Analysis of Some Linear and Nonlinear Time Series Models

Ponnuthurai Ainkaran

A thesis submitted in fulfillment of the requirements
for the degree of Master of Science

School of Mathematics and Statistics
University of Sydney

August 2004

Abstract

This thesis considers some linear and nonlinear time series models. In the linear case, the analysis of a large number of short time series generated by a first order autoregressive type model is considered. The conditional and exact maximum likelihood procedures are developed to estimate parameters. Simulation results are presented and compare the bias and the mean square errors of the parameter estimates. In Chapter 3, five important nonlinear models are considered and their time series properties are discussed. The estimating function approach for nonlinear models is developed in detail in Chapter 4 and examples are added to illustrate the theory. A simulation study is carried out to examine the finite sample behavior of these proposed estimates based on the estimating functions.

Acknowledgements

I wish to express my gratitude to my supervisor Dr. Shelton Peiris for guiding me in this field and for his inspiration , encouragement, constant guidance, unfailing politeness and kindness throughout this master's programme, as well as his great generosity with his time when it came to discussing issues involved in this work. There are also many others whom I list below who deserve my fullest appreciation.

I would like to take this opportunity to thank Professor John Robinson and Dr Marc Raimondo for their guidance in this research during the period in which my supervisor was away from Sydney in 2002.

Associate Professor Robert Mellor of the University of Western Sydney gave me invaluable guidance towards my research on linear time series which appears in Chapter2 and his help is greatly appreciated.

I am also indebted to Professor A. Thavaneswaran of University of Manitoba, Canada, for giving me the opportunity of working with him on the nonlinear time series described in Chapter 4 of this thesis during his visits to Sydney University.

I owe sincere thanks to the statistics research group for their many helpful discussions and suggestions leading to the improvement of the quality of my research.

The Head and staff of the School of Mathematics and Statistics gave their constant support during my master's programme, for which I wish to express my gratitude.

I also appreciate the valuable suggestions, comments and proof-reading provided by my friends Chitta Mylvaganam and Jeevanantham Rajeswaran which contributed towards improving the literary quality of this thesis.

I am also indebted to my parents, teachers and all my friends who contributed to my personal and academic development up to this stage.

Last but not least in importance to me are my wife Kema and two children (Sivaram & Suvedini) without whose understanding, love and moral support, I could not have completed this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Notation and Definitions

Time Series Analysis is an important technique used in many observational disciplines, such as physics, engineering, finance, economics, meteorology, biology, medicine, hydrology, oceanography and geomorphology. This technique is mainly used to infer properties of a system by the analysis of a measured time record (data). This is done by fitting a representative model to the data with an aim of discovering the underlying structure as closely as possible. Traditional time series analysis is based on assumptions of linearity and stationarity. However, there has been a growing interest in studying nonlinear and nonstationary time series models in many practical problems. The first and the simplest reason for this is that many real world problems do not satisfy the assumptions of linearity and/or stationarity. For example, the financial markets are one of the areas where there is a greater need to explain behaviours that are far from being even approximately linear. Therefore, the need for the further development of the theory and applications for nonlinear models is essential.

In general time series analysis, it is known that there are a large number of nonlinear features such as cycles, asymmetries, bursts, jumps, chaos, thresholds, heteroscedasticity and mixtures of these have to be taken into account. A problem arises directly from a suitable definition of the nonlinear model because not every model is linear. This class clearly encompasses a large number of possible choices.

Furthermore, forecasting the future values of an observed time series is an important phenomenon for many real world problems. It provides a good basis for production planning and technical decisions. Forecasting means extrapolating the observations available up to time t to predict observations at future times. Forecasting methods are mainly classified into qualitative and quantitative techniques, which are based on unscientific and mathematical and/or statistical models respectively. The quantitative techniques are more important than qualitative techniques for future planning.
In this thesis, we consider some linear and nonlinear time series models and discuss various extensions and methods of parameter estimation. The estimating function approach, in particular, is considered in detail. A simulation study is carried out to verify the finite sample properties of the proposed estimates.

Below we give some basic definitions in time series that will be used in later chapters.

Definition 1.1. A time series is a set of observations on X_t, each being recorded at a specific time t, $t \in (0, \infty)$.

Notation 1.1. A discrete time series is represented as $\{X_t : t \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the set of integers (index set).

Let Ω be a sample space and $\mathcal{F}$ be the class of subsets of Ω.

Definition 1.2. If $\mathcal{F}$ satisfies the following conditions then it is called a σ-algebra (or σ-Field):

(i) $\Omega \in \mathcal{F}$

(ii) If $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$

(iii) If a finite or infinite sequence $\{A_i\} \in \mathcal{F}$ then $\bigcup_i A_i \in \mathcal{F}$.

Suppose that $\mathcal{P}$ is a real valued function, $\mathcal{P} : \Omega \longrightarrow [0, 1]$ (given a σ-algebra $\mathcal{F}$ of subsets of Ω) satisfying

(i) $\mathcal{P}$ (Ω)=1

(ii) for $A \in \mathcal{F}$, $\mathcal{P}$(A)$\geq$0

(iii) for a finite or infinite sequence

$\{A_i\} \in \mathcal{F}$ such that $A_i \bigcap A_j = \emptyset$, $i \neq j$, $\mathcal{P}$ $(\bigcup_i A_i) = \sum_i \mathcal{P}$ (A_i).

A function $\mathcal{P}$ satisfies the above conditions is called a probability measure on σ-algebra $\mathcal{F}$ of the sample space Ω. The ordered triple $(\Omega, \mathcal{F}, \mathcal{P})$ is called a probability space.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let $\mathrm{L}^2(\Omega, \mathcal{F}, \mathcal{P})$ or simply $\mathrm{L}^2(.)$ be the space of all random variables with finite first and second order moments. Then it can be seen that $\mathrm{L}^2(.)$ is a Hilbert space, if the inner product and the *norm* are defined by

$$\langle \xi, \mu \rangle = E(\xi\mu) \ and \ \|\xi\|^2 = E(\xi^2); \ \xi, \mu \in \mathrm{L}^2(.)$$

respectively (here we restrict our approach to the real valued random variables only). In other words, $\mathrm{L}^2(.)$ is a complete inner product space. That is, an inner product $\langle \xi, \mu \rangle$, defined, which assigns scalar values to a pair of vectors $\xi, \mu \in \mathrm{L}^2(.)$ satisfies the following:

(i) $\langle \alpha\xi + \beta\mu, \theta \rangle = \alpha\langle \xi, \theta \rangle + \beta\langle \mu, \theta \rangle$,

where $\xi, \mu, \theta \in \mathrm{L}^2(.)$ and α, β are scalars.

(*ii*) $\langle \xi, \xi \rangle \geq 0$ with equality if and only if $\xi = 0$.

Completeness means that every Cauchy sequence converges in the same space. The Hilbert space concept is useful in the development of time series theory and will be considered in later chapters.

Let $\{X_t; t \in T\}$ denote a stochastic process, where T is the index set. If T is a discrete set then the process is *discrete* and in this work, we consider only discrete processes. Clearly, a discrete time series is a discrete stochastic process. A stochastic process is determined by the set of finite dimensional probability distributions,

$$F(x_1, \cdots, x_n; t_1, \cdots, t_n) = Pr\{X_{t_1} \leq x_1; \cdots; X_{t_n} \leq x_n\}$$

for any arbitrary set of values $(t_1, t_2, \cdots, t_n)$. The finite probability distributions completely determine the probability structure of a stochastic process (Kolmogorov,1933).

Definition 1.3. A stochastic process (discrete) is said to be strictly stationary if the probability distribution of the process is invariant under translation of the index, i.e., the joint probability distributions of $(X_{t_1}, \cdots, X_{t_n})$ is identical to that of $(X_{t_1+k}, \cdots, X_{t_n+k})$, for all $n \in \mathbb{Z}^+$, $(t_1, \cdots, t_n) \in T$, $k \in \mathbb{Z}$. i.e.

$$F(x_1, \cdots, x_n; t_1, \cdots, t_n) = F(x_1, \cdots, x_n; t_1 + k, \cdots, t_n + k) \quad (1.1)$$

for all $n \in \mathbb{Z}^+$, $(t_1, \cdots, t_n) \in T$ and any $k \in \mathbb{Z}$.

Definition 1.4. A stochastic process $\{X_t\}$ is said to be Gaussian process if and only if the probability distribution associated with any set of time points is multivariate normal.

In particular, if the multivariate moments $E(X_{t_1}^{s_1} \cdots X_{t_n}^{s_n})$ depends only on the time differences, the process is called stationary of order s, where $s = s_1 + \cdots + s_n$.

Note that, the second order stationarity is obtained by setting $s = 2$ and this weak stationarity asserts that the mean μ is a constant (i.e. independent of t) and the covariance function $\gamma_{t\tau}$ is dependent only on the time difference. That is,

$$E(X_t) = \mu, \text{ for all } t$$

and

$$Cov(X_t, X_\tau) = E[(X_t - \mu)(X_\tau - \mu)] = \gamma_{|t-\tau|}, \text{ for all } t, \tau.$$

This is denoted as $\gamma(h)$, where the time difference $h = |t - \tau|$ is called the *lag*.

Definition 1.5. The autocorrelation function (acf) of a stationary process $\{X_t\}$ is a function whose value at lag h is

$$\rho(h) = \gamma(h)/\gamma(0) = Corr(X_t, X_{t+h}), \textit{ for all } t, h \in \mathbb{Z}, \tag{1.2}$$

where $\gamma(h)$ is the autocovariance function (acvf) of the process at lag h.

Most of the probability theory of time series is concerned with stationary time series.The important part of the analysis of time series is the selection of a suitable probability model for the data. The simplest kind of time series $\{X_t\}$ is the one in which the random variables $X_t, t = 0, \pm 1, \pm 2, \cdots$, are uncorrelated and identically distributed with zero mean and variance σ^2. Ignoring all properties of the joint distribution of $\{X_t\}$ except those which can be deduced from the moments $E(X_t)$ and $E(X_t X_{t+h})$, such processes having mean 0 and autocovariance function

$$\gamma(h) = \begin{cases} \sigma^2, \text{ if } h = 0, \\ 0, \text{ if } h \neq 0. \end{cases} \tag{1.3}$$

Definition 1.6. The process $\{\varepsilon_t\}$ is said to be white noise with mean 0 and variance σ^2, written

$$\{\varepsilon_t\} \sim WN(0, \sigma^2) \tag{1.4}$$

if and only if $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables with zero mean and covariance function in (1.3).

If the random variables ε_t are independently and identically distributed (iid) with mean 0 and variance σ^2 then we shall write

$$\{\varepsilon_t\} \sim IID(0, \sigma^2). \tag{1.5}$$

A wide class of stationary time series models can be generated by using white noise as forcing terms in a set of linear difference equations.

This leads to the notion of an autoregressive-moving average (ARMA) process, which plays a main role in time series analysis. This will be described in detail in Chapter 2.

Below we briefly define the likelihood criterion.

Definition 1.7. Let $\{X_t\}$ be a Gaussian process with mean μ. The corresponding exact log-likelihood function is defined as the function ,

$$-2L = n \log 2\pi + \log|\Gamma| + (\boldsymbol{X} - \mu\,\boldsymbol{V})'\Gamma^{-1}(\boldsymbol{X} - \mu\,\boldsymbol{V}), \tag{1.6}$$

where $\mathbf{X}' = (X_1, \cdots, X_n)$, $\mathbf{V}$ is an $n \times 1$ column vector of 1's and Γ is the covariance matrix of $\mathbf{X}$ (Γ is a function of unknown parameters). Minimization of the right hand side of (1.6) with respect to parameters lead to maximum likelihood estimates of the parameters.

We now state some results based on the Kalman filtering algorithm which will be used in later chapters.

1.2. Recursive estimation methods and Kalman filtering

Recursive estimators are aimed at tracking time varying parameters. It is desirable to make the calculations recursively to save computation time. In this section we describe some standard recursive algorithms. In a recursive estimation, the parameter estimate Φ_t (a vector of parameter at time t) can be obtained as a function of the previous estimate Φ_{t-1} and of the current (or new) measurements.

1.2.1. Kalman filtering. The Kalman filter is a recursive algorithm derived by Kalman (1960) to provide estimates of parameters in the state space model of a time series or a linear dynamic system disturbed by the Gaussian white noise. There are several state-space models which were proposed by several authors, and each consists of two equations for a process $\{\mathbf{Y_t}\}$. Suppose that an observed vector series $\{\mathbf{Y_t}\}$ can be written in terms of an observed state vector $\{\mathbf{X_t}\}$ (of dimension v). This first equation is known as the *observation equation* and is given by

$$\mathbf{Y_t} = \mathbf{G_t X_t} + \mathbf{W_t} \ , \ t = 1, 2, \cdots , \tag{1.7}$$

where $\{\mathbf{W_t}\} \sim WN(0, R_t)$ and$\{\mathbf{G_t}\}$ is a sequence of $w \times v$ matrices. The second equation known as state equation determines evolution of the state $\mathbf{X_t}$ at time 't' in terms of the previous state $\mathbf{X_{t-1}}$ and the noise term $\mathbf{V_t}$. This state equation is given by

$$\mathbf{X_t} = \mathbf{H_t X_{t-1}} + \mathbf{V_t} \ , \ t = 2, 3, \cdots , \tag{1.8}$$

where $\mathbf{H_t}$ is a sequence of $v \times v$ matrices, $\{\mathbf{V_t}\} \sim WN(0, Q_t)$ and $\{\mathbf{V_t}\}$ is uncorrelated with $\{\mathbf{W_t}\}$. It is assumed that the initial state $\mathbf{X_1}$ is uncorrelated with all the noise terms. In general,

$$\mathbf{X_t} = f_t(\mathbf{X_1}, \mathbf{V_1}, \cdots, \mathbf{V_{t-1}})$$

and

$$\mathbf{Y_t} = g_t(\mathbf{X_1}, \mathbf{V_1}, \cdots, \mathbf{V_{t-1}}, \mathbf{W_t}).$$

Note that the state-space representation given by (1.7) and (1.8) is not unique. This state-space representation is important for developing some important results in both linear and nonlinear time series models.

In Chapter 2, we discuss a class of linear time series driven by Autoregressive Integrated Moving Average (ARIMA) models and discuss recent contributions to the literature.

CHAPTER 2

Linear Time Series Models

In this chapter we consider an important parametric family of stationary time series, the autoregressive moving average (ARMA) process which plays a key role in modelling of time series data. We first summarize the key properties of ARMA(p,q) processes.

2.1. Autoregressive Moving Average Process

Definition 2.1. The process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is said to be an ARMA(p,q) *process if* $\{X_t\}$ *is stationary and if for every t,*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \qquad (2.1)$$

where $\{\varepsilon_t\} \sim WN(0, \sigma^2)$.

We say that $\{X_t\}$ is an ARMA(p,q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p,q) process.

The equation (2.1) can be written symbolically in a compact form

$$\phi(B) X_t = \theta(B) \varepsilon_t, \; t = 0, \pm 1, \pm 2, \cdots, \qquad (2.2)$$

where B is the back-shift operator defined by $B^j X_t = X_{t-j}$, $j \geq 0$ with $B^0 = I$ (identity operator) and $\phi(\cdot)$ and $\theta(\cdot)$ are the p^{th} and q^{th} degree polynomials in B such that

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \qquad (2.3)$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q. \tag{2.4}$$

The polynomials $\phi(\cdot)$ and $\theta(\cdot)$ are called *autoregressive* and *moving average* polynomials respectively of the difference equation (2.1). If $\phi(z) \equiv 1$ (i.e. p=0) then

$$X_t = \theta(B)\varepsilon_t,\ t = 0, \pm 1, \pm 2, \cdots, \tag{2.5}$$

and the process is said to be a *moving-average process of order q* (or MA(q)).
If $\theta(z) \equiv 1$ (i.e. q=0) then

$$\phi(B)X_t = \varepsilon_t,\ t = 0, \pm 1, \pm 2, \cdots, \tag{2.6}$$

and the resulting process is said to be an *autoregressive process of order p* (or AR(p)).

Note : The process in (2.2) is stationary and invertible when the roots of $\phi(z) = 0$ and $\theta(z) = 0$ are outside the unit circle.

For an ARMA(p,q) process in (2.2), the equivalent autoregressive representation of $\{X_t\}$ is

$$\pi(B)X_t = \varepsilon_t,$$

where $\pi(z) = \theta^{-1}(z)\phi(z)$.
Since $\pi(z)\theta(z) = \phi(z)$, we have the infinite AR polynomial $\pi(z) =$

$1 - \sum_{i=1}^{\infty} \pi_i z^i$ by recursively substituting

$$\begin{aligned} \pi_1 &= \theta_1 + \phi_1 \\ \pi_2 &= \theta_2 - \phi_2 + \theta_1^2 + \theta_1 \phi_1 \qquad (2.7) \\ &\vdots \qquad (2.8) \end{aligned}$$

Also any stationary ARMA(p,q) model in (2.2) can be written as

$$X_t = \psi(B)\varepsilon_t,$$

where $\psi(z) = \phi^{-1}(z)\theta(z)$. The infinite MA polynomial $\psi(z) = 1 + \sum_{i=1}^{\infty} \psi_i z^i$ and formulate it by recursively substituting on $\psi(z)\phi(z) = \theta(z)$,
i.e.

$$\begin{aligned} \psi_1 &= \phi_1 + \theta_1 \\ \psi_2 &= \phi_2 + \theta_2 + \phi_1(\phi_1 + \theta_1) \qquad (2.9) \\ &\vdots \end{aligned}$$

Example 2.1. Consider an ARMA(1,1) process with zero mean given by

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}.$$

Write the infinite AR representation of the above as

$$\pi(B)X_t = \varepsilon_t,$$

where the infinite AR polynomial $\pi(z) = 1 - \sum_{i=1}^{\infty} \pi_i z^i$ satisfies the equation

$$\theta(z)\pi(z) = \phi(z).$$

Now it is easy to see that

$$\begin{aligned}\pi_1 &= \theta+\phi \\ \pi_i &= (\theta+\phi)(-\theta)^{i-1},\ i=2,3,\cdots.\end{aligned}$$

Similarly we can write the infinite MA polynomial $\psi(z)$ with the following coefficients

$$\begin{aligned}\psi_1 &= \theta+\phi \\ \psi_i &= (\theta+\phi)(\phi)^{i-1},\ i=2,3,\cdots.\end{aligned}$$

The assumptions of stationarity in many applications of time series is too restrictive and this led us to consider developments in non-stationary time series. The non-stationarity of a process may be due to the changes in the mean and/or variance with time. Certain classes of time series are characterized by a behaviour in which, apart from a local level and/or a local trend, one part of the series behaves like the others. Such series are called homogeneous non-stationary (see Abraham and Ledolter (1993),p.225). Homogeneous non-stationary sequences can be transformed into stationary sequences by successive differencing (see Box and Jenkins(1976)); i.e., by considering $\nabla X_t, \nabla^2 X_t, \cdots$, where $\nabla = I - B$ is the difference operator. A generalization of the class (2.1) incorporating the above type of non-stationarity is given by the class of autoregressive integrated moving average (ARIMA) processes and is defined as follows:

Definition 2.2. If d is a non-negative integer, then $\{X_t\}$ is said to be an ARIMA(p,d,q) process if

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t, \{\varepsilon_t\} \sim WN(0, \sigma^2), \tag{2.10}$$

where $\phi(.)$ and $\theta(.)$ are stationary AR and invertible MA polynomials of degrees p and q respectively.

