(\xi)$. From (1.2.1), this implies that

$$H(w) = w\beta - \delta|w|^\alpha\theta\tan(\pi\alpha/2) \qquad \text{if } \alpha \neq 1$$

$$= w\beta - \frac{2\delta w\theta}{\pi}\ln(|w|)) \qquad \text{if } \alpha = 1$$

Choosing w_3 and w_4 such that $w_3 \neq w_4$, it implies that

$$\frac{H(w_k)}{w_k} = \beta - \delta|w_k|^{\alpha-1}\theta\tan(\pi\alpha/2) \qquad \text{if } \alpha \neq 1 \ \text{ and } \ k = 3, 4$$

$$= \beta - \frac{2\delta\theta}{\pi}\ln(|w_k|)) \qquad \text{if } \alpha = 1 \ \text{ and } k = 3, 4) \tag{1.2.4}$$

Solving the two equations from (1.2.4) simultaneously with α, δ and $H(w_k)$ replaced by their estimates $\hat{\alpha}$, $\hat{\delta}$ and $\hat{H}(w_k)$ respectively leads to the following results:

$$\hat{\beta} = \frac{\hat{H}(w_3)|w_4|^{\hat{\alpha}} - \hat{H}(w_4)|w_3|^{\hat{\alpha}}}{w_3|w_4|^{\hat{\alpha}-1} - w_4|w_3|^{\hat{\alpha}}} \qquad \text{if } \alpha \neq 1$$

$$= \frac{w_3\hat{H}(w_3)\ln|w_3| - w_4\hat{H}(w_4)\ln|w_4|}{w_3w_4(\ln|w_3| - \ln|w_4|)} \qquad \text{if } \alpha = 1,$$

$$\hat{\theta} = \frac{w_3\hat{H}(w_4) - w_4\hat{H}(w_3)}{\hat{\delta}[w_3|w_4|^{\hat{\alpha}} - w_4|w_3|^{\hat{\alpha}}]\tan(\pi\hat{\alpha}/2)} \qquad \text{if } \alpha \neq 1$$

$$= \frac{w_3\hat{H}(w_4) - w_4\hat{H}(w_3)}{2\hat{\delta}w_3w_4[\ln|w_3| - \ln|w_4|]/\pi} \qquad \text{if } \alpha = 1.$$

Estimates of the parameters for the multivariate stable distributions are obtained using a similar approach based on the sample characteristic function.

Press (1972b) derived association measures for jointly distributed random variables for multivariate stable distributions while Griffiths (1972) and De Silva (1978) studied some properties of multivariate stable distributions and in analogy with the finite variance case, they obtained the correlation coefficient for the bivariate symmetric stable random variables.

Basic limit properties of stable variables

The central limit theorem for finite variance variables states that if we have a sequence of iid random variables $\{\xi_1, \xi_2, ..., \xi_n\}$ such that $E(\xi_1) = \mu$ and $Var(\xi_1) = \sigma^2$, then

$$\frac{1}{\sigma/\sqrt{n}} \sum_{i=1}^{n} (\xi_i - \mu) \xrightarrow{d} Z,$$

where $\xrightarrow{d}$ represents convergence in distribution and Z is a standard Gaussian distribution.

A similar property exists for infinite variance stable random variables under certain conditions. In this case we suppose that the sequence of stable iid random variables $\{\xi_i, i = 1, 2, ..., n\}$ are in the domain of attraction of the stable law with characteristic exponent $\alpha \in (0,2)$ and distribution F. Suppose that there exist constants a_n and b_n defined as

$$a_n = \inf \{x: [1 - F(x) + F(-x)] \leq n^{-1}\}$$

and

$$b_n = F(a_n) - F(-a_n).$$

Then the following convergence limit exists:

$$\frac{1}{a_n} \sum_{i=1}^{n} (\xi_i - b_n) \xrightarrow{d} S_\alpha,$$

where S_α is a stable random variable with characteristic exponent $\alpha \in (0,2)$. The sequence of random variables $\{\xi_i\}$ is then said to be in the domain of attraction of the stable law. Such characterization for stable random variables facilitates the study of the limiting distributions of parameter estimates for processes with stable errors. These properties will be utilized in Chapter two when considering optimal parameter estimates for stable processes.

1.3. Model identification

Process identification involves various steps, the most basic being the determination of the class of processes to which a given time series belongs. In particular, one has to determine whether a given time series is generated by a linear process, a nonlinear process, an infinite variance process or some other non-Gaussian process or a combination of some of these processes. This initial identification stage is either based on prior information about the generating process or on subjective deductions from the timeplot of the series followed by some confirmatory objective tests. As an example, if on the basis of the timeplot a given series is suspected to have been generated by a nonlinear process, then the next step is to test for nonlinearity. Once the form of the process is established, the next important step is to determine the specific set in the selected class to which the series belongs. Thus again if the series is initially identified as having been possibly generated by a nonlinear model, it now has to be determined as to which set of nonlinear models this series belongs. This is then followed by the determination of the order of the model in the selected set.

Although there is no known specific paradigm for assessing distinctions between time series belonging to the different classes, a lot of research has been done for establishing the identity of a given series within a particular class especially the class of finite variance linear processes. Simple graphical tools for identification such as those developed by Box and Jenkins (1976) as well as the objective techniques of Akaike (1969, 1970) have played a key role in this. Identification within the class of nonlinear models is still in its infant stages.

The class of infinite variance stable processes has been identified using two major tools, the most basic one being the presence of bursts of observations in the series. Simulation results based on time series observations generated by an infinite variance process depict bursts of observations and moreso as the characteristic exponent of the generating stable random variable decreases. A more objective approach was used by Fama (1965) and Mandelbrot (1963) when modeling stock market prices. This is based on the evaluation of the tail probabilities for the given series of observations. Large tail probabilities in the observations makes the finite variance assumption inappropriate. This is due to the fact that inferential techniques for observations from finite variance processes are based on asymptotic theory which are based on the requirement that second order properties exist. Thus even for large samples, the choice of a finite sample model may not be a feasible one.

These tools do not necessarily provide complete evidence to conclude that the generating mechanism is based on stable random variables since for example the presence of many outliers in a finite variance series could lead to the conclusion that it is infinite in its variance. When dealing with outliers, it is usually assumed that the outlying observations are independent of the generating mechanism. This is not the case for stable variables since the bursts in the observations are part of the generating process. Thus the presence of many outliers is bound to lead to an erroneous conclusion regarding the distribution of the process. A more effective tool for establishing the distinction between data from the finite and infinite variance processes need to be developed. In this dissertation, we assume that we are dealing with data generated by an infinite variance process.

Once a given series has been identified as having been generated by an infinite variance process, an area that has been receiving increasing attention is that of order determination particularly in the class of linear models. Some of the early work in this direction was carried out by Rosenfeld (1976) who used simulation studies on the application of the Box and Jenkins (1976) identification techniques to a limited set of infinite variance stable processes and concluded that such techniques are still appropriate for identification of infinite variance stable processes. For a wider range of stable processes particularly when $\alpha < 1.5$, the Box and Jenkins approach might be inadequate when used as an identification tool and thus this is an area which requires more research.

In the finite variance case, suppose that ξ_t is a k^{th} order autoregressive (AR) process given by the difference equation:

$$\xi_t = \sum_{j=1}^{k} \phi_j \xi_{t-j} + \eta_t,$$

where $\phi(0) = 1$, $\{\eta_t, t=1,2,...\}$ is a sequence of i.i.d random variables with zero mean and variance σ^2 and the polynomial $\Phi(z) = 1+\psi_1 z + \psi_2 z^2 +....+ \psi_p z^p$ has roots outside the unit circle (stationarity condition). Suppose that the observations, $(\xi_1, \xi_2, ...,\xi_N)$ are available, Akaike (1969,1970) defined the Final Prediction Error (FPE) technique for determining the order k of the model such that the one step ahead prediction error is minimized as

$$FPE(k) = E[\xi(k) - \hat{\xi}(k)]^2 \quad = \sigma_k^2(1 + \frac{k}{N})$$

where $\hat{\xi}(k)$ is the predictor of ξ_{t+1} based on the model of order k. This is obtained as

$$\hat{\xi}(k) = \sum_{j=1}^{k} \hat{\phi}_j \xi_{t-j}$$

where the sequence of coefficients $\{\hat{\phi}_j\}$ are the least squares estimates of $\{\phi_j\}$. By replacing σ_k^2 with its unbiased estimate $N\hat{\sigma}_k^2/(N - k)$, we obtain

$$FPE(k) = \hat{\sigma}_k^2 \{\frac{N + k}{N - k}\} = \hat{\sigma}_k^2\left(1 + \frac{k}{N}\right)\left(1 - \frac{k}{N}\right)^{-1}$$

where

$$\hat{\sigma}_k^2 = \frac{1}{N}\sum_{t=1}^{N} \sum_{j=0}^{k} \hat{\phi}_j(k)\xi_{t-j}, \qquad \hat{\phi}_o(k) = 1 \text{ for all k.}$$

Bhansali and Downham (1977) obtained a generalized form of the FPE criterion as

$$\mathrm{FPE}_{\alpha}(k) = \hat{\sigma}_k^2\left(1 + \frac{\alpha k}{N}\right) \tag{1.3.1}$$

where $\alpha > 0$ is a fixed constant and derived the asymptotic distribution. Bhansali (1988) studied the Final Prediction Criteria (FPE_{α}) in the context of infinite variance stable errors and showed that this leads to a consistent model identification. In essence, the use of this FPE approach amounts to assuming that the variance is a consistent measure of the dispersion in the infinite variance case. It would thus be imperative to consider alternative dispersion measures which would be adequate particularly for moderate sample sizes with heavy tails.

Another order determining technique for infinite variance stable processes was studied by Knight (1990) where he showed that the Akaike Information Criterion (AIC) consistently determines the order of an autoregressive model.

1.4. Estimation of model coefficients

Model coefficient estimation in time series is a crucial step in attaining some of the major goals in time series analysis and this in particular is due to the fact that efficient estimation of the coefficients leads to effective forecasts. In the literature, several approaches to the estimation of model coefficients for infinite variance processes have been discussed. These include robust estimation for moving average processes (Onishchenko, 1990), the least Gamma deviation estimates (Liu, 1987), the least absolute deviation approach (Gross and Steiger, 1979) and the usual least squares technique whose asymptotic properties have been widely studied for autoregressive and regression models. We briefly review some of these techniques in this section.

Least squares estimates.

Suppose the observations $y_1, y_2, ..., y_n$ are available from a p^{th} order stationary autoregressive model given as

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + U_t \quad t \in (-\infty, \infty) \tag{1.4.1}$$

where the sequence $\{U_t\}$ is of independent and identically distributed random variables which have a symmetric stable distribution with characteristic exponent $\alpha \in (0,2)$. The least squares estimate for the vector of coefficients $\Phi = (\phi_1, \phi_2, ..., \phi_p)$ is obtained as

$$\hat{\Phi} = \left[\sum_{t=2}^{n} \mathbf{Y}_{t-1} \mathbf{Y}_{t-1}^{T} \right]^{-1} \sum_{t=2}^{n} y_t \mathbf{Y}_{t-1} \tag{1.4.2}$$

where $\mathbf{Y}_{t-1} = (y_{t-1}, y_{t-2}, ..., y_{t-p})^T$.

Under the assumption that the sequence of random variables $\{U_t\}$ are in the domain of attraction of the stable law, Kanter and Steiger (1974) showed that the least squares estimates of the model parameters as in equation (1.4.2) is consistent, i.e., converges in probability. Yohai and Maronna (1977) extended the distribution of $\{U_t\}$ to the class whereby $E[\log^+(U_t)] < \infty$ and also showed that the estimate given in (1.4.2) is consistent. Hannan and Kanter (1977) also showed the almost sure convergence of the least squares estimates using a slightly different approach from that of Yohai and Maronna (1977).

Least absolute deviation estimate (L1 estimate)

Let ξ_t be a p^{th} order autoregressive process given by the difference equation

$$\xi_t = \sum_{i=1}^{p} \phi_i \xi_{t-i} + \eta_t,$$

where $\phi_0 = 1$, $\{\eta_t, t=1,2,...\}$ is a sequence of i.i.d random variables and the coefficients satisfy the stationarity condition. Let the observations $(\xi_1, \xi_2, ..., \xi_n)$ are available. To estimate the model parameters, the following function is minimized with respect to ϕ_i:

$$f_N(\phi) = \frac{1}{N} \sum_{t=p+1}^{N} |\xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i}|. \tag{1.4.3}$$

Gross and Steiger (1979) showed that the resulting estimates of ϕ_i , $i = 1, 2, ..., p$ from equation (1.4.3) are strongly consistent when $\alpha > 1$.

Least gamma estimate

This is a more general estimate for which the least squares and the least absolute deviation estimates are special cases. Liu (1987) showed that the estimates of the coefficients obtained when the observations $(\xi_1, \xi_2, ..., \xi_N)$ are available by minimizing the following function:

$$f_N(\phi) = \frac{1}{N} \sum_{t=1}^{N} \left| \xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i} \right|^{\gamma}$$

(where $\gamma > 0$) are strongly consistent. By applying measure theoretic techniques, he proved the weak convergence of the resulting estimates of the parameters.

Robust estimation of parameters of a moving average process with regression terms have also been discussed by Onishchenko (1990) when the innovations have a stable distribution. Blattberg and Sargent (1971) obtained linear estimates of the parameters for the simple regression model while Cline (1989) showed that the usual least squares estimates for the regression model are consistent. Kanter and Steiger (1974) obtained the screened ratio estimates and also discussed the least squares estimates for the regression and the autoregressive models.

1.5. Basics of estimating equations

Some fundamental ideas on estimating functions were initially proposed by Godambe (1960) and Durbin (1960). Godambe's ideas were later extended and generalized in several directions by a number of authors for example by Bhapkar (1972), Godambe and Thompson (1984), Godambe (1976, 1985), and Thavaneswaran and Abraham (1988). An interesting aspect that arises from the theory of estimating functions is the ability to estimate model parameters without the requirement that the underlying distribution be known.

Consider a discrete time stochastic process $\{\xi_t , t \in I\}$ defined on a probability space (Ω , A , F) where I is the set of all positive integers. Assume that the observed values of ξ_t belong to the real space $\Re^n$ while the parameter $\theta \in \Theta$, where Θ is a compact subset of $\Re$. Let F_t^{ξ} be the sigma field generated by ξ up to time t.

Following Godambe (1960, 1976), any real function $g_n(\xi_1, \xi_2, ..., \xi_n ; \theta)$ of the observations and the parameter θ such that,

(a) $\dfrac{\partial g_n(\xi,\theta)}{\partial\theta}$ exists for all $\theta \in \Theta$,

(b) $E[g_n(\xi,\theta)]$ is differentiable with respect to θ under the integral sign and

(c) $E[\dfrac{\partial g_n(\xi,\theta)}{\partial\theta}]^2$ is finite for all $\theta \in \Theta$

is called a regular unbiased estimating function if $E_F[g(\xi_1 , ..., \xi_n ; \theta(F))] = 0$, where F is a subset of a class of distributions in $\Re^n$.

Suppose G is the class of unbiased estimating functions such that $G = \{g: g = g(\xi_1, \xi_2, ..., \xi_n ; \theta)\}$. In the class G, the estimating function g^0 is the optimum estimating function if $g^0 \in G$ and

$$E\left[\frac{g^0}{E(\partial g^0/\partial\theta)}\right]^2 \le E\left[\frac{g}{E(\partial g/\partial\theta)}\right]^2$$

for all $g \in G$ and all $\theta \in \Theta$. Estimation of the parameter θ is then accomplished by solving $g^o = 0$.

Godambe (1960) established an optimality criterion for independent and identically distributed (iid) random variables and showed that the optimum estimating equation is the maximum likelihood equation when estimating a single parameter, i.e., he showed that if the set of iid random variables $\{\xi_1, \xi_2, ..., \xi_n\}$ each having a probability density function $f(\xi,\theta)$ where θ is a parameter, then the estimating function given by

$$\sum_{t=1}^{n} \partial(\log f(\xi_t,\theta))/\partial\theta = 0$$

is optimal.

Durbin (1960) on the other hand discussed the optimality criteria for linear estimating functions under the time series regression setting while Bhapkar (1972) considered the efficiency measure for an optimal estimating function and extended Godambe's (1960) work to the multiparameter case. Godambe and Thompson (1974) discussed the optimality of estimating functions in the presence of a nuisance parameter while Godambe (1976) showed that the maximum likelihood equation is optimal, with the optimality criterion being independent of the conditioning. The optimality criteria was then applied in selecting a parameter that can be estimated most efficiently in Godambe and Thompson (1984) .

Consider a class L of estimating functions g satisfying the above conditions and having the form

$$g = \sum_{t=2}^{n} h_t a_{t-1}$$

where the function h_t is a set of uncorrelated random variables of the form $h_t = \xi_t - E[h_t \mid F^{\xi}_{t-1}]$ and thus implying that $E[h_t \mid F^{\xi}_{t-1}] = 0$, $t = 1,2, \ldots, n$ while a_{t-1} is a function of $\xi_1, \ldots, \xi_{t-1}$ and θ for $t = 2,\ldots, n$. Godambe (1985) obtained the finite sample optimal estimates for stochastic processes based on the optimal estimating function in the class L of estimating functions without making any specific probability distribution assumptions on the random variables. Specifically, he proved the following theorem.

Theorem 1.5.1

In the class L of unbiased estimating functions g , the optimum estimating function g^0 is one that minimizes

$$E\left[\frac{g^2}{\left(E(\partial g/\partial\theta)\right)^2}\right] \tag{1.5.1}$$

and this is given by

$$g^0 = \sum_{t=2}^{n} h_t a^0_{t-1} \tag{1.5.2}$$

where

$$a^0_{t-1} = \left[\frac{E\left(\frac{\partial h_t}{\partial\theta} \mid F^{\xi}_{t-1}\right)}{E\left(h_t^2 \mid F^{\xi}_{t-1}\right)}\right] \tag{1.5.3}$$

A sketch of the proof is as follows. The result in equation (1.5.3) is obtained by minimization of (1.5.1) taking into account the requirement that $E[h_t | F^{\xi}_{t-1}] = 0$ for all t. The optimality of g^0 is shown by substituting the value of a^0_{t-1} given in (1.5.3) into equation (1.5.2). The optimal estimate of the parameter of interest is then evaluated by setting (1.5.2) to zero.

In the literature, Godambe's theory of estimating equations has been applied to estimate model parameters for various classes of stochastic models as well as for the estimation of missing values. Thavaneswaran and Abraham (1988) applied Godambe's optimality criteria to study the optimal estimates for nonlinear time series models. The following example illustrates the application of this procedure to the finite variance case.

Example 1.1

Consider the first order Random coefficient model (See Nicolls and Quinn, 1982) for the process ξ_t given by the equation

$$\xi_t = (\phi + a_t)\xi_{t-1} + \eta_t$$

where $\{a_t\}$ and $\{\eta_t\}$ are mutually independent error random variables having common variance Θ_η. A unique optimal estimate of the forecast error variance can be evaluated by letting the variable h_t to be

$$h_t = (\xi_t - \phi\xi_{t-1})^2 - \Theta_\eta (1 + \xi_{t-1}^2)$$

where

$$\mathrm{Var}(\xi_t | F_{t-1}^{\xi}) = \Theta_\eta(\xi_{t-1}^2 + 1) .$$

The optimal estimate of the variance $\sigma_\eta^2 = \Theta_\eta$ is obtained as

$$\hat{\sigma}_\eta^2 = \sum_{t=2}^{n} \left[\frac{(\xi_t - \phi\xi_{t-1})^2}{(n-2)(\xi_{t-1}^2 + 1)} \right]$$

This optimal estimate of the variance has got a weighting factor as in contrast to the mean square error. The implication from this is that the optimal estimates are bound to have a smaller variance measure as compared to the least squares estimates particularly in the case of nonlinear models. Such an observation was also deduced by Thavaneswaran and Abraham (1988). In chapter two, this estimation criterion is developed for infinite variance stable sequences and is then applied to estimate coefficients for various linear models. It is also applied in the estimation of missing observations in Chapter three.

1.6. Time series forecasting

Forecasting is one of the crucial goals in time series analysis. In the Gaussian case, forecasts are the conditional expectation. This is not the case for infinite variance stable processes. An example of this is the minimum dispersion (See definition 1.2.5) linear predictor for the ARMA(1,1) model with parameters ϕ and θ and having infinite variance errors (See Cline and Brockwell, 1985). Cambanis and Soltani (1982) approached the problem by minimizing the expected absolute error and evaluated linear predictors using the spectral and moving average representations while Cline and Brockwell (1985) obtained minimum dispersion linear predictors. In the case of Gaussian processes, linear predictors have also been discussed extensively (See for example Box and Jenkins, 1976).

Consider the autoregressive model given in equation (1.4.1), Cline and Brockwell (1985) obtained linear forecasts by minimizing the dispersion. This was achieved by assuming that the forecast $\hat{Y}$ is a linear function of the observed values. The coefficients in $\hat{Y}$ are then estimated by minimizing the dispersion of the forecast error i.e., minimize $\mathrm{disp}(Y_t - \hat{Y}_t)$ with respect to the parameters.

A slightly different approach was proposed by Miamee and Pourahmadi (1988) to obtain linear predictors. They evaluated their forecasts by using the prediction decomposition of the projected value. Their results were restricted to the case where $\alpha \in (1,2]$ so as to facilitate the evaluation of the conditional expectation.

The forecasting techniques of Cline and Brockwell (1985) and Miamee and Pourahmadi (1988) would be expected to be most effective when $\alpha > 1$ since the case when it is less than one would not necessarily lead to unique projections on the predictor space.

1.7. Irregularly observed data

Irregularities in observations usually occur when a set of values are not observed. In the case of finite variance processes, irregularities could also arise as a result of deleted outliers. This is not the case when analyzing infinite variance processes since they are necessarily characterized by bursts in outlying observations. It is thus assumed here that there are no missing values.

The need for accurate estimation of missing values in irregularly observed time series arises quite often and as a result of this, there has been a growing interest in the estimation of missing values for linear models (Brockwell and Davis, 1987). Alternative techniques for handling irregularly observed data have been discussed in Parzen (1984), Jones (1980) and Nassiuma and Thavaneswaran (1992). Estimation of missing values has unfortunately been restricted to processes with finite variance (See for example Miller and Ferreiro (1984) and Abraham and Thavaneswaran (1991). This is evidently due to the prevalently conceived idea of 'inappropriateness' when considering infinite variance processes since this goes against the well established theories for finite variance variables.

In the literature on estimation of missing values for finite variance processes, two methods are used. The first one assumes that the missing value is a parameter and uses the least squares criteria. The second one assumes that the missing value is a random variable and the estimates are obtained by prediction algorithms. Abraham and Thavaneswaran (1991) applied both methods and demonstrated the superiority of the optimal method. In the case of processes with the stable errors, the least squares method is not optimal in the Gauss- Markov sense. The prediction algorithm would similarly not be used to estimate the missing values since the second moments do not exist and thus the covariance function cannot be defined. We give the general approach for estimating missing observations for finite variance processes in the following illustration.

Consider the ARMA(p,q) processes given as

$$\Phi(B)\xi_t = \Theta(B)\eta_t$$

where $\Phi(B)$ and $\Theta(B)$ are the autoregressive (AR) and the moving average (MA) operators such that the zeros of the polynomials $\Phi(z) = 1+ \phi_1 z + \phi_2 z^2 + \ ....+ \phi_p z^p$ and $\Theta(z) = 1+ \theta_1 z + \theta_2 z^2 + ...+\theta_q z^q$ lie in the region $|z| > 1$ (stationarity and invertibility conditions) and $\{\eta_t\}$ is a sequence of iid random variables. A possible representation of the estimate for a singe missing observation ξ_m is as a function of the inverse autocorrelation function. Such an estimate can be obtained as

$$\hat{\xi}_m = -\sum_{i=1}^{\infty} \rho_i^* (\xi_{m+i} + \xi_{m-i})$$

where ρ_i^* is the inverse autocorrelation function at the i^{th} lag (Chatfield, 1979) given as

$$\rho_i^* = \frac{-\psi_i + \sum_{j=1}^{\infty} \psi_j \psi_{j+i}}{\sum_{j=o}^{\infty} \psi_j^2},$$

where ψ_j are the coefficients in the MA representation of the process ξ_t.

In the simple case of the AR(1) model with parameter ϕ, and $\rho_1^* = - \phi$. Thus the estimate for the m^{th} missing value is obtained as

$$\xi_m = \frac{\phi}{1 + \phi^2} (\xi_{m+1} + \xi_{m-1}).$$

In the case of infinite variance processes, the structure of the approximate joint variation function makes this approach inapplicable. Some techniques for the estimation of missing observations when the processes are driven by infinite variance stable errors are discussed in Chapter three.

1.8. Nonstationarity in Time Series

A stochastic process which is not constant in its first or second order properties is said to be nonstationary. Linear nonstationary processes for which the first order properties vary over time (homogeneous nonstationarity) are usually differenced in order to attain stationarity and these lead to ARIMA processes studied by Box and Jenkins (1976). Nonstationary processes with time varying second order properties (heterogeneous nonstationarity) are appropriately transformed to attain stationarity (See for example Priestley (1988)). However, suitable transformations do not always exist and hence alternative ways of studying nonstationary processes is by using ARMA models in three possible forms, (i) with time varying coefficients (See Whittle, 1965), (ii) with time varying variance or (iii) with unit roots. These approaches to dealing with nonstationarity are not adequate for infinite variance processes except under some conditions since none of the moments may exist.

The continued realization that for many practical situations, the Gaussian laws are inadequate and the stable laws may be more appropriate in representing some processes as well as the curiosity posed by non-Gaussian stable laws has led to an increased interest in the study of infinite variance processes. However, past work on infinite variance processes has been restricted to the requirement that the processes be stationary.

Prediction for nonstationary processes with stable errors and index $\alpha = 2$ were initially discussed by Whittle (1965) and Abdrabbo and Priestley (1967). When $0 < \alpha < 2$, the variance does not exist and in recent papers by Chan and Tran (1989), Phillips (1990) and Chan (1990), some limiting properties of parameter estimates for the first order unit root and nearly-stationary autoregressive processes with infinite variance are studied. Thus the concept of nonstationarity for infinite variance processes is basically still in its infant stage and requires further study. This is pursued in Chapter four where some properties of nonstationary processes are discussed.

