MODELING AND ANALYSIS OF SOME HEAVY TAILED TIME SERIES

Thesis Submitted to the
Cochin University of Science and Technology
for the Award of the Degree of

Doctor of Philosophy

under the Faculty of Science

by

Hareesh G.

Department of Statistics

Cochin University of Science and Technology

Cochin-682022

AUGUST 2010

To My Teachers

CERTIFICATE

Certified that the thesis entitled **"Modeling and Analysis of Some Heavy Tailed Time Series"** is a bonafide record of work done by Shri. Hareesh G. under my guidance in the Department of Statistics, Cochin University of Science and Technology and that no part of it has been included anywhere previously for the award of any degree or title.

Cochin- 22,
27 August 2010.

Dr. N. Balakrishna
Professor,
Department of Statistics,
Cochin University of
Science and Technology.

DECLARATION

This thesis contains no material which has been accepted for the award of any other Degree or Diploma in any University and to the best of my knowledge and belief, it contains no material previously published by any other person, except where due references are made in the text of the thesis.

Cochin- 22, **Hareesh G.**

27 August 2010.

Acknowledgments

I hereby acknowledge the people whose involvement, direct or indirect, helped in this thesis seeing the light of the day.

I wish to express my deep sense of respect and gratitude to my supervising guide and Head of the Department, Professor N. Balakrishna, who has been a constant source of inspiration during the course of my Ph.D work. He has always been patient towards my shortcomings and kept encouraging me to work in a better way. Without his help and support, perhaps, I would have not been able to write this thesis.

I take this opportunity to record my sincere respect and heartiest gratitude to H.V. Srinivasa Rao, Director, ISSA, who had kindly agreed to continue my Ph.D work. Without his kind cooperation, it would have been almost impossible for me to continue research work with the job.

I am obliged to Professor V.K. Ramachandran Nair and K.R. Muraleedharan Nair former Heads, Department of Statistics, Cochin University of Science and Technology (CUSAT) for their extensive support.

I wish to express my sincere thanks to all faculty members of Department of Statistics, CUSAT for their valuable comments and encouragement. I offer my regards to all non-teaching staff of the Department of Statistics, CUSAT for their kind cooperation.

I also express my regards to all officers and staff of ISSA for their kind support.

I express my gratitude to all my family members who have provided me their helping hands in each and every walk of my life.

I owe my appreciation and thankfulness to Council for Scientific and Industrial Research (CSIR), Government of India, for providing me financial support during the initial phase of my Ph.D. at CUSAT.

I owe a debt of gratitude to all my teachers, research scholars and friends who influenced me, a gratitude that can never be paid back in totality, and perhaps be paid back only in part, if I can impart some of what I have learned over the years.

Hareesh G.

Contents

Chapter 1

Chapter 1. Introduction

The contents of this thesis are on various aspects of modeling and analysis of finite mean time series with symmetric stable distributed innovations. Classical time series analysis, generally known as Box and Jenkins time series approach, includes model identification, parameter estimation, diagnostic checking and forecasting (for details see Box et al. (1994) and Brockwell and Davis (1987)). Mathematical theory of classical time series analysis is based on the assumption that the error variances are finite. In recent years there is a great deal of attention in modeling non-Gaussian time series which includes time series with heavy tailed innovations. Symmetric stable distributions are widely used to model heavy tailed variables as stated by Adler et al. (1998), Gallagher (2001), Shao and Nikias (1993). In many practical instances, communication (Stuck and Kleiner (1974)), economics and finance (Fama (1965)), network traffic (Willinger et al. (1998)), tele traffic (Resnick (1997)), data shows sharp spikes or occasional bursts of outlying observations. Heavy tailed distributions can be used to model such series and stable distribution is a good candidate in the family of heavy tailed distributions. Stable distributions are widely used in signal processing especially modeling impulsive signal (see Shao and Nikias (1993)). A broad and increasingly important class of non Gaussian phenomenon encountered in practice can be characterized by its impulsive nature. Signals and noise in this class are more likely to exhibit sharp spike or occasionally bursts of

outlying observations than one would expect from normally distributed signals. As a result, their density functions decay in the tail less rapidly than the Gaussian density function. Underwater acoustic signals, low frequency atmospheric noise and man-made noise have all been found to belonging to this class, (Nikias and Shao (1995)). It is for this type of signals that stable distribution provides a useful theoretical tool. The stable law is a direct generalization of the Gaussian distribution and in fact includes Gaussian as a special case. The tail of stable density is heavier than that of Gaussian density. Stable distribution is characterized by four parameters: $\alpha \in (0, 2]$, measuring the tail thickness (thicker tails for smaller values of the parameter), $\theta \in [-1, 1]$ determining the degree and the sign of asymmetry, $\gamma > 0$ (scale) and $\beta \in R$ (location). To denote stable distribution with parameters α, θ, γ and β we will use the notation $S_\alpha(\beta, \theta, \gamma)$. In addition stable distribution is very flexible as a modeling tool in that its parameter α $(0 < \alpha \leq 2)$, that controls the heaviness of its tails. A smaller positive value of α indicates several impulsiveness, while a value of α close to 2 indicates a more Gaussian type of behavior. We can find a lot of applications of symmetric stable distributions in time series modeling, (see Adler et al. (1998), Nikias and Shao (1995), Gordon et al. (2003)) eg: intensity and duration of rainfalls analyzed in environmetrics, activity time of CPUs and network traffic or noise in degraded audio samples in engineering, impulsive signal and noise modeling etc. Now we will discuss some motivating examples to illustrate the applications of stable distributions in modeling some real data.

Economics and Finance:

(i) Nolan (1999) used stable distributions to model daily exchange rate data for 15 different currencies which were recorded (in U.K. pounds) over a 16 year period (2 January 1980 to 21 May 1996). The data was transformed by $y_t = \ln(x_{t+1}/x_t)$ giving $n = 4,274$ data values.

(ii) McCulloch (1997) analyzed forty years (January 1952-December 1992) of monthly stock price data from the Center for Research in Security Prices (CRSP). The data set consists of 480 values of the CRSP value weighted stock index, including

dividends, and adjusted for inflation. The data was analyzed using stable distribution.

(iii) Buckle (1995) fitted a stable distribution for a return series on Abbey National Shares.

(iv) Qiou and Ravisankar (1998) fitted a second order autoregressive model with stable innovations to study the real data set which consists of 394 observations on daily stock prices of a retail store.

Radar Noise:

(i) Nolan (1999) fitted a stable distribution for the in phase components of sea clutter radar noise. This is a very large data set with $n = 320,000$ pairs of data points.

(ii) Lagha and Bensebti (2007) used stable distributions to model the weather precipitation echoes detected by a weather pulse Doppler radar.

Environmetrics:

(i) Pierce (1997) proposed positive alpha stable distributions to model inherently positive quantities such as energy or power. One example he uses is the power in ocean waves (hourly wave data obtained from National Oceanographic and Atmospheric Administration (NOAA) web site) which is proportional to the square of the wave height.

(ii) Gallagher (2001) fits stable auto-regressive model to global sea surface temperature (SST) data.

Signal processing:

(i) Kidmose (2000) shows that class of stable distributions provides a better model for audio signals, than the Gaussian distributed model.

(ii) Kidmose (2000) shows that class of stable distributions provides a better model for audio signals, than the Gaussian distributed model.Tsakalides and Nikias (1998) studied the direction of arrival (DOA) estimation based on stable assumption.

Image Processing:

(i) Tsakalides et al. (2001) considered symmetric alpha stable distributions for modeling the wavelet transform coefficients of sub band images.

(ii) Achim et al. (2001) employed stable distribution for the removal of speckle noise in synthetic aperture radar (SAR) images.

Aerospace Applications:

(i) Gordon et al. (2003) used stable innovations models and Kalman- Levy filter for tracking manoeuvering targets.

Model identification in Gaussian time series analysis is generally carried out using autocorrelation and partial autocorrelation functions. Autocorrelation and partial autocorrelation cannot be defined in stable processes due to the non-existence of second order moments. This also prevents us from defining power spectral density, which is a classical tool for frequency domain analysis of time series. Mathematical theory of Gaussian time series is matured but the corresponding theory in the heavy tailed time series is in its infant stage. In order to develop a theory for stable processes we have to utilize some other dependency measures which can be well defined in this context. We have to explore the applicability of these alternative measures to handle the problems of model identification, parameter estimation and forecasting.

Time series models are widely used in various applications in science and engineering. In many applications the observed time series is considered as a signal plus noise model or observed time series is an actual time series plus a measurement noise. Another important objective of classical time series analysis is to extract signal and noise component from a signal plus noise model. Wiener Kolmogorov filtering and Kalman filtering are the

popular classical methods used for this purpose. But both these methods require finite second order moments. So we need some parallel filtering techniques for infinite variance models. These limitations motivated us to develop a generalized signal extraction filter for heavy tailed processes. Signal extraction filters entail the knowledge of signal and noise parameters. Estimation of signal and noise parameters from an observed signal under heavy tailed assumption is another important problem in this context.

1.1 Outline of the thesis

In Chapter 2 we provide a theoretical background for our proposed study. Here we describe the time series analysis in both finite variance and infinite variance set up. In finite variance case we assume the time series models with Gaussian innovation distribution and in infinite variance case we assume stable innovation distribution. In the present chapter we surveyed the theoretical developments for time series analysis based on stable assumptions and organized them parallel to the developments in classical set up. We can see the limitations of classical time series methods under stable assumptions. Alpha stable distributions and time series models with symmetric stable distributed innovations are discussed in this chapter. Here we introduce another concept known as the *tail covariance* (for details see Sornette and Ide (2001), Bouchaud et al. (1998)), which is a generalized measure of covariance in multivariate stable distributions with heavy tailed index, $\alpha < 2$. Linear prediction theory for infinite variance processes are also discussed in this chapter. Another important tool for analyzing stable time series data is the *auto-covariation function*. In this chapter we explore the application of auto-covariation function and sample auto-covariation function in time series analysis. Generalized Yule-Walker equations based on auto-covariation function are also discussed in this chapter.

Chapter 3 is devoted to study the properties of signal extraction models under the assumption that signal/noise are generated by symmetric stable processes. The optimum

filter is obtained by the method of minimum dispersion discussed by Cline and Brockwell (1985). The problem can be stated as below.
The observed data process Y_t is often depicted as a combination of signal X_t and noise N_t as follows:

$$Y_t = X_t + N_t, t = 1, 2, \ldots . \tag{1.1}$$

The signal X_t and noise N_t are assumed to follow stationary autoregressive moving average of order (p, q)($ARMA(p, q)$) models with symmetric stable innovation distributions. Further X_t and N_t are assumed to be independent of each other. The objective here is to use the data on Y_t to estimate the unobserved component series X_t and N_t . Signal and noise can be estimated by applying a linear filter $W(B)$ to the observed signal Y_t as follows

$$\begin{aligned} \hat{X}_t &= W(B)Y_t\,, \\ \hat{N}_t &= Y_t - \hat{X}_t = (1 - W(B))Y_t, \end{aligned} \tag{1.2}$$

where $W(B) = \sum_j w_j B^j$ and B is the back-shift operator.
Signal extraction error, ζ_t can be defined as,

$$\zeta_t = X_t - \hat{X}_t = (1 - W(B))X_t - W(B)N_t.$$

Signal extraction procedure consists of finding an optimal filter which minimizes the signal extraction error. In finite variance case optimal filter is the one which minimizes the mean square error, where as in the case of symmetric stable process we propose minimum dispersion criteria. For a finite mean process, the optimal filter weights, w_j which minimizes the error dispersion is the solution of the system of equations,

$$\frac{\partial Disp(\zeta_t)}{\partial w_k} = 0, \quad k = 0, 1, 2, ..., . \tag{1.3}$$

The proposed filter has been generalized to doubly infinite and asymmetric filters studied

in the literature for finite variance processes. We introduce a finite length filtering algorithm based on Kalman-Levy filtering discussed by Sornette and Ide (2001). This can be considered as an improvement over the infinite length minimum dispersion filter discussed above. Kalman-Levy filter and predictor can be expressed as a finite linear function of observed sequence as follows

$$\widehat{\mathbf{X}}_k = w_0 + \sum_{j=1}^{k} w_j \mathbf{Y}_j. \tag{1.4}$$

The performance of new filter is compared with their Gaussian counterparts by simulation. The main results of this chapter are published in a paper by Balakrishna and Hareesh (2009).

In Chapter 4 we study the parameter estimation of a stable autoregressive signal observed in a symmetric stable noisy environment. Autoregressive parameters of this model are estimated using a modified version of extended Yule-Walker method (see Davila (1998)) based on sample auto-covariation function. To minimize the bias of extended Yule-Walker estimates, it is suggested that a large number of extended Yule-Walker equations to be included for estimation. Auto-covariation functions in the extended Yule-Walker equations are replaced by their respective estimates. This replacement introduces some estimation error in this model. We represent these equations in the form of linear regression model. The proposed estimate $\hat{\phi}$ for the autoregressive parameter is obtained using ordinary least square regression model, and is given by

$$\hat{\phi} = (\hat{\Delta}'_{p',p}\hat{\Delta}_{p',p})^{-1}\hat{\Delta}'_{p',p}\hat{T}_{p'}, \tag{1.5}$$

where $\hat{\Delta}'_{p,p'} = [\hat{\lambda}(i-j+p)], i = 1, ..., p', j = 1, .., p$, $\hat{T}_{p'} = [\hat{\lambda}(i)]_{i=p+1}^{p+p'}$ and $\hat{\lambda}(.)$ is the estimates of auto-covariation function. The scale parameters of innovation and noise sequences are estimated using method of moments.

One limitation of the covariation based estimation is that the covariation matrix is not necessarily non-singular. The present study highlights this problem and proposes a

generalized solution to this problem using Moore-Penrose pseudo inverse. Singular value decomposition helps to identify and eliminate the singular values ν_i, of the auto-covariation matrix $\widehat{\Delta}_m$, which are close to zero. This will introduce the matrix $\widehat{\Delta}_p$ of rank, $p \leq m$. Based on this matrix we can propose a modified version of the generalized Yule-Walker estimate defined by,

$$\widehat{\phi}_* = (\overline{\widehat{\Delta}}_p)^+ \widehat{T}_m, \tag{1.6}$$

where,

$$(\overline{\widehat{\Delta}}_p)^+ = V \begin{bmatrix} \widehat{\Lambda}^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T$$

denotes the Moore-Penrose pseudo inverse of $\overline{\widehat{\Delta}}_p$ (see Stewart (1973), Rao (1973)). Asymptotic results of the proposed Yule-Walker estimates are studied. The proposed methods are illustrated through the data simulated from autoregressive signals with symmetric stable innovations. The new technique is applied to analyze the time series of sea surface temperature anomaly. Part of the results in this chapter is reported in Balakrishna and Hareesh (2010a, 2010b).

In Chapter 5 we introduce the concept of partial auto-covariation function (PcovF) for stable autoregressive time series model, a measure similar to PACF in the finite variance time series. We generalize the Durbin-Levinson algorithm in stable autoregressive models in terms of partial auto-covariation and use it for model identification. We propose a new information criteria for consistent order selection similar to Akaike Information Criteria. Concept of Partial auto-covariation is based on the linear prediction theory of stable processes by Cline and Brockwell (1985). We consider the vectors $\Phi'_k = (\phi'_1, \phi'_2, ..., \phi'_k)$, where $\phi'_i = \phi_i$ for $i \leq m$ and $\phi'_i = 0$ when $i > m$. The lag k partial auto-covariation, ϕ_{kk} is defined as the kth component of the vector

$$\Phi'_k = \Delta_k^{-1} T_k, \tag{1.7}$$

where, $\Delta_k = [\lambda(i-j)]_{i,j=1}^{k}$ is an $k \times k$ matrix of auto-covariation function $\lambda(.)$ with $\lambda(0) = 1$ and $T_k = (\lambda(1), ..., \lambda(k))'$.

The well known Durbin-Levinson algorithm has been generalized to fit stable autoregressive model consecutively increasing order to the observed time series data. Based on this we can estimate autoregressive parameters and partial auto-covariation function recursively. We can also derive an expression for mean absolute deviation of the prediction error in terms of the proposed partial auto-covariation function.

In this chapter we introduce a new Information Criterion, similar to Akaike Information Criterion (AIC) for order selection of autoregressive models, which can be defined as,

$$IC(k) = N^{2/\beta} \ln(\widehat{\gamma}_u(k)) + 2k, \text{ forsome, } \beta > \frac{\alpha}{\alpha - 1}, \tag{1.8}$$

where, $\widehat{\gamma}_u(k)$ is the mean absolute deviation of prediction error and α is the heavy tailed index. The order estimate $\widehat{m}$ is,

$$\widehat{m} = arg \min_{1<k\leq K(N)} IC(k). \tag{1.9}$$

Under some condition we have shown that for large samples the proposed order selection is consistent. That is,

$$\widehat{m} \xrightarrow{p} m \; as \; N \to \infty.$$

Simulation results show that the proposed information criteria perform better than AIC in both Gaussian and stable auto-regressive models in terms of model identification. Part of the results in this chapter is reported in Balakrishna and Hareesh (2010a, 2010b).

Chapter 6 discusses the frequency estimation of sinusoidal signal observed in symmetric stable noises using the modified version of generalized Yule-Walker estimate. Yule-Walker based spectrum estimation is widely used in Gaussian signal processing. Though the classical power spectral density does not exist in the stable signal environment, we can

still define power transfer function corresponding to the power spectral density (see Kluppelberg and Mikosch (1993)). In the present study we focus on the estimation of power transfer function using the proposed Generalized Yule-Walker method. Frequency estimators are obtained from the pole of the estimated power transfer function. The power transfer function estimate can be written as,

$$\widehat{S}(\omega) = \frac{1}{\widehat{\phi}(\omega)\widehat{\phi}^{*}(\omega)} \tag{1.10}$$

where, $\widehat{\phi}^{*}$ is the complex conjugate of $\widehat{\phi}$.

$$\widehat{\phi}(\omega) = 1 + \sum_{k=1}^{m} \widehat{\phi}_k \exp(-ik\omega),$$

and noting, $\widehat{z}_k = \widehat{r}_k \exp(-i\widehat{\omega})$, as an estimate of pole of power transfer function $\widehat{S}(\omega)$, we can estimate frequency component $\widehat{\omega}$ from this estimated pole.

Another important problem discussed in this chapter is that of identifying the number of frequency components in an observed signal. Number of frequency components depends on the order of autoregressive model. So we can modify the order estimation criteria using the decomposition method (see Castanie (2006)). The information criteria depends on the singular values of the auto-covariation matrix. Proposed modified Information Criteria is,

$$IC(k) = N^{2/\beta}(m-k)\ln\left(\frac{(\prod_{t=k+1}^{m}\widehat{\nu}_t)^{\frac{1}{m-k}}}{\frac{1}{m-k}\sum_{t=k+1}^{m}\widehat{\nu}_t}\right) + k(2m-k), \tag{1.11}$$

where, $k = 1, 2, ..., m-1$ and $p < m < N$. We can also define these criteria using the eigen values β_t of the auto-covariation matrix by replacing $\widehat{\nu}_t$ by $\widehat{\beta}_t$ in equation (1.11). Part of the results in this chapter is reported in Balakrishna and Hareesh (2010b).

Chapter 2

Basic Concepts

2.1 Introduction

Classical time series approach includes the modeling and analysis of finite variance linear time series models (Box et al. (1994) and Brockwell and Davis (1987)). This approach is generally known as Box and Jenkins time series analysis. Mathematical theory of classical time series analysis is based on the assumption that the error variances are finite. Time series analysis may be carried out either in time domain or in frequency domain. The time domain theory is generally motivated by the presumption that correlation between adjacent points in time is best explained in terms of the dependence of the current value on the past values. The time domain approach focuses on modeling some future value of a time series as a parametric function of the current and past values. Autocorrelation function (ACF) and partial autocorrelation function (PACF) are major tools used for analyzing the serial dependency of time series data. On the other hand, the frequency domain approach assumes that the primary characteristics of interest in time series analysis relate to periodic or systematic sinusoidal variations found naturally in most data. Frequency domain properties of time series can be well explained using spectral density function. One main objective of time series analysis is to predict the future behavior of the time series based on its past values. Minimum mean square prediction is the most popular method in

this direction. Before going for prediction we have to identify a proper stochastic model to the time series and then estimate its parameters. Estimation and forecasting are two important problems in time series analysis. Autocorrelation and partial autocorrelation function (PACF) plots are graphical approaches for model identification. Another popular tool for model identification is the Akaike information criteria (AIC). Yule-Walker method is a well accepted estimation procedure in classical time series model.

In recent years there has been a great deal of attention on modeling non-Gaussian time series which includes time series with heavy tailed innovations. Symmetric stable distributions are widely used to model heavy tailed variables as stated by Adler et al. (1998), Gallagher (2001), Shao and Nikias (1993). Most of the analysis techniques of classical time series models entails the knowledge of ACF. Autocorrelation function cannot be defined in time series based on symmetric stable distributions. Many authors used sample autocorrelation for stable time series due to its limiting properties (Adler et al. (1998), Davis and Resnick (1985)). Another important tool for stable time series is auto-covariation function (AcovF), which can be mathematically defined in stable processes (Gallagher (2001), Shao and Nikias (1993)). These functions are used for model identification and parameter estimation of some stable time series models, which will be discussed in this chapter.

Mathematical theory and methods of classical time series analysis will be discussed in the second section of this chapter. We discuss the classical time series models and some tools and techniques used for its analysis. Symmetric stable distributions and processes are defined in the third section. We surveyed the theoretical developments for time series analysis based on stable assumptions and organized them parallel to the developments in classical set up in Section 2.4. Linear prediction theory of some stable processes is addressed in Section 2.5. Last section covers some limit theorems used in our study.

2.2 Box and Jenkins time series approach

In this section we briefly discuss the linear stationary time series models with *finite variance* used in the classical set up and its analysis based on Box and Jenkins approach. Time series analysis starts with selection of a suitable mathematical model (or class of models) for the data. To allow for the possibly unpredictable nature of future observations, it is natural to suppose that each observation x_t is a realization of certain random variable X_t. The time series $\{x_t, t \in T_0\}$ is then a realization of the family of random variables $\{X_t, t \in T_0\}$. These considerations suggest modeling the data as a realization (or part of a realization) of a stochastic process $\{X_t, t \in T\}$ where $T \supseteq T_0$. Now we need to define stochastic process and its realization to get more clarity to the previous discussion.

Definition 2.2.1. A **stochastic process** is a family of random variables $\{X_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. The set T is an **index set**, of time points, such as $\{0, \pm 1, ...\},\ \{0, 1, ...\},\ [0,\ \infty),\ (-\infty,\ \infty)$.

In the present study, t will typically be discrete and vary over the integers $t = 0, 1, 2, ...,$ or some subset of the integers. From the definition of random variable we note that for each fixed $t \in T,\ X_t$ is a function $X_t(.)$ on the set Ω. On the other hand, for each fixed $\omega \in \Omega,\ X_.(\omega)$ is a function on T.

Definition 2.2.2. The functions $\{X_.(\omega),\ \omega \in \Omega\}$ on T are known as the **realizations** or **sample-paths** of the process $\{X_t, t \in T\}$.

Remark 2.2.3. We shall use the term time series to mean both the data and the process of which it is a realization.

Definition 2.2.4. Let $\mathcal{F}$ be the set of all vectors $\{t = (t_1, ..., t_n) \in T^n :\ t_1 < t_2 < ... < t_n,\ n = 1, 2, ...\}$. Then the **finite dimensional distribution functions** of $\{X_t,\ t \in T\}$ are the functions $\{F_t(.),\ t \in \mathcal{F}\}$ defined for $t = (t_1, ..., t_n)$ by

$$F_t(x) = P(X_{t_1} < x_1, ..., X_{t_n} < x_n), \quad x = (x_1, ..., x_n) \in \mathbb{R}^n.$$

Before going for the model description we start with some definitions and the basic tools for time series analysis such as autocorrelation, partial autocorrelation and spectrum. Most of the definitions in this section are taken from Brockwell and Davis (1987), Box et al. (1994) and Shumway and Stoffer (2006).

Definition 2.2.5. If $\{X_t, t \in T\}$ is a process such that $V(X_t) < \infty$ for each $t \in T$, then the **auto-covariance function** $\gamma_x(.,.)$ of $\{X_t\}$ is defined by

$$\gamma_x(r,s) = cov(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))], \quad r,s \in T. \tag{2.1}$$

Definition 2.2.6. The time series $\{X_t, t \in \mathbb{Z}\}$, with index set $\mathbb{Z} = \{0, \pm 1, \pm 2, ...\}$, is said to be **covariance stationary** or **weak stationary** if

(i) $E|X_t|^2 < \infty, \quad t \in \mathbb{Z}$,
(ii) $E(X_t) = \mu, \quad t \in \mathbb{Z}$,
and
(iii) $\gamma(r,s) = \gamma(r+t, s+t), \quad r,s,t \in \mathbb{Z}$ is a function of $|r-s|$ only.

If $\{X_t, t \in \mathbb{Z}\}$, is covariance stationary then $\gamma_x(r,s) = \gamma_x(r-s,0)$ for all $r,s \in \mathbb{Z}$. So the auto-covariance function of a covariance stationarity process can be redefined as the function of just one variable (lag),

$$\gamma(k) = \gamma(k,0) = cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)], \tag{2.2}$$

where $\mu = E(X_t)$. The function $\gamma(.)$ will be referred to as the auto-covariance function of $\{X_t\}$ and $\gamma(k)$ as its value at lag k. Now we can state some elementary properties of the auto-covariance function defined in (2.2).

Property 2.2.7. If $\gamma(.)$ is the auto-covariance function of a covariance stationary process $\{X_t, t \in \mathbb{Z}\}$, then,

(i) $\gamma(0) \geq 0$,

(ii) $|\gamma(h)| \leq \gamma(0), \quad h \in \mathbb{Z},$

and

(iii) $\gamma(h) = \gamma(-h), \quad h \in \mathbb{Z}.$

Property 2.2.8. A real-valued even function defined on the set $\mathbb{Z}$ of all integers is non-negative definite if and only if it is the auto-covariance function of a stationary time series.

Definition 2.2.9. Autocorrelation function at lag k is defined by

$$\rho(k) = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sqrt{E(X_t - \mu)^2 \; E(X_{t+k} - \mu)^2}}$$

For a covariance stationary process, the formula becomes $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$, where $\gamma(0) = \sigma_x^2 = E(X_t - \mu)^2$.

Definition 2.2.10. The process $\{X_t\}$, is said to be **Gaussian** if the finite dimensional distribution functions of $\{X_t\}$ are all multivariate normal.

Definition 2.2.11. The time series $\{X_t, t \in \mathbb{Z}\}$, is said to be **strictly stationary** if the joint distributions of $(X_{t_1}, ..., X_{t_m})$ and $(X_{t_1+k}, ..., X_{t_m+k})$ are the same for all positive integer m and for all $t_1, ..., t_m, k \in \mathbb{Z}$.

Remark 2.2.12. A strictly stationary process with finite second order moment is covariance stationary. The converse of this statement is not true in general (Brockwell and Davis (1987), page 13). However a Gaussian process is stationary in strict as well as weak stationary.

A matrix associated with a stationary process is defined as the covariance matrix of for random variables $(X_1, X_2, ..., X_n)$ made at n successive times and is given by,

$$\Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-2) & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-3) & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(1) & \gamma(0) \end{bmatrix}$$

The matrix can also be expressed in terms of auto-correlation function. That is, $\Gamma_n = \sigma_x^2 \, R_n$, where R_n is obtained by replacing $\gamma(.)$ in Γ_n by $\rho(.)$. It can be shown that both these matrices are positive definite for any stationary process (Brockwell and Davis (1987), Box et al. (1994)).

In practice we have a finite time series $X_1, X_2, ..., X_N$ of N observations, from which we can only obtain estimates of the mean μ and autocorrelations. One of the most satisfactory estimates of these functions have been discussed by Box et al. (1994) and is defined as follows:

Definition 2.2.13. An estimate of the $k-$th lag autocorrelation $\rho(k)$ is

$$\widehat{\rho}(k) = \frac{\widehat{\gamma}(k)}{\widehat{\gamma}(0)}, \tag{2.3}$$

where,

$$\widehat{\gamma}(k) = \sum_{t=1}^{N-k} (X_t - \overline{X})(X_{t+k} - \overline{X}), k = 0, 1, 2, ..., K, \tag{2.4}$$

is the estimate of the auto-covariance $\gamma(k)$, and $\overline{X}$ is the sample mean of the time series. The function $\widehat{\rho}(k)$ defined in (2.3) may be called the sample autocorrelation function.

So far we have discussed time series in the time domain. Now we will discuss it in the frequency domain. The idea that a time series is composed of periodic components, appearing in proportion to their underlying variances, is fundamental in the spectral representation of stationary processes. In other words, any stationary time series may be thought

of, approximately, as the random superposition of sines and cosines oscillating at various frequencies. Spectral density function is a mathematical tool for analyzing the periodic behavior of stationary time series. We will define it using some spectral representation theorems (for details see Shumway and Stoffer (2006)).

Theorem 2.2.14. *A function $\gamma(k)$, for $k = 0, \pm 1, \pm 2, ...$ is Hermitian non-negative definite if and only if it can be represented as*

$$\gamma(k) = \int_{-1/2}^{1/2} e^{2\pi i \omega k} dF(\omega), \tag{2.5}$$

where, $F(\omega)$ is a monotone non-decreasing function which is right continuous, bounded in $[-1/2, 1/2]$, and uniquely determined by the conditions $F(-1/2) = 0, F(1/2) = \gamma(0)$.

Proof. See Shumway and Stoffer (2006), page 534-535. □

Theorem 2.2.14 states that in particular, if $\{X_t\}$ is stationary with auto-covariance $\gamma(k)$, then there exist a unique monotonically increasing function $F(\omega)$, called the **spectral distribution function**, that is bounded, with $F(-\infty) = F(-1/2) = 0$, and $F(\infty) = F(1/2) = \gamma(0)$ such that (2.5) is true.

Theorem 2.2.15. *If $\gamma(k)$, is the auto-covariance function of a stationary process, $\{X_t\}$, with*

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty,$$

*then the **spectral density** of $\{X_t\}$ is given by,*

$$f(\omega) = \sum_{k=-\infty}^{\infty} \gamma(k)\, e^{-2\pi i \omega k}. \tag{2.6}$$

Proof. See Shumway and Stoffer (2006), page 537. □

A more important situation we use repeatedly is the one covered by Theorem 2.2.15, where it is shown that, subject to absolute summability of the auto-covariance, the spectral distribution function is absolutely continuous with $dF(\omega) = f(\omega)d\omega$, and the representation (2.6) becomes a motivation for the property given below.

Property 2.2.16. If the auto-covariance function, $\gamma(k)$, of a stationary process satisfies

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty,$$

then it has the representation,

$$\gamma(k) = \int_{-1/2}^{1/2} e^{2\pi i\omega k} f(\omega)d\omega, k = 0, \pm 1, \pm 2, ..., \tag{2.7}$$

as the inverse transform of the spectral density, which has the representation as shown in (2.6).

Auto-covariance function $\gamma(k)$ and the spectral density function $f(\omega)$ contain the same information about the underlying process. The auto-covariance function expresses information in terms of lags, whereas the spectral density expresses the same information in frequencies. So $\gamma(k)$ is a classical tool for analyzing time series in time domain where as $f(\omega)$ entails the same in frequency domain. More properties of this function, its estimation and applications are extensively discussed in Shumway and Stoffer (2006), Brockwell and Davis (1987), Box et al. (1994). Now we describe some standard linear time series models and their properties.

Definition 2.2.17. White noise process is a sequence $\{a_t\}$ of uncorrelated random variable with mean zero and constant variance σ_a^2. A particularly useful white noise series is Gaussian white noise, wherein the a_t are independent normal random variables, with mean zero and variance σ_a^2 or more briefly, $a_t \sim iidN(0, \sigma_a^2)$.

It is well known that a stochastic processes $\{X_t\}$ can be represented as the output from

a linear filter, whose input is white noise $\{a_t\}$. That is

$$\begin{aligned} X_t - \mu &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + ... \\ &= \textstyle\sum_{j=1}^{\infty} \psi_j a_{t-j}, \end{aligned} \tag{2.8}$$

where, μ is the common mean of X_t and ψ_j's are suitable constants. We assume the process $\{X_t\}$ is a zero mean process unless it specified. For $\{X_t\}$ defined by (2.8) to a weakly stationary process, it is necessary for the coefficients ψ_j to be absolutely summable, that is, for $\sum_{j=1}^{\infty} |\psi_j| < \infty$. The model (2.8) implies that, under suitable conditions X_t is also a weighted sum of past values of X_t's plus a white noise a_t, that is

$$\begin{aligned} X_t &= \pi_1 X_{t-1} + \pi_2 X_{t-2} + ... + a_t \\ &= \textstyle\sum_{j=1}^{\infty} \pi_j X_{t-j} + a_t. \end{aligned} \tag{2.9}$$

If we define $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, then we can show that, $\pi(B) = \psi^{-1}(B)$, and B is the shift operator defined as $B^k X_n = X_{n-k},\ \psi_0 = \pi_0 = 1$.

Definition 2.2.18. Consider a special case of (2.9), in which only first p of the weights are non zero. The model is known as **autoregressive model** of order p ($AR(p)$), which may be written as,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + a_t. \tag{2.10}$$

Model (2.10) can be represented as $\phi(B)X_t = a_t$, where the polynomial $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$. An $AR(p)$ process $\{X_t\}$ is stationary, if the roots of the equation $\phi(z) = 0$ lie outside the unit circle.

Definition 2.2.19. Consider a special case of (2.8), in which only first of the q weights are non zero. The model is known as **moving average model** of order q ($MA(q)$), which may be written as,

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + ... + \theta_q a_{t-q}. \tag{2.11}$$

Model (2.11) can be represented as $X_t = \theta(B)a_t$, where the polynomial $\theta(B) = 1 + \theta_1 B +$

$\theta_2 B^2 + ... + \theta_q B^q$. Moving average process is always stationary.

Definition 2.2.20. Autoregressive Moving Average ($ARMA(p, q)$) model is the combination of **autoregressive and moving average models** which may be defined as

$$\phi(B)X_t = \theta(B)a_t, \tag{2.12}$$

where, the polynomials $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q$. An $ARMA(p, q)$ process $\{X_t\}$ is stationary, if the roots of the equation $\phi(z) = 0$ lie outside the unit circle.

Definition 2.2.21. An $ARMA(p, q)$ model, $\phi(B)X_t = \theta(B)a_t$, is said to be **causal**, if the time series $\{X_t; t = 0, \pm 1, \pm 2, ...\}$ can be written as a one-sided linear process:

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \psi(B)a_t,$$

where, $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$; we set $\psi_0 = 1$.

Property 2.2.22. An $ARMA(p, q)$ model is **causal** if and only if $\phi(z) \neq 0$ for $|z| \leq 1$. The coefficients of $\psi(B)$ can be determined by solving

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \ |z| \leq 1.$$

Another way to phrase this property is that an $ARMA$ process is causal only when the roots of $\phi(z)$ lie outside the unit circle; that is, $\phi(z) = 0$ only when $|z| > 1$.