If d=0 , then the process is stationary and reduces to an ARMA(p,q) process in (2.2).

We now consider the State-space representation and Kalman filtering of linear time series. These results will be used in later chapters.

2.2. State-space representation and Kalman filtering of ARMA models

Kalman filtering and recursive estimation has important application in time series. In this approach the time series model needs to be rewritten in a suitable state-space form. Note that this state-space representation of a time series is not unique.
A state-space representation for an ARMA process is given below:

Example 2.2. Let $\{Y_t\}$ be a causal ARMA(p,q) process satisfying $\phi(B)Y_t = \theta(B)\varepsilon_t$.

Let

$$Y_t = \theta(B)X_t \tag{2.11}$$

and

$$\phi(B)X_t = \varepsilon_t. \tag{2.12}$$

Equations (2.11) and (2.12) are called the observation equation and state equation respectively. These two equations are equivalent to the original ARMA(p,q) process since

$$\begin{aligned}\phi(B)Y_t &= \phi(B)\theta(B)X_t \\ &= \theta(B)\phi(B)X_t \\ &= \theta(B)\varepsilon_t.\end{aligned}$$

Therefore an ARMA(p,q) process can be represented in a state-space form as given in (2.11) and (2.12).

Now consider a general state-space representation of process

$$\text{State}: \ \phi(B)X_t = \sigma v_t \tag{2.13}$$

$$\text{Observation}: \ Y_t = \theta(B)X_t + w_t, \tag{2.14}$$

where the state $\{X_t\}$ is a real process , $\{v_t\}$ and $\{w_t\}$ are 0 mean white noise processes with time dependent variances $Var(v_t) = Q_t$ and $Var(w_t) = R_t$ and covariance $Cov(v_t, w_t) = S_t$.

It is assumed that the past values of the states and observations are uncorrelated with the present errors. To make the derivation simpler, assume that the two noise processes are uncorrelated (i.e., $S_t = 0$) and constant variances ($Q_t = Q$ and $R_t = R$).
Let $\hat{X}_{t|t}$ be the prediction of X_t given $Y_t, \cdots, Y_1$. i.e.

$$\hat{X}_{t|t} = E[\mathbf{X}_t|\mathcal{F}^y_{t-1}],$$

where $\mathcal{F}^y_{t-1}$ is the σ-algebra generated by $Y_t, \cdots, Y_1$. Let $d = max(p, q+1)$ and write $\mathbf{X_t} = (X_t, X_{t-1}, \cdots, X_{t-d+1})'$ as

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{V}_t, \tag{2.15}$$

where we pad ϕ or θ with zeros as needed.

Now the observation process in a vector form is

$$\mathbf{Y}_t = \mathbf{\Theta}\mathbf{X}_t + \mathbf{W}_t, \tag{2.16}$$

where

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

and

$$\Theta = \begin{pmatrix} 1 & \theta_1 & \ldots & \theta_d \end{pmatrix}.$$

Then the one step ahead estimates of $\hat{\mathbf{X}}_{t|t-1}$ are

$$\begin{aligned} \hat{\mathbf{X}}_{t|t-1} &= E[\mathbf{X}_t|\mathcal{F}^y_{t-1}] \\ &= E[(\mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{V}_t)|\mathcal{F}^y_{t-1}] \\ &= E[\Phi\mathbf{X}_{t-1}|\mathcal{F}^y_{t-1}] \\ &= \Phi\hat{\mathbf{X}}_{t-1|t-1} \end{aligned} \tag{2.17}$$

and

$$\begin{aligned} \mathbf{P}_{t|t-1} &= Cov(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t-1}) \\ &= Cov(\mathbf{\Phi}\mathbf{X}_{t-1} + \mathbf{V}_t - \mathbf{\Phi}\hat{\mathbf{X}}_{t-1|t-1}) \\ &= \mathbf{\Phi} Cov(\mathbf{X}_{t-1} - \hat{\mathbf{X}}_{t-1|t-1})\mathbf{\Phi}' + V(\mathbf{V}_t) \\ &= \mathbf{\Phi}\mathbf{P}_{t-1|t-1}\mathbf{\Phi}' + \mathbf{Q}, \end{aligned} \tag{2.18}$$

where $\mathbf{Q} = cov(\mathbf{V}_t)$ which is denoted as $V(\mathbf{V}_t)$ for convenience .
Equations (2.17) and (2.18) give the updated filtered estimates and

$$\hat{\mathbf{X}}_{t|t} = \hat{\mathbf{X}}_{t|t-1} + \kappa_t \mathbf{I}_t \tag{2.19}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} + \kappa_t \mathbf{\Theta} \mathbf{P}_{t|t-1}, \tag{2.20}$$

where $\mathbf{I}_t = \mathbf{Y}_t - \Theta\hat{\mathbf{X}}_{t|t-1}$ is known as the innovation at time t, and the Kalman gain is

$$\kappa_t = \mathbf{P}_{t|t-1}\mathbf{\Theta}'(\mathbf{\Theta}\mathbf{P}_{t|t-1}\mathbf{\Theta}' + \mathbf{R})^{-1}. \tag{2.21}$$

Derivations of (2.19) and (2.20) are as follows:
First split $\mathcal{F}_t^y$ into two orthogonal sub spaces $\tilde{\mathbf{Y}}_{t|t-1}$ and $\mathcal{F}_{t-1}^y$. Assuming that the mean of Y_t is zero,

$$\begin{aligned}
\hat{\mathbf{X}}_{t|t} &= E[\mathbf{X}_t|\mathcal{F}_t^y] \\
&= E[\mathbf{X}_t|\tilde{\mathbf{Y}}_{t|t-1}, \mathcal{F}_{t-1}^y] \\
&= E[\mathbf{X}_t|\tilde{\mathbf{Y}}_{t|t-1}] + E[\mathbf{X}_t|\mathcal{F}_{t-1}^y] \\
&= \kappa_t\tilde{\mathbf{Y}}_{t|t-1} + \hat{\mathbf{X}}_{t|t-1},
\end{aligned}$$

where $\tilde{\mathbf{Y}}_{t|t-1}$ is known as the innovation $\mathbf{I}_t$ at time t.
The orthogonality condition

$$Cov(\mathbf{X}_t - \kappa_t\tilde{\mathbf{Y}}_{t|t-1}, \tilde{\mathbf{Y}}_{t|t-1}) = E[(\mathbf{X}_t - \kappa_t\tilde{\mathbf{Y}}_{t|t-1})\tilde{\mathbf{Y}}'_{t|t-1}] = 0$$

implies

$$\kappa_t = Cov(\mathbf{X}_t, \kappa_t\tilde{\mathbf{Y}}_{t|t-1})V(\tilde{\mathbf{Y}}_{t|t-1})^{-1}.$$

Consider

$$
\begin{aligned}
\tilde{\mathbf{Y}}_{t|t-1} &= \mathbf{Y}_t - E[\mathbf{Y}_t|\mathcal{F}^y_{t-1}] \\
&= \mathbf{Y}_t - E[(\mathbf{\Theta}\mathbf{X}_t + \mathbf{W}_t)|\mathcal{F}^y_{t-1}] \\
&= \mathbf{Y}_t - \mathbf{\Theta}E[\mathbf{X}_t|\mathcal{F}^y_{t-1}] \\
&= \mathbf{\Theta}\mathbf{X}_t + \mathbf{W}_t - \mathbf{\Theta}\hat{\mathbf{X}}_{t|t-1} \\
&= \mathbf{\Theta}\tilde{\mathbf{X}}_{t|t-1} + \mathbf{W}_t,
\end{aligned}
$$

where $\tilde{\mathbf{X}}_{t|t-1} = \mathbf{X}_t - \hat{\mathbf{X}}_{t|t-1}$.

Splitting $\mathbf{X}_t$ into orthogonal parts,

$$
\begin{aligned}
\kappa_t &= Cov(\hat{\mathbf{X}}_{t|t-1} + \tilde{\mathbf{X}}_{t|t-1}, \mathbf{\Theta}\tilde{\mathbf{X}}_{t|t-1} + \mathbf{W}_t)V(\mathbf{\Theta}\tilde{\mathbf{X}}_{t|t-1} + \mathbf{W}_t)^{-1} \\
&= Cov(\tilde{\mathbf{X}}_{t|t-1}, \mathbf{\Theta}\tilde{\mathbf{X}}_{t|t-1})(\mathbf{\Theta}\mathbf{P}_{t|t-1}\mathbf{\Theta}' + \mathbf{R})^{-1} \\
&= \mathbf{P}_{t|t-1}\mathbf{\Theta}'(\mathbf{\Theta}\mathbf{P}_{t|t-1}\mathbf{\Theta}' + \mathbf{R})^{-1}.
\end{aligned}
$$

Therefore one has,

$$
\begin{aligned}
\mathbf{P}_{t|t} &= V(\tilde{\mathbf{X}}_{t|t}) \\
&= V(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t}) \\
&= V(\mathbf{X}_t - \hat{\mathbf{X}}_{t|t-1} - \kappa_t\tilde{\mathbf{Y}}_{t|t-1}) \\
&= V(\tilde{\mathbf{X}}_{t|t-1} - \kappa_t\tilde{\mathbf{Y}}_{t|t-1}) \\
&= V(\tilde{\mathbf{X}}_{t|t-1}) - cov(\tilde{\mathbf{X}}_{t|t-1}, \kappa_t\tilde{\mathbf{Y}}_{t|t-1}) - cov(\kappa_t\tilde{\mathbf{Y}}_{t|t-1}, \tilde{\mathbf{X}}_{t|t-1}) \\
&\quad + V(\kappa_t\tilde{\mathbf{Y}}_{t|t-1}) \\
&= \mathbf{P}_{t|t-1} - cov(\tilde{\mathbf{X}}_{t|t-1}, \kappa_t\mathbf{\Theta}\tilde{\mathbf{X}}_{t|t-1}) - cov(\kappa_t\mathbf{\Theta}\tilde{\mathbf{X}}_{t|t-1}, \tilde{\mathbf{X}}_{t|t-1}) \\
&\quad + \kappa_t V(\tilde{\mathbf{Y}}_{t|t-1})\kappa_t'
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbf{P}_{t|t-1} - v(\tilde{\mathbf{X}}_{t|t-1})\kappa_t'\mathbf{\Theta}' - \kappa_t\mathbf{\Theta}v(\tilde{\mathbf{X}}_{t|t-1}) + cov(\tilde{\mathbf{X}}_{t|t-1}, \tilde{\mathbf{Y}}_{t|t-1})\kappa_t' \\
&= \mathbf{P}_{t|t-1} - \kappa_t\mathbf{\Theta}\mathbf{P}_{t|t-1},
\end{aligned}
$$

where $\mathbf{P}_{t|t-1}$ is called the conditional variance of the one step ahead prediction error.

Example 2.3. The observations of a time series are $x_1, \cdots, x_n$,and the mean μ_n is estimated by $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$.

If a new point x_{n+1} is measured, we can update μ_n, but it is more efficient to use the old value of $\hat{\mu}_n$, and make a small correction using x_{n+1}. The correction is easy to derive, since

$$
\hat{\mu}_{n+1} = \frac{1}{n+1}\sum_{i=1}^{n+1} x_i = \frac{n}{n+1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}x_{n+1}\right),
$$

and so $\hat{\mu}_{n+1}$ can be written as $\hat{\mu}_{n+1} = \hat{\mu}_n + \kappa(x_{n+1} - \hat{\mu}_n)$, where $\kappa = \frac{1}{n+1}$ is called the gain factor.

Similarly, we can express the variance

$$
\hat{\sigma}_{n+1}^2 = (1-\kappa)\hat{\sigma}_n^2 + \kappa(1-\kappa)(x_{n+1} - \hat{\mu}_n)^2.
$$

Example 2.4. Consider MA(1) model

$$
Y_t = \varepsilon_t + \theta\varepsilon_{t-1}. \tag{2.22}
$$

We can write this model in state space form as

$$
\begin{aligned}
Y_t &= \begin{pmatrix} 1 & \theta \end{pmatrix}\mathbf{X}_t \\
&= \mathbf{\Theta}\mathbf{X}_t,
\end{aligned}
$$

where $\mathbf{X}_t = (\varepsilon_t\ \varepsilon_{t-1})'$ with

$$
\begin{aligned}
\mathbf{X}_t &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \mathbf{X}_{t-1} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \varepsilon_t \\
&= \mathbf{\Phi} \mathbf{X}_{t-1} + \mathbf{a}\varepsilon_t \ .
\end{aligned}
$$

From the equations (2.17)-(2.21), the Kalman gain is

$$
\kappa_t = \frac{\mathbf{P}_{t|t-1}\mathbf{\Theta}'}{R + \mathbf{\Theta}'\mathbf{P}_{t|t-1}\mathbf{\Theta}}
$$

and the recursion for the parameter estimates

$$
\begin{aligned}
\hat{\mathbf{X}}_{t|t-1} &= \mathbf{\Phi}\hat{\mathbf{X}}_{t-1|t-1} \qquad \text{and} \\
\hat{\mathbf{X}}_{t|t} &= \hat{\mathbf{X}}_{t|t-1} + \frac{\mathbf{P}_{t|t-1}\mathbf{\Theta}'}{\mathbf{R} + \mathbf{\Theta}'\mathbf{P}_{t|t-1}\mathbf{\Theta}} (\mathbf{Y}_t - \mathbf{\Theta}'\hat{\mathbf{X}}_{t|t-1}).
\end{aligned}
$$

The corresponding variances are given by the recursive equation

$$
\begin{aligned}
\mathbf{P}_{t|t-1} &= \mathbf{\Phi}\mathbf{P}_{t-1|t-1}\mathbf{\Phi}' + \mathbf{a}\mathbf{a}'\sigma^2 \\
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \frac{\mathbf{P}_{t|t-1}\mathbf{\Theta}\mathbf{\Theta}'\mathbf{P}_{t|t-1}}{\mathbf{R} + \mathbf{\Theta}'\mathbf{P}_{t|t-1}\mathbf{\Theta}} \ .
\end{aligned}
$$

Note : In linear stationary time series it is known that the conditional means, $E(X_t|X_{t-1}, X_{t-2}, \cdots) = E(X_t|\mathcal{F}_{t-1}^X)$ and the conditional variances, $Var(X_t|X_{t-1}, X_{t-2}, \cdots) = Var(X_t|\mathcal{F}_{t-1}^X)$ are constants, where $\mathcal{F}_{t-1}^X$ is the σ-field generated by $X_{t-1}, X_{t-2}, \cdots, X_1$.

Theory and applications of ARMA type time series models are well developed when n,the number of observations, is large. However, in many real world problems one observes short time series (n is small) with many replications. And, in general, for short time series one cannot rely on the usual procedures of estimation or asymptotic theory . The motivation and applications of this type can be found in

Cox and Solomon (1988), Rai et. al. (1995), Hjellvik and Tjϕstheim (1999), Peiris,Mellor and Ainkaran(2003) and Ainkaran, Peiris, and Mellor (2003).

In the next section, we consider the analysis of such short time series.

2.3. Analysis of short time series.

It is known that there are many situations, especially in medical research where one observes several very short time series. For example, to assess the durability of a mitral valve repair, repeated echocardiograms are obtained over a period of time and the ejection fraction is recorded. Hence, here, we have a large number of patients each with a shorter time series. In this situation, it is reasonable to suspect serial correlation among observations collected for a patient. In this case, it is reasonable to begin with a simple AR(1) process and extend the maximum likelihood estimation (mle) procedures (exact and conditional) to estimate ϕ as this parameter plays an important role in many practical situations.

Consider an AR(1) process with mean μ is given by

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t, \tag{2.23}$$

where $|\phi| < 1$ and $\mu = E(X_t)$ are constants. Assume that ϕ and μ remain unchanged for each series with auto-covariance function at lag k, $\gamma_k = \dfrac{\sigma^2\phi^k}{1-\phi^2}$. Let Γ be a symmetric $n \times n$ matrix given by

$$\Gamma = \frac{\sigma^2}{1-\phi^2}\begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{n-1} & \dots & \dots & \dots & 1 \end{pmatrix}. \tag{2.24}$$

We now describe estimation of the parameters of (2.23) using maximum-likelihood criteria. Assuming $\{\varepsilon_t\}$ is Gaussian white noise (i.e. iid $N(0, \sigma^2)$), the exact log-likelihood function of (2.23) based on n observations is

$$\begin{aligned} -2L = n\log(2\pi\sigma^2) - \log(1-\phi^2) + \{(X_1-\mu)^2 \\ + \sum_{t=2}^{n}[X_t - \mu - \phi(X_{t-1}-\mu)]^2\}/\sigma^2. \end{aligned} \tag{2.25}$$

To estimate the parameters ϕ, μ and σ^2, we need a suitable optimization algorithm to maximize the function (2.25).

The corresponding conditional log-likelihood function based on (2.23) can be written as

$$-2L = (n-1)\log(2\pi\sigma^2) + \sum_{t=2}^{n}[X_t - \mu - \phi(X_{t-1}-\mu)]^2/\sigma^2. \tag{2.26}$$

Maximizing the log likelihood, L in (2.26) with respect to the parameters ϕ, μ, and σ^2, we have

$$\hat{\phi} = \frac{\sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1} - (n-1)\sum_{t=2}^{n} X_t X_{t-1}}{(\sum_{t=2}^{n} X_{t-1})^2 - (n-1)\sum_{t=2}^{n} X_{t-1}^2}, \tag{2.27}$$

$$\hat{\mu} = \frac{\sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1}^2 - \sum_{t=2}^{n} X_{t-1} \sum_{t=2}^{n} X_t X_{t-1}}{((n-1)\sum_{t=2}^{n} X_{t-1}^2 - (\sum_{t=2}^{n} X_{t-1})^2)(1-\hat{\phi})}, \tag{2.28}$$

and

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^{n}(X_t - \hat{\mu} - \hat{\phi}(X_{t-1}-\mu))^2}{(n-1)}. \tag{2.29}$$

Note: $\hat{\mu}$ can be reduced to

$$\hat{\mu} = \frac{\sum_{t=2}^{n}(X_t - \hat{\phi}X_{t-1})}{(n-1)(1-\hat{\phi})}.$$

Now suppose that n is small and there are m independent replications on (2.23) satisfying

$$X_{it} - \mu_i = \phi_i(X_{i,t-1} - \mu_i) + \varepsilon_{it}. \tag{2.30}$$

In our analysis we assume that $\phi_i = \phi$ and $\mu_i = \mu$ for all $i = 1, \cdots, m$.

2.3.1. Conditional likelihood estimation. Let the series X_{it} be independent of the series X_{jt} for each $i \neq j$. Assuming $\{\varepsilon_{it}\}$ is a Gaussian white noise (ie. ε_{it} are iid $N(0, \sigma^2)$), the corresponding conditional log-likelihood function based on (2.30) can be re-written as

$$-2L = m(n-1)\log(2\pi\sigma^2) + \sum^{*}[X_{it} - \mu - \phi(X_{i,t-1} - \mu)]^2/\sigma^2, \tag{2.31}$$

where $\sum^* = \sum_{i=1}^{m}\sum_{t=2}^{n}$.