1.9. Layout of Dissertation

In this dissertation, three major inferential problems pertaining to processes with infinite variance are discussed. In chapter two, the problem of estimation of parameters in linear and nonlinear time series models is considered. An optimality criterion of an estimating function when the process is driven by infinite variance stable innovations is developed and the asymptotic properties of the resulting optimal estimates are studied. Recursive optimal estimates are also derived. Applications to parameter estimation in AR, MA, and ARMA models is discussed. In particular, the consistency of optimal estimates of the AR(p) parameters as well as their recursive forms are discussed.

The estimation of missing observations for processes driven by infinite variance stable innovations is discussed in chapter three. Various techniques are considered and these include the application of the extended optimality criteria discussed in chapter two to estimate the missing values. A linear interpolation approach to the estimation of missing observations is also developed in this chapter. Estimates of missing observations based on minimizing the λ^{th} norm ($0 < \lambda < 2$) are discussed. The advantages and disadvantages of each estimating technique is also discussed.

Chapter four is concerned with nonstationary processes. Three cases of nonstationarity are considered: the case of a unit root, the time varying coefficients model and the case of time varying dispersion model. Prediction and coefficient estimation for the processes with time varying coefficients are discussed and properties of an optimal estimate for AR(1) model with unit root are studied.

CHAPTER TWO

2. PARAMETER ESTIMATION FOR INFINITE VARIANCE MODELS USING ESTIMATING EQUATIONS

2.1. Introduction

Estimation using the least squares method is a popular approach for a wide range of models, both stochastic and nonstochastic. In the finite variance case, the Gauss-Markov theorem guarantees optimality of the least squares estimates in the minimum variance sense and properties of such estimates are well known. Other estimating criteria based on the finite variance assumption and which have been extensively studied in the literature include the method of moments, the maximum likelihood criterion, and the minimum absolute deviation technique. The application of some of these methods in estimating parameters of processes whose observations follow the stable laws would normally seem 'inappropriate' since neither the first nor higher order moments necessarily exist and thus the resulting estimates may not have the same optimality properties.

A number of studies have been carried out to justify the application of the above techniques to the infinite variance case and these have particularly focused on the limiting properties of the estimates. These are properties that are based on the assumption that the available sample are large in size. The least squares estimates for the coefficients of the autoregressive model have for example been shown to converge under the stable law assumption on the errors (See Yohai and Maronna (1977); Hannan and Kanter (1977); Kanter and Steiger (1974)). The least absolute deviation estimates have similarly been shown to converge (Gross and Steiger, 1979).

The optimality criterion of estimating equations based on independent and identically distributed (i.i.d) random variables was developed by Godambe (1960) where he showed the relationship between the optimum estimating equation and the maximum likelihood equation when estimating a single parameter. Durbin (1960) on the other hand established the optimality criteria for linear estimating functions under the vector valued time series regression setting and Bhapkar (1972) discussed the efficiency measure for an optimal estimating function and extended Godambe's (1960) work to the multiparameter case. Based on these developments, Godambe (1985) developed an optimality criterion for the finite sample stochastic setting. This was applied by Thavaneswaran and Abraham (1988) to the estimation of parameters in nonlinear models where they showed that the resulting estimates have a smaller variance as compared to the least squares estimates.

This chapter deals with parameter estimation for processes driven by infinite variance stable innovations. We propose to extend Godambe's (1985) theory of optimal estimating functions for stochastic processes with finite variance to processes with infinite variance errors by minimizing the process dispersion, which corresponds to minimizing the tail probability of the innovations. The process innovations are assumed to be independent random variables having an infinite variance stable distribution with characteristic exponent α $(1 < \alpha \leq 2)$ and scale parameter c(t), which is assumed to be a general function that changes with time.

In section 2.2 we give some well known results on the conditional expectations of symmetric stable processes. The extended sufficiency conditions for the existence of an optimal estimating function g^o in the set G of unbiased estimating functions are laid out in section 2.3. In Section 2.4, the strong consistency of the optimal estimate is discussed while recursive optimal estimates are derived in Section 2.5. In Section 2.6, we consider the application of the optimality criterion to estimate parameters for the AR(p) process driven by an infinite variance stable innovation sequence. Recursive optimal estimates for the parameters are also discussed in this Section. A simulation study is used to examine the relative closeness of the optimal estimates of the parameters to their true values. The performance of the optimal estimates relative to the least squares estimates is shown through simulation studies.

Section 2.7 deals with the moving average (MA) process of order q which is driven by infinite variance stable innovations. Optimal estimates for the model parameters such that the dispersion is minimized are similarly obtained. In Section 2.8, a procedure for evaluating optimal estimates for the general ARMA(p,q) process is similarly discussed.

2.2. Conditional expectations

A fundamental requirement in the development of the optimality criteria for an optimal estimating function for finite variance random variables is the existence of the conditional mean and variance. In the infinite variance setting, we similarly require the existence of the conditional dispersion. We thus start by examining stable random variables which are jointly distributed before developing the optimality criteria. We consider a set of independent random variables $\{\eta_t: t = 1, 2, ..., n\}$ having scale parameters c(t) and common characteristic exponent $\alpha > 1$. It is well known that linear combinations of the stable random variables η_t are also stable with characteristic exponent α. The following Theorems are thus focused on such linear combinations of independent random variables.

Theorem 2.2.1

Suppose the random variables ξ and χ are given as the following linear combinations of independent stable random variables, $\{\eta_t\}$, i.e.,

$$\xi = \sum_{t=1}^{n} a_t \eta_t \quad \text{and} \quad \chi = \sum_{t=1}^{n} b_t \eta_t$$

where $\{a_t\}, \{b_t\} \in \Re$ are non random coefficients such that at least one value in the set $\{b_t\}$ is not zero and η_t is as defined above. Then the conditional expectation of ξ given χ is obtained as $E(\xi|\chi) = \lambda\chi$ where the constant λ is obtained as

$$\lambda = \frac{\sum_{t=1}^{n} c(t) a_t |b_t|^{\alpha-1}}{\sum_{t=1}^{n} c(t) |b_t|^{\alpha}}$$

Proof:

Let the characteristic function for each η_t be given by $\Phi_t(\omega) = \exp(-c(t)|\omega|^{\alpha})$. Then the characteristic function m(u,v) of the joint vector (ξ,χ) is obtained as

$$\ln[m(u,v)] = \sum_{t=1}^{n} \ln[\Phi_t(a_t u + b_t v)].$$

Now since $\alpha > 1$, we have

$$\frac{\partial}{\partial u}[\ln(m(u,v))]_{u=0} = \frac{\sum_{t=1}^{n} c(t) b_t |a_t|^{\alpha-1}}{\sum_{t=1}^{n} c(t)|a_t|^{\alpha}} \frac{d}{dv} \ln(m(0,v))$$

and thus a sufficient condition for the conditional expectation to be $\lambda\chi$ is that

$$\frac{\partial}{\partial v} \ln[m(u,v)] = \lambda \frac{d}{dv} \ln[m(0,v)]$$

for all v.

The conditional dispersion of ξ given χ is shown in the following Theorem.

Theorem 2.2.2

Suppose the random variables ξ and χ are as given in Theorem 2.2.1, then

(a) The random variable h given as $h = \xi - E(\xi|\chi)$ simplifies to the form

$$h = \sum_{t=1}^{n} (a_t - \lambda b_t)\eta_t.$$

and has a stable distribution with characteristic exponent α and dispersion c(h) which corresponds to the tail probability (Cline, 1983). This is obtained as

$$c(h) = \sum_{t=1}^{n} |a_t - \lambda b_t|^{\alpha} c(t).$$

(b) The conditional distribution of h given χ is stable with characteristic exponent α and dispersion given by

$$\mathrm{disp}(h|\chi) = \sum_{t=1}^{n} |a_t - (\lambda + \lambda_1) b_t|^{\alpha} c(t)$$

where $$\lambda_1 = \frac{\sum_{t=1}^{n} (a_t - \lambda b_t) |b_t|^{\alpha} c(t)}{\sum_{t=1}^{n} |b_t|^{\alpha} c(t)}.$$

Proof:

The result in (a) simply follows from the characterization of the stable random variables and definition (1.2.5) of the dispersion while (b) is obtained by noting that

$$\mathrm{disp}(h|\chi) = \mathrm{disp}(h - E(h|\chi))$$

$$= \mathrm{disp}\left(\sum_{t=1}^{n} [a_t - (\lambda + \lambda_1) b_t] \eta_t \right)$$

$$= \sum_{t=1}^{n} |a_t - (\lambda + \lambda_1) b_t|^{\alpha} c(t)$$

The constant λ_1 is obtained as in Theorem 2.2.1.

Related results have been extensively discussed by Lukacs and Laha (1964). Thavaneswaran and Thompson (1991) have applied these ideas in the study of filtering and interpolation problems for infinite variance stable processes.

2.3. Optimality criterion for parameter estimation

Suppose the sequence of random variables $\{\eta_t ; t = 1,2,...\}$ is symmetric stable with characteristic index $\alpha \in (1,2]$ and is defined on the probability space (Ω , A , F). Suppose also that the sequence $\{\xi_t , t \in I\}$, where I is the set of all positive integers, is defined on the above probability space and assume that the observed values of ξ_t belong to the real space $\Re^n$. Let $(\xi_1,\xi_2,\xi_3,...,\xi_n)$ be a sequence of random variables such that the conditional distribution of ξ_t given the past values is stable with mean $f(\xi_1,\xi_2,\xi_3,...\xi_{t-1},\theta)$ and its dispersion (denoted disp) is proportional to c(t), i.e.,

$$E[\xi_t | F^{\xi}_{t-1}] = f(\xi_1,\xi_2,\xi_3,...\xi_{t-1},\theta)$$

and

$$\text{disp}(\xi_t | F^{\xi}_{t-1}) = Kc(t)$$

where $K \in \Re$ is a constant and F^{ξ}_{t-1} is the sigma field up to time t - 1. Proceeding as in Godambe (1985), we consider a class of unbiased estimating functions

$$G_n = \left\{ g_n : g_n = \sum_{t=2}^{n} a_{t-1} h_t \right\}$$

where a_{t-1} is a function of $(\xi_1,\xi_2,\xi_3,...\xi_{t-1})$ and θ , i.e., a_{t-1} is F^{ξ}_{t-1} measurable, while the sequence of independent random variables $\{h_t\}$ are obtained as

$$h_t = \xi_t - E[\xi_t | F^{\xi}_{t-1}].$$

We will henceforth drop the subscript n on G and g and it will be understood that they are based on samples of size n unless otherwise specified.

In analogy with the finite variance case, among the estimating functions $g \in G$, the estimating function g^o is said to be optimal if it minimizes the dispersion function of the quotient given by

$$\mathrm{disp}\left(g/E[(\partial g/\partial \theta)|F^{\xi}_{t-1}]\right) \tag{2.3.1}$$

at $g=g^o$. We assume here that the conditional expectation $E[(\partial g/\partial \theta)|F^{\xi}_{t-1}]$ exists for all $g \in G$. When $\alpha = 2$, the variance is finite and this condition is the same as that of Godambe (1985). In the following theorem, a sufficient condition for the above optimality is given.

Theorem 2.3.1

In the class G of unbiased estimating functions, a sufficient condition for the estimating function g^o to be optimal is that

$$\mathrm{disp}(g - g^o) = \mathrm{disp}(g^o) + \mathrm{disp}(g) + K\, E[(\partial g/\partial \theta)|F^{\xi}_{t-1}] \tag{2.3.2}$$

for all $(g, g^o) \in G$ and for a real constant K.

Proof:

From equation (2.3.2), it follows by standardization using $E[(\partial g/\partial \theta)|F^{\xi}_{t-1}]$ that

$$\mathrm{disp}\left(\frac{g}{E[(\partial g/\partial \theta)|F^{\xi}_{t-1}]} - \frac{g^o}{E[(\partial g^o/\partial \theta)|F^{\xi}_{t-1}]}\right) = \mathrm{disp}\left(\frac{g}{E[(\partial g/\partial \theta)|F^{\xi}_{t-1}]}\right)$$

$$+\, \mathrm{disp}\left(\frac{g^o}{E[(\partial g^o/\partial \theta)|F^{\xi}_{t-1}]}\right) + K\, E[(\partial g/\partial \theta)|F^{\xi}_{t-1}]\, \frac{1}{E[(\partial g/\partial \theta)|F^{\xi}_{t-1}]}\ . \tag{2.3.3}$$

When $g = g^o$, equation (2.3.3) reduces to the form:

$$0 = 2\,\mathrm{disp}(g^o/E[(\partial g^o/\partial\theta)|F^{\xi}_{t-1}]) + K \tag{2.3.4}$$

and thus

$$K = -2\,\mathrm{disp}(g^o/E[(\partial g^o/\partial\theta)|F^{\xi}_{t-1}]).$$

We can then write equation (2.3.3) as

$$\mathrm{disp}\left(\frac{g}{E[(\partial g/\partial\theta)|F^{\xi}_{t-1}]}\right) + (1-2)\mathrm{disp}\left(\frac{g^o}{E[(\partial g^o/\partial\theta)|F^{\xi}_{t-1}]}\right) \tag{2.3.5}$$

$$= \mathrm{disp}\left(\frac{g}{E[(\partial g/\partial\theta)|F^{\xi}_{t-1}]}\right) - \mathrm{disp}\left(\frac{g^o}{E[(\partial g^o/\partial\theta)|F^{\xi}_{t-1}]}\right)$$

for all $g \in G$. The result then follows by recognizing that disp(.) is a positive function and thus

$$\mathrm{disp}\left(\frac{g}{E[(\partial g/\partial\theta)|F^{\xi}_{t-1}]}\right) \geq \mathrm{disp}\left(\frac{g^o}{E[(\partial g^o/\partial\theta)|F^{\xi}_{t-1}]}\right).$$

In the next Theorem, we further establish the specific form of the optimal estimating function g^o among the class of linear estimating functions $g \in G$.

Theorem 2.3.2

In the class G of unbiased linear estimating functions, the optimal estimating function g^o is obtained as

$$g^o = \sum_{t=2}^{n} a^o_{t-1} h_t \tag{2.3.6}$$

where

$$a^{o}_{t-1} = \left| \frac{E[(\partial h_t/\partial\theta)|F^{\xi}_{t-1}]}{disp(h_t|F^{\xi}_{t-1})} \right|^{1/(\alpha-1)} sgn(h_t) \tag{2.3.7}$$

and $|x|sgn(x) = x$.

proof:

For each $g \in G$, the dispersion of g is obtained as

$$disp(g) = \sum_{t=2}^{n} |a_{t-1}|^{\alpha} c(t) \tag{2.3.8}$$

where c(t) is the conditional dispersion of h_t given the observations up to time t-1. Similarly, obtaining the derivative of g with respect to the parameter θ leads to the following relation:

$$E[(\partial g/\partial\theta)|F^{\xi}_{t-1}] = \sum_{t=2}^{n}\left[a_{t-1}E\left(\frac{\partial h_t}{\partial\theta}|F^{\xi}_{t-1}\right)sgn(h_t) + \left(\frac{\partial a_{t-1}}{\partial\theta}|F^{\xi}_{t-1}\right)E\left(h_t|F^{\xi}_{t-1}\right)\right]$$

and since $E\left(h_t|F^{\xi}_{t-1}\right) = 0$, we can write this as

$$E[(\partial g/\partial\theta)|F^{\xi}_{t-1}] = \sum_{t=2}^{n}\left[a_{t-1}E\left(\frac{\partial h_t}{\partial\theta}|F^{\xi}_{t-1}\right)\right]sgn(h_t)\,. \tag{2.3.9}$$

From equation (2.3.4) of Theorem 2.3.1, it implies that when k=-2, we obtain the dispersion of the optimal estimating function as

$$disp(g^o) = E[(\partial g^o/\partial\theta)|F^{\xi}_{t-1}].$$

Based on this, we can equate equations (2.3.8) and (2.3.9) when $g = g^o$ and this leads to

$$\sum_{t=2}^{n} |a_{t-1}^{o}|^{\alpha} c(t) = \sum_{t=2}^{n} \left[a_{t-1}^{o} E\left(\frac{\partial h_t}{\partial \theta} | F_{t-1}^{\xi} \right) \right] sgn(h_t).$$

The result then follows by solving

$$|a_{t-1}^{o}|^{\alpha} c(t) = a_{t-1}^{o} E\left(\frac{\partial h_t}{\partial \theta} | F_{t-1}^{\xi} \right) sgn(h_t).$$

Remark: when $\alpha = 2$, the result for a_{t-1}^{o} reduces to

$$a_{t-1}^{o} = E\Big((\partial h_t / \partial \theta) | F_{t-1}^{\xi}] \Big) / E[(h_t^2) | F_{t-1}^{\xi}]$$

which is the same as that of Godambe (1985).

As an illustration, we consider the optimal estimation of the parameters $\phi_1, \phi_2, \ldots, \phi_k$ for a linear process ξ_t of order k given by

$$\xi_t = \Phi^T U(F_{t-1}^{\xi}) + \eta_t \tag{2.3.10}$$

where

$$\Phi = (\phi_1, \phi_2, \ldots, \phi_k)^T,\ U(F_{t-1}^{\xi}) = [u(F_{t-1}^{\xi}), u(F_{t-2}^{\xi}), \ldots, u(F_{t-k}^{\xi})]^T$$

and η_t is a stable random variable having characteristic exponent α and scale parameter c(t). The vector $U(F_{t-1}^{\xi})$ is a linear or nonlinear bounded function of the observations up to time t-1 and A^T represents the transpose of the vector A. We assume that ξ_t admits a moving average representation of the form

$$\xi_t = \sum_{i=0}^{\infty} \pi_i \eta_{t-i}$$

for some π_i such that

$$\sum_{i=0}^{\infty} |\pi_i|^{\alpha} < \infty.$$

Based on the optimality criteria developed in the previous section, we consider the class of vector valued estimating functions $\mathbf{G}_n$ based on a sample of size n and given by

$$\mathbf{G}_n = \{\mathbf{g}_n: \mathbf{g}_n = \sum_{t=k+1}^{n} \mathbf{a}_{t-1} h_t\}.$$

Following the criteria in Theorem 2.3.1, the optimal estimating function vector $\mathbf{g}_n^0$ is given by

$$\mathbf{g}_n^0 = \sum_{t=k+1}^{n} \mathbf{a}_{t-1}^0 h_t$$

where

$$h_t = \xi_t - \sum_{i=1}^{k} \phi_i u(F_{t-i}^{\xi})$$

and the components in the vector $\mathbf{a}_{t-1}^{o}$ are given by

$$a_{t-i}^{o} = \left| \frac{E[(\partial h_t/\partial\theta)|F_{t-i}^{\xi}]}{\operatorname{disp}(h_t|F_{t-i}^{\xi})} \right|^{1/(\alpha-1)} \operatorname{sgn}(h_t), \quad i = 1, 2, ..., k.$$

These can be written in vector form as

$$h_t = \xi_t - \Phi^T U(\xi_{t-1})$$

and the vector

$$\mathbf{a}^{o}_{t-1}=\left[\left|u(F^{\xi}_{t-1})/c(t)\right|^{1/(\alpha-1)},\left|u(F^{\xi}_{t-2})/c(t)\right|^{1/(\alpha-1)}, \ldots,\left|u(F^{\xi}_{t-k})/c(t)\right|^{1/(\alpha-1)}\right]^{T}$$

where for simplicity, $\mathbf{U}(\xi_{t-1}) = \mathbf{U}(F^{\xi}_{t-1})$. This leads to the optimal estimating equation in vector form as

$$\mathbf{g}^{0}_{n} = \sum_{t=k+1}^{n} \mathbf{a}^{0}_{t-1}(\xi_t - \Phi^{T}\mathbf{U}(\xi_{t-1})).$$

Each component in the vector $\mathbf{g}^{0}_{n}$ is obtained as

$$g^{0}_{n}(j) = \sum_{t=k+1}^{n} \left|\frac{u(F^{\xi}_{t-j})}{c(t)}\right|^{1/(\alpha-1)}\left(\xi_t - \sum_{i=1}^{k}\phi_i u(F^{\xi}_{t-i})\right)\mathrm{sgn}(u(F^{\xi}_{t-j})) \quad , \text{ for } j=1,2,\ldots k.$$

The parameters are then estimated by solving $\mathbf{g}^{0}_{n} = 0$.

Let the column vector of optimal estimates of the parameters be $\Phi^{o} = (\phi^{o}_{1}, \phi^{o}_{2}, \ldots, \phi^{o}_{k})^{T}$ and define the square matrix $\mathbf{C}_n$ as

$$\mathbf{C}_n = \sum_{t=k+1}^{n} \mathbf{a}^{0}_{t-1}[\mathbf{U}(\xi_{t-1})]^{T}$$

which has value in the j^{th} row and i^{th} column given as

$$c_{ji} = \sum_{t=k+1}^{n} \left|\frac{u(F^{\xi}_{t-i})}{c(t)}\right|^{1/(\alpha-1)} u(F^{\xi}_{t-j})\mathrm{sgn}(u(F^{\xi}_{t-j}))$$

where i,j = 1, 2,..., k.

Let the vector $\mathbf{D}_n$ be defined by

$$\mathbf{D}_n = \left[\left(\sum_{t=k+1}^{n} \left|\frac{u(F^{\xi}_{t-1})}{c(t)}\right|^{1/(\alpha-1)} \xi_t, \sum_{t=k+1}^{n} \left|\frac{u(F^{\xi}_{t-2})}{c(t)}\right|^{1/(\alpha-1)} \xi_t, \ldots, \sum_{t=k+1}^{n} \left|\frac{u(F^{\xi}_{t-k})}{c(t)}\right|^{1/(\alpha-1)} \xi_t \right) \right]^T$$

$$= \sum_{t=k+1}^{n} \mathbf{a}^0_{t-1} \xi_t.$$

The optimal estimator Φ^o for the parameter Φ is then obtained by solving

$$\mathbf{C}_n \Phi = \mathbf{D}_n.$$

In the one parameter case, the optimal estimate ϕ^o is obtained as

$$\phi^o = \frac{\sum_{t=2}^{n} |u(F^{\xi}_{t-1})/c(t)|^{1/(\alpha-1)} \xi_t \operatorname{sgn}(u(F^{\xi}_{t-1}))}{\sum_{t=2}^{n} |u(F^{\xi}_{t-1})/c(t)|^{\alpha/(\alpha-1)} \operatorname{sgn}(u(F^{\xi}_{t-j}))}. \tag{2.3.11}$$

Example: (Simple linear regression)

Consider the simple linear regression model driven by infinite variance stable errors. Suppose that $\{(Y_i, X_i),\ i=1,2,\ldots,n\}$ is a set of pairwise observations with X being the independent variable and Y the response variable. The following regression model is then fitted to the data:

$$Y_i = \theta X_i + \eta_i \quad i = 1,2,..,n$$

where the sequence $\{\eta_i, i=1,2,..,n\}$ is of independent symmetric stable random variables having dispersion c(i). We define the sequence of independent innovations h_i to be $h_i = Y_i - \theta X_i$ and by Theorem 2.3.2, we have

$$a_i^o = |X_i/c(i)|^{1/(\alpha-1)} \mathrm{sgn}(X_i).$$

Substituting this into equation (2.3.6), we obtain

$$g^o = \sum_{i=1}^{n} |X_i/c(i)|^{1/(\alpha-1)} (Y_i - \theta X_i) \mathrm{sgn}(X_i)).$$

Solving $g^o = 0$ for θ then leads to the optimal estimate of the parameter θ as

$$\theta^o(\alpha) = \left[\sum_{i=1}^{n} |X_i/c(i)|^{\alpha/(\alpha-1)} \mathrm{sgn}(X_i) \right]^{-1} \sum_{i=1}^{n} |X_i/c(i)|^{1/(\alpha-1)} Y_i \mathrm{sgn}(X_i).$$

It is easy to see that the optimal estimate turns out to be the same as the minimum dispersion linear estimator of Blattberg and Sargent (1971) when c(i) is constant for all i.

2.4. Asymptotic properties of the optimal estimate

We first of all recall some of the definitions and basic properties of random variables having regularly varying tails. Further expositions of the definitions and basic properties given here can be found in Feller (1971).

Definition 2.4.1

Consider a positive function $L(\eta)$. Such a function is said to vary slowly at infinity if for all positive η, the following limit holds:

$$\lim_{t\to\infty} \frac{L(t\eta)}{L(t)} \to 1.$$

Definition 2.4.2

A sequence of random variables η is said to be in the domain of attraction of the stable law with characteristic exponent α if there exists a slowly varying function $L(\eta)$ such that as $\eta \to \infty$,

$$1 - F(\eta) \approx p\eta^{-\alpha}L(\eta)$$

where $$p = \lim_{\eta\to\infty} \frac{[1 - F(\eta)]}{1 - F(\eta) + F(-\eta)}.$$

A limiting property for infinite variance stable random variables which is similar to the central limit theorem for the finite variance case exists under certain conditions. This is given as follows:

Let the sequence of iid random variables $\{\xi_i, i = 1, 2, ..., n\}$ be defined in the domain of attraction of the stable law with characteristic exponent $\alpha \in (0,2)$ and distribution function F. Suppose that there exist constants a_n and b_n defined as

$$a_n = \inf\{x: [1 - F(x) + F(-x)] \leq n^{-1}\}$$

and
$$b_n = E(\xi_i I_{[|\xi_i| \leq a_n]})$$

where $I_{[|\xi_i| \leq a_n]}$ is an indicator variable taking on values 0 or 1. The following limit then exists:

$$\frac{1}{a_n}\sum_{i=1}^{n}(\xi_i - b_n) \xrightarrow{d} S_\alpha,$$

where S_α is a stable random variable with characteristic exponent α. When $\alpha > 1$, b_n is simply the expected value $E(\xi_i)$. A similar limiting distribution is obtained for the case of a linear function given by

$$\xi_t = \sum_{i=1}^{\infty}\theta_i\eta_{t-i}.$$

Let the sequence of iid random variables $\{\eta_t, t = 1, 2,...\}$ satisfy the following relations:

(a) $P(|\eta_t| > x) = x^{-\alpha}L(x)$

(b) $\dfrac{P(\eta_t > x)}{P(|\eta_t| > x)} \rightarrow p$ and $\dfrac{P(\eta_t \leq -x)}{P(|\eta_t| > x)} \rightarrow q,$ for $x > 0$

and $p \in (0,1)$, $q = 1 - p$, and $L(x)$ is a slowly varying function at infinity. Then there exists constants a_n and b_n such that

$$a_n^{-1}\left(\sum_{t=1}^{n}\xi_t - n\sum_{i=0}^{\infty}\theta_i b_n\right) \xrightarrow{d} \left(\sum_{i=0}^{\infty}\theta_i\right)S_\alpha.$$

Considering the general model given in (2.3.10) as

$$\xi_t = \Phi^T U(F^{\xi}_{t-1}) + \eta_t,$$

the least squares estimates for the vector of coefficients Φ based on the observations $(\xi_1, \xi_2, ..., \xi_n)$ is

$$\hat{\Phi} = \left[\sum_{t=2}^{n}[U(F^{\xi}_{t-1})][U(F^{\xi}_{t-1})]^T\right]^{-1}\left[\sum_{t=2}^{n}\xi_t[U(F^{\xi}_{t-1})]^T\right]. \tag{2.4.1}$$

Kanter and Steiger (1974) showed that in the case when $U(F^{\xi}_{t-1}) = (\xi_{t-1}, \xi_{t-2}, ..., \xi_{t-k})$, the above least squares estimators of the model parameters converge in probability. Yohai and Maronna (1977) extended the distribution of $\{\eta_t\}$ to the case whereby $E[\log^+(\eta_t)] < \infty$ and also showed that the estimate given in (2.4.1) converges almost surely. Hannan and Kanter (1977) also showed the consistency of the least squares estimates using a slightly different approach from that of Yohai and Maronna (1977).