Definition 2.2.23. An $ARMA(p, q)$ model, $\phi(B)X_t = \theta(B)a_t$, is said to be **invertible**, if the time series $\{X_t; t = 0, \pm 1, \pm 2, ...\}$ can be written as

$$\pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = a_t,$$

where, $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, and $\sum_{j=0}^{\infty} |\pi_j| < \infty$; we set $\pi_0 = 1$.

Property 2.2.24. An $ARMA(p,q)$ model is **invertible** if and only if $\theta(z) \neq 0$ for $|z| \leq 1$. The coefficients of $\pi(B)$ can be determined by solving

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \ |B| \leq 1.$$

Another way to phrase this property is that an $ARMA$ process is invertible only when the roots of $\theta(z)$ lie outside the unit circle.

From the definition we can see that, $ARMA(p,q)$ model reduces to $AR(p)$ model when $q = 0$ and similarly it reduces to $MA(q)$ model when $p = 0$.

Property 2.2.25. Autocorrelation function, $\rho(.)$ of a stationary $AR(p)$ process with finite second order moments, follows the Yule-Walker equations specified by,

$$\rho(k) = \sum_{i=1}^{p} \phi_i \rho(k-i), \ k \geq 1. \tag{2.13}$$

These equations can be used to estimate the AR parameters $\phi_1, ..., \phi_p$ by replacing the $\rho(k)$ by the sample ACF. The resulting estimates referred to as the Yule-Walker estimates.

Property 2.2.26. The $MA(q)$ process is stationary with mean zero and auto-correlation function

$$\rho(k) = \begin{cases} \sum_{j=0}^{q-k} \theta_j \theta_{j+k}, & 0 \leq k \leq q \\ 0 & k > q. \end{cases}$$

The cutting of $\rho(k)$ after q lags is the signature of the $MA(q)$ model. So the auto-correlation function (ACF) provides a considerable amount of information about the order of dependence when the process is a moving average process. If the process, however, is ARMA or AR, the ACF alone tells us little about the orders of dependence. Hence, it is worthwhile pursuing a function that will behave like the ACF of MA models, but for AR models, namely, the partial autocorrelation function (PACF) is more useful. To formally define the PACF, we need linear prediction theory of stationary process. We will discuss this in next section.

Time series prediction: In prediction, the goal is to forecast the future values of a time series, $X_{n+m}, m = 1, 2, ...,$ based on the data collected up to the present, $X = (X_n, X_{n-1}, ..., X_1)$. Throughout this section, we will assume $\{X_t\}$ is stationary and the model parameters are known. The problem of model parameter estimation will be discussed later. The theory of minimum mean square error (MMSE) forecast for linear time series (process) provides us the result that the $m-$step ahead forecast $\widehat{X}_{n+m}$ is the conditional expectation given by

$$\widehat{X}_{n+m} = E(X_{n+m}|X_n, X_{n-1}, ..., X_1).$$

When we are dealing with linear time series the predictor will be a linear function of the past observations and may be represented as,

$$\widehat{X}_{n+m} = l_0 + \sum_{j=1}^{n} l_j X_{n+m-j}, \tag{2.14}$$

where $l_0, l_1, ..., l_n$ are real numbers. Linear predictors of the form (2.14) that minimize the mean square prediction error are called best linear predictors (BLPs). If the process is Gaussian, minimum mean square error predictors and best linear predictors are the same (Shumway and Stoffer (2006), page 111).

Property 2.2.27. Best Linear Prediction (BLP) for Stationary Processes: Given data $(X_1, X_2, ..., X_n)$, the best linear predictor, $\widehat{X}_{n+m} = l_0 + \sum_{j=1}^{n} l_j X_j$, of X_{n+m}, for $m \geq 1$, is found by solving

$$E[(\widehat{X}_{n+m} - X_{n+m})X_k] = 0, k = 0, 1, ..., n, \tag{2.15}$$

where $X_0 = 1$.

The equations specified in (2.15) are called the prediction equations, which can be used to solve for the coefficients $(l_0, l_1, ..., l_n)$.

For mean-zero stationary time series, let $\widehat{X}_k$ denote the regression of X_k on $(X_{k-1}, X_{k-2}, ..., X_1)$, which we write as

$$\widehat{X}_k = l_1 X_{k-1} + l_2 X_{k-2} +, ..., + l_{k-1} X_1. \tag{2.16}$$

No intercept term is needed in (2.16) because the mean of X_k is zero. In addition, let $\widehat{X}_0$ denote the regression of X_0 on $X_1, X_2, ..., X_{k-1}$, then

$$\widehat{X}_0 = l_1 X_1 + l_2 X_2 + ... + l_{k-1} X_{k-1}. \tag{2.17}$$

The coefficients, $l_1, l_2 ..., l_{k-1}$ in (2.17) are the same as those in (2.16). Based on these equations, partial autocorrelations can be defined as follows:

Definition 2.2.28. The **partial autocorrelation function (PACF)** of a stationary process, $\{X_k\}$, denoted $\phi_{k,k}$, for $k = 1, 2, ...$, is defined by

$$\phi_{1,1} = corr(X_1, X_0) = \rho(1), \tag{2.18}$$

and

$$\phi_{k,k} = corr(X_k - \widehat{X}_k, X_0 - \widehat{X}_0), k \geq 2. \tag{2.19}$$

Both $(X_k - \widehat{X}_k)$ and $(X_0 - \widehat{X}_0)$ are uncorrelated with $\{X_1, X_2, ..., X_{k-1}\}$. By stationarity, the PACF, $\phi_{k,k}$, is the correlation between X_t and X_{t-k} obtained by fixing the effect of $X_{t-1}, X_{t-2}, ..., X_{t-(k-1)}$.

Consider, first, one-step-ahead prediction. That is, given $(X_1, X_2, ..., X_n)$, we wish to forecast the value of the time series at the next time point, X_{n+1} by assuming an $AR(n)$ model for X_{n+1}. The BLP of X_{n+1} is

$$\widehat{X}_{n+1} = \phi_{n,1} X_n + \phi_{n,2} X_{n-1} + ... + \phi_{n,n} X_1. \tag{2.20}$$

Using the best linear prediction for stationary process, prediction equations (2.15) assure that the coefficients $\phi_{n,1}, \phi_{n,2}, ..., \phi_{n,n}$ satisfy the Yule-Walker equations,

$$\sum_{j=1}^{n} \phi_{n,j} \gamma(k-j) = \gamma(k), k = 1, 2, ..., n. \tag{2.21}$$

The Yule-Walker equations (2.21) can be written using matrix notation as

$$\Gamma_n \Phi_n = \gamma_n, \tag{2.22}$$

where, $\Phi_n = (\phi_{n,1}, \phi_{n,2}, ..., \phi_{n,n})'$, $\Gamma_n = [\gamma(i-j)]_{i,j=1}^{n}$ is an $n \times n$ matrix and $\gamma_n = (\gamma(1), ..., \gamma(n))'$.

If Γ_n is nonsingular, Φ_n is unique, and is given by

$$\Phi_n = \Gamma_n^{-1} \gamma_n. \tag{2.23}$$

It is sometimes convenient to write the one-step-ahead forecast in vector notation

$$\widehat{X}_{n+1} = \Phi_n X, \tag{2.24}$$

where $X = (X_n, X_{n-1}, ..., X_1)$. The mean square error is

$$P_{n+1} = E(\widehat{X}_{n+1} - X_{n+1})^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n. \tag{2.25}$$

For ARMA models in general, the prediction equations will not be as simple as in the pure AR case (see Shumway and Stoffer (2006), page 113). In addition, for n large, the use of (2.23) is prohibitive because it requires the inversion of a large matrix. There are, however, iterative solutions that do not require any matrix inversion. In particular, we mention the recursive solution due to Levinson (1947) and Durbin (1960). A detailed description of this algorithm is given in Shumway and Stoffer (2006), Page 113.

Definition 2.2.29. Durbin Levinson Algorithm: Equations (2.23) and (2.25) can be solved iteratively as follows:

$$\begin{aligned} \phi_{0,1} &= 0, \\ \widehat{P}_0 &= \gamma(0). \end{aligned} \tag{2.26}$$

For, $n \geq 1$

$$\begin{aligned} \phi_{n,n} &= \frac{\rho(n)-\sum_{j=1}^{n-1}\phi_{n-1,k}\rho(n-k)}{1-\sum_{j=1}^{n-1}\phi_{n-1,k}\rho(k)}, \\ P_{n+1} &= P_n(1-\phi_{n,n}^2), \end{aligned} \tag{2.27}$$

where, for $n \geq 2$,

$$\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n}\phi_{n-1,n-k},\ k = 1, 2, ..., n-1. \tag{2.28}$$

Durbin Levinson Algorithm is an efficient algorithm in modern time series analysis. This recursive algorithm can be used to estimate partial autocorrelations, Yule-Walker estimates of autoregressive parameters, forecast, forecast error etc.

Model identification: The primary tools for model identification are the plots of autocorrelation and the partial autocorrelation. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behavior of these plots when the order is known. Autocorrelation function of an autoregressive process of order p tail off, its partial autocorrelation function has a cut off after lag p. On the other hand the autocorrelation function of moving average process cuts off after lag q, while its partial autocorrelation tails off after lag q. If both autocorrelation and partial autocorrelation tail off, a mixed process is suggested. Furthermore, the autocorrelation function for a mixed process, contains a p-th order AR component and q-th order moving average component, and is a mixture of exponential and damped sine waves after the first $q-p$ lags. The PACF for a mixed process is dominated by a mixture of exponential and damped sine waves after the first $q-p$ lags.

Specifically, for an $AR(1)$ process, the sample autocorrelation function should have an exponentially decreasing behavior. However,the sample auto-correlation function for higher-order AR processes are often a mixture of exponentially decreasing and damped sinusoidal components. For higher-order autoregressive processes, the sample autocorrelation needs to be supplemented with a partial autocorrelation plot. The partial autocorrelation of an $AR(p)$ process becomes zero at lag $p+1$ and greater, so we examine the sample

partial autocorrelation function to see if there is evidence of a departure from zero. This is usually determined by placing a sample partial autocorrelation plot and corresponding $(1-\alpha)100\%$ confidence level of white noise process, for a given α.

The autocorrelation function of a $MA(q)$ process becomes zero at lag $q+1$ and greater, so we examine the sample autocorrelation function to see where it essentially becomes zero. We do this by placing the $(1-\alpha)100\%$ confidence interval for the sample autocorrelation function on the sample autocorrelation plot.

Prediction error variance decrease when the order increases. When the theoretical order is reached, this variance remains constant. Most of the order selection criteria follow this concept. However, in practice we have to use estimate of the prediction error variance for model identification. When the order is over estimated the bias of the estimated prediction error variance increases. Model selection criteria should take both these aspects in to account. One of the first criteria integrate these two aspects, proposed by Akaike (1970) was the Final Prediction Error (FPE): the estimated error which minimizes:

$$FPE(k) = \frac{N+k}{N-k}\widehat{\sigma}^2(k), \tag{2.29}$$

where $\widehat{\sigma}^2(k)$ is the estimated prediction error variance of autoregressive models with order k. The autoregressive parameters and corresponding prediction error variances are estimated using Durbin-Levinson algorithm. Bias of the estimated prediction error variance of $AR(k)$ model for large is N given by $\sigma^2(\frac{k}{N})$ and which can be used to correct the identification criteria.

One of the most well known criteria proposed by Akaike (1974) is Akaike Information Criteria (AIC) which included the bias correction and may be defined as

$$AIC(k) = N\ln(\widehat{\sigma}^2(k)) + 2k, \tag{2.30}$$

where, $\widehat{\sigma}^2(k)$ is the estimated prediction error variance of autoregressive models of order k.

Roughly speaking, the first term on the right hand side of AIC is a measure of goodness

of fit of the model to the data and the second term is a penalty function, which penalizes higher dimensional models. Given a set of candidate models, the selection is typically made by choosing the model that minimizes the adopted criterion function among all candidate models. The autoregressive parameters and corresponding prediction error variances are estimated using Durbin-Levinson algorithm. Best choice of the order is the one which minimizes the $AIC(k)$. Order estimate is not consistent in finite variance case. There are some other order selection criteria such as Bias corrected AIC (AICc), Bayesian Information Criteria (BIC) etc, for details one can refer Shumway and Stoffer (2006), page 54. A more general description of these methods for model identification of ARMA models can be seen in Brockwell and Davis (1987), page 293.

Parameter estimation: Throughout this section, we assume we have N observations, $X_1, ..., X_N$, from an $ARMA(p, q)$ process in which, initially, the order parameters, p and q, are known. Our goal is to estimate the parameters, $\phi_1, ..., \phi_p$, $\theta_1, ..., \theta_q$, and σ_a. We have already discussed the problem of order selection in the pervious section. We begin with the method of moments to estimate the parameters. We immediately see that, if X_t, is not centered, $E(X_t) = \mu$ then the method of moments estimator of μ is the sample average, $\overline{X}$. Thus, while discussing method of moments, we will assume $\mu = 0$. Although the method of moments can produce good estimators, they can sometimes lead to suboptimal estimators. We limited our study to the case in which the method leads to optimal (efficient) estimators, that is, $AR(p)$ models. This is because, given initial conditions, $AR(p)$ models are linear models, and the Yule-Walker estimators are essentially least squares estimators.

For a stationary $AR(p)$ process, $\{X_t\}$ defined by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + a_t, \tag{2.31}$$

the auto-covariance function follows the Yule-Walker equations specified by,

$$\begin{aligned} &\gamma(k) = \textstyle\sum_{i=1}^{p} \phi_i \gamma(k-i),\ k = 1, 2, ..., p, \\ &\sigma_a = \gamma(0) - \textstyle\sum_{i=1}^{p} \phi_i \gamma(i). \end{aligned} \tag{2.32}$$

The prediction equations (2.32) can be written in matrix notation as

$$\begin{aligned} \Gamma_p \Phi_p &= \gamma_p, \\ \sigma_a^2 &= \gamma(0) - \Phi' \gamma_p, \end{aligned} \tag{2.33}$$

where, $\Phi_p = (\phi_1, \phi_2, ..., \phi_p)'$, $\Gamma_p = [\gamma(i-j)]_{i,j=1}^p$ is an $p \times p$ matrix and $\gamma_p = (\gamma(1), ..., \gamma(p))'$. Using the method of moments, we replace $\gamma(k)$ in (2.33) by $\widehat{\gamma}(k)$, and solve for

$$\begin{aligned} \widehat{\Phi}_p &= \widehat{\Gamma}_p^{-1} \widehat{\gamma}_p, \\ \widehat{\sigma}_a^2 &= \widehat{\gamma}(0) - \widehat{\Phi}' \widehat{\gamma}_p. \end{aligned} \tag{2.34}$$

These are called the Yule-Walker estimators. For calculation purposes, it is sometimes more convenient to work with the sample autocorrelation function. By factoring $\widehat{\gamma}(0)$ in (2.34), we can write the Yule-Walker estimates as

$$\begin{aligned} \widehat{\Phi}_p &= \widehat{R}_p^{-1} \widehat{\rho}_p, \\ \widehat{\sigma}_a^2 &= \widehat{\rho}(0) - \widehat{\Phi}' \widehat{\rho}_p, \end{aligned} \tag{2.35}$$

where, $R_p = [\widehat{\rho}(i-j)]_{i,j=1}^p$ is an $p \times p$ matrix and $\widehat{\rho}_p = (\widehat{\rho}(1), ..., \widehat{\rho}(p))'$. For $AR(p)$ models, if the sample size is large, the Yule-Walker estimators are approximately normally distributed, and $\widehat{\sigma}_a^2$ is close to the true value of σ_a^2.

Large Sample Results for Yule-Walker Estimators: The asymptotic ($N \to \infty$) behavior of the Yule-Walker estimators in the case of causal AR(p) processes is as follows: (for a proof see Shumway and Stoffer (2006), page 530). As $N \to \infty$,

$$\begin{aligned} &\sqrt{N}(\widehat{\Phi}_p - \Phi_p) \xrightarrow{L} N(0, \sigma_a^2 \Gamma_p^{-1}), \\ &\text{and} \quad \widehat{\sigma}_a^2 \xrightarrow{p} \sigma_a^2. \end{aligned} \tag{2.36}$$

The Durbin-Levinson algorithm, (2.26)-(2.28), can be used to calculate $\widehat{\Phi}_p$ without inverting Γ_p or R_p, by replacing $\rho(k)$ by $\widehat{\rho}(k)$ in the algorithm. In running the algorithm, we will iteratively calculate the $k \times 1$ vector, $\widehat{\Phi}_k = (\widehat{\phi}_{k,1}, \widehat{\phi}_{k,2}, ..., \widehat{\phi}_{k,k})'$, for $k = 1, 2, ...$. Thus, in

addition to obtaining the desired forecasts, the Durbin-Levinson algorithm yields $\widehat{\phi}_{k,k}$, the sample PACF. Using (2.36), we can get the following result.

Large Sample Distribution of the PACF: For a causal $AR(p)$ process, as $N \to \infty$,

$$\sqrt{N}\widehat{\phi}_{k,k} \xrightarrow{L} N(0,1), \ fork > p. \tag{2.37}$$

If we use the method of moments for MA or ARMA models, we will not get optimal estimators because in such processes moment equations are nonlinear in the parameters (Shumway and Stoffer (2006), page 124). Maximum likelihood and least square estimation procedures are another popular method in linear time series estimation (For details one can refer Shumway and Stoffer (2006), Brockwell and Davis (1987), Box et al. (1994)).

2.3 Alpha Stable Distributions and Processes

Stable distributions are recommended to model the data when it shows a tendency to follow heavy tailed distributions. Theoretical justifications for using stable distribution as a basic statistical modeling tool come from the Generalized Central Limit Theorem (GCLT) discussed by Shao and Nikias (1993), just like the central limit theorem in Gaussian distribution. The theorem stated below shows that if the finite variance assumption is dropped, the only possible resulting limit distributions are stable.

Theorem 2.3.1. *A random variable η is said to be stable, or to have a stable distribution, if for every positive integer n there exist constants $a_n > 0$ and b_n, such that the sum $\eta_1 + \eta_2 + ... + \eta_n$ has the distribution as that of $a_n\eta + b_n$ for $n \to \infty$ and all iid random variables $\eta_1, \ \eta_2, ..., \eta_n$, with the same distribution as η.*

Proof. See Breiman (1968), page 199. □

If an observed signal or noise can be thought of as the sum or a result of large number of independent and identically distributed effects, then the generalized central limit theorem suggests that a stable model may be appropriate. This characteristic of the stable

distribution is one of the main reasons why the stable distribution is suitable for modeling signals and noise of impulsive nature. Definition and properties of alpha stable distributions and processes are discussed in detail by, Zolotarev (1986), Samorodnitsky and Taqqu (1994), Brockwell and Davis (1987). The stable distribution is specified by its characteristic function, $\Phi(t) = E(exp(it\eta))$ and is given by,

$$\Phi(t) = \begin{cases} \exp(it\beta - \gamma|t|^{\alpha}(1 - i\theta\frac{t}{|t|}\tan(\frac{\pi\alpha}{2}))), \ if \ \alpha \neq 1 \\ \exp(it\beta - \gamma|t|^{\alpha}(1 - 2i\theta\frac{t}{\pi|t|}\ln(|t|))), \ if \ \alpha = 1 \end{cases} \tag{2.38}$$

which depends on four parameters: $\alpha \in (0, 2]$, measuring the tail thickness (thicker tails for smaller values of the parameter), $\theta \in [-1, 1]$ determining the degree and sign of asymmetry, $\gamma > 0$ (scale) and $\beta \in R$ (location). To denote stable distribution with parameters α, θ, γ and β we will use the notation $S_{\alpha}(\beta, \theta, \gamma)$.

As in the Gaussian case, a random variable X with $S_{\alpha}(\beta, \theta, \gamma)$ distribution can be standardized to produce

$$Z = \frac{X - \beta}{\gamma} \sim S_{\alpha}(0, \theta, 1). \tag{2.39}$$

Stable family of distributions have a closed form density function in a very few cases: $\alpha = 2$, corresponding to the normal distribution, $\alpha = 1$ and $\beta = 0$, yielding the Cauchy distribution, and $\alpha = 1/2$ and $\beta = 0$, for the Levy distribution. Another difficulty in this class of distributions is that moments of order greater than α do not exist, except when $\alpha = 2$. Standard estimation procedures will fail due to the above mentioned limitations of these classes of distributions. When $\alpha = 2$, the normal distribution has well understood asymptotic tail properties. The tail probabilities in the non Gaussian stable case are asymptotically known.

Definition 2.3.2. Tail approximation: Let $X \sim S_{\alpha}(\beta, \theta, \gamma)$ with $0 < \alpha < 2$, $-1 < \theta < 1$, then as $x \to \infty$

$$\begin{aligned} P(X > x) &\sim \gamma^{\alpha} C(1 + \theta) x^{-\alpha} \\ f(x|\alpha, \beta, \theta, \gamma) &\sim \alpha\gamma^{\alpha} C(1 + \theta) x^{-(\alpha+1)} \end{aligned} \tag{2.40}$$

The parameter α is the heavy tailed index and C is the scale factor. The statement

$h(x) \sim g(x)$ as $x \to a$ means that $\lim_{x \to a} \frac{h(x)}{g(x)} = 1$.

Pareto distributions are a class of probability laws with upper tail probabilities given exactly by the right hand side of (2.40). The term stable Paretian laws is used to distinguish between the fast decay of the Gaussian law and the Pareto-like tail behavior in the $\alpha < 2$ case. A general distribution is said to be heavy tailed if its tails are heavier than exponential distributions. Heavy tails refer to a slow, power-like decay of a tail of a distribution function. When $\beta = 0$ and $\theta = 0$ the distribution is referred to as a symmetric stable distribution and is defined below.

Definition 2.3.3. A random variable η is said to have a **symmetric stable distribution** and we denote it by $\eta \sim S_\alpha(\gamma)$ if its characteristic function is of the form

$$\varphi_\eta(v) = \exp(-\gamma|v|^\alpha),$$

where, $\alpha \in (0, 2]$, measuring the tail thickness, $\gamma > 0$ the scale (dispersion) parameter. Note that $E|\eta|^k < \infty$, if $k < \alpha$. We assume that $E|\eta| < \infty$. In other words, we restrict α to be in the interval $(1, 2]$. In the present study our main focus is limited to the finite mean symmetric stable random variable.

Multivariate stable distributions is achieved by constructing a linear sum of p independent stable random variables. A $p-$dimensional alpha stable random vector ε, can be written as a linear sum of p independent stable random variables:

$$\varepsilon = G\zeta, \tag{2.41}$$

where ζ is the vector representation of p independent alpha stable variables and $G \in \mathbb{R}^{p \times p}$ represents the linear relation between ε and ζ. Then, ε is characterized by the heavy tailed parameter and tail covariance matrix, which is a generalized measure of covariance in stable distributions with $\alpha < 2$ (Sornette and Ide (2001), Bouchaud et al. (1998)). We can define the Tail Covariance as follows:

Definition 2.3.4. Let ζ_i be the i-th component of the random vector ζ defined in equation (2.41), then the **Tail covariance** B is defined as the matrix of scale factors of the distribution of all products $\zeta_i\zeta_j$. The matrix B can be represented as

$$B = G^{[\alpha/2]}\, C\, G^{T[\alpha/2]}, \tag{2.42}$$

where, C is a diagonal matrix containing the scale factors of the p independent random variables ζ and the operator $\{.\}^{[\beta]}$ is defined as,

$$G_{ij}^{[\beta]} = |G_{ij}|^{\beta}\, sign(G_{ij}), \tag{2.43}$$

where,

$$sign(x) = \begin{cases} 1, \text{ for } x > 0 \\ -1, \quad x < 0 \\ 0, \quad x = 0. \end{cases}$$

Given a tail-covariance matrix, B, G and C can be obtained directly via diagonalization.

Brockwell and Davis (1987) discussed stable $ARMA$ model and its properties, and various statistical problems associated with an $ARMA$ processes. The stable $ARMA(p, q)$ process can be described through some propositions given by Brockwell and Davis (1987).

Proposition 2.3.5. *Let $\{a_t\}$ be an iid sequence of symmetric stable random variables specified in Definition 2.3.3. If $\{\psi_j\}$ is a sequence of constants such that*

$$\sum_{j=-\infty}^{\infty} |\psi_j|^{\delta} < \infty, \;\; \delta \in (0, 1]$$

then the infinite series,

$$\sum_{j=-\infty}^{\infty} \psi_j a_j,$$

converges absolutely with probability one.

Proof. See Brockwell and Davis (1987), page 480. □

The process defined by

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j}, \tag{2.44}$$

where $\{\psi_j\}$ and $\{a_t\}$ satisfy the assumptions of Proposition 2.3.5, exists with probability one and is strictly stationary. In particular if the coefficients ψ_j are chosen so that $\psi_j = 0$ for $j < 0$ and

$$\sum_{j=-\infty}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| < 1, \tag{2.45}$$

where $\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q$ and $\phi(z) = 1 + \phi_1 z + ... + \phi_p z^q \neq 0$ for $|z| \leq 1$, then we can show that $\{X_t\}$ as defined by equation (2.44) satisfies the $ARMA$ equations $\phi(B)X_t = \theta(B)a_t$. This results can be stated as a proposition.

Proposition 2.3.6. *Let $\{a_t\}$ be an iid sequence of symmetric stable random variables specified in Definition 2.3.3. Then if $\theta(.)$ and $\phi(.)$ are polynomials such that $\phi(z) \neq 0$ for $|z| < 1$, the difference equations*

$$\phi(B)X_t = \theta(B)a_t, \tag{2.46}$$

have the unique strictly stationary solution,

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j},$$

where the coefficients $\{\psi_j\}$ are determined by the relation (2.45)*. If in addition $\phi(z)$ and $\theta(z)$ have no common zeros, then the process* (2.46) *is invertible if and only if $\theta(z) \neq 0$ for $|z| \leq 1$.*

Proof. See Brockwell and Davis (1987), page 481. □

Remark 2.3.7. The ARMA process expressed in (2.46) is known as **stable autoregressive moving average ($ARMA(p, q)$) model** and it reduces to **stable autoregressive**

($AR(p)$) model when $\theta(B) = 1$ or $q = 0$ and similarly it reduces to **stable moving average ($MA(q)$) model** when $\phi(B) = 1$ or $p = 0$.

2.4 Analysis of Stable Time Series Models

In this section we surveyed the existing tools and techniques for analyzing stable time series models. Autocorrelation function cannot be defined for stable time series due to the lack of finite second order moments. But many researchers have used sample auto-correlation functions for time series analysis of stable models (see Adler et al. (1998) and reference there for). The main motivation of these studies is based on a limit Theorem of sample auto-covariation function of stable moving average process by Davis and Resnick (1986). The Theorem can be stated as follows:

Theorem 2.4.1. *Let $\{a_t\}$ be an iid symmetric sequence of α-stable random variables and let $\{X_t\}$ be the strictly stationary process defined by,*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j},$$

where, $\sum_{j=-\infty}^{\infty} |j||\psi_j|^\delta < \infty$ for some $\delta \in (0, \alpha) \cap [0, 1]$. It then follows immediately that the sample correlation function

$$\hat{\rho}(l) = \sum_{t=1}^{n-l} X_t X_{t+l} / \sum_{t=1}^{n} X_t^2,\ l > 0$$

converges in probability to the analogue of the correlation function defined by

$$\rho(l) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+l} / \sum_{j=-\infty}^{\infty} \psi_j^2.$$

That is for $\delta > \alpha$,

$$n^{1/\delta}(\hat{\rho}(l) - \rho(l)) \xrightarrow{p} 0.$$

Also

$$(n/\ln(n))^{1/\alpha}(\hat{\rho}(l) - \rho(l)) \xrightarrow{L} Y_l$$

where

$$Y_l = \left(\sum_{j=1}^{\infty} |\rho(l+j) + \rho(l-j) - 2\rho(j)\rho(l)|^{\alpha}\right)^{1/\alpha} \frac{U}{V}.$$

Here U and V are independent stable random variable with V is positive with $V \sim S_{\alpha}(0, 1, C_{\alpha/2}^{-2/\alpha})$ and $U \sim S_{\alpha}(0, 0, C_{\alpha}^{1/\alpha})$, where

$$C_{\alpha} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\frac{\pi\alpha}{2})}, & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} & \text{if } \alpha = 1. \end{cases}$$

Proof. See Davis and Resnick (1986). □

Adler et al. (1998) discussed the importance of this result in the analysis of stable time series models. Model identification and parameter estimation of stable autoregressive model has been studied parallel to the classical Box and Jenkins setup using sample autocorrelation function. Simulation results discussed in Adler et al. (1998) indicate that the sample ACF is a satisfactory tool for stable time series analysis.

From Theorem 2.4.1 we can show that the Yule-Walker estimates defined in (2.35) in terms of sample autocorrelation function of a stable $AR(p)$ model follows the distribution,

$$(n/\ln(n))^{1/\alpha}(\hat{\phi} - \phi) \xrightarrow{L} D(Y_1, ..., Y_p), \tag{2.47}$$

where, D is a $p \times p$ matrix of partial derivative of vector function $\psi(z) = R_p(z)^{-1}z$. Here $R_p(z) = [z_{|i-j|}]_{i,j=1}^{p}$, $z_0 = 1$ and $\phi = \psi(\rho)$.

The limiting distribution of the sample PACF is now given by (2.47), which is in general complicated. However, when $p = 0$, the right hand side of (2.47) reduces to U/V, which is the same limit as for the sample ACF of white noise. This result can be used to identify the order of autoregressive model from the sample PACF plot and corresponding confidence

level of white noise process. In practice distribution of U/V cannot be computed theoretically, simulation or numerical methods are the only solution. This is one of the limitations of this method compared to the classical setup.

Bhansali (1988) and Knight (1989) used Akaike information criterion (AIC) for order selection of stable autoregressive models. AIC is defined by

$$AIC(k) = N \ln(\widehat{\sigma}^2(k)) + 2k, \tag{2.48}$$

where N is the sample size, and $\widehat{\sigma}^2(k)$ is the estimate of the innovation variance obtained from the Yule-Walker estimates for $k-$th order autoregressive sequence. The order estimate $\widehat{p}$ is,

$$\widehat{p} = arg \min_{1<k\leq K(N)} AIC(k), \tag{2.49}$$

where $K(N)$ is an acceptance upper bound for p, is the corresponding estimate of the order p. In classical setup AIC is not a consistent order selection criteria, but Knight (1989) shows that this criteria is consistent for heavy tailed situation.

Auto-covariation is an alternative measure employed in the place of autocorrelation (Gallagher (2001)) which requires only the finite absolute mean. In other words, we restrict α to be in the interval $(1, 2]$. Now we can see the definition and properties of this function.

Definition 2.4.2. For a zero mean stationary process, with finite absolute moment, the auto-covariation function of lag k is defined as

$$\lambda_x^{\alpha}(k) = Acov(X_n, X_{n-k}) = \frac{E(X_n {X_{n-k}}^{[\beta-1]})}{E(X_{n-k}^{[\beta]})}, \text{ forsome } \beta < \alpha \tag{2.50}$$

where, $\{\}^{[\beta]}$, is same as defined in (2.43). For the sake of simplicity in our study, we consider $\beta = 1$, so the auto-covariation function becomes,

$$\lambda_x(k) = Acov(X_n, X_{n-k}) = \frac{E(X_n sign(X_{n-k})}{E|X_{n-k}|}. \tag{2.51}$$

Covariation function is not symmetric in k , that is, $\lambda_x(k) \neq \lambda_x(-k)$. Some authors use the term covariation to define the numerator of equation (2.50). That is

$$[X_n, X_{n-k}]_\alpha = E(X_n {X_{n-k}}^{[\beta-1]}). \tag{2.52}$$

Proposition 2.4.3. *Let* $(X,\ Y_1,\ Y_2)$ *be jointly alpha stable,* $\alpha > 1$*, with* Y_1 *and* Y_2 *independent. Then*

$$[X, Y_1 + Y_2]_\alpha = [X, Y_1]_\alpha + [X, Y_1]_\alpha.$$

Proof. See Samorodnitsky and Taqqu (1994), page 93. □

Gallagher (2001) studied the limiting behavior of sample auto-covariation function. This study leads us to utilize sample auto-covariation as a tool in time series analysis of stable processes and is defined as

$$\widehat{\lambda}_x(k) = \frac{\sum_{n=s}^{r} X_n sign(X_{n-k})}{\sum_{n=1}^{N} |X_n|}, \tag{2.53}$$

where N is the sample size, $s = max(1, 1+k)$ and $r = min(N, N+k)$. Then following Theorems by Gallagher (2000) gives the asymptotic behavior of sample auto-covariation function.

Theorem 2.4.4. *If* X_n *is a stationary sequence with* $E|X_n| < \infty$*, then as* $N \to \infty$*,*

$$\hat{\lambda}(l) \ \stackrel{as}{\to}\ \lambda(l).$$

Proof. See Gallagher (2000). □

Theorem 2.4.5. *Let* $\{X_n\}$ *be a strictly stationary stable ARMA process with infinite order moving average representation,*

$$X_n = \sum_{j=0}^{\infty} \psi_j a_{n-j};\ \sum_{j=0}^{\infty} j|\psi_j|^\delta < \infty$$

and for any l *let* $\lambda_l = (\lambda(-l), ..., \lambda(l))$ *and* $\widehat{\lambda}_l = (\widehat{\lambda}(-l), ..., \widehat{\lambda}(l))$.

(i) *If* $E|a_n|^2 < \infty$ *then,*

$$N^{1/2}(\widehat{\lambda}_l - \lambda_l) \xrightarrow{L} (E|X_1|)^{-1}\mathrm{X},$$

where, X *is a multivariate normal random vector.*

(ii) *If* $a_1 \sim S_\alpha(\gamma)$ *with* $\alpha > 1$,

$$N^{1-1/\alpha}(\widehat{\lambda}_l - \lambda_l) \xrightarrow{L} (E|X_1|)^{-1}\mathrm{S},$$

where, S *has a multivariate stable distribution.*

Proof. See Gallagher (2000). □

Remark 2.4.6. The limiting vector S can be represented as $S = W[S_1, ..., S_{2l}]'$, where $S_1, ..., S_{2l}$ are iid skewed stable random variables and the matrix W is given by $W = (E|X_1|)^2 AVA'$.