Maximizing the log-likelihood, (2.31) with respect to the parameters ϕ, μ, and σ^2, we have

$$\hat{\phi} = \frac{\sum^* X_{it}\sum^* X_{i,t-1} - m(n-1)\sum^* X_{it}X_{i,t-1}}{(\sum^* X_{i,t-1})^2 - m(n-1)\sum^* X_{i,t-1}^2}, \tag{2.32}$$

$$\hat{\mu} = \frac{\sum^* X_{it}\sum^* X_{i,t-1}^2 - \sum^* X_{i,t-1}\sum^* X_{it}X_{i,t-1}}{(m(n-1)\sum^* X_{i,t-1}^2 - (\sum^* X_{i,t-1})^2)(1-\hat{\phi})}, \tag{2.33}$$

and

$$\hat{\sigma}^2 = \frac{\sum^*(X_{it} - \hat{\mu} - \hat{\phi}(X_{i,t-1} - \mu))^2}{m(n-1)}. \tag{2.34}$$

Furthermore, $\hat{\mu}$ can be reduced to

$$\hat{\mu} = \frac{\sum^*(X_{it} - \hat{\phi}X_{i,t-1})}{m(n-1)(1-\hat{\phi})}.$$

Equations (2.32), (2.33) and (2.34) can be simplified using a vector notation as follows:

Define $\mathbf{V}_1$ as a row vector of 1's of order $1 \times m$ and $\mathbf{V}_2$ as a column vector of 1's of order $(n-1) \times 1$. Let $\mathbf{A}$ be the $m \times n$ matrix of all (X_{it})'s $(i = 1, \cdots m; t = 1, \cdots n)$ and let $\mathbf{A}_1$ and $\mathbf{A}_2$ be two matrices of order $m \times (n-1)$ excluding the first column and the last column of $\mathbf{A}$ respectively. Let $sum(\mathbf{A})$ be the sum of all elements of the matrix $\mathbf{A}$. Then, one has

$$\sum^{*} X_{it} = \mathbf{V}_1\mathbf{A}_1\mathbf{V}_2 = sum(\mathbf{A}_1),$$

and

$$\sum^{*} X_{i,t-1} = \mathbf{V}_1\mathbf{A}_2\mathbf{V}_2 = sum(\mathbf{A}_2).$$

Also,

$$\sum^{*} X_{it}X_{i,t-1} = \mathbf{V}_1(\mathbf{A}_1 * \mathbf{A}_2)\mathbf{V}_2 = sum(\mathbf{A}_1 * \mathbf{A}_2),$$

and

$$\sum^{*} X_{i,t-1}^2 = \mathbf{V}_1(\mathbf{A}_2 * \mathbf{A}_2)\mathbf{V}_2 = sum(\mathbf{A}_2^2),$$

where $\mathbf{A} * \mathbf{B}$ denotes the matrix formed by the product of the corresponding elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ and $\mathbf{A} * \mathbf{A} = \mathbf{A}^2$. Now the corresponding conditional ml estimates can be written as

$$\begin{aligned} \hat{\phi} &= \frac{(\mathbf{V}_1\mathbf{A}_1\mathbf{V}_2)(\mathbf{V}_1\mathbf{A}_2\mathbf{V}_2) - m(n-1)\mathbf{V}_1(\mathbf{A}_1 * \mathbf{A}_2)\mathbf{V}_2}{(\mathbf{V}_1\mathbf{A}_2\mathbf{V}_2)^2 - m(n-1)\mathbf{V}_1(\mathbf{A}_2 * \mathbf{A}_2)\mathbf{V}_2} \\ &= \frac{sum(\mathbf{A}_1)sum(\mathbf{A}_2) - m(n-1)sum(\mathbf{A}_1 * \mathbf{A}_2)}{sum(\mathbf{A}_2)^2 - m(n-1)sum(\mathbf{A}_2^2)} \end{aligned} \quad (2.35)$$

$$\begin{aligned} \hat{\mu} &= \frac{(\mathbf{V}_1\mathbf{A}_1\mathbf{V}_2)(\mathbf{V}_1(\mathbf{A}_2 * \mathbf{A}_2)\mathbf{V}_2) - (\mathbf{V}_1\mathbf{A}_2\mathbf{V}_2)(\mathbf{V}_1(\mathbf{A}_1 * \mathbf{A}_2)\mathbf{V}_2)}{(m(n-1)\mathbf{V}_1(\mathbf{A}_2 * \mathbf{A}_2)\mathbf{V}_2 - (\mathbf{V}_1\mathbf{A}_2\mathbf{V}_2)^2)(1-\hat{\phi})} \\ &= \frac{sum(\mathbf{A}_1)sum(\mathbf{A}_2^2) - sum(\mathbf{A}_2)sum(\mathbf{A}_1 * \mathbf{A}_2)}{(m(n-1)sum(\mathbf{A}_2^2) - sum(\mathbf{A}_2)^2)(1-\hat{\phi})} \end{aligned} \quad (2.36)$$

$$\hat{\sigma}^2 = (1 - \hat{\phi}^2)\hat{\mu}^2 + P_1 + P_2, \tag{2.37}$$

where

$$P_1 = \frac{\mathbf{V}_1(\mathbf{A}_1 * \mathbf{A}_1)\mathbf{V}_2) + \hat{\phi}^2\mathbf{V}_1(\mathbf{A}_2 * \mathbf{A}_2)\mathbf{V}_2}{m(n-1)} = \frac{sum(\mathbf{A}_1^2) - \hat{\phi}^2 sum(\mathbf{A}_2^2)}{m(n-1)}$$

and

$$P_2 = \frac{2\hat{\mu}(1-\hat{\phi})(\hat{\phi}(\mathbf{V}_1\mathbf{A}_2\mathbf{V}_2) - (\mathbf{V}_1\mathbf{A}_1\mathbf{V}_2)) - 2\hat{\phi}(\mathbf{V}_1(\mathbf{A}_1 * \mathbf{A}_2)\mathbf{V}_2)}{m(n-1)},$$

$$= \frac{2\hat{\mu}(1-\hat{\phi})(\hat{\phi} sum(\mathbf{A}_2) - sum(\mathbf{A}_1)) - 2\hat{\phi}(sum(\mathbf{A}_1 * \mathbf{A}_2))}{m(n-1)}.$$

Using these equations (2.35), (2.36) and (2.37), it is easy to estimate the parameters for a given replicated short series. Denote the corresponding vector of the estimates by $\hat{\delta}_1$, where $\hat{\delta}_1 = (\hat{\phi}_1, \hat{\mu}_1, \hat{\sigma}_1^2)'$.

Now we look at the exact maximum likelihood estimation procedure.

2.3.2. Exact Likelihood Estimation. Consider a stationary normally distributed AR(1) time series $\{X_{it}\}$ generated by (2.23). Let $\mathbf{X_T}$ be a sample of size $T = mn$ from (2.30) and let $\mathbf{X_T} = (\mathbf{X_1}, \cdots, \mathbf{X_m})'$, where $\mathbf{X_i} = (X_{i1}, \cdots, X_{in})'$. That is, the column $\mathbf{X_T}$ represents the vector of mn observations given by

$$\mathbf{X_T} = (X_{11}, \cdots X_{1n}, X_{21}, \cdots X_{2n}, \cdots, X_{m1}, \cdots, Xmn)'.$$

Then it is clear that $\mathbf{X_T} \sim N_T(\mu\mathbf{V}, \mathbf{\Sigma})$, where $\mathbf{V}$ is a column vector of 1's of order $T \times 1$ and Σ is the covariance matrix (order $T \times T$) of $\mathbf{X_T}$. From the independence of $\mathbf{X_i}$'s, $\mathbf{\Sigma}$ is a block diagonal matrix such that $\mathbf{\Sigma} = \text{diag}(\mathbf{\Omega})$, where $\mathbf{\Omega}$ is the covariance matrix (order $n \times n$) of any single series $\mathbf{X_i}$ given in (2.30).

The corresponding log-likelihood function can be written as

$$-2L = mn\log(2\pi\sigma^2) - m\log(1-\phi^2) + \sum_{i=1}^{m}\{(X_i1-\mu)^2 + \sum_{t=2}^{n}[X_{it} - \mu - \phi(X_{i,t-1}-\mu)]^2\}/\sigma^2. \quad (2.38)$$

Equation (2.38) is equivalent to

$$-2L\sigma^2 = mn\sigma^2\log(2\pi\sigma^2) - m\sigma^2\log(1-\phi^2) + \sum_{i=1}^{m}(X_{i1}-\mu)^2 + (1+\phi^2)$$

$$\sum_{i=1}^{m}\sum_{t=2}^{n-1}(X_{it}-\mu)^2 + \sum_{i=1}^{m}(X_{i,n}-\mu)^2 - 2\phi\sum_{i=1}^{m}\sum_{t=1}^{n-1}(X_{it}-\mu)(X_{i,t+1}-\mu).$$

To estimate the parameters ϕ, μ and σ^2, one needs a suitable optimization algorithm to maximize the log-likelihood function given in (2.38). As we have the covariance matrix $\mathbf{\Sigma} = diag(\mathbf{\Omega})$ in terms of the parameters ϕ and σ^2, the exact mle's can easily be obtained by choosing an appropriate set of starting up values for the optimization algorithm. Denote the corresponding vector of the estimates by $\hat{\delta}_2$, where $\hat{\delta}_2 = (\hat{\phi}_2, \hat{\mu}_2, \hat{\sigma}_2^2)'$. Furthermore, there are some alternative estimation procedures available via recursive methods.

In the next section we compare the finite sample properties of $\hat{\delta}_1$ and $\hat{\delta}_2$ via a simulation study with corresponding asymptotic results.

2.3.3. Finite Sample Comparison of $\hat{\delta}_1$ and $\hat{\delta}_2$. The properties of $\hat{\delta}_1$, and $\hat{\delta}_2$ are very similar to each other for large values of m, especially the bias and the mean square error (mse). However, there are slight differences for small values of m. It can be seen that the asymptotic covariance matrix of $\hat{\delta}_1$ based on (2.31) is

$$Cov(\hat{\delta}_1) = \frac{1}{m(n-1)} \begin{pmatrix} 1-\phi^2 & 0 & 0 \\ 0 & \frac{\sigma^2}{(1-\phi)^2} & 0 \\ 0 & 0 & 2\sigma^4 \end{pmatrix}.$$

The corresponding matrix for the exact mle of $\hat{\delta}_2$ based on (2.38) is

$$Cov(\hat{\delta}_2) = \frac{1}{ac-d^2} \begin{pmatrix} c & 0 & -d \\ 0 & \frac{ac-d^2}{b} & 0 \\ -d & 0 & a \end{pmatrix},$$

where

$$a = \frac{m[\phi^2(3-n)+n-1]}{(1-\phi^2)^2}, \; b = \frac{m(1-\phi)[n-(n-2)\phi]}{\sigma^2},$$

$$c = \frac{mn}{2\sigma^4}, \; d = \frac{m\phi}{\sigma^2(1-\phi^2)}.$$

Hence, it is clear that asymptotically (for large m) $\hat{\phi}$ is normal with mean ϕ and variance $\dfrac{c}{ac-d^2}$.

For example, when $m = 100$ and $n = 5$ ($\phi = .8, \; \mu = 4, \; \sigma^2 = 2$), we have

$$Cov(\hat{\delta}_1)_t = \begin{pmatrix} 0.0009 & 00.0000 & 0.0000 \\ 0.0000 & 0.0278 & 0.0000 \\ 0.0000 & 0.0000 & 0.0800 \end{pmatrix}$$

$$Cov(\hat{\delta}_2)_t = \begin{pmatrix} 0.0005 & 00.0000 & -0.0001 \\ 0.0000 & 0.0385 & 0.0000 \\ -0.0001 & 0.0000 & 0.0080 \end{pmatrix}$$

where t stands for corresponding theoretical values.

Under the null hypothesis of no serial correlation, ie. $\phi = 0$, the null distribution of $\hat{\phi}$ is asymptotically normal with mean 0 and variance $\frac{1}{m(n-1)}$. Next section reports a simulation study.

2.3.4. A Simulation Study. We first generate a sample of 100 observations from (2.30) for a given set of parameters ϕ, μ, and σ^2 using S-plus and repeat this m times. Now let us pick the last five columns of this matrix (of order $m \times 100$) and take a matrix (of order $m \times 5$) as our sample. Using this sample and equations (2.32), (2.33) and (2.34) let us compute $\hat{\delta}_1$ based on the conditional argument. The exact likelihood estimates of parameters ($\hat{\delta}_2$) are obtained by numerically maximizing (2.38) using the Newton-Raphson method. For $n = 5$, we repeat the simulation and estimation using both ML procedures for different values m and k. At the end of each estimation, we compute the mean and variance of $\hat{\delta}_1$ and $\hat{\delta}_2$. Further the bias and mean square error (mse) of $\hat{\delta}_1$ and $\hat{\delta}_2$ are obtained for comparison. Let $\hat{\delta}_{i.j}$ stand for the j^{th} estimate of the vector $\hat{\delta}_i$. ; i=1,2.

Then

$$\text{bias}_i = \frac{1}{k}\sum_{j=1}^{k}(\hat{\delta}_{i.j} - \delta_i), \tag{2.39}$$

and

$$\text{mse}_i = \frac{1}{k}\sum_{j=1}^{k}(\hat{\delta}_{i.j} - \delta_i)^2. \tag{2.40}$$

Below we tabulate these results for various values of m and k. Each table has four parts consisting of:

- the true values of ϕ and means of estimated values of $\hat{\phi}_1$ and $\hat{\phi}_2$,
- the variances of estimated values of $\hat{\phi}_1$ and $\hat{\phi}_2$,
- the biases of the estimates $\hat{\phi}_1$ and $\hat{\phi}_2$,
- mean square errors of the estimates $\hat{\phi}_1$ and $\hat{\phi}_2$.

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.7986,4.0019,0.9959)	(0.7964,3.9897,0.9991)
(0.7,5,1.5)	(0.6973,5.0137,1.5013)	(0.6973,5.0049,1.4890)
(0.6,6,2.0)	(0.5938,6.0090,1.9928)	(0.5967,6.0030,1.9920)
(0.5,5,1.0)	(0.4927,4.9930,1.0040)	(0.4961,5.0012,1.0006)
(0.4,8,5.0)	(0.3975,7.9734,5.0216)	(0.3967,8.0039,4.9618)
(0.3,10,7.0)	(0.2952,10.0001,6.9853)	(0.2973,10.0086,6.9624)
(0.2,4,3.0)	(0.1952,4.0052,3.0048)	(0.1999,4.0008,2.9898)
(0.1,5,2.0)	(0.1007,4.9965,1.9879)	(0.1006,5.0013,1.9888)

TABLE 1. Simulated means of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=100, k=300)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.0009,0.0613,0.0040)	(0.0005,0.0196,0.0042)
(0.7,5,1.5)	(0.0012,0.0489,0.0104)	(0.0010,0.0173,0.0107)
(0.6,6,2.0)	(0.0018,0.0282,0.0201)	(0.0013,0.0168,0.0166)
(0.5,5,1.0)	(0.0021,0.0104,0.0051)	(0.0018,0.0056,0.0045)
(0.4,8,5.0)	(0.0021,0.0400,0.1308)	(0.0019,0.0239,0.0901)
(0.3,10,7.0)	(0.0022,0.0394,0.2150)	(0.0020,0.0236,0.1932)
(0.2,4,3.0)	(0.0025,0.0124,0.0471)	(0.0022,0.0091,0.0316)
(0.1,5,2.0)	(0.0023,0.0067,0.0199)	(0.0023,0.0048,0.0143)

TABLE 2. Simulated variances of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=100, k=300)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(-0.0014,0.0019,-0.0041)	(-0.0036,-0.0103,-0.0009)
(0.7,5,1.5)	(-0.0027,0.0137,0.0013)	(-0.0027,0.0049,-0.0110)
(0.6,6,2.0)	(-0.0062,0.0090,-0.0072)	(-0.0033,0.0030,-0.0080)
(0.5,5,1.0)	(-0.0073,-0.0070,0.0040)	(-0.0039,0.0012,0.0006)
(0.4,8,5.0)	(-0.0025,-0.0266,0.0216)	(-0.0033,0.0039,-0.0382)
(0.3,10,7.0)	(-0.0048,0.0001,-0.0147)	(-0.0027,0.0086,-0.0376)
(0.2,4,3.0)	(-0.0048,0.0052,0.0048)	(-0.0001,0.0008,-0.0102)
(0.1,5,2.0)	(0.0007,-0.0035,0.0121)	(0.0006,0.0013,-0.0112)

TABLE 3. Simulated bias of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=100, k=300)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.0009,0.0613,0.0041)	(0.0005,0.0197,0.0042)
(0.7,5,1.5)	(0.0012,0.0491,0.0104)	(0.0010,0.0173,0.0108)
(0.6,6,2.0)	(0.0018,0.0283,0.0201)	(0.0013,0.0168,0.0166)
(0.5,5,1.0)	(0.0021,0.0104,0.0051)	(0.0018,0.0056,0.0045)
(0.4,8,5.0)	(0.0021,0.0407,0.1312)	(0.0019,0.0240,0.0915)
(0.3,10,7.0)	(0.0022,0.0394,0.2152)	(0.0020,0.0237,0.1946)
(0.2,4,3.0)	(0.0025,0.0124,0.0471)	(0.0022,0.0091,0.0317)
(0.1,5,2.0)	(0.0023,0.0067,0.0201)	(0.0023,0.0048,0.0145)

TABLE 4. Simulated mse of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=100, k=300)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.7988,3.9959,0.9979)	(0.7995,4.0003,0.9983)
(0.7,5,1.5)	(0.6988,4.9975,1.4959)	(0.6999,4.9994,1.4986)
(0.6,6,2.0)	(0.5996,6.0032,1.9988)	(0.5985,6.0003,2.0019)
(0.5,5,1.0)	(0.5009,4.9973,1.0009)	(0.4983,5.0007,0.9999)
(0.4,8,5.0)	(0.4012,7.9934,4.9918)	(0.4000,8.0058,4.9963)
(0.3,10,7.0)	(0.2986,9.9990,7.0031)	(0.2995,10.0005,7.0058)
(0.2,4,3.0)	(0.1992,3.9975,2.9954)	(0.2006,4.0014,3.0006)
(0.1,5,2.0)	(0.1003,5.0002,2.0029)	(0.0986,5.0024,1.9965)

TABLE 5. Simulated means of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=500, k=500)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.0002,0.0135,0.0010)	(0.0001,0.0040,0.0009)
(0.7,5,1.5)	(0.0003,0.0077,0.0021)	(0.0002,0.0035,0.0021)
(0.6,6,2.0)	(0.0003,0.0069,0.0042)	(0.0002,0.0033,0.0034)
(0.5,5,1.0)	(0.0004,0.0019,0.0011)	(0.0003,0.0010,0.0008)
(0.4,8,5.0)	(0.0004,0.0065,0.0274)	(0.0004,0.0037,0.0194)
(0.3,10,7.0)	(0.0004,0.0075,0.0481)	(0.0004,0.0049,0.0411)
(0.2,4,3.0)	(0.0005,0.0023,0.0090)	(0.0005,0.0018,0.0065)
(0.1,5,2.0)	(0.0005,0.0013,0.0037)	(0.0004,0.0010,0.0032)