Based on the concept of optimal estimating functions, we consider the strong consistency of the optimal estimate ϕ^o given in (2.3.11) for the one parameter case.

Before we show the convergence of ϕ^o for the one parameter case, we state the following Lemma due to Chartterji (1969) (a simpler version and relevant to our setup here).

Theorem 2.4.1 (Chatterji (1969)

Suppose the measurable function W and the sequence of measurable functions$\{W_n, n > 1\}$ are such that W is defined in the L^{λ} space where $\lambda \in (1,2)$, and

$$P(|W_n| \geq z) \leq P(|W| \geq z) \text{ almost surely for } z \geq 0.$$

Then in the L^{λ} space,

$$\lim_{n \to \infty} n^{-1/\lambda} \sum_{i=1}^{n} (W_i - E(W_i | F^{W}_{i-1})) = 0 \qquad \text{almost surely,}$$

where F^{W}_{i-1} is the sigma field up to time i - 1.

A basic implication of this Theorem is that the sequence of functions $\{W_n, n > 1\}$is also measurable in the L^{λ} space and this facilitates the evaluation of the conditional expectation. The convergence for the above general functions will be utilized in showing the convergence of the optimal estimates as in the following Theorem for the one parameter case.

Theorem 2.4.2

Let the process $\{\xi_t\}$ be given as

$$\xi_t = \phi u(F^{\xi}_{t-1}) + \eta_t.$$

where $u(F^{\xi}_{t-1})$ is a bounded function. Let the corresponding optimal estimate for the parameter ϕ be

$$\phi^{o} = \frac{\sum_{t=2}^{n} |u(F^{\xi}_{t-1})/c(t)|^{1/(\alpha-1)} \xi_t \operatorname{sgn}(u(F^{\xi}_{t-1}))}{\sum_{t=2}^{n} |u(F^{\xi}_{t-1})/c(t)|^{\alpha/(\alpha-1)} \operatorname{sgn}(u(F^{\xi}_{t-j}))}. \tag{2.4.2}$$

Then the estimate ϕ^{o} converges almost surely to the true parameter ϕ.

Proof:

The deviation of the optimal estimate ϕ^o from the true parameter ϕ can be written as

$$\phi^O - \phi = \frac{\sum_{t=2}^{n} |u(F_{t-1}^{\xi})/c(t)|^{1/(\alpha-1)} \xi_t \operatorname{sgn}(u(F_{t-1}^{\xi})) - \phi \sum_{t=2}^{n} |u(F_{t-1}^{\xi})/c(t)|^{\alpha/(\alpha-1)} \operatorname{sgn}(u(F_{t-1}^{\xi}))}{\sum_{t=2}^{n} |u(F_{t-1}^{\xi})/c(t)|^{\alpha/(\alpha-1)} \operatorname{sgn}(u(F_{t-1}^{\xi}))}$$

$$= \frac{\sum_{t=2}^{n} |u(F_{t-1}^{\xi})/c(t)|^{1/(\alpha-1)} [\xi_t - \phi u(F_{t-1}^{\xi})] \operatorname{sgn}(u(F_{t-1}^{\xi}))}{\sum_{t=2}^{n} |u(F_{t-1}^{\xi})/c(t)|^{\alpha/(\alpha-1)} \operatorname{sgn}(u(F_{t-1}^{\xi}))}$$

$$= \frac{\sum_{t=2}^{n} |u(F_{t-1}^{\xi})/c(t)|^{1/(\alpha-1)} \eta_t \operatorname{sgn}(u(F_{t-1}^{\xi}))}{\sum_{t=2}^{n} |u(F_{t-1}^{\xi})/c(t)|^{\alpha/(\alpha-1)} \operatorname{sgn}(u(F_{t-1}^{\xi}))} \qquad (2.4.3)$$

where

$$\eta_t = \xi_t - \phi u(F_{t-1}^{\xi}).$$

To now prove the consistency of the optimal estimate, it basically requires us to show that $A_n(\phi^o - \phi)$, (where A_n is a regularly varying function of the form $n^{-1/\vartheta}$ and $\vartheta > \alpha$), converges to zero almost surely as $n \to \infty$. Suppose that for simplicity, c(t) is a constant for all t, then equation (2.4.3) can be written as

$$\phi^{o} - \phi = \frac{\sum_{t=2}^{n} |u(F^{\xi}_{t-1})|^{1/(\alpha-1)} \eta_t \operatorname{sgn}(u(F^{\xi}_{t-1}))}{\sum_{t=2}^{n} |u(F^{\xi}_{t-1})|^{\alpha/(\alpha-1)} \operatorname{sgn}(u(F^{\xi}_{t-1}))}. \tag{2.4.4}$$

Now since η_t and $u(F^{\xi}_{t-1})$ are independent and $E(\eta_t|\xi_{t-1}, \xi_{t-2}, \ldots) = 0$ and because $1/(\alpha-1) > \alpha$ implying that $|u(F^{\xi}_{t-1})|^{1/(\alpha-1)}$ is stable with index $\alpha(\alpha-1) < 2$ (Logan et al., 1973), it follows that $|u(F^{\xi}_{t-1})|^{1/(\alpha-1)}\eta_t$ is stable with index $\vartheta = \min(\alpha, \alpha(\alpha-1))$ and thus from Lemma 2.4.1,

$$\lim_{n \to \infty} n^{-1/\vartheta} \sum_{t=2}^{n} |u(F^{\xi}_{t-1})|^{1/(\alpha-1)} \eta_t = 0 \qquad \text{almost surely.}$$

Next, we need to show that the denominator of the second component of equation (2.4.4) converges to infinity. We note here that since $|u(F^{\xi}_{t-1})|^{1/(\alpha-1)}$ is similarly bounded by, say a constant $M > 0$, it implies that

$$\sum_{t=2}^{n} |u(F^{\xi}_{t-1})|^{\alpha/(\alpha-1)}$$

is similarly bounded by $(n-1)M$. Thus the folowing limit exist almost surely:

$$\lim_{n \to \infty} n^{-1/\kappa} \sum_{t=2}^{n} |u(F^{\xi}_{t-1})|^{\alpha/(\alpha-1)} = \infty \qquad \text{almost surely}$$

where $\alpha^{-1} < k < (\alpha - 1)^{-1}$.

Combining the two results implies that there exists a real value $q = k - \vartheta$ such that ϕ^o converges almost surely.

2.5. Recursive optimal estimates

It is often the case that observations are obtained recursively and this calls for estimates based on previous estimates and the current observation. This is particularly appropriate for dynamic systems. In this Section, a recursive form of the optimal estimate is established.

Consider a set of n observations from the process ξ_t generated by a nonlinear model given in equation (2.3.10). The optimal estimating equation is obtained as

$$g_n^0 = \sum_{t=2}^{n} a_{t-1}^0 (\xi_t - \Phi^T U(\xi_{t-1})),$$

where the vector

$$a_{t-1}^o = \left[|u(F_{t-1}^{\xi})/c(t)|^{1/(\alpha-1)}, |u(F_{t-2}^{\xi})/c(t)|^{1/(\alpha-1)}, \ldots, |u(F_{t-k}^{\xi})/c(t)|^{1/(\alpha-1)} \right]^T.$$

Let the optimal estimate based on the n observations be denoted as Φ_n^o . This is then evaluated as

$$\Phi_n^o = \left(\sum_{t=2}^{n} a_{t-1}^0 [U(\xi_{t-1})]^T \right)^{-1} \left(\sum_{t=2}^{n} a_{t-1}^0 \xi_t \right).$$

When the n+1 observation becomes available, its optimal estimate based on Φ_n^o and the new observation ξ_{n+1} is evaluated as follows:

$$\Phi_{n+1}^o = \left(\sum_{t=2}^{n+1} a_{t-1}^0 [U(\xi_{t-1})]^T \right)^{-1} \left(\sum_{t=2}^{n+1} a_{t-1}^0 \xi_t \right).$$

Thus the difference between the two estimates can be evaluated as

$$\Phi^o_{n+1} - \Phi^o_n = K_{n+1} a^0_n \left(\xi_n - (\Phi^o_n)^T [U(\xi_{t-1})]\right)$$

where

$$K^{-1}_{n+1} = \sum_{t=2}^{n+1} a^0_{t-1} [U(\xi_{t-1})]^T$$

$$= K^{-1}_n + a^0_n [U(\xi_n)]^T$$

The special case of a model with one parameter leads to the following recursive estimate:

$$\phi^o_{n+1} = \phi^o_n + \frac{K_n |u(F^{\xi}_n)|^{1/(\alpha-1)} [\xi_{n+1} - \phi^o_n u(F^{\xi}_n)]}{1 + K_n |u(F^{\xi}_n)|^{\alpha/(\alpha-1)}}$$

$$= \phi^o_n + \frac{K_n |u(F^{\xi}_n)|^{1/(\alpha-1)} \eta_{n+1}}{1 + K_n |u(F^{\xi}_n)|^{\alpha/(\alpha-1)}} \qquad (2.5.1)$$

where

$$K^{-1}_n = \sum_{t=2}^{n} |u(F^{\xi}_{t-1})|^{\alpha/(\alpha-1)}, \quad \eta_{n+1} = \xi_{n+1} - \phi^o_n u(F^{\xi}_n)$$

and

$$K_{n+1} = \frac{K_n}{1 + K_n |u(F^{\xi}_n)|^{\alpha/(\alpha-1)}}.$$

The consistency of the recursive estimate in the one parameter case is shown in the following Theorem.

Theorem 2.5.1

Suppose the sequence $\{\eta_t, t = 1,2,...\}$ is of symmetric stable random variables and the recursive optimal estimate based on n+1 observations is as given in (2.5.1). Then the estimate in equation (2.5.1) is strongly consistent.

Proof:

The value of K_{n+1} can be obtained in the form

$$K_{n+1} = \frac{K_0}{1 + K_0 \sum_{t=2}^{n+1} |u(F^{\xi}_{t-1})|^{\alpha/(\alpha-1)}}$$

where $K_0 = 1$ and thus the result in (2.5.1) can be written as

$$\phi^o_{n+1} = \phi^o_0 + \frac{K_0 \sum_{t=2}^{n+1} |u(F^{\xi}_{t-1})|^{1/(\alpha-1)} \eta_t}{1 + K_0 \sum_{t=2}^{n+1} |u(F^{\xi}_{t-1})|^{\alpha/(\alpha-1)}}$$

$$= \phi^o_0 + \frac{K_0 \sum_{t=2}^{n+1} |u(F^{\xi}_{t-1})|^{1/(\alpha-1)} \eta_t}{1 + K_0 \sum_{t=2}^{n+1} |u(F^{\xi}_{t-1})|^{\alpha/(\alpha-1)}} \tag{2.5.2}$$

where is $\vartheta = \min(\alpha, \alpha(\alpha\text{-}1))$, and $\xi_0 = 1$. Using the same argument as in Theorem 2.4.2, it follows that the second component in the denominator of equation (2.5.2) converges to infinity and the numerator converges to zero and hence the result.

2.6. Optimal estimation for the AR process

Consider a p^{th} order autoregressive (AR(p)) model satisfying the difference equation

$$\xi_t = \sum_{i=1}^{p} \phi_i \xi_{t-i} + \eta_t, \quad t \in \{0,1,2,...\} \tag{2.6.1}$$

where $\{\eta_t;\ t = 1,2,..,n\}$ is a sequence of independent and identically distributed (iid) random variables defined in the domain of attraction of the stable law with characteristic exponent α and $\{\phi_i\ ;\ i=1,2,..,p\}$ is the set of model coefficients satisfying stationarity conditions. When η_t has finite variance, the unique stationary solution of (2.6.1) is written as

$$\xi_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}\,,\ \psi_0 = 1. \tag{2.6.2}$$

In the case when $\{\eta_t\}$ is a sequence of iid infinite variance stable random variables, the stationarity condition on the coefficients as well as the restriction that

$$\sum_{k=0}^{\infty} |\psi_i|^{\alpha} < \infty \tag{2.6.3}$$

guarantees the existence of a unique stationary solution to the equation given in equation (2.6.1) and this is the same solution as that in (2.6.2).

In this section, we consider the stationary AR(p) process which satisfies condition (2.6.3). Suppose the observations $(\xi_1, \xi_2, ..., \xi_n)$ are available from an AR(p) process $\{\xi_t\}$ given in (2.6.1). A common approach for the estimation of the model parameters $\{\phi_i,\ i = 1, 2, ..., p\}$ is to minimize the sum of the squared errors, i.e., minimize the function

$$H_n(\phi) = \sum_{t=p+1}^{n} [\xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i}]^2 . \qquad (2.6.4)$$

This leads to the least squares estimators for the parameters. Properties of the resulting estimates have been studied by for example Yohai and Maronna (1977) and Hannan and Kanter (1977). They show that if the iid sequence $\{\eta_t\}$ has an infinite variance stable distribution with exponent α, then as $n \to \infty$, the least squares estimate of $\Phi = (\phi_1, \phi_2, ..., \phi_p)^T$ converges almost surely.

Another procedure for estimating $\{\phi_i ; i=1,2,..,p\}$ is the screened ratio estimate $\lambda(i)$ (Kanter and Steiger, 1974) which is defined from the relation

$$E\,(\xi_{n+1}|\xi_n) = \lambda(i)\xi_n$$

when $E\,|\xi_n| < \infty$ and $E\,(\xi_{n+1}|\xi_n) = \lambda\xi_n$ almost surely.

It is to be noted that the least squares procedure is particularly appropriate for the Gaussian case whereby the second order properties are clearly defined and is thus optimal; by the Gauss-Markov Theorem. In the stochastic set up with finite variance such as for the AR(p) process, the conditional least squares estimate is obtained and this, as shown in Godambe (1985), leads to an estimating function which is inferior to the optimal estimating function. It is thus natural to consider the extension of Godambe's ideas to the infinite variance case. We thus apply the theory developed in the earlier sections to obtain optimal estimates for the parameters of the AR(p) model.

Consider the AR(p) model given in (2.6.1). It is well known that when the characteristic exponent α of η_t is greater than one, the first moments exist. The conditional expectations for the process ξ_t are thus obtained using the results in section 2.2 as follows.

$$E\,(\xi_t | F^{\xi}_{t-1}) = \sum_{i=1}^{p} \phi_i \xi_{t-i}\,. \tag{2.6.5}$$

This implies that the set of variables h_t which are obtained as

$$h_t = \xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i}$$

are zero centered and mutually independent when η_t is zero centred. We can thus consider a class of vector valued unbiased estimating functions $\mathbf{G_n}$ given by

$$\mathbf{G_n} = \{\mathbf{g_n}: \mathbf{g_n} = \sum_{t=2}^{n} \mathbf{a}_{t-1} h_t\}.$$

Following Theorem 2.3.1, the optimal estimating function is obtained as

$$\mathbf{g}^0_n = \sum_{t=p+1}^{n} \mathbf{a}^0_{t-1} h_t$$

where $\mathbf{a}^0_{t-1}$ is a vector such that the i^{th} element is given by

$$a^0_{t-i} = |\xi_{t-i}/c(t)|^{1/(\alpha-1)} \mathrm{sgn}(\xi_{t-i}).$$

This leads to the following optimal equations in the vector $\mathbf{g}^0_n$:

$$g^0_n(j) = \sum_{t=p+1}^{n} \left|\frac{\xi_{t-j}}{c(t)}\right|^{1/(\alpha-1)} \left(\xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i}\right) \mathrm{sgn}(\xi_{t-j})$$

for $j = 1, 2, ..., p$. The estimating function vector $\mathbf{g}_n^0 = \mathbf{0}$ is then solved for the parameters. When $c(t) = 1$, the optimal estimates are evaluated as follows:

Let $\Phi^o = (\phi_1^o, \phi_2^o, ..., \phi_p^o)^T$ be the vector of parameters and define the transpose of the matrix Λ_n and the vector $\mathbf{D}_n$ respectively as

$$\Lambda_n^T = \begin{bmatrix} \sum_{t=p+1}^{n} \left|\frac{\xi_{t-1}}{c(t)}\right|^{\alpha/(\alpha-1)} & \sum_{t=p+1}^{n} \left|\frac{\xi_{t-2}}{c(t)}\right|^{1/(\alpha-1)} \xi_{t-1} & \cdot & \cdot & \cdot & \sum_{t=p+1}^{n} \left|\frac{\xi_{t-p}}{c(t)}\right|^{1/(\alpha-1)} \xi_{t-1} \\ \sum_{t=p+1}^{n} \left|\frac{\xi_{t-1}}{c(t)}\right|^{1/(\alpha-1)} \xi_{t-2} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum_{t=p+1}^{n} \left|\frac{\xi_{t-1}}{c(t)}\right|^{1/(\alpha-1)} \xi_{t-p} & \sum_{t=p+1}^{n} \left|\frac{\xi_{t-2}}{c(t)}\right|^{1/(\alpha-1)} \xi_{t-p} & \cdot & \cdot & \cdot & \sum_{t=p+1}^{n} \left|\frac{\xi_{t-p}}{c(t)}\right|^{\alpha/(\alpha-1)} \end{bmatrix}$$

and

$$\mathbf{D}_n = \left[\left(\sum_{t=p+1}^{n} \left|\frac{\xi_{t-1}}{c(t)}\right|^{1/(\alpha-1)} \xi_t, \sum_{t=p+1}^{n} \left|\frac{\xi_{t-2}}{c(t)}\right|^{1/(\alpha-1)} \xi_t, ..., \sum_{t=p+1}^{n} \left|\frac{\xi_{t-p}}{c(t)}\right|^{1/(\alpha-1)} \xi_t \right) \right]^T.$$

The optimal estimator for the set of coefficients in Φ is then obtained as

$$\Phi^o = \Lambda_n^{-1} \mathbf{D}_n. \tag{2.6.6}$$

Note that when $\alpha = 2$ and $c(t)=1$, this reduces to the conditional least squares estimate. It is also easy to see that when $p=1$, an optimal estimate of the parameter for an AR(1) model is obtained as

$$\phi^o = \frac{\sum_{t=2}^{n} |\xi_{t-1}/c(t)|^{1/(\alpha-1)} \xi_t \operatorname{sgn}(\xi_{t-1})}{\sum_{t=2}^{n} |\xi_{t-1}/c(t)|^{\alpha/(\alpha-1)} \operatorname{sgn}(\xi_{t-1})} \tag{2.6.7}$$

Recursive optimal estimates

Observations obtained recursively require that estimates be based on previous estimates and current observation. In this section, a recursive form of the optimal estimate for the AR(1) model is established.

Suppose n observations are available from the one parameter AR process ξ_t. Let the optimal estimate of ϕ based on the n observations be ϕ_n^o. When the $(n+1)^{th}$ observation becomes available, the optimal estimate based on ϕ_n^o and the new observation ξ_{n+1} is evaluated as follows:

$$\phi_{n+1}^o = \phi_n^o + \frac{K_n|\xi_n|^{1/(\alpha-1)}\,[\xi_{n+1} - \phi_n^o\xi_n]}{1 + K_n|\xi_n|^{\alpha/(\alpha-1)}}$$

$$= \phi_n^o + \frac{K_n|\xi_n|^{1/(\alpha-1)}\eta_{n+1}}{1 + K_n|\xi_n|^{\alpha/(\alpha-1)}} \qquad (2.6.8)$$

where

$$K_n^{-1} = \sum_{i=1}^{n-1}|\xi_i|^{\alpha/(\alpha-1)}\ , \quad \eta_{n+1} = \xi_{n+1} - \phi_n^o\xi_n$$

and

$$K_{n+1} = \frac{K_n}{1 + K_n|\xi_n|^{\alpha/(\alpha-1)}}$$

The convergence of of the recursive estimate is proved in the following Theorem.

Theorem 2.6.2

Suppose the sequence $\{\eta_t\}$ is of symmetric stable random variables. Then the estimate in (2.6.8) is strongly consistent.

Proof:

The value of K_{n+1} can be written in the form

$$K_{n+1} = \frac{K_0}{1 + K_o \sum_{i=1}^{n+1} |\xi_i|^{\alpha/(\alpha-1)}}$$

where $K_0 = 1$ and thus the result in (2.6.8) can be obtained as

$$\phi^o_{n+1} = \phi^o_0 + \frac{K_0 \sum_{j=2}^{n+1} |\xi_{j-1}|^{1/(\alpha-1)} \eta_j}{1 + K_0 \sum_{j=2}^{n+1} |\xi_{j-1}|^{\alpha/(\alpha-1)}}$$

$$= \beta + \frac{K_0 \sum_{j=2}^{n+1} |\xi_{j-1}|^{1/(\alpha-1)} \eta_j}{1 + K_0 \sum_{j=2}^{n+1} |\xi_i|^{\alpha/(\alpha-1)}} \qquad (2.6.9)$$

where $\xi_0 = 1$. Now since $1/(\alpha-1) > \alpha$, it follows that $|\xi_{j-1}|^{1/(\alpha-1)}$ is in the domain of attraction of the stable law with index $\alpha(\alpha-1)$ and is thus bounded (Logan et al. 1973). We also note that ξ_{j-1} and η_j are independent and thus $|\xi_{t-1}|^{1/(\alpha-1)}\eta_t$ is stable with index $\vartheta = \min(\alpha, \alpha(\alpha-1))$ and by Lemma 2.4.1,

$$\lim_{n \to \infty} n^{-1/\vartheta} \sum_{t=2}^{n} |\xi_{t-1}|^{1/(\alpha-1)} \eta_t = 0 \qquad \text{almost surely.}$$

Similarly by the same arguments, the denominator of equation (2.6.9) converges to infinity, i.e., for $\alpha^{-1} < k < (\alpha - 1)^{-1}$,

$$\lim_{n \to \infty} n^{-1/\kappa} \sum_{t=2}^{n} |\xi_{t-1}|^{\alpha/(\alpha-1)} = \infty \qquad \text{almost surely}$$

Note that if $\{\eta_t\}$ is not zero centered then we can use $\eta_t^* = \eta_t - E(\eta_t)$ since $E(\eta_t)$ exists when $\alpha > 1$.

2.6.1. Simulation results

In this section, we give some results for the optimal estimates and least squares estimates based on simulated observations from the AR(1) model. Observations for the AR(1) model were based on infinite variance stable random variable having characteristic exponent $\alpha \in [1.1,1.9]$. The Fortran subroutine GGSTA was used to generate the stable random variables. The first 200 observations were discarded. Values of the model parameters $\phi \in (0.2,0.9)$ are used for samples of sizes n = 50, 150, 500 and 1000. Estimates of the parameter were obtained as the mean of either 100 or 500 simulated samples for each of the sample sizes. The simulation results are given in Tables 2.1- 2.7.

Conclusions that emerge from these simulation results are as follows. First, it is observed that there is convergence of both the optimal and the least squares estimate as seen from their behavior when the sample size increases. It is also observed that optimal estimate is most suitable when the parameter is small, α is small and the sample size is small. The least squares approach overestimates the parameter when α and ϕ are small. For larger values of α, none of the methods has an advantage over the other.

The use of higher or a lower values of the exponent than the true one in the estimation leads respectively to overestimation or underestimation of the parameter. Change in the number of samples from 500 to 100 has no evident effect on the optimal estimates as well as the least squares estimates. The mean square deviation of the estimates are also given.