The matrix A is defined as, $A = \begin{pmatrix} I_l & c_1 & 0 \\ 0 & c_2 & I_l \end{pmatrix}$, where I_l is the $l \times l$ identity matrix, 0 is a matrix with all zero entries, $c_1 = (-\lambda(l), ..., -\lambda(1))'$ and $c_2 = (-\lambda(-1), ..., -\lambda(-l))'$ and

$$V = E\left(Y_0 Y_0' + \left(\sum_{t=1}^{N} Y_t\right) Y_0' + Y_0 \left(\sum_{t=1}^{N} Y_t\right)'\right),$$

where $Y_t = (sign(X_t)\, X_{t+l} - E(sign(X_t)\, X_{t+l}), ..., sign(X_t)\, X_{t-l} - E(sign(X_t)\, X_{t-l}))'$

Theorem 2.4.7. *If* X_n *is given as in Theorem 2.4.5 with* $a_t \sim S_\alpha(\gamma)$ *then*

$$N^{1-1/\alpha}(\hat{\lambda}(l) - \lambda(l)) \Rightarrow \sigma S,$$

where S has a univariate stable distribution,

$$\lambda(l) = \frac{\sum_{j=l}^{\infty} \psi_j |\psi_{j-l}|^{\alpha-1} sign(\psi_{j-l})}{\sum_{j=0}^{\infty} |\psi_j|^{\alpha}},$$

and

$$\sigma = (\sum_{j=0}^{\infty} |\psi_j|^{\alpha})^{-1/\alpha}.$$

Proof. See Gallagher (2000). □

From (2.13) we have seen that the autocorrelation function of a finite variance autoregressive process follows Yule-Walker equation. Parallel to this property, Kanter and Steiger (1974) and Nikias and Shao (1995) defined a Generalized Yule-Walker (GYW) equation based on auto-covariation function. For an $AR(p)$ model GYW equations can be written as,

$$\lambda_x(k) = \sum_{i=1}^{p} \phi_i \lambda_x(k-i),\ k \geq 1. \tag{2.54}$$

Solution of this equation provides the generalized Yule-Walker estimates of autoregressive parameter $\phi_i\,, i = 1, ..., p$ in terms of the sample auto-covariation function. Gallagher (2000) used this function for model identification of moving average models. Similar to autocorrelation function, we can extend the scope of this function and some of its properties for model identification of autoregressive models. In chapter 5 we discuss this issue by introducing partial auto-covariation function and generalized Durbin-Levinson algorithm. We have proposed new information criteria parallel to AIC based on this study.

2.5 Linear prediction problems in stable processes

We will start with some definitions before going to the detailed description. Most of the definitions in this section is taken from Bobrowski (2005), Kreyszig (1989).

Definition 2.5.1. Let $\mathbb{X}$ be a set; its elements will be denoted x, y, z, etc. A triplet $(\mathbb{X}, +, .)$, where $+$ is a map $+ : \mathbb{X} \times \mathbb{X} \to \mathbb{X}, (x, y) \to x + y$ and $.$ is a map $. : \mathbb{R} \times \mathbb{X} \to \mathbb{X}, (\alpha, x) \to \alpha x$, is called a (real) **linear space** if the following conditions are satisfied:

$(a1)$ $(x + y) + z = x + (y + z)$, for all $x, y, z \in \mathbb{X}$,
$(a2)$ there exists $\Theta \in \mathbb{X}$ such that $x + \Theta = x$, for all $x \in \mathbb{X}$,
$(a3)$ for all $x \in \mathbb{X}$ there exists an $x' \in \mathbb{X}$ such that $x + x' = \Theta$,
$(a4)$ $x + y = y + x$, for all $x, y \in \mathbb{X}$,
$(m1)$ $\alpha(\beta x) = (\alpha\beta)x$, for all $\alpha,\ \beta \in \mathbb{R}, x \in \mathbb{X}$,
$(m2)$ $1x = x$, for all $x \in \mathbb{X}$,
(d) $\alpha(x + y) = \alpha x + \alpha y$, *and* $(\alpha + \beta)x = \alpha x + \beta x$ for all $\alpha,\ \beta \in \mathbb{R}$ and $x, y \in \mathbb{X}$.

Definition 2.5.2. Let $\mathbb{X}$ be a linear space. A function $||.|| : \mathbb{X} \to \mathbb{R}$, $x \to ||x||$ is called a norm, if for all $x, y \in \mathbb{X}$ and $\alpha \in \mathbb{R}$

$(n1)$ $||x|| \geq 0$,
$(n2)$ $||x|| = 0$, iff $x = \Theta$,
$(n3)$ $||\alpha x|| = |\alpha|||x||$,
$(n4)$ $||x + y|| < ||x|| + ||y||$.

A pair $(\mathbb{X}, ||.||)$, where $\mathbb{X}$ is a linear space and $||.||$ is a norm in $\mathbb{X}$ called a normed linear space, and for simplicity we say that $\mathbb{X}$ itself is a **normed linear space** (or just normed space).

Note that if $||.||$ is a norm, then $d(x, y) = ||x - y||$ is a metric. This means that $(\mathbb{X}, d)$ is a **metric space**.

A **subspace** $(Y, \widetilde{d})$ of (X, d) is obtained if we take a subset $Y \subset X$ and restrict d to $Y \times Y$.

Definition 2.5.3. A sequence $(x_n),\ n \geq 1$ of elements of a normed linear space $\mathbb{X}$ is said to be **Cauchy sequence** if for all $\epsilon > 0$ there exists an $n_0 = n_0(\epsilon)$ such that $d(x_n, x_m) = ||x_n - x_m|| < \epsilon$, for all $n, m \geq n_0$.

We can prove that every convergent sequence is Cauchy. A metric space is termed **complete** if every Cauchy sequence of its elements is convergent.

Definition 2.5.4. If every Cauchy sequence in a normed linear space $\mathbb{X}$ is convergent, $\mathbb{X}$ is called a **Banach space**. We may say that a Banach space is a complete normed linear space.

Now let us define linear space of stable processes and covariation norm.

Definition 2.5.5. Let $X = \{X_t,\ t \in T\}$ be a stochastic process with underlying probability space $(\Omega, \mathcal{F}, P)$ such that $\{X_t \in L^p(\Omega, \mathcal{F}, P)\}$, for all $t \in T$, where $1 < p < \infty$, and let $l(X)$ be the space of all finite linear combination of $\{X_t,\ t \in T\}$. Then we call X a p-th order process and define a norm of $l(X)$ by

$$\|X\|_p = (E|X|^p)^{1/p}, \quad X \in l(X_t).$$

The linear space $L(X)$ of the process X is the completion of $l(X)$ with respect to this norm. That is, in $L^p(\Omega, \mathcal{F}, P)$.

When $\alpha > 1$, covariation introduces a norm on a linear space of alpha stable random variables, which is known as covariation norm (Samorodnitsky and Taqqu (1995), page 95), and is defined by

$$\|X\|_\alpha = (E|X|^\alpha)^{1/\alpha}.$$

From the definition we can show that the covariation norm is equal to the scale parameter of alpha stable distributions (Samorodnitsky and Taqqu (1995), page 95).

Property 2.5.6. If $X \sim S_\alpha(\gamma)$ with $\alpha > 1$, then $\|X\|_\alpha = \gamma$.

Linear prediction problem in stable process is complex due to the fact that the linear space of stable processes is Banach space when $1 < \alpha < 2$, (Bobrowski, 2005, page 51). When $1 < \alpha < 2$ and $p < \alpha$, the p-th order moment exists and similar to covariance in the

Gaussian case, we can use covariation defined in (2.52), and it reduces to covariance when $\alpha = 2$ (Cambanis and Miller (1981), Samorodnitsky and Taqqu (1995)). Cambanis and Miller (1981) studied the linear estimation problems of stable processes. Some observations and theorems are stated below.

Consider $\{\theta, X_t, t \in T\}$ with unknown parameter θ as a symmetric stable process, $1 < \alpha < 2$ and $L(X_t, t \in T)$ the linear space of symmetric stable process $X = \{X_t, t \in T\}$. The regression estimate of θ based on X_t is given by $\widehat{\theta} = E(\theta|X_t)$ is not in general linear and except for Gaussian case it necessarily does not belong to $L(X_t, t \in T)$. When T consists of one point, or T is a finite set and random variables X_t are independent then $E(\theta|X_t)$ is linear (Cambanis and Miller (1981)). Regression estimates are linear in some other cases, that is discussed below.

Minimum Dispersion Criteria: Under the minimum dispersion criterion, the best estimate of a alpha stable random variable in the linear space of observations is the one that minimizes the dispersion of the estimation error. Recall that the dispersion (scale parameter) of a stable random variable plays an analogous role of the variance. For example, the larger the dispersion of a stable random variable is, the more it spreads around the median. Thus, by minimizing the dispersion we minimize the average magnitude of estimation errors.

The generic linear estimation problem of stable processes based on minimum dispersion criteria can be formulated as follows. This result is analogous to the projection theorem in Banach space (Singer (1970), Hill (2003)). Before going to the details we need to define two concepts, James orthogonality and metric projection.

Definition 2.5.7. Let (x, y) be any arbitrary random variables of some Banach space $\mathbb{B}$. The random variable y is **James orthogonal** to x, whenever

$$\|y + \lambda x\| > \|y\|$$

for every real scalar $\lambda \in \mathbb{R}$, and is denoted by $y \perp_J x$.

Banach space norms $\|.\|$ may be supported by arbitrarily many semi-inner products $[,]$. However, $(x, y) \in \mathbb{B}$, if y is orthogonal to x there exists one inner-product that supports $[y, x] = 0$.

Consider arbitrary subspaces, that is, $U, V \subseteq \mathbb{B}$, $\sigma(V) \subset \sigma(U)$, where $\sigma(V)$ denotes the sigma algebra induced by the elements of V . For some element $u \in U$, we say $v \in V$ is the "best predictor" of u with respect to V if and only if

$$\|u - v\| \leq \|u - w\|$$

for every element $w \in V$. Since $\mathbb{B}$ is a Banach space, the predictor v exists and is unique.

Definition 2.5.8. Metric projection operator is a maps $P : U \to V$ such as $P(u|V) = v$. The projection $P(u|V)$ is identically the best predictor of u.

Orthogonality conditions:

(i) The element $v \in V$ satisfies $P(u|V) = v$ if and only if $(u - v) \perp_J V$.

(ii) For every $w \in V$, and for a unique $[,]$, $(u - v) \perp_J V$ if and only if $[u - v, w] = 0$.

The linear estimate of θ based on X_t under minimum dispersion criterion is defined as the best approximation to θ in $L(X)$. The problem is to find a random variable $\widehat{\theta}$ in $L(X_t, t \in T)$ with minimum distance from θ. That is,

$$\|\theta - \hat{\theta}\|_\alpha = \inf_{Z \in L(X_t, t \in T)} \|\theta - Z\|_\alpha \tag{2.55}$$

or equivalently

$$E|\theta - \hat{\theta}|^p = \inf_{X \in L(X_t, t \in T)} E|\theta - X|^p, \text{for} 0 < p < \alpha. \tag{2.56}$$

The estimate $\hat{\theta}$ is denoted by $l(\theta|X)$. Since $L(X_t)$ is a Banach space, $\hat{\theta}$ exist and is unique for $1 < \alpha < 2$. The linear estimate $\hat{\theta}$ is obtained by a metric projection of θ onto the Banach space $L(X_t, t \in T)$ (see Singer (1970), Shao and Nikias (1993), Hill (2003)). For $1 < \alpha < 2$, $\hat{\theta}$ is also uniquely determined by either of the following equations based on orthogonality conditions (Cambanis and Miller (1981))

$$\begin{aligned}[\xi,\, \hat{\theta} - \theta]_\alpha &= 0, \text{ for all } \xi \in L(X) \\ [X_t, \hat{\theta} - \theta]_\alpha &= 0, \text{ for all } t \in T.\end{aligned} \tag{2.57}$$

This is analogous to the orthogonality principle used extensively in the linear estimation problem of second-order processes (see Hill (2003)). When $\alpha = 2$, the equations in (2.57) are linear and thus closed-form solutions exist for θ. For $\alpha < 2$, it is highly nonlinear and hard to solve for θ. Cambanis and Miller (1981), Samorodnitsky and Taqqu (1994) and Cline and Brockwell (1985) studied the linear estimation problems of stable processes using the minimum dispersion criteria. Some observations and theorems are stated below. Samorodnitsky and Taqqu (1994) state one of the regression properties of the stable random variables which is stated in the following theorem. This result is analogous to the result stated in (2.57).

Theorem 2.5.9. *If $X_0, X_1, ..., X_n$ are jointly alpha stable distributed random variables with $1 < \alpha \leq 2$ then,*

$$E(X_{n+1}|X_1, X_2, ..., X_n) = \phi_1 X_1 + ... + \phi_n X_n, \tag{2.58}$$

if and only if

$$\sum_{i=1}^{n} \phi_i \lambda(j-i) = \lambda(j),\ j = 1, ..., n. \tag{2.59}$$

Proof. See Samorodnitsky and Taqqu (1994), page 176-177. □

Cline and Brockwell (1985) defined a linear predictor with infinite past of stable $ARMA(p, q)$ process using minimum dispersion criteria. The problem is to find an optimal predictor for X_{n+k}, $k \geq 1$, of the form $\sum_{j=1}^{\infty} \alpha_j X_{n-j+1}$, which is a linear predictor based on an infinite past. If $\{X_n\}$ is a pure autoregressive (AR(p)) and $n > p$, the truncated predictor $\sum_{j=1}^{n} \alpha_j X_{n-j+1}$ will in fact be optimal.

Thus if $a_t's$ are iid symmetric stable random variables with index α and if $\sum_{j=-\infty}^{\infty} |\psi_j|^\alpha < \infty$,

then $Y = \sum_{j=-\infty}^{\infty} \psi_j a_j$ is also symmetric stable with

$$disp(Y) = \sum\nolimits_{j=-\infty}^{\infty} |\psi_j|^{\alpha}. \tag{2.60}$$

We have already seen that the the stable ARMA process $\{X_n\}$ can be expressed as the moving average $X_n = \sum_{j=-\infty}^{\infty} \psi_{n-j} a_j$ with $\psi_j = 0,\ j < 0$.

Hence $disp(X_t) = \sum_{j=-\infty}^{\infty} |\psi_j|^{\alpha}$. If $Y = \sum_{j=-\infty}^{\infty} \psi_j a_j$ then we define the minimum error dispersion linear predictor of Y (based on $X_1, ..., X_n,$) to be the linear combination $\widehat{Y} = b_1 X_1 + ... + b_n X_n = \mathbf{b}'\mathbf{X}$, which minimizes

$$disp(Y - \widehat{Y}) = \sum\nolimits_{j=-\infty}^{\infty} |\psi_j - (b_1 \psi_{n-j} + ... + b_n \psi_{1-j})|^{\alpha}.$$

In the k-step ahead forecasting case we take $Y = X_{n+k}$, and minimize

$$disp(X_{n+k} - \widehat{X}_{n+k}) = \sum\nolimits_{j=0}^{k-1} |\psi_j|^{\alpha} + \sum\nolimits_{j=k}^{\infty} |\psi_j - (b_1 \psi_{n-j} + ... + b_n \psi_{1-j})|^{\alpha}.$$

For a linear process driven by symmetric stable noise, the prediction error for any linear predictor also has symmetric stable distribution. The minimum dispersion prediction error has the distribution with the smallest scale and hence is optimal. The procedure is easily extended to more general linear processes, since it requires only the knowledge of coefficients of the process and of the tail index α of the noise distribution. Following theorem by Cline and Brockwell (1985) indicates that the prediction error dispersion is roughly proportional to the probability of a large prediction error. A corollary of this is that among linear predictors, the minimum dispersion predictor is optimal in the sense that it minimizes the probability of large prediction errors.

Theorem 2.5.10. *Suppose $\{a_j\}$ are independent and identically distributed symmetric stable random variables and $Y = \sum_{j=-\infty}^{\infty} \psi_j a_j$ where $\sum_{j=-\infty}^{\infty} |\psi_j|^{\delta} < \infty$, for some $\delta < min(1, \alpha)$.*

Then Y exists almost surely (is absolutely convergent) and

$$\lim_{t\to\infty} \frac{P[|Y| > t]}{P[|a_1| > t]} = disp(Y) = \sum\nolimits_{j=-\infty}^{\infty} |\psi_j|^{\alpha}. \tag{2.61}$$

Proof. See Cline and Brockwell (1985). □

The following results of Cline and Brockwell (1985) will be seen to be identical to the corresponding results for least square prediction of the finite variance process.

Lemma 2.5.11. *Fix $\delta < \min(1, \alpha)$ and let $\{X_t\}$ be the stable $ARMA(p, q)$ process defined by (2.46). Let $\{v_j\}$, $\{\rho_j\}$ are sequences of constants and S_* be the class of random variables of the form $\sum\limits_{j=n+1}^{\infty} \rho_j a_j + \sum\limits_{j=1}^{\infty} v_j a_{n+1-j}$, where,*

$$\sum_{j=n+1}^{\infty} |\rho_j|^{\delta} < \infty \text{ and } \sum_{j=n+1}^{\infty} |v_j|^{\delta} < \infty.$$

Then for each $Y \in S_*$, the set

$$P_{\infty}Y = \left\{ \sum_{j=1}^{\infty} \beta_j X_{n+1-j} : disp(Y - \sum_{j=1}^{\infty} v_j X_{n+1-j}) \, is \, \text{minimum} \right\}$$

consist of exactly one element. For

$$Y = \sum_{j=n+1}^{\infty} \rho_j a_1 + \sum_{j=1}^{\infty} v_j X_{n+1-j},$$

this element is $Y^* = \sum\limits_{j=1}^{\infty} v_j X_{n+1-j}$. Further more the mapping $Y \to Y^*$ is linear on S_*.

Proof. See Cline and Brockwell (1985). □

Remark: For symmetric stable process with $\alpha > 1$, we have

$$Y^* = E(Y|X_n, X_{n-1}, ...).$$

(Cambanis and Miller 1981).

Theorem 2.5.12. *For the stable ARMA process there exists a unique minimum dispersion linear predictor X^*_{n+k} for X_{n+k} , $k \geq 1$, based on the infinite past $X_n, X_{n-1}, ...$ .This predictor satisfies the recursive relation*

$$X^*_{n+k} = \sum_{j=1}^{k-1} \psi_j X^*_{n+k-j} + \sum_{j=k}^{\infty} \psi_j X_{n+k-j}. \tag{2.62}$$

Proof. See Cline and Brockwell (1985). □

In practical situations we have only finite number of data points $X_1, X_2, ..., X_n$. For any $Y \in S_*$ one can use the truncated predictor $Y^*(n) = \sum_{j=1}^{n} v_j X_{n+1-j}$, where v_j is defined as in Lemma 2.5.11. The truncated predictor is in fact optimal when the process is purely autoregressive and n is large enough, that is when X_n satisfy

$$X_n = \phi_1 X_{n-1} - ... - \phi_p X_{n-p} + a_n \tag{2.63}$$

and $n \geq p$.

Lemma 2.5.13. *Let S_* be the class of random variables of the form $Y = Z + \nu' X_n$ for some $\nu \in R^n$ and $Z = \sum_{j=n+1}^{\infty} \rho_j u_j$ such that Z exists. Then for each $Y \in S_*$, the set*

$$P_n Y = \{\beta' X_n : disp(Y - a' X_n) \text{ is minimum}\}$$

consists of exactly one variable. For $Y = Z + \nu' X_n$ this unique variable is $\widehat{Y} = \nu' X_n$. Furthermore, the mapping $Y \to \hat{Y}$ is linear in S_*.

Proof. See Cline and Brockwell (1985). □

Corollary 2.5.14. *For the process* (2.63)*, provided $n \geq p$, there exist a unique minimum dispersion linear predictor $\widehat{X}_{n+k}$ for $X_{n+k}, k \geq 1$ in terms of $X_1, X_2, ..., X_n$. This predictor*

satisfy the recursive relationship

$$\hat{X}_{n+k} = \phi_1 \hat{X}_{n+k-1} + \phi_2 \hat{X}_{n+k-2} + ... + \phi_1 \hat{X}_{n+k-p} \tag{2.64}$$

with initial conditions $\hat{X}_j = X_j$ *for* $1 \leq j \leq n$.

Proof. See Cline and Brockwell (1985). □

From Corollary 2.5.14 and for $\alpha > 1$ we can say that,

$$\hat{X}_{n+k} = E(X_{n+k}|X_{n+k-1}, ..., X_{n+k-p}),$$

which is the minimum dispersion predictor of X_{n+k} given $X_{n+k-1}, ..., X_{n+k-p}$.

2.6 Some theoretical results

Definitions and theorems stated in this section are taken from Brockwell and Davis (1987).

Definition 2.6.1. We say that X_n converges in probability to X, written $X_n - X = o_p(1)$ or $(X_n \xrightarrow{p} X)$, if for every $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Definition 2.6.2. We say that the sequence X_n is bounded in probability (or tight), written $X_n = O_p(1)$, if for every $\epsilon > 0$, there exist $\delta(\epsilon) \in (0, \infty)$ such that,

$$P(|X_n| > \delta(\epsilon)) < \epsilon \text{ for all } n.$$

Definition 2.6.3. Converges in probability and order in probability: If $X_n = \{X_{ni}\}$ is a sequence of $k-$diamentional random vectors then,

$$\text{(i) } X_n = o_p(a_n) \text{if and only if } a_n^{-1} X_{ni} = o_p(1).$$

(ii) $X_n = O_p(a_n)$if and only if $a_n^{-1}X_{ni} = O_p(1)$.

Proposition 2.6.4. *If $\{X_n\}$ and $\{Y_n\}$ are two sequences of random $k-$vectors such that $X_n - Y_n = o_p(1)$ and $X_n \xrightarrow{L} X$, then $Y_n \xrightarrow{L} X$.*

Proof. See Brockwell and Davis (1987), page 198. □

Proposition 2.6.5. *If $\{X_n\}$ is a sequence of $k-$diamentional random vectors such that $X_n \xrightarrow{p} X$ and if $g : R^k \to R^m$ is a continuous mapping, then $g(X_n) \xrightarrow{p} g(X)$.*

Proof. See Brockwell and Davis (1987), page 193. □

Proposition 2.6.6. *If $\{X_n\}$ is a sequence of random $k \times 1$ vectors such that*

$$X_n - a = O_p(r_n),$$

where $a \in R^k$ and $r_n \to 0$ as $n \to \infty$. If g is a function from R^k into R such that the derivatives $\frac{\partial g}{\partial x_i}$ are continuous in a neighborhood $N(a)$ of a, then,

$$g(X_n) = g(a) + \sum_{i=1}^{k} \frac{\partial g}{\partial x_i}(a)(X_{ni} - a_i) + o_p(r_n).$$

Proof. See Brockwell and Davis (1987), page 195. □

The results described in this chapter are used to establish the statistical properties of the estimators and forecasts in the forth coming chapters.

Chapter 3

Statistical Signal Extraction using Stable Processes

3.1 Introduction

Statistical signal extraction has attracted the attention of researchers in areas as diverse as engineering, medicine, economics, finance and climatology, to name but a few. A lot of techniques are available, including Wiener Kolmogorov filtering, Kalman filtering, principle component analysis and wavelet analysis (see Pollock (2005)) for statistical signal extraction. Most of the theoretical developments in statistical signal extractions assume that the signal and/or noise follow certain auto-regressive moving average (ARMA) models with Gaussian or other exponential family of innovation distributions with finite second and higher order moments. Then the signals are expressed as linear filters of the observations, where the optimum filter weights are obtained using the method of minimum mean square error (MMSE) (see Bell and Martin (2004) for details).

Signal extraction problem can be stated as below. The observed data process $\{Y_t\}$ is often depicted as a combination of signal $\{X_t\}$ and noise $\{N_t\}$ as follows:

$$Y_t = X_t + N_t. \tag{3.1}$$

Here we assume that the signal $\{X_t\}$ and noise $\{N_t\}$ are independent. The objective is to use the data on $\{Y_t\}$ to estimate the unobserved components $\{X_t\}$ and $\{N_t\}$.

The modern theory of statistical signal extraction was formulated independently by Wiener (1949) and Kolmogorov (1941). The two proponent of the theory established it in different ways. Wiener worked primarily in the frequency domain, whereas Kolmogorov took a time domain approach to the problem. The unification of two approaches has been provided by Whittle (1983). Another important theory of signal extraction was introduced by Kalman (1960) and Kalman and Bucy (1961), which dealt with the filtering and forecasting of time varying linear stochastic systems. This approach proposes a state space model for the system. The formulation of Kalman and Bucy was sufficiently general to subsume the theory of Wiener and Kolmogorov. Econometricians commonly cite the paper of Burridge and Wallis (1988) in the context of signal extraction. The latter shows how, in the case of a stationary ARMA process, the Kalman filter converges asymptotically to the backward looking Wiener Kolmogorov filter applied to a semi-infinite data series. There has recently been a great deal of interest in estimating components from time series using signal extraction procedure. Signal extracting results for optimal (MMSE) linear estimators of the components were given in the stationary case by Wiener Kolmogorov Filter. Hannan (1967) extended the procedure to the case of non stationary signal and stationary noise. Bell (1984a) gave a more general treatment that covered the case were both signal and noise are non stationary. Bell and Martin (2004) studied the asymmetric signal extraction filter and mean square error of non stationary signal and noise model. These studies dealt with estimation of signal and noise from an infinite realization of the observed data.

In the previous discussion we have seen that the standard models for statistical signal extraction assume that the signal and noise are generated by linear Gaussian processes or other exponential family of innovation distributions with finite second and higher order moments. These models may not be suitable to model when the data show sharp spikes or occasional bursts of outlying observations discussed in the previous chapters. Heavy tailed distribution such as stable distribution can be considered as a good candidate. Now let us

discuss some of the works carried out in this direction. Stuck and Kleiner (1974) generalized Kalman filter in the case of state space models with stable distributed noise. Cline and Brockwell (1985) introduced a prediction theory for stable ARMA models. Breton and Musiela (1993) proposed the generalization of Kalman filter to models with infinite variance. They studied the problem of optimal linear estimation for continuous time processes. The optimal filter is given by recursive equations which reduce to the classical Kalman-Bucy equations when the system is driven by independent white noises. Shao and Nikias (1993) discussed the issues of signal extraction problems in stable processes and proposed an adaptive Wiener filter for stable processes. Bidarkota and McCulloch (1998) developed a univariate state space model with symmetric stable shocks for monthly inflation in the United States . The non-Gaussian state space model is estimated by the Sorenson-Alspach filtering algorithm. McCulloch and Bidarkota (2003) discussed various signal extraction problems for processes with stable distributed innovations. Sornette and Ide (2001) introduced Kalman-Levy filter which is a generalization of Kalman filter for heavy tailed processes. Gordon et al. (2003) used stable innovations models and Kalman-Levy filter for tracking manoeuvering targets. These works motivated us to generalize the signal extraction problem in the case of heavy tailed innovations.

In the present chapter our main objective is to discuss the properties of the signal extraction model specified by a signal plus noise model when the signal and noise are assumed to follow stationary ARMA models with symmetric stable innovations. The estimated signals are expressed as linear filters of the observations, where the optimum filter weights are obtained using the method of minimum dispersion criteria. Signal extraction filter based on minimum dispersion criteria, which minimize dispersion of signal extraction error. For the sake of simplicity we start with a semi infinite filter and then extend the scope of this filter to the case of doubly infinite filter and asymmetric filter discussed in the literature. Infinite length filters have limited applications in many situations so we derived a finite length filter using Kalman-Levy filter and identified the pattern of the filter weights. This has been achieved using a state space representation of the signal plus noise model. Per-

formance of the proposed filters are compared with Wiener Kolmogorov filter and Kalman filter through simulation.

The chapter is split in to six sections : In the second section we discuss the statistical models for signal extraction and classical signal extraction procedures. Section 3.3 introduces the mathematical representation of signal and noise processes based on symmetric stable assumption. Section 3.4 includes the minimum dispersion signal extraction criteria and we discuss how symmetric stable distribution can be embedded in to this frame work. Section 3.5 consists of finite length signal extraction filtering based on Kalamn-Levy filter. Last section contains some simulation results.

3.2 Statistical Models for Signal Extraction

Suppose that the observations are generated by the model (3.1). In the next subsection we briefly discuss the signal extraction problem of stationary signal and noise processes in both time and frequency domain.

3.2.1 Wiener Kolmogorov filtering theory

Classical time series methods consist of estimating the signal X_t as a linear combination of an observed signal Y_t, specified by,

$$\widehat{X}_t = \sum_{k=-\infty}^{\infty} w_k Y_{t-k} = W(B)Y_t, \tag{3.2}$$

where, $W(B) = \sum_j w_j B^j$, is a polynomial in lag B. The optimal linear filter coefficients w_k are obtained using minimum mean square error criteria. That is, in general we seek a set of filter coefficients w_k which minimizes the mean squared error of the estimator, say,

$$MSE = E[(\widehat{X}_t - \sum_{k=-\infty}^{\infty} w_k Y_{t-k})^2]. \tag{3.3}$$

This problem was originally solved by Kolmogorov (1941) in time domain and by Wiener (1949) in frequency domain. From minimum mean square error method, we can write,

$$E[(\widehat{X}_t - \sum\nolimits_{k=-\infty}^{\infty} w_k Y_{t-k})Y_{t-r}] = 0,$$

for $r = 0, \pm 1, \pm 2, ...$, which leads to

$$\sum\nolimits_{k=-\infty}^{\infty} w_k \gamma_{yy}(r-k) = \gamma_{xy}(r), \tag{3.4}$$

to be solved for the filter coefficients. Equation (3.4) can be represented as

$$W(B)\gamma_{yy}(B) = \gamma_{xy}(B), \tag{3.5}$$

where, $\gamma_{yy}(B) = \sum_k \gamma_{yy}(k)B^k$ and $\gamma_{xy}(B) = \sum_k \gamma_{xy}(k)B^k$. Substituting the spectral representations discussed in Theorem 2.2.15 for the auto-covariance functions into the above equation (3.4) and identifying the spectral densities through the uniqueness of the Fourier transform produces the following relation,

$$W(\omega)f_{yy}(\omega) = f_{xy}(\omega), \tag{3.6}$$

where, $W(\omega) = \sum w_k e^{i\,\omega_k}$ and w_k are Fourier transform pairs. Since signal and noise components are independent, we can show that $f_{xy}(\omega) = f_{xx}(\omega)$ and $f_{yy}(\omega) = f_{xx}(\omega) + f_{nn}(\omega)$. Hence the optimum filter would be the Fourier transform of

$$W(\omega) = \frac{1}{1 + \frac{f_{nn}(\omega)}{f_{xx}(\omega)}}, \tag{3.7}$$

where the second term in the denominator is just the inverse of the signal to noise ratio, say,

$$SNR(\omega) = \frac{f_{xx}(\omega)}{f_{nn}(\omega)}.$$

Signal-to-noise ratio (often abbreviated SNR or S/N) is a measure used to quantify how much a signal has been corrupted by noise. It is defined as the ratio of signal power to the noise power corrupting the signal. signal-to-noise ratio compares the level of a desired signal to the level of background noise. The higher the ratio, the less obtrusive the background noise is.

The result shows the optimum filters can be computed for this model if the signal and noise spectra are both known or if we can assume knowledge of the signal-to-noise ratio $SNR(\omega)$ as function of frequency. If the signal-to-noise ratio is known, the optimal filter can be computed by the inverse transform of the function $W(\omega)$ (for details see Shumway and Stoffer (2005), page 252).

From these discussions we have seen that the signal extraction filters are functions of spectral density or auto-covariance function of $\{Y_t\}$, $\{X_t\}$, and $\{N_t\}$. Filter weights w_j or frequency response function, $W(\omega)$ can be computed from equation (3.4) and (3.6). Now we will assume that the signal and/or noise follow auto-regressive moving average (ARMA) models defined in equation (2.12) with independent white noise Gaussian innovation processes $\{a_t\}$ and $\{b_t\}$ having respective variances σ_a^2 and σ_b^2. They are described in terms of the models as:

$$\begin{aligned} \phi_x(B)X_t &= \theta_x(B)a_t, \\ \phi_n(B)N_t &= \theta_n(B)b_t. \end{aligned} \tag{3.8}$$

From (3.1) and (3.8) we can represent Y_t as,

$$\phi(B)Y_t = \theta(B)e_t. \tag{3.9}$$

Here we assume that $\{X_t\}$, $\{N_t\}$ are zero mean processes, so the mean of $\{Y_t\}$ is also zero. The series $\{e_t\}$ is a white noise sequence with variance σ_e^2. From above two equations we can see that,

$$\phi(B) = \phi_x(B)\phi_n(B).$$

From equation (3.5), the optimum signal extraction filter $W(B)$ can be derived as

$$W(B) = \frac{\sigma_a^2}{\sigma_e^2} \frac{\theta_x(B)\theta_x(B^{-1})\phi_n(B)\phi_n(B^{-1})}{\theta(B)\theta(B^{-1})}. \tag{3.10}$$

The optimum filter in the frequency domain $W(\omega)$ can be obtained as,

$$W(\omega) = \frac{\sigma_a^2}{\sigma_e^2} \frac{|\theta_x(\omega)|^2|\phi_n(\omega)|^2}{|\theta(\omega)|^2}. \tag{3.11}$$

Example: To illustrate the signal extraction problem, let us assume that $Y_t = X_t + b_t$ and the signal X_t follows an $AR(1)$ model

$$X_t = \phi X_{t-1} + a_t,$$

where, $\{a_t\}$ and $\{b_t\}$ are mutually independent zero mean white Gaussian noise process with respective variances σ_a^2 and σ_b^2. From the model (3.9) we can show that

$$Y_t - \phi Y_{t-1} = a_t + b_t - \phi b_{t-1}. \tag{3.12}$$

Right hand side of (3.12) has variance $\sigma_a^2 + (1+\phi^2)\sigma_b^2$ and lag-1 auto-covariance $-\phi\sigma_b^2$, and the covariances of higher lag are zero. It is thus an $MA(1)$ model and the observed process $\{Y_t\}$ follows an $ARMA(1,1)$ model,

$$Y_t - \phi Y_{t-1} = e_t + \theta e_{t-1}. \tag{3.13}$$

From the above two equations we can show that,

$$\begin{aligned} &\sigma_a^2 + (1+\phi^2)\sigma_b^2 = (1+\theta^2)\sigma_e^2, \\ &\text{and} \qquad -\phi\sigma_b^2 = \theta\sigma_e^2. \end{aligned}$$

Now we can calculate,

$$\frac{\sigma_a^2}{\sigma_e^2} = \frac{\theta^2\phi + (1+\phi^2)\theta + \phi}{\phi}.$$

Using these results in equations (3.10) and (3.11) we get,

$$\begin{aligned} W(B) &= w_0 \textstyle\sum_{k=0}^{\infty} \theta^k B^k, \\ W(\omega) &= w_0 \textstyle\sum_{k=0}^{\infty} \theta^k e^{ik\omega}, \end{aligned} \tag{3.14}$$

where, $w_0 = \frac{\theta^2\phi+(1+\phi^2)\theta+\phi}{\phi(1-\theta^2)}$. From the above example it is seen that the obtained filter has infinite length.

Signal extraction problem of non-stationary signal and stationary noise was discussed by Hannan (1970), Sobel (1967) and Cleveland and Tiao (1976). Bell (1984a) gave a general treatment to the case when both signal and noise are non-stationary. Most of these methods provide estimates of signal and noise based on an infinite realization of an observed signal $\{Y_t\}$, a case that applies approximately when the observed series is sufficiently long. Linear MMSE estimators based on finite sample of observed series has been studied by Ansley and Kohn (1985), Kohn and Ansley (1987) and Bell and Hillmer (1988). Bell and Martin (2004) studied the asymmetric signal extraction problem of ARIMA component models. Another important signal extraction procedure is the well-known Kalman filtering and it is based on state space models. This method solves the difficulties associated with the infinite length filters. Next we will discuss the details of Kalman filtering for finite length signal extraction.