TABLE 6. Simulated variances of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=500, k=500)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(-0.0012,-0.0041,-0.0021)	(-0.0005,0.0003,-0.0017)
(0.7,5,1.5)	(-0.0012,-0.0025,-0.0041)	(-0.0001,-0.0006,-0.0014)
(0.6,6,2.0)	(-0.0004,0.0032,-0.0012)	(-0.0015,0.0003,0.0019)
(0.5,5,1.0)	(0.0009,-0.0027,0.0009)	(-0.0017,0.0007,-0.0001)
(0.4,8,5.0)	(0.0012,-0.0066,-0.0082)	(0.0000,0.0058,-0.0037)
(0.3,10,7.0)	(-0.0014,-0.0010,0.0031)	(-0.0005,0.0005,0.0058)
(0.2,4,3.0)	(-0.0008,-0.0025,-0.0046)	(0.0006,0.00014,0.0006)
(0.1,5,2.0)	(0.0003,0.0002,0.0029)	(-0.0014,0.0024,-0.0035)

TABLE 7. Simulated bias of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=500, k=500)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.0002,0.0135,0.0010)	(0.0001,0.0040,0.0009)
(0.7,5,1.5)	(0.0003,0.0077,0.0021)	(0.0002,0.0035,0.0021)
(0.6,6,2.0)	(0.0003,0.0069,0.0042)	(0.0003,0.0033,0.0034)
(0.5,5,1.0)	(0.0004,0.0019,0.0011)	(0.0003,0.0010,0.0008)
(0.4,8,5.0)	(0.0004,0.0066,0.0275)	(0.0004,0.0037,0.0194)
(0.3,10,7.0)	(0.0004,0.0075,0.0481)	(0.0004,0.0049,0.0411)
(0.2,4,3.0)	(0.0005,0.0023,0.0090)	(0.0005,0.0018,0.0065)
(0.1,5,2.0)	(0.0005,0.0013,0.0037)	(0.0004,0.0010,0.0033)

TABLE 8. Simulated mse of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=500, k=500)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.7998,40030,0.9985)	(0.7998,4.0005,0.9986)
(0.7,5,1.5)	(0.7000,5.0020,1.4981)	(0.6992,5.0007,1.4996)
(0.6,6,2.0)	(0.5995,5.9986,1.9993)	(0.5998,5.9980,1.9986)
(0.5,5,1.0)	(0.4995,4.9999,0.9994)	(0.4998,5.0004,1.0000)
(0.4,8,5.0)	(0.3993,7.9997,4.9981)	(0.4001,7.9992,5.0006)
(0.3,10,7.0)	(0.3003,9.9999,6.9903)	(0.3000,9.9979,6.9999)
(0.2,4,3.0)	(0.1999,3.9995,2.9972)	(0.1993,4.0003,2.9961)
(0.1,5,2.0)	(0.0992,5.0000,1.9980)	(0.1003,4.9999,1.9997)

TABLE 9. Simulated means of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=1000, k=1000)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.0001,0.0063,0.0005)	(0.0001,0.0019,0.0004)
(0.7,5,1.5)	(0.0001,0.0043,0.0011)	(0.0001,0.0017,0.0009)
(0.6,6,2.0)	(0.0002,0.0032,0.0020)	(0.0001,0.0016,0.0016)
(0.5,5,1.0)	(0.0002,0.0010,0.0005)	(0.0002,0.0006,0.0004)
(0.4,8,5.0)	(0.0002,0.0034,0.0127)	(0.0002,0.0021,0.0104)
(0.3,10,7.0)	(0.0003,0.0035,0.0234)	(0.0002,0.0023,0.0200)
(0.2,4,3.0)	(0.0002,0.0011,0.0046)	(0.0002,0.0008,0.0036)
(0.1,5,2.0)	(0.0002,0.0006,0.0020)	(0.0003,0.0005,0.0018)

TABLE 10. Simulated variances of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=1000, k=1000)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(-0.0002,0.0030,-0.0015)	(-0.0002,0.0005,-0.0014)
(0.7,5,1.5)	(0.0000,0.0020,-0.0019)	(-0.0008,0.0007,-0.0004)
(0.6,6,2.0)	(-0.0005,-0.0014,-0.0007)	(-0.0002,-0.0020,-0.0014)
(0.5,5,1.0)	(-0.0005,-0.0001,-0.0006)	(-0.0002,0.0004,0.0000)
(0.4,8,5.0)	(-0.0007,-0.0003,-0.0019)	(0.0001,-0.0008,0.0006)
(0.3,10,7.0)	(0.0003,-0.0001,-0.0092)	(0.0000,-0.0021,0.0001)
(0.2,4,3.0)	(-0.0001,-0.0005,-0.0028)	(-0.0007,0.00003,-0.0039)
(0.1,5,2.0)	(-0.0008,0.0000,-0.0020)	(0.0003,-0.0001,-0.0003)

TABLE 11. Simulated bias of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=1000, k=1000)

δ	$\hat{\delta}_1$,Conditional mle	$\hat{\delta}_2$,Exact mle
(0.8,4,1.0)	(0.0001,0.0063,0.0005)	(0.0001,0.0019,0.0005)
(0.7,5,1.5)	(0.0001,0.0043,0.0011)	(0.0001,0.0017,0.0009)
(0.6,6,2.0)	(0.0002,0.0032,0.0020)	(0.0001,0.0016,0.0016)
(0.5,5,1.0)	(0.0002,0.0010,0.0005)	(0.0002,0.0006,0.0004)
(0.4,8,5.0)	(0.0002,0.0034,0.0127)	(0.0002,0.0021,0.0104)
(0.3,10,7.0)	(0.0003,0.0035,0.0235)	(0.0002,0.0023,0.0200)
(0.2,4,3.0)	(0.0002,0.0011,0.0046)	(0.0002,0.0008,0.0036)
(0.1,5,2.0)	(0.0002,0.0006,0.0020)	(0.0003,0.0005,0.0018)

TABLE 12. Simulated mse of $\hat{\delta}_1$ and $\hat{\delta}_2$ (m=1000, k=1000)

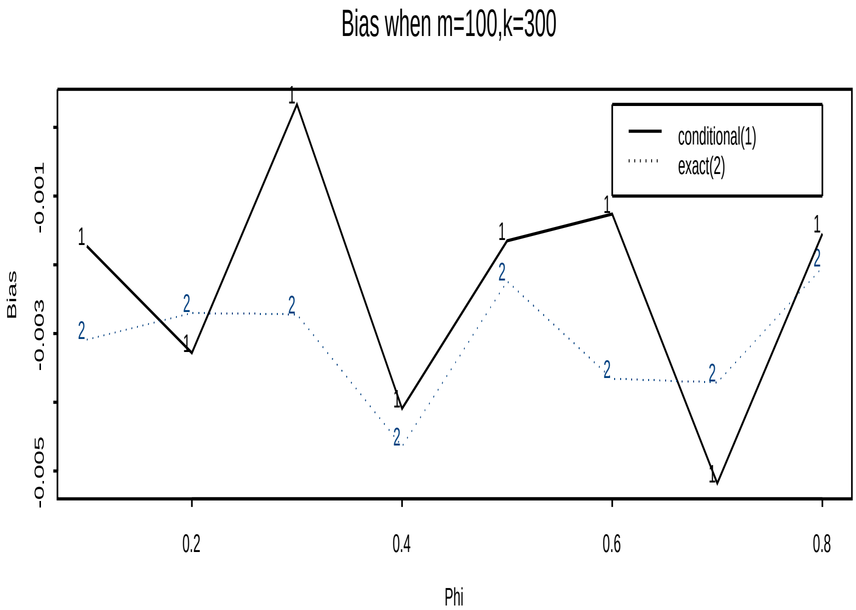
For $\phi = 0.8$, $\mu = 4$ and $\sigma^2 = 2$, we compute the covariance matrices of $\hat{\delta}_1$ and $\hat{\delta}_2$ based on our simulations for $m = 100$ and $n = 5$. We denote these matrices by $Cov(\hat{\delta}_1)_s$ and $Cov(\hat{\delta}_2)_s$ for convenience ('s' stands for simulation).

$$Cov(\hat{\delta}_1)_s = \begin{pmatrix} 0.0009 & 00.0012 & 0.0000 \\ 0.0012 & 0.1367 & 0.0000 \\ 0.0000 & 0.0000 & 0.0180 \end{pmatrix}$$

$$Cov(\hat{\delta}_2)_s = \begin{pmatrix} 0.0005 & 00.0000 & -0.0009 \\ 0.0000 & 0.0391 & 0.0000 \\ -0.0009 & 0.0000 & 0.0160 \end{pmatrix}$$

We now can conclude that in each case (conditional mle and exact mle) $\text{var}(\hat{\phi}_i)_t \approx \text{var}(\hat{\phi}_i)_s$ for i=1,2 (t and s stand for theoretical and simulation respectively) and for large m. It is clear that, for large m, the loss of information under the conditional likelihood procedure is negligible (recall that this is the conclusion for $m = 1$ and n is large).

Below we give graphical representations of the bias (Figure 1) and the mse (Figure 2) of $\hat{\phi}$ in both conditional and exact cases for a given m and k. This graphs further confirm the above conclusion.



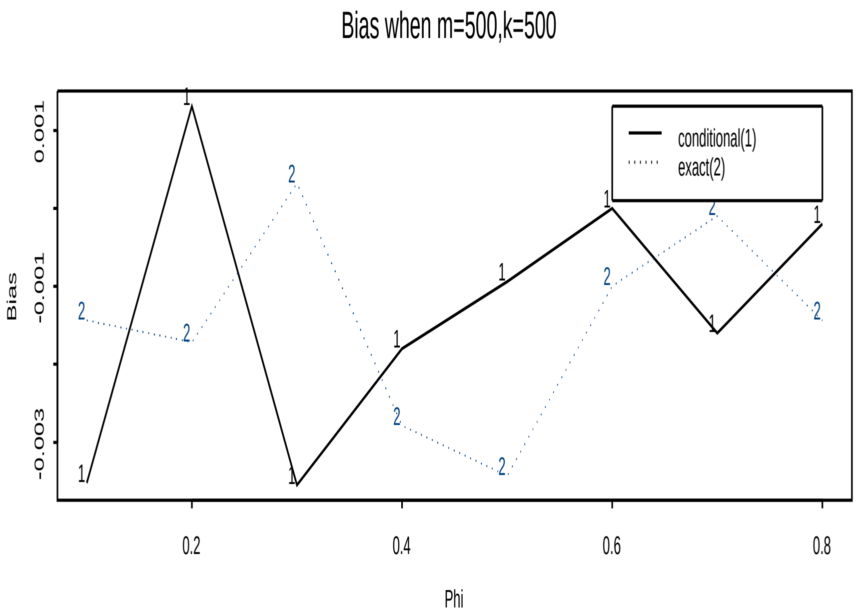


FIGURE 1. Comparison of the bias of $\hat{\phi}_1$ and $\hat{\phi}_2$ ($\mu = 5$ and $\sigma^2 = 1$)

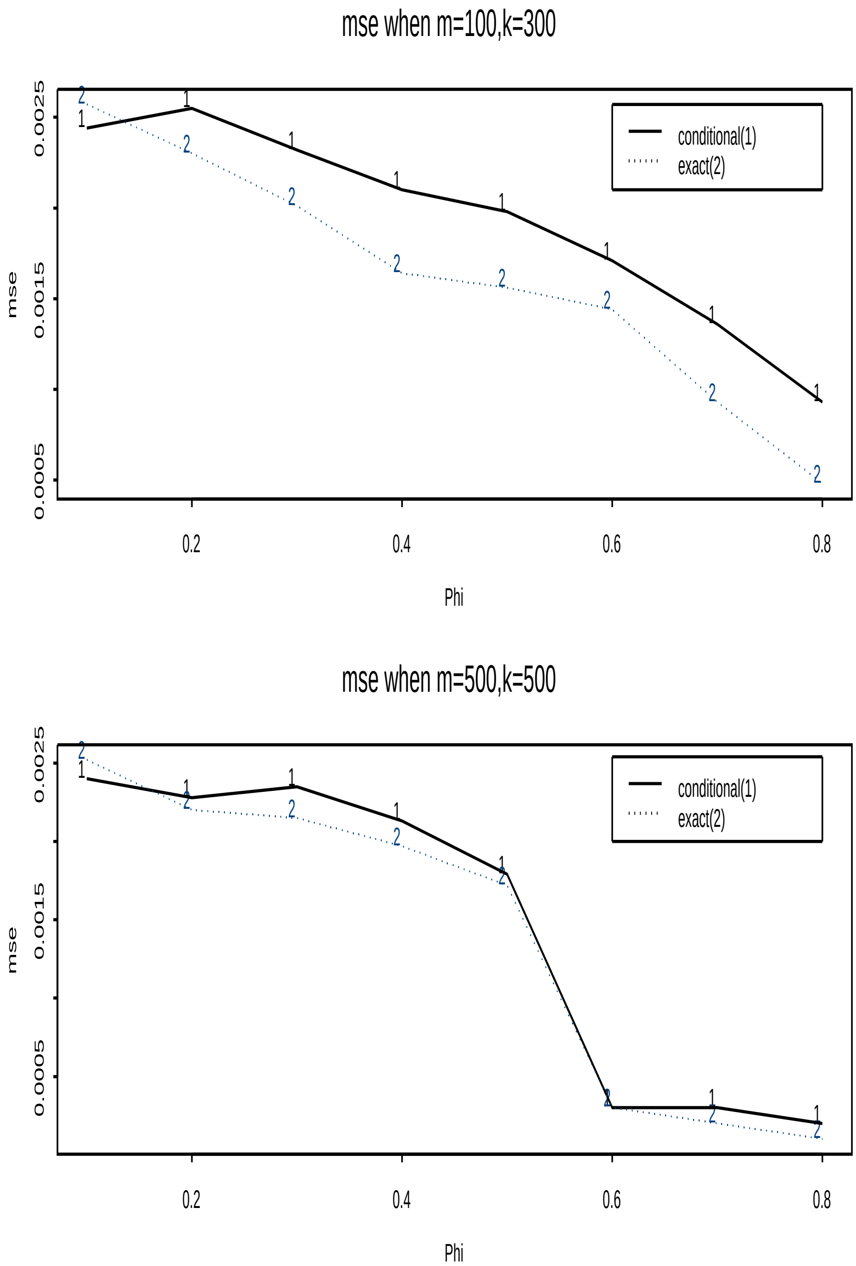


FIGURE 2. Comparison of the mse of $\hat{\phi}_1$ and $\hat{\phi}_2$ ($\mu = 5$ and $\sigma^2 = 1$)

Now we describe a method of estimation due to Cox and Solomon (1988) for the sake of comparison.

2.3.5. The approach due to Cox and Solomon (1988). Suppose that $\{X_1, \ldots, X_n\}$ be a sample with serial correlation ρ at lag one. When n is small, define

$$\hat{\rho} = \frac{Q}{S}, \tag{2.41}$$

where the lag one sum of product

$$Q = \sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X}) = \sum_{t=1}^{n-1} X_t X_{t+1} - (n+1)\bar{X}^2 + \bar{X}(X_1 + X_n),$$

$$S = \sum_{t=1}^{n}(X_t - \bar{X})^2, \ and \ \bar{X} = \frac{\sum_{t=1}^{n} X_t}{n}.$$

Note: For large n, $\hat{\rho} \longrightarrow \phi$.

For the i^{th} sample, denote these values by (Q_i, S_i),$i = 1, \cdots, m$, where the i^{th} sample is $\{X_{i1}, X_{i2}, \cdots, X_{in}\}$.

Pooling numerators and denominators of (2.41), let

$$\hat{\phi} = \sum_{i=1}^{m} Q_i / \sum_{i=1}^{m} S_i. \tag{2.42}$$

Cox and Solomon (1988) have shown that the statistic given in (2.42) is more efficient than $\hat{\phi}^* = \sum_{i=1}^{m}(Q_i/S_i)$ when the variance σ^2 is constant.

To derive the distribution of $\hat{\phi}$ consider the following lemma:

Lemma: Let$X_1, \ldots, X_n$, be independent identically distributed standard normal variables, and $\mathbf{X} = (X_1, \ldots, X_n)'$, then

(i) $E(\mathbf{X}'\mathbf{A}\mathbf{X}) = tr(\mathbf{A})$

(ii) $\text{var}(\mathbf{X}'\mathbf{A}\mathbf{X}) = 2tr(\mathbf{A}^2)$

(iii) $\text{cov}(\mathbf{X}'\mathbf{A}\mathbf{X}, \mathbf{X}'\mathbf{B}\mathbf{X}) = 2tr(\mathbf{AB})$

where $\mathbf{A}$,$\mathbf{B}$ are arbitrary constant matrices.

Proof:see Searle(1971) , p.57.

To obtain the asymptotic distribution of the test statistic $\hat{\phi}$ under the null hypothesis $H_0 : \rho = 0$, (i.e. the observations are independent) and standard normal variables, we can use the above Lemma. For example, writing $Q = \mathbf{X}'\mathbf{A}\mathbf{X}$ and $S = \mathbf{X}'\mathbf{B}\mathbf{X}$ for suitably chosen matrices $\mathbf{A}$ and $\mathbf{B}$, it can be seen that

$E(Q) = -(n-1)/n, \quad \text{var}(Q) = (n^3 - 3n^2 + 2n + 2)/n^2,$

$E(S) = n - 1, \quad \text{var}(S) = 2(n-1), \quad \text{and } \text{cov}(Q, S) = -2(n-1)/n.$

It follows after further calculation that under the null hypothesis the approximate mean and variance of the limiting normal distribution of $\hat{\phi}$ are $-1/n$ and $[(n+1)(n-2)^2]/[mn^2(n-1)^2]$ respectively.

Using the estimator of ϕ given in (2.42), the corresponding null distribution (for large m) of $\hat{\phi}$ when $n = 5$ is normal with mean $\frac{-1}{5}$ and variance $\frac{27}{200m}$. (See Cox and Solomon (1988), p.147). Although this has a smaller variance than the previous one given in section 2.3.3 (i.e.$\frac{1}{4m}$), it has a considerable bias.

As described before, the assumption of linearity is not reasonable in many applications. It is recognized by many researchers that the conditional means and conditional variances are not constant in some applications as in the linear theory. These types of models are called nonlinear time series models and the next Chapter is devoted to discussing some of their properties in detail.

Chapter 3

Nonlinear Time Series Models

As many applications in financial data are nonlinear, nonlinear models are appropriate for forecasting and accurately describing returns and volatility. Since there are an enormous number of nonlinear models available for modelling and forecasting economic time series, choosing the best model for a particular application is daunting (Franses and Van Dijk (2000),p.2).

In financial time series asymmetric behaviour is often displayed . An example of this behavior is that large negative returns appear more frequently than large positive returns. Another example is that large negative returns are often a prelude to a period of substantial volatility, while large positive returns are less so. Such asymmetries should be incorporated in a time series model for description and out-of-sample forecasting to avoid forecasts that are always too low or too high.

Non-linear time series analysis is a rapidly developing area and there have been major developments in model building and forecasting. In this chapter, we discuss five popular models in practice and describe their statistical properties. We first consider the class of bilinear time series models.