Table 2.1: Optimal and Least squares estimates for the parameter $\phi = 0.1$ of an AR(1) model when the number of simulations N=100.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.07817 (4.763E-04)	0.09755 (5.999E-06)	0.09640 (1.296E-05)	0.09643 (1.275E-05)
	1.6	0.1278 (7.732E-04)	0.0978 (4.795E-06)	0.0823 (3.131E-04)	0.0922 (6.034E-05)
Least Squares Estimate	1.1	0.1692 (4.784E-03)	0.1049 (2.353E-05)	0.09384 (3.796E-05)	0.09708 (8.528E-06)
	1.6	0.1585 (3.421E-03)	0.1222 (4.941E-04)	0.0794 (4.233E-04)	0.0892 (1.171E-04)

Table 2.2: Optimal and Least squares estimates for the parameter $\phi = 0.3$ of an AR(1) model when the number of simulations N=500.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.2	0.2728 (7.396E-04)	0.2953 (2.251E-05)	0.2952 (2.278E-05)	0.2959 (1.680E-05)
	1.4	0.2756 (5.946E-04)	0.2922 (6.105E-05)	0.2926 (5.447E-05)	0.2970 (8.991E-06)
	1.8	0.3548 (3.004E-03)	0.3066 (4.381E-05)	0.2779 (4.857E-04)	0.2942 (3.347E-05)
Least Squares Estimate	1.2	0.3668 (4.459E-03)	0.3092 (8.412E-05)	0.2949 (2.565E-05)	0.2976 (5.565E-06)
	1.4	0.3655 (4.296E-03)	0.3162 (2.625E-04)	0.2889 (1.234E-04)	0.2947 (2.777E-05)
	1.8	0.3493 (2.436E-03)	0.3111 (1.245E-04)	0.2807 (3.706E-04)	0.2924 (5.798E-05)

Table 2.3: Optimal and Least squares estimates for the parameter $\phi = 0.6$ of an AR(1) model when the number of simulations N=500.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.5799 (4.053E-04)	0.5976 (5.658E-06)	0.5966 (1.130E-05)	0.5966 (1.114E-05)
	1.9	0.6307 (9.465E-04)	0.5742 (6.626E-04)	0.5952 (2.305E-05)	0.6077 (5.958E-05)
Least Squares	1.1	0.6392 (1.538E-03)	0.6053 (2.837E-05)	0.5991 (7.883E-07)	0.5981 (3.573E-06)
	1.9	0.6309 (9.560E-04)	0.5749 (6.317E-04)	0.5967 (1.107E-05)	0.6070 (4.897E-05)

Table 2.4: Optimal and Least squares estimates for the parameter $\phi = 0.8$ of an AR(1) model when the number of simulations N=500.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.7746 (6.447E-04)	0.7988 (1.341E-06)	0.8007 (5.425E-05)	0.8007 (5.629E-05)
	1.4	0.7608 (1.536E-03)	0.8055 (3.018E-05)	0.8043 (1.901E-05)	0.8070 (4.968E-05)
	1.9	0.8205 (4.205E-04)	0.7738 (6.854E-04)	0.8058 (3.395E-05)	0.8142 (2.026E-04)
Least Squares Estimate	1.1	0.7934 (4.308E-05)	0.8039 (1.517E-05)	0.7997 (1.107E-07)	0.7983 (2.772E-06)
	1.4	0.7974 (6.568E-06)	0.7943 (3.279E-05)	0.7960 (1.600E-05)	0.7984 (2.485E-06)
	1.9	0.8245 (6.018E-04)	0.7758 (5.852E-04)	0.8087 (7.507E-05)	0.8148 (2.187E-04)

Table 2.5: Optimal estimates for the parameter $\phi = 0.1$ of an MA(1) model when the true value of $\alpha = 1.2$.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.2	0.0705 (8.721E-04)	0.0952 (2.280E-05)	0.0951 (2.371E-05)	0.0958 (1.756E-05)
	1.5	0.0993 (4.499E-07)	0.0976 (5.733E-06)	0.0953 (2.164E-05)	0.0981 (3.408E-06)

Table 2.6: Optimal estimates for the parameter $\phi = 0.8$ of an MA(1) model when the true value of $\alpha = 1.9$.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.2	0.8136 (1.864E-04)	0.7846 (2.357E-04)	0.7627 (1.388E-03)	0.7805 (3.813E-04)
	1.9	0.8205 (4.205E-04)	0.7738 (6.854E-04)	0.8058 (3.395E-05)	0.8142 (2.026E-04)

Table 2.7: Optimal estimates for the parameter $\phi = 0.1$ of an MA(1) model when the true value of $\alpha = 1.4$.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.2	0.0524 (2.263E-03)	0.0858 (2.029E-04)	0.0906 (8.868E-05)	0.0918 (6.671E-05)
	1.4	0.0758 (5.847E-04)	0.0911 (7.846E-05)	0.0906 (8.728E-05)	0.0950 (2.482E-05)

2.7. Optimal estimation for the moving average process

Let $\{\xi_t\}$ be a moving average process of order q (MA(q)) defined by the following difference equation

$$\xi_t = \sum_{i=1}^{q} \theta_i \eta_{t-i} + \eta_t, \quad t = 1,2,3,... \tag{2.7.1}$$

where $\{\eta_t\}$ is the innovation sequence of iid symmetric stable random variables having characteristic exponent $\alpha \in (1,2]$. We assume that the process ξ_t is invertible, i.e, there exists constants π_i such that the MA(q) process $\{\xi_t\}$ has the following representation

$$\eta_t = \sum_{i=0}^{\infty} \pi_i(\theta) \xi_{t-i}, \quad t = 1,2,....$$

The coefficients $\{\pi_i\}$ in the above representation can be uniquely obtained by the following power series expansion:

$$[\Theta(z)]^{-1} = \sum_{i=0}^{\infty} \pi_i(\theta) z^i$$

where $\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + ... + \theta_q z^q = 0$ for all z such that $|z| > 1$ (the invertibility condition holds). A set of independent random variables can thus be obtained as

$$h_t = \xi_t + \sum_{i=1}^{\infty} \pi_i(\theta) \xi_{t-i}$$

Following the optimality criteria stipulated in Section 2.3, we consider the class of vector valued unbiased estimating functions

$$G_n = \{g_n: g_n = \sum_{t=k}^{n} a_{t-1} h_t\}$$

(for large k) from which the optimal estimating function is obtained as

$$g^0 = \sum_{t=k}^{n} a^0_{t-1} h_t .$$

The vector $\mathbf{a}^0_{t-1}$ is obtained such that each component is of the form

$$a^0_{t-i} = \left| \frac{E[(\partial h_t/\partial \theta_i) | F^{\xi}_{t-1}]}{\mathrm{disp}(h_t | F^{\xi}_{t-1})} \right|^{1/(\alpha-1)} \mathrm{sgn}(E[(\partial h_t/\partial \theta_i) | F^{\xi}_{t-1}])$$

$$= |\pi^{(1)}_i(\theta)/c(t)|^{1/(\alpha-1)} \mathrm{sgn}(\pi^{(1)}_i(\theta))$$

where $\pi^{(1)}_i(\theta)$ is the derivative of $\pi_i(\theta)$ with respect to θ_i. This leads to the set of optimal functions $g^0_n(j)$, j = 1,2,..q where

$$g^0_n(j) = \sum_{t=k}^{n} \left[|\pi^{(1)}_j(\theta)/c(t)|^{1/(\alpha-1)} (\xi_t + \sum_{i=1}^{\infty} \pi_i(\theta)\xi_{t-i}) \right] \mathrm{sgn}(\pi^{(1)}_j(\theta)).$$

The coefficients $\{\pi_j(\theta), j=1,2,...,q\}$ are then evaluated from $\mathbf{g}^0_n = \mathbf{0}$. Unfortunately this is a complicated nonlinear function of θ and it is difficult to obtain an analytical form for the estimate of θ_i from it . This procedure is illustrated in the following example for the MA(1) model

Example: Optimal estimate for the MA(1) process

Consider an MA(1) process ξ_t give by the equation

$$\xi_t = \theta\eta_{t-1} + \eta_t$$

with $\{\eta_t\}$ as specified for equation (2.7.1). In this case $\pi_i(\theta) = \theta^i$ and hence

$$h_t = \xi_t + \sum_{i=1}^{\infty}\theta^i\xi_{t-i}$$

and the value for a^0_{t-1} is then obtained as

$$a^0_{t-1} = \left|\sum_{i=1}^{\infty} i\theta^{i-1}\xi_{t-i}/c(t)\right|^{1/(\alpha-1)} \operatorname{sgn}(\xi_{t-i})$$

$$\cong \left|\sum_{i=1}^{n} i\theta^{i-1}\xi_{t-i}/c(t)\right|^{1/(\alpha-1)} \operatorname{sgn}(\xi_{t-i}) \qquad \text{(for large n).}$$

The optimal estimating function g_n can thus be written as

$$g^0_n = \sum_{t=k}^{n}\left[\left|\sum_{i=1}^{\infty} i\theta^{i-1}\xi_{t-i}/c(t)\right|^{1/(\alpha-1)}\left(\xi_t + \sum_{i=1}^{\infty}\theta^i\xi_{t-i}\right)\right] \operatorname{sgn}(\xi_{t-i})$$

and the optimal estimate of θ is obtained by solving $g^0_n = 0$. This is relatively difficult to solve even for $i = 1,2$ and thus a simpler alternative approach has to be developed for practical purposes.

In the finite variance case, nonlinear techniques are usually used to evaluate the coefficients of the MA(q) process. A relatively more feasible approach for the infinite variance setting is proposed for the MA(q) process based on the following steps:

Step 1: Evaluate the least squares estimate $\hat{\xi}_t(L)$ of ξ_t from the set of observations $\xi_1, \xi_2,$ .., ξ_n as a preliminary value. This is achieved by first obtaining the yule walker estimate for $\theta = (\theta_1, \theta_2, ..., \theta_q)$ and then substituting it into the model equation to obtain

$$\hat{\xi}_t(Y) = \sum_{i=1}^{q} \hat{\theta}_i \eta_{t-i}.$$

Step 2: Use the result in step 1 to obtain the preliminary innovations recursively as

$$\hat{\eta}_t(L) = \xi_t - \hat{\xi}_t(L)\text{ , where } \eta_o = 0.$$

Step 3: Evaluate the optimal estimates of the parameter θ by letting

$$h_t = \xi_t - \sum_{i=1}^{q} \theta_i \hat{\eta}_{t-i}(L).$$

This leads to the optimal estimating function vector

$$g_n^0 = \sum_{t=q+1}^{n} a_{t-1}^0 h_t$$

where the vector $\mathbf{a}_{t-1}^0$ has components given by

$$a_{t-i}^0 = |\hat{\eta}_{t-i}(L))/c(t)|^{1/(\alpha-1)}\, \mathrm{sgn}(\hat{\eta}_{t-i}(L)) \quad \text{and} \quad i = 1,2,\ldots..,q.$$

Let the i,j component of the matrix $\mathbf{B}_n$ be

$$bij = \sum_{t=q+1}^{n} \left|\frac{\hat{\eta}_{t-i}(L)}{c(t)}\right|^{1/(\alpha-1)} \hat{\eta}_{t-j}(L))$$

and suppose the vector $\mathbf{D}_n$ is given by

$$\mathbf{D}_n = \left[\left(\sum_{t=q+1}^{n} \left|\frac{\hat{\eta}_{t-1}(L)}{c(t)}\right|^{1/(\alpha-1)} \xi_t,\ \sum_{t=q+1}^{n} \left|\frac{\hat{\eta}_{t-2}(L)}{c(t)}\right|^{1/(\alpha-1)} \xi_t, \ldots,\ \sum_{t=q+1}^{n} \left|\frac{\hat{\eta}_{t-q}(L)}{c(t)}\right|^{1/(\alpha-1)} \xi_t \right) \right]^T$$

The optimal estimator for the set of coefficients in Φ is then obtained using Lemma 2.4.1 as

$$\Theta^o = \mathbf{B}_n^{-1}\,\mathbf{D}_n. \tag{2.6.6}$$

Note that when $\alpha = 2$ and c(t)=1, this reduces to the conditional least squares estimate. It is also easy to see that when q=1, an optimal estimate of the parameter for an MA(1) model is obtained as

$$\hat{\theta}^o = \frac{\sum_{t=2}^{n} |\hat{\eta}_{t-1}(L)/c(t)|^{1/(\alpha-1)} \xi_t \, \mathrm{sgn}(\hat{\eta}_{t-1}(L))}{\sum_{t=2}^{n} c(t) |\hat{\eta}_{t-1}(L)/c(t)|^{\alpha/(\alpha-1)}}$$

Step 4: Using the optimal estimate Θ^0 obtained in Step 3, evaluate the optimal estimate $\hat{\xi}_t(o)$ of ξ_t as

$$\hat{\xi}_t(o) = \sum_{i=1}^{q} \hat{\theta}_i^o \, \hat{\eta}_{t-i}(L)$$

This is then used to obtain the optimal estimate of the innovations $\{\eta_t,\ t = 1, 2, .., n\}$ as $\hat{\eta}_t(o) = \xi_t - \hat{\xi}_t(o)$. Step 3 is then repeated using $\hat{\eta}_t(o)$ and a more refined estimate of Θ is obtained.

Simulation results

The following results based on the steps outlined above were obtained by simulating values for the MA(1) process using the Fortran subroutine GGSTA for different values of the parameter θ. Each estimate was obtained as a mean value based on 500 samples of size 50, 150, 500 and 1000. The first 200 values in the simulation were deleted. Results were similarly obtained for different values of the exponent α and these are given in Tables 2.8 - 2.10.

The superiority of the optimal estimate over the least squares estimate is quite evident particularly for high values of θ and for low sample sizes. Convergence of the least squares estimate is very slow especially when θ is large.

When the value of the index α is incorrectly specified at the estimation stage, the parameter is either overestimated or underestimated by the optimal estimate depending on whether the incorrect α is respectively higher or lower than the true α. This is shown in Tables 2.11 - 2.13 when $\theta = 0.1$, 0.5 and 0.8.

Table 2.8: Optimal and Least squares estimates for the parameter $\theta = 0.2$ of an MA(1) model when number of samples N=500.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.1848 (2.309E-04)	0.1975 (5.996E-06)	0.1964 (1.260E-05)	0.1965 (1.240E-05)
	1.4	0.1862 (1.902E-04)	0.1915 (7.264E-05)	0.1907 (8.669E-05)	0.1950 (2.500E-05)
	1.8	0.2855 (7.308E-03)	0.2211 (4.472E-04)	0.1788 (4.474E-04)	0.1887 (1.280E-04)
Least Squares Estimate	1.1	0.2549 (3.012E-03)	0.1975 (6.361E-06)	0.1886 (1.304E-04)	0.1913 (7.489E-05)
	1.4	0.2557 (3.113E-03)	0.2101 (1.030E-04)	0.1810 (3.618E-04)	0.1877 (1.516E-04)
	1.8	0.2382 (1.459E-03)	0.2095 (9.015E-05)	0.1749 (6.302E-04)	0.1845 (2.413E-04)

Table 2.9: Optimal and Least squares estimates for the parameter $\theta = 0.5$ of an MA(1) model

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.2	0.4788 (4.475E-04)	0.4957 (2.034E-05)	0.4961 (1.551E-05)	0.4967 (1.060E-05)
	1.5	0.5243 (5.887E-04)	0.4937 (3.965E-05)	0.4884 (1.340E-04)	0.4952 (2.278E-05)
	1.8	0.5898 (8.059E-03)	0.5229 (5.260E-04)	0.4807 (3.736E-04)	0.4903 (9.366E-05)
Least Squares Estimate	1.2	0.4475 (3.045E-03)	0.4090 (8.280E-03)	0.4013 (9.749E-03)	0.4033 (9.353E-03)
	1.5	0.4501 (2.485E-03)	0.4223 (6.031E-03)	0.3953 (1.096E-02)	0.4014 (9.703E-03)
	1.8	0.4572 (1.827E-03)	0.4238 (5.800E-03)	0.3941 (1.120E-02)	0.4026 (9.483E-03)

Table 2.10: Optimal and Least squares estimates for the parameter $\theta = 0.8$ of an MA(1) model.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.7581 (1.755E-03)	0.7992 (6.815E-07)	0.8005 (2.841E-07)	0.8006 (3.197E-07)
	1.5	0.7688 (9.738E-04)	0.7815 (3.410E-04)	0.7840 (2.570E-04)	0.7886 (1.288E-04)
	1.9	0.8045 (1.996E-05)	0.7532 (2.181E-03)	0.7285 (5.105E-03)	0.7303 (4.852E-03)
Least Squares Estimate	1.1	0.5205 (0.0781)	0.4933 (0.0940)	0.4931 (0.0942)	0.4933 (0.0941)
	1.5	0.5313 (0.0722)	0.5092 (0.0845)	0.4883 (0.0972)	0.4931 (0.0942)
	1.9	0.5593 (0.0579)	0.5080 (0.0852)	0.4958 (0.0925)	0.4994 (0.0903)

Table 2.11: Optimal estimates for the parameter $\theta = 0.2$ of an MA(1) model when the true value of $\alpha = 1.6$.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.3	0.1559 (1.947E-03)	0.1770 (5.274E-04)	0.1832 (2.819E-04)	0.1883 (1.363E-04)
	1.6	0.2562 (3.158E-03)	0.2019 (3.471E-06)	0.1825 (3.057E-04)	0.1917 (6.818E-05)
	1.9	0.2898 (8.071E-03)	0.2229 (5.246E-04)	0.1803 (3.860E-04)	0.1897 (1.068E-04)

Table 2.12: Optimal estimates for the parameter $\theta = 0.5$ of an MA(1) model when the true value of $\alpha = 1.5$.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.4607 (1.547E-03)	0.4742 (6.628E-04)	0.4891 (1.182E-04)	0.4892 (1.163E-04)
	1.5	0.5243 (5.887E-04)	0.4937 (3.965E-04)	0.4884 (1.340E-04)	0.4952 (2.278E-05)
	1.9	0.5927 (8.599E-03)	0.5208 (4.344E-04)	0.4837 (2.650E-04)	0.4915 (7.152E-05)

Table 2.13: Optimal estimates for the parameter $\theta = 0.8$ of an MA(1) model when the true value of $\alpha = 1.4$.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.7306 (4.818E-03)	0.7884 (1.341E-04)	0.7991 (7.214E-07)	0.7992 (6.078E-07)
	1.4	0.7507 (2.431E-03)	0.7896 (1.069E-04)	0.7932 (4.542E-05)	0.7970 (8.687E-06)
	1.8	0.8051 (2.596E-05)	0.7830 (2.875E-04)	0.7637 (1.314E-03)	0.7640 (1.295E-03)

2.8. Optimal estimates for the ARMA process

Consider the process $\{\xi_t\}$ from the autoregressive moving average process (ARMA(p,q)) satisfying the difference equation

$$\Phi(B)\xi_t = \Theta(B)\eta_t$$

where $\Phi(B)$ and $\Theta(B)$ are the autoregressive and moving average operators satisfying the stationarity and invertibility conditions respectively while $\{\eta_t\}$ is a sequence of iid symmetric stable random variables. Under these conditions, the process ξ_t has the moving average and the autoregressive representations respectively as

$$\xi_t = \sum_{j=0}^{\infty}\psi_j(\theta,\phi)\eta_{t-j} \tag{2.8.1}$$

$$\xi_t = \sum_{k=1}^{\infty}\pi_k(\theta,\phi)\xi_{t-k} + \eta_t \tag{2.8.2}$$

where $\{\psi_j(\theta,\phi)\}$ and $\{\pi_k(\theta,\phi)\}$ are the coefficients generated from the power series expansion of $\Theta(B)/\Phi(B)$ and $\Phi(B)/\Theta(B)$ respectively. A unique stationary solution to the ARMA process is obtained as in equation (2.8.1).

When η_t has finite variance, the parameter estimates for the ARMA(p,q) model are usually obtained using nonlinear techniques or by solving the Yule-Walker equations. For the infinite variance case, we consider estimating the model parameters using an optimal estimating function. The sequence of independent random variables h_t can now be obtained as

$$h_t = \xi_t - \sum_{i=1}^{\infty}\pi_i(\theta,\phi)\xi_{t-i}$$

Let $\beta = (\beta_1, \beta_2, ..\beta_{p+q})$ be the set of parameters. In the class G of estimating functions, the optimal estimate of the vector $\mathbf{a}_{t-1}$ has components

$$a^0_{t-i} = |\pi^{(1)}_i(\beta)/c(t)|^{1/(\alpha-1)} \, \mathrm{sgn}(\pi^{(1)}_i(\beta))$$

where $\pi_i(\theta,\phi)$ is written for simplicity as $\pi_i(\beta)$ and $\pi^{(1)}_i(\beta)$ is the derivative of $\pi_i(\beta)$ with respect to β_i. This leads to the f optimal estimating function vector $\mathbf{g}^0_n$ with components

$$g^0_n(i) = \sum_{t=1}^{n} \left[|\pi^{(1)}_i(\beta)/c(t)|^{1/(\alpha-1)}(\xi_t + \sum_{i=1}^{\infty}\pi_i(\beta)\xi_{t-i}) \right] \mathrm{sgn}(\pi^{(1)}_i(\beta)) \tag{2.8.4}$$

where i = 1, 2, ..., p+q.

Solving $\mathbf{g}^0_n = \mathbf{0}$ for the parameters leads to the optimal estimates of $\{\theta_i\}$ and $\{\phi_i\}$ being obtained. As in the moving average case, this is difficult to solve because of the nonlinearity of the model coefficients in $\pi(\beta)$. An alternative procedure would be to follow the steps outlined in Section 2.7. In Step 4, Θ^o is now replaced by the optimal estimate of β which is a function of θ and ϕ.

CHAPTER THREE

3. ESTIMATION OF MISSING OBSERVATIONS

3.1. Introduction

Irregularly observed data arises quite often in time series and this is partly due to errors in data collection or because of misplaced data. In the case of finite variance sequences, irregularities could also arise as a result of deleted values which in most cases are outliers. This is not the case when analysing infinite variance time series since they are necessarily characterized by bursts of outlying observations. The presence of irregularities in observations usually leads to inaccuracies when analysing the data using procedures tailored for regularly observed data. The need for accurate estimation of missing values in irregularly observed time series is thus a necessity and this has led to a growing number of techniques for the estimation of missing values for linear as well as nonlinear processes (Brockwell and Davis, 1987; Miller and Ferreiro, 1984; Abraham and Thavaneswaran, 1991). The diversity of the studies in Parzen (1984) point to some of the problems associated with irregular observations. The technique developed by Pourahmadi (1989) also provides an alternative approach by which one can obtain estimates of missing values.

One method in the literature on estimation of missing values assumes that the missing values are parameters and uses the least squares criteria to obtain estimates for them. A second method is based on the assumption that the missing values are random variables and the estimates are then obtained by prediction algorithms (See Abraham and Thavaneswaran, 1991). The analysis of processes with infinite variance has been receiving a lot of attention as situations which are more adequately represented by such processes are recognized. For time series having infinite variance stable errors, neither the least squares method nor the prediction algorithm may be appropriate to estimate missing values. From the current literature, the estimation of missing values has unfortunately been restricted to processes with finite variance except in cases where general formulations that might not be of much practical use have been given (Pourahmadi, 1984). There is thus the need to establish approaches (which are optimal in some sense) that lead to simple analytical representations for estimating missing values for the infinite variance case.

In this chapter, estimates of missing observations are obtained when the process is driven by symmetric stable innovations having characteristic exponent α ($1 < \alpha \leq 2$) are obtained. This is achieved using three possible techniques. First, we consider the missing value as a parameter and apply the optimality criterion for infinite variance stable processes developed in chapter two to estimate the missing values. This is discussed in Section 3.2. By making use of a projection theorem in the L^{λ} metric, a modified interpolation technique is attained and this is also used to obtain linear estimates of missing values in Section 3.3. A third estimation technique is discussed in Section 3.4 where a simple dispersion minimization procedure which is an analogue of the variance minimization leads to possible estimates of the missing values. This is based on the well known fact that the dispersion is proportional to the λ^{th} norm ($0 < \lambda \leq \alpha$) of the L^{λ} space and thus by minimizing the λ^{th} norm and considering the missing values as parameters, we obtain their estimates. In this final missing value estimation technique, two separate cases ($\lambda > 1$ or $\lambda \leq 1$) are considered.

3.2. Optimal estimation of missing observations

Among the procedures utilized for the estimation of missing values, treatment of missing observations as parameters was initially studied by Ferreiro (1987) in the context of linear models for the finite variance case. This was later extended to nonlinear models by Abraham and Thavaneswaran (1991). The case of missing observations for infinite variance time series models was recently studied by Thavaneswaran and Thompson (1991) by considering the missing value as a random variable and evaluating the conditional expectations.

In this Section, we consider the missing observation as a parameter and by making use of the optimality criterion for infinite variance stable processes, an optimal estimate of the missing value in the minimum dispersion sense is obtained. Illustrations from linear time series models are given. First, some results discussed in chapter two are recalled. It was shown that the optimal estimate of the parameter θ is obtained by solving

$$\sum_{t=1}^{n} a_{t-1}^{o} h_t = 0 \tag{3.2.1}$$

where $\{h_t\}$ is a sequence of innovations of the form

$$h_t = \xi_t - f(\theta, F_{t-1}^{\xi}))$$

while a_{t-1}^{o} is given by

$$a_{t-1}^{o} = \left| \frac{E[(\partial h_t/\partial\theta)|F_{t-1}^{\xi}]}{\mathrm{disp}(h_t|F_{t-1}^{\xi})} \right|^{1/(\alpha-1)} \mathrm{sgn}(E[(\partial h_t/\partial\theta)|F_{t-1}^{\xi}]).$$

This implies that when the missing observation ξ_m, denoted by β_m is to be optimally estimated, it should similarly satisfy (3.2.1). This is illustrated using examples for the single missing value as well as for two missing values where we now consider a_{t-1} and h_t as a function of β_m and F^{ξ}_{t-1}. We also let $c(t) = 1$ for simplicity.

Single missing value

Example 3.2.1: AR(1) model

Consider the AR(1) model given by $\xi_t = \phi\xi_{t-1} + \eta_t$ where ϕ is the model parameter and η_t is symmetric stable and having scale parameter $c(t) = 1$. Suppose that from a set of n possible observations $\{\xi_1, \xi_2, ..., \xi_m, ..., \xi_n\}$, the m^{th} value ($m < n$) is missing. Denoting it as the parameter β_m, it is clear that $h_t = \xi_t - \phi\xi_{t-1}$ and thus $a^o_{t-1} = 1 - |\phi|^{1/(\alpha-1)}$. Thus the optimal estimate of β_m is obtained from (3.2.1) by solving

$$(\beta_m - \phi\xi_{m-1}) - |\phi|^{1/(\alpha-1)}(\xi_{m+1} - \phi\beta_m) = 0.$$

This leads to the estimate

$$\beta^o_m = \frac{\phi\xi_{m-1} + |\phi|^{1/(\alpha-1)}\xi_{m+1}}{(1 + |\phi|^{\alpha/(\alpha-1)})}.$$

This reduces to a well known result for the finite variance case when $\alpha = 2$.

Example 3.2.2: AR(2) model

Suppose that the process ξ_t is given by the difference equation $\xi_t = \phi_1\xi_{t-1} + \phi_2\xi_{t-2} + \eta_t$ where ϕ_1 and ϕ_2 are the model parameters and η_t is as in example 3.2.1. From a set of n possible observations $\{\xi_1, \xi_2, ..., \xi_m, ..., \xi_n\}$, we suppose that the m^{th} value $(m < n)$ is missing and denoting it as the parameter β_m. Its optimal estimate is similarly obtained by solving

$$(\beta_m - \phi_1\xi_{m-1} - \phi_2\xi_{m-2}) - |\phi_1|^{1/(\alpha-1)}(\xi_{m+1} - \phi_1\beta_m - \phi_2\xi_{m-1})$$

$$- |\phi_2|^{1/(\alpha-1)}(\xi_{m+2} - \phi_1\xi_{m+1} - \phi_2\beta_m) = 0.$$

This leads to the following solution for β_m:

$$\beta_m^o = [(1 + |\phi_1|^{\alpha/(\alpha-1)} + |\phi_2|^{\alpha/(\alpha-1)})]^{-1}\Big[\phi_2\xi_{m-2} + (\phi_1 - \phi_2|\phi_1|^{1/(\alpha-1)})\xi_{m-1}$$

$$+(|\phi_1|^{1/(\alpha-1)} - \phi_1|\phi_2|^{1/(\alpha-1)})\xi_{m+1} - |\phi_2|^{1/(\alpha-1)}\xi_{m+2}\Big].$$

The case when $\alpha = 2$ leads to the estimated value as

$$\beta_m^o = \frac{\phi_2\xi_{m-2} + \phi_1(1 - \phi_2)(\xi_{m-1} + \xi_{m+1}) - \phi_2\xi_{m+2}}{[1 + \phi_1^2 + \phi_2^2]}.$$

A nonparametric approach of taking the average of ξ_{m-1} and ξ_{m+1} would lead to an estimate having a higher variance than the optimal estimate since it is not based on minimizing the scale parameter.