3.2.2 State space representation and Kalman filtering

Kalman filter is widely used for studying the linear dynamic systems in almost all areas of science and technology. A very general model that seems to subsume a whole class of special cases of interest in much the same way that linear regression does is the state-space model or the dynamic linear model, which was introduced in Kalman (1960) and

Kalman and Bucy (1961). Although the model was originally introduced as a method primarily for uses in aerospace-related research, it has been applied to modeling data from economics, medicine and the soil sciences (Shumway and Stoffer (2006), page 330). An excellent modern treatment of time series analysis based on the state space model is the text by Durbin and Koopman (2001). A paper by Meinhold and Singapurwalla (1983) made this topic easily understandable by statistician through Bayesian inference. The formulation of Kalman and Bucy was sufficiently general to subsume the theory of Wiener and Kolmogorov. This had been clearly apparent from the beginning. However, it has taken a while for these two approaches to statistical signal extraction to be unified, and each continues to have its separate adherents. Nevertheless, it is now widely recognized that they stand on a common ground. There have been numerous offerings that have demonstrated the connection. Kalman filtering have had a profound impact in time series analysis and its application areas (Box et al. (1994), page 164, Brockwell and Davis(1987), page 447 , Shumway and Stoffer (2006), page 330, Koopman and Harvey(2003)).

The state-space model or dynamic linear model (DLM), in its basic form, employs an order one, vector auto-regression as the state equation,

$$\mathbf{X}_{t+1} = Z_t \mathbf{X}_t + H_t \eta_t, \tag{3.15}$$

where the state equation determines the rule for the generation of the $p \times 1$ state vector $\mathbf{X}_t$ from the past $p \times 1$ state $\mathbf{X}_{t-1}$, for time points $t = 1, ..., n$. The matrix Z_t is known as state transition matrix and H_t is a non-random matrix and both are assumed to be known. We assume the η_t are $p \times 1$ independent and identically distributed, zero-mean normal vectors with covariance matrix Q_t. In the DLM, we assume the process starts with a normal vector $\mathbf{X}_0$ that has mean μ_0 and $p \times p$ covariance matrix Σ_0.

The DLM, however, adds an additional component to the model in assuming that we do not observe the state vector $\mathbf{X}_t$ directly, but only a linear transformed version of it with noise added, say

$$Y_t = T_t \mathbf{X}_t + G_t \varepsilon_t, \tag{3.16}$$

where T_t is a $q \times p$ measurement or observation matrix and the matrix G_t is assumed to be known; equation (3.16) is called the observation equation. The additive observation noise ε_t is assumed to be white and Gaussian with $q \times p$ covariance matrix R_t. In addition, we usually assume, for simplicity, η_t and ε_t are uncorrelated.

Example: Assume that an observed series Y_t follows model (3.1) with a stationary $AR(p)$ model $\phi(B)X_t = a_t$ and a white noise sequence N_t. Now we can study state space representation of this model as follows.

If we define,

$$Z = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

and define $\mathbf{X}_t = [X_t, X_{t-1}, ..., X_{t-p+1}]^T$, $H = [1, 0.....,0]^T$, $T = [1, 0, ..., 0]$, then we can write

$$\begin{aligned} \mathbf{X}_t &= Z\mathbf{X}_{t-1} + Ha_t, \\ Y_t &= T\mathbf{X}_t + N_t. \end{aligned} \tag{3.17}$$

Similarly causal invertible $ARMA(p, q)$ process with observational noise can be represented as a state space model (see Brockwell and Davis (1987), page 453). Primary aim of the analysis involving state space model would be to produce estimates of the unobserved state variable $\mathbf{X}_t$ based on the observed sequence $\mathbf{Y}_k = \{y_1, y_2, ..., y_k\}$, to time k. When $k < t$ the problem is called prediction or forecasting. When $k = t$, the problem is called filtering, and when $k > t$, the problem is called smoothing. The solution to these problems is accomplished via the Kalman filter and predictor.

Kalman filtering is a recursive method which includes mainly two steps: (i) estimate

the future state through prediction step and (ii) current state estimation through filtering (update) step. These two estimation steps are accomplished through some set of recursive equations. A predictive estimator is an estimator of the state vector at time t based on observation up to and including time $t-1$. Kalman filter computes the predictive estimator $\widehat{\mathbf{X}}_t$, the MMSE estimator of the state variable $\mathbf{X}_t$ conditional on the observations $\mathbf{Y}_t = \{y_1, y_2, ..., y_{t-1}\}$. In prediction step we also calculate the covariance matrix of the estimation error, $\widehat{\mathbf{P}}_{t|t-1}$. From this we can write, $\widehat{\mathbf{X}}_{t|t-1} = E(\mathbf{X}_t|y_1, y_2, ..., y_{t-1})$ and $\widehat{\mathbf{P}}_{t|t-1} = E((\mathbf{X}_{t|t-1} - \widehat{\mathbf{X}}_{t|t-1})'(\mathbf{X}_{t|t-1} - \widehat{\mathbf{X}}_{t|t-1}))$. Similarly the current state estimate and corresponding error covariance matrix can be obtained through filtering step. Current state estimates are defined as $\widehat{\mathbf{X}}_{t|t} = E(\mathbf{X}_t|y_1, y_2, ..., y_t)$ and $\widehat{\mathbf{P}}_{t|t} = E((\mathbf{X}_{t|t} - \widehat{\mathbf{X}}_{t|t})'(\mathbf{X}_{t|t} - \widehat{\mathbf{X}}_{t|t}))$. It is known that, starting from some appropriate initial values $\widehat{\mathbf{X}}_{0|0}$ and $\widehat{\mathbf{P}}_{0|0}$, the optimal state estimates are given through the following recursive relations:

$$\begin{aligned} \widehat{\mathbf{X}}_{t|t-1} &= Z_t \widehat{\mathbf{X}}_{t-1|t-1}, \\ \widehat{\mathbf{P}}_{t|t-1} &= Z_t \widehat{\mathbf{P}}_{t-1|t-1} Z_t' + H_t Q_t H_t', \end{aligned} \tag{3.18}$$

with,

$$\begin{aligned} \widehat{\mathbf{X}}_{t|t} &= \widehat{\mathbf{X}}_{t|t-1} + K_t(\mathbf{Y}_t - T_t \widehat{\mathbf{X}}_{t|t-1}), \\ \widehat{\mathbf{P}}_{t|t} &= \widehat{\mathbf{P}}_{t|t-1} - K_t T_t \widehat{\mathbf{P}}_{t|t-1}, \end{aligned} \tag{3.19}$$

where,

$$K_t = \widehat{\mathbf{P}}_{t|t-1} T_t' (T_t \widehat{\mathbf{P}}_{t|t-1} T_t' + G_t R_t G_t')^{-1}. \tag{3.20}$$

The matrix K_t is known as Kalman gain. A paper by Meinhold and Singapurwalla (1983) familiarize this concept in terms of sequential Bayesian approach and derived these equations in terms of sequential Bayes estimate. Box et al. (1994) discussed the state space representation of $ARIMA$ model and proposed an exact forecasting using Kalman filter. This representation can be used for model specification and maximum likelihood estimation. In the up coming section we will focus on the signal extraction problem while assuming that the signal and noise processes follow certain stable time series models.

3.3 Stable time series models and signal extraction

Our objective in this section is to discuss the properties of the signal extraction model specified by (3.1) when the signal and noise are assumed to follow stationary $ARMA(p,q)$ models with symmetric stable innovation distributions defined in Proposition 2.3.6. We have seen that the signal extraction filters are functions of spectral density or auto-covariance function in classical set up. But these functions are ill-defined in infinite variance processes. Classical MMSE technique cannot be used due to the non-existence of second order moments of the symmetric stable distribution. We use the method of minimum dispersion to obtain the optimal filter for the signal, since the dispersion is well defined in stable distribution. Linear prediction theory of stable processes has been utilized for this purpose. Our main motivation in this chapter is to discuss the problems associated with signal extraction using stable processes and development of minimum dispersion signal extraction filter parallel to the classical methods. Detailed description of the signal extraction model is given below:

We assume that the observed process $\{Y_t\}$ follows the model (3.1) and signal $\{X_t\}$ and noise $\{N_t\}$ follow stable $ARMA$ (p,q) processes. The model structure is similar to (3.8) with $\{a_t\}$ and $\{b_t\}$ are mutually independent symmetric stable noise processes with scale parameters λ_a and λ_b respectively. This in turn implies that $\{X_t\}$ and $\{N_t\}$ are independent. As we discussed earlier our main objective is to obtain an estimate $\widehat{X}_t$ and $\widehat{N}_t$ of X_t and N_t by filtering $\{Y_t\}$ as

$$\begin{aligned} \widehat{X}_t &= W(B)Y_t\,, \\ \widehat{N}_t &= Y_t - \hat{X}_t = (1 - W(B))Y_t, \end{aligned} \tag{3.21}$$

where $W(B) = \sum_j w_j B^j$. Signal extraction error, ζ_t can be defined as,

$$\zeta_t = X_t - \widehat{X}_t = (1 - W(B))X_t - W(B)N_t.$$

We summarize the above discussion in the following proposition.

Proposition 3.3.1. *For an estimate $\hat{X}_t$ of the form* (3.21) *for X_t, the estimation error is given by*

$$\zeta_t = (1 - W(B))X_t - W(B)N_t. \tag{3.22}$$

The two components of ζ_t are mutually independent if and only if the components of Y_t are mutually independent.

Proof. The two components of ζ_t in the representation (3.22) are linear function of X_t and N_t. Thus the independence of X_t and N_t is equivalent to that of components of ζ_t. □

Now we an define the dispersion of signal extraction error and minimum dispersion criteria for optimum filter weights w_j.

3.4 Signal extraction filters using Minimum dispersion criteria

Signal extraction procedure consists of finding an optimal filter which minimizes the signal extraction error. In finite variance case optimal filter is the one which minimizes the MSE where as in symmetric stable process we propose minimum dispersion criteria. When signal extraction error process has a symmetric stable distribution, the minimization of error dispersion is equivalent to minimization of the scale parameter of the error distribution (see Brockwell and Davis, 1987, page 486). From Proposition 2.3.6, we have seen that stable $ARMA$ process can be written as an infinite order moving average process. So the moving average representation for the signal and noise respectively written as

$$X_t = \sum_{j=0}^{\infty} \psi_j^x a_{t-j}, \;\; N_t = \sum_{j=0}^{\infty} \psi_j^n b_{t-j}, \tag{3.23}$$

where ψ_j^x and ψ_j^n are the weights obtained for X_t and N_t respectively. Thus by (3.22), the error process may be derived as

$$\begin{aligned}\zeta_t &= \sum_{j=0}^{\infty} w_j Y_{t-j} - X_t \\ &= \sum_{j=0}^{\infty} w_j (X_{t-j} + N_{t-j}) - X_t \\ &= \sum_{j=0}^{\infty} w_j (\sum_{k=0}^{\infty} \psi_k^x a_{t-j-k} + \sum_{k=0}^{\infty} \psi_k^n b_{t-j-k}) - \sum_{k=0}^{\infty} \psi_k^x a_{t-k}.\end{aligned}$$

That is,

$$\zeta_t = \sum_{j=0}^{\infty} (\sum_{k=0}^{j} w_k \psi_{j-k}^x - \psi_j^x) a_{t-j} + \sum_{j=0}^{\infty} \sum_{k=0}^{j} w_k \psi_{j-k}^n b_{t-j}. \tag{3.24}$$

Using (3.24) and the distributional properties of a_t and b_t stated in Theorem 2.5.10 we can show that the dispersion of the error process is

$$Disp(\zeta_t) = \sum_{j=0}^{\infty} |\sum_{k=0}^{j} w_k \psi_{j-k}^x - \psi_j^x|^\alpha \lambda_a + \sum_{j=0}^{\infty} |\sum_{k=0}^{j} w_k \psi_{j-k}^n|^\alpha \lambda_b. \tag{3.25}$$

Finding optimal filter is equivalent to finding the weights w_k in $W(B)$ which minimizes (3.25). In general the solution does not have a closed form, but it gives some satisfactory results for some special cases. For this, however, we need the following results.

Lemma 3.4.1. *If $a, b > 0$ and $\alpha > 0$, the function $g(x) = a|x|^\alpha + b|x-c|^\alpha$ has its minimum value at x_0, where*

$$x_0 = \begin{cases} c, & if,\ \alpha \le 1, \frac{a}{b} \le 1, \\ 0, & if,\ \alpha \le 1, \frac{a}{b} > 1, \\ \frac{c}{1+(\frac{a}{b})^{1/\alpha-1}}, & if,\ \alpha > 1, \end{cases}$$

and x_0 is unique if $\frac{a}{b} \neq 1$ or $\alpha > 1$. The minimum value of g is

$$g(x_0) = \begin{cases} |c|^\alpha \min(1, \frac{a}{b}), & if, \alpha \le 1, \\ a|c|^\alpha (1 + (\frac{a}{b})^{1/\alpha-1})^{1-\alpha} if, \alpha > 1. \end{cases}$$

Proof. Define, $h(x) = \frac{g(x)}{b}$, since $b > 0$, the optimum solution will be unaltered by minimizing $h(x)$. This function is similar to the objective function defined by Cline and Brockwell (1985) in their Lemma 3.1 and hence the proof follows with similar steps. Now we have,

$$h(x) = a'|x|^{\alpha} + |x - c|^{\alpha},$$

$$h'(x) = \alpha[a'[x]^{\alpha-1} + [x - c]^{\alpha-1}],\ h''(x) = \alpha(\alpha - 1)[a'|x|^{\alpha-2} + |x - c|^{\alpha-2}],$$

where, $a' = \frac{a}{b}$. We also define $[x]^{\alpha} = |x|^{\alpha} sgn(x)$. Suppose $c > 0$ and $x \neq 0, x \neq c$, then we can show that for $x < 0, h'(x) < 0$ and for $x > c, h'(x) > 0$. Also the derivative do not exist at $x = 0$ and $x = c$. Thus h is minimized on $[0, c]$.

If $\alpha \leq 1$, $h''(x) \leq 0$, so the minimum must be either 0 or at c. We can show that $h(c) \leq h(0)$ if and only if $a' \leq 1$.

If $\alpha > 1$, then h' is continuous on $[0, c]$ and h'' is non-negative. Thus $h'(x_0) = 0$ gives the point of minimum. We can easily show that the point of minimum, $x_0 = \frac{c}{1+(a')^{1/\alpha-1}}$. and the proof is similar if $c < 0$. □

Theorem 3.4.2. *For $1 < \alpha \leq 2$, the optimal filter weights,w_j which minimizes* (3.25)*, is the solution of the system of equations,*

$$\frac{\partial Disp(\zeta_t)}{\partial w_k} = 0, \quad k = 0, 1, 2, ..., . \tag{3.26}$$

When $\alpha \leq 1$, general expressions do not exist.

Proof. From the proof of Lemma 3.4.1 we can see that, if $1 < \alpha \leq 2$, then the function $\frac{\partial Disp(\zeta_t)}{\partial w_k}$ is continuous on $[0, \psi_j^x]$ and $\frac{\partial^2 Disp(\zeta_t)}{\partial^2 w_k}$ is non-negative. Thus, $\frac{\partial Disp(\zeta_t)}{\partial w_k} = 0$ gives the point of minimum. □

From the above discussion it is clear that, we have to adopt some numerical methods for getting optimum filter weights and the signal estimate. For a moving average process the dispersion function defined in (3.25) reduces to a finite sum. Optimum filter weights

defined in equation (3.26) limited to the solution of some finite set of equations. We address this problem in our simulation studies.

Now we can consider the case, when $\alpha = 2$, that is equivalent to the Gaussian case. If we assume that a_t and b_t are zero mean processes, then from (3.25) we can show that the dispersion of ζ_t reduces to variance of ζ_t. In this case signal extraction filter defined in (3.26) reduces to minimum mean squared error filter. When $\alpha = 2$, the proposed filter reduces to the semi-infinite Wiener Kolmogorov filter defined in (3.10). So Minimum dispersion criteria can be considered as a generalization of MMSE.

Asymmetric Filter and Doubly Infinite Filter

The semi-infinite filter discussed so far in this section may be generalized to doubly infinite and asymmetric filters studied in the literature. The former uses future as well as the past of $\{Y_t\}$ for estimating X_t but the latter filter estimates X_t based on given data up through $u = t - m$, for finite m. In order to apply this method we can modify our filter as:

$$W(B) = \sum\nolimits_{-m}^{\infty} w_j B^j.$$

Similar to (3.25) dispersion of signal extraction error process may be written as,

$$Disp(\zeta_t) = \sum_{j=-m}^{\infty} \mid \sum_{k=-m}^{j} w_k \psi_{j-k}^{x} - \psi_j^{x} |^{\alpha} \lambda_a + \sum_{j=-m}^{\infty} \mid \sum_{k=-m}^{j} w_k \psi_{j-k}^{n} |^{\alpha} \lambda_b. \tag{3.27}$$

From Theorem 3.4.2, the optimum filter, which minimizes $Disp(\zeta_t)$ can be obtained as,

$$\frac{\partial Disp(\zeta_t)}{\partial w_k} = 0, \quad k = -m, -m+1, ..., . \tag{3.28}$$

Doubly infinite filter is a symmetric filter, which can be obtained by letting $m \to \infty$ and it reduces to semi-infinite filter when, $m = 0$. When $\alpha = 2$ the error dispersion in (3.27) reduces to the mean square error and the optimal filter reduces to the asymmetric Wiener

Kolmogorov filter (see Bell and Martin (2004)),

$$W(B) = F^m \frac{\sigma_b^2 \phi_x(B)\phi_n(B)}{\sigma^2 \theta(B)} [\frac{|\theta_x(B)|^2 \phi_n(F)}{\theta(F)\phi_x(B)} B^m]_+, \tag{3.29}$$

where, $F = B^{-1}$ and the notation $[\ ]_+$ indicates only terms with non-negative powers of B are retained and those with positive powers of $F = B^{-1}$ are dropped.

So the proposed signal extraction filter using minimum dispersion criteria generalize the classical Wiener Kolmogorov class of filters for linear stationary time series models with infinite variance innovations. Infinite length filters derived out of these criteria are not feasible in many practical problems. So we would like to derive a finite length filter using Kalman Leavy filtering. We address this problem in the next section.

3.5 Signal extraction using Kalman-Levy filter

So far we have discussed the infinite length filter, but in practice we have only finite length of observations. In this section we introduce a finite length filtering algorithm based on state space representation and Kalman-Levy filtering. This can be considered as an improvement over the infinite length minimum dispersion filter defined in the previous section. The classical approach of Kalman filtering assumes that the underlying models are linear and the innovations are Gaussian. Kalman-Levy filter is a generalized version of Kalman filter for heavy tailed processes (Stuck and Kleiner(1974), Sornette and Ide (2001)). In the present section we discuss the finite length signal extraction filter for such processes with symmetric stable noise. A linear dynamic system of state variable $\mathbf{X}_k$ can be represented as

$$\mathbf{X}_{k+1} = \mathbf{Z}_k \mathbf{X}_k + \eta_{k+1},\ k = 0, 1, 2, \ldots , \tag{3.30}$$

here the state equation determines the rule for the generation of the $p \times 1$ state vector $\mathbf{X}_{k+1}$ from the past $p \times 1$ state $\mathbf{X}_k$, for time points $k = 0, ..., n$. The matrix $\mathbf{Z}_k$ is known as state transition matrix assumed to be known. We further assume the η_t are $p \times 1$ independent

and identically distributed, zero-mean symmetric stable sequence with scale factor matrix B_η. The observations $\mathbf{Y}_k$ follow the equation:

$$\mathbf{Y}_k = \mathbf{T}_k\mathbf{X}_k + \varepsilon_k, \tag{3.31}$$

where T_k is a $q \times p$ measurement or observation matrix and assumed to be known. The additive observation noise ε_k is assumed serially independent symmetric stable random sequence with $q \times p$ scale factor matrix B_ε. In addition, we assume, η_k and ε_k are mutually independent.

For $1 < \alpha \leq 2$, the predictor of state variable is defined as $\mathbf{X}_{k|k-1} = E(\mathbf{X}_k|\mathbf{Y}_{k-1})$ and filter is $\mathbf{X}_{k|k} = E(\mathbf{X}_k|\mathbf{Y}_k)$. The Kalman-Levy filtering algorithm by Sornette and Ide (2001) provides a sequential procedure for estimating the unobserved state variable $\mathbf{X}_k$ and the solution is obtained by sequential prediction and filtering. The prediction equation is given as

$$\mathbf{X}_{k|k-1} = \mathbf{Z}_{k-1}\mathbf{X}_{k-1|k-1}. \tag{3.32}$$

The error associated with prediction can be written as an auto-regressive process,

$$\mathbf{X}_{k|k-1} - \mathbf{X}_k = \mathbf{Z}_{k-1}(\mathbf{X}_{k-1|k-1} - \mathbf{X}_{k-1}) - \eta_k; \tag{3.33}$$

this is a linear combination of multivariate alpha stable random variables and from equation (2.41) we can write,

$$\begin{aligned} \mathbf{X}_{k-1|k-1} - \mathbf{X}_{k-1} &= G_{k-1|k-1}\zeta_{k-1|k-1}, \\ \eta_k &= G_k^\eta \zeta_k^\eta, \end{aligned} \tag{3.34}$$

with associated scale factor matrices $C_{k-1|k-1}$ and C_k^η. The tail covariance of the prediction error can thus be written as,

$$\begin{aligned} B_{k|k-1} = (\mathbf{Z}_{k-1}G_{k-1|k-1})^{[\alpha/2]}\, C_{k-1|k-1}\, (\mathbf{Z}_{k-1}G_{k-1|k-1})^{T[\alpha/2]} \; + \\ (G_k^\eta)^{[\alpha/2]}\, C_k^\eta\, (G_k^\eta)^{T[\alpha/2]}. \end{aligned} \tag{3.35}$$

The Kalman update (filter) equation is given as

$$\mathbf{X}_{k|k} = \mathbf{X}_{k|k-1} + K_k(\mathbf{Y}_k - \mathbf{T}_k\mathbf{X}_{k|k-1}), \tag{3.36}$$

where K_k is the Kalman Levy gain. The error associated with filtering can be written as,

$$\mathbf{X}_{k|k} - \mathbf{X}_k = (I - K_k\mathbf{T}_k)(\mathbf{X}_{k|k-1} - \mathbf{X}_k) + K_k\varepsilon_k. \tag{3.37}$$

In the case of filtering, we can write,

$$\begin{aligned} &\mathbf{X}_{k|k-1} - \mathbf{X}_k = G_{k|k-1}\zeta_{k|k-1}, \\ &\varepsilon_k = G_k^{\varepsilon}\zeta_k^{\varepsilon}, \end{aligned} \tag{3.38}$$

with associated scale factor matrices $C_{k|k-1}$ and C_k^{ε}. The tail covariance of the filtering error can thus be written as,

$$\begin{aligned} B_{k|k} = &(G_{k|k-1} - K_k\mathbf{T}_kG_{k|k-1})^{[\alpha/2]}\ C_{k|k-1}\ (G_{k|k-1} - K_k\mathbf{T}_kG_{k|k-1})^{T[\alpha/2]}\ + \\ &(K_kG_k^{\varepsilon})^{[\alpha/2]}\ C_k{}^{\varepsilon}\ (K_kG_k^{\varepsilon})^{T[\alpha/2]}. \end{aligned} \tag{3.39}$$

Diagonal elements of the tail covariance matrix consist of the scale factor of the error process. Minimizing the trace of the tail covariance matrix is equivalent to minimizing the scale factor of the filtering error process. So it corresponds to the minimum dispersion criteria. The Kalman-Levy gain K_k is obtained by minimizing the trace of the tail covariance matrix $B_{k|k}$. That is, K_k is the solution of a set of N non-linear simultaneous equations,

$$\begin{aligned} &\textstyle\sum_{q=1}^{L}(\sum_{m=1}^{L} K_{im}G_{mq}^{\varepsilon})^{[\alpha-1]}G_{jq}^{\varepsilon}C_q^{\varepsilon} \\ &-\textstyle\sum_{p=1}^{N}(G_{ip} - \sum_{m=1}^{L} K_{im}T_{mp})^{[\alpha-1]}T_{jp}C_p\ = 0, \end{aligned} \tag{3.40}$$

for each i, j where, $G = G_{k|k-1}$, $C = C_{k|k-1}$ and $T = \mathbf{T}_kG_{k|k-1}$. Here N and L are the dimensions of the state vector and measurement vector respectively and the subscripts refer to matrix elements.

For the sake of illustration we consider a univariate case. For example, if we consider $\mathbf{Z}_k = \phi$ and $\mathbf{T}_k = 1$ the above model reduces to a stable-$AR(1)$ signal plus symmetric stable noise model described in McCulloch (1998). The prediction equations become,

$$\begin{aligned} \mathbf{x}_{k|k-1} &= Z_{k-1}\mathbf{x}_{k-1|k-1}, \\ B_{k|k-1} &= |\mathbf{Z}_{k-1}|^{\alpha} B_{k-1|k-1} + \lambda^{\eta}. \end{aligned} \tag{3.41}$$

In this set of equations $\mathbf{x}_{k|k-1}$ determines the forecast of $\mathbf{x}_k$ from a given initial condition $\mathbf{x}_{0|0}$, $B_{k+1|k}$ determines the prediction error dispersion with an initial error dispersion $B_{0|0}$ and λ^{η} is the scale factor of η.

This forecast is then used to find new filter $\mathbf{x}_{k|k}$ and filtering error dispersion $B_{k|k}$ which will be mixed with the observed information $\mathbf{y}_k$ and given by

$$\begin{aligned} \mathbf{x}_{k|k} &= \mathbf{x}_{k|k-1} + K_k(\mathbf{y}_k - \mathbf{T}_k\mathbf{x}_{k|k-1}), \\ B_{k|k} &= |1 - K_k\mathbf{T}_k|^{\alpha} B_{k|k-1} + |K_k|^{\alpha}\lambda^{\epsilon}, \end{aligned} \tag{3.42}$$

where, λ^{ϵ} is the scale factor of ϵ. The term, K_k is called Kalman-Levy gain and can be obtained by minimizing the scale factor of the filtering error process and is given by

$$K_k = T_k/(1 + (\Delta_k)^{\frac{\alpha}{\alpha-1}}),$$

with modified relative error ratio,

$$\Delta_k = \frac{(\lambda_k^{\varepsilon})^{1/\alpha}}{\mathbf{T}_k(\lambda_{k|k-1})^{1/\alpha}}.$$

Now we can study the finite length filtering algorithms based on (3.32) and (3.36). From the models (3.30) and (3.31) the finite length filter may be defined as

$$\widehat{\mathbf{X}}_k = w_0 + \sum\nolimits_{j=1}^{k} w_j \mathbf{Y}_j, \tag{3.43}$$

where $w_j,\ j = 0, 1, 2...k$ are the filter weights whose expressions are to be obtained using

Kalman-Levy filter. In this case the signal extraction problem can be divided into that of prediction and filtering. Under the prediction problem we estimate the future state of the signal from a given initial value and the observed signal. The Kalman-Levy predictor and filter defined in (3.32) and (3.36) for the models (3.30) and (3.31) may be represented in a finite length filter form. The results are stated in the following propositions.

Proposition 3.5.1. *The Kalman-Levy predictor for the models* (3.30) *and* (3.31) *is given by* (3.32) *and may be expressed as,*

$$\boldsymbol{X}_{k+1|k} = \sum_{j=1}^{k} L_j \widetilde{K}_j \, \boldsymbol{Y}_j + L_0 \boldsymbol{X}_{1|0}, \tag{3.44}$$

where, $\widetilde{K}_j = Z_j K_j$, $L_k = I$, $L_j = N_k N_{k-1} ... N_{j+1}$, $j = 0, 1, 2, ..., k-1$ *with* $N_j = Z_j - \widetilde{K}_j T_j$. *Comparing* (3.43) *and* (3.44) *we get,* $w_j = L_j \widetilde{K}_j$ *for* $j = 1, 2..., k$ *and* $w_0 = L_0 \boldsymbol{X}_{1|0}$.

Proof. From (3.32) and (3.36), we can write,

$$\mathbf{X}_{k+1|k} = \widetilde{K}_k \mathbf{Y}_k + N_k \mathbf{X}_{k|k-1},$$

where, $N_k = \mathbf{Z}_k - \widetilde{K}_k \mathbf{T}_k$. By simple recursion it follows that,

$$\mathbf{X}_{k|k-1} = \widetilde{K}_{k-1} \mathbf{Y}_{k-1} + N_k N_{k-1} \widetilde{K}_{k-2} \mathbf{Y}_{k-2} + N_k N_{k-1} N_{k-2} \widetilde{K}_{k-3} \mathbf{Y}_{k-3} + ... +$$

$$N_k N_{k-1} N_{k-2} ... N_2 \widetilde{K}_1 y_1 + N_k N_{k-1} N_{k-2} ... N_1 x_{1|0}.$$

If we choose, $L_k = I$, $L_j = N_k N_{k-1} ... N_{j+1}$, we get the required result. □

The filtering problem deals with the estimation of the present state of the signal from a given initial condition and the observed signal at that time.

Proposition 3.5.2. *The Kalman-Levy filter (update) for the model* (3.30) *and* (3.31) *is*

given by (3.36) *may be written as*

$$\boldsymbol{X}_{k|k} = \sum_{j=1}^{k} L_j \widetilde{K}_j \, \boldsymbol{Y}_j + L_0 \boldsymbol{X}_{1|0}, \tag{3.45}$$

where, $\widetilde{K}_j = Z_j K_j$, *with* $\boldsymbol{Z}_k = I$, $L_k = I$, $L_{k-1} = I - K_k \boldsymbol{T}_k$ *and* $L_j = L_{k-1} N_{k-1} ... N_{j+1}$, $j = 0, 1, 2, ..., k-2$, $N_j = Z_j - \widetilde{K}_j T_j$, *then comparing* (3.43) *and* (3.46) *we get* $w_j = L_j \widetilde{K}_j$ *for* $j = 1, 2..., k$ *and* $w_0 = L_0 \boldsymbol{X}_{1|0}$.

Proof. From (3.36), we can write,

$$\mathbf{X}_{k|k} = K_k \mathbf{Y}_k + (I - K_k \mathbf{T}_k) \mathbf{X}_{k|k-1}.$$

From the proof of Proposition 3.5.1, we have,

$$\mathbf{X}_{k|k-1} = \widetilde{K}_{k-1} \mathbf{Y}_{k-1} + N_k N_{k-1} \widetilde{K}_{k-2} \mathbf{Y}_{k-2} + N_k N_{k-1} N_{k-2} \widetilde{K}_{k-3} \mathbf{Y}_{k-3} + ...+$$

$$N_k N_{k-1} N_{k-2} ... N_2 \widetilde{K}_1 \mathbf{Y}_1 + N_k N_{k-1} N_{k-2} ... N_1 \mathbf{X}_{1|0}.$$

If we choose, $\mathbf{Z}_k = I$, $L_k = I$, $L_{k-1} = I - K_k \mathbf{T}_k$, $L_j = L_{k-1} N_{k-1} ... N_{j+1}$, $j = 0, 1, 2, ..., k-2$, $N_j = Z_j - \widetilde{K}_j T_j$, and $w_0 = L_0 \mathbf{X}_{1|0}$, we get the required result. □

3.6 Simulation

Before going to the details of the simulation study and analysis, let us start with the simulation of stable random variable. For this we need the following results by Samorodnitsky and Taqqu (1994), page 42.

Proposition 3.6.1. *Let* γ *be uniform on* $(-\pi/2, \pi/2)$ *and let* W *be exponential with mean 1. Assume* γ *and* W *independent. Then*

$$X = \frac{sin(\alpha \, \gamma)}{cos(\gamma)^{(1/\alpha)}} \left[\frac{cos((1-\alpha)\gamma)}{W} \right]^{(1-\alpha)/\alpha} \tag{3.46}$$

is distributed like $S_\alpha(1)$.

If X is generated by (3.46) then $\lambda_a\, X$, provides $S_\alpha(\lambda_a)$ random variable.

Now we can move on to the simulation study. Suppose that an observed time series Y_t evolves according to equation (3.1) and the unobserved signal X_t, is a moving average process order q defined by

$$X_t = a_t + \theta_1 a_{t-1} + ... + \theta_q a_{t-q}. \tag{3.47}$$

Also assume that $a_t \sim S_\alpha(\lambda_a)$ and $N_t \sim S_\alpha(\lambda_\epsilon)$ where, λ_a and λ_ϵ are respectively the dispersion parameter of a_t and N_t. Our main objective is to extract the signal X_t from the given observed signal Y_t . This problem can be solved by applying the methods discussed in Section 3.4. We simulate the data from the above model by taking $q = 4$, and $\theta_1 = 0.7$, $\theta_2 = 0.4$, $\theta_3 = 0.2$, $\theta_4 = 0.1$,$\alpha = 1.5$, $\lambda_a = 5$, $\lambda_e = 3$. The symmetric stable innovation sequences $\{a_t\}$ and $\{N_t\}$ are generated using the method described in Proposition 3.6.1. The system of non-linear equations solved using the Levenberg-Marquardt algorithm discussed by Kelley (1999).

Performance of minimum dispersion filters are compared with the MMSE Filter in terms of their error sum of squares and dispersion which are summarized in Table 3.1 for different sample sizes. The results show the improvement of signal extraction of stable models under minimum dispersion criteria. Error sum of square and error dispersion of these filters show the improvement of minimum dispersion filter against minimum mean squared error filter. Figures. 3.1, 3.2 and 3.3 give the plots correspond to actual signal (simulated using model (3.47)) , observed signal under symmetric stable noise (simulated using model (3.1)), estimated signal obtained through minimum dispersion filter and MMSE filter. Figure 3.4 shows the plot of the signal extraction error derived out of these filters. Figures 3.5 and 3.6 give the plot for the finite length filtering signals and error plot respectively. Here we assume that the signal X_t follows an AR (1) model with $\rho = 0.7$, innovation sequences $a_t \sim S_\alpha(\lambda_a)$ and noise $N_t \sim S_\alpha(\lambda_\epsilon)$. The filtering is done using Kalmn-Levy filter and

Kalman filter.

Sample Size	Filter	Dispersion	Error Sum of Squares
50	MMSE	116.0728	223.0761
	Minimum Dispersion	113.0230	212.4247
100	MMSE	225.3723	282.0869
	Minimum Dispersion	218.7023	258.6477

Table 3.1: Comparison of error sum of squares and the dispersion

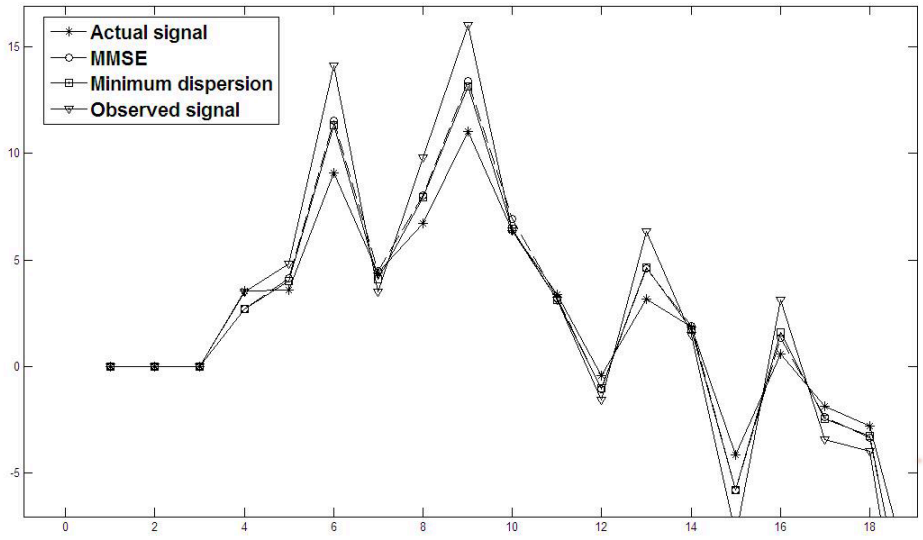

Figure 3.1: Time series Plot of actual signal, observed signal, estimated signal using minimum dispersion filter and MMSE filter

Simulation studies result that minimum dispersion filter introduced in this chapter performs well in signal extraction problems, when signal and noise processes assumed to follows some stable time series models. Time series plots indicate that the proposed procedures are competent enough to handle signal extraction of heavy tailed processes. In the present chapter we assumed the parameters of signal and noise models are known. Next chapter we address the estimation of signal and noise parameters based on an observed signal.