3.1. Bilinear Models

Wiener(1958) considered a nonlinear relationships between an input U_t and an output X_t (both observable) using Volterra series expansion given by

$$X_t = \sum_{i=0}^{\infty} \alpha_i U_{t-i} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \alpha_{ij} U_{t-i} U_{t-j} + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \alpha_{ijk} U_{t-i} U_{t-j} U_{t-k} + \cdots . \tag{3.1}$$

From a given finite realization of a process, one cannot estimate the parameters $\{\alpha_i\}, \{\alpha_{ij}\}, \{\alpha_{ijk}\}, \cdots$ efficiently. To overcome this difficulty Granger and Anderson (1978) have introduced a class of nonlinear models in the time series context assuming $U_t = \varepsilon_t$ (unobservable) satisfying

$$\phi(B) X_t = \theta(B)\varepsilon_t + \sum_{i=1}^{r}\sum_{j=1}^{s} \beta_{ij} X_{t-i}\varepsilon_{t-j} \ , \tag{3.2}$$

where $\phi(B)$ and $\theta(B)$ are p^{th} order AR and q^{th} order MA polynomials on back shift operator B as given in (2.2) and β_{ij} are constants. This is an extension of the (linear) ARMA model obtained by adding the nonlinear term $\sum_{i=1}^{r}\sum_{j=1}^{s} \beta_{ij} X_{t-i}\varepsilon_{t-j}$ to the right hand side. In literature, the model (3.2) is called a bilinear time series model of order (p, q, r, s) and denoted $BL(p, q, r, s)$.

In their monograph Granger and Anderson (1978) have considered the statistical properties of the model $BL(1, 0, 1, 1)$. Subba Rao(1981) has analyzed the model $BL(p, 0, p, 1)$ and obtained some interesting time series properties.

Consider the bilinear model $BL(p,0,p,1)$ is given by

$$\phi(B)X_t = \varepsilon_t + \left(\sum_{i=1}^{p} \beta_{i1} X_{t-i}\right) \varepsilon_{t-1} \ . \tag{3.3}$$

Following Subba Rao (1981), we show that (3.3) can be written in the state space form. Define $p \times p$ matrices A and B such that

$$\mathbf{A}=\begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \ \mathbf{B}=\begin{pmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{p-1,1} & \beta_{p1} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Let the vector $\mathbf{H}'=(1,0,\cdots,0)$ of $(1 \times p)$ and the random vector $\mathcal{X}_t{}'=(X_t, X_{t-1}, \cdots, X_{t-p+1})$ of order $(1 \times p)$. Using the above notation we rewrite the model (3.3) in the form

$$\mathcal{X}_t = \mathbf{A}\mathcal{X}_{t-1} + \mathbf{B}\mathcal{X}_{t-1}\varepsilon_{t-1} + \mathbf{H}\varepsilon_t, \tag{3.4}$$

$$X_t = \mathbf{H}'\mathcal{X}_t. \tag{3.5}$$

The above representations (3.4) and (3.5) together are called the state space representation of the bilinear model BL(p,0,p,1). The representations (3.4) and (3.5) taken together are a vector form of the bilinear model BL(p,0,p,1) and we denote this as VBL(p) for convenience. We extend this approach to obtain the state space representation of $BL(p,0,p,q)$ which can be obtained as follows:
Define the matrices

$$\mathbf{B}_j = \begin{pmatrix} \beta_{1j} & \beta_{2j} & \cdots & \beta_{p-1,j} & \beta_{pj} \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \ j = 1, \cdots, q$$

and write the state space representation as

$$\begin{aligned} \mathcal{X}_t &= \mathbf{A}\mathcal{X}_{t-1} + \sum_{j=1}^{q} \mathbf{B}_j \mathcal{X}_{t-1}\varepsilon_{t-j} + \mathbf{H}\varepsilon_t, \quad (3.6) \\ X_t &= \mathbf{H}'\mathcal{X}_t. \quad (3.7) \end{aligned}$$

This form can be denoted as $VBL(p, q)$.

We use this approach to write the general bilinear model $BL(p, q, r, s)$ in the state space form. First we define the matrices of order $(l + 2) \times (l + 2)$ (l=max(p,q,r,s)) as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \phi_1 & \phi_2 & \cdots & \phi_l & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

$$\mathbf{B}_j = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \theta_j & \beta_{1j} & \beta_{2j} & \cdots & \beta_{lj} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \ (j = 1, \cdots, \max(q, s)),$$

and the vector $\mathbf{H}'=(0, 1, 0, \cdots, 0)$. Let $\mathcal{X}_t$ be the random vector given by, $\mathcal{X}_t' = (1, X_t, X_{t-1}, \cdots, X_{t-l})$ of $1 \times (l+2)$. Now the state space representation of the general $BL(p,q,r,s)$ can then be expressed as

$$\mathcal{X}_t = \mathbf{A}\mathcal{X}_{t-1} + \sum_{j=1}^{max(q,s)} \mathbf{B}_j \mathcal{X}_{t-1}\varepsilon_{t-j} + \mathbf{H}\varepsilon_t, \tag{3.8}$$

$$X_t = \mathbf{H}'\mathcal{X}_t, \tag{3.9}$$

where $\theta_j = 0,\ j > q$; $\beta_{ij} = 0,\ j > s$ or $i > r$; $X_{t-i} = 0,\ i > p$.

Example 3.1. Consider the BL(1,1,1,1) model

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} + \beta X_{t-1}\varepsilon_{t-1}. \tag{3.10}$$

A state space representation of (3.10)is

$$\begin{pmatrix} 1 \\ X_t \\ X_{t-1} \end{pmatrix} = \mathbf{A} \begin{pmatrix} 1 \\ X_{t-1} \\ X_{t-2} \end{pmatrix} + \mathbf{B} \begin{pmatrix} 1 \\ X_{t-1} \\ X_{t-2} \end{pmatrix} \varepsilon_{t-1} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \varepsilon_t$$

$$\text{and } X_t = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ X_t \\ X_{t-1} \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ \theta & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

A graphical representation of a series generated by (3.10) with $\phi = 0.8$, $\theta = 0.7$ and $\beta = 0.6$ is given in Figure 3. The *acf* and *pacf* are given in Figures 4 and 5 respectively.

FIGURE 3. Plot of BL(1,1,1,1) ($\phi = 0.8$, $\theta = 0.7$ and $\beta = 0.6$).

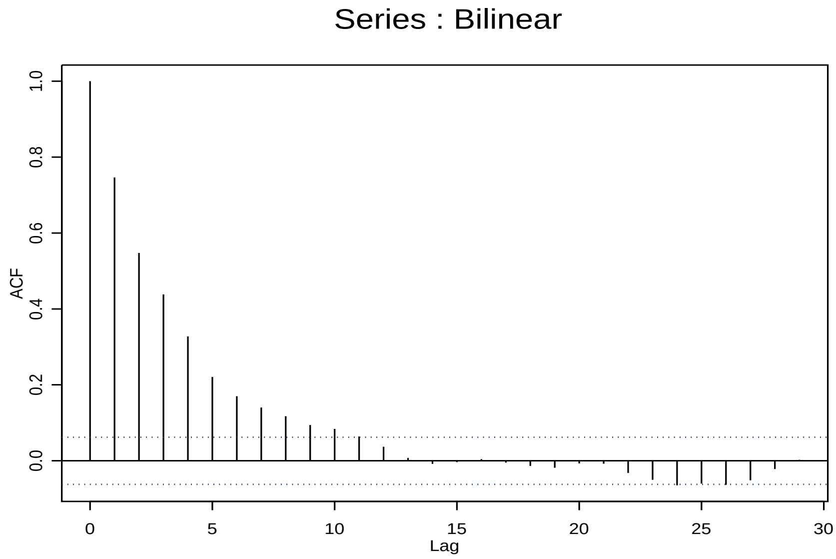


FIGURE 4. acf plot of BL(1,1,1,1) ($\phi = 0.8$, $\theta = 0.7$ and $\beta = 0.6$).

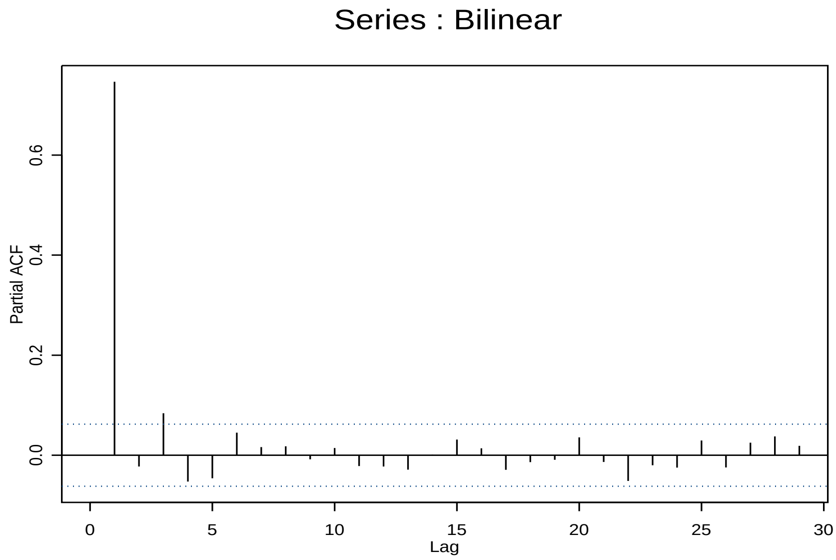


FIGURE 5. pacf plot of BL(1,1,1,1) ($\phi = 0.8$, $\theta = 0.7$ and $\beta = 0.6$).

Stensholt and Tjøstheim (1987) have considered a class of multiple bilinear models, and shown that the space time bilinear (STBL) model is a special form of a multiple bilinear model. This model and its identification have been proposed and discussed by Dai and Billard (1998). Recently, Dai and Billard (2003) have considered the problem of the parameter estimation for the space time bilinear model. A conditional maximum likelihood estimation procedure was provided through the use of a Newton-Raphson numerical optimization algorithm with the assumption that the model is stationary and invertible.

In Chapter 4 we consider the estimation function approach due to Godambe (1985) to estimate parameters. In section 3.2 we consider the class of random coefficient autoregressive models.

3.2. Random Coefficient Autoregressive (RCA) Models

RCA models are defined by allowing random additive perturbations $\{\beta_i(t)\}$ for the autoregressive coefficients of ordinary AR models. That is, a stochastic process $\{X_t\}$ is said to follow an RCA model of order k ($RCA(k)$), if X_t satisfies an equation of the form

$$X_t - \sum_{i=1}^{k}(\phi_i + \beta_i(t))X_{t-i} = \varepsilon_t, \tag{3.11}$$

where

(i) $\{\epsilon_t\}$ and $\{\beta_i(t)\}$ are zero mean square integrable independent processes with constant variances σ_ϵ^2 and σ_β^2;

(ii) $\beta_i(t)(i = 1, 2, \cdots, k)$ are independent of $\{\epsilon_t\}$ and $\{X_{t-i}\}; i \geq 1$;

(iii) $\phi_i,\ i = 1, \cdots, k$, are the parameters to be estimated;

Conlisk (1974), (1976) has derived conditions for the stability of RCA models. Robinson (1978) has considered statistical inference for the RCA model. Nicholls and Quinn (1982), proposed a method of testing

$$H_0 : \sigma_\beta^2 = 0 \quad vs \quad H_1 : \sigma_\beta^2 > 0 \tag{3.12}$$

based on likelihood criterion. Ramanathan and Rajarshi (1994) have suggested a test based on the least square residuals. Both Nicholls-Quinn and Ramanathan-Rajarshi assumed $\beta_i(t)$ and ε_t are independent. This assumption of independence plays an important role in deriving the limiting distribution of the test statistics.

Recently, Lee (1998) has considered a generalized first RCA model, which includes RCA(1) (independence of $\beta(t)$ and ε_t is not required),

$$X_t = (\phi + \beta(t))X_{t-1} + \varepsilon_t, \tag{3.13}$$

where $(\beta(t), \varepsilon_t)$ is an iid random vectors with $E(\beta(t)) = E(\varepsilon_t) = 0$, $E(\beta^2(t)) = \sigma_\beta^2$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $\text{cov}(\beta(t), \varepsilon_t) = \psi\sigma_\beta$ with $|\psi| < \sigma_\varepsilon$. Assuming the Gaussianity of $(\beta(t), \varepsilon_t)$, Lee has developed a locally best invariant (LBI) test for $\phi = 1$. The LBI test statistic $Z_T(\hat{\phi})$ for the test (3.12) with the estimate $\hat{\phi}$ of ϕ for the model (3.13), is given by

$$Z_T(\phi) = \begin{cases} \upsilon^{-4}(\phi)T(T+2)\sum_{t=1}^{T}(X_t - X_{t-1})^2X_{t-1}^2 - \upsilon^{-2}(\phi)T\sum_{t=2}^{T}X_{t-1}^2, \\ \qquad\qquad \text{if } T = 2n, \\ \upsilon^{-4}(\phi)T(T+2)\sum_{t=1}^{T}(X_t - X_{t-1})^2X_{t-1}^2 - \upsilon^{-2}(\phi)\frac{T(T+1)}{T-1}\sum_{t=2}^{T}X_{t-1}^2, \\ \qquad\qquad \text{if } T = 2n+1, \end{cases} \tag{3.14}$$

with $\nu^2(\phi) = \sum_{t=2}^{T}(X_t - \phi X_{t-1})^2 + X_1^2$, where $n = 1, 2, \cdots$.

The test statistic is shown to be asymptotically standard normal under the null hypothesis H_0. Under the alternative hypothesis H_1, the test statistic is shown to diverge to ∞ in probability. This asserts the consistency of the test for coefficient.

We now consider the state space representation of RCA models. Rewrite the equation (3.11) as follows:

$$X_t - \sum_{i=1}^{k}\phi_i X_{t-i} = \sum_{i=1}^{k}\beta_i(t)X_{t-i} + \varepsilon_t. \tag{3.15}$$

To write a state space form of the RCA model consider a similar approach as in the bilinear model.
Define the following matrices:

$$\mathbf{A} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{k-1} & \phi_k \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

of order $(k \times k)$, the vector $\mathbf{H}'=(1, 0, \cdots, 0)$ of $(1 \times k)$ and the random vectors $\beta'(t) = (\beta_1(t), \cdots, \beta_k(t))$ and $\mathcal{X}_t'=(X_t, X_{t-1}, \cdots, X_{t-k+1})$ of order $(1 \times k)$. The state space representation of the RCA model (3.11) is expressed as

$$\mathcal{X}_t = \mathbf{A}\mathcal{X}_{t-1} + \mathbf{B} \ \mathrm{diag}(\mathcal{X}_{t-1}\beta'(t)) + \mathbf{H}\varepsilon_t, \quad (3.16)$$

$$X_t = \mathbf{H}'\mathcal{X}_t, \quad (3.17)$$

where diag($\mathbf{A}$) stands for a vector formed by the diagonal elements of the corresponding matrix $\mathbf{A}$. Using this state space form, we can write a recursive estimation algorithm based on Kalman filtering.

Example 3.2. Consider the RCA(1) model given by

$$X_t = (\phi + \beta(t))X_{t-1} + \varepsilon_t, \quad (3.18)$$

where
(i) $\{\epsilon_t\}$ and $\{\beta(t)\}$ are zero mean square integrable independent processes with constant variances σ^2_ϵ and σ^2_β,

(ii) $\beta(t)$ are independent of $\{\epsilon_t\}$ and $\{X_{t-1}\}$,

(iii) ϕ is the parameter to be estimated.

Let $\{\epsilon_t\}$ and $\{\beta(t)\}$ be independent IID $N(0,1)$ variates.

State space form of 3.18 is:

$$\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \mathbf{A}\begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \mathbf{B}\ \mathrm{diag}(\begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix}\begin{pmatrix} \beta(t) & 0 \end{pmatrix}) + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\varepsilon_t$$

and

$$X_t = \begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} \phi & 0 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$.

The graphical representation of simulated series of (3.18) with $\phi = 0.8$ is given in Figure 6. The *acf* and the *pacf* are given in Figures 7 and 8 respectively.

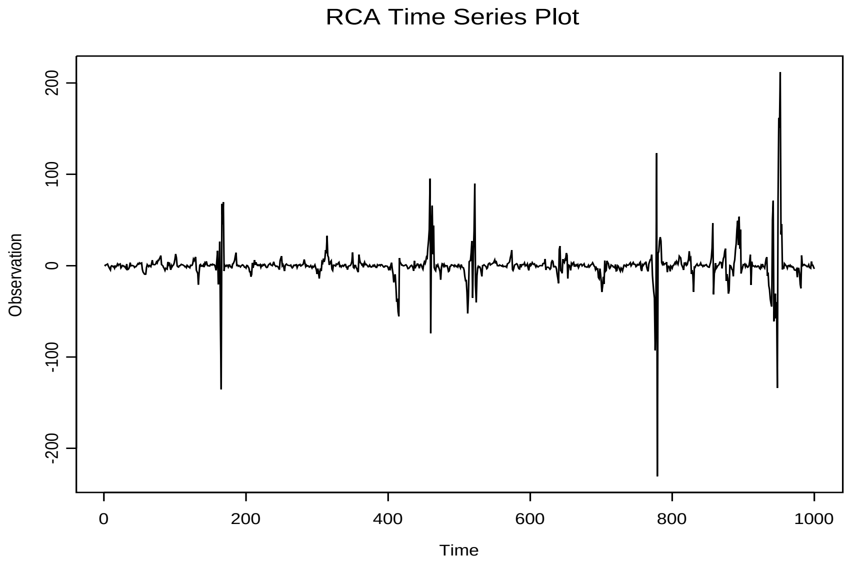


FIGURE 6. Plot of RCA(1) ($\phi = 0.8$).

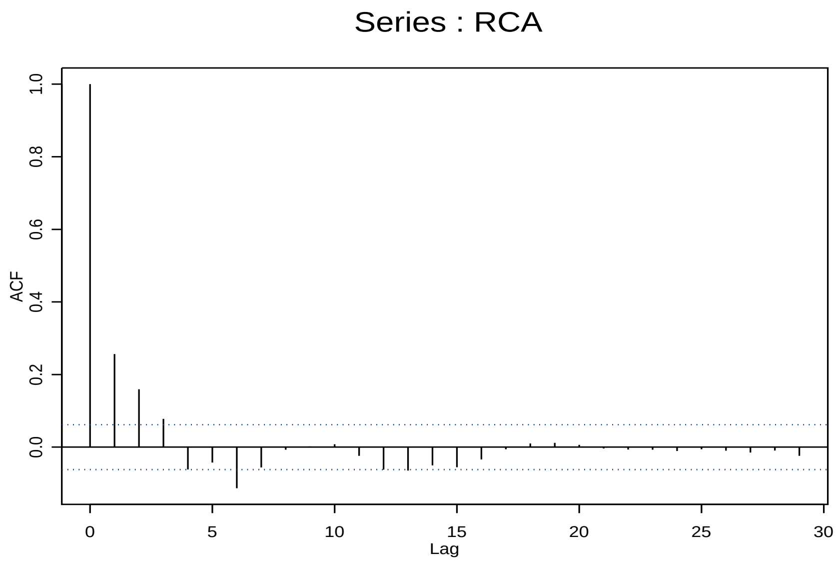


FIGURE 7. acf plot of RCA(1) ($\phi = 0.8$.)