Two missing values

Example 3.2.3: AR(1) model

The AR(1) model given by $\xi_t = \phi\xi_{t-1} + \eta_t$ as in example 3.2.1 is considered. It is similarly assumed that from a set of n possible observations $\{\xi_1, \xi_2, ..., \xi_m, ..., \xi_{m+k}, ..., \xi_n\}$, the m^{th} value and the $(m + |k|)^{th}$ (where $|k| \geq 1$ and $0 < m + |k| < n$) values are missing. These are denoted as the parameters β_m and $\beta_{m+|k|}$. When $|k| = 1$, this reduces to the case of two consecutive values. For the above AR(1) model, we suppose that k=1and thus the observations ξ_m and ξ_{m+1} are missing. Using the optimality criterion as in the previous examples, the missing values are obtained by solving the following equations simultaneously:

$$(\beta_m - \phi\xi_{m-1}) - |\phi|^{1/(\alpha-1)}(\xi_{m+1} - \phi\beta_m) = 0$$

$$(\beta_{m+1} - \phi\beta_m) - |\phi|^{1/(\alpha-1)}(\xi_{m+2} - \phi\beta_{m+1}) = 0$$

The optimal estimates of β^o_m and β^o_{m+1} are then evaluated from the following equation:

$$\begin{pmatrix} \beta^o_m \\ \beta^o_{m+1} \end{pmatrix} = [(1 + |\phi|^{\alpha/(\alpha-1)})^2 - |\phi|^{(2\alpha-1)/(\alpha-1)}]^{-1}$$

$$\begin{bmatrix} 1 + |\phi|^{\alpha/(\alpha-1)} & |\phi|^{\alpha/(\alpha-1)} \\ \phi & 1 + |\phi|^{\alpha/(\alpha-1)} \end{bmatrix} \begin{bmatrix} \phi\xi_{m-1} \\ |\phi|^{1/(\alpha-1)}\xi_{m+2} \end{bmatrix}$$

and the optimal estimates are obtained as

$$\beta^o_m = \frac{(1 + |\phi|^{\alpha/(\alpha-1)})\phi\xi_{m-1} + |\phi|^{(\alpha+1)/(\alpha-1)}\xi_{m+2}}{[(1 + |\phi|^{\alpha/(\alpha-1)})^2 - |\phi|^{(2\alpha-1)/(\alpha-1)}]}$$

$$\beta^{o}_{m+1} = \frac{\phi^2\xi_{m-1} + (1 + |\phi|^{\alpha/(\alpha-1)})|\phi|^{1/(\alpha-1)}\xi_{m+2}}{[(1 + |\phi|^{\alpha/(\alpha-1)})^2 - |\phi|^{(2\alpha-1)/(\alpha-1)}]}\;.$$

When $|k| \geq 2$, this leads to nonconsecutive missing values which are estimated in a similar way to the single missing value case.

3.3. Linear estimation of missing values

In this Section, we start by recalling some of the results on the Wold and predictive decompositions of infinite variance stable processes. This is then followed by the development of linear estimation technique of missing observations.

Decomposition of stochastic processes

The concept of decomposition of stationary processes can be discussed in the context of the Wold and predictive decompositions. Such decompositions are easily developed for second order stationary processes using the orthogonality property of the innovation sequence (See for example Chapter 5 of Brockwell and Davis, 1987). The infinite variance processes on the other hand do not have the orthogonality property for the innovation sequence but by ascribing independence and finiteness of the p^{th} mean, Cambanis et al. (1988) established an analoque of the Wold decomposition while Miamee and Pourahmadi (1988) discussed the Wold and predictive decompositions.

Consider the stationary symmetric stable process $\{\xi_t : t = 0,1,2,..\}$ defined on the probability space (Ω,S,P) and driven by a sequence of iid stable random variables $\{\eta_t\}$ having characteristic index α $(1 < \alpha \leq 2)$ and constant scale parameter which for simplicity we can assume is one. Suppose $S_t = \bar{sp}\ \{\xi_s : s \leq t\}$ is a closed subspace in the norm of $L^{\alpha}(\Omega,S,P)$. If the stationary process $\{\xi_t\}$ is nondeterministic, then by the Wold decomposition Theorem, every such process can be decomposed as

$$\xi_t = W_t + \sum_{i=0}^{\infty} \beta_i \eta_{t-i}\,, \quad \beta_0 = 1 \tag{3.3.1}$$

where the scale parameter (dispersion) of ξ_t converges, i.e.,

$$\sum_{i=0}^{\infty} c|\beta_i|^{\alpha} < \infty.$$

The process W_t is stationary and contained in the S_t. A simple proof for this result is given in Theorem 5.2 of Miamee and Pourahmadi (1988).

In the forecasting case, it is well known that if S_t is defined in the Hilbert space where $\alpha = 2$, then for $h \geq 1$, the h-steps ahead linear forecast ξ_{t+h} based on the past values $\{\xi_t, \xi_{t-1}, \xi_{t-2}, ..., \xi_1\}$ is the orthogonal projection of ξ_{t+h} onto the subspace S_t such that the error variance $E(\xi_{t+h} - \hat{\xi}_{t+h})^2$ is a minimum. In the case of infinite variance stable processes having exponent $\alpha \in (0,2)$, a unique projection can similarly be defined (See Cline and Brockwell, 1985). From the decomposition given in (3.3.1), the future value ξ_{t+h} based on the infinite past is obtained as

$$\xi_{t+h} = W_{t+h} + \sum_{i=0}^{h-1} \beta_i \eta_{t+h-i} + \sum_{i=h}^{\infty} \beta_i \eta_{t+h-i}.$$

Using this innovation representation, the projection of ξ_{t+h} over the space S_t is evaluated as

$$P_{S_t}(\xi_{t+h}) = \hat{\xi}_{t+h} = W_{t+h} + \sum_{i=h}^{\infty} \beta_i \eta_{t+h-i} \tag{3.3.2}$$

where W_{t+h} is a function of $\{\xi_t, \xi_{t-1}, ...\}$. The forecast error is thus given by

$$\xi_{t+h} - \hat{\xi}_{t+h} = \sum_{i=0}^{h-1} \beta_i \eta_{t+h-i}. \tag{3.3.3}$$

An alternative representation of equations (3.3.2) and (3.3.3) as functions of the observations is given by Cline and Brockwell (1985). The corresponding forecast error dispersion is thus evaluated as

$$\mathrm{disp}(\xi_{t+h} - \hat{\xi}_{t+h}) = \sum_{i=0}^{h-1} c|\beta_i|^{\alpha}.$$

Suppose now that $\{\xi_t\}$ is an ARMA(p,q) process given by the equation

$$\Phi(B)\xi_t = \Theta(B)\eta_t$$

where $\Phi(B)$ and $\Theta(B)$ are the autoregressive (AR) and the moving average (MA) operators satisfying the stationarity and invertibility conditions. The sequence $\{\eta_t\}$ is of independent nonstationary stable random variables with characteristic exponent $\alpha \in (0,2)$. It is evident that under these conditions, the process ξ_t admits a moving average representation and a subsequent Wold decomposition as in (3.3.1).

Linear interpolation and estimation of missing values

Suppose we have one value ξ_m missing out of a set of an arbitrarily large number of possible observations n generated from an ARMA process ξ_t as given in the previous section. Let the subspace S_m^* be the allowable space of linear estimators of ξ_m based on the observed values $(\xi_t, \xi_{t-1}, .., \xi_{m+1}, \xi_{m-1}, .., \xi_1.)$ i.e., $S_m^* = \bar{sp}\ \{\xi_t\colon t \leq n, t \neq m\}$ where the sample size n is assumed large. The projection of ξ_m onto S_m^* (denoted $P_{S_m^*}^{\xi_m}$) such that the dispersion of the error of the estimate (which we write as $\mathrm{disp}(\xi_m - P_{S_m^*}^{\xi_m})$) is a minimum would thus simply be the minimum dispersion linear interpolator. Direct computation of the projection of ξ_m onto S_m^* is complicated since the subspaces $S_1 = \mathrm{sp}\{\xi_{m-1}, \xi_{m-2}, \ldots\}$ and S_m^* are not independent of each other. We thus consider evaluating the projection onto two disjoint subspaces of S_m^*. To achieve this, we express S_m^* as a direct sum of the subspace S_1 and another subspace, say S_* , such that $S_m^* = S_1 \oplus S_*$. A possible subspace is $S_* = sp\{\xi_i - \hat{\xi}_i : i \geq m+1\}$ where $\hat{\xi}_i$ is based on the values $\{\xi_{m-1}, \xi_{m-2}, \ldots\}$. The existence of the subspaces S_1 and S_* is shown in the following Lemma.

Lemma 3.3.1

Suppose $\{\xi_t\}$ is a nondeterministic stationary process defined on the probability space $(\Omega,\mathbf{B},P)$. Then the subspaces S_1 and S_* defined in the norm of the L^α as above are such that $S_m^* = S_1 \oplus S_*$.

Proof:

Suppose $\xi_* \in S_m^*$, then ξ_* can be represented as

$$\xi_* = Z + \sum_{i=m+1}^{n} a_i\xi_i = (Z + \sum_{i=m+1}^{n} a_i\hat{\xi}_i) + \sum_{i=m+1}^{n} a_i(\xi_i - \hat{\xi}_i) \qquad (3.3.4)$$

where $Z \in S_1$. Clearly, the two components on the right hand side of the equality in (3.3.4) are disjoint and independent and hence the result.

The best linear estimator of ξ_m can now be evaluated as the projection onto the two subspaces S_1 and S_* such that

$$\mathrm{disp}(\xi_m - P_{S_m^*}^{\xi_m})$$

is minimized, i.e.,

$$\xi_m^* = P_{S_m^*}^{\xi_m} = P_{S_1}^{\xi_m} + P_{S_*}^{\xi_m} = \hat{\xi}_m + P_{S_*}^{\xi_m}. \qquad (3.3.5)$$

Note here that if we restrict the characteristic exponent to the case where $\alpha > 1$, it clearly implies that

$$P_{S_1}^{\xi_m} = E(\xi_m | \xi_{m-1}, \xi_{m-2}, \ldots),$$

(See Cambanis and Miller, 1981). A problem that now arises is how to evaluate the projection represented by the second component in (3.3.5). This can be represented as

$$P_{S_*}^{\xi_m} = \{ \sum_{k=m+1}^{n} a_k(\xi_k - \hat{\xi}_k) : \mathrm{disp}(\xi_m - P_{S_*}^{\xi_m}) \text{ is a minimum}\} \tag{3.3.6}$$

where the coefficients $\{a_k : k \geq m+1\}$ are estimated such that the error dispersion of the estimate is minimized. This is achieved as follows. Using equations (3.3.5) and (3.3.6), we obtain

$$\xi_m^* = \hat{\xi}_m + \sum_{k=m+1}^{n} a_k(\xi_k - \hat{\xi}_k). \tag{3.3.7}$$

The resulting error for the estimate is evaluated as

$$\xi_m - \xi_m^* = (\xi_m - \hat{\xi}_m) - \sum_{k=m+1}^{n} a_k(\xi_k - \hat{\xi}_k). \tag{3.3.8}$$

Suppose that the scale parameter of η_t denoted as $\mathrm{disp}(\eta_t) = c$ for all $t \in T$. Then we can derive the dispersion of the error of the estimate. We start by defining an approximate value for the joint dispersion (codispersion) function which is the analog of the covariance function in the finite variance case.

Let X and Y be two dependent symmetric stable random variables with characteristic index $\alpha > 0$. In analogy with finite variance case where the covariance between two variable X and Y is given by

$$\mathrm{Cov}(X,Y) = \frac{1}{2}\,[\mathrm{Var}(X + Y) - \mathrm{Var}(X) - \mathrm{Var}(Y)],$$

we define an approximate value for the codispersion between X and Y in the case of infinite variance processes as

$$\mathrm{codisp}\,(X,Y) = [\mathrm{disp}(X+Y) - \mathrm{disp}(X) - \mathrm{disp}(Y)]/\alpha.$$

This approximation of the codispersion is utilized in the interpolation procedure for estimating missing observations.

In the following Theorem, minimum dispersion estimates of the coefficients $\{a_k\}$ in (3.3.8) are obtained by assuming that the forecast $\hat{\xi}_k$ is based on the infinite past and n is large so that the resulting estimates are optimal in the minimum dispersion sense.

Theorem 3.3.1

Let the innovation $\eta_m^* = \xi_m - \xi_m^*$ be as given in (3.3.8) where ξ_m^* is based on an infinite past. The minimum dispersion estimate of a_k is then obtained as

$$\hat{a}_k = \frac{|\beta_{k-m}|^{1/(\alpha-1)}}{\left|\sum_{i=0}^{k-m}|\beta_i|^\alpha + |\beta_{k-m}|^\alpha\right|^{1/(\alpha-1)} - |\beta_{k-m}|^{\alpha/(\alpha-1)}}\,\mathrm{sgn}(\beta_{k-m}) \text{ if } 1 < \alpha < 2$$

$$= -1 \text{ if } -c\sum_{k=m+1}^{n}\sum_{i=0}^{k-m}|\beta_i|^\alpha - c\left\{|1 - \beta_{k-m}|^\alpha - |\beta_{k-m}|^\alpha\right\} < -1 \text{ and } 0 < \alpha \le 1$$

$$= 0 \text{ if } -c\sum_{k=m+1}^{n}\sum_{i=0}^{k-m}|\beta_i|^\alpha - c\left\{|1 - \beta_{k-m}|^\alpha - |\beta_{k-m}|^\alpha\right\} > -1 \text{ and } 0 < \alpha \le 1$$

(3.3.9)

where β_{k-m} is the coefficient on η_k in the summation

$$\sum_{i=0}^{h-1}\beta_i\eta_{k-i}.$$

Proof.

Using (3.3.3), we can write (3.3.8) as

$$\xi_m - \xi_m^* = (\xi_m - \hat{\xi}_m) - \sum_{k=m+1}^{n} a_k\left[\sum_{i=0}^{h-1}\beta_i\eta_{k-i}\right] \tag{3.3.10}$$

where h = k+1-m. Evaluating the dispersion of error given in (3.3.10) leads to

$$\mathrm{disp}(\xi_m - \xi_m^*) = c + c\sum_{k=m+1}^{n}|a_k|^\alpha\left[\sum_{i=0}^{k-m}|\beta_i|^\alpha\right] - \mathrm{Codisp}\left\{\xi_m - \hat{\xi}_m \ , \ \sum_{k=m+1}^{n} a_k\sum_{i=0}^{k-m}\beta_i\eta_{k-i}\right\}$$

$$= c + c\sum_{k=m+1}^{n}|a_k|^\alpha\sum_{i=0}^{k-m}|\beta_i|^\alpha - c\left\{|1 + a_k\beta_{k-m}|^\alpha - |a_k\beta_{k-m}|^\alpha - 1\right\} \tag{3.3.11}$$

The above dispersion is convex for $\alpha \in (1,2)$ and thus by differentiating (3.3.11) with respect to a_k, setting the derivative to zero and solving leads to the result when $1 < \alpha \leq 2$. The case when $0 < \alpha \leq 1$ results by observing possible values of the dispersion in (3.3.11) and in this case $a_k = o$ leads to the minimum.

Colorrary:

Suppose the conditions of theorem 3.3.1 are satisfied. Then the minimum dispersion estimate of ξ_m is obtained as

$$\xi_m^* = \hat{\xi}_m + \sum_{k=m+1}^{n} \frac{|\beta_{k-m}|^{1/(\alpha-1)}}{\left|\sum_{i=0}^{k-m}|\beta_i|^\alpha + |\beta_{k-m}|^\alpha\right|^{1/(\alpha-1)} - |\beta_{k-m}|^{\alpha/(\alpha-1)}} (\xi_k - \hat{\xi}_k) \text{ if } 1 < \alpha \leq 2$$

$$= \hat{\xi}_m + \sum_{k=m+1}^{n} (\xi_k - \hat{\xi}_k) \text{ if } -c \sum_{k=m+1}^{n} \sum_{i=0}^{k-m} |\beta_i|^\alpha - c\left\{|1 - \beta_{k-m}|^\alpha - |\beta_{k-m}|^\alpha\right\} < -1 \text{ and } 0 < \alpha \leq 1$$

$$= \hat{\xi}_m \quad \text{if } -c \sum_{k=m+1}^{n} \sum_{i=0}^{k-m} |\beta_i|^\alpha - c\left\{|1 - \beta_{k-m}|^\alpha - |\beta_{k-m}|^\alpha\right\} > -1 \text{ and } 0 < \alpha \leq 1 \text{ and } 0 < \alpha \leq 1.$$

(3.3.12)

This result simply follows by substituting (3.3.9) into (3.3.7).

Applications

Consider the following AR(p) process ξ_t given by the equation

$$\Phi(B)\xi_t = \eta_t \tag{3.3.13}$$

where $\Phi(B)$ is the AR operator satisfying stationarity conditions. The sequence $\{\eta_t\}$ is of i.i.d stable random variables having characteristic exponent α ($\alpha \in (0,2)$). Under these conditions, the above model can be written as a moving average process, i.e.,

$$\xi_t = [\Phi(B)]^{-1}\eta_t = \Big\{ 1 + (\phi_1 + \phi_2 + ... + \phi_p)B + (\sum_{i>j}^{p} \sum_{j=1}^{p} \phi_i\phi_j + \sum_{j=1}^{p} \phi_i^2)B^2 + \dots \Big\}\eta_t \tag{3.3.14}$$

Thus to evaluate a_k in (3.3.9), we need to note that values of β_k are obtained from (3.3.14) as follows

$$\beta_0 = 1, \qquad \beta_1 = \phi_1 + \phi_2 + ... + \phi_p\,,$$

$$\beta_2 = \sum_{i>j}^{p} \sum_{j=1}^{p} \phi_i\phi_j + \sum_{j=1}^{p} \phi_i^2 \quad , ...$$

and then substituting these into (3.3.12) leads to the required estimate of the missing value.

Example 3.1: AR(1) model

When ξ_t is generated by an AR(1) process

$$\xi_t = \phi\xi_{t-1} + \eta_t$$

satisfying conditions as for the process in (3.3.13) and having the moving average representation as

$$\xi_t = \eta_t + \phi\eta_{t-1} + \phi^2\eta_{t-2} + \ldots..$$

Clearly the best linear predictor of the missing value is obtained as

$$\hat{\xi}_m = \phi\xi_{m-1}.$$

Estimates of a_k are obtained by noting that in equation (3.3.11), $\beta_k = \phi^k$ and then substituting this into equation (3.2.9). Similarly using equation (3.3.12), the estimate of the missing value is obtained as

$$\xi_m^* = \phi\xi_{m-1} + \frac{|\phi|^{1/(\alpha-1)}}{\left|1 + |\phi|^\alpha + |\phi|^\alpha\right|^{1/(\alpha-1)} - |\phi|^{\alpha/(\alpha-1)}}(\xi_{m+1} - \phi^2\xi_{m-1})$$

$$+ \frac{|\phi|^{2/(\alpha-1)}}{\left|1 + |\phi|^\alpha + |\phi^2|^\alpha + |\phi^2|^\alpha\right|^{1/(\alpha-1)} - |\phi^2|^{\alpha/(\alpha-1)}}(\xi_{m+2} - \phi^3\xi_{m-1}) + \ldots$$

$$\cong \phi\xi_{m-1} + \frac{|\phi|^{1/(\alpha-1)}}{\left|1 + |\phi|^\alpha + |\phi|^\alpha\right|^{1/(\alpha-1)} - |\phi|^{\alpha/(\alpha-1)}}(\xi_{m+1} - \phi^2\xi_{m-1})$$

$$+ \frac{|\phi|^{2/(\alpha-1)}}{\left|1 + |\phi|^\alpha + |\phi^2|^\alpha + |\phi^2|^\alpha\right|^{1/(\alpha-1)} - |\phi^2|^{\alpha/(\alpha-1)}}\xi_{m+2}\,.$$

when the index $\alpha = 2$, the above estimator simplifies to the form

$$\xi_m^* = \phi x_{m-1} + \frac{\phi}{1 + \phi^2}(\xi_{m+1} - \phi^2\xi_{m-1}) + \frac{\phi^2}{1 + \phi^2 + \phi^4}(\xi_{m+2} - \phi^3\xi_{m-1}) + ..$$

$$\cong \frac{\phi}{1 + \phi^2}(\xi_{m+1} + \xi_{m-1}).$$

This is a well known result which is a truncated form of the general result obtained here.

Example 3.2: AR(2) model

When p=2 in equation (3.3.13), we have an AR(2) process. The result in (3.3.12) thus simplifies to the form

$$\xi_m^* = \phi_1\xi_{m-1} + \phi_2\xi_{m-2}$$

$$+ \frac{|\phi_1|^{1/(\alpha-1)}}{\left|1 + |\phi_1|^{\alpha} + |\phi_1|^{\alpha}\right|^{1/(\alpha-1)} - |\phi_1|^{\alpha/(\alpha-1)}}(\xi_{m+1} - \phi_1(\phi_1\xi_{m-1} + \phi_2\xi_{m-2})$$

$$+ \frac{|\phi_2+\phi_1^2|^{1/(\alpha-1)}}{\left|1+|\phi_1|^{\alpha}+|\phi_2+\phi_1^2|^{\alpha}+|\phi_2+\phi_1^2|^{\alpha}\right|^{1/(\alpha-1)} - |\phi_2+\phi_1^2|^{\alpha/(\alpha-1)}}\Big[\xi_{m+2} - \phi_1^2(\phi_1\xi_{m-1} + \phi_2\xi_{m-2})$$

$$- 2\phi_1\phi_1\xi_{m-1} - \phi_2^2\xi_{m-2}\Big] + \ldots$$

Again it is of interest to note that when $\alpha = 2$, the above result simplifies to the form

$$\xi_m^* \cong \frac{\phi_1(\xi_{m-1} + \xi_{m+1}) + \phi_2\xi_{m-2} + (\phi_2 + \phi_1^2)\xi_{m+2}}{1 + \phi_1^2 + (\phi_2 + \phi_1^2)^2}.$$

In the case of a moving average process, $\beta_k = \theta_k$. Using the autoregressive representation, the future values are evaluated in a similar way as for the AR process and the values of the missing value estimates follow.

Two missing values

In this Section, we consider the situation where two consecutive observations are missing. The case of nonconsecutive missing values follows easily from the previous results and so we do not discuss it here.

Suppose two consecutive values ξ_k and ξ_{k+1} are missing. Lemma 3.3.1 would still be valid although now we could use a vector form such that $\xi_m = (\xi_k, \xi_{k+1})^T$ while the matrix

$$a_s = \begin{bmatrix} a_k & 0 \\ 0 & a_{k+1} \end{bmatrix}$$

and H^T is the row vector of H. We can thus write equation (3.3.7) as

$$\xi_m^* = \hat{\xi}_m + \sum_{s=k+2}^{n} \mathbf{a}_s(\xi_s - \hat{\xi}_s) \tag{3.3.15}$$

The innovation for the missing value given in (3.3.8) would now be a column vector

$$\xi_m - \xi_m^* = \xi_m - \hat{\xi}_m - \sum_{s=k+2}^{n} \mathbf{a}_s(\xi_s - \hat{\xi}_s) \tag{3.3.16}$$

where $\xi_s - \hat{\xi}_s$ is also a column vector.

Theorem 3.3.2

The estimate of $\mathbf{a}_s$ obtained by minimizing the dispersion of (3.3.16) is

$$\hat{\mathbf{a}}_s = \mathbf{I}\frac{|\beta_{s-k}|^{1/(\alpha-1)}}{\left|\sum_{i=0}^{s}|\beta_i|^{\alpha} + |\beta_{s-k}|^{\alpha}\right|^{1/(\alpha-1)} - |\beta_{s-k}|^{\alpha/(\alpha-1)}}\mathrm{sgn}(\beta_{s-k})$$

$$= -\mathbf{I} \text{ if } \quad -\mathbf{c}\sum_{s=k+2}^{n}\sum_{i=0}^{s-k}|\beta_i|^{\alpha} - \mathbf{c}\left\{|1 - \beta_{s-k}|^{\alpha} - |\beta_{s-k}|^{\alpha}\right\} < -1 \text{ and } 0 < \alpha \leq 1$$

$$= \mathbf{0} \quad \text{if} \quad -\mathbf{c}\sum_{k=m+1}^{n}\sum_{i=0}^{s-k}|\beta_i|^{\alpha} - \mathbf{c}\left\{|1 - \beta_{s-k}|^{\alpha} - |\beta_{s-k}|^{\alpha}\right\} > -1 \text{ and } 0 < \alpha \leq 1$$

(3.3.17)

where $\mathbf{I}$ is a 2x2 identity matrix and $\mathbf{0}$ is a zero matrix. β_{k-m} is as in Theorem 3.3.1.

Proof.

Using (3.3.15), we can write (3.3.16) as

$$\xi_m - \xi_m^* = (\xi_m - \hat{\xi}_m) - \sum_{s=k+2}^{n} \mathbf{a}_s \sum_{i=0}^{s} \beta_i \eta_{s+k+i}$$

Evaluating the dispersion of (3.3.16) leads to

$$\mathrm{disp}(\xi_m - \xi_m^*) = \mathbf{c} + \mathbf{c} \sum_{s=k+2}^{n} |\mathbf{a}_s|^{\alpha} \sum_{i=0}^{s-k} |\beta_i|^{\alpha} \ - \mathrm{Codisp}\{(\xi_m - \xi_m^*) \ , \ \sum_{s=k+2}^{n} \mathbf{a}_s \sum_{i=0}^{s-k} \beta_i \eta_{s-i}\}$$

$$= \mathbf{c} + \mathbf{c} \sum_{s=k+2}^{n} |\mathbf{a}_s|^{\alpha} \sum_{i=0}^{s-k} |\beta_i|^{\alpha} - \mathbf{c} \left\{ |\mathbf{I} + \mathbf{a}_s \beta_{s-k}|^{\alpha} - |\mathbf{a}_s \beta_{s-k}|^{\alpha} - \mathbf{1} \right\} \tag{3.3.18}$$

where $\qquad |\mathbf{a}_s|^{\alpha} = \begin{bmatrix} |a_k|^{\alpha} & 0 \\ 0 & |a_{k+1}|^{\alpha} \end{bmatrix},$

$\mathbf{c} = (c, c)^T$ and η_{m+h-i} is a column vector.