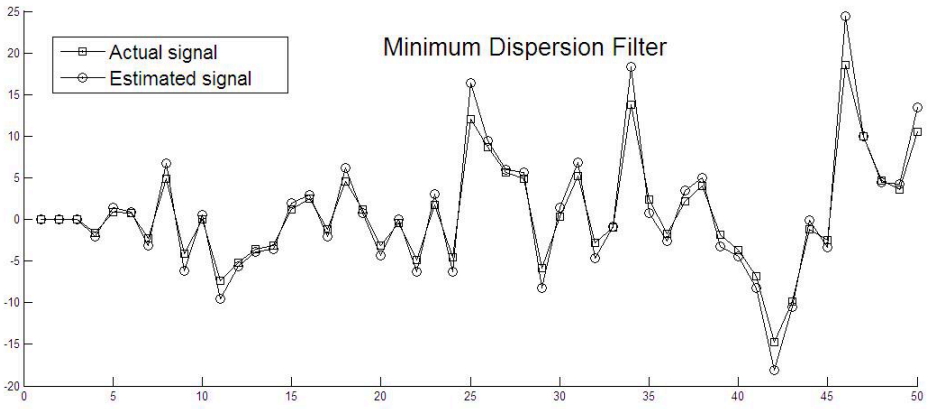

Figure 3.2: Time series Plot of actual signal and estimated signal using minimum dispersion criteria

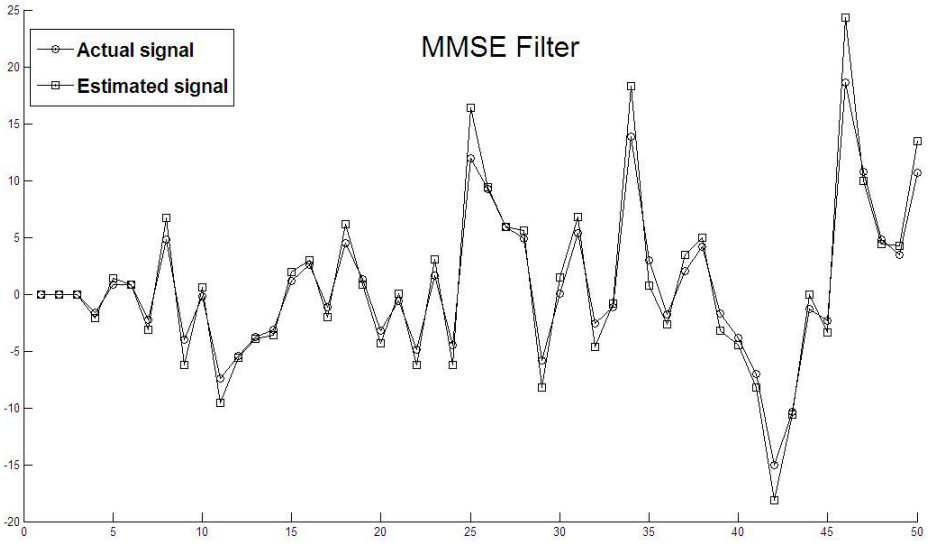

Figure 3.3: Time series Plot of actual signal and estimated signal using MMSE criteria

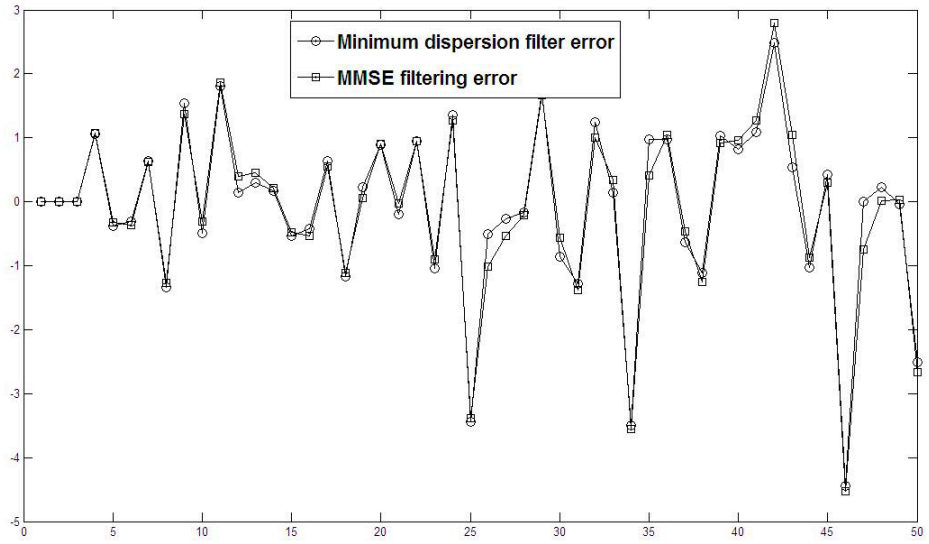

Figure 3.4: Error Plots corresponds to minimum dispersion and MMSE criteria

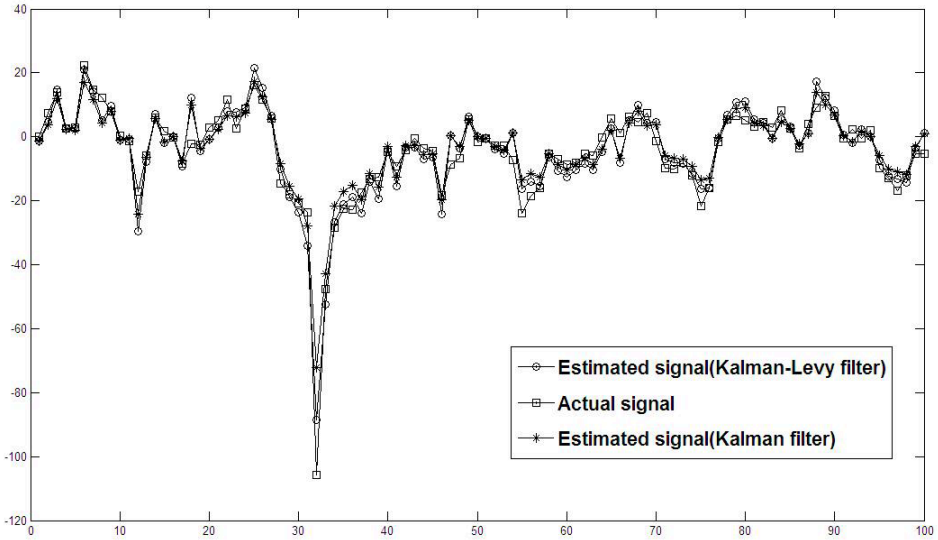

Figure 3.5: Actual signal and finite length filter using Kalman-Levy filter and Kalman filter

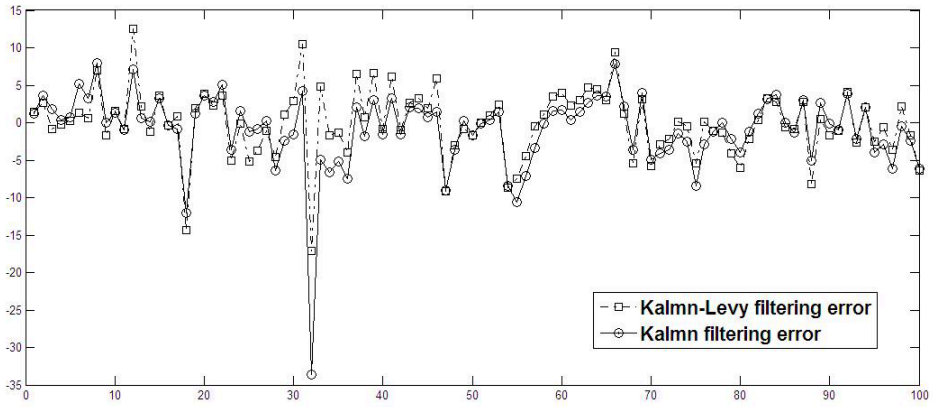

Figure 3.6: Error Plot corresponds to Kalman-Levy filter and Kalman filter

Chapter 4

Stable Autoregressive Models and Signal Estimation

4.1 Introduction

Time series models are popular tools in statistical signal processing, where the signal and noise are assumed to be generated by certain autoregressive moving average models. In the present chapter we study the parameter estimation of the signals generated by autoregressive models with symmetric stable innovations. Adler et al. (1998) studied time series analysis of stable processes using sample ACF and PACF. Samorodnitsky and Taqqu (1994) discussed covariation function to study the dependency structure of sequence of stable random variates. Gallagher (2001) shows that confidence intervals coming from the sample covariation have better coverage probabilities than those coming from the sample correlations. If the process is symmetric stable, the sample covariation has a stable distribution, while the sample correlation appears to converge very slowly to its asymptotic distribution (See Gallagher et al. (2003)). Gallagher (2001) introduced a method for fitting stable autoregressive models using the auto-covariation function, where he uses it in the place of ACF for classical time series. Estimation of autoregressive parameters from a signal plus Gaussian noise has been extensively studied by Gingras (1982) and Chan and Langford

(1980) using a system of extended Yule-Walker (EYW) equations. Nikias and Shao (1995) introduced generalized Yule-Walker (GYW) equations based on auto-covariation function. These two studies motivated us to estimate the parameters of a stable autoregressive signal observed in a symmetric stable noise environment using a modified version of the EYW equations in terms of sample auto-covariation function. In order to minimize the bias in the signal estimation, we use a large number of generalized Yule-Walker equations. Rate of convergence of the estimates are better than the Yule-Walker estimates based on sample auto-correlation function when the process is in the domain of normal attraction. An initial estimate of the noise parameter and innovation parameter is estimated using method of moments. These initial estimates have been used to extract signal and noise from the observed signal by applying Kalman-Levy filter discussed in chapter 3. We can again estimate the signal and noise parameters from the extracted signal and noise sequences. One of the limitations of the covariation based estimation is that the covariation matrix is not necessarily non-singular. The present study highlights this problem and proposes a generalized solution to this problem using Moore-Penrose pseudo inverse.

The chapter is organized as follows: In second section we address the definition and properties of stable autoregressive signals. Section 4.3 deals with the estimation of autoregressive parameters and innovation parameters from a signal plus noise model. Section 4.4 consists of the modification of generalized Yule-Walker estimate using singular value decomposition. Section 4.5 contains simulation studies. In the Last section the methods are used to analyze the sea surface temperature data.

4.2 Stable Autoregressive Signals

In this section we present stable autoregressive signal plus symmetric stable noise model and its properties. Let $\{Y_n\}$ be the noisy measurement of a stable $AR(p)$ signal $\{X_n\}$

defined in (2.63) observed at time n, then the models may be written as,

$$\begin{aligned} X_n &= \phi_1 X_{n-1} + \phi_2 X_{n-2} + ... + \phi_p X_{n-p} + \eta_n, \\ Y_n &= X_n + \varepsilon_n. \end{aligned} \tag{4.1}$$

The measurement error $\varepsilon_n \sim S_\alpha(\gamma_\varepsilon)$, are assumed to be serially independent and also independent of $\{X_n\}$ at all time points. So we also have ε_n and $\{\eta_n\}$ are mutually independent at all time points. We also assume that both $\{\varepsilon_n\}$ and $\{\eta_n\}$ are zero mean processes. A detailed discussion on the applications of such models in the context of signal estimation may be found in McCulloch (1998).

From equation (4.1) we can see that the generalized Yule-Walker estimates derived from (2.54) based on Y_n will be biased due to the presence of measurement noise. A range of approaches have been introduced in finite variance case to solve this problem (see Davila (1998) and the references there of). Here we focus on to generalize the extended Yule-Walker estimation in terms of auto-covariation function. This leads to the estimation of the stable autoregressive signal parameters based on the EYW equations using sample auto-covariation function (AcovF).

From (4.1) we can write,

$$Y_n + \phi_1 Y_{n-1} + \phi_2 Y_{n-2} + ... + \phi_p Y_{n-p} = \eta_n - \varepsilon_n - \phi_1 \varepsilon_{n-1} - ... - \phi_p \varepsilon_{n-p}. \tag{4.2}$$

Multiplying through out the equation (4.2) by $sign(Y_{n-p-k})$ to obtain,

$$\begin{aligned} &Y_n sign(Y_{n-p-k}) + \phi_1 Y_{n-1} sign(Y_{n-p-k}) + ... + \phi_p Y_{n-p} sign(Y_{n-p-k}) = \\ &\eta_n sign(Y_{n-p-k}) - \varepsilon_n sign(Y_{n-p-k}) - \phi_1 \varepsilon_{n-1} sign(Y_{n-p-k}) - ... - \phi_p \varepsilon_{n-p} sign(Y_{n-p-k}). \end{aligned} \tag{4.3}$$

On taking expectation in equation (4.3) and divided by $E|Y_n|$ we obtain the equation,

$$\lambda_y(p+k) = \sum_{j=1}^{p} \phi_j \, \lambda_y(p+k-j), \quad k = 1, ..., p, \tag{4.4}$$

where, $\lambda_y(.)$ is the auto-covariation function of Y_n defined by (2.51). Note that, the expected value of right hand side of equation (4.3) vanishes when $k > 0$, since y_{n-k-1} can only involve the innovations η_j and ε_j up to time $n-k-1$, which are independent of η_n and ε_n. We know that the autoregressive parameter vector $\phi = (\phi_1, \phi_2, ..., \phi_p)$ is a function of $\lambda(k)$ and this relation can be used to find $\hat{\phi}$ based on $\hat{\lambda}(k)$, for $k = 0, 1, 2, ..., N-1$. To minimize the bias of EYW estimates, it is suggested that a large number of EYW equations be included for estimation. So we have to choose an optimum number p' of equations, such that p' is probably large compared to p but small compared to sample size. When the noise dispersion is small, an AR model of order $p' = 2p$ might be appropriate but as the noise dispersion increases, a higher p is required for satisfactory approximation. A method for choosing such best p' has been discussed by Politis (2009). One can choose $p' = \max(p, [cN])$, where $c \in (0, 1)$. Reasonable choice of c should lie in the interval [0.1 0.2] for practical sample size of order 100 to 1000. The Extended Yule-Walker equations can be modified as,

$$\lambda_y(k) = \sum_{j=1}^{p} \phi_j \, \lambda_y(p + k - j), \quad k = 1, ..., p', \tag{4.5}$$

for some $p \leq p' < N$. In practice $\lambda_y(k)$ and p are unknown. Estimation of p is a problem of order identification. We address this problem in Chapter 5. In the next section we discuss the estimation of $\lambda_y(k)$ and subsequently the estimation of the AR parameters ϕ_i , $i = 1, ..., p$.

4.3 Signal Estimation

This section discusses the parameter estimation of autoregressive signals observed in a stable noise environment through the models described in equation (4.1). The observed signal Y_n can be represented as an $ARMA$ model and the parameters can be estimated using generalized Yule-Walker equations, by Nikias and Shao (1995). We look at the problem in a different angle as mentioned in the previous section. The system of equations (4.5) can

be expressed as

$$\Delta_{p',p}\,\phi = T_{p'}, \tag{4.6}$$

where,

$$\Delta_{p',p} = \begin{pmatrix} \lambda_y(p) & \cdots & \lambda_y(1) \\ \vdots & \ddots & \vdots \\ \lambda_y(p+p'-1) & \cdots & \lambda_y(p') \end{pmatrix},\ \phi = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}$$

and,

$$T_{p'} = \begin{pmatrix} \lambda_y(p+1) \\ \vdots \\ \lambda_y(p+p') \end{pmatrix}.$$

But in practice $\Delta_{p',p}$, $T_{p'}$ and p are unknown. Order p of the autoregressive term, can be chosen according to the order of observed signal Y_n, which will be discussed in the next chapter. Estimate of $\lambda_y(.)$ is obtained by its sample version, $\hat{\lambda}_y(.)$ defined as in (2.53)(see Gallagher 2000). From, Equation (4.6), we may write

$$\hat{T}_{p'} = \hat{\Delta}_{p',p}\,\phi + \mathrm{Z}. \tag{4.7}$$

The terms, $\hat{\Delta}_{p',p}$ and $\hat{T}_{p'}$ are obtained by replacing the $\lambda_y(.)$ values by its estimate $\hat{\lambda}_y(.)$ in equation (4.6) where as Z is the corresponding error. From Theorem 2.4.5, we have, $\hat{\lambda}(k) = \lambda(k) + O_P(N^{\frac{1}{\alpha}-1})$ for any $k \leq p'$. This in turn gives, $Z = O_P(N^{\frac{1}{\alpha}-1})$. Hannan and Kanter (1977) used least square estimation to estimate the parameters of stable autoregressive models. Cline (1989)studied the consistency for least squares regression estimators with infinite variance data. These studies motivate us to propose an estimate $\hat{\phi}$ for the vector of autoregressive parameters using ordinary least square regression method in equation (4.7), and is given by

$$\hat{\phi} = (\hat{\Delta}'_{p',p}\hat{\Delta}_{p',p})^{-1}\hat{\Delta}'_{p',p}\hat{T}_{p'}, \tag{4.8}$$

if the matrix $\hat{\Delta}'_{p',p}\hat{\Delta}_{p',p}$ is non-singular. If the matrix is singular we can handle the problem in terms of Moore-Penrose pseudo inverse. The details are discussed in the next section.

In the present study we assume the covariation matrix is non-singular unless it specified.

Note: When, $p' = p$ the estimate $\hat{\phi}$ reduces to that obtained from the Extended Yule-Walker Equations.

Theorem 4.3.1. *Define a function*

$$g(w) = (\hat{\Delta}'_{p',p}(w)\Delta_{p',p}(w))^{-1}\hat{\Delta}'_{p',p}(w)\, w,$$

and assume that $g(w)$ is continuously differentiable. Then, as $N \to \infty$,

$$\begin{aligned} &\text{(i)} \ \ \widehat{\phi} \xrightarrow{p} \phi, and, \\ &\text{(ii)} \ \ \hat{\phi} = \phi + O_p(N^{\frac{1}{\alpha}-1}). \end{aligned}$$

Proof. (i) From the Theorem 2.4.4, we may write $\hat{\lambda}(k) = \lambda(k) + o_p(1)$, and from the definition of g we have $\phi = g(\lambda)$ and $\hat{\phi} = g(\hat{\lambda})$. Now the proof of $\widehat{\phi} \xrightarrow{p} \phi$ follows by the continuous mapping theorem stated in Proposition 2.6.5.

(ii) Applying Theorem 2.4.5 we can show that, $\hat{\lambda}(k) = \lambda(k) + O_P(N^{\frac{1}{\alpha}-1})$ for any $k \leq p'$. It is enough to show that, $N^{1-\frac{1}{\alpha}}(\hat{\phi} - \phi) = O_p(1)$

That is,

$$P\{N^{1-\frac{1}{\alpha}}|\hat{\phi} - \phi| > \delta(\epsilon)\} < \epsilon.$$

Take, $N^{1-\frac{1}{\alpha}} = a_n$ and consider

$$\begin{aligned} &P\{a_n|\hat{\phi} - \phi| > \delta(\epsilon)\} \\ &= P\{a_n|g(\hat{\lambda}) - g(\lambda)| > \delta(\epsilon)\}, \\ &\leq P\{a_n|g(\hat{\lambda}) - g(\lambda)| > \delta(\epsilon),\ a_n|\lambda| \leq K,\ a_n|\hat{\lambda}| \leq K\} \ + \\ &P\{(a_n|\hat{\lambda}| > K) \cup (a_n|\lambda| > K)\}. \end{aligned}$$

Since the function g is uniformly continuous on $\{\lambda : a_n|\lambda| \leq K\}$, there exist $\gamma(\epsilon) > 0$, such

that for all N,

$$\{a_n|g(\hat{\lambda}) - g(\lambda)| > \delta(\epsilon),\ a_n|\hat{\lambda}| \leq K,\ a_n|\lambda| \leq K\} \subseteq \{a_n|\hat{\lambda} - \lambda| > \gamma(\epsilon)\}.$$

Hence,

$$\begin{aligned}
&P\{a_n|\hat{\phi} - \phi| > \delta(\epsilon)\} \\
&\leq P\{a_n|\hat{\lambda} - \lambda| > \gamma(\epsilon)\} + P\{(a_n|\lambda| > K)\} + \\
&P\{(a_n|\hat{\lambda}| > K/2)\} + P\{(a_n|\hat{\lambda} - \lambda| > K/2)\}
\end{aligned}$$

Now given any $\delta > 0$, we can choose K to make each of the second and third terms less than $\delta/4$. Then since $a_n(\hat{\lambda} - \lambda) = O_p(1)$, each of the first and fourth terms is also bounded by $\delta/4$ for large N. Consequently, $g(\hat{\lambda}) = g(\lambda) + O_p(N^{\frac{1}{\alpha}-1})$. □

Theorem 4.3.2. *If g is defined as in Theorem 4.3.1 and F is the $p \times (p + p' - 1)$ matrix of partial derivatives of the vector function g, with respect to λ at $\hat{\lambda}$, then as $N \to \infty$,*

$$N^{1-\frac{1}{\alpha}}(\widehat{\phi} - \phi) \xrightarrow{L} FS,$$

where $S = [S_1\, S_2 \ldots S_{p+p'-1}]'$, defined as in 2.4.5.

Proof. From the Theorem 2.4.5 we can show that, $\hat{\lambda}(k) = \lambda(k) + O_P(N^{\frac{1}{\alpha}-1})$. Take, $N^{\frac{1}{\alpha}-1} = a_n$.

Now applying Proposition 2.6.6, we can write,

$$g(\widehat{\lambda}) = g(\lambda) + \sum_{i=1}^{p+p'-1} \frac{\partial g(\lambda)}{\partial \lambda_i}(\widehat{\lambda}_i - \lambda_i) + o_p(a_n).$$

or equivalently,

$$g(\widehat{\lambda}) - g(\lambda) = F(\widehat{\lambda} - \lambda) + o_p(a_n).$$

Multiplying both side by ${a_n}^{-1}$ we obtain,

$$ {a_n}^{-1}(g(\widehat{\lambda}) - g(\lambda)) = {a_n}^{-1}F(\widehat{\lambda} - \lambda) + o_p(1). $$

From Theorem 2.4.5 and 2.4.7 we have, ${a_n}^{-1}F(\widehat{\lambda} - \lambda) \Rightarrow FS$. Hence from Proposition 2.6.4 it follows that of ${a_n}^{-1}(g(\widehat{\lambda}) - g(\lambda)) \Rightarrow FS$. □

The following procedure suggests a method of estimating scale parameters and heavy tail index of the innovation and noise sequence in model (4.1). Estimation procedure entails two steps; we propose an initial estimate of the scale parameter obtained by the method of moment and heavy tail index using Hill estimation method in the first step.

For symmetric stable random variable with scale γ,

$$ E|X| = \frac{\pi\,\gamma}{2\,\Gamma(1 - 1/\alpha)}. \tag{4.9} $$

For the sake of simplicity we can illustrate this procedure by assuming that X_n follows $AR(1)$ model in (4.1). From (4.1) we can show that,

$$ Y_n - \phi_1\,Y_{n-1} = U_n + \varepsilon_n - \phi_1\,\varepsilon_{n-1} = \eta_n. \tag{4.10} $$

From the assumption of independence of U_n and ε_n, we can show that, scale parameter γ_η of η_n, as,

$$ \gamma_\eta = \gamma_u + (1 + |\phi_1|^\alpha)\gamma_\varepsilon, $$

That is,

$$ \frac{E|\eta_t|2\,\Gamma(1 - 1/\alpha)}{\pi} = \gamma_u + (1 + |\phi_1|^\alpha)\gamma_\varepsilon. \tag{4.11} $$

From proposition 2.4.3 we can show that the auto-covariation between η_n and η_{n-1} as,

$$ [\eta_n, \eta_{n-1}]_\alpha = -\phi_1\,E|\varepsilon_n|, \; = -\phi_1\,\frac{\pi\,\gamma_\varepsilon}{2\,\Gamma(1 - 1/\alpha)}. \tag{4.12} $$

Define $\gamma(0) = E|\eta_n|$ and $\gamma(1) = [\eta_n, \eta_{n-1}]_\alpha = E(\eta_n sign(\eta_{n-1}))$, then we can write (4.11) and (4.12) respectively as,

$$\frac{\gamma(0)2\,\Gamma(1-1/\alpha)}{\pi} = \gamma_u + (1+|\phi_1|^\alpha)\gamma_\varepsilon, \tag{4.13}$$

and

$$\gamma(1) = -\phi_1\,\frac{\pi\,\gamma_\varepsilon}{2\,\Gamma(1-1/\alpha)}. \tag{4.14}$$

Let α be known and $\hat{\phi}_1$ be the estimate of ϕ_1 obtained from (4.8). If we define $\hat{\gamma}(0) = \frac{1}{N}\sum_{n=1}^{N}|\eta_n|$ and $\hat{\gamma}(1) = \frac{1}{N}\sum_{n=1}^{N}\eta_n sign(\eta_{n-1})$, then the moment estimates of γ_u and γ_ε are obtained by solving the equations (4.13) and (4.14).

If α is unknown we can still estimate it by Hill estimate proposed by Pictel et al. (1998) and replace α by its estimate in equations (4.13) and (4.14).

In the second step we apply Kalman-Levy filter discussed in the previous chapter to extract signal (X_n) and noise (ε_n) sequences from the observed signal (Y_n). This step entails the knowledge of the parameters γ_u, γ_ε and α values. These values can be replaced by the estimated values obtained from the first step. An estimate of the innovation sequence $\{U_n\}$ and noise sequence $\{\varepsilon_n\}$ can be obtained from the extracted signal sequence $\hat{X}_n$ as,

$$\hat{U}_n = \hat{X}_n - \hat{\phi}_1\,\hat{X}_{n-1}, \tag{4.15}$$

and

$$\hat{\varepsilon}_n = Y_n - \hat{X}_n.$$

Scale parameter of the extracted sequence $\hat{U}_n$ and $\hat{\varepsilon}_n$ is estimated using method of moments discussed above.

Heavy tailed index can be estimated by the following estimate given by Gallagher (2001),

that is,

$$\hat{\alpha} = \frac{\log(1 - \hat{\phi}_1 \hat{\lambda}_{-1})}{\log(\sum_{n=1}^{N} \hat{U}_n / \sum_{n=1}^{N} |\hat{X}_n|)}.$$

Note: For an $AR(p)$ model expressions similar to (4.11) and (4.12) can be obtained as,

$$\frac{\gamma(0) 2\,\Gamma(1 - 1/\alpha)}{\pi} = \gamma_u + (1 + |\phi_1|^\alpha + ... + |\phi_p|^\alpha)\gamma_\varepsilon, \tag{4.16}$$

and

$$\gamma(k) = -\phi_k \, \frac{\pi\,\gamma_\varepsilon}{2\,\Gamma(1 - 1/\alpha)}, k = 1, ..., p. \tag{4.17}$$

The moment equations defined in (4.16)and (4.17) are nonlinear in its parameters. So the estimation may be computationally expensive and inaccurate. For better performance we recommend this estimate as an initial estimate for the Kalman-Levy filtering in the second step.

4.4 Modified generalized Yule-Walker estimation

The covariation matrix discussed earlier is not nonsingular in general. This makes the estimation problem difficult, and introduces error in the estimation. Further, it poses problems for studying the asymptotic properties of the estimate $\hat{\lambda}$. A method for sorting out this problem is discussed below. Before going to details let us define Moore-Penrose pseudo inverse (see Stewart (1973), Rao (1973))

Definition 4.4.1. Moore-Penrose pseudo inverse defined a generalized inverse as a matrix A^+ satisfying the properties

(i) $AA^+A = A$,

(ii) $A^+AA^+ = A^+$,

(iii) $(AA^+)^* = AA^+$,

(iv) $(A^+A)^* = A^+A$.
such an inverse exist and is unique.

Let $\widehat{\Delta}_m$ be an estimate of the covariation matrix obtained by replacing the $\lambda_y(.)$ by its sample version, $\widehat{\lambda}_y(.)$. We can apply the singular value decomposition method of Stewart (1973) to the auto-covariation matrix, and is given by ,

$$\widehat{\Delta}_m = U\Lambda V^T. \tag{4.18}$$

In (4.18) U and V are orthogonal matrices of size $m \times m$, and

$$\Lambda = diag(\nu_1, \nu_2, ..., \nu_m),\ \nu_1 \geq \nu_2 \geq, ... \geq, \nu_m, \tag{4.19}$$

where $\nu_1,\ \nu_2, ..., \nu_m$ are singular values of $\widehat{\Delta}_m$. Singular value decomposition helps to identify and eliminate the singular values ν_i, which are close to zero. After eliminating singular values the resulting matrix will have rank $p \leq m$ and given by

$$\overline{\widehat{\Delta}}_p = UDV^T \tag{4.20}$$

where,

$$D = \begin{bmatrix} \widehat{\Lambda} & 0 \\ 0 & 0 \end{bmatrix}$$

with, $\widehat{\Lambda} = diag(\nu_1, \nu_2, ..., \nu_p)$. The estimate $\overline{\widehat{\Delta}}_p$ is likely to be, in general, a better estimate than $\widehat{\Delta}_m$ of Δ_m . Now the generalized Yule-Walker equation can be modified in terms of $\overline{\widehat{\Delta}}_p$ and $\widehat{T}_m$, the equality holds only approximately.

$$\overline{\widehat{\Delta}}_p\ \phi_* \approx \widehat{T}_m \tag{4.21}$$

The generalized Yule-Walker estimate can be obtained as,

$$\widehat{\phi}_* = (\overline{\widehat{\Delta}}_p)^+ \, \widehat{T}_m \tag{4.22}$$

where,

$$(\overline{\widehat{\Delta}}_p)^+ = V \begin{bmatrix} \widehat{\Lambda}^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T$$

denotes the Moore-Penrose pseudo inverse of $\overline{\widehat{\Delta}}_p$.

Theorem 4.4.2. *Define a function $g(w) = (\Delta_m(w))^+ \, w$, and assume that $g(w)$ is continuously differentiable. Then from (4.22), as $N \longrightarrow \infty$, $\widehat{\phi}_* \xrightarrow{p} \phi_*$ and, $N^{1-\frac{1}{\alpha}}(\widehat{\phi}_* - \phi_*) \xrightarrow{L} FS$, where $S = [S_1, ..., S_{2p-1}]$ and F is the $m \times 2p-1$ matrix of partial derivatives of vector function g , with respect to λ at $\widehat{\lambda}$.*

Proof. From Theorem 2.4.5, we may write ,

$$\widehat{\lambda(k)} = \lambda(k) + o_p(1)$$

and we also have $\phi_* = g(\lambda)$ and $\widehat{\phi}_* = g(\widehat{\lambda})$. Proof of $\widehat{\phi}_* \xrightarrow{p} \phi_*$ is directly for the continuous mapping theorem stated in Proposition 2.6.5. By using the mean value theorem we can write

$$\widehat{\phi}_* - \phi_* = F(\widehat{\lambda} - \lambda) + o_n(\widehat{\lambda} - \lambda)$$

where, F is the matrix specified in the statement of the theorem. Applying the steps similar to those of the proof of Theorem 4.3.1 we get the required result. □

Note:

1. When $p = m$, we get

$$\overline{\widehat{\Delta}}_p^{+} = \widehat{\Delta}_m^{+} = \widehat{\Delta}_m^{-1}$$

and the corresponding estimator reduces to the generalized Yule-Walker estimator.

2. The estimate defined in equation (4.22) is an extended version of generalized Yule-Walker estimate. Since the Moore-Penrose pseudo inverse is unique, the estimate also has the same property.

4.5 Simulation

To study the signal parameter estimation problem we generate 10000 realizations of $AR(1)$ signal X_t with sample sizes, 100, 1000, 5000 and parameters $\phi = 0.6$, $\gamma_\eta = 5$, and $\alpha = 1.7$ using the first model in (4.1). The observed signal Y_n is obtained using the second model (4.1) with measurement noise parameters $\gamma_\varepsilon = 2$. That is

$$\begin{aligned} X_n &= \phi X_{n-1} + \eta_n, \\ Y_n &= X_n + \varepsilon_n. \end{aligned} \tag{4.23}$$

where, $\varepsilon_n \sim S_\alpha(\gamma_\varepsilon)$ and $\eta_n \sim S_\alpha(\gamma_\eta)$. The symmetric stable sequences are generated using the algorithm discussed in Section 3.6.

We carried out the estimation procedure in four steps.

(i) First step we estimate the autoregressive parameters based on the estimate proposed in equation (4.8) with $p' = 4$.

(ii) An initial estimate of γ_u and γ_ε are obtained using equation (4.13) and (4.14) in the second step.

(iii) We consider an initial value of α in this step. Based on these values the signal and noise components are extracted using the Kalman-Levy filter in the third step.

(iv) Signal and noise parameters are modified using the extracted components in the last step.

Step (iii) and (iv) are continued until the estimates get converged. These procedures repeated for all 10000 realizations and the Mean and the Mean Absolute Deviation (MAD) of the estimated parameters are obtained out of this simulation.The results are given in Table 4.1.

From the Table 4.1, we can see that the mean absolute deviation of the estimate reduces when the sample size increases. Simulation results show that the proposed estimation method is a satisfactory tool for signal estimation. In the next section we validate our method based on a real data set.

Sample Size	Parameters	True Values	Estimated Values	MAD
100	α	1.7	1.9900	1.8815
	γ_η	5	4.9797	1.1275
	γ_ε	2.0	2.0224	0.3941
	ϕ	0.6	0.6800	0.1661
1000	α	1.7	1.9800	0.6385
	γ_η	5	5.1040	0.6197
	γ_ε	2	1.9773	0.2274
	ϕ	0.6	0.7215	0.1344
5000	α	1.7	1.7800	0.3860
	γ_η	5	5.1030	0.3980
	γ_ε	2	1.9686	0.1474
	ϕ	0.6	0.7248	0.1292

Table 4.1: Mean and Mean Absolute Deviations (MAD) for the signal and noise parameters estimated based on the methods discussed in section 4.

4.6 Analysis of Global Sea Surface Temperature Time Series Data

We analyze a data set consists of global average Sea Surface Temperature (SST) anomaly in the month of November from 1845 to 2006. The data was downloaded from *http* :

//jisao.washington.edu/data/global_sstanomts. The anomaly is the difference between observed data and its global average. The data set is multiplied by 100 ($C * 100$). Figure 4.1 shows the time series plot of the anomaly data.