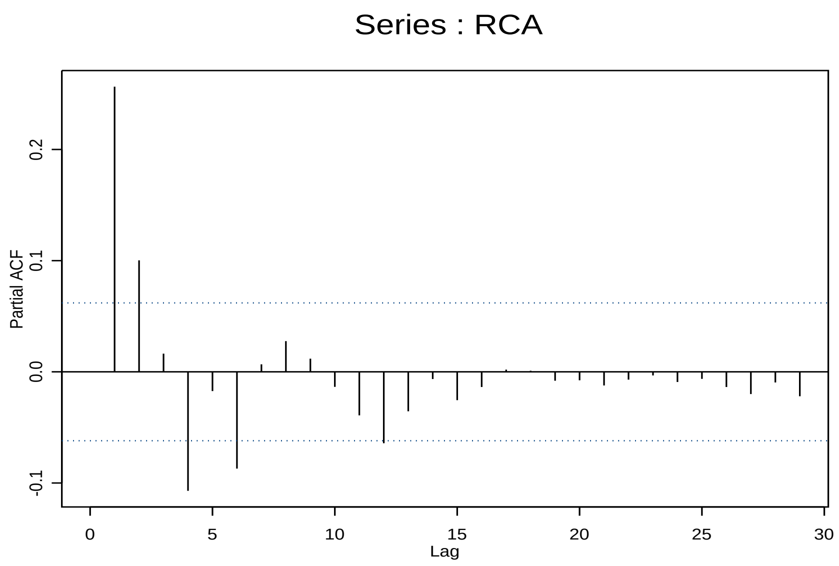


FIGURE 8. pacf plot of RCA(1) ($\phi = 0.8$).

Next section considers the class of doubly stochastic models.

3.3. Doubly Stochastic Models

Tjøstheim (1986) considered the class of doubly stochastic time series models as another generalization of an ARMA(p,q) family given in (2.1) which is obtained by replacing the parameters of (2.1) by stochastic processes. The motivation behind this class of models is that in many practical cases, the underlying mechanism described by the parameters of (2.1) expected to change in a non deterministic manner.

A general class of doubly stochastic time series model of order (p,q) is given by

$$X_t - \sum_{i=1}^{p} \phi_i(f_i(\mathcal{F}_{t-1}^X))X_{t-i} = \varepsilon_t + \sum_{i=1}^{q} \theta_i(g_i(\mathcal{F}_{t-1}^X))\varepsilon_{t-i}, \qquad (3.19)$$

where $\{\phi_i\}$ and $\{\theta_i\}$ are the parameter processes, $f_i(\mathcal{F}_{t-1}^X), i = 1. \cdots, p$ and $g_i(\mathcal{F}_{t-1}^X), i = 1. \cdots, q$ are functions where $\mathcal{F}_{t-1}^X$ is the σ-algebra generated by $\{X_i, i \leq t-1\}$. Assume that $\{X_t\}$, $\{\varepsilon_t\}$, $\{\phi_i\}$ and $\{\theta_i\}$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

The model described by (3.19) encompasses many time series models that have been proposed in the literature.

(1) If the functions $f_i(\mathcal{F}_{t-1}^X)$, $i = 1, \cdots, p$ and $g_i(\mathcal{F}_{t-1}^X)$, $i = 1, \cdots, q$, and the sequences $\{\phi_i\}$ and $\{\theta_i\}$ are constants then the model (3.19) is an ARMA model.

(2) If the functions $f_i(\mathcal{F}_{t-1}^X)$, $i = 1, \cdots, p$ are constants and $\theta_i = 0, i = 1, \cdots, q$, while the process $\{\phi_i\}$ consists of iid variables and independent of $\{\varepsilon_t\}$ and $\{X_{t-i}\}$, then the model (3.19) is a RCA model of order p.

Some other forms of (3.19) and the processes of $\{\phi_t\}$ are introduced by Tjøstheim (1986), Pourahmadi(1986), Karlson(1990), and Holst(1994).
Another form of (3.19) is AR(1)-MA(q) doubly stochastic process satisfying

$$\begin{aligned} X_t &= \phi_t X_{t-1} + \varepsilon_t, \\ \phi_t &= \phi + e_t + b_1 e_{t_1} + \cdots + b_q e_{t-q}, \end{aligned} \tag{3.20}$$

where ϕ, $b_1, \cdots, b_q$ are real constants, $\{\phi_t\}$, $\{\varepsilon_t\}$, $\{e_t\}$ are random sequences defined on the common probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and$\{\varepsilon_t\}$ and $\{e_t\}$ are mutually independent noise processes with zero means and the finite variances σ_ε^2, σ_e^2 respectively.
Lu(1998) considered this AR(1)-MA(q) doubly stochastic model (3.20) and obtained the conditions for the existence of higher-order stationary solutions.

We write the model (3.19) in a state space form.
Define the matrices

$$\mathbf{A} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_q \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

of order $(p \times p)$, $(p \times q)$ respectively, the vector $\mathbf{H}'$=$(1, 0, \cdots, 0)$ of $(1 \times p)$, the random vectors

$$\mathbf{f}'(\mathcal{F}_{t-1}^X) = (f_1(\mathcal{F}_{t-1}^X), \cdots, f_p(\mathcal{F}_{t-1}^X)),$$

$$\mathcal{X}_t' = (X_t, X_{t-1}, \cdots, X_{t-p+1}) \text{ of } (1 \times p),$$

$$\mathbf{g}'(\mathcal{F}_{t-1}^X) = (g_1(\mathcal{F}_{t-1}^X), \cdots, g_q(\mathcal{F}_{t-1}^X))$$

and $\mathcal{E}_t' = (\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-q+1})$ of order $(1 \times q)$. Then the state space representation for the model (3.19) is

$$\mathcal{X}_t = \mathbf{A} \text{ diag}(\mathcal{X}_{t-1}\mathbf{f}'(\mathcal{F}_{t-1}^X)) + \mathbf{B} \text{ diag}(\mathcal{E}_{t-1}\mathbf{g}'(\mathcal{F}_{t-1}^X)) + \mathbf{H}\varepsilon_t, \tag{3.21}$$

$$X_t = \mathbf{H}'\mathcal{X}_t. \tag{3.22}$$

Example 3.3. Consider a doubly stochastic model with order (1,0) given by

$$X_t = \phi X_{t-1}^2 + \varepsilon_t + \theta X_{t-1}\varepsilon_{t-1}, \tag{3.23}$$

where ε_t is a sequence of iid standard normal variates. State space form of (3.23) is:

$$\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \mathbf{A} \begin{pmatrix} X_{t-1}^2 \\ X_{t-2}^2 \end{pmatrix} + \mathbf{B} \begin{pmatrix} \varepsilon_{t-1} \\ 1 \end{pmatrix} X_{t-1} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \varepsilon_t$$

and

$$X_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} \phi & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} \theta & 0 \\ 0 & 1 \end{pmatrix}$.

The graphical representation of simulated series of the model (3.23) with $\phi = 0.2$ and $\theta = 0.1$ is given in Figure 9. The *acf* and the *pacf* are also given in Figures 10 and 11 respectively.

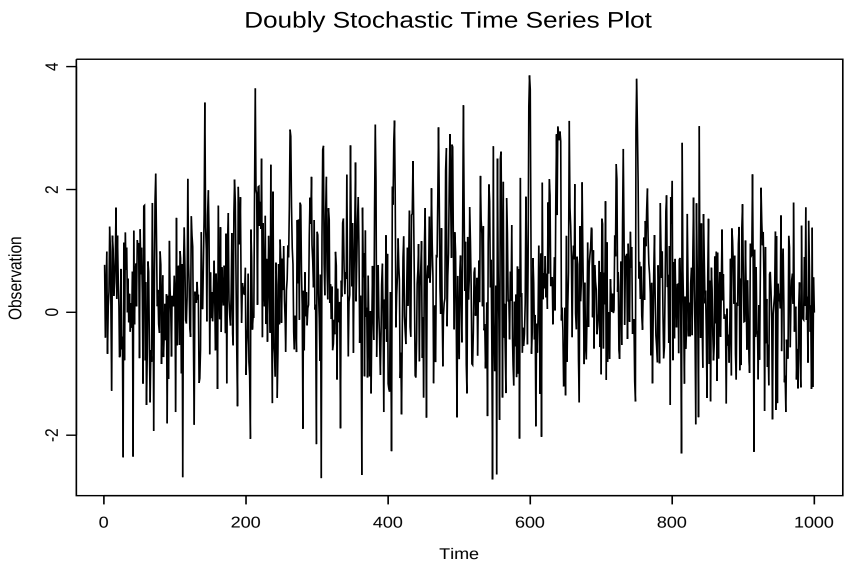


FIGURE 9. Plot of doubly stochastic(1,1) ($\phi = 0.2$ and $\theta = 0.1$).

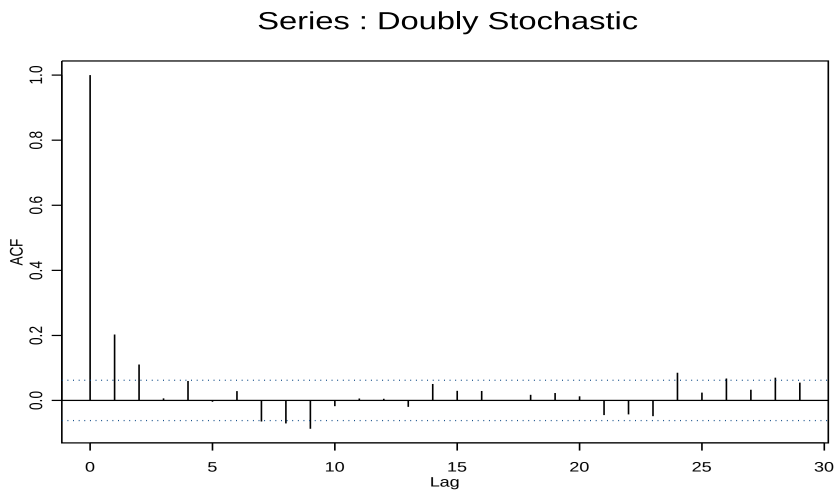


FIGURE 10. acf plot of doubly stochastic(1,1) ($\phi = 0.2$ and $\theta = 0.1$).

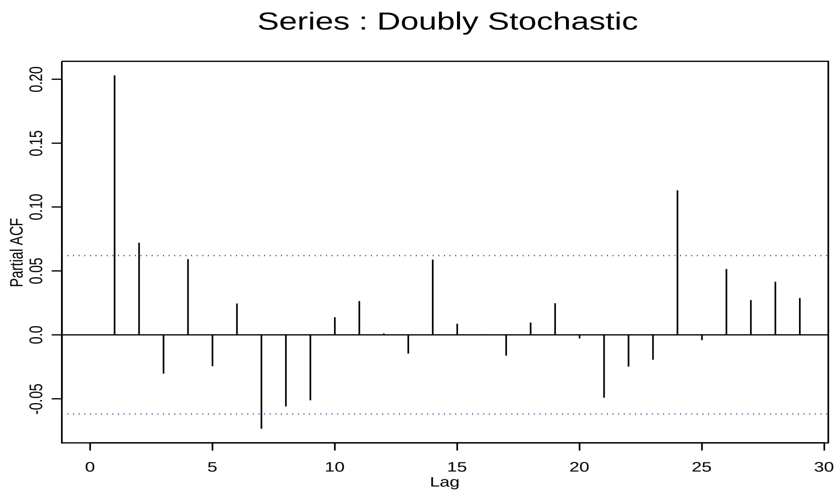


FIGURE 11. pacf plot of doubly stochastic(1,1) ($\phi = 0.2$ and $\theta = 0.1$).

Section 3.4 considers the class of threshold autoregressive models.

3.4. Threshold Autoregressive (TAR) Models

The TAR model was introduced by Tong (1978) as an alternative model for describing piecewise time series, which have been further developed by Tong (1983, 1990) and Tsay (1989). These models have several characteristics that can be used for general judgement and identification such as periodic cycling, up-and-down oscillations and sudden rises or falls. These characteristics are difficult to describe by the conventional linear ARIMA models.

Tong (1978) defined a time series $\{X_t\}$ as the $TAR1(l; k_1, \cdots, k_l)$ if it can be given by the following piecewise linear sub models,

$$X_t = \begin{cases} \phi_{01} + \sum_{i=1}^{k_1} \phi_{i1} X_{t-i} + \varepsilon_t, & \text{if } r_0 < X_{t-d} < r_1, \\ \phi_{02} + \sum_{i=1}^{k_2} \phi_{i2} X_{t-i} + \varepsilon_t, & \text{if } r_1 < X_{t-d} < r_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ \phi_{0l} + \sum_{i=1}^{k_l} \phi_{il} X_{t-i} + \varepsilon_t, & \text{if } r_{l-1} < X_{t-d} < r_l, \end{cases} \tag{3.24}$$

where, $k_1, k_2, \cdots, k_l$ are orders of l AR models; $\{\phi_{ij}\}(i = 1, \cdots, k_j, j = 1, \cdots, l)$ are AR coefficients; $\{\varepsilon_t\}$ is a stationary white noise sequence with zero mean and finite variances σ^2. The values $r_i(i = 0, 1, \cdots, l)$ are called the threshold values with $-\infty = r_0 < r_1 < \cdots < r_l = \infty$ and $d(\in \mathbb{Z}^+)$ is the delay (or lag) parameter (see Tong (1983), p.54) (see also Tsay(1989), Chan (1993), Chen (1998) and Wu and Chang (2002)). These models are also called self-existing threshold autoregressive models (SETAR).

If $D_j = (r_{j-1}, r_j], j = 1, \cdots, l$ and $\mathbb{R} = \bigcup_{j=1}^{l} D_j$ then (3.24) can be written as

$$X_t = \sum_{j=1}^{l} \left(\phi_{0j} + \sum_{i=1}^{k_j} \phi_{ij} X_{t-i} \right) I(X_{t-d} \in D_j) + \varepsilon_t, \tag{3.25}$$

where $I(\cdot)$ denotes the indicator function.

A state space representation of the model (3.25) is

$$\begin{aligned} \mathcal{X}_t &= \sum_{j=1}^{l} \mathbf{A}_j \mathcal{X}_{t-1} I(X_{t-d} \in D_j) + \mathbf{H}\varepsilon_t, & (3.26) \\ X_t &= \mathbf{H}' \mathcal{X}_t, & (3.27) \end{aligned}$$

where

$$\mathbf{A_j} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \phi_{0j} & \phi_{1j} & \phi_{2j} & \cdots & \phi_{k_j j} & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

are of order $(k_j + 2) \times (k_j + 2)$, the vector $\mathbf{H}'=(0, 1, 0, \cdots, 0)$ and the random vector are $\mathcal{X}_t{}'=(1, X_t, X_{t-1}, \cdots, X_{t-k_j})$ of $1 \times (k_j + 2)$.
Chen (1998) has introduced a two-regime generalized threshold autoregressive (GTAR) model as

$$X_t = \begin{cases} \phi_{01} + \sum_{i=1}^{k_1} \phi_{i1} X_{t-i} + \sum_{i=1}^{k} \theta_{i1} Y_{it} + \varepsilon_{1t}, & \text{if } Y_{i,t-d} \le r, \\ \phi_{02} + \sum_{i=1}^{k_2} \phi_{i2} X_{t-i} + \sum_{i=1}^{k} \theta_{i2} Y_{it} + \varepsilon_{2t}, & \text{if } Y_{i,t-d} > r, \end{cases} \tag{3.28}$$

where r is a real number and d is the threshold lag of the model.The sequences $\{\varepsilon_{it}\}, i = 1, 2$ are iid normal variates with zero mean and

variances $\sigma_i^2, i = 1, 2$ respectively and $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ are independent. $\{Y_{1,t}, \cdots, Y_{k,t}\}$ denotes the exogenous variables in regime i. This GTAR model (3.28) is sufficiently flexible to accommodate some practical models. For example, if the exogenous variables
$\{Y_{1t}, \cdots, Y_{kt}\}$ are deleted then it is reduced to a TAR model.

Example 3.4. Consider a TAR(2;1,1) model

$$X_t = \begin{cases} \phi_0 + \phi_1 X_{t-1} + \varepsilon_t & \textit{if } X_{t-1} \geq 0, \\ \phi_0 - \phi_1 X_{t-1} + \varepsilon_t & \textit{otherwise,} \end{cases}$$

where ε_t is a sequence of iid standard normal variates.
State space form of (3.29) is:

$$\begin{pmatrix} 1 \\ X_t \\ X_{t-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \phi_0 \\ 0 \end{pmatrix} + \mathbf{A} \begin{pmatrix} 1 \\ X_{t-1} \\ X_{t-2} \end{pmatrix} I_{X_{t-1}} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \varepsilon_t$$

and

$$X_t = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ X_t \\ X_{t-1} \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and the indicator function

$$I_{X_t} = \begin{cases} 1 & \text{if} \quad X_t \geq 0, \\ -1 & \text{if} \quad X_t < 0. \end{cases}$$

The graphical representation of simulated series of the model (3.29) with $\phi_0 = 1$ and $\phi_1 = 0.8$, is given in Figure 12. The *acf* and the *pacf* are given in Figures 13 and 14 respectively.

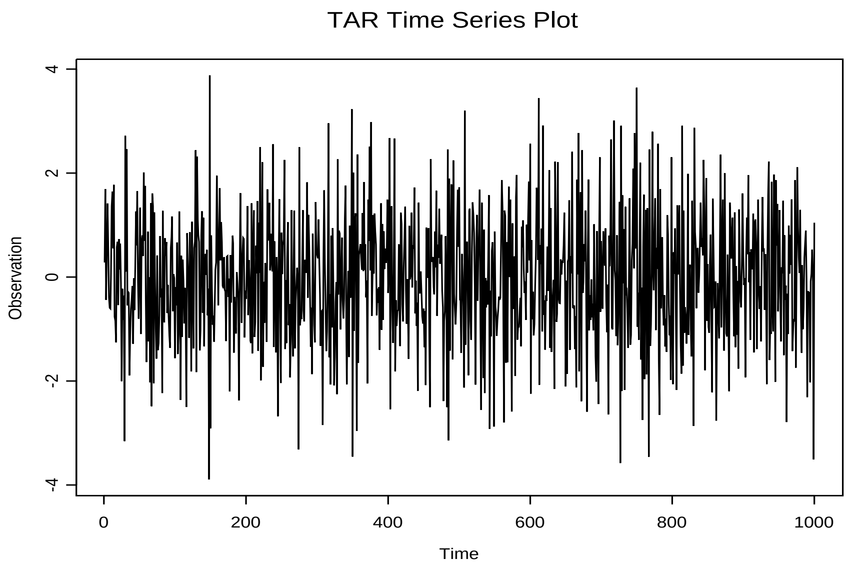


FIGURE 12. Plot of TAR(2;1,1) ($\phi_0 = 1$ and $\phi_1 = 0.8$).

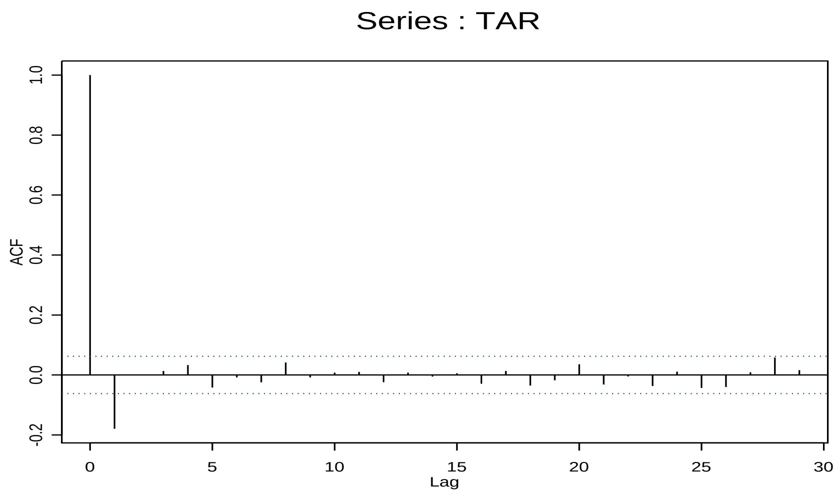


FIGURE 13. acf plot of TAR(2;1,1) ($\phi_0 = 1$ and $\phi_1 = 0.8$).