The above dispersion is convex for $\alpha \in (1,2)$ and thus by differentiating (3.3.18) with respect to $\mathbf{a}_s$, setting the differential to zero ans solving, we obtain

$$|\mathbf{a}_s|^{\alpha-1} \sum_{i=0}^{s} |\beta_i|^{\alpha} - \beta_{s-k} |\mathbf{I} + \mathbf{a}_s \beta_{s-k}|^{\alpha-1} - |\mathbf{a}_s|^{\alpha-1} |\beta_{s-k}|^{\alpha} = 0$$

where **I** is similarly a 2x2 identity matrix. The result then easily follows. The case when $0 < \alpha \leq 1$ follows by examining possible values of the dispersion in (3.3.18).

Corollary 3.3.2

Let the estimate of the missing value be as given in (3.3.15). The minimum dispersion estimate of ξ_m is obtained as

$$\xi_m^* = \hat{\xi}_m + \mathbf{I}\left\{ \sum_{s=k+2}^{n} \frac{|\beta_{S-k}|^{1/(\alpha-1)}}{\left|\sum_{i=0}^{s-k}|\beta_i|^{\alpha} + |\beta_{S-k}|^{\alpha}\right|^{1/(\alpha-1)} - |\beta_{S-k}|^{\alpha/(\alpha-1)}} (\xi_s - \hat{\xi}_s) \right\} \text{ if } 1 < \alpha \leq 2$$

$$= \hat{\xi}_m + \sum_{s=k+2}^{n} (\xi_k - \hat{\xi}_k) \text{ if } -\mathbf{c} \sum_{s=k+2}^{n} \sum_{i=0}^{s-k} |\beta_i|^{\alpha} - \mathbf{c}\left\{ |1 - \beta_{s-k}|^{\alpha} - |\beta_{s-k}|^{\alpha} \right\} < -\mathbf{1} \text{ and } 0 < \alpha \leq 1$$

$$= \hat{\xi}_m \quad \text{if } -\mathbf{c} \sum_{s=k+2}^{n} \sum_{i=0}^{s-k} |\beta_i|^{\alpha} - \mathbf{c}\left\{ |1 - \beta_{s-k}|^{\alpha} - |\beta_{s-k}|^{\alpha} \right\} > -\mathbf{1} \text{ and } 0 < \alpha \leq 1 \text{ and } 0 < \alpha \leq 1$$

(3.3.19)

where $\mathbf{1} = (1,1)^T$ and $\mathbf{I}$ is a 2x2 identity matrix.

This result follows when (3.3.17) is substituted into (3.3.15).

Example 3.3

Consider the AR(1) process ξ_t such that

$$\xi_t = \phi\xi_{t-1} + \eta_t$$

satisfies the conditions for the process in (3.2.13) and has the moving average representation

$$\xi_t = \eta_t + \phi\eta_{t-1} + \phi^2\eta_{t-2} + \ldots..$$

Suppose for a given set of observations, two consecutive values ξ_k, ξ_{k+1} are missing. The minimum dispersion estimates for the missing values would be obtained from equations (3.3.17) and (3.3.19) by again noting that in this case, $\beta_i = \phi^i$.

$$\xi_m^* = \begin{bmatrix} \phi\xi_{k-1} \\ \phi^2\xi_{k-1} \end{bmatrix}$$

$$+ \frac{|\phi|^{1/(\alpha-1)}}{\left|1 + |\phi|^\alpha + |\phi|^\alpha\right|^{1/(\alpha-1)} - |\phi|^{\alpha/(\alpha-1)}} \begin{bmatrix} \xi_{k+2} - \phi^3\xi_{k-1} \\ \xi_{k+2} - \phi^4 x_{k-1} \end{bmatrix}$$

$$+ \frac{|\phi|^{2/(\alpha-1)}}{\left|1+|\phi|^\alpha+|\phi^2|^\alpha+|\phi^2|^\alpha\right|^{1/(\alpha-1)} - |\phi^2|^{\alpha/(\alpha-1)}} \begin{bmatrix} \xi_{k+3} - \phi^4\xi_{k-1} \\ \xi_{k+3} - \phi^5\xi_{k-1} \end{bmatrix} + \ldots$$

When $\alpha = 2$, the above result simplifies to

$$\xi_m^* = \begin{bmatrix} \phi\xi_{k-1} \\ \phi^2\xi_{k-1} \end{bmatrix}$$

$$+ \frac{\phi^2}{1 + \phi^2 + \phi^4} \begin{bmatrix} \xi_{k+2} - \phi^3\xi_{k-1} \\ \xi_{k+2} - \phi^4\xi_{k-1} \end{bmatrix} + \frac{\phi^4}{1 + \phi^2 + \phi^4 + \phi^6} \begin{bmatrix} \xi_{k+3} - \phi^4\xi_{k-1} \\ \xi_{k+3} - \phi^5\xi_{k-1} \end{bmatrix} + \ldots$$

$$\cong \frac{\phi}{1 + \phi^2 + \phi^4} \begin{bmatrix} \phi\xi_{k+2} + (1 + \phi)\xi_{k-1} \\ \phi\xi_{k+2} + \phi\xi_{k-1} \end{bmatrix}.$$

Generalization to more than two consecutive missing values would follow from the above results although this would get be more complicated as the number of parameters increase.

3.4. Least lambda estimates

Consider the following AR(p) process

$$\Phi(B)\xi_t = \eta_t$$

where $\Phi(B)$ is the AR operator satisfying stationarity conditions and the sequence $\{\eta_t\}$ is a sequence of i.i.d zero centered random variables with common variance σ^2. A common procedure for the estimation of the parameters ϕ_i, i = 1, 2, .., p is to minimize the function

$$g_n(\phi) = \sum_{t=2}^{n} [\xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i}]^2$$

which leads to the least squares estimates for the parameters. An alternative procedure is based on minimizing

$$S_n(\phi) = \sum_{t=2}^{n} |\xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i}| .$$

This leads to the least absolute deviation estimates. When estimating missing observations, the above functions $g_n(\phi)$ and $S_n(\phi)$ are minimized when the missing value is considered as a parameter. In this section, a more general function $f_n(.)$ which includes $g_n(\phi)$ and $S_n(\phi)$ as special cases is minimized. Such a function can be written as

$$f_n(\beta,\xi) = \frac{1}{n} \sum_{t=2}^{n} |\xi_t - \sum_{i=1}^{p} \phi_i \xi_{t-i}|^\lambda \tag{3.4.1}$$

where $0 < \lambda \leq \alpha$ and $\alpha \in (0,2]$.

Consider a stationary process $\{\xi_t\}$ which is driven by symmetric stable innovations $\{\eta_t\}$ and is defined on the probability space (Ω,S,P). Let the innovation sequence $\{\eta_t\}$ consist of i.i.d stable random variables with a common characteristic exponent α. Suppose that ξ_t is an AR(p) process. Then it is well known that the function $f_n(\beta,\xi)$ given in equation (3.4.1) converges to

$$E|\xi_t - \sum_{i=1}^{p}\phi_i\xi_{t-i}|^\lambda,$$

which is the dispersion of the process. This is a widely applied result particularly when ξ_t is defined in the Hibert space and extension to the L^λ space $\lambda \in (0,2)$ follows by use of the egordic Theorem.

Single Missing Value

Consider the the series $\{\xi_t\ t=1,2,.,m,..n\}$ such that the m^{th} observation $(1 \leq m < n)$ is not observed. Suppose that ξ_t is a p^{th} order autoregressive given by

$$\xi_t = \sum_{i=1}^{p}\phi_i\xi_{t-i} + \eta_t \tag{3.4.2}$$

where $\{\eta_t\}$ is the sequence of independent zero centered symmetric stable random variables with common characteristic exponent α $(0 < \alpha \leq 2)$ and the model parameters $\{\phi_i, i = 1, 2, .., p\}$ are assumed to be known and satisfy the stationarity conditions.

We aim to estimate the missing value $\beta = \xi_m$ by minimizing the dispersion of the innovation sequence with respect to β. This is achieved by minimizing the λ^{th} norm of this sequence. Let the function $f_n(\beta,\xi)$ be defined as above. We consider the unobserved value

ξ_m as the parameter β and aim to minimize the above function with respect to the unknown parameter β.

Two cases arise when $f_n(\beta,\xi)$ is minimized depending on the value of the characteristic exponent. The two cases are considered separately.

Case 1: $(1 < \lambda \leq 2)$

Here, $f_n(\beta, \xi)$ which we simply write as $f_n(\beta)$ is convex in β so the optimal value of β is obtained by solving $f_n^{(1)}(\beta, \xi) = 0$ where $f^{(1)}(.)$ is the derivative of $f(.)$ with respect to the first argument. This leads to the function

$$\lambda|\beta - \sum_{i=1}^{p}\phi_i\xi_{m-i}|^{\lambda-1} - \phi_1\lambda|\xi_{m+1} - \phi_1\beta - \sum_{i=2}^{p}\phi_i\xi_{m-i} - \ldots |^{\lambda-1} = 0.$$

To evaluate the estimate of β when $p > 1$ requires the use of minimization subroutines but when the order $p = 1$, an analytical form of the estimate is obtained as

$$\beta^* = \frac{\phi\xi_{m-1} + \xi_{m+1}(\phi)^{1/\lambda-1}\,\mathrm{sgn}(\phi)}{[1 + (\phi)^{\lambda/\lambda-1}]}. \tag{3.4.3}$$

It is of interest to note that when $\lambda = 2$, the above estimate simplifies to the form

$$\beta^* = \frac{\phi(\xi_{m-1} + \xi_{m+1})}{1 + \phi^2}$$

which is a well known formula for the estimate of a missing observation for the AR(1) process.

Case 2: $(0 < \lambda \leq 1)$

In this case $f(\beta,\xi_t)$ is not necessarily convex and may not have a unique minimum with respect to β. When p = 1, the estimate of the missing values is obtained as

$$\beta^* = \phi\xi_{m-1} \qquad \text{if } |\phi| > \sqrt{\xi_{m+1}/\xi_{m-1}}$$

$$= \xi_{m+1}/\phi \qquad \text{if } |\phi| < \sqrt{\xi_{m+1}/\xi_{m-1}} \tag{3.4.4}$$

It should be noted that these estimates are subject to the initial restrictions imposed on the parameters of the process.

Two missing values

Suppose the values ξ_k and ξ_{k+m} are missing from the set $\{\xi_t : t=1, 2, ..., k, .., k+m, .. n\}$ where $1 \leq k < k+m < n$. We aim to obtain estimates of the two missing values by minimizing the norm given in (3.4.1) as a function of the two missing values which are similarly considered as parameters β_1 and β_2. For the evaluations, we consider the case when m = 1.

When $1 < \lambda \leq 2$, the convexity of the function $f(\beta_1, \beta_2,\xi_t)$ enables us to set the derivatives $f^{(1)}(\beta_1, \beta_2)$ to zero and solve for β_1 and β_2 i.e.,

$$\frac{d(f(\beta_1, \beta_2))}{d\beta_1} = \lambda|\beta_1 - \sum_{i=1}^{p}\phi_i\xi_{k-i}|^{\lambda-1} - \phi_1\lambda|\beta_2 - \phi_1\beta_1 - \sum_{i=1}^{p}\phi_i\xi_{k-i}|^{\lambda-1}$$

$$- \phi_2\lambda|\xi_{k+2} - \phi_1\beta_2 - \phi_2\beta_1 - \sum_{i=1}^{p}\phi_i\xi_{k-i}|^{\lambda-1} - \dots = 0$$

$$\frac{d(f(\beta_1, \beta_2))}{d\beta_2} = \lambda|\beta_2 - \phi_1\beta_1 - \sum_{i=1}^{p}\phi_i\xi_{k-i}|^{\lambda-1} - \phi_1\lambda|\xi_{k+2} - \phi_1\beta_2 - \phi_2\beta_1 - \sum_{i=1}^{p}\phi_i\xi_{k-i}|^{\lambda-1} - \cdots = 0.$$

Solving the above equations simultaneously for the two missing values is difficult and numerical algorithms would be required. However, for the case when $p = 1$, the derivatives are obtained as

$$\frac{d(f(\beta_1, \beta_2))}{d\beta_1} = \lambda|\beta_1 - \phi\xi_{k-1})|^{\lambda} - \phi\lambda|\beta_2 - \phi\beta_1|^{\lambda-1} = 0$$

$$\frac{d(f(\beta_1, \beta_2))}{d\beta_2} = \lambda|\beta_2 - \phi\beta_1|^{\lambda-1} - \phi\lambda|\xi_{k+2} - \phi\beta_2|^{\lambda-1} = 0$$

and solving the above two equations simultaneously leads to the following solutions:

$$\beta_1^* = \frac{1}{G}\left[(1 + (\phi_{n-k})^{\lambda/(\lambda-1)})\phi\xi_{k-1} + (\phi)^{2/(\lambda-1)}\xi_{k+2}\right]\mathrm{sgn}(\phi)$$

$$\beta_2^* = \frac{1}{G}\left[\phi^2\xi_{k-1} + (\phi)^{1/(\lambda-1)}(1 + (\phi)^{\lambda/(\lambda-1)})\xi_{k+2}\right]\mathrm{sgn}(\phi)$$

where $$G = [1 + (\phi)^{1/(\lambda-1)} + (\phi)^{2\lambda/(\lambda-1)}]$$

When $\lambda = 2$, these simplify to

$$\beta_1^* = \frac{(1+\phi)\phi\xi_{k-1} + \phi^2\xi_{k+2}}{1 + \phi + \phi^4}$$

and

$$\beta_2^* = \frac{\phi^2\xi_{k-1} + \phi + (1 + \phi^2)\xi_{k+2}}{1 + \phi + \phi^4}$$

For nonconsecutive missing values, the analytical form of the estimates would be simpler. Extension to s missing values for the AR(1) model follows easily using the above procedure .

When $0 < \lambda \leq 1$, the function $f(\beta_1,\beta_2,\xi_t)$ is not convex and may not posses a unique minimum. Thus when p=1, we would consider the values of β_1 and β_2 in the intervals $[\phi\xi_{k-1}, \beta_2/\phi]$ and $[\phi\beta_1, \xi_{k+2}/\phi]$ respectively. This leads to a single feasible set of estimates as

$$\beta_1^* = \phi\xi_{k-1} \quad \text{and} \quad \beta_2^* = \phi^2\xi_{k-1}$$

under the restriction that $|\phi| < 1$. Other possible regions would violate this requirement on the parameter.

Remarks:

(i) When $\alpha > 1$, the analytical form of the optimal estimate of the missing observation for the AR(1) is the same as the least λ estimate both of which are truncated forms of the linear estimate. This basically implies that the three methods esimate the missing values by approximately the same margins.

(ii) In the case when $\alpha < 1$, the linear and the least λ estimates of the missing observation are slightly different depending on the restrictions on the parameters for the model.

CHAPTER FOUR

4. NONSTATIONARY PROCESSES WITH INFINITE VARIANCE

4.1. Introduction

A sequence $\{\xi_t\}$ is said to be stationary in a general sense if its statistical properties such as the mean and the variance are constant relative to changes in time. Strict stationarity is attributed to such a sequence if for all $\{t_i \,;\, i = 1,2,3,...,n\}$, the joint probability distribution of $\{\xi(t_1), \xi(t_2), ..., \xi(t_n)\}$ is the same as that of $\{\xi(t_1+k), \xi(t_2+k), ..., \xi(t_n+k)\}$ for all integers k and all t_i , $i = 1,2,3,...,n$. Sequences which possess stationary first and second order properties are said to be 'weakly' or 'second order' stationary and this property can be ascribed to most finite variance processes including the well known class of Gaussian processes. This is not so for the case of infinite variance processes whose second order properties and possibly the first order properties are not defined. The notion of stationarity in the case of infinite variance processes can only be viewed in the strict sense since their second order properties do not exist. Processes which are not stationary are said to be nonstationary.

The major essence of the stationarity requirement in time series analysis is to give a sense of stability to the series since for most series, only single realizations are available. Lack of stationarity thus invalidates the identification of the process and consequently the applicability of the analytical results such as forecasts based on such results since these would be termed as being 'unstable'.

Lack of stationarity in linear processes occurs as a result of various anomalies such as the presence of a trend in the series which clearly implies a lack of stationarity in the first order properties. It is also often the case that the second order properties such as the variance are not constant and this too leads to nonstationarity. The presence of explosive roots (roots which are inside the unit circle) similarly results in nonstationarity.

In the literature on nonstationary sequences, various techniques such as the differencing of the series, data transformation and use of autoregressive moving average (ARMA) models with time varying coefficients have been suggested as possible solutions to counteracting nonstationarity (Box and Jenkins, 1976; Priestley, 1981). The differencing of a series is aimed at removing the nonstationarity in the first order properties as manifested by the presence of a trend or explosive roots in the model. This technique has led to the fitting of the class of autoregressive intergrated moving average (ARIMA) models which have been widely studied (Box and Jenkins, 1976).

A possible remedy to the presence of nonstationary second order properties is to transform the data using techniques such as the power transformations of Box and Cox (1964) for some series. Unfortunately, suitable transformations do not always exist or if they do, retransformation of the results does not always lead to optimal results on the original scale (Granger and Andersen, 1976; Nassiuma and Ordhoukhani, 1991). A case in point would be for example an ARIMA model with unstable second order properties. The constant coefficients in such models make these models inappropriate for removing the second order nonstationarity and a possible alternative approach studied in the literature is the use of ARMA models with time varying coefficients (See Whittle, 1965 and Niemi, 1983).

Stationarity has so far been studied in the Hilbert space context but the presence of nonstationarity is a problem that is undesirable even in infinite variance processes. It is the goal in this Chapter to show that under some mild conditions, solutions can be obtained for nonstationary sequences with infinite variance. This is motivated by the continued realization that for many practical situations, the Gaussian laws are not fully adequate and the stable laws may be more appropriate (See for example Fama, 1965; Granger and Orr, 1972; Dumouchel, 1973). The increasing interest in the analysis of stable processes thus calls for their study particularly when they are characterized by nonstationary behavior. Studies on nonstationary infinite variance processes include the work of Chan and Tran (1989), Chan (1990), Phillips (1990), Nassiuma (1991) and Nassiuma (1992). Chan and Tran (1989) studied the properties of the least squares estimates for the AR(1) model having unit roots and driven by infinite variance innovations while Chan (1990) considers some limiting properties of least squares parameter estimates for the first order autoregressive processes with a 'nearly' unit root and similarly driven by infinite variance stable innovations. Phillips (1990) discusses limiting properties for the regression case with unit root and infinite variance errors.

In this chapter, general nonstationary autoregressive moving average processes are studied. Such processes can be represented by the following difference equations:

$$\Phi(B)\xi_t = \Theta(B)\eta_t. \tag{4.1}$$

$$\Phi_t(B)\xi_t = \Theta_t(B)\eta_t. \tag{4.2}$$

The major distinction between the two processes is that in (4.1), the model coefficients are constant and satisfy the stationarity and invertibility conditions while the scale parameter for the innovations is time varying. The model coefficients in (4.2) are time varying but the scale parameter is assumed constant. The coefficients in (4.2) satisfy the stationarity and invertibility for each time period t. A simple example of such a model is the random coefficients autoregressive model of Nicolls and Quinn (1982).

When the variance of the random variable η_t is finite, prediction for nonstationary processes of the type given in (4.1) have been discussed by Whittle (1965) and Abdrabbo and Priestley (1967).

In this chapter, the infinite variance case is studied for the models in (4.1) and (4.2). Properties of first order autoregressive models having unit roots are discussed in Section 4.2 while the conditions under which unique solutions to (4.1) and (4.2) exist are given in Sections 4.3 and 4.4 respectively. Linear forecasts are evaluated and by analogy with the finite variance case when the characteristic exponent α of η_t is 2, the forecast error is given. In Section 4.4, the estimation of model coefficients in the case when the coefficients are time varying is studied for the AR(1) model. Some of the results discussed in Sections 4.3 and 4.4 have been published by Nassiuma (1991) and Nassiuma (1992).

4.2. Processes having unit roots

Consider the process ξ_t generated by a p^{th} order autoregressive process given by the difference equation

$$\Phi(B)\xi_t = \eta_t, \quad t \in \{0,1,2,...\} \tag{4.2.1}$$

where $\{\eta_t;\ t = 1,2,..,n\}$ is a sequence of independent and identically distributed (i.i.d) random variables and the zeros of the polynomial $\Phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \ ....+ \phi_p z^p$ lie in the region $|z| > 1$. This requirement is often not satisfied by one or more roots and this leads to nonstationarity in the process. Properties of least squares parameter estimates for autoregressive processes having unit roots and explosive roots have been widely studied in the finite variance context by several authors for example Fuller (1976), and Basu and Roy (1989).

Nonstationary processes driven by infinite variance stable innovations have recently been studied by Chan and Tran (1989), Chan (1990) and Phillips (1990). Their studies have mainly focussed on properties of least squares estimates of the AR(1) model parameter when the root is one or close to one.

Alternative parameter estimates for the autoregressive process have recently been developed and these include the least absolute deviation estimates (Gross and Steiger, 1979) and the least gamma estimates proposed by Liu (1987). These have been shown to be consistent. The optimal estimate, developed in Chapter 2 under the assumption that the sequence $\{\eta_t\}$is of infinite variance stable random variables, was similarly shown to be a consistent estimator of the parameter under the stationarity assumption.

Consider now a sequence $\{\xi_t\}$ generated by the AR(1) process and given by

$$\xi_t = \phi\xi_{t-1} + \eta_t$$

where $\{\eta_t\}$ is a sequence of infinite variance stable random variables having characteristic exponent $\alpha \in (1,2)$. The optimal estimate of the coefficient ϕ was obtained in chapter 2 as

$$\phi^o = \sum_{t=2}^{n} |\xi_{t-1}|^{1/(\alpha-1)}\xi_t \mathrm{sgn}(\xi_{t-1}) / \sum_{t=2}^{n} |\xi_{t-1}|^{\alpha/(\alpha-1)}\mathrm{sgn}(\xi_{t-1}) \qquad (4.2.2)$$

when the scale parameter of h_t is constant, i.e., $c(t) = 1$ for all $t = 1,2,...n$. Our interest in this section centers on the case when $|\phi| = 1$.

Consider a positive function $L(x)$. Such a function is said to vary slowly at infinity if for all positive x, the following limit exists:

$$\lim_{t\to\infty} \frac{L(tx)}{L(t)} \to 1.$$

A necessary and sufficient condition for the sequence of iid random variables $\{\eta_t\}$ having distribution function F to be in the domain of attraction of the stable law with characteristic exponent α is that there exists a slowly varying function $L(x)$ such that as $x \to \infty$,

$$1 - F(x) \approx px^{-\alpha}L(x)$$

where
$$p = \lim_{\eta\to\infty} \frac{P(\eta > x)}{P(|\eta| > x)}.$$

Let the sequence of iid random variables $\{\eta_t, t = 1, 2, ..., n\}$ be defined in the domain of attraction of the stable law with characteristic exponent $\alpha \in (0,2)$ and distribution F. Suppose that there exists constants a_n and b_n defined as

$$a_n = \inf\{x: [1 - F(x) + F(-x)] \le n^{-1}\}$$

and $$b_n = E(\eta_t I_{[|\eta_t| \leq a_n]})$$

where $I_{[|\eta_t| \leq a_n]}$ is an indicator variable taking on values 0 or 1.

The following limit then exists:

$$\frac{1}{a_n}\sum_{t=1}^{n}(\eta_t - b_n) \xrightarrow{d} S_\alpha,$$

where S_α is a stable random variable with characteristic exponent α (Feller, 1971). When $\alpha > 1$, b_n is obtained as $E(\eta_t)$.

We show the strong consistency of the optimal estimate in the following Theorem.

Theorem 4.2.1

Consider the optimal estimator given in (4.2.2) for the parameter of the AR(1) model having $|\phi| = 1$. Let the sequence η_t be such that $a_n(\eta_n+\eta_{n-1} + \ldots + \eta_1)$ (where a_n is a regularly varying function of the form $n^{-\upsilon}$ and $\upsilon < \alpha$) be in the domain of attraction of the stable law with index $\alpha > 1$. Then the optimal estimator is strongly consistent, i.e., as $n \rightarrow \infty$,

$$n^{-\vartheta}(\phi^o - 1) \rightarrow 0 \qquad \text{almost surely.}$$

Proof:

When $|\phi| = 1$, this implies that the AR(1) model can be written as

$$\xi_t = \xi_{t-1} + \eta_t$$

It is then clear that

$$\phi^0 - 1 = \frac{\sum_{t=2}^{n} |\xi_{t-1}|^{1/(\alpha-1)}(\xi_t - \xi_{t-1})\mathrm{sgn}(\xi_{t-1})}{\sum_{t=2}^{n} |\xi_{t-1}|^{\alpha/(\alpha-1)}\mathrm{sgn}(\xi_{t-1})}$$

$$= \frac{\sum_{t=2}^{n} |\xi_{t-1}|^{1/(\alpha-1)}\eta_t\mathrm{sgn}(\xi_{t-1})}{\sum_{t=2}^{n} |\xi_{t-1}|^{\alpha/(\alpha-1)}\mathrm{sgn}(\xi_{t-1})}. \tag{4.2.3}$$

It is clear that $|\xi_{t-1}|^{1/(\alpha-1)}$ is a bounded function and also η_t and ξ_{t-1} are independent. It thus follows from Lemma 2.4.1 that for $\delta = \min(\alpha, \alpha(\alpha-1))$,

$$\frac{1}{n^\delta}\sum_{t=2}^{n} |\xi_{t-1}|^{1/(\alpha-1)}\eta_t\mathrm{sgn}(\xi_{t-1})$$

converges almost surely to zero as $n \to \infty$. By a similar argument, we note that the denominator converges almost surely to infinity. Combination of the two results then leads to the Theorem.

Simulation results

Observations for the AR(1) process when $\phi = 1$ were generated using the Fortran subroutine GGSTA. The first 200 observations were deleted. Mean values of the optimal and least squares estimates based on 500 samples each of sizes 50, 150, 500 and 1000 were evaluated and the results are given in Tables 4.1, 4.2 and 4.3 respectively.

It is observed that both the optimal estimate and the least squares estimate the parameter with approximately the same error margin.

There is minimal change in both the least squares and optimal estimates when a wrong value of the characteristic index is used instead of the true one. This basically is expected considering the minimal differences between the optimal and the least squares estimates.

Table 4.1: Optimal and Least squares estimates for the parameter $\phi = 1$ of an AR(1) model when the number of simulations N=500.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.1	0.9877 (1.499E-04)	1.0026 (6.778E-06)	0.9926 (5.394E-05)	0.9926 (5.397E-05)
	1.2	0.9926 (5.501E-05)	1.002 (2.703E-06)	0.9953 (2.200E-05)	0.9956 (1.894E-05)
Least Squares Estimate	1.1	0.9980 (3.593E-06)	1.000 (9.648E-09)	0.9990 (1.028E-06)	0.9990 (1.414E-06)
	1.2	0.9993 (5.092E-07)	1.0012 (1.446E-06)	0.9993 (4.675E-07)	0.9994 (3.609E-07)

Table 4.2: Optimal and Least squares estimates for the parameter $\phi = 1$ of an AR(1) model when the number of simulations N=100.