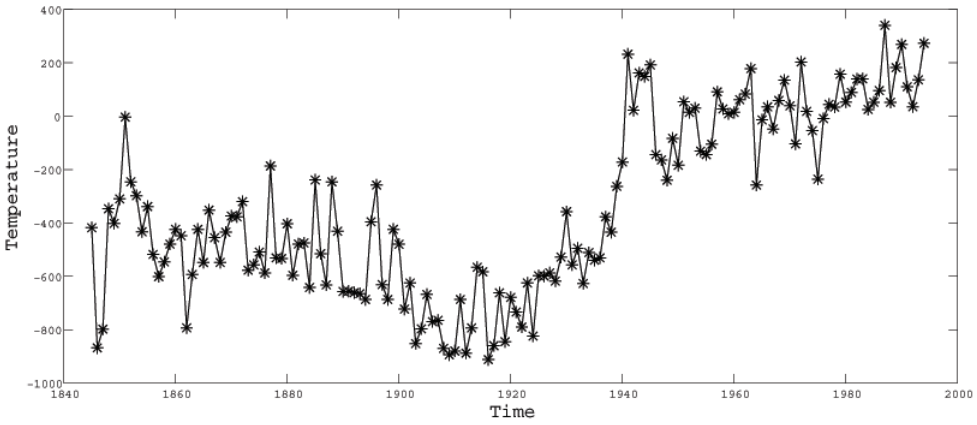

Figure 4.1: Time series plot of Sea Surface Temperature Data

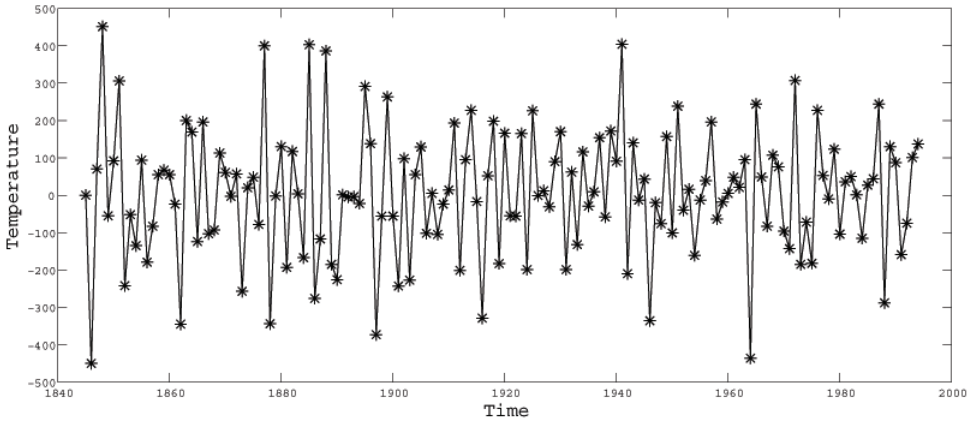

Figure 4.2: Time series plot of first order difference Data

Time series plot indicates that the data shows some increasing trend so it cannot be assumed to be a realization of stationary stochastic process. The differenced data plot in Figure 4.2 shows stationarity but the time series show some sharp spikes, which may be an indication of heavy tails. We plot the auto-covariation and partial auto-covariation of the differenced data and observe that the auto-covariation function tails off and the partial

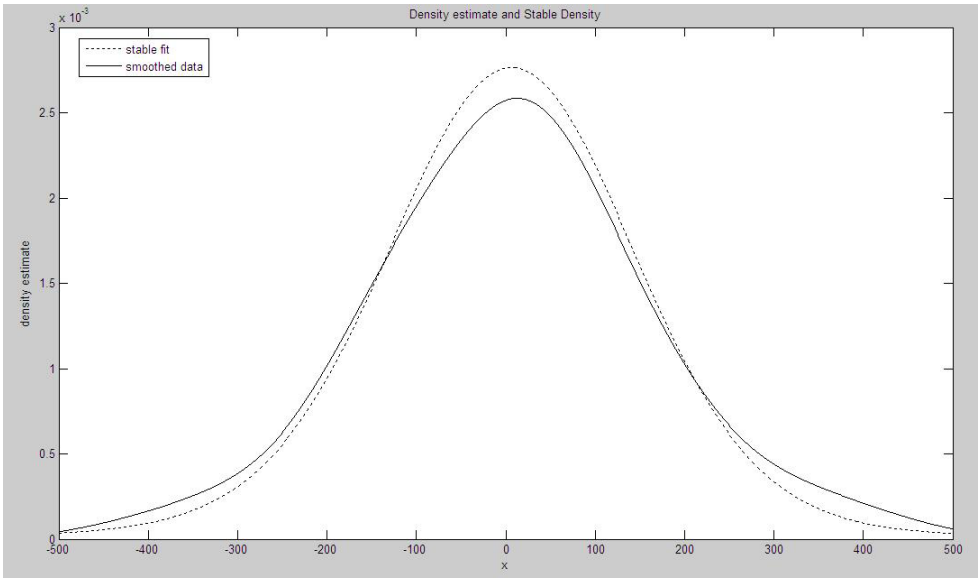

Figure 4.3: Stable density and estimated density of the estimated innovations from stable model

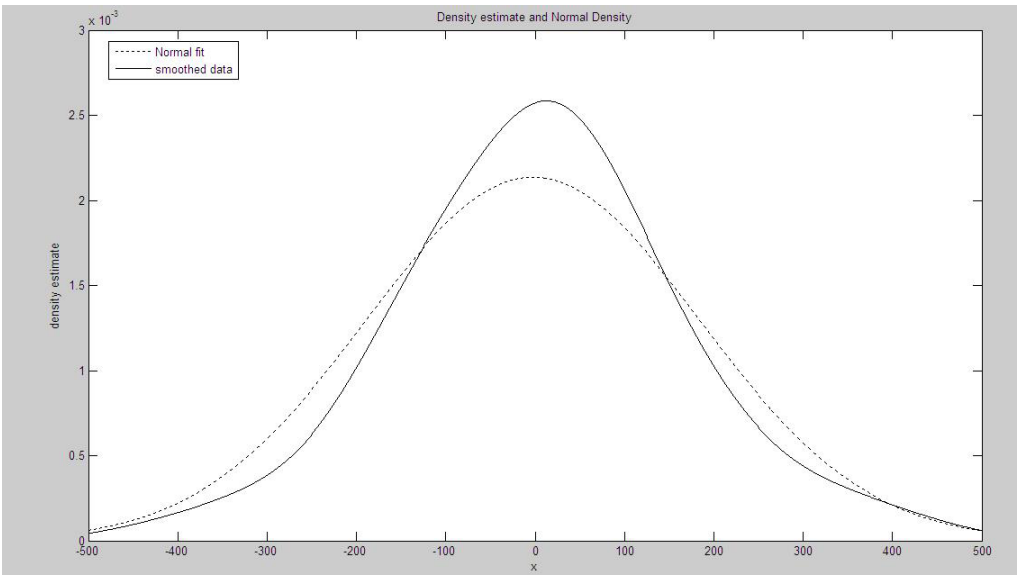

Figure 4.4: Normal density and estimated density of the estimated innovations from normal model

auto-covariation function cuts off after three lags. Higher order observed AR signals can be written as a lower order AR signal plus noise model. Based on this, we fit a stable AR (1) signal plus noise model (4.23) to this data. The autoregressive parameters are estimated

by the proposed method using sample AcovF and ACF. Estimated AR parameter, $\widehat{\phi} = -0.4750$. We estimate the innovation sequence and noise sequence based on the method discussed in Section 4.3. For an iid sequence, the AcovF and PcovF are zero for all $k \neq 0$. Since $\hat{\lambda}(k) \to \lambda(k)$ as $N \to \infty$, the sample versions of AcovF and PcovF should be close to zero. It is observed that the AcovF and PcovF of the estimated sequences are close to zero.

Parameters	MLE	SCF	QE	MME
α	1.8	1.78	1.69	1.85
γ_η	102	100	98	88.4
γ_ε	3.6	3.5	3.18	2.99

Table 4.2: Estimated parameters of the noise and innovation sequence.

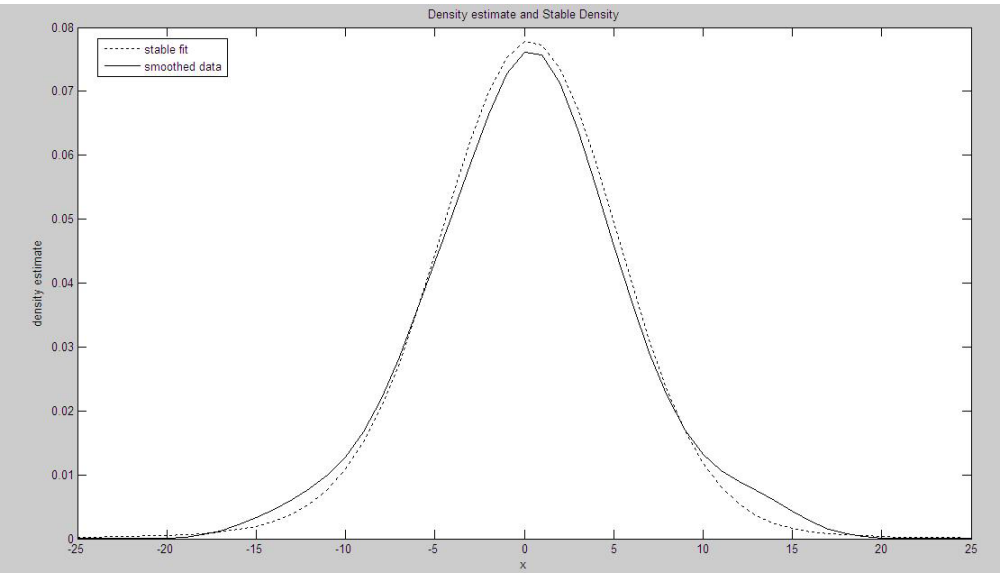

Figure 4.5: Stable density and estimated density of the estimated noise from stable model

Parameters of the innovation sequence are estimated using Gaussian and stable assumptions. Here we use four well known methods to estimate the stable parameters which are Maximum Likelihood Estimation (MLE), Sample Characteristic Function (SCF), Quantile Estimation (QE) and our proposed Method of Moment Estimation (MME)). Estimation under the first three methods have been carried out using the program stable.exe of Nolan

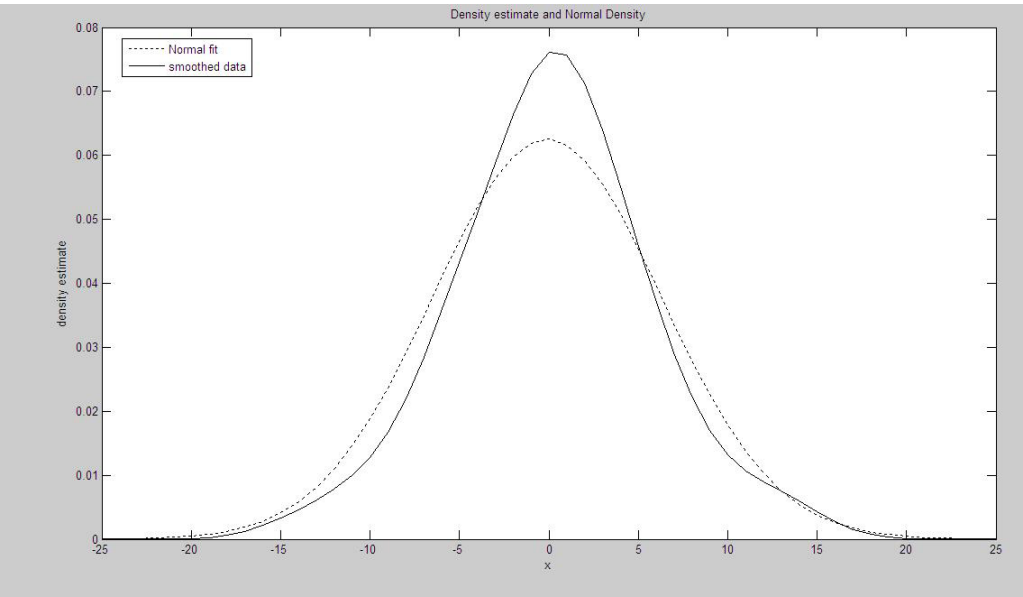

Figure 4.6: Stable density and estimated density of the estimated noise from normal model

(2005); the results are listed in Table 4.2. For Gaussian model, $\sigma_\eta = 132$, and $\sigma_\varepsilon = 3.5$ are the estimated standard deviation of the innovation sequence and noise sequence respectively. Figures 4.3, 4.4, 4.5 and 4.6 show the plots of stable and normal density functions with the estimated parameter, respectively along with a smoothed density estimate from the estimated innovation and noise sequences. Figure 4.3 and 4.5 are respectively the density plot of estimated innovation and noise sequences under stable assumption. Figure 4.4 and 4.6 are respectively the density plot of estimated innovation and noise sequences under normal assumption. The stable density fit and density plot are obtained using the program stable.exe of Nolan (2005). The density estimate under the assumption of stable models provides a better fit. The computation and simulation studies have been carried out using MATLAB software.

Chapter 5

Model Identification techniques for Stable Autoregressive Models

5.1 Introduction

The twin problems of model identification and estimation of Gaussian time series models are extensively studied by Box and Jenkins (1976), Brockwell and Davis (1987). Model identification problems can be generally classified in to graphical approach and information approach. Graphical model identification approach of Gaussian time series is generally carried out using autocorrelation and partial autocorrelation plot. In the present chapter we study the model identification techniques for stable time series models. Theoretical properties of autocorrelation and partial autocorrelation cannot be studied in this set up due to the lack of second order moments. Theorem 2.4.1 of Chapter 2 gives the limiting behavior of sample auto-correlation function in stable set up. Keeping these results in mind Adler et al. (1998) explored the application of sample auto-correlation and partial auto-correlation plot for model identification of stable auto-regressive models. Gallagher (2001) used sample auto-covariation function to identify moving average models based on its limiting behavior. In the present work we introduce the concept of partial auto-covariation function (PcovF) for time series with heavy tailed marginals, a measure similar

to PACF in the Gaussian time series. To introduce the concept of PcovF, we use the ideas from the linear prediction theory discussed in Chapter 2. In this chapter our focus is to study the time series analysis of infinite variance autoregressive processes using auto-covariation and partial auto-covariation functions. We generalize the Durbin-Levinson algorithm in stable autoregressive models in terms of partial auto-covariation and use it for model identification. Simulation studies show that sample PcovF performs better than sample ACF for model identification in autoregressive process in the domain of attraction of normal laws. The problem of over fitting can be avoided by imposing a cost for increasing the number of parameters in the fitted models. One way of doing this for pure auto-regressive model is to minimize the final prediction error (FPE) of Akaike (1970). Akaike information criteria (AIC) is one of the popular methods in model identification (see Akaike (1974), Brockwell and Davis (1987), Shibata (1976)). This method minimizes the prediction error variance and it can be written as a function of partial auto-correlation function. This approach is widely used for model identification of infinite variance time series even though the variance is ill defined in this case (see Bhansali (1988), Knight (1989), Adler et al. (1998), Burridge and Hristova (2007)). In this chapter we propose a new information criteria based on the minimization of prediction error dispersion, which can be defined mathematically for infinite variance processes. This criterion can be expressed as a function of partial auto-covariation function. Theoretical studies show that the proposed method is consistent in model identification under certain conditions. Simulation results substantiate the consistency properties in model identification techniques and it performs better than AIC in both Gaussian and stable auto-regressive models.

The chapter is organized as follows: In Section 5.2 we discuss the definition and properties of Partial auto-covariation function (PcovF). Generalization of Durbin-Levinson algorithm for fitting stable auto-regressive model is mentioned in Section 5.3. We propose a new model identification criterion and its theoretical performance in Section 5.4 . Section 5. 5 contains simulation studies.

5.2 Partial Auto-covariation Function for AR Models

Cambanis and Miller (1981) studied the linear estimation problems of alpha stable processes. Cline and Brockwell (1985) defined a finite linear prediction for a stable autoregressive process. The results are observed to be identica to the corresponding results for least squares prediction of the process, with finite variance.

In this chapter we consider an autoregressive process of order m and is defined as,

$$X_n = \phi_1 X_{n-1} + \phi_2 X_{n-2} + ... + \phi_m X_{n-m} + \eta_n. \tag{5.1}$$

From Lemma 2.5.11 we can shows that, for symmetric stable processes with $\alpha > 1$, the minimum dispersion predictor of X_{n+k} is given by

$$\hat{X}_{n+k} = E(X_{n+k}|X_{n+k-1}, X_{n+k-2}, ...). \tag{5.2}$$

The Corollary 2.5.14 shows that, for an $AR(m)$ process, with $m \leq n$, there exists a unique minimum dispersion linear predictor $\hat{X}_{n+k}$ for $X_{n+k}(k \geq 1)$ in terms of $X_{n+k-1}, X_{n+k-2}, ..., X_{n+k-m}$ and it follows the recursive relationship

$$\hat{X}_{n+k} = \phi_1 \hat{X}_{n+k-1} + \phi_2 \hat{X}_{n+k-2} + ... + \phi_m \hat{X}_{n+k-m}, \tag{5.3}$$

with initial conditions $\hat{X}_j = X_j$ for $1 \leq j \leq n$. From (5.2) and (5.3), for $1 < \alpha \leq 2$,

$$\hat{X}_{n+k} = E(X_{n+k}|X_{n+k-1}, ..., X_{n+k-m}),$$

which is the minimum dispersion predictor of X_{n+k} based on $(X_{n+k-1}, ..., X_{n+k-m})$. Now let us define **partial auto-covariation function (PcovF)** as follows.

Definition 5.2.1. Let X_n be a zero mean stationary alpha stable process with auto-covariation function $\lambda(k)$ such that $\lambda(k) \to 0$, as $k \to \infty$ and suppose that $\phi_{nj}, j =$

$1, ..., n, n = 1, 2, ...$ are the coefficients in the representation:

$$E(X_{n+1}|X_1, X_2, ..., X_n) = \sum_{j=1}^{n} \phi_{nj} X_{n-j+1}.$$

Then from Theorem 2.5.9, we get

$$\begin{bmatrix} \lambda(0) & \dots & \lambda(1-n) \\ \vdots & \ddots & \vdots \\ \lambda(n-1) & \dots & \lambda(0) \end{bmatrix} \begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn} \end{bmatrix} = \begin{bmatrix} \lambda(1) \\ \vdots \\ \lambda(n) \end{bmatrix} \tag{5.4}$$

and the **partial auto-covariation** $\tau(n)$ of $\{X_n\}$ at lag n is defined as

$$\tau(n) = \phi_{nn}, n \geq 1.$$

If $\lambda(.)$ matrix on the LHS of (5.4) is non-singular then, ϕ_{nn} can be determined by this equation.

Remark 5.2.2. If $\{X_n\}$ is either an $AR(1)$ process or an $AR(2)$ process, then the estimates of AR parameter are unique and strongly consistent. In either case $\lambda(.)$ matrix on the LHS of (5.4) is non-singular. In both cases the estimates of this matrix are almost surely nonsingular if $P(X_n = 0) = 0$. If X_t is an $AR(1)$ process LHS matrix becomes $\lambda(0) = 1$. If X_t is an $AR(2)$ process, $\lambda(.)$ matrix is nonsingular unless $\lambda(1) = \lambda(-1) = \pm 1$.

The following is an example to illustrate the existence of a non-singular $\lambda(.)$ matrix in the case of a stationary $AR(3)$ model. Let η_n be an iid sequence of symmetric stable random variables and define

$$X_n = 0.7126X_{n-1} + 0.1868X_{n-2} - 0.0007X_{n-3} + \eta_n. \tag{5.5}$$

The associated $\lambda(.)$ matrix in (5.4) becomes

$$\Delta = \begin{bmatrix} 1 & 0.6274 & 0.5158 \\ 0.8294 & 1 & 0.6274 \\ 0.7774 & 0.8294 & 1 \end{bmatrix}.$$

Clearly the matrix is non-singular as $|\Delta| = 0.2191$.

If $\lambda(.)$ matrix in (5.4) is singular then the estimation problem can be handled in terms of Moore-Penrose pseudo inverse discussed in Section 4.4.

Now, we discuss the problem of estimating partial auto-covariation function based on sample values and study the asymptotic properties of the estimates. We have seen the limiting behavior of sample autocorrelation function of moving average processes with symmetric stable innovations in Theorem 2.4.1. This result provides the theoretical support to make use of sample correlation as a tool in time series analysis of alpha stable processes, especially model identification (see Adler et al. (1998)). Important aspect of the covariation function is that we can define it in stable process environment, where correlation is ill defined. Theorems 2.4.4, 2.4.5 and 2.4.7 illustrate the limiting behavior of sample auto-covariation function in a moving average process with alpha stable innovations. We use the application of these theorems to study the limiting distribution of partial auto-covariation function.

In order to study the partial auto-covariation function (PcovF) in the autoregressive case, we consider the vectors $\Phi_k' = (\phi_1', \phi_2', ..., \phi_k')$, where $\phi_i' = \phi_i$ for $i \leq m$ and $\phi_i' = 0$ when $i > m$. The partial auto-covariation at lag k , ϕ_{kk} is defined as the $k-$th component of the vector

$$\Phi_k' = \Delta_k^{-1} T_k, \tag{5.6}$$

where, $\Delta_k = [\lambda(i-j)]_{i,j=1}^k$ is an $k \times k$ matrix with $\lambda(0) = 1$ and $T_k = (\lambda(1), ..., \lambda(k))'$.

Sample PcovF at lag k, $\widehat{\phi}_{kk}$ is defined as the $k-$th component of the vector

$$\widehat{\Phi}_k' = \widehat{\Delta}_k^{-1} \widehat{T}_k, \tag{5.7}$$

where, $\widehat{\Delta}_k$ and $\widehat{T}_k$ are respectively Δ_k and T_k, whose elements are replaced by the corresponding sample quantities. Simulation results show that these measures perform well in model identification and estimation of autoregressive models with heavy tailed innovations.

Theorem 5.2.3. *With $\widehat{\Phi}'_k$ defined by (5.7), as $N \to \infty$*

$$\widehat{\Phi}'_k \xrightarrow{p} \Phi'_k$$

and

$$N^{1-\frac{1}{\alpha}}(\widehat{\Phi}'_k - \Phi'_k) \xrightarrow{L} FS, \tag{5.8}$$

where $S = [S_1, S_2, ..., S_{2k-1}]'$ has stable distribution defined in 2.4.5 for $\alpha \in (1, 2]$ and F is the $k \times 2k-1$ matrix of partial derivatives of a continuously differenciable vector function $g(w) = \Delta_k^{-1}(w)\, w$.

Proof. From the equation (5.6), we can identify a function $g(.)$ as

$$\Phi_k = g(\lambda).$$

Then by the mean value theorem

$$\hat{\Phi}_k - \Phi_k = F(\hat{\lambda} - \lambda) + o_n(\hat{\lambda} - \lambda),$$

where, F is the matrix specified in the statement of the theorem. If we define $\Delta_k(w) = [w_{i-j}]_{i,j=1}^k$, $w_0 \equiv 1$, the proof follows similar to the proof of Theorems 4.3.1 and 4.3.2. □

Theorem 5.2.3 gives the limiting distribution of sample partial auto-covariation function. Now we can switch over to solve the model identification problem using PcovF. Gallagher (2001) used auto-covariation function to identify moving average models. We can easily show that PcovF is zero after lag m, in $AR(m)$ models which is similar to that of PACF in the Gaussian setup, (see Brockwell and Davis (1987)). This result can be used to identify the order of autoregressive models. Auto-covariation function of white noise process is zero.

When $m = 0$, right hand side of (5.8) reduces to the marginal distribution of S, this is the same limit as that of the auto-covariation function of white noise process. The model identification problem reduces to that of testing null hypothesis that, which of the $\phi_{n,n}$ are zero (see Gallagher (2006)).

5.3 Durbin-Levinson algorithm for fitting stable autoregressive model

Durbin-Levinson recursion method of estimation is widely used in finite variance time series set up in terms of autocorrelation function (see Brockwell and Davis (1987)). In this section we extend this method for the infinite variance model such as autoregressive model with stable innovation using partial auto-covariation function. From the definition of partial auto-covariation we can write the predictor $\widehat{X}_{n+1}$ of X_{n+1} as a linear function of its past values.

$$\widehat{X}_{n+1} = \phi_{n,1}X_n + \phi_{n,2}X_{n-1} + ... + \phi_{n,n}X_1. \tag{5.9}$$

The absolute mean of prediction error can be denoted by $\gamma_u(n)$. Thus,

$$\gamma_u(n) = E(|X_{n+1} - \widehat{X}_{n+1}|). \tag{5.10}$$

clearly $\gamma_u(0) = \gamma(0) = E|X_n - E(X_n)|$.

If $\gamma(0) > 0$ and $\Delta_1, \Delta_2, ..., \Delta_n$ are nonsingular, one can generalize Durbin-Levinson algorithm, which is a recursive scheme for computing $\phi_{n,1}, \phi_{n,2}, ...$ and $\gamma_u(1), \gamma_u(2), ...$ for $n = 1, 2, ...$. The algorithm can be stated as a proposition and a corollary as follows:

Proposition 5.3.1. *(**Generalized Durbin-Levinson Algorithm**) If $\{X_t\}$ is a zero mean stationary process with $\gamma(0) > 0$ and $\lambda(h) \to 0$ as $h \to \infty$, then $\phi_{n,j}$ defined in* (5.9),

satisfy $\phi_{1,1} = \lambda_1, \phi_{1,-1} = \lambda_{-1}$,

$$\phi_{s,s} = \frac{\lambda(s) - \phi_{s-1,1}\lambda(s-1) - \phi_{s-1,2}\lambda(s-2) - ... - \phi_{s-1,s-1}\lambda(1)}{1 - \phi_{s-1,1}\lambda(-1) - \phi_{s-1,2}\lambda(-2) - ... - \phi_{s-1,s-1}\lambda(1-s)}, \tag{5.11}$$

$$\phi_{s,-s} = \frac{\lambda(-s) - \phi_{s-1,-1}\lambda(1-s) - \phi_{s-1,-2}\lambda(2-s) - ... - \phi_{s-1,1-s}\lambda(-1)}{1 - \phi_{s-1,-1}\lambda(1) - \phi_{s-1,-2}\lambda(2) - ... - \phi_{s-1,1-s}\lambda(s-1)},$$

$$\phi_{s,r} = \phi_{s-1,r} - \phi_{s,s}\phi_{s-1,r-s}, s = 1, 2, ..., N,$$

$$\phi_{s,-r} = \phi_{s-1,-r} - \phi_{s,-s}\phi_{s-1,s-r}, s = 1, 2, ..., N.$$

Proof. From equation (5.4) we can write,

$$\Delta_n \ \phi_n = T_n \tag{5.12}$$

The basic idea of the recursion is to find the solution ϕ_{n+1} for the $(n+1)$st order case from the solution Φ_n for the nth order case. Unfortunately the Δ_n is not symmetric compared to the correlation matrix but diagonal elements are equal.

To begin the proof, consider the case $n = 1$. Equation (5.12) becomes

$$\Delta_1 \ \Phi_1 = T_1,$$

$$\lambda(0) \ \phi_{1,1} = \lambda(1),$$

$$\phi_{1,1} = \lambda(1),$$

since $\lambda(0) = 1$.

The recursion will now be developed by evaluating Equation (5.12) for order $n+1$:

$$\begin{bmatrix} \lambda(0) & \lambda(-1) & \dots & \lambda(-n+1) & \lambda(-n) \\ \lambda(1) & \lambda(0) & \dots & \lambda(-n+2) & \lambda(-n+1) \\ \vdots & \ddots & \vdots & & \\ \lambda(n-1) & \lambda(n-2) & \dots & \lambda(0) & \lambda(-1) \\ \lambda(n) & \lambda(n-1) & \dots & \lambda(1) & \lambda(0) \end{bmatrix} \begin{bmatrix} \phi_{n,1} \\ \phi_{n,2} \\ \vdots \\ \phi_{n,n-1} \\ \phi_{n,n} \end{bmatrix} = \begin{bmatrix} \lambda(1) \\ \lambda(2) \\ \vdots \\ \lambda(n) \\ \lambda(n+1) \end{bmatrix}.$$

The matrices Δ_{n+1} and Δ_n are related by the relation

$$\Delta_{n+1} = \begin{bmatrix} & & & \begin{bmatrix} \lambda(-n) \\ \lambda(-n+1) \\ \vdots \\ \lambda(-1) \end{bmatrix} \\ & \Delta_n & & \\ & & & \\ \begin{bmatrix} \lambda(n) & \lambda(n-1) & \dots & \lambda(1) \end{bmatrix} & & & [\lambda(0)] \end{bmatrix}$$

and the vector T_{n+1} related to T_n by

$$T_{n+1} = \begin{bmatrix} T_n \\ \lambda(n+1) \end{bmatrix}$$

Now let us define some vectors,

$$\phi_n = \begin{bmatrix} \phi_{1,1} \\ \phi_{1,2} \\ \vdots \\ \phi_{1,n} \end{bmatrix}, N_m = \begin{bmatrix} \lambda(-1) \\ \lambda(-2) \\ \vdots \\ \lambda(-n) \end{bmatrix}, \quad \lambda_n^+ = \begin{bmatrix} \lambda(n) \\ \lambda(n-1) \\ \vdots \\ \lambda(1) \end{bmatrix}, \quad \lambda_n^- = \begin{bmatrix} \lambda(-n) \\ \lambda(-n+1) \\ \vdots \\ \lambda(-1) \end{bmatrix}$$

and also we have,

$$T_n = \begin{bmatrix} \lambda(1) \\ \lambda(2) \\ \vdots \\ \lambda(n) \end{bmatrix}$$

So, the matrix Δ_{n+1} may be written as,

$$\Delta_{n+1} = \begin{bmatrix} \Delta_n & \lambda_n^- \\ (\lambda_n^+)^T & \lambda(0) \end{bmatrix}.$$

Lets now represent the $(n+1)$st-order parameter vector ϕ_{n+1} in terms of the nth order vector ϕ_n, a correction term k_{n+1} and a correction vector ε_n as

$$\phi_{n+1} = \begin{bmatrix} \phi_n \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_n \\ k_{n+1} \end{bmatrix}. \tag{5.13}$$

Equation (5.13) for order $n+1$ is therefore represented in terms of the nth order equation as

$$\begin{bmatrix} \Delta_n & \lambda_n^- \\ (\lambda_n^+)^T & \lambda(0) \end{bmatrix} \left\{ \begin{bmatrix} \phi_n \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_n \\ k_{n+1} \end{bmatrix} \right\} = \begin{bmatrix} T_n \\ \lambda(n+1) \end{bmatrix}$$

which, sorting the equations out, implies

$$\Delta_n \phi_n + \Delta_n \varepsilon_n + \lambda_n^- k_{n+1} = T_n, \tag{5.14}$$

and

$$(\lambda_n^+)^T \phi_n + (\lambda_n^+)^T \varepsilon_n + \lambda(0) k_{n+1} = \lambda(n+1). \tag{5.15}$$

Now let's define two more vectors,

$$\beta_n = \begin{bmatrix} \phi_{1,-n} \\ \phi_{1,-n+1} \\ \vdots \\ \phi_{1,-1} \end{bmatrix}, \quad \alpha_n = \begin{bmatrix} \phi_{1,-1} \\ \phi_{1,-2} \\ \vdots \\ \phi_{1,-n} \end{bmatrix}$$

and let

$$\Delta_n \beta_n = \lambda_n^-. \tag{5.16}$$

From (5.12) and (5.14) we can show that,

$$\Delta_n \varepsilon_n + \lambda_n^- k_{n+1} = 0. \tag{5.17}$$

Substituting (5.16) into (5.17) we get,

$$\varepsilon_n = -k_{n+1}\beta_n. \tag{5.18}$$

We will now use this and Equation (5.15) to find k_{n+1} and thus have all the components needed to complete the recursion.

Pre-multiplying both sides of Equation (5.18) by $(\lambda_n^+)^T$ gives the scalar relation

$$(\lambda_n^+)^T \varepsilon_n = -(\lambda_n^+)^T \beta_n k_{n+1} \tag{5.19}$$

$$(\lambda_n^+)^T \beta_n = T_n^T \gamma_n = T_n^T (\Delta_n^T)^{-1} N_n,$$

and

$$\phi_n{}^T N_n = T_n^T \gamma_n = T_n^T (\Delta_n^{-1} T_n)^T N_n = T_n^T (\Delta_n^T)^{-1} N_n.$$

That is,

$$(\lambda_n^+)^T \beta_n = {\phi_n}^T N_n.$$

Substituting this in (5.19) we get

$$(\lambda_n^+)^T \varepsilon_n = -{\phi_n}^T N_n k_{n+1}. \tag{5.20}$$

So equation (5.15) becomes,

$$(\lambda(0) - \phi_n^T N_n) k_{n+1} = \lambda(n+1) - \lambda_n^+ \phi_n$$

From this we get the expression for k_{n+1} as,

$$k_{n+1} = \frac{\lambda(n+1) - \lambda_n^+ \phi_n}{\lambda(0) - \phi_n^T N_n}. \tag{5.21}$$

Equations (5.13), (5.18), and (5.21) therefore define a recursion formula for finding the $n+1$st-order model parameters $\phi_{k,n+1}$ in terms of the previously-obtained nth order solution. We get a starting point to the recursion from $n = 1$ case discussed above. □

Now let us derive an expression for absolute mean for the prediction error as a corollary to this result. For that we need the following result by Gallagher (2001).

Proposition 5.3.2. *For a causal stable $AR(m)$ process, $\{X_n,\}$ the following equation holds*

$$[E|\eta_n|/E|X_n|]^\alpha = 1 - \phi_1\lambda(-1) - ... - \phi_m\lambda(-m). \tag{5.22}$$

Proof. see Gallagher (2001). □

From (5.3) we have seen that for an $AR(m)$ model, the best linear predictor satisfy the recursion relation

$$\hat{X}_{n+1} = \phi_1\hat{X}_n + \phi_2\hat{X}_{n-1} + ... + \phi_m\hat{X}_{n-m+1}.$$

Comparing this with model (5.9) we have

$$\phi_{nj} = \phi_j \text{ for } j = 1, 2, ..., m \text{ and}$$
$$\phi_{nj} = 0 \text{ for } j > m.$$

From the definition of $AR(m)$ model the prediction error u_{n+1}can be seen as,

$$u_{n+1} = X_{n+1} - \hat{X}_{n+1} = \eta_{n+1}$$
$$X_{n+1} = \hat{X}_{n+1} + u_{n+1}.$$

From Proposition 5.3.2 we have

$$[E|u_n|/E|X_n|]^\alpha = 1 - \phi_{n1}\lambda(-1) - ... - \phi_{nm}\lambda(-m). \tag{5.23}$$

Corollary 5.3.3. *If $\gamma(0) > 0$ for a stable autoregressive models for $m = 1, 2, ..., N-1$, and, $\phi_{k,k}, \phi_{k,-k}$ are defined as in Proposition 5.3.1 then,*

$$\gamma_u(1) = \gamma(0)(1 - \phi_{1,1}\phi_{1,-1})^{1/\alpha}$$

and,

$$\gamma_u(m) = \gamma(0)(\prod_{k=1}^{m}(1 - \phi_{k,k}\phi_{k,-k}))^{1/\alpha}. \tag{5.24}$$

Proof. The result can be proved by induction. To begin with consider the case of $n = 1$. From equation (5.22), we have

$$[\gamma_u(1)/\gamma(0)]^\alpha = 1 - \phi_{1,1}\lambda(-1),$$
$$= 1 - \phi_{1,1}\phi_{1,-1}.$$

Assume that the result hold for $n = m$.