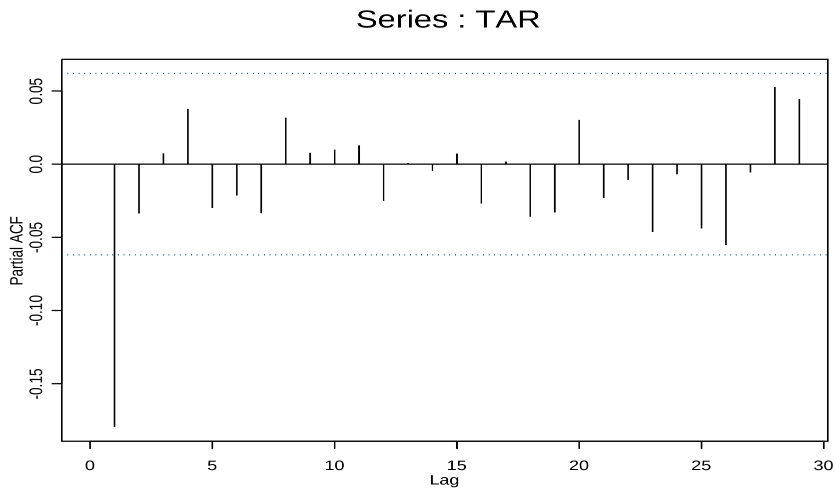


FIGURE 14. pacf plot of TAR(2;1,1) ($\phi_0 = 1$ and $\phi_1 = 0.8$).

Next we consider autoregressive conditional heteroscedasticity models as this plays a significant role in financial modelling.

3.5. Autoregressive Conditional Heteroscedasticity (ARCH) Models

The class of ARCH model, introduced by Engle (1982), is very useful for modelling economic data processing high volatility. In the last two decades, statistical methodologies for time series models with ARCH errors have been developed. Weiss (1994) considered a class of ARMA models with ARCH errors and studied statistical inference for those models (see also Lee and Hanson (1994) and Lumsdaine(1996)).

The ARCH(p) model for a sequence of random variables $\{X_t\}$ on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ satisfies the following random difference equations:

$$\begin{aligned} X_t &= \varepsilon_t \sigma_t, \\ \sigma_t^2 &= \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2, \end{aligned} \tag{3.29}$$

where $\phi_0 > 0$, $\phi_i \geq 0, i = 1, \cdots, p$, $\{\varepsilon_t\}$ is an iid white noise defined on the same probability space and ε_t independent of $\{X_s, s < t\}$. Clearly this model (3.29) can be written as (Cox, Hinkley and Barndorff-Nielsen(1996), p.8)

$$\begin{aligned} X_t &= \varepsilon_t \sigma_t, \\ X_t^2 = \sigma_t^2 + (X_t^2 - \sigma_t^2) &= \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2 + v_t, \end{aligned} \tag{3.30}$$

where $v_t = X_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1)$ is a martingale difference and ε_t is Gaussian with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$. On the second order stationarity, Engle (1982) obtained the necessary and sufficient conditions in the usual way. i.e. all the roots of the characteristic polynomial

$$\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p \tag{3.31}$$

lie inside the unit circle. This condition is equivalent to

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1. \tag{3.32}$$

This basic univariate ARCH model has been extended in a number of directions. Using the fact that ARCH(p) model is an AR(p) model for the squared variables of the series as in (3.30), the state space representation of (3.29) is given by

$$\mathcal{X}_t = \mathbf{A}\mathcal{X}_{t-1} + \mathbf{H}v_t, \tag{3.33}$$

$$X_t^2 = \mathbf{H}'\mathcal{X}_t, \tag{3.34}$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_p & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

of order $(p+2)\times(p+2)$, the vector $\mathbf{H}'$=$(0, 1, 0, \cdots, 0)$, and the random vector $\mathcal{X}_t'$=$(1, X_t^2, X_{t-1}^2, \cdots, X_{t-p}^2)$ of $1 \times (p+2)$.

Example 3.5. Consider a ARCH(1) model given by

$$X_t = \sigma_t\varepsilon_t,$$

$$\sigma_t^2 = \phi_0 + \phi_1 X_{t-1}^2, \tag{3.35}$$

where ε_t is a sequence of iid standard normal variates. State space form of (3.35) is

$$\begin{pmatrix} 1 \\ X_t^2 \\ X_{t-1}^2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} 1 \\ X_{t-1}^2 \\ X_{t-2}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} v_t$$

and

$$X_t = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ X_t \\ X_{t-1} \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ \phi_0 & \phi_1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

and $v_t = X_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1)$ is a martingale difference.

The graphical representation of a simulated series of the model (3.35) with $\phi_0 = 1$ and $\phi_1 = 0.8$ is given in Figure 15. The *acf* and the *pacf* are given in Figures 16 and 17 respectively.

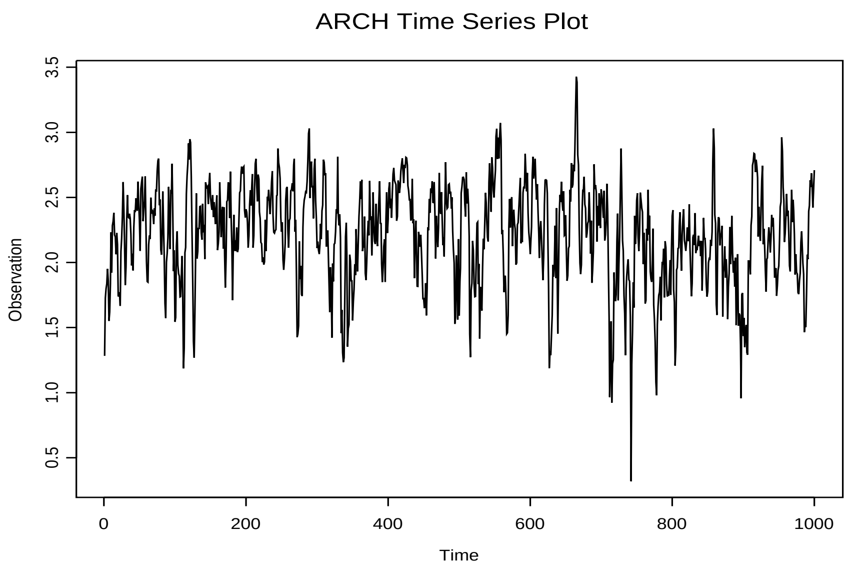


FIGURE 15. Plot of ARCH(1) ($\phi_0 = 1$ and $\phi_1 = 0.8$).

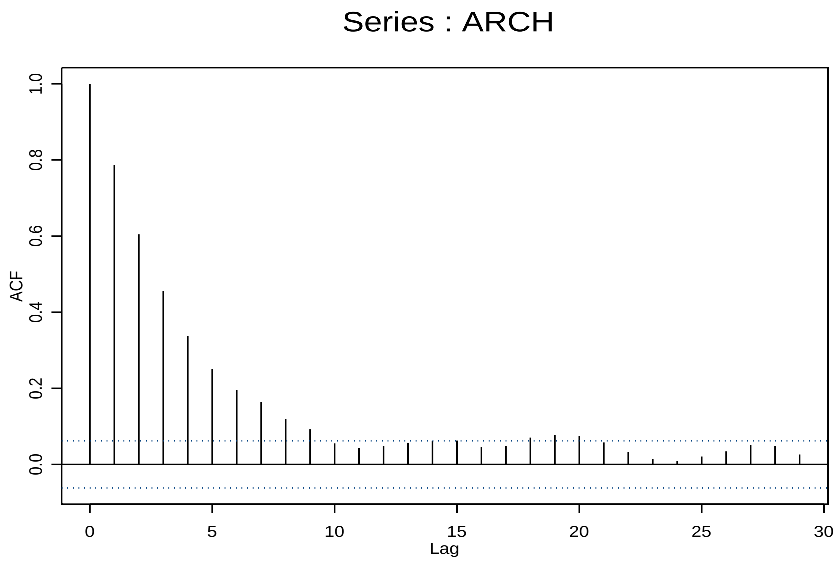


FIGURE 16. acf plot of ARCH(1) ($\phi_0 = 1$ and $\phi_1 = 0.8$).

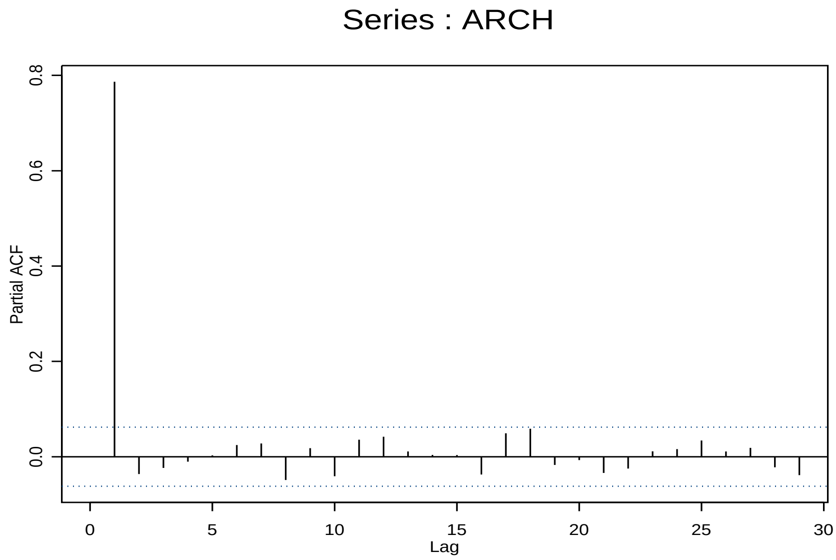


FIGURE 17. pacf plot of ARCH(1) ($\phi_0 = 1$ and $\phi_1 = 0.8$).

Bollerslev (1986) extends the class of ARCH models to include moving average terms and this is called the class of generalized ARCH (GRACH) (also see Taylor (1986)). This class is generated by (3.29) with σ_t^2 is generated by

$$\sigma_t^2 = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2 + \sum_{i=1}^{q} \theta_i \sigma_{t-i}^2 \ . \tag{3.36}$$

(See Tong (1990), p.116) Let $v_t = X_t^2 - \sigma_t^2$ be the martingale difference with variance σ_v^2. Following Cox, Hinkley and Barndorff-Nielsen(1996), equation (3.36) together with $X_t = \sigma_t \varepsilon_t$ can be written as

$$\begin{aligned}
X_t^2 - v_t &= \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2 + \sum_{j=1}^{q} \theta_j \sigma_{t-j}^2 \\
X_t^2 - \sum_{i=1}^{p} \phi_i X_{t-i}^2 &= \phi_0 + v_t + \sum_{j=1}^{q} \theta_j (X_{t-j}^2 - v_{t-j}) \\
X_t^2 - \sum_{i=1}^{p} \phi_i X_{t-i}^2 - \sum_{j=1}^{q} \theta_j X_{t-j}^2 &= \phi_0 + v_t - \sum_{j=1}^{q} \theta_j v_{t-j} \\
\alpha(B) X_t^2 &= \phi_0 + \theta(B) v_t, \qquad (3.37)
\end{aligned}$$

where $\alpha(B) = 1 - \sum_{i=1}^{r} \alpha_i B^i$, $\alpha_i = \phi_i + \theta_i$, $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ and $r = max(p, q)$.

It is clear that $\{X_t^2\}$ has an ARMA(r,q) representation with the following assumptions:

(A.1) all the roots of the polynomial $\alpha(B)$ lie outside of the unit circle.

(A.2) $\sum_{i=0}^{\infty} \psi_i^2 < \infty$, where the $\psi_i's$ are obtained from the relation $\psi(B)\alpha(B) = \theta(B)$ with $\psi(B) = 1+\sum_{i=1}^{\infty} \psi_i B^i$.

These assumptions ensure that the X_t^2 process is weakly stationary. In this case, the autocorrelation function of X_t^2 will be exactly the same as that for a stationary ARMA(r, q) model.

Consider the GARCH(p,q) of the form of ARMA(r,q) of X_t^2 given in (3.37)

$$X_t^2 - \sum_{i=1}^{r} \alpha_i X_{t-i}^2 = \phi_0 + v_t - \sum_{i=1}^{q} \theta_i v_{t-i} \ . \qquad (3.38)$$

The state space representation of (3.38) is

$$\mathcal{X}_t = \mathbf{A}\mathcal{X}_{t-1} + \mathbf{B}\mathcal{V}_t \quad (3.39)$$

$$X_t = \mathbf{H}'\mathcal{X}_t, \quad (3.40)$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \phi_0 & \alpha_1 & \cdots & \alpha_r & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & -\theta_1 & \cdots & -\theta_q & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

are matrices of order $(r+2)\times(r+2)$ and $(r+2)\times(q+2)$ respectively; the vector $\mathbf{H}' = (0, 1, 0, \cdots, 0)$ of order $1\times(r+2)$, the random vector $\mathcal{X}'_t = (1, X_t^2, X_{t-1}^2, \cdots, X_{t-r}^2)$ of $1 \times (r+2)$ and $\mathcal{V}'_t = (v_t, v_{t-1}, \cdots, v_{t-q-1})$ of order $1 \times (q+2)$.

Example 3.6. Consider a GARCH(1,1) model

$$X_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \phi_0 + \phi_1 X_{t-1}^2 + \theta_1 \sigma_{t-1}^2 \quad (3.41)$$

where ε_t is a sequence of IID standard normal variates. State space form of (3.41) is

$$\begin{pmatrix} 1 \\ X_t^2 \\ X_{t-1}^2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} 1 \\ X_{t-1}^2 \\ X_{t-2}^2 \end{pmatrix} + \mathbf{B} \begin{pmatrix} v_t \\ v_{t-1} \\ v_{t-2} \end{pmatrix}$$

$$\text{and } X_t^2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ X_t^2 \\ X_{t-1}^2 \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ \phi_0 & \phi_1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & \theta_1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

The graphical representation of simulated series of the model (3.41) with $\phi_0 = 1$, $\phi_1 = 0.1$ and $\theta_1 = 0.1$ is given in Figure 18. The *acf* and the *pacf* are given in Figures 19 and 20 respectively.

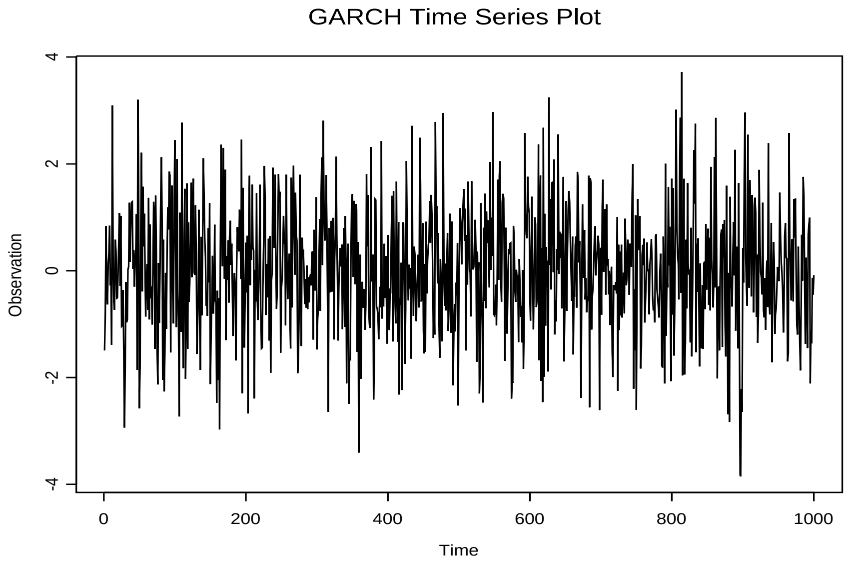


FIGURE 18. Plot of GARCH(1,1) ($\phi_0 = 1$, $\phi_1 = 0.1$ and $\theta_1 = 0.1$).

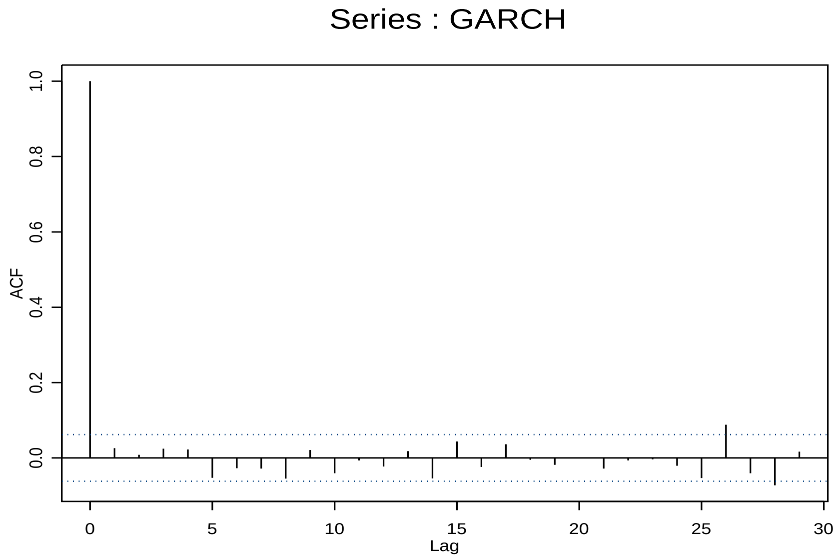


FIGURE 19. acf plot of GARCH(1,1) ($\phi_0 = 1$, $\phi_1 = 0.1$ and $\theta_1 = 0.1$).

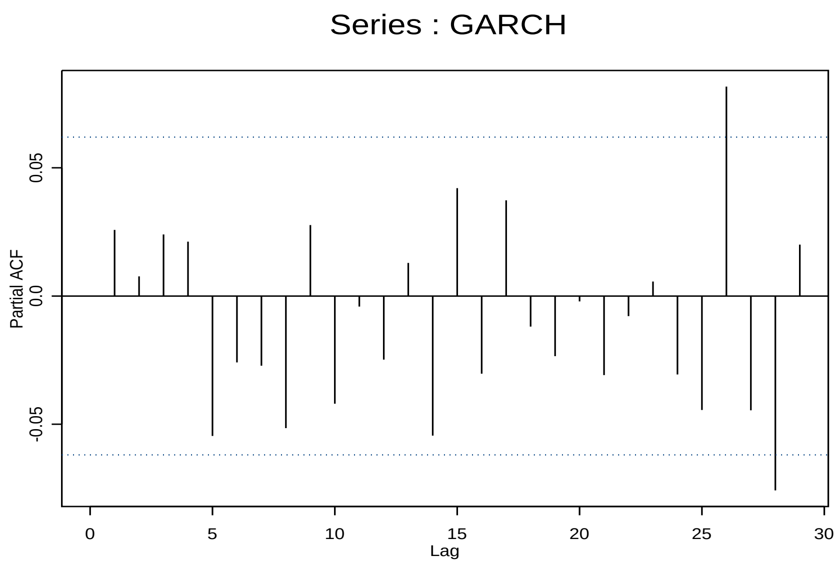


FIGURE 20. pacf plot of GARCH(1,1) ($\phi_0 = 1$, $\phi_1 = 0.1$ and $\theta_1 = 0.1$).