	Alpha (α)	n=50	n=150	n=500	n=1000
Optimal Estimate	1.4	0.9994 (4.006E-07)	1.0001 (7.091E-07)	0.9982 (3.097E-06)	0.9992 (6.153E-07)
	1.6	1.0044 (1.911E-05)	1.0033 (1.066E-05)	0.9998 (5.670E-08)	1.0001 (2.185E-08)
	1.9	1.0109 (1.191E-04)	1.0050 (2.545E-05)	1.0007 (4.860E-07)	1.0005 (2.801E-07)
Least Squares Estimate	1.4	1.0022 (4.946E-06)	1.0028 (8.355E-06)	0.9999 (2.579E-09)	1.0001 (1.421E-08)
	1.6	1.0055 (3.016E-05)	1.0040 (1.593E-05)	1.0004 (1.777E-07)	1.0004 (1.891E-07)
	1.9	1.0110 (1.221E-04)	1.0051 (2.641E-05)	1.0008 (6.190E-07)	1.0006 (3.252E-07)

4.3. ARMA processes with nonstationary innovations

In this section, we study some properties of the autoregressive moving average (ARMA) process given in equation (4.1) where $\Phi(B)$ and $\Theta(B)$ are the autoregressive (AR) and the moving average (MA) operators satisfying the stationarity and invertibility conditions. The sequence $\{\eta_t\}$ is of infinite variance stable random variables with characteristic exponent $\alpha \in (0, 2)$ and time varying scale parameter c(t).

Let (Ω, F, P) be a probability space and let $L^{\alpha}(\Omega)$ be the space of all real valued random variables η_t defined on the probability space such that $E|\eta_t|^{\alpha} < \infty$. Assume that ξ_t^* is a solution to the stochastic difference equation (4.1) defined on the probability space and let $S_t(\xi^*)$ and $S_t(\eta)$ be the subspaces of the space $L^{\alpha}(\Omega)$ spanned by the random variables ξ_s^* and η_s (where $s \in T = (0, \pm1, \pm2,...)$ and $s \leq t$) respectively. Under the assumptions on $\Phi(B)$ and $\Theta(B)$ given above, it follows that the process ξ_t given in equation (4.1) is admits a moving average representation and is invertible and thus the subspaces $S_t(\xi^*)$ and $S_t(\eta)$ are equal. Under these conditions, a stochastic difference equation as given in (4.1) thus represents an ARMA(p,q) process if it admits a solution $\{\xi_t^*: t \in T\}$ such that $\xi_t^* \in S_t(\eta)$. Under some mild conditions, a solution to the difference equation (4.1) exists. These are given in the following Theorem.

Theorem 4.3.1

Suppose that for the linear process ξ_t given in equation (4.1), the assumptions on the polynomials $\Phi(B)$ and $\Theta(B)$ are satisfied. Let the following further conditions also be satisfied;

(a) $E|\eta_t|^{\alpha} < \infty$ for all $t \in T$ and $\alpha \in (0,2)$

and

$$\text{(b)} \sum_{s=0}^{\infty} |\psi_j|^{\alpha} c(t\text{-}j) < \infty$$

where c(t) is the scale parameter for η_t and $\psi(B) = \theta(B)/\Phi(B)$. Then the process ξ_t has the infinite moving average process representation of the form

$$\xi_t^* = \sum_{j=0}^{\infty} \psi_j \eta_{t-j} \tag{4.3.1}$$

which constitutes a solution to the process in (4.1).

Proof:

Let the sequence of random variables $\{w_t\}$ be defined by the moving average process $w_t = \theta(B)\eta_t$. The above conditions then imply that $E|w_t|^{\alpha}$ is uniformly bounded on the set T=(0, ±1, ±2, ...) and thus there exists a solution ξ_t^* for $\Phi(B)\xi_t = w_t$ given by

$$\xi_t^* = [\Phi(B)]^{-1} w_t = \sum_{j=0}^{\infty} \psi_j^* w_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$$

as stipulated in the Theorem.

A solution to equation (4.1) implies that such a model could be utilized for prediction purposes. Now suppose the solution ξ_t^* as obtained above is a unique solution to the process given in (4.1) and assume that conditions (a) and (b) of Theorem 4.2.1 are satisfied for $\alpha \in (0,2)$. Let $S(\xi^*)$ be the subspace spanned by the set of random variables $\{\xi_s^* : s \in T\}$. An optimal linear h steps ahead forecast (such that the forecast error dispersion is minimized) of ξ_{t+h}^* for $h > 0$ is the projection of the subspace $S(\xi^*)$ (which is spanned by $\overset{*}{\xi}_s$ for $s \in T$) onto the subspace $S_t(\xi^*)$, $t = 1,2,\ldots,n$ (also known as the observation space) i.e., the optimal linear predictor $\hat{\xi}^*_t(h)$ of ξ^*_{t+h} is given by

$$\hat{\xi}^*_t(h) = P_{S_t(\xi^*)}\, \xi^*_{t+h}$$

where

$$P_{S_t(\xi^*)}\, \overset{*}{\xi}_{t+h} = \overset{*}{\xi}_{t+h} \quad \text{if } h \le 0$$

$$= \overset{*}{\xi}_t(h) \quad \text{if } h > 0$$

It is to be noted that in the finite variance case with $\alpha = 2$, the corresponding projection from the L^2 space is orthorgonal. However, this is not necessarily the case under the infinite variance setting. When $\alpha < 2$, a similar projection is feasible and the distance measure to be minimized now corresponds to the dispersion (Cline and Brockwell, 1985). In evaluating the above projection for the case when $\alpha \in (0,1)$, the process $\xi_t \in L^{\alpha}(\Omega)$ does not have a unique projection on the closed subspace $S_t(\xi^*)$, $t = 1,2,\ldots,n$ of $L^{\alpha}(\Omega)$. The existence of the h steps ahead forecast is shown in the following Theorem and the dispersion of the forecast error is also given.

Theorem 4.3.2

Let $\overset{*}{\xi}_t$ be the solution to the difference equation given in (4.1) and suppose further that the conditions of Theorem 4.3.1 apply. Then

(a) For $h > 0$, the optimal forecast $\overset{*}{\xi}_t(h)$ (with minimum dispersion) for ξ_{t+h} when the observations $(\xi_1, \xi_2, \ldots, \xi_t)$ are available is obtained as

$$\hat{\xi}^*_t(h) = \sum_{j=h}^{\infty} \psi_j \eta_{t+h-j}, \quad t = 1, 2, .., n$$

$$= \sum_{s=0}^{t-1} \pi_s \xi_{t+h-s}$$

(b) The forecast error is obtained as

$$\omega_{t+h} \;=\; \xi^*_{t+h} - \hat{\xi}^*_t(h) = \sum_{j=0}^{h-1} \psi_j \eta_{t+h-j}$$

with corresponding dispersion as

$$\mathrm{disp}(\omega_{t+h}) = \sum_{j=0}^{h-1} |\psi_j|^{\alpha}\, c(t-j)$$

Proof:

The assumptions that $\xi^*_t \in S_t(\eta)$ and $\eta_t \in S_t(\xi^*)$ imply that the subspaces $S_t(\xi^*)$ and $S_t(\eta)$ are also the same and thus the linear projection $P_{S_t(\xi^*)}\, \xi^*_{t+h} = P_{S_t(\eta^*)}\, \xi^*_{t+h}$. Thus from (4.3.1), we have

$$\xi^*_{t+h} = \sum_{j=0}^{\infty} \psi_j \eta_{t+h-j}.$$

The expression for the forecast error follows from (a) and by use of the results in Theorem 4.3.1. The dispersion of the forecast error simply follows by definition.

As an illustration, we consider the first order autoregressive process ξ_t given by

$$\xi_t = \phi \xi_{t-1} + \eta_t,\; t = 0, \pm 1, \pm 2 \ldots$$

where ϕ is a given parameter with $|\phi| < 1$ and $\{\eta_t\}$ is a sequence of symmetric stable random variables. It is clear that under these conditions, the above process admits an infinite moving average representation of the form

$$\overset{*}{\xi}_t = \sum_{s=0}^{\infty} \psi_s \eta_{t-s}\,, \tag{4.3.2}$$

where ψ_s are the coefficients in the expansion $\psi(B) = 1/\Phi(B)$.

This is basically the solution for ξ_t and consequently from Theorem 4.3.2, the h-steps ahead forecast value when the observations $(\xi_1, \xi_2,..., \xi_n)$ are available is obtained as

$$\overset{*}{\xi}_{n+h} = (1-\phi B)^{-1}\eta_{n+h} = \sum_{s=0}^{\infty} \psi_s \eta_{n+h-s} \tag{4.3.3}$$

where $\psi_0 = 1$.

Suppose we now require to obtain the forecast value $\hat{\xi}_{n+h}$ as a linear function of the observed values, then for the AR(1) case this obtained as

$$\hat{\xi}_{n+h} = \phi^h \xi_n \tag{4.3.4}$$

For this forecast to be optimal in the minimum dispersion sense, we require that ϕ be obtained such that the dispersion, $\mathrm{disp}(\overset{*}{\xi}_{n+h}-\hat{\xi}_{n+h})$ is minimized. By writing (4.3.4) in terms of η_{n-s}, we obtain

$$(\overset{*}{\xi}_{n+h}-\hat{\xi}_{n+h}) = \sum_{s=0}^{\infty} \psi_s \eta_{t+h-s} - \sum_{s=0}^{n-1} a(s)\eta_{n-s}$$

$$= \eta_{n+h} + \sum_{s=1}^{h-1} \psi_s \eta_{n+h-s}) + \sum_{s=0}^{\infty} (\psi_s - a(s))\eta_{n-s})$$

where a(s) are the coefficients in the moving average representation of ξ_n.

Since the scale parameter of η_t is c(t), we have

$$\operatorname{disp}(\xi^*_{n+h}-\hat{\xi}_{n+h}) = c(n+h) + \sum_{s=1}^{h-1} |\psi_s|^\alpha c(n+s) \; + \sum_{s=0}^{\infty} |\psi_s - a(s)|^\alpha \, c(n+s)$$

when the scale parameter is constant say c(t)=1 for all t, the above dispersion function reduces to the result of Cline and Brockwell (1985) which can be obtained from

$$D(\alpha) = \operatorname{disp}(\xi^*_{n+h}-\hat{\xi}_{n+h}) = 1 + \sum_{s=1}^{h-1} |\psi_s|^\alpha + \sum_{s=0}^{\infty} |\psi_s - a(s)|^\alpha \, .$$

This reduces to $(1 - |\phi|^{\alpha h})/(1 - |\phi|^{\alpha})$ in the case of the AR(1) model.

It is easy to see that

$$D(\alpha) \geq 1 + \sum_{s=1}^{h-1} |\psi_s|^\alpha$$

with the equality holding if and only if $\psi_s = a(s)$. This implies that the optimal minimum dispersion forecast for the AR(1) model is given by

$$\hat{\xi}_{n+h} = \phi^h \xi_n.$$

The only evident distinction between this result and that from the finite variance case is the method applied in estimating the parameters ψ_s.

4.4. ARMA Processes With Time Varying Coefficients

The ARIMA models studied by Box and Jenkins (1976) are appropriate for nonstationary processes which have explosive roots. These models take care of the nonstationarity that arises due to time varying first order properties such as the mean but they would not be appropriate when nonstationarity in the second order properties are present since the model coefficients are constant. A more appropriate approach would thus be to consider a more general group of ARMA processes with time varying coefficients. In the finite variance setting, such processes have been studied by Whittle (1965) and Abdrabbo and Priestley (1967). In this section, we consider the infinite variance nonstationary processes with time varying coefficients as given in equation (4.2).

We assume that $\{\eta_t\}$ in equation (4.2) is a sequence of infinite variance stable random variables with characteristic exponent α $(0 < \alpha < 2)$ and constant scale parameter c. We also assume that the autoregressive (AR) and the moving average (MA) operators $\Phi_t(B)$ and $\Theta_t(B)$ respectively are such that the zeroes of the polynomials $\Phi_t(z) = 1 + \phi_{1t}z + \phi_{2t}z^2 + ... + \phi_{pt}z^p$ and $\Theta_t(z) = 1 + \theta_{1t}z + \theta_{2t}z^2 + ... + \theta_{qt}z^q$ lie in the region $|z| >$ for each t.

Let (Ω, F, P) be a probability space with corresponding metric space for all real valued random variables η_t on the probability space being $L^{\alpha}(\Omega)$. Suppose the sequence of random variables η_t is such that $E|\eta_t|^{\lambda} < \infty$. Let ξ_t^* be a solution to the stochastic difference equation (4.2) defined on the probability space and let $H_t(\xi^*)$ and $H_t(\eta)$ be the subspaces of the metric space $L^{\lambda}(\Omega)$ spanned by the random variables ξ_s^* and η_s, for $s \in T = (0, \pm 1, \pm 2, ...)$ and $s \leq t$ respectively. Under the assumptions on the zeroes of $\Phi_t(m)$ and $\Theta_t(m)$, it implies that the two subspaces are equal, i.e., $H_t(\xi^*) = H_t(\eta)$. A stochastic difference equation given by (4.2) thus represents an ARMA(p,q) process if it admits a unique solution $\{\xi_t^*; t \in T\}$ such that $\xi_t^* \in H_t(\eta)$. Thus the conditions for the existence of a solution to the process given by equation (4.2) can be summed up as in the following Theorem:

Theorem 4.4.1

Suppose that for the linear process ξ_t given in equation (4.2), the following further conditions are satisfied;

(a) $E|\eta_t|^{\alpha} < \infty$ for all $t \in T$ and $\alpha \in (0,2)$

(b) $\sum_{s=0}^{\infty} |\psi(t,s)|^{\alpha} < \infty$

where $\psi(t,s)$ are the coefficients from $\Theta_t(B)/\Phi_t(B)$. Then the following moving average process is a solution to equation (4.2).

$$\overset{*}{\xi}_t = \sum_{s=0}^{\infty} \psi(t,s)\eta_s \tag{4.4.1}$$

Proof:

Let the sequence of random variables w_t be given by $w_t = \theta_t(B)\eta_t$. Then the above conditions imply that the dispersion of w_t obtained from $E|w_t|^{\alpha}$ is uniformly bounded on the set $T=(0, \pm1, \pm2, \ldots)$ and thus by the assumption on $\Phi_t(B)$ and $\Theta_t(B)$, there exists a solution $\overset{*}{\xi}_t$ for $\Phi_t(B)\xi_t = w_t$ given by (4.4.1).

Linear Forecasting

The process ξ_t as specified in equation (4.2) is defined in the $L^{\alpha}(\Omega)$ space and thus the linear projections in this space constitute a possible set of forecasts. Such forecasts when evaluated in the $L^2(\Omega)$ space lead to the conditional expectation for Gaussian processes but this is not necessarily the case for the set of infinite variance processes as shown here.

Let ξ^*_t be a unique solution to the process given in equation (4.2) and assume that the conditions of Theorem 4.4.1 are satisfied for $\alpha \in (0, 2)$. Then an optimal linear h-steps ahead forecast of ξ^*_{n+h} for $h > 0$ given the observations $(\xi_1, \xi_2, \dots, \xi_n)$ is obtained as the projection of the space $H(\xi^*)$ onto $H_n(\xi^*)$ (where $H(\xi^*)$ is the subspace spanned by the set of random variables $\{\xi^*_s : s \in T\}$ and $H_n(\xi^*)$ is the observation space) such that the forecast error dispersion is minimized. We can thus write the h-steps ahead optimal forecast $\hat{\xi}^*_n(h)$ of ξ^*_{n+h} as

$$\hat{\xi}^*_n(h) = P_{H_n(\xi^*)}\, \xi^*_{n+h}$$

where

$$\begin{aligned} P_{H_n(\xi^*)}\, \xi^*_{n+h} &= \xi^*_{n+h} \quad \text{if } h \le 0 \\ &= \xi^*_n(h) \quad \text{if } h > 0. \end{aligned}$$

Again we note that for $\alpha \in (0,1)$, the process $\xi_t \in L^{\alpha}(\Omega)$ does not admit a unique projection on the closed subspace $H(.)$ of $L^{\alpha}(\Omega)$ and thus the resulting solution is not unique. The evaluation of the optimum linear forecast is shown in the next Theorem.

Theorem 4.4.2

Suppose ξ^*_t is the solution to (4.2). It then follows as in Theorem 4.3.2, that

(a) For $h > 0$, the optimal forecast (with minimum dispersion) for ξ^*_{n+h} is

$$\xi^*_n(h) = \sum_{s=0}^{\nu} \psi(n+h,s)\eta_s, \qquad \text{where } \eta_s = \xi^*_{s+1} - \hat{\xi}^*_s(1)$$

(b) The forecast error is obtained as

$$\eta_{n+h} = \xi^*_{n+h} - \hat{\xi}^*_n(h) = \sum_{i=0}^{\eta-1} \psi(n+h,n+h-i)\eta_{n+h-i}$$

and the dispersion for the forecast error is obtained as

$$\mathrm{disp}(\eta_{n+h}) = \sum_{i=0}^{h-1} |g(n+h,n+h-i)|^{\lambda}.$$

Proof:

The proof of part (a) is easy to see since under the assumptions that $\xi^*_t \in H_t(\eta)$, i.e., ξ^*_t is measurable in the closed subspace $H_t(\eta)$ of η and similarly $\eta_t \in H_t(\xi^*)$.This implies that $H_t(\xi^*) = H_t(\eta)$ and thus $P_{H_n(\xi^*)}\, \xi^*_{n+h}$ and $P_{H_n(\eta^*)}\, \xi^*_{n+h}$ such that the forecast error dispersion is minimized are equal. Similar arguments as those in Theorem 4.3.2 readily lead to the result. Part (b) follows easily from (a) and the forecast error dispersion simply follows by definition.

Example 1: AR(1) Model

Suppose the process ξ_t is given by

$$\xi_t = \phi(t)\xi_{t-1} + \eta_t,\ t = 0, \pm 1, \pm 2 \ldots$$

where $\{\phi(t)\}$ is a given sequence with $|\phi(t)| < 1$ for all $t \in T$, $\{\eta_t\}$ is an infinite variance stable sequence of random variables having characteristic exponent $\alpha \in (0,2)$. It can then be shown that the solution of ξ_t exists and is given by

$$\overset{*}{\xi}_t = \sum_{s=0}^{\infty} \psi(t,s)\eta_{t-s} \tag{4.4.2}$$

Thus

$$\overset{*}{\xi}_{t+h} = \sum_{s=h}^{\infty} \psi(t+h,s)\eta_{t+h-s}$$

$$= \sum_{s=0}^{t} \psi(t+h,s)\eta_s + \sum_{s=t+1}^{h} \psi(t+h,s)\eta_s \tag{4.4.3}$$

where

$$\psi(t,s) = 1 \qquad s = 0$$

$$= \prod_{i=0}^{s-1} \phi(t-i) \qquad s > 0$$

Suppose that the observations $\{\xi_1, \xi_2, \ldots, \xi_n\}$ are available and a linear forecast $\hat{\xi}_{n+h}$ is required as a linear function of the observed values, this is obtained as

$$\hat{\xi}_{n+h} = \sum_{s=0}^{n-1} a(t,s)\xi_{n-s} \,. \tag{4.4.4}$$

For this forecast to be optimal in the minimum dispersion sense, we require that a(t,s) be obtained such that $\mathrm{disp}(\overset{*}{\xi}_{n+h}-\hat{\xi}_{n+h})$ is minimized. By writing (4.4.4) in terms of η_{n-s}, we obtain

$$(\overset{*}{\xi}_{n+h}-\hat{\xi}_{n+h}) = \sum_{s=0}^{\infty}(\psi(n+h,s) - b(n,s))\eta_{n+h-s}$$

$$= \eta_{n+h} + \phi(t)\eta_{n+h-1} + \ldots + \prod_{i=0}^{h-1}\phi(n-i)\eta_{n+1}$$

$$= \eta_{n+h} + \sum_{s=1}^{h}\prod_{i=0}^{s-1}\phi(n-i)\eta_{n+h-i} + \sum_{s=0}^{\infty}(\psi(n,s)-b(n,s))\eta_{n-s})$$

and the error dispersion is given by

$$\mathrm{disp}(\overset{*}{\xi}_{n+h}-\hat{\xi}_{n+h}) = c+\sum_{s=1}^{h}\left|\prod_{i=0}^{s-1}\phi(n-i)\right|^{\alpha}c+\sum_{s=0}^{\infty}|\psi(n,s)-b(n,s)|^{\alpha}c.$$

Denoting the dispersion as $K(\alpha)$, it is also easy to see that

$$K(\alpha) \geq c + \sum_{s=1}^{h}\left|\prod_{i=0}^{s-1}\phi(n-i)\right|^{\alpha}c$$

with the equality holding if and only if $\psi(t,s) = b(t,s)$. Thus the optimal minimum dispersion forecast can be evaluated as

$$\hat{\xi}_{n+h} = \sum_{s=0}^{n-1} \psi(n,s)\xi_{n-s}.$$

In the case of the first order autoregressive model with fixed parameters, we have $\phi(t) = \phi$ and $\hat{\xi}_n(h) = \phi^h \xi_n$ for all t. An advantage of such a stationary case is thus the considerable reduction in the number of parameters that have to be estimated.

Example 2: ARMA (1,1) Model.

Suppose that the sequence $\{\xi_t\}$ is generated by an ARMA(1,1) process give by the difference equation

$$\xi_t = \phi(t)\xi_{t-1} + \theta(t)\eta_{t-1} + \eta_t$$

where $\{\phi(t)\}$ and $\{\theta(t)\}$ are given sequences with $|\phi(t)| < 1$ and $|\theta(t)| < 1$ for all t, $\{\eta_t\}$ is a sequence of random variables with an infinite variance stable distribution. The above process can be written as

$$\xi_t = [1 + (\phi(t) + \theta(t))B + \{\phi(t)(\phi(t-1) + \theta(t-1))\}B^2$$

$$+ \{\phi(t)\phi(t-1)(\phi(t-2) + \theta(t-2))\}B^3 + \dots]\eta_t$$

$$= \sum_{s=0}^{\infty} \psi(t,s)\eta_{t-s},$$

where

$$\psi(t,s) = 1 \qquad s = 0$$

$$= \phi(t) + \theta(t) \qquad s = 1$$

$$= \prod_{i=0}^{s-2} \phi(t-i)(\phi(t-i-1) + \theta(t-i-1)) \qquad s > 1.$$

It is clear that a solution to the process exists when the conditions of Theorem 4.4.1 apply. An optimal linear forecast would be obtained using a similar approach as for the AR(1) process.

4.4.1. Estimation of model coefficients

A possible estimate for stationary infinite variance linear processes is the least squares estimate whose properties have been discussed by among others Hannan and Kanter (1977) and Yohai and Maronna (1977) for the autoregressive process. Other techniques include the best linear estimate of Blattberg and Sargent (1971) for the regression model and also the least absolute deviation estimate of Gross and Steiger (1979) for the autoregressive process.

In this Section, we discuss the estimation of the parameters in the nonstationary ARMA model given in equation (5.2) by minimizing the norm in analogy with the minimization of the variance in the finite variance setting. First we consider stationary processes with infinite variance and observe their relation to the nonstationary case. When $\alpha \in (1,2)$, the resulting estimates are compared with the optimal estimates discussed in chapter two.

Stationary processes with infinite variance

Consider the AR(1) model given as

$$\xi_t = \phi\xi_{t-1} + \eta_t \qquad (4.4.5)$$

where the sequence of innovations $\{\eta_t\}$ is as defined before and the coefficient ϕ is such that $|\phi| < 1$. Under these conditions, the above AR(1) process is stationary. To estimate the coefficient in the above process, the function $g(\phi,\xi) = E(|\xi_t - \phi\xi_{t-1}|^\alpha)$ may be minimized. Now as an estimate of the parameter, we consider the function $g_n(\phi,\xi)$ given as

$$g_n(\phi,\xi) = \frac{1}{n}\sum_{t=2}^{n}|\xi_t - \phi\xi_{t-1}|^\alpha . \tag{4.4.6}$$

By the egordic theorem, the above function $g_n(\phi,\xi)$ converges almost surely to $g(\phi,\xi)$ as the sample size n tends to infinity. When $1 < \alpha \leq 2$, $g_n(\phi,\xi) = g_n(\phi)$ is convex and thus the resulting estimate would be obtained by differentiating $g_n(\phi)$ with respect to ϕ and then solving the resulting estimating equation for ϕ. This leads to

$$\frac{\alpha}{n}\sum_{t=2}^{n}\xi_{t-1}|\xi_t - \phi\xi_{t-1}|^{\alpha-1} = 0$$

from which we have a solution for ϕ as

$$\phi_1^* = \frac{\sum_{t=2}^{n}\xi_t|\xi_{t-1}|^{1/(\alpha-1)}}{\sum_{t=2}^{n}|\xi_{t-1}|^{\alpha/\alpha-1}}\operatorname{sgn}(\xi_{t-1})$$

where $|x|\operatorname{sgn}(x) = x$. It is to be noted that this result corresponds to the optimal estimate for the AR(1) model discussed in the earlier chapters when the scale parameter c(t) is constant.

In the case when $0 < \alpha \leq 1$, the minimization of $g_n(\phi)$ entails consideration of the points of the function $g_n(\phi)$ in the interval $(0, \phi\xi_{t-1})$. This leads to a second estimate of the coefficient ϕ as

$$\phi_2^* = \sum_{t=2}^{n} \xi_t/\xi_{t-1}.$$

Thus the resulting estimate can be given as

$$\phi^* = \frac{\sum_{t=2}^{n} \xi_t |\xi_{t-1}|^{1/(\alpha-1)}}{\sum_{t=2}^{n} |\xi_{t-1}|^{\alpha/\alpha-1}} \operatorname{sgn}(\xi_{t-1}) \qquad \text{if} \quad 1 < \alpha \leq 2$$

$$= \sum_{t=2}^{n} \xi_t/\xi_{t-1} \qquad \text{if} \quad 0 < \alpha \leq 1.$$

Note that when $\alpha = 1$, the resulting estimate is the least absolute deviation estimate discussed by Gross and Steiger (1979).

We next consider the estimation of the coefficients for the nonstationary case.