That is,

$$[\gamma_u(m)/\gamma(0)]^{\alpha} = \prod_{k=1}^{m}(1 - \phi_{k,k}\phi_{k,-k}).$$

Now we need to prove the result for $n = m + 1$, let us write

$$[\gamma_u(m+1)/\gamma(0)]^{\alpha} = 1 - \phi_{m+1,1}\lambda(-1) - ... - \phi_{m+1,m+1}\lambda(-m-1)$$

Using the equation (5.11) we can write,

$$\begin{aligned}
&1 - \phi_{m+1,1}\lambda(-1) - ... - \phi_{m+1,m+1}\lambda(-m-1) \\
&= 1 - (\phi_{m,1} - \phi_{m+1,m+1}\phi_{m,-m})\lambda(-1) - (\phi_{m,2} - \phi_{m+1,m+1}\phi_{m,-m+1})\lambda(-2) - \\
&... - (\phi_{m,m} - \phi_{m+1,m+1}\phi_{m,-1})\lambda(-m) - \phi_{m+1,m+1}\lambda(-m).
\end{aligned}$$

Since,

$$(\lambda_n^+)^T\beta_n = {\phi_n}^T N_n,$$

we can write,

$$[\gamma_u(m+1)/\gamma(0)]^{\alpha} = 1 - \phi_{m1}\lambda(-1) - ... - \phi_{mm}\lambda(-m)(1 - \phi_{m+1,m+1}\phi_{m+1,-m-1})$$

$$\begin{aligned}
&= [\gamma_u(m)/\gamma(0)](1 - \phi_{m+1,m+1}\phi_{m+1,-m-1}) \\
&= (\prod_{k=1}^{m}(1 - \phi_{k,k}\phi_{k,-k}))(1 - \phi_{m+1,m+1}\phi_{m+1,-m-1}).
\end{aligned}$$

There fore,

$$[\gamma_u(m+1)/\gamma(0)]^{\alpha} = (\prod_{k=1}^{m+1}(1 - \phi_{k,k}\phi_{k,-k}))^{1/\alpha}$$

This completes the proof. □

If $\widehat{\gamma}(0) > 0$ we can fit a stable autoregressive signals of order ($n < N$) by means of

generalized Yule-Walker equations. The fitted $AR(m)$ process is

$$X_n = \widehat{\phi}_1 X_{n-1} + \widehat{\phi}_2 X_{n-2} + ... + \widehat{\phi}_m X_{n-m} + \eta_n,\ \eta_n \sim S_\alpha(\frac{2\Gamma(1-1/\alpha)}{\pi}\widehat{\gamma}_u(m)).$$

Durbin-Levinson recursion can be used to fit stable autoregressive models of successively increasing order to the observed time series data. This will provide an estimate of autoregressive parameters and innovation dispersion. So the recursive equations derived in proposition 5.3.1 can be restated in terms of the estimates $\widehat{\phi}_m,\ \widehat{\gamma}_u$, and the algorithm is described in the following proposition.

Proposition 5.3.4. *(Generalized Durbin-Levinson Algorithm for fitting stable autoregressive model): If $\widehat{\gamma}(0) > 0$ then the fitted stable autoregressive models for $m = 1, 2, ..., N-1$, can be determined recursively from the relations, $\widehat{\phi}_{1,1} = \widehat{\lambda}_1, \widehat{\phi}_{1,-1} = \widehat{\lambda}_{-1}$, $\widehat{\gamma}_u(1) = \widehat{\widehat{\gamma}}(0)(1 - \widehat{\phi}_{1,1}\widehat{\phi}_{1,-1})^{1/\alpha}$*

$$\widehat{\phi}_{m,m} = \frac{\widehat{\lambda}(m) - \widehat{\phi}_{m-1,1}\widehat{\lambda}(m-1) - \widehat{\phi}_{m-1,2}\widehat{\lambda}(m-2) - ... - \widehat{\phi}_{m-1,m-1}\widehat{\lambda}(1)}{1 - \widehat{\phi}_{m-1,1}\widehat{\lambda}(-1) - \widehat{\phi}_{m-1,2}\widehat{\lambda}(-2) - ... - \widehat{\phi}_{m-1,m-1}\widehat{\lambda}(1-m)}, \tag{5.25}$$

$$\widehat{\phi}_{m,-m} = \frac{\widehat{\lambda}(-m) - \widehat{\phi}_{m-1,-1}\widehat{\lambda}(1-m) - \widehat{\phi}_{m-1,-2}\widehat{\lambda}(2-m) - ... - \widehat{\phi}_{m-1,1-m}\widehat{\lambda}(-1)}{1 - \widehat{\phi}_{s-1,-1}\widehat{\lambda}(1) - \widehat{\phi}_{m-1,-2}\widehat{\lambda}(2) - ... - \widehat{\phi}_{m-1,1-m}\widehat{\lambda}(m-1)},$$

$$\widehat{\phi}_{m,r} = \widehat{\phi}_{m-1,r} - \widehat{\phi}_{m,m}\widehat{\phi}_{m-1,r-m},\ r = 1, 2, .., m-1,$$

$$\widehat{\phi}_{m,-r} = \widehat{\phi}_{m-1,-r} - \widehat{\phi}_{m,-m}\widehat{\phi}_{m-1,m-r},\ r = 1, 2, .., m-1.$$

and,

$$\widehat{\gamma}_u(m) = \widehat{\gamma}(0)(\prod_{k=1}^{m}(1 - \widehat{\phi}_{k,k}\widehat{\phi}_{k,-k}))^{1/\alpha}. \tag{5.26}$$

5.4 A new model identification criteria

Prediction error variance decreases when the order increases. When the theoretical order is reached, this variance remains constant. Most of the order selection criteria follow this concept (see Brockwell and Davis (1987)). The most well known criteria proposed by Akaike (1974) is Akaike Information Criteria (AIC) which includes a bias correction term and is defined as

$$AIC(k) = N \ln(\widehat{\sigma}^2(k)) + 2k, \tag{5.27}$$

where, $\widehat{\sigma}^2(k)$ is the estimated prediction error variance of autoregressive models of order k. The autoregressive parameters and corresponding prediction error variances are estimated using Durbin-Levinson algorithm. Best choice of the order is the one which minimizes the $AIC(k)$. Order estimator is not consistent in finite variance case. For stationary autoregressive models with infinite variance, consistency of the autoregressive order selected by minimizing a version of the AIC, expressed as a function of the Yule-Walker (Y W) estimate for the "innovation variance", is established by Knight (1989). Estimation of innovation variance is meaningless in infinite variance model though the performance is good. In the present work we explore the prediction error dispersion to study the model identification problem. The autoregressive parameters and corresponding prediction error dispersion are estimated using the modified version of Durbin-Levinson algorithm (5.11). An Information Criterion, similar to AIC, can be defined as,

$$IC(k) = N^{2/\beta} \ln(\widehat{\gamma}_u(k)) + 2k, \text{ for some, } \beta > \frac{\alpha}{\alpha - 1}. \tag{5.28}$$

The order estimate $\widehat{m}$ is,

$$\widehat{m} = arg \min_{1<k\leq K(N)} IC(k). \tag{5.29}$$

Simulation results show that the proposed information criterion performs well in the identification of stable autoregressive models. Next we study the consistency of the order selection criteria.

Lemma 5.4.1. *For an AR(m) process defined by* (5.1)*, and from Theorem 5.2.3, we have as* $N \to \infty$

$$N^{2/\beta}(\widehat{\phi}_{k,k}\widehat{\phi}_{k,-k}) \xrightarrow{p} 0.$$

for some, $m < k \leq K(N)$.

Proof. It is enough to prove that

$$\hat{\phi}_{k,k} - \phi_{k,k} = o_p(N^{\frac{1}{\beta}}).$$

From Theorem 5.2.3 we have,

$$N^{1-\frac{1}{\alpha}}(\hat{\phi}_{k,k} - \phi_{k,k}) = O_p(1).$$

That is, for a given $\epsilon > 0$, there exists a δ' and N_0 such that

$$P\{|N^{1-\frac{1}{\alpha}}(\hat{\phi}_{k,k} - \phi_{k,k})| > \delta'(\varepsilon)\} < \varepsilon \ \forall N > N_0.$$

That is,

$$\begin{aligned}\varepsilon &> P\{|N^{1-\frac{1}{\alpha}-\frac{1}{\beta}+\frac{1}{\beta}}(\hat{\phi}_{k,k} - \phi_{k,k})| > \delta'(\varepsilon)\},\\ &= P\{|N^{\frac{1}{\beta}}(\hat{\phi}_{k,k} - \phi_{k,k})| > \tfrac{\delta'(\varepsilon)}{N^{1-\frac{1}{\alpha}-\frac{1}{\beta}}}\}.\end{aligned}$$

Take, $\delta = \frac{\delta'}{N^{1-\frac{1}{\alpha}-\frac{1}{\beta}}}$ and

$$1 - \frac{1}{\alpha} - \frac{1}{\beta} > 0, \ \ i.e, \ \beta > \frac{\alpha}{\alpha - 1}.$$

Then,

$$P\{|N^{\frac{1}{\beta}}(\hat{\phi}_{k,k} - \phi_{k,k})| > \delta\} < \varepsilon\,, \ \beta > \frac{\alpha}{\alpha - 1} \ \forall N > N_0.$$

That is,

$$N^{1/\beta}(\widehat{\phi}_{k,k} - \phi_{k,k}) = o_p(1).$$

Similarly we can show that

$$N^{1/\beta}(\widehat{\phi}_{k,-k} - \phi_{k,-k}) = o_p(1).$$

We also have $\phi_{k,k} = 0$ for $k > m$. This completes the proof. □

Theorem 5.4.2. *Let $\{X_n\}$ be a stationary $AR(m), m \geq 1$, process defined by (5.1), with $1 < \alpha \leq 2$, the upper bound $K(N) > m$, and $\widehat{m}$ be defined by (5.29). If*

(a)$\liminf_{N\to\infty} N^{2/\beta}\phi_{m,m}\phi_{m,-m} > 2m$ *and*

(b)$N^{2/\beta}\widehat{\phi}_{k,k}\widehat{\phi}_{k,-k} = o_p(1)$ *for $k > m$*

then,

$$\widehat{m} \xrightarrow{p} m \text{ as } N \to \infty.$$

Proof. Since $\widehat{m}$ is an integer valued, $\widehat{m} \xrightarrow{p} m$ is equivalent to $P(\widehat{m} = m) \to 1$ as $N \to \infty$. We have

$$\widehat{\gamma}_u(k) = \widehat{\gamma}(0)\left(\prod_{j=1}^{k}(1 - \widehat{\phi}_{j,j}\widehat{\phi}_{j,-j})\right)^{1/\alpha},$$

where $\widehat{\gamma}(0) = \frac{1}{N}\sum_{n=1}^{N}|X_n|$.

$$P(\widehat{m} < m) \leq P\{\min_{0\leq k<m} IC(k) \leq IC(m)\},$$

and since,

$$\min_{0\leq k<m} IC(k) \geq N^{2/\beta}\sum_{l=1}^{m-1}\frac{1}{\alpha}\ln(1 - \widehat{\phi}_{l,l}\widehat{\phi}_{l,-l}) + N^{2/\beta}\ln\left(\widehat{\gamma}(0)\right),$$

$$\begin{aligned} P(\widehat{m} < m) &\leq P\{\ln(1 - \widehat{\phi}_{m,m}\widehat{\phi}_{m,-m}) \geq \tfrac{-2m\alpha}{N^{2/\beta}}\} \\ &= P\{(1 - \widehat{\phi}_{m,m}\widehat{\phi}_{m,-m}) \geq e^{\frac{-2m\alpha}{N^{2/\beta}}}\} \\ &\leq P\{\tfrac{1}{\alpha}N^{2/\beta}\widehat{\phi}_{m,m}\widehat{\phi}_{m,-m} \leq 2m\} \end{aligned}$$

But,

$$N^{2/\beta}\widehat{\phi}_{m,m}\widehat{\phi}_{m,-m} = (N^{1/\beta}\phi_{m,m} + o_p(1))(N^{1/\beta}\phi_{m,-m} + o_p(1)).$$

So,

$$\limsup_{N\to\infty} P\{N^{2/\beta}\widehat{\phi}_{m,m}\widehat{\phi}_{m,-m} \leq 2m\} = 0,$$

since,

$$\liminf_{N\to\infty} N^{2/\beta}\widehat{\phi}_{m,m}\widehat{\phi}_{m,-m} > 2m.$$

Thus, $P(\widehat{m} < m) \to 0$. We can show that,

$$\begin{aligned} P(\widehat{m} > m) &\leq P\{IC(k) \leq IC(k-1) \text{ for some } m < k \leq K(N)\} \\ &\leq P\{N^{2/\beta} \min_{m<k\leq K(N)} \ln(1 - \phi_{k,k}\phi_{k,-k}) < -2\}. \end{aligned}$$

From Lemma 5.4.1, and assumption of the theorem we have,

$$N^{2/\beta} \max_{m<k\leq K(N)} \phi_{k,k}\phi_{k,-k} \xrightarrow{p} 0$$

so,

$$N^{2/\beta} \min_{m<k\leq K(N)} \ln(1 - \phi_{k,k}\phi_{k,-k}) \xrightarrow{p} 0.$$

Therefore,

$$P(\widehat{m} > m) \to 0.$$

This concludes the proof. □

5.5 Simulation

We simulate time series data using two different autoregressive models

$$X_n = 0.9X_{n-1} + U_n, \tag{5.30}$$

and

$$X_n = 0.5X_{n-1} + 0.3X_{n-2} + U_n. \tag{5.31}$$

The symmetric stable innovation sequences U_n are generated using the algorithm discussed in Section 3.6. For both models we simulate 1000 time series each having size N=1000 for 9 different values of $\alpha \in (1, 2]$ and follow the steps given below:

1. Plot sample partial covariation ($\hat{\tau}(k)$) at 10 lags and compare with 0.025 and 0.975 quantiles of $N^{\frac{1}{\alpha}-1}S$. The comparison criteria obtained under the null hypothesis $\tau(k) = 0$. (quantiles for stable distributions can be found in Samorodnitsky and Taqqu (1994)).

2. Plot $\tilde{\rho}(k) = (N/\log N)^{(1/\alpha)}\hat{\rho}(k)$ at 10 lags and compare with 0.025 and 0.975 quantiles of the distribution of U/V given in Adler et al. (1998).

Order of autoregressive models correctly identified has been counted at each strategy. The results corresponds to $AR(1)$ and $AR(2)$ models are respectively shown in Table 5.1 and 5.2.

From the simulation results we observe that, $\hat{\tau}(k)$ perform consistently well in $AR(1)$ case. When α close to 2, the partial auto covariation based identification perform well in both $AR(1)$ and $AR(2)$ case. At the same time its performance is not satisfactory when α approaches to 1 in an AR(2)case. From these studies we can conclude that sample PcovF is an acceptable tool for model identification in stable autoregressive processes. Figure 5.1 is the plot of partial auto-covariation function of simulated AR(1) series.

In this section we also study the performance of the proposed information criteria in model identification problem. In order to study the performance of model identification

criteria, we simulate time series data using three different autoregressive models

$$X_n = 0.7X_{n-1} + U_n, \tag{5.32}$$

$$X_n = 0.5X_{n-1} - 0.4X_{n-2} + U_n, \tag{5.33}$$

and

$$X_n = 0.5X_{n-1} + 0.4X_{n-2} - 0.35X_{n-2} + U_n. \tag{5.34}$$

The symmetric stable innovation sequences U_n are generated with dispersion parameters $\gamma_u = 1$ and two different heavy tailed indices $\alpha = 1.5$ and $\alpha = 1.7$. We estimate the partial auto-covariation function and the proposed information criteria using the Durbin-Levinson algorithm. Under each model assumption we count the number of correct identification out of 1000 time series each having size N=500, 1000,2000 and 5000.

We compare the performance of the new information criteria with the well-known Akaike Information Criteria (AIC). Results shows that the proposed information criteria outperform over AIC in most of the cases. The results are shown in Table 5.3 and 5.4. We also study the performance of the proposed criteria in Gaussian autoregressive models and the results in Table 5.5 show that our new criteria performs much better than the Akaike Information criteria. Figure 5.2 is the plot of $IC(k)$ as a function of order k for autoregressive models fitted to the simulated AR(3) series.

So far we have discussed the model identification and parameter estimation of stable autoregressive models. Now we can explore some of its applications in signal processing. Next chapter we will discuss some challenging signal processing problems based on our present studies.

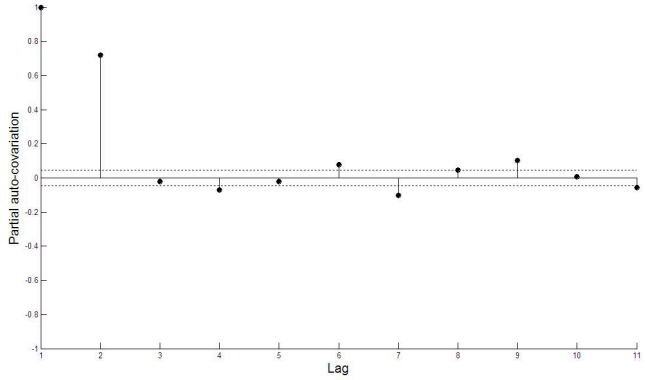

Figure 5.1: Partial auto-covariation plot of simulated AR(1) series

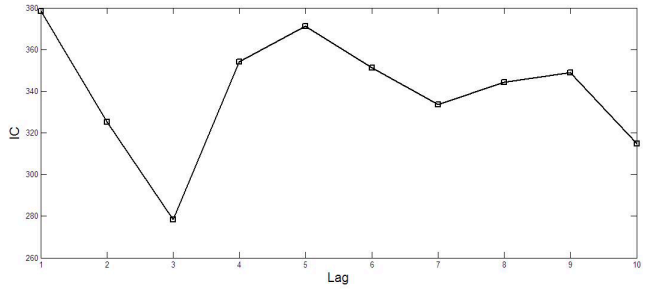

Figure 5.2: Proposed model selection criteria IC (k) as a function of order k for autoregressive models fitted to the simulated AR(3) series

Sample Size	alpha	Based on PACF	Based on PcovF
500	1.1	94.1	97.1
	1.2	92.9	97.3
	1.3	92.8	96.3
	1.4	91.5	95.5
	1.5	89.1	91.0
	1.6	83.1	82.5
	1.7	74.0	72.8
	1.8	3.8	65.1
	1.9	3.8	57.6
1000	1.1	92.6	98.1
	1.2	91.2	97.2
	1.3	90.5	97.0
	1.4	89.6	95.1
	1.5	85.1	90.3
	1.6	83.0	80.5
	1.7	78.2	68.4
	1.8	5.8	58.9
	1.9	7.4	54.0
5000	1.1	85.5	98.8
	1.2	85.9	97.7
	1.3	88.3	96.9
	1.4	85.9	94.2
	1.5	85.1	90.2
	1.6	85.1	81.8
	1.7	80.5	60.2
	1.8	8.0	40.7
	1.9	13.4	34.3

Table 5.1: Percentage of simulated series out of 1000 that were correctly identified as coming from $AR(1)$ process using above steps 1 and 2.

Sample Size	alpha	Based on PACF	Based on PcovF
500	1.1	95.7	0.3
	1.2	96.6	0.01
	1.3	96.8	3.4
	1.4	94.7	23.9
	1.5	92.3	66.7
	1.6	86.7	79.8
	1.7	78.2	76.9
	1.8	4.2	70.9
	1.9	4.4	65.9
1000	1.1	99.3	0.8
	1.2	94.2	1.1
	1.3	94.0	8.0
	1.4	92.2	61.9
	1.5	90.7	87.2
	1.6	86.9	86.4
	1.7	82.4	76.4
	1.8	5.2	65.9
	1.9	6.7	59.4
5000	1.1	91.1	1.6
	1.2	91.8	5.0
	1.3	91.1	80.0
	1.4	88.4	95.4
	1.5	87.7	92.9
	1.6	85.2	86.9
	1.7	87.9	69.7
	1.8	8.9	50.5
	1.9	13.2	43.5

Table 5.2: Percentage of simulated series out of 1000 that were correctly identified as coming from $AR(2)$ process using above steps 1 and 2.

Model	Sample Size	Based on IC	Based on AIC
AR(1)	500	99.6	81.6
	1000	99.9	82.7
	2000	99.8	84.9
	5000	99.8	86.2
AR(2)	500	48.6	83.7
	1000	80.3	83.3
	2000	93.9	84.2
	5000	98.2	85.0
AR(3)	500	70.2	81.9
	1000	91.0	84.6
	2000	94.8	84.3
	5000	95.0	86.2

Table 5.3: Percentage of correct identification out of 1000 stable time series with $\alpha = 1.5$.

Model	Sample Size	Based on IC	Based on AIC
AR(1)	500	98.8	82.3
	1000	99.2	86.9
	2000	99.6	85.9
	5000	99.6	87.8
AR(2)	500	72.2	83.0
	1000	90.7	85.5
	2000	95.4	85.1
	5000	97.4	91.1
AR(3)	500	82.7	83.0
	1000	85.8	85.6
	2000	81.4	86.3
	5000	76.9	89.6

Table 5.4: Percentage of correct identification out of 1000 stable time series with $\alpha = 1.7$.

Model	Sample Size	Based on IC	Based on AIC
AR(1)	500	99.7	70.7
	1000	99.8	69.7
	2000	100.0	71.3
	5000	100.0	71.7
AR(2)	500	98.6	68.2
	1000	99.7	70.1
	2000	99.8	71.0
	5000	100.0	71.3
AR(3)	500	99.2	74.2
	1000	100.0	71.5
	2000	100.0	71.6
	5000	100.0	72.6

Table 5.5: Percentage of correct identification out of 1000 Gaussian time series

Chapter 6

Application of stable time series models in statistical signal processing

6.1 Introduction

Time series methods and models are widely used in modeling and analysis of signal and noise processes. Most of these studies assume that the measurement noise is driven by Gaussian model. Spectral density function is the basic tool for analyzing the periodic behavior of the signal. Estimating spectral density from an observed signal is an important problem in statistical signal processing and in general time series analysis. Periodogram is generally used as a non-parametric estimate of spectral density function but it is not an unbiased and consistent estimator. Parametric spectrum analysis techniques assume some time series models to the signals and estimate their theoretical spectral density function. Frequency estimate of the signals can be obtained from the estimated spectrum. Frequency estimation of multiple sinusoids from noisy measurements has been extensively studied in statistical signal processing (see Castanie (2006), Stoica et al. (1989)). When the signal is observed in impulsive noise environment, performance of the models under Gaussian assumption is poor. As we discussed in the previous chapters stable distributions are recommended in this situations even though spectral density function cannot be defined

mathematically. Kluppelberg and Mikosch (1993) discussed the power transfer function for infinite variance case which is similar to classical spectral density in finite variance case. This function determines the model completely and hence it can be considered as an important tool for analyzing stable processes. Kluppelberg and Mikosch (1993) show that normalized periodogram can be used to construct a consistent estimate of power transfer function in infinite variance case. Mikosch et al. (1995) used it for the parameter estimation of stable ARMA process with infinite variance. Frequency estimation of symmetric stable signals have been studied by Altinkaya et al. (2002) and Nandi et al. (2002) using subspace and periodogram method respectively. Some other works in this area include Tsakalides and Nikias (1998) where they discussed frequency estimation based on the noise subspace method using auto-covariation function. Liu and Mendel (2001) investigated the same problem for those signals that consist of circular signals in symmetric stable noise by using the fractional lower order moments (FLOMs). In the present study we focus on the parametric spectrum estimation of multiple sinusoids from symmetric stable noisy measurements. The model description is parallel to that of Altinkaya et al. (2002), but we use parametric spectral methods for estimation. The problem reduces to the estimation of the power transfer function of stable autoregressive model. Here we discuss the statistical properties of the generalized Yule-Walker estimates in the spectrum and frequency estimation problem.

Another important problem in statistical signal processing is to identify the number of frequency components in an observed signal. A signal model consists of q sinusoidal frequency component and which follows an autoregressive model of order $2q$. So the identification of number of components reduces to the model identification of AR model. Usually AR models of higher orders are widely used in this context. So in the present study we modify the proposed information criteria for order selection of higher order AR models using singular values and eigen values of the sample auto-covariation matrix.

The rest of this chapter is divided in to two parts: In the first part we describe the frequency estimation of sinusoidal signal observed in a symmetric stable noisy environment.

In the second part we study the identification of number of frequency component in an observed signal.

6.2 Multiple sinusoidal signal plus symmetric stable noise model

Consider a signal model involving of q sinusoids,

$$x_n = \sum_{k=1}^{q} A_k \sin(\omega_k n + \theta_k) \tag{6.1}$$

observed in additive symmetric stable noise

$$y_k = x_k + u_k, k = 1, 2..., N \tag{6.2}$$

where $A_k, k = 1, ..., q$ and $\omega_k, k = 1, ..., q$ are the unknown real constants, which are respectively amplitude and angular frequency, whereas $\theta_k, k = 1, ..., q$ are phase of the sinusoids assumed to be realizations of random variables, distributed uniformly and independently over $(0, 2\pi]$. The observed signal $\{y_n\}$ and noise $\{u_n\}$ are realizations of the observation sequence $\{Y_n\}$ and the independent and identically distributed symmetric stable noise sequence $\{U_n\}$, respectively and N is the sample size. The signal process $\{x_n\}$ follows autoregressive equation given below (see Altinkaya et al. (2002)).

$$\phi(B)x_n = 0, \tag{6.3}$$

where $\phi(B)$ is a polynomial of degree $2q$ given by

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_{2q} B^{2q} = \prod_{k=1}^{q} (1 - 2cos\omega_k B + B^2). \tag{6.4}$$

From (6.2) and (6.3), we have

$$\phi(B)y_n = \phi(B)u_n. \tag{6.5}$$

Observed process, y_n in (6.5) follows $ARMA\ (2q, 2q)$ model.

Model (6.5) can be approximated by a stable autoregressive model of order $m \geq 2q$,

$$Y_n = \phi_1 Y_{n-1} + \phi_2 Y_{n-2} + ... + \phi_m Y_{n-m} + U_n. \tag{6.6}$$

Choice of m depends upon the noise dispersion, $m = 2q$ might be an appropriate order for a small noise dispersion. Some order selection methods will be discussed in Section 6.4. Autoregressive parameters in model (6.5) play an important role in the spectrum analysis and frequency estimation problem (Stoica et al. (1989), Castanie (2006)). Basic idea is first to estimate the autoregressive parameters of ARMA process $\{y_n\}$ and then to determine the frequency $\{\omega_k\}$, by computing poles of the estimated AR spectrum or zeros of the estimated AR polynomial. Now we look at the problems associated with the estimation of AR parameters. In finite variance case we have various AR spectrum estimation techniques, popular one is Yule-Walker technique. Stoica et al. (1989) introduced many advanced versions of this technique which includes over-determined Yule-Walker estimator, under-determined Yule-Walker estimator and higher-order Yule-Walker estimator. The basic fact behind all these studies is that the estimation accuracy may increase considerably with increased number of Yule-Walker equations. This is parallel to our estimation study discussed in Chapter 4 based on generalized Yule-Walker equation. In the present study we discuss the estimation of autoregressive parameters in model (6.5) by increasing the number of generalized Yule-Walker equations.

6.3 Spectrum Analysis for Stable Processes

Classical spectrum analysis is classified in to parametric and non-parametric methods. In Parametric spectral methods, we derive the spectrum of an autoregressive series and

then estimate the unknown parameters involved in it. No assumption is made about the parametric form of the spectral density in non-parametric set up.

6.3.1 periodogram based spectrum analysis

Kluppelberg and Mikosch (1993,1994) studied the asymptotic behavior of periodogram. The results obtained indicate that the self-normalized periodogram behave like a periodogram in infinite variance linear processes. Let us consider the linear process defined by (2.44), that is

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j u_{t-j}, \tag{6.7}$$

where $\{u_t\}$ is an innovation sequence of iid symmetric stable random variables for $\alpha \in (0, 2]$. Self-normalized periodogram may be defined as follows,

$$\widetilde{I}_Y(\omega) = \frac{I_Y(\omega)}{\gamma_Y},\ \omega \in [-\pi,\ \pi] \tag{6.8}$$

where, $I_Y(\omega)$ is the periodogram defined as,

$$I_Y(\omega) = N^{-2/\alpha} \left| \sum_{t=1}^{N} Y_t e^{-i\omega t} \right|^2,$$

and

$$\gamma_Y = N^{-2/\alpha} \sum_{t=1}^{N} Y_t^2.$$

Theorem 3.1 of Kluppelberg and Mikosch (1994) states that, under some assumptions on the weight function and for $\alpha \in (0, 2]$, smoothed version of $\widetilde{I}_Y(\omega)$ converges in probability to $\frac{|\psi(\omega)|^2}{\psi^2}$ where

$$|\psi(\omega)|^2 = |\sum_{j=-\infty}^{\infty} \psi_j e^{-ij\omega}|^2,\ \omega \in [-\pi,\ \pi], \tag{6.9}$$

denotes the **power transfer function** of the linear filter (ψ_j) and

$$\psi^2 = \sum_{j=-\infty}^{\infty} \psi_j^2.$$

In infinite variance case the so-called **power transfer function** $|\psi(\omega)|^2$ corresponds to the classical spectral density and it determines the model completely (Kluppelberg and Mikosch (1993)).

6.3.2 Parametric spectrum analysis

In classical time series set up U_n is considered to be a Gaussian noise and we assume that the observed sequence $\{Y_n\}$ follows an autoregressive process. Spectrum of an autoregressive process of order m is defined as (Box et al. (1994), Brockwell and Davis(1987), Shumway and Stoffer (2006)),

$$S(\omega) = \frac{\sigma_u^2}{\phi(\omega)\phi^*(\omega)} \tag{6.10}$$

where

$$\phi(\omega) = 1 + \sum_{k=1}^{m} \phi_k \exp(-ik\omega)$$

and $\phi^*(.)$ is the complex conjugate of $\phi(.)$. In practical situations we have to estimate this spectrum from an observed signal. From the definition of autoregressive spectrum we can see that the problem of estimating spectrum reduces to that of identification of order of autoregressive model, estimation of AR parameters and innovation variance. In Chapter 2 we addressed the order identification and parameter estimation of autoregressive models under classical time series set up. Akaike information criteria and minimum description length (MDL) are widely used to identify the order of autoregressive model. Parameters of the autoregressive model can be estimated using some higher order Yule-Walker equations (for details see Stoica et al. (1989), Castanie (2006)).

In the present chapter our emphasis is on stable autoregressive model and its spectral

properties. If we consider the model (6.6) as a stable autoregressive model of order m then we cannot define spectral density to this model. However we can define power transfer function, which plays the same role as that of power spectral density in stable time series models. Properties of the stable ARMA model discussed in Chapter 2 can be used to define the power transfer function in infinite variance stable process. From Proposition 2.3.6 and equations (2.44), (2.45) the power transfer function of a stable autoregressive process of order m may be defined as,

$$S(\omega) = \frac{1}{\phi(\omega)\phi^*(\omega)} \tag{6.11}$$

where

$$\phi(\omega) = 1 + \sum_{k=1}^{m} \phi_k \exp(-ik\omega).$$

The estimates of autoregressive parameters can be obtained by the Yule-Walker method discussed in previous chapters using sample autocorrelation function or auto-covariation function with proper modification. As we mentioned in Chapter 4 the generalized Yule-Walker estimates derived from (2.54) based on Y_n will be biased due to the presence of measurement noise. To minimize the bias of EYW estimates, we suggested that a large number of EYW equations to be included for estimation. So we have to choose an optimum number m' of equations, such that m' is probably large compared to m but small compared to the sample size. Castanie (2006), page 156 recommended the value of m' should be at most of the order $\frac{N}{2}$ when the unbiased estimate of the correlation is used. We have employed the same approach in terms of covariation in Chapter 4. An estimate $\widehat{\phi}$ for the autoregressive parameter vector is obtained using ordinary least square regression of generalized Yule-Walker equations given in (4.8) and is written as,

$$\widehat{\phi}_a = (\widehat{\Delta}'_{m',m}\widehat{\Delta}_{m',m})^{-1}\widehat{\Delta}'_{m',m}\widehat{T}_{m'}. \tag{6.12}$$

We can apply the singular value decomposition to the auto-covariation matrix because it proves a greater numerical stability. This will also provide flexibility while using auto-

covariation matrix. This will go parallel to the method discussed in Section 4.4 of Chapter 4. Proposed estimate can be considered as a modified version of the ordinary least square generalized Yule-Walker (MLSGYW) estimates.

Here we apply singular value decomposition to the matrix $\widehat{\Delta}_{p',p}$, that is,

$$\widehat{\Delta}_{m',m} = U\Lambda V^T. \tag{6.13}$$

In (6.13) U and V are orthogonal matrices of respective orders $m' \times m'$ and $m \times m$. Here Λ is a matrix of size $m' \times m$, which can be written as,

$$\Lambda = \begin{bmatrix} \Lambda_{m,\,m} \\ 0_{m'-m,\,m} \end{bmatrix},$$

where $\Lambda_{m,\,m}$ is a diagonal matrix.

Now let us define a matrix $\widehat{\Lambda}_{m,m}$, based on singular values ν, of $\widehat{\Delta}_{m',m}$ as follows:

$$\widehat{\Lambda}_{m,m} = diag(\nu_1, \nu_2, ..., \nu_m),\ \nu_1 \geq \nu_2 \geq, ..., \geq \nu_m. \tag{6.14}$$

$$\overline{\widehat{\Delta}}_{m',m} = U\overline{\widehat{\Lambda}}V^T. \tag{6.15}$$

where,

$$\overline{\widehat{\Lambda}} = \begin{bmatrix} \widehat{\Lambda}_{m,m} \\ 0 \end{bmatrix}.$$

Now we can modify the estimate of ϕ as,

$$\widehat{\phi}_b = (\overline{\widehat{\Lambda}})^+ \widehat{T}_{m'} \tag{6.16}$$

where,

$$(\overline{\widehat{\Lambda}})^{+} = V \begin{bmatrix} \widehat{\Lambda}_{m,m}^{-1} \\ 0 \end{bmatrix} U^{T}$$

denotes the Moore-Penrose pseudo inverse of $\overline{\widehat{\Lambda}}$.

An estimate of the power transfer function can be written as,

$$\widehat{S}(\omega) = \frac{1}{\widehat{\phi}(\omega)\widehat{\phi}^{*}(\omega)} \tag{6.17}$$

where,

$$\widehat{\phi}(\omega) = 1 + \sum_{k=1}^{m} \widehat{\phi}_{bk} \exp(-ik\omega).$$

This estimate is known as the parametric estimate of power transfer function. In finite variance set up parametric spectrum estimation of this type is known as maximum entropy spectrum estimation.