Nonstationary processes with infinite variance

Consider an AR(1) process defined in (4.4.5) with the coefficient ϕ being time varying, i.e.,

$$\xi_t = \phi_t \xi_{t-1} + \eta_t$$

such that $|\phi_t| < 1$ for $t = 1,2,\ldots,n$. The set of parameters $\{\phi_t,\ t = 1,2,\ldots,n\}$ are similarly estimated by minimizing $g_n(\phi_t)$ with respect to ϕ_t. When $1 < \alpha \leq 2$, the estimating function is obtained as

$$\alpha \xi_{t-1} |\xi_t - \phi_t \xi_{t-1}|^{\alpha-1} = 0$$

and this leads to the estimator as

$$\phi_t^*(1) = \xi_t/|\xi_{t-1}|^{1/(\alpha-1)}.$$

A similar result would be obtained when the optimality criterion of Chapter 2 is applied. In the case when $0 < \alpha \leq 1$, the resulting estimator is obtained as

$$\phi_t^*(2) = \xi_t/\xi_{t-1}.$$

This type of estimate is not very useful since its error is zero. An alternative estimate would be obtained using the optimality criterion developed in Chapter 2 for $\alpha > 1$. We illustrate this using the random coefficient autoregressive model (Nicolls and Quinn, 1982) which is a special case of a model with time varying coefficients.

The first order random coefficient autoregressive model is given by

$$\xi_t = (\phi + a_t)\xi_{t-1} + \eta_t.$$

The optimal estimate for the parameter ϕ is obtained based on the observations $(\xi_1, \xi_2, ..., \xi_n)$ by noting that

$$h_t = \xi_t - \phi\xi_{t-1} \quad \text{and} \quad a_{t-1}^o = |\xi_{t-1}|^{1/(\alpha-1)}\mathrm{sgn}(\xi_{t-1})/(|\xi_{t-1}|^\alpha + 1)$$

and thus

$$\phi^o = \sum_{t=2}^{n} \frac{|\xi_{t-1}|^{1/(\alpha-1)}\xi_t \mathrm{sgn}(\xi_{t-1})}{1 + |\xi_{t-1}|^\alpha}\left[\sum_{t=2}^{n} \frac{|\xi_{t-1}|^{\alpha/(\alpha-1)}\mathrm{sgn}(\xi_{t-1})}{1 + |\xi_{t-1}|^\alpha}\right]^{-1}$$

Based on this estimate, the error values η_t and a_t can then be estimated as

$$\eta_t^o = \xi_t - \phi^o\xi_{t-1}$$

and

$$a_t^o = \eta_t^o/\xi_{t-1}.$$

The random coefficient $\phi_t = \phi + a_t$ is then obtained as

$$\phi_t^o = \phi^o + a_t^o.$$

CHAPTER V

5. SUMMARY AND FUTURE RESEARCH

In this dissertation, we have developed a criterion for an estimating function to be optimal in the minimum dispersion sense when the generating process is driven by infinite variance stable errors. This has then been applied to the estimation of parameters for the autoregressive and moving average processes. The resulting simulation results indicate that for the autoregressive process of order one, the optimal estimate converges as the sample size increases and is superior to the least squares estimate when the sample size, the index α and the parameter to be estimated are small. In the case of the moving average process, the optimal estimate performs better and converges faster than the least squares estimate.

Three techniques for estimation of missing observations were also developed. The first one based on the concept of optimal estimating functions is limited to the case when $\alpha > 1$. An advantage of this technique is that simple analytical forms of the estimate can be obtained even when the order of the model is high. The second method for evaluating linear estimates of missing observations is functionally different from the optimal estimate although at $\alpha = 2$, its truncated form is the same as the least squares estimate. The third method is based on the minimization of the lambda norm and the biggest drawback for it is that its analytical form cannot be obtained when the order of the model is more than one. It thus requires the use of minimization algorithms.

In the third part of the dissertation, properties of nonstationary processes have been studied. Particular interest was focussed on three forms of nonstationarity. The convergence and the limit distribution of the optimal estimate for the AR(1) process having a unit root was established. Conditions for the existence of a solution for a process with innovations that have a nonconstant scale parameter were given. Processes that have varying coefficients were also studied and conditions under which a solution exists were developed. The estimation of nonconstant coefficients was also discussed.

The development of practical inferential techniques for analysing time series models with infinite variance stable errors is still in its infant stage and many areas need further research. We have assumed in this dissertation that the process has already been identified as an infinite variance process and the model order is given but this is not usually the case in practice. Thus the development of objective techniques for distinguishing between finite variance and infinite variance processes as well as methods for determining the order of a model for an infinite variance process are necessary. Another area that requires further research is the extension of the optimality criterion of an estimating function to the case when $\alpha \in (0,1)$. Finally the application of the techniques developed here to real data would be of great interest for further investigation.

BIBLIOGRAPHY

Abdrabbo,N.A. and M.B. Priestley (1967). On the prediction of nonstationary processes. *JRSS*. B 29, 570-585.

Abraham, B. and A.Thavaneswaran (1991). A nonlinear Time series model and estimation of missing observations. *Ann. Inst. Stat. Math.* **43**, 493-504.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statis. Math.,* 21, 243-247.

Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Statis. Math.,* 22, 203-217.

Basu,A.K.and S.S.Roy (1989). On asymptotic prediction problems for autoregressive models with explosive and other roots. *Calcutta Statistical Association Bulletin,* 38, 149 - 150.

Bhansali, R.J. (1988). Consistent order determination for processes with infinite variance. *JRSS, Ser. B*, 50, 46-60.

Bhansali, R.J. and D.Y. Downham (1977). Some properties of the order of an autoregressive model selected by generalization of Akaike's EPF Criterion. *Biometrika* , 64, 547-551.

Bhapkar, V.P (1972). On a Measure of Efficiency of an Estimating Equation. *Sankya., Ser A.,* 34, 467-472

Blattberg, R. and T. Sargent (1971). Regression with non gaussian stable disturbances: Some sampling results. *Econometrika* , 39, 501-510.

Box,G.E.P. and D.R. Cox (1964). An analysis of transformations. JRSS, Ser. B, 26, 211 - 252

Box, G.E.P. and G.M. Jenkins (1976). *Time Series: Forecasting and Control.* San Francisco. Holden Day.

Brockwell, P.J. and R.A. Davis (1987). *Time series: Theory and methods*.Springer Verlag. New York.

Cambanis, S. and G. Miller (1981). Linear problems in p^{th} order and stable processes. *SIAM J. Appl. Math.*, 41, 43 - 69.

Cambanis, S., Hardin Jr, C.D.and A. Weron (1988). Innovations and Wold decomposition of stable processes. *Prob. Th. Rel. Fields* , 79, 1- 27.

Cambanis, S and A.R Soltani (1982). Prediction of stable processes. Spectral and moving average representations. *Tech. Rept. 11, Centre for stochastic processes,* University of North Carolina,Chapel Hill.

Chan, N.G. (1990). Inference for near-intergrated time series with infinite variance. *JASA* 85, 1069-1074.

Chan, N.G. and L.T. Tran (1989). On the first order autoregressive process with infinite variance. *Econometric Theory* , 5, 354 - 362.

Chatfield, C. (1979). Inverse autocorrelations. *JRSS, Ser A,* 142, 363-377.

Chatterji, S.D. (1969).An L^p - convergence Theorem. Ann. Math. Stats., 40, 1068-1070.

Cline, D.B.H. (1983). Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Ph.D. dissertation. Colorado State University.

Cline, D.B.H. (1988). Joint stable attraction of two sums of products. *J. Multivariate Analy.* , 25, 272 - 285.

Cline,D.B.H. and P.J. Brockwell (1985). Linear Prediction of ARIMA processes with infinite variance. *Stoc. Processes and their Appl.* , 19, 281-296.

Cline,D.B.H. (1989). Consistency for least squares regression estimators with infinite variance data. *Journal of Statis. Plann. and Inferen.,* 23, 163-179.

Davis, R.A. and S. Resnick (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statisti.*, 14, 533 - 558.

De Haan, L. (1970). *On regular variation and its application to weak convergence of sample extremes*. Mathematisch Centrum, Amsterdam.

De Silva, B.M (1978). A class of multivariate symmetric stable distributions. *J. Multi. Anal.* 8, 335-345.

Dumouchel,W.H. (1973). On the asymptotic normality of the maximum likelihood estimate when sampling from a stable distribution. *Annals Statist.* , 1, 948 - 957.

Durbin,J. (1960). Estimation of parameters in time series regression models. *JRSS, Ser. B*, 22, 139 - 153.

Fama,E. (1965). The behavior of stock market prices. *J. Business* , 38, 34-105.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2. Wiley.

Fuller, W.A. (1976). *Introduction to statistical time series.* John Wiley. New York.

Ferreiro,O. (1987). Methodologies for the estimation of missing observations in time series. *Statist. Probab. Lett.,* 5, 65 - 69.

Gaver, D.P. and P.A.W. Lewis (1980). First order Gamma sequences and point processes. *Adv. Appl. Prob.,* 12, 727 - 745.

Godambe, V.P. (1960). An optimum property of regular maximum likelihood equations. *Ann Math Stat.,* 31, 1208-1211.

Godambe, V.P. (1976). Conditional likelihood and unconditional optimal estimating equations. *Biometrika,* 63, 277 - 284.

Godambe, V.P.(1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* , 72, 319-328.

Godambe, V.P. and M. Thompson (1974). Estimating equations in the presence of a nuisance parameter. Ann. Statists. 2, 568-571.

Godambe, V.P. and M. Thompson (1984). Robust estimation through estimating equations. *Biometrika,* 71, 115 - 125.

Granger,C.W.J. and A.P.Andersen (1978). *An introduction to bilinear time series models.* Vandenhoeck and Ruprecht. Gottingen.

Granger,C.W.J. and D. Orr (1972). Infinite variance and research strategy in time series analysis. *JASA.*, 67, 275-285.

Griffiths, R.C (1972). Linear dependence in bivariate distributions. *Austral. J. Statis.*, 14, 182-187.

Gross, S. and W.L. Steiger (1979). Least absolute deviation estimates in autoregression with infinite variance. *J. Appl. Prob.*, 16, 104-116.

Haggan, V. and T. Ozaki (1981). Modelling non-linear random vibrations in using an amplitude - dependent autoregressive time series model. *Biometrika*, 68, 189 - 196.

Hannan, E.J. and M. Kanter (1977). Autoregressive processes with infinite variance. *J. Appl. Prob.*, 14, 411-415.

Jacobs, P.A. and P.A.W. Lewis (1977). A mixed autoregressive moving average exponential sequence and point processes (EARMA 1,1). *Adv. Appl. Prob.*, 9, 87 - 104.

Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics, 22, 389-395.

Kanter,M. (1972). Linear sample spaces and stable processes. *J. Funct. Analy.*, 9, 441 - 456.

Kanter, M. and W.L. Steiger (1974). Regression and autoregression with infinite variance. *Adv. Appl. Prob.*, 6, 768-783.

Knight,K. (1989). Limit theory for autoregressive parameter estimates in an infinite variance random walk. *Canadian J. Statist.*, 17, 261 - 278.

Knight,K. (1990). Consistency of Akaike,s information criterion for infinite variance autoregressive processes. *Annals Statist.*, 17, 824 - 840.

Lawrence, A.J. and P.A.W. Lewis (1980). The exponential autoregressive moving average EARMA(p,q) process. *JRSS, Ser. B*, 42, 150 - 161.

Liu,J. (1987). Infinite variance and nonlinear time series models. Unpublished Ph.d. dissertation. Colorado State University.

Liu, J. (1989). A simple condition for the existence of some stationary bilinear time series. *J. Time Series Anal.*, 10, 33-39.

Logan, B.F., Mallows, C.L., Rice, S.O. and L.A. Shepp (1973). Limit distribution of self normalized sums. *Annals of Prob.*, 1, 788 - 809.

Lukacs,E. and R. Laha (1964). Applications of characteristic functions. Hanfer Publishing Company. New York.

Mandelbrot, B. (1963). The variation of speculative prices. *J. Business* , 36, 394-419.

Miamee,A.G. and Pourahmadi,M. (1988). Wold decomposition, prediction and parametrization of stochastic processes with infinite variance.*Prob. Theory Rel. Fields.,* 79, 145-164.

Miller, H.D. (1967). A note on sums of independent random variables with infinite first moment. *Ann. Math. Statist.,* 38, 751-758.

Miller R.B. and Ferreiro,O. (1984). A strategy to complete a time series with missing observations. Lecture notes in Statistics Vol. 25. *Springer verlag*. New York.

Nassiuma, D. (1991). A note on nonstationary ARMA processes with infinite variance. *Commun. in Statist. Theor.* 20, 1832-1836.

Nassiuma, D. (1992). Nonstationary ARMA processes with infinite variance. *J. Time Series Analysis* (To appear).

Nassiuma, D. and N. Ordhukhani (1991). Forecasting of some transformed series. *Commun. in Statist. Theor.,* 20, 3207 - 3220.

Nassiuma, D. and A. Thavaneswaran (1992). Smoothed estimates for nonlinear time series models with irregular data. *Commun. Statists. Theory,* 21 (to appear).

Nicolls,D.F. and B.G. Quinn (1982). *Random coefficient autoregressive models: An introduction. Lecture notes in statistics No. 11.* Springer. New York.

Niemi,H. (1983). On effects of a nonstationary noise on ARMA models. *Scand. J. Statist.,* 10, 11-17.

Onishchenko, V.F. (1990). Estimation of parameters of the moving average in the case of infinite variance. *SIAM: Theory of Prob. and its Appli.*, 34, 335-340.

Parzen,E. (ed) (1984). *Proceedings of time series analysis of irregularly observed data. Lecture notes in statistics No. 25*. Springer. New York.

Paulson,A.S; E.W. Holcomb and R.A. Leitch (1975). The estimation of the stable laws. *Biometrika* , 62, 163-170).

Phillips, P.C. (1990). Time series regression with unit root and infinite variance. *Econometric Theory*, 6, 44 - 62.

Pourahmadi, M (1984). On minimality and interpolation of harmonizable stable processes. *SIAM J. Appl. Math.*, 44, 1023 - 1030.

Pourahmadi, M (1989). Estimation and interpolation of missing values of a stationary time series. *J. Time Series Anal.*, 10, 149-169).

Press, S.J (1972a). Estimation of univariate and multivariate stable distributions. *JASA* , 67, 842-846.

Press, S.J (1972b). Multivariate stable distributions. *J. Multi. Anal.*, 2, 444-462.

Priestley,M.B. (1980). State dependent models: A general approach to time series analysis. *J. Time ser. Analy.*, 1, 47 - 71.

Priestley,M.B. (1981). *Spectral analysis and time series, Vol. I and II*. Academic Press. London.

Priestley,M.B. (1988). Nonlinear and Nonstationary Time series analysis. *Academic Press*. London.

Rosenfeld, G. (1976). Identification of time series with infinite variance. Appl. Statist., 25, 147 - 153.

Shiryayev (1984). *Probability. Graduate texts in Mathematics. Vol. 95*. springer. New York.

Stuck, B.W., (1978). Minimum error dispersion linear filtering of scaler symmetric stable processes. *IEEE Trans. A.C.*, 23, 507-509.

Stuck,B.W. and B. Kleiner (1974). A statistical analysis of telephone noise. *The Bell System Technical Journal,* 53, 1263-1320.

Thavaneswaran, A and B. Abraham (1988). Estimation for nonlinear time series models using estimating equations. *J. of Time series Anal.,* 9, 99-108

Thavaneswaran, A and M. Thompson (1991). Filtering and Interpolation with infinite variance. (Preprint).

Tjosteim,D. (1986). Estimation in nonlinear time series models. *Stoch. Processes Appl.,* 21, 251 - 273.

Tong, L. (1983). Threshold models in non-linear time series analysis. Lecture notes in Statistics, Vol. 21. Springer Verlag.

Whittle,P. (1965). Recursive relations for predictors of nonstationary processes. *JRSS , Ser. B,* 27, 523-532.

Yohai,V.J. and R.A. Maronna (1977). Asymptotic behavior of least squares estimates for autoregressive processes with infinite variances. *Ann. Statis.,* 5, 554-560.

Zolotarev, V.M. (1986). *One Dimensional Stable Distribution.* American Math. Society.

APPENDIX

1. Generation of observations for the AR(1) process

```
// JOB
// EXEC FORT7CLG
      INTEGER      NR
      REAL    PHI,X(510),ALPHA,BPRIM,R(510)
      DOUBLE PRECISION DSEED1
      NR=511
C  INPUT THE MODEL PARAMETER.
      PHI=0.4
      X(0)=0
      DSEED1=123457.D0
C  INPUT THE INDEX AND THE SCALE PARAMETER.
      ALPHA=1.2
      BPRIM=0
      CALL GGSTA(DSEED1,ALPHA,BPRIM,NR,R,S)
      DO 1200 I=1,NR
C  GENERATE VALUES FOR THE AR(1) MODEL.
      X(I+1)=PHI*X(I)+R(I+1)
      WRITE(*,*) X(I+1)
1200  CONTINUE
      STOP
      END
```

2. Generation of observations for the MA(1) process

```
// JOB
// EXEC FORT7CLG
      INTEGER      NR
      REAL    Z,X(649),ALPHA,BPRIM,R(650)
      DOUBLE PRECISION DSEED1,DSEED2
      NR=511
C     INPUT THE MA COEFFICIENTTHETA.
      THETA=0.6
      X(1)=0
      R(1)=0
      DSEED1=123457.D0
C     INPUT THE INDEX AND SCALE PARAMETER.
      ALPHA=0.5
      BPRIM=0
      CALL GGSTA(DSEED1,ALPHA,BPRIM,NR,R)
      DO 1200 I=1,NR
C     GENERATE VALUES FOR THE MA(1) MODEL.
      X(I+1)=THETA*R(I)+R(I+1)
      WRITE(*,*) X(I+1)
1200  CONTINUE
      STOP
      END
```

3. Generation of observations for the ARMA(1,1) process

```
// JOB
// EXEC FORT7CLG
      INTEGER      NR
      REAL     PHI,X(510),ALPHA,BPRIM,R(510)
      DOUBLE PRECISION DSEED1
      NR=511
C     INPUT THE MODEL PARAMETERS PHI AND THETA.
      PHI=0.4
      THETA=0.3
      X(0)=0
      R(1)=0
      DSEED1=123457.D0
C     INPUT THE INDEX AND SKEWNESS PARAMETER.
      ALPHA=0.4
      BPRIM=0
      CALL GGSTA(DSEED1,ALPHA,BPRIM,NR,R,S)
      DO 1200 I=1,NR
      X(I+1)=PHI*X(I)+THETA*R(I)+R(I+1)
      WRITE(*,*) X(I+1)
 1200 CONTINUE
      STOP
      END
```

4. Estimation of the parameter for the AR(1) process

```
1. // JOB 'T=10M'
2. // EXEC FORT7CLG
3.      INTEGER      NR,N
4.      REAL    PHI,X(1202),ALPHA,BPRIM,R(1202),S(1202),E4
5.      REAL    Y(1202),W(1202),L(1202),C(1202),E1,E2,E3,U4
6.      REAL    Q(1202),U1,U2,U3,D(1202),V1,V2,V3,V4,V5,V6
7.      REAL    OP1,OP2,OP3,OP4,LS1,LS2,LS3,LS4,V7,V8,ALP
8.      DOUBLE PRECISION DSEED1
9. C    INPUT THE SAMPLE SIZE NR,NUMBER OF SAMPLES N AND
9.1 C   PARAMETER PHI.
10.     NR=1202
11.     N=500
12.     PHI=0.1
13.     OP4=0
14.     LS4=0
15.     V1=0
16.     V2=0
17.     OP1=0
18.     OP2=0
19.     OP3=0
20.     LS1=0
21.     LS2=0
22.     LS3=0
23.     V3=0
24.     V4=0
```

```
      V5=0
      V6=0
      V7=0
      V8=0
C GENERATING 500 SAMPLES EACH OF SIZE N FROM A STABLE
C DISTRIBUTION WITH PARAMETER ALPHA.
C
      DO 1700 V=1,N
      X(1)=0
      Y(200)=0
      W(200)=0
      L(200)=0
      C(200)=0
      D(200)=0
      DSEED1=123457.D0
      ALPHA=1.2
      ALP=1.2
      BPRIM=0
      CALL GGSTA(DSEED1,ALPHA,BPRIM,NR,R,S)
      DO 2000 K=1,NR
      S(K)=ABS(R(K))
C GENERATING VALUES FOR THE AR(1) MODEL WITH PARAMETER
C PHI.
       X(K+1)=PHI*X(K)+R(K+1)
2000  CONTINUE
      DO 2400 I=200,1200
      Q(I)=ABS(X(I))
```

```
      Y(I+1)=Y(I)+(X(I+1)*X(I)/Q(I))*(Q(I))**(1/(ALP-1))
      W(I+1)=W(I)+(Q(I))**(ALP/(ALP-1))
      L(I+1)=L(I)+(X(I))*X(I+1)
      C(I+1)=C(I)+(X(I))*X(I)
      D(I+1)=D(I)+X(I)
 2400 CONTINUE
      E1=Y(251)/W(251)
      E2=Y(351)/W(351)
      E3=Y(701)/W(701)
      E4=Y(1201)/W(1201)
      U1=L(251)/C(251)
      U2=L(351)/C(351)
      U3=L(701)/C(701)
      U4=L(1201)/C(1201)
C EVALUATING OPTIMAL AND LEAST SQUARES ESTIMATES
C OF THE PARAMETER FOR SAMPLE SIZES 50,150,500 AND 1000.
      OP1=OP1+E1/N
      OP2=OP2+E2/N
      OP3=OP3+E3/N
      OP4=OP4+E4/N
      LS1=LS1+U1/N
      LS2=LS2+U2/N
      LS3=LS3+U3/N
      LS4=LS4+U4/N
      V1=V1+((E1-PHI)**2)/N
      V2=V2+((E2-PHI)**2)/N
      V3=V3+((E3-PHI)**2)/N
```

```
      V4=V4+((E4-PHI)**2)/N
      V5=V5+((U1-PHI)**2)/N
      V6=V6+((U2-PHI)**2)/N
      V7=V7+((U3-PHI)**2)/N
      V8=V8+((U4-PHI)**2)/N
 1700 CONTINUE
C     OPTIMAL ESTIMATES.
      WRITE(*,*) OP1,OP2,OP3,OP4
C     MEAN SQUARE ERRORS OF OPTIMAL ESTIMATES.
      WRITE(*,*) V1,V2,V3,V4
C     LEAST SQUARES ESTIMATES.
      WRITE(*,*) LS1,LS2,LS3,LS4
C     MEAN SQUARE ERRORS OF LEAST SQUARES ESTIMATES.
      WRITE(*,*) V5,V6,V7,V8
      STOP
      END
```

5. Estimation of the parameter for the MA(1) process

```
// JOB 'T=10M'
// EXEC FORT7CLG
      INTEGER      NR,N
      REAL   THETA,X(1202),ALPHA,BPRIM,R(1202),S(1202),E4
      REAL   Y(1202),W(1202),L(1202),C(1202),E1,E2,E3,U4
      REAL   Q(1202),E(1202),U(1202),U1,U2,U3,D(1202),V1
      REAL   A1(1203),LS(1203),F(1203),V2,V3,V4,V5,V6,V7
      REAL   V8,OP1,OP2,OP3,OP4,LS1,LS2,LS3,LS4,ALP
      DOUBLE PRECISION DSEED1
C INPUT THE SAMPLE SIZE NR AND THE NUMBER OF SAMPLES N.
      NR=1202
      N=500
C  INPUT THE MODEL PARAMETER THETA.
      THETA=0.8
      OP4=0
      LS4=0
      V1=0
      V2=0
      OP1=0
      OP2=0
      OP3=0
      LS1=0
      LS2=0
      LS3=0
      V3=0
```

```
      v4=0
      V5=0
      V6=0
      V7=0
      V8=0
      DO 2000 J=1,N
      R(1)=0
      X(1)=0
      Y(200)=0
      W(200)=0
      L(200)=0
      C(200)=0
      D(200)=0
      F(200)=0.1
      DSEED1=123457.D0
C  INPUT THE EXPONENT ALPHA.
      ALPHA=1.4
      ALP=1.4
      BPRIM=0
      CALL GGSTA(DSEED1,ALPHA,BPRIM,NR,R,S)
      DO 1200 K=1,NR
      S(K)=ABS(R(K))
C  GENERATE VALUES FOR THE MA(1) AND OBTAIN THE LS ESTIMATE.
      X(K+1)=THETA*R(K)+R(K+1)
 1200 CONTINUE
      DO 1000 I=200,NR
      A1(I)=ABS(X(I))
```

```
      Q(I)=ABS(F(I))
      L(I+1)=L(I)+(X(I))*X(I+1)
      C(I+1)=C(I)+(X(I))*X(I)
      D(I+1)=D(I)+X(I)
      LS(I+1)=L(I+1)/C(I+1)
C EVALUATE THE ERROR ESTIMATE F.
      F(I+1)=X(I+1)-LS(I+1)*F(I)
      Y(I+1)=Y(I)+(X(I+1)*F(I)/Q(I))*Q(I)**(1/(ALP-1))
      W(I+1)=W(I)+(Q(I))**(ALP/(ALP-1))
1000  CONTINUE
C OBTAIN THE OPTIMAL AND LS ESTIMATES.
      E1=Y(251)/W(251)
      U1=L(251)/C(251)
      E2=Y(351)/W(351)
      U2=L(351)/C(351)
      E3=Y(701)/W(701)
      U3=L(701)/C(701)
      E4=Y(1201)/W(1201)
      U4=L(1201)/C(1201)
      OP1=OP1+E1/N
      OP2=OP2+E2/N
      OP3=OP3+E3/N
      OP4=OP4+E4/N
      LS1=LS1+U1/N
      LS2=LS2+U2/N
      LS3=LS3+U3/N
      LS4=LS4+U4/N
```

```
      V1=V1+((E1-THETA)**2)/N
      V2=V2+((E2-THETA)**2)/N
      V3=V3+((E3-THETA)**2)/N
      V4=V4+((E4-THETA)**2)/N
      V5=V5+((U1-THETA)**2)/N
      V6=V6+((U2-THETA)**2)/N
      V7=V7+((U3-THETA)**2)/N
      V8=V8+((U4-THETA)**2)/N
2000  CONTINUE
C OPTIMAL ESTIMATES OF THETA AND THEIR MEAN SQUARE ERRORS
      WRITE(*,*) OP1,OP2,OP3,OP4
      WRITE(*,*) V1,V2,V3,V4
C     LS ESTIMATES OF THETA AND THEIR MEAN SQUARE ERRORS
      WRITE(*,*) LS1,LS2,LS3,LS4
      WRITE(*,*) V5,V6,V7,V8
      STOP
      END
```