6.4 Frequency Estimation

Nandi et al. (2002) studied the estimation of frequencies in the presence of heavy tailed errors, namely errors with symmetric stable distribution. They proposed an approximate least square estimate(ALSE) of the frequency component by maximizing the periodogram function

$$I_Y(\omega) = \frac{2}{N} \left| \sum_{t=1}^{N} Y_t e^{-i\omega t} \right|^2 ,$$

with respect to ω. If $\widetilde{\omega}$ maximizes $I_Y(\omega)$, then $\widetilde{\omega}$ is called the ALSE of ω. They have also shown that the proposed estimator is strongly consistent under certain conditions. Periodogram based spectral estimates are not of high resolution estimates in general and it is one of the crude methods in frequency estimation. So the parametric methods are

recommended in many applications.

In this section we explore the application of parametric spectral analysis and frequency estimation in infinite variance set up. Now we can study the estimation of parameters of an autoregressive model and the frequency estimates are derived by the poles of the estimated power transfer function.

Substituting these estimates in equation (6.17) we get the power transfer function estimate $\widehat{S}(\omega)$. Now we can study the estimation of frequencies of poles of the model. When $2q < m$, we can estimate only q frequencies. From equation (6.4), we have,

$$\phi(z) = \sum_{k=0}^{2q} \phi_k z^{-k} = \prod_{k=1}^{2q} (1 - z_k z^{-1}) \tag{6.18}$$

$$\text{that is, } \phi(z) = \prod_{k=1}^{q} (1 - r_k \exp(i\omega) z^{-1})(1 - r_k \exp(-i\omega) z^{-1}). \tag{6.19}$$

We will note, $\widehat{z}_k = \widehat{r}_k \exp(-i\widehat{\omega})$, $k = 1, 2, ..., q$ as the estimates of the poles obtained from $\sum_{k=0}^{2q} \widehat{\phi}_k z^{-k}$. We can find out the asymptotic distribution of the frequency estimate $N^{1-\frac{1}{\alpha}}(\omega - \widehat{\omega})$,where $\widehat{\omega} = (\widehat{\omega}_1, \widehat{\omega}_2, ..., \widehat{\omega_q})$ and $\omega = (\omega_1, \omega_2, ..., \omega_q)$, from that of the statistics $N^{1-\frac{1}{\alpha}}(\phi - \widehat{\phi})$. The frequency estimate $\widehat{\omega}$ is observed as a function of $\widehat{\phi}$, that is, $\widehat{\omega} = \xi(\widehat{\phi})$, and the function is continuously differentiable, so that the asymptotic distribution of $N^{1-\frac{1}{\alpha}}(\omega - \widehat{\omega})$, can be obtained using the results of Theorem 6.4.1.

Theorem 6.4.1. *Assume that $\widehat{\omega}_k$ obtained by the zeros of $\phi(z)$ which corresponds to the $\widehat{\phi}_b$ give by* (6.16)*. Then,*

$$\widehat{\omega} \xrightarrow{p} \omega$$

and,

$$N^{1-\frac{1}{\alpha}}(\widehat{\omega} - \omega) \xrightarrow{L} PHFS \tag{6.20}$$

where,

$$P = \left(\begin{array}{ccc|ccc} \frac{\gamma_1}{\gamma_1^2+\delta_1^2} & \cdots & 0 & \frac{-\gamma_1}{\gamma_1^2+\delta_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\gamma_q}{\gamma_l^2+\delta_l^2} & 0 & \cdots & \frac{-\gamma_q}{\gamma_l^2+\delta_l^2} \end{array} \right)$$

$$H = (h_1, \ldots, h_q, g_1, \ldots, g_q)' \; and, for \;\; k = 1, ..., q$$

$$\begin{aligned} \gamma_k &= [r_k^{-2} \sin(\omega_k), \ldots, m r_k^{-m-1} \sin(m\omega_k)] \; \phi \\ \delta_k &= [r_k^{-2} \cos(\omega_k), \ldots, m r_k^{-m-1} \cos(m\omega_k)] \; \phi \\ g_k &= [r_k^{-2} \sin(\omega_k), \ldots, r_k^{-m-1} \sin(m\omega_k)] \\ h_k &= [r_k^{-2} \cos(\omega_k), \ldots, r_k^{-m-1} \cos(m\omega_k)] \end{aligned}$$

Proof. Let

$$\widehat{r}_k \exp(\pm\widehat{\omega_k}), k = 1, 2, \ldots q,$$

denote the q smallest modulus roots of $\hat{\phi}(z)$. Similar to Theorem 4.4.2 we can show that $\widehat{\phi_b} \xrightarrow{p} \phi$. Since the frequency estimate $\widehat{\omega} = \xi(\widehat{\phi})$ and the function ξ is continuous, the proof of, $\widehat{\omega} \xrightarrow{p} \omega$ follows by the continuous mapping theorem.

Using Taylor's series expansion we can show that

$$\begin{aligned} 0 = Re(\hat{\phi}(\hat{r}_k e^{i\hat{\omega}_k})) = Re(\hat{\phi}(r_k e^{i\omega_k})) + \frac{\partial Re\{\hat{\phi}(re^{i\omega})\}}{\partial r}|_{\substack{r=r_k \\ \omega=\omega_k}} (\hat{r}_k - r_k) \; + \\ \frac{\partial Re\{\hat{\phi}(re^{i\omega})\}}{\partial \omega}|_{\substack{r=r_k \\ \omega=\omega_k}} (\hat{\omega}_k - \omega_k) + O(N^{\frac{1}{\alpha}-1}). \end{aligned} \tag{6.21}$$

Similarly,

$$0 = Im(\hat{\phi}(\hat{r}_k e^{i\hat{\omega}_k})) = Im(\hat{\phi}(r_k e^{i\omega_k})) + \frac{\partial Im\{\hat{\phi}(re^{i\omega})\}}{\partial r}|_{\substack{r=r_k \\ \omega=\omega_k}} (\hat{r}_k - r_k) \; + \\ \frac{\partial Im\{\hat{\phi}(re^{i\omega})\}}{\partial \omega}|_{\substack{r=r_k \\ \omega=\omega_k}} (\hat{\omega}_k - \omega_k) + O(N^{\frac{1}{\alpha}-1}). \tag{6.22}$$

$$\begin{aligned} \frac{\partial Re\{\hat{\phi}(re^{i\omega})\}}{\partial r}|_{\substack{r=r_k \\ \omega=\omega_k}} &= [r_k^{-2}\cos(\omega_k), \ldots, mr_k^{-m-1}\cos(m\omega_k)]\ \hat{\phi} \\ \frac{\partial Re\{\hat{\phi}(re^{i\omega})\}}{\partial \omega}|_{\substack{r=r_k \\ \omega=\omega_k}} &= -[r_k^{-2}\sin(\omega_k), \ldots, mr_k^{-m-1}\sin(m\omega_k)]\ \hat{\phi} \\ \frac{\partial Im\{\hat{\phi}(re^{i\omega})\}}{\partial r}|_{\substack{r=r_k \\ \omega=\omega_k}} &= [r_k^{-2}\sin(\omega_k), \ldots, mr_k^{-m-1}\sin(m\omega_k)]\ \hat{\phi} \\ \frac{\partial Im\{\hat{\phi}(re^{i\omega})\}}{\partial \omega}|_{\substack{r=r_k \\ \omega=\omega_k}} &= [r_k^{-2}\cos(\omega_k), \ldots, mr_k^{-m-1}\cos(m\omega_k)]\ \hat{\phi}. \end{aligned} \tag{6.23}$$

Since $(\hat{\phi} - \phi) = O(N^{\frac{1}{\alpha}-1})$ dominant term in (6.21) and (6.22) is not affected if we replace $\hat{\phi}$ by ϕ in (6.23). Thus we get,

$$0 = Re(\hat{\phi}(e^{i\omega_k})) + \delta_k(\widehat{r}_k - r_k) - \gamma_k(\widehat{\omega}_k - \omega_k) + O(N^{\frac{1}{\alpha}-1}), \tag{6.24}$$

and

$$0 = Im(\hat{\phi}(e^{i\omega_k})) + \gamma_k(\widehat{r}_k - r_k) + \delta_k(\widehat{\omega}_k - \omega_k) + O(N^{\frac{1}{\alpha}-1}). \tag{6.25}$$

On solving (6.24) and (6.25) (recall $\phi(r_k e^{i\omega_k}) = 0$) we can write,

$$\widehat{\omega}_k - \omega_k = \frac{\gamma_k h_k - \delta_k g_k}{\delta_k^2 + \gamma_k^2}(\hat{\phi} - \phi) + O(N^{\frac{1}{\alpha}-1}),\ k = 1, ..., q. \tag{6.26}$$

Since $\phi(r_k e^{i\omega_k}) = 0$, we can show that $h_k\phi = 0$ and $g_k\phi = 0$. We can write (6.26) in more compact form,

$$\widehat{\omega} - \omega = PH(\hat{\phi} - \phi) + O(N^{\frac{1}{\alpha}-1}).$$

Similar to the second part of Theorem 4.4.2 we can show that

$$N^{1-\frac{1}{\alpha}}(\widehat{\omega} - \omega) \xrightarrow{L} PHFS.$$

□

Note:

1. One can also use sample autocorrelation based Yule-Walker method to estimate the autoregressive parameters and the corresponding power transfer function (see Adler et al. (1997)). Similar to the above theorems, we can show that the resulting estimates also have the asymptotic properties,

$$(\frac{N}{\log(N)})^{\frac{1}{\alpha}}(\widehat{\omega}-\omega) \xrightarrow{L} PHDS, \tag{6.27}$$

where D is defined as in equation (2.47).

6.5 Number of frequency components in an observed signal

Estimating the number of frequency components is an important problem in Signal Processing. Estimation of q of the model (6.1) has been considered by several authors in the Signal Processing and Time Series literature for few years (Kundu (2002), Kundu and Nandi (2005), Quinn (1989), Wang (1993), Quinn and Thomson (1991), Sakai (1990, 1993)). Kundu (2002) developed a method using the penalty function technique. Kundu and Nandi (2005) studied the estimation of number of components of the fundamental frequency model when all the adjacent harmonics are present. They proposed a consistent penalty function based information criteria. In the present section we explore the scope of the information criteria proposed in Chapter 5 for order estimation using the decomposition method by Castanie (2006), page 165.

In Section 6.2 we have seen the relation between number of frequency components of an observed signal and the order of autoregressive model. If number of frequency components of an observed signal is q, then the order of auto-regressive polynomial is $p = 2q$. So the problem reduces to the estimation of the order of auto-regressive model. This is a very popular method in signal processing literature. The information criteria depend on the singular values of the auto-covariation matrix. The proposed Information criteria is,

$$IC(k) = N^{2/\beta}(m-k)\ln\left(\frac{(\prod_{t=k+1}^{m}\widehat{\nu}_t)^{\frac{1}{m-k}}}{\frac{1}{m-k}\sum_{t=k+1}^{m}\widehat{\nu}_t}\right) + k(2m-k), \tag{6.28}$$

where, $k = 1, 2, ..., m-1$ and $p < m < N$. Simulation studies show that the modified $IC(k)$ performs well in order identification.

We can also define these criteria according to the eigen values τ_t of the auto-covariation matrix by replacing $\widehat{\nu}_t$ by $\widehat{\tau}_t$ in equation (6.28) to get

$$IC(k) = N^{2/\beta}(m-k)\ln\left(\frac{(\prod_{t=k+1}^{m}\widehat{\tau}_t)^{\frac{1}{m-k}}}{\frac{1}{m-k}\sum_{t=k+1}^{m}\widehat{\tau}_t}\right) + k(2m-k), \tag{6.29}$$

where, $k = 1, 2, ..., m-1$ and $p < m < N$. Theoretical performance of this proposed criterion can be considered as a future problem. We can also extend the scope of this function parallel to minimum description length (MDL) criteria and Bayesian information criteria (BIC) along with proper penalty functions.

6.6 Simulation

Suppose that an observed time series $\{y_t\}$ evolves according to the equation (6.2) and the unobserved signal $\{x_t\}$, is a sinusoidal signal which can be represented as follows:

$$x_t = A_1\sin(\omega_1 t + \phi_1) + A_2\sin(\omega_2 t + \phi_2),\ t = 1, 2, ... \tag{6.30}$$

Also assume that measurement noise u_t in (6.2) follows $\sim S_\alpha(\lambda_u)$, where λ_u is the dispersion parameter of u_t. Our main objective is to estimate the frequency component of the signal $\{x_t\}$ from the given observed signal $\{y_t\}$. This problem can be solved by applying the methods discussed in Section 6.4. We simulate the above model by taking 100 time series each having size n=50, 100, 200 for 3 different set of frequencies (0.1315,0.4025), (0.1215,0.3525), (0.2123,0.4321) and $\alpha = 1.5$. The symmetric stable innovation sequences u_t are generated using the algorithm discussed in Section 3.6. Mean value and the mean

square error of the estimates are computed. Simulation result show that the estimates obtained using the proposed method performs well and consistent. The simulation results are displayed in Table 6.1.

Number of frequencies	Percentage of correct identification (sample size)
1	83.6(50)
2	83.3(50)
3	63.7(50)
1	92.0(100)
2	85.6(100)
3	90.8(100)

Table 6.1: Percentage of sinusoidal frequencies were correctly identified out of 100 trial.

Sample size	Actual Frequencies		Estimated Frequencies		Mean squared Errors	
	f1	f2	f1	f2	f1	f2
50	0.1315	0.4025	0.1315	0.3877	0.00059	0.00062
100			0.1319	0.3841	0.00016	0.00089
200			0.1345	0.3827	0.00027	0.0006
50	0.1215	0.3525	0.1247	0.3927	0.00082	0.0012
100			0.1214	0.3926	0.00043	0.0019
200			0.1234	0.3912	0.00012	0.0016
50	0.2123	0.4321	0.2144	0.4437	0.0011	0.0086
100			0.2134	0.4672	0.00079	0.0069
200			0.2146	0.4801	0.00079	0.0006

Table 6.2: Mean and Mean Square error of the estimated frequencies

The estimated power spectral density obtained through the proposed method is displayed in Figure 6.1.

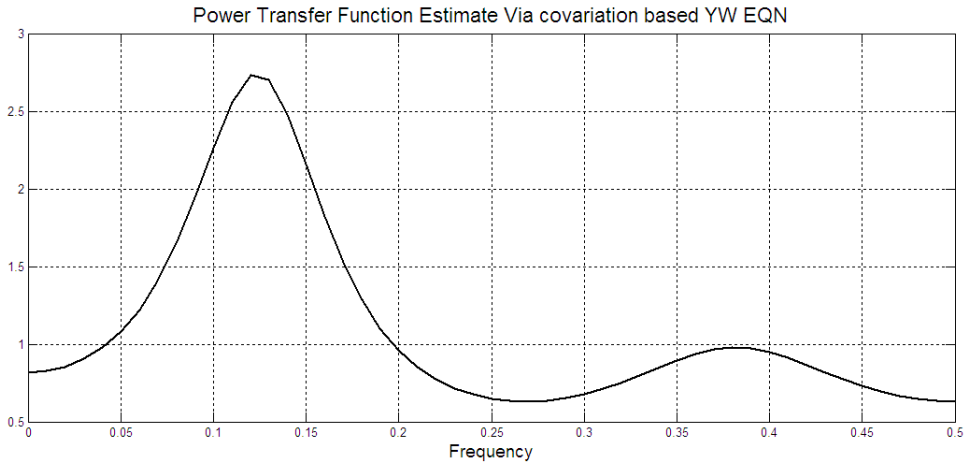

Figure 6.1: Estimated Power Transfer Function based on the frequencies (0.1315, 0.4025)

Chapter 7

Conclusions and Further Research

The thesis has covered various aspects of modeling and analysis of finite mean time series with symmetric stable distributed innovations. Time series analysis based on Box and Jenkins methods are the most popular approaches where the models are linear and errors are Gaussian. We highlighted the limitations of classical time series analysis tools and explored some generalized tools and organized the approach parallel to the classical set up. In the present thesis we mainly studied the estimation and prediction of signal plus noise model. Here we assumed the signal and noise follow some models with symmetric stable innovations.

We start the thesis with some motivating examples and application areas of alpha stable time series models. Classical time series analysis and corresponding theories based on finite variance models are extensively discussed in second chapter. We also surveyed the existing theories and methods correspond to infinite variance models in the same chapter.

We present a linear filtering method for computing the filter weights assigned to the observation for estimating unobserved signal under general noisy environment in third chapter. Here we consider both the signal and the noise as stationary processes with infinite variance innovations. We derived semi infinite, double infinite and asymmetric signal extraction filters based on minimum dispersion criteria. Finite length filters based on Kalman-Levy filters are developed and identified the pattern of the filter weights. Simulation studies

show that the proposed methods are competent enough in signal extraction for processes with infinite variance.

Parameter estimation of autoregressive signals observed in a symmetric stable noise environment is discussed in fourth chapter. Here we used higher order Yule-Walker type estimation using auto-covariation function and exemplify the methods by simulation and application to Sea surface temperature data. We increased the number of Yule-Walker equations and proposed a ordinary least square estimate to the autoregressive parameters. Singularity problem of the auto-covariation matrix is addressed and derived a modified version of the Generalized Yule-Walker method using singular value decomposition.

In fifth chapter of the thesis we introduced partial covariation function as a tool for stable time series analysis where covariance or partial covariance is ill defined. Asymptotic results of the partial auto-covariation is studied and its application in model identification of stable auto-regressive models are discussed. We generalize the Durbin-Levinson algorithm to include infinite variance models in terms of partial auto-covariation function and introduce a new information criteria for consistent order estimation of stable autoregressive model.

In chapter six we explore the application of the techniques discussed in the previous chapter in signal processing. Frequency estimation of sinusoidal signal observed in symmetric stable noisy environment is discussed in this context. Here we introduced a parametric spectrum analysis and frequency estimate using power transfer function. Estimate of the power transfer function is obtained using the modified generalized Yule-Walker approach. Another important problem in statistical signal processing is to identify the number of sinusoidal components in an observed signal. We used a modified version of the proposed information criteria for this purpose.

Nonlinear filtering such as particle filtering and sequential Bayesian filtering are some alternative to the classical signal extraction methods. Tanizaki (2001) discussed maximum likelihood estimation of non-linear non-Gaussian state space models. He used monte carlo optimization procedure for maximization of the likelihood function. This procedure simultaneously estimates signal, noise process and corresponding parameters in the model. Qiou

and Ravisankar (1998) proposed a likelihood function for stable ARMA process. Both these papers motivate us to propose a generalized non linear signal extraction procedure. This can be considered as a future work in this direction.

In previous chapters we have seen some difficulties associated with the covariation based estimation. One major reason is that the covariation function is not symmetric. This makes expensive computation in the generalized Durbin-Levinson algorithm. Reduction in the complexity of generalized Durbin-Levinson using some decomposition method can also be considered as a future work.

There is a great deal of interest in the spectrum analysis and frequency estimation using pseudo spectral techniques. Multiple signal classification (MUSIC) and covariation-based multiple signal classification (ROC-MUSIC) (Tsakalides and Nikias (1998)) are some of the existing techniques in this direction. Statistical properties of these estimates are not well studied in literature. We can also extend the scope of the model identification criteria proposed in this chapter to minimum description length (MDL) criteria and Bayesian information criteria (BIC) with proper penalty function. Further research in this direction is in progress.

Bibliography

[1] Achim, A., Bezerianos, A. and Tsakalides, P. (2002). *SAR image denoising: a multiscale robust statistical approach.* IEEE Proceedings. 14-th International Conference on Digital Signal Processing (DSP 2002), Santorini, Greece II, 1235 (1238).

[2] Adler, R., Feldman, R. and Gallagher, C. (1998). *Analysing stable time series.* In: Adler, R.J., Feldman, R.E., Taqqu, M.S. (Eds.), A Practical Guide to Heavy Tails. BirkhTauser, Boston.

[3] Akaike, H.(1970). *Statistical predictor identification.* Annals of Institute of Statistical Mathathematics, 22, 203-217.

[4] Akaike, H. (1974). *A new look at the statistical model identification.* IEEE Transactions on automatic control, 19 716-723.

[5] Altinkaya, M. A., Delic, H., Sankur, B. and Anarim, E. (2002). *Subspace-based frequency estimation of sinusoidal signals in alpha stable noise.* Signal Processing, 82, 1807-1827.

[6] Ansley, C.F. and Kohn, R. (1985). *Estimation, filtering and smoothing in state space models with incompletely specified initial conditions.* Annals of Statistics 13, 1286-1316.

[7] Balakrishna, N. and Hareesh, G. (2009). *Statistical Signal Extraction Using Stable Processes.* Statistics and Probability Letters, 79, 851-856.

[8] Balakrishna, N. and Hareesh, G. (2010a). *Stable Autoregressive Models and Signal Estimation.* (communicated).

[9] Balakrishna, N. and Hareesh, G. (2010b). *Stable Autoregressive Models and Time Series Analysis.* (communicated).

[10] Bell, W. R. (1984a). *Signal Extraction for Nonstationary Time Series.* Annals of Statistics, 12, 646-664.

[11] Bell, R. and Hillmer, S. C. (1988). *A Matrix Approach to Signal Extraction and Likelihood Evaluation of ARIMA component Time Series Models.* Research Report Number 88/22, Statistical Research Division, Beauro of the Census.

[12] Bell, W. R. and Martin, D. E. (2004). *Computation of Asymmetric Signal Extraction Filters and Mean Squared Error for ARIMA Component Models* . Journal of Time Series Analysis, 25, 603-635.

[13] Bhansali, R. J. (1988). *Consistent order determination for processes with infinite variance.* Journal of Royal Statistical Society, Sories B, 50, 46-60.

[14] Bidarkota, P. and McCulloch, J.H. (1998). *Optimal Univariate Inflation Forecasting with Symmetric Stable. Shocks'.* Journal of Applied Econometrics, 13, 659-70.

[15] Bobrowski, A. (2005). *Functional Analysis for Probability and Stochastic Processes.* Cambridge University Press.

[16] Bouchaud, J.P., Sornette, D., Walter, C. and Aguilar, J. P. (1998). *Taming large events: Optimal portfolio theory for strongly fluctuating assets.* International Journal of Theoretical and Applied Finance 1, 25-41.

[17] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis, Forecasting and Control.* San Fransisco: Holden Day.

[18] Breiman, L. (1968). *Probability.* Reading, MA: Addison-Wesley.

[19] Breton, A. L. and Musiela, M. (1993). *A generalization of the Kalman filter to models with infinite variance.* Stochastic Processes and their Applications, 47 (1), 75-94.

[20] Brockwell, P.J. and Davis, R.A. (1987).*Time Series: Theory and Methods*, New York: Springer-Verlag.

[21] Buckle, D. J. (1995). *Bayesian inference for stable distributions.* Journal of the American Statistical Association, 90, 605-613.

[22] Burridge, P. and Hristova, D. (2007). *Consistent estimation and order selection for non-stationary autoregressive processes with stable innovations.* Journal of Time Series Analysis, 29 (4), 695-718.

[23] Burridge, P. and Wallis, K.F., (1988). *Prediction theory for autoregressivemoving average processes.* Econometric Review, 7, 65-95.

[24] Cambanis, S. and Miller, G. (1981). *Linear problems in order stable processes.* SIAM Journal of Applied Mathematics, 41 (1), 43-61 .

[25] Castanie, F. (2006). *Spectral Analysis: parametric and non-parametric digital methods.* ISTE Ltd.

[26] Chan, Y. T. and Langford, R. (1980). *Spectral estimation via the higher order Yule Walker equations.* IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-30, 689-698.

[27] Cleveland, W.P. and Tiao, G.C. (1976). *Decomposition of seasonal time series: a model for the Census X-11 program.* Journal of the American Statistical Association, 71, 581-587.

[28] Cline, D.B.H. and Brockwell, P.J. (1985). *Linear Prediction of ARMA processes With Infinite Variance.* Stochastic processes and their Applications 19, 281-296

[29] Cline, D.B.H. (1989). *Consistency for least squares regression estimators with infinite variance data.* Journal of Statistical Planning and Inference, 23, 163-179.

[30] Davila, C.E (1998). *A Subspace Approach to Estimation of Autoregressive Parameters from Noisy Measurements.* IEEE Transactions on Signal Processing, 46 (2), 531-534.

[31] Davis, R. A. and Resnick, S. (1986). *Limit theory for the sample covariance and correlation functions of moving averages.* Annals of Statistics, 14, 533-558.

[32] Durbin, J. (1960). *The Fitting of Time-Series Models.* Revue de l'Institut International de Statistique / Review of the International Statistical Institute, 28 (3), 233-244.

[33] Durbin, J. and Koopman, J. (2001). *Time series analysis by state space methods.* Oxford University Press.

[34] Fama, E. F. (1965). *The behavior of stock-market prices.* Journal of business, UChicago Press, 38, 34-105.

[35] Gallagher, C. (2000). *Estimating the auto-covariation from stationary heavy-tailed data.* with applications to time series modeling, Department of Mathematical Sciences Technical Report, Gal2. Clemson University.

[36] Gallagher, C. (2001). *A method for fitting stable autoregressive models using the auto-covariation function.* Statistics and Probability Letters, 53, 381 - 390.

[37] Gallagher, C. and Okuyama, T. (2003). *Some differences in the rates of convergence of the sample covariance and covariation functions.* www.interstat.statjournals.net.

[38] Gallagher, C. (2006). *Detecting Dependence in Heavy-Tailed Time Series using Portmanteau-type dependence tests.* International Mathematical Forum, 1 (10), 455-469.

[39] Gingras, D. (1982). *Estimation of autoregressive parameter from observations of a noise-corrupted autoregressive time series.* In Proc. ICASSP. Paris, France.

[40] Gordon, N., Percival, J. and Robinson, M. (2003). *The Kalman-Levy filter and heavy-tailed models for tracking manoeuvring targets.* In Proceedings of the Sixth International Conference of Information Fusion, 2, 1024-1031.

[41] Hannan, E. J. (1970). *Multiple Time Series,* New York: Wiley.

[42] Hannan, E. J. and Kanter, M. (1977). *Autoregressive Processes with Infinite Variance.* Journal of Applied Probability, 14 (2),411-41.

[43] Hill, J. B. (2003). *Strong Orthogonal Decompositions and Non-Linear Impulse Response Functions for Infinite Variance Processes.* Canadian Journal of Statistics, 34 (3), 453 - 473.

[44] Kalman, R.E. (1960). *A new approach to linear filtering and prediction problems.* Transactions ASME Journal of Basic Engineering, Series D, 82, 35-45.

[45] Kalman, R.E. and Bucy, R.S. (1961). *New results in linear filtering and prediction theory.* Transactions ASME Journal of Basic Engineering, Series D, 83, 95-107.

[46] Kanter, M. and Steiger, W.L. (1974). *Regression and autoregression with infinite variance.* Advances in Applied Probability, 6, 768-783.

[47] Kelley, C. T. (1999). *Iterative Methods for Optimization.* SIAM Frontiers in Applied Mathematics.

[48] Kidmose, P. (2000). *Alpha-Stable Distributions in Signal Processing of Audio Signals.* www.ei.dtu.dk/staff/kidmose/documents/kidmose00-2.pdf.

[49] Kluppelberg, C. and Mikosch, T. (1993). *Spectral estimates and stable processes.* Stochastic Processes and their Applications, 47, 323-344.

[50] Kluppelberg, C. and Mikosch, T. (1994). *Some limit theory for the self-normalised periodogram of stable processes.* Scandinavian Journal of Statistics, 21, 485-491.

[51] Knight, K. (1989). *Consistency of Akaikes Information Criterion for Infinite Variance Autoregressive Processes.* The Annals of Statistics, 17 (2), 824-840.

[52] Kohn, R. and Ansley, C.F. (1987). *Signal extraction for finite nonstationary time series*, Biometrika. 74 (2), 411-21.

[53] Kolmogorov, A.N., (1941). *Interpolation and Extrapolation.* Bulletin of the Academy of Sciences of the USSR, Mathematical Series, 5, 3-14.

[54] Koopman, J., Harvey A., (2003). *Computing observation weights for signal extraction and filtering.* Journal of Economic Dynamics and Control, 27 (7), 1317-1333.

[55] Kreyszig, E. (1989). *Introductory functional analysis with applications.* John Wiley and Sons. Inc.

[56] Kundu, D. (2002). *Estimating the number of signals in the presence of white noise.* Journal of Statistical Planning and Inference, 90, 57-68

[57] Kundu, D. and Nandi, S. (2005). *Estimating the Number of Components of the Fundamental Frequency Model.* Journal of Japan Statististical Society, 35 (1), 41-59

[58] Lagha, M., and Bensebti, M., (2007). *Weather Radar signals with Alpha Stable Distributions.* http:// hal.archives-ouvertes.fr/docs/00/19/76/17/PDF/Weather-Radar-Signals-with-Alpha-Stable-Distributions.pdf.

[59] Levinson, N. (1947). *The Wiener (root mean square) error criterion in filter design and prediction.* Journal of Mathematical Physics, 25, 262-278.

[60] Liu, T.H. and Mendel, J. M. (2001). *A subspace-based direction finding algorithm using fractional lower order statistics.* IEEE Transactions on Signal Processing. 49 (8), 1605-1613.

[61] McCulloch, J.H. (1997). *Measuring Tail Thickness to Estimate the Stable Index Alpha: A Critique.* Journal of Business and Economic Statisics, 15, 74-81.

[62] McCulloch, J.H. (1998). *Linear regression with stable disturbances.* In: Adler, R.J., Feldman, R.E., Taqqu, M.S. (Eds.), A Practical Guide to Heavy Tails. BirkhTauser, Boston, 359-376.

[63] McCulloch, J.H. and Bidarkota, P. (2003). *Signal Extraction can Generate Volatility Clusters.* http://www.econ.ohio-state.edu/jhm/gvc1hv.pdf

[64] Meinhold, R. J. and Singpurwalla, N.D. (1983). *Understanding the Kalman filter.* The American Statistician, 37 (2), 123-127.

[65] Mikosch, T., Gadrich, T., Klppelberg, C. and Adler, R. J. (1995). *Parameter estimation for ARMA models with infinite variance innovations*, The Annals of Statistics, 23 (1), 305-326.

[66] Nandi, S., Iyer, S. K. and Kundu, D. (2002). *Estimation of frequencies in presence of heavy tail errors.* Statistics and Probability Letters 58, 265-282.

[67] Nikias, C.L. and Shao, M. (1995). *Signal Processing With -stable Distributions and Applications.* John Wiley and Sons.

[68] Nolan, J. P. (1999). *Fitting data and assessing goodness of fit with stable distributions.* In Proceedings of the Conference on Applications of Heavy Tailed Distributions in Economics Engineering and Statistics, American University, Washington, DC.

[69] Nolan, J. (2005). http://academic2.american.edu/ jpnolan/stable/stable.exe.

[70] Pictel, V. O., Dacorogna, M.M. and Muller, A. U. (1998). *Hill, Bootstrap and Jackknife Estimators for Heavy tails.* A Practical Guide to Heavy Tails. BirkhTauser, Boston, 283-310.

[71] Pierce, R. D.(1997). *Application of the positive alpha stable distribution.* IEEE Signal Processing Workshop on Higher Order Statistics.

[72] Politis, D. N. (2009). *Robust fitting of autoregressive models.* Economics Letters, 102 (2), 128-131.

[73] Pollock, D.S.G. (2005). *Introduction to the special issue on statistical signal extraction and filtering.* Computational Statistics and Data Analysis, 50 (9), 2137-2145.

[74] Qiou, Z. and Ravisankar, N. (1998). *Bayesian Inference for Time Series with Stable Innovations.* Journal of Time Sereries Analysis, 19 (2), 235-249.

[75] Quinn, B. G. (1989). *Estimating the number of terms in a sinusoidal regression.* Journal of Time Sereries Analysis, 10, 71-75.

[76] Quinn, B. G. and Thomson, P. J. (1991). *Estimating the frequency of a periodic function.* Biometrika, 78 (1), 65-74.

[77] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications.* New York: Wiley.

[78] Resnick, S.I. (1997). *Heavy tail modeling and teletraffic data.* The Annals of Statistics, 25 (5), 1805-1869.

[79] Sakai, H. (1990). *An application of a BIC-type method to harmonic analysis and a new criterion for order determination of an AR process.* IEEE Transactions on Acoustics, Speech and Signal Processing, 38, 999-1004.

[80] Sakai, H. (1993). *The determination of the number of terms in a multichannel sinusoidal regression.* Journal of Time Sereries Analysis, 14, 621-628.

[81] Samorodnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes.* Chapman-hall, New York.

[82] Shao, M. and Nikias, C.L. (1993). *Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications.* Proceedings of the IEEE, 81 (7), 986-1010 .

[83] Shibata, R. (1976). *Selection of the order of an autoregressive model by Akaike's information criteria.* Biometrica, 63, 117-126.

[84] Shumway, R.H. and Stoffer D.S. (2006). *Time Series Analysis and Its Applications With R Examples.* Springer, USA.

[85] Singer, I. (1970). *Best Approximation in Normed Linear Space by Elements of Linear Subspaces.* Springer, New York.

[86] Sobel, E. (1967). *Prediction of a noise-distorted, multivariate, non-stationary signal.* Journal of Applied Probability, 4, 330-342.

[87] Sornette, D. and Ide, K. (2001). *The Kalman Levy Filter.* Physica D, 151, 142-174.

[88] Stewart, G.W. (1973). *Introduction to Matrix Computations.* Academic, New York

[89] Stocia P., Soderstrom T. and Funan T. (1989). *Asymptotic properties of the Higher Order Yule-Walker Estimates of Sinusoidal Frequencies.* IEEE Transactions of Acoustics Speech and Signal Processing, 37 (11), 1721-1734.

[90] Stuck, B. W. and Kleiner, B. (1974). *A statistical analysis of telephone noise.* Bell System Technical Journal, 53 (7), 1263-1320.

[91] Tanizaki, H. (2001). *Estimation of unknown parameters in nonlinear and non-Gaussian state-space models.* Journal of Statistical Planning and Inference, 96, (2001), 301-323.

[92] Tsakalides, P. and Nikias, C.L (1998). *Deviation from Normality in Statitical Signal Processing: Parameter Estimation with Alpha Stable Distributions.* A Practical Guide to Heavy Tails. BirkhTauser, Boston.

[93] Tsakalides, P., Raspanti, R. and Nikias, C. L. (2001). *Angle/doppler estimation in heavy-tailed clutter backgrounds.* IEEE Transactions on Aerospace and Electronic Systems, 35 (2), 419-436.

[94] Wang, H. (1993). *An AIC type estimator for the number of cosinusoids.* Journal of Time Sereries Analysis, 14, 433-440.

[95] Whittle, P. (1983). *Prediction and Regulation by Linear Least Squire Methods.* Second revised ed. Basil Blackwell, Oxford.

[96] Wiener, N. (1949). *Extrapolation, interpolation and smoothing of stationary time series.* Report on the Service Research Project DIC - 6037. MIT Technology press, Wiley, New York.

[97] Willinger, W. Paxson, V. and Taqqu, M.S. (1998). *Self-similarity and heavy tails: Structural modeling of network traffic.* A Practical Guide to Heavy Tails. BirkhTauser, Boston.

[98] Zolotarev, V. (1986). *One-Dimensional Stable Distributions.* Providence, RI: American Mathematical Society.