Another extension of this class is called AR-ARCH models. Ha and Lee (2002) considered an RCA model for $\{X_t\}$ of the form with $\{\varepsilon_t\}$ following an ARCH(p) model satisfying

$$X_t = (\phi + \beta(t))X_{t-1} + \varepsilon_t, \quad |\phi| < 1, \tag{3.42}$$

where $\beta(t)$ are iid random variables and $\{\varepsilon_t\}$ is an ARCH(p) process, such that

(1) $\varepsilon_t = \eta_{t-1}\xi_t, \quad \eta_{t-1}^2 = \phi_0 + \phi_1\varepsilon_{t-1}^2 + \cdots + \phi_p\varepsilon_{t-p}^2$
(ξ_t are iid random variables with zero mean and unit variance, $\phi_0 > 0$, $\phi_i \geq 0, i = 1, \cdots, p$, and ξ_t is independent of $\{\eta_s, s < t\}$,

(2) $\{\beta(t)\}$ and $\{\xi_t\}$ are independent iid random variables with $E[\beta(t)] = 0,\ E[\beta(t)^2] = \sigma_\beta^2 > 0$ and $E\xi_t^4 < \infty$.

The process generated by (3.42) is called the AR-ARCH process. Diebolt and Guégan (1991) considered the class of β-ARCH model given by

$$X_t = \phi X_{t-1} + (\phi_0 + \phi_1 X_{t-1}^{2\beta})\varepsilon_t, \tag{3.43}$$

where $\{\varepsilon_t\}$ is a sequence of iid random variables with zero mean and finite variance and $\phi, \phi_0, \phi_1, \beta$ are the parameters.

To illustrate a state space representation of β-ARCH model,let

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 & 0 \\ \phi_0 & \phi_1 \end{pmatrix},$$

$\mathcal{X}_t' = (1, X_t)$ and $\mathbf{H}' = (1, 0)$. The state space representation of (3.43) is

$$\begin{aligned} \mathcal{X}_t &= \mathbf{A}\mathcal{X}_{t-1} + \mathbf{B}\mathcal{X}_{t-1}^{2\beta}\varepsilon_t \\ X_t &= \mathbf{H}'\mathcal{X}_t, \end{aligned}$$

where $\mathcal{X}_t^{2\beta} = (1, X_t^{2\beta})'$.

Kurtosis is important in GARCH modelling and below we consider some associated results.

3.5.1. Kurtosis of GARCH Models. We now consider the kurtosis of the GARCH(p, q) model for the calculation of the fourth order moment of the noise which will be used in the following:

Definition 3.1. For any random variable X with finite fourth moments, the kurtosis is defined by $\dfrac{E(X-\mu)^4}{[Var(X)]^2}$.

If the process $\{\varepsilon_t\}$ is normal then the process $\{X_t\}$ defined by equations (3.36) together with $X_t = \sigma_t \varepsilon_t$ is called a normal GARCH (p, q) process. Denote the kurtosis of the GARCH process by $K^{(X)}$. In order to calculate the kurtosis $K^{(X)}$ in terms of the ψ weights, we have the following theorem.

Theorem 3.1. For the GARCH(p, q) process specified by (3.38), under the stationarity assumptions and finite fourth moment, the kurtosis $K^{(X)}$ of the process is given by:

a) $K^{(X)} = \dfrac{E(\varepsilon_t^4)}{E(\varepsilon_t^4) - [E(\varepsilon_t^4) - 1] \sum\limits_{j=0}^{\infty} \psi_j^2}$,

b) (i) The variance of the X_t^2 process is $\gamma_0^{X^2} = \sigma_v^2 \sum\limits_{j=0}^{\infty} \psi_j^2$,

(ii) The $k - lag$ autocovariance of the X_t^2 process is

$$\gamma_k^{X^2} = \sigma_v^2 \sum_{j=0}^{\infty} \psi_{k+1} \psi_j \text{ and for } k \geq 1,$$

(iii) The $k - lag$ autocorrelation is given by $\rho_k^{X^2} = \dfrac{\sum\limits_{j=0}^{\infty} \psi_{k+j} \psi_j}{\sum\limits_{j=0}^{\infty} \psi_j^2}$.

c) For a normal GARCH(p, q) process $K^{(X)} = \dfrac{3}{3 - 2 \sum\limits_{j=1}^{\infty} \psi_j^2}$.

Proof:

(a)

The variable ε_t is assumed to have finite fourth moments, mean zero and variance one, and hence by (3.36) one has

$$E(X_t) = E(\sigma_t \varepsilon_t) = \sigma_t E(\varepsilon_t) = 0. \tag{3.44}$$

and $v_t = X_t^2 - \sigma_t^2$.

Stationary process in (3.38) can be written as an infinite MA process $X_t^2 = \sum_{j=0}^{\infty} \psi_j v_{t-j}$ and hence (with $psi_0 = 1$) we have (see the example(2.1)).

$$\begin{aligned} \mathrm{var}(X_t^2) &= \sigma_v^2[1 + \psi_1^2 + \psi_2^2 + \cdots], && (3.45) \\ \sigma_v^2 &= E(X_t^4) - E(\sigma_t^4) && (3.46) \\ &= E(\sigma_t^4 \varepsilon_t^4) - E(\sigma_t^4) \\ &= E(\sigma_t^4)E(\varepsilon_t^4) - E(\sigma_t^4) \\ &= E(\sigma_t^4)[E(\varepsilon_t^4) - 1], && (3.47) \\ \mathrm{var}(X_t^2) &= E(\sigma_t^4)[E(\varepsilon_t^4) - 1][1 + \psi_1^2 + \psi_2^2 + \cdots]. && (3.48) \end{aligned}$$

Moreover, from (3.36) it follows that,

$$\begin{aligned} \mathrm{var}(X_t^2) &= E(X_t^4) - [E(X_t^2)]^2 \\ &= E(\varepsilon_t^4 \sigma_t^4) - [E(\sigma_t^2)]^2 \\ &= E(\varepsilon_t^4)E(\sigma_t^4) - [E(\sigma_t^2)]^2. && (3.49) \end{aligned}$$

Equating (3.48) and (3.49) one has

$$E(\varepsilon_t^4)E(\sigma_t^4) - [E(\sigma_t^2)]^2 = E(\sigma_t^4)[E(\varepsilon_t^4) - 1][1 + \psi_1^2 + \psi_2^2 + \cdots]. \tag{3.50}$$

Now

$$
\begin{aligned}
K^{(X)} &= \frac{E(X_t^4)}{[E(X_t^2)]^2} \\
&= \frac{E(\varepsilon_t^4)E(\sigma_t^4)}{[E(\sigma_t^2\varepsilon_t^2)]^2} \\
&= \frac{E(\varepsilon_t^4)E(\sigma_t^4)}{[E(\sigma_t^2)]^2}. \qquad (3.51)
\end{aligned}
$$

From (3.50),

$$
\frac{E(\sigma_t^4)}{[E(\sigma_t^2)]^2} = \frac{1}{E(\varepsilon_t^4) - [E(\varepsilon_t^4) - 1][1 + \psi_1^2 + \psi_2^2 + \cdots]} \qquad (3.52)
$$

and hence (3.51) and (3.52) complete the proof of part(a) (see Bai, Russell and Tiao (2003) and Thavaneswaran, Appadoo and Samanta (2004)).

(b) (i)

$$
\begin{aligned}
\text{var}(X_t^2) &= \text{var}(\psi(B)v_t) \\
&= \text{var}(\sum_{j=0}^{\infty} \psi_j v_{t-j}) \\
&= \sum_{j=0}^{\infty} \psi_j^2 \text{var}(v_{t-j}) \\
\text{i.e.,} \gamma_0^{X^2} &= \sigma_v^2 \sum_{j=0}^{\infty} \psi_j^2. \qquad (3.53)
\end{aligned}
$$

(ii) Using the definition of autocovariance,

$$\begin{aligned}
\gamma_k^{X^2} &= \text{cov}(X_{t+k}^2, X_t^2) \\
&= \text{cov}[\psi(B)v_{t+k}, \psi(B)v_t] \\
&= \text{cov}[\sum_{j=0}^{\infty} \psi_j v_{t+k-j}, \sum_{j=0}^{\infty} \psi_j v_{t-j}] \\
&= \sum_{j=0}^{\infty} \psi_{k+j}\psi_j \text{var}(v_{t-j}).
\end{aligned}$$

$$\text{i.e.} \qquad \gamma_k^{X^2} = \sigma_v^2 \sum_{j=0}^{\infty} \psi_{k+j}\psi_j. \tag{3.54}$$

(iii) Using the definition of the autocorrelation,

$$\begin{aligned}
\rho_k^{X^2} &= \text{Corr}(X_{t+k}^2, X_t^2) \\
&= \frac{\text{cov}(X_{t+k}^2, X_t^2)}{\text{var}(X_t^2)} \\
&= \frac{\gamma_k^{X^2}}{\gamma_0^{X^2}}.
\end{aligned}$$

$$\text{i.e.} \qquad \rho_k^{X^2} = \frac{\sum_{j=0}^{\infty} \psi_{k+j}\psi_j}{\sum_{j=0}^{\infty} \psi_j^2}. \tag{3.55}$$

(c) Proof of this part follows from the fact that for a standard normal variate $E(\varepsilon_t^4) = 3$ and part (a) (see Franses and Van Dijk (2000) p.146).

Example 3.7. Consider GARCH(1,1) process

$$X_t = \sigma_t \varepsilon_t, \;\; \sigma_t^2 = \phi X_t^2 + \theta \sigma_{t-1}^2. \tag{3.56}$$

Let $v_t = X_t^2 - \sigma_t^2$ be the martingale difference with variance σ_v^2, then (3.56) can be rewritten as an ARMA(1,1) model of X_t^2. Consider the equation for ARMA(1,1) process,

$$X_t^2 - \alpha X_{t-1}^2 = v_t - \beta v_{t-1},$$

where $\alpha = \phi + \theta$ and $\beta = -\theta$. Using the ψ-weights in the example (2.1) and the variance formula in (3.53),

$$\begin{aligned}
\gamma_0 &= \sigma_v^2 \sum_{j=0}^{\infty} \psi_j^2 &(3.57)\\
&= \sigma_v^2 (1 + \psi_1^2 + \psi_2^2 + \cdots + \cdots) \\
&= \sigma_v^2 (1 + (\alpha + \beta)^2 \left[1 + \alpha^2 + \alpha^4 + \alpha^6 + \cdots . \right]) \\
&= \sigma_v^2 \left[\frac{1 + 2\alpha\beta + \beta^2}{1 - \alpha^2} \right] \\
&= \sigma_v^2 \left[\frac{1 - 2\phi\theta - \theta^2}{1 - (\phi + \theta)^2} \right]. &(3.58)
\end{aligned}$$

Similarly using the autocovariance formula in (3.54), the lag 1 auto covariance is

$$\begin{aligned}
\gamma_1 &= \sigma_v^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} \\
&= \sigma_v^2 [\psi_0 \psi_1 + \psi_1 \psi_2 + \cdots] \\
&= \frac{\sigma_v^2 (\alpha + \beta)(1 + \alpha\beta)}{1 - \alpha^2} \\
&= \frac{\sigma_v^2 \phi (1 - \phi\theta - \theta^2)}{1 - (\phi + \theta)^2} &(3.59)
\end{aligned}$$

for $k \geq 2$, $\gamma_k = \alpha\gamma_{k-1}$.

i.e.

$$\rho_k = \begin{cases} 1 & k = 0 \\ \dfrac{(\alpha+\beta)(1+\alpha\beta)}{1+2\alpha\beta+\beta^2} & k = 1 \\ \alpha\rho_{k-1} & k \geq 1 \end{cases}$$

$$= \begin{cases} 1 & k = 0 \\ \dfrac{\phi(1-\phi\theta-\theta^2)}{1-2\phi\theta-\theta^2} & k = 1 \\ (\phi+\theta)\rho_{k-1} & k \geq 1 \end{cases} \tag{3.60}$$

Example 3.8. Consider the ARCH(1) model of the form (p=1 in (3.29) and $\phi_0 = 0$*)*

$$X_t = \varepsilon_t\sigma_t, \qquad \sigma_t^2 = \phi X_{t-1}^2. \tag{3.61}$$

Let $v_t = X_t^2 - \sigma_t^2$ be the martingale difference with variance σ_v^2, then

$$\begin{aligned} X_t^2 - v_t &= \phi X_{t-1}^2 \\ X_t^2 &= \phi X_{t-1}^2 + v_t. \end{aligned}$$

To calculate the $\psi-$weights, use the following relationship

$$(1-\phi B)(1+\psi_1 B+\psi_2 B+\cdots) = 1,$$

and get

$$\psi_j = \phi^j, j = 1, 2, \cdots.$$

By part (c) of the theorem (3.1), the kurtosis

$$K^{(X)} = \frac{3}{3 - 2\sum_{j=1}^{\infty} \psi_j^2} = \frac{3(1-\phi^2)}{3-5\phi^2}.$$

Example 3.9. Consider the normal GARCH(1,1) model of the form (3.56) and from the example (2.1)

$$\psi_1 = \alpha + \beta = \phi$$

$$\psi_j = (\alpha+\beta)\alpha^{j-1} = \phi(\phi+\theta)^{j-1}, \ j = 1, 2, \cdots .$$

$$\sum_{j=1}^{\infty} \psi_j^2 = \frac{\phi^2}{1-(\phi+\theta)^2}.$$

By part (c) of the theorem (3.1), the kurtosis

$$K^{(X)} = \frac{3}{3 - 2\sum\limits_{j=1}^{\infty} \psi_j^2} = \frac{3[1-(\phi+\theta)^2]}{3-5(\phi+\theta)^2} .$$

Chapter 4 considers an approach to parameter estimation for some nonlinear models based on estimating functions developed by Godambe (1985).

CHAPTER 4

Estimating Functions and Applications

4.1. Estimating Functions

In this chapter we obtain optimum estimating functions for nonlinear time series models and discuss the parameter estimation problem in detail.

Let $X_1, \cdots, X_n$ be a finite sample of a time series $\{X_t\}$ and let $\mathcal{F}$ be the class of probability distributions F on $\mathbb{R}^n$. Suppose that $\theta = \theta(F),\ F \in \mathcal{F}$ is a real parameter and write

$$g = g\{X_1, \cdots, X_n; \theta(F)\},\ F \in \mathcal{F}.$$

A function g of this form is called an estimating function. This function is called an unbiased estimating function if

$$E_F[g\{X_1, \cdots, X_n; \theta(F)\}] = 0,\ F \in \mathcal{F}. \tag{4.1}$$

Now we state the following regularity conditions for the class of unbiased estimating functions satisfying (4.1).

Definition 4.1. Any real valued function g of the random variates $X_1, \cdots, X_n$ and the parameter θ satisfying (4.1) is called a *regular unbiased estimating function* if,

(i) g is measurable on the sample space with respect to a measure μ ,

(ii) $\partial g/\partial\theta$ exists for all $\theta = \theta(F)$,

(iii) $\int gpd\mu$ is differentiable under the integral sign, where p is the probability density of the sample.

Denote the standardized form of the estimating function g by g_s. i.e. $g_s = g/E_F(\partial g/\partial\theta)$ and $(g_1 - g_2)_s = (g_1 - g_2)/\left\{E_F(\partial g/\partial\theta) - E_F(\partial g/\partial\theta)\right\}$ (see Godambe (1976), p.278).

Definition 4.2. A regular unbiased estimating function $g^* \in \mathcal{G}$ is said to be optimum if

$$E_F\{(g_s^*)^2\} \leq E_F\{(g_s)^2\}, \ \textit{for all } g \in \mathcal{G} \textit{ and } F \in \mathcal{F}. \tag{4.2}$$

This is equivalent to

$$\frac{E_F\{(g^*)^2\}}{\{E_F(\partial g^*/\partial\theta)\}^2} \leq \frac{E_\theta\{g^2\}}{\{E_F(\partial g/\partial\theta_1)\}^2}, \text{ for all } g \in \mathcal{G} \text{ and for all } F \in \mathcal{F}.$$

We state and prove the following theorem proposed by Godambe (1976) in order to establish the uniqueness of the optimum estimating function.

Theorem 4.1. If $g_1^* \in \mathcal{G}$ and $g_2^* \in \mathcal{G}$ are both optimum, then

$$g_{1s}^* = g_{2s}^* \tag{4.3}$$

Proof: Let

$$E_F\{(g_{1s}^*)^2\} = E_F\{(g_{2s}^*)^2\} = M_F^2. \tag{4.4}$$

Then

$$E_F\{(g_{1s}^* - g_{2s}^*)^2\} = 2M_F^2 - 2E_F(g_{1s}^* g_{2s}^*), \text{ for all } F \in \mathcal{F}. \tag{4.5}$$

Since $g_1^*,\ g_2^* \in \mathcal{G}$, it follows that $g_1^* + g_2^*,\ g_1^* - \ g_2^* \in \mathcal{G}$ provided

$$|\{E_F(\partial g_1^*/\partial\theta)\}| \neq |\{E_F(\partial g_2^*/\partial\theta)\}|.$$

Then from (4.2), we have

$$E_F[\{(g_1^*+g_2^*)_s\}^2] \geq M_F^2,\ E_F[\{(g_1^*-g_2^*)_s\}^2] \geq M_F^2, \text{ for all } F \in \mathcal{F}. \quad (4.6)$$

From (4.4) and (4.6),

$$E_F(g_1^* g_2^*) \geq M_F^2 E_F(\partial g_1^*/\partial\theta) E_F(\partial g_2^*/\partial\theta),$$

$$-E_F(g_1^* g_2^*) \geq -M_F^2 E_F(\partial g_1^*/\partial\theta) E_F(\partial g_2^*/\partial\theta).$$

Hence

$$E_F(g_1^* g_2^*) = M_F^2 E_F(\partial g_1^*/\partial\theta) E_F(\partial g_2^*/\partial\theta), \text{ for all } F \in \mathcal{F}. \quad (4.7)$$

Substituting (4.7) in (4.5) we have

$$E_F[\{(g_1^* - g_2^*)_s\}^2] = 0, \text{ for all } F \in \mathcal{F} \quad (4.8)$$

and this implies (4.3).

Now we consider a way of constructing an estimating function as follows:

Let h_t be a real valued function of $X_1, \cdots, X_t$ and θ such that

$$E_{t-1,F}[h_t\{X_1, \cdots, X_t; \theta(F)\}] = 0\ , t = 1, \cdots, n,\ F \in \mathcal{F}, \quad (4.9)$$

where $E_{t-1,F}(\cdot)$ denotes the expectation holding the first $t-1$ values $X_1, \cdots, X_{t-1}$ fixed. Denote

$$E_{t-1,F}(\cdot) \equiv E_{t-1}(\cdot),\ E_{0,F}(\cdot) \equiv E_F(\cdot) \equiv E(\cdot).$$

For a specific function h_t $(= h_t\{X_1, \cdots, X_t; \theta(F)\})$, $t = 1, \cdots, n$, assume $\mathcal{F}$ to be a class of distributions with property (4.9). In applications of this in stochastic processes a special form of the function h_t is

$$h_t = X_t - E_{t-1}(X_t). \tag{4.10}$$

For example an AR(1) process satisfying $X_t = \theta X_{t-1} + \varepsilon_t$ we have $E_{t-1}(X_t) = \theta X_{t-1}$ and hence $h_t = \varepsilon_t$. The general motivation underlying the transformation $X_t - E_{t-1}(X_t)$ was to obtain variates which if not independent are uncorrelated.This is achieved by computing the residual by subtracting from X_t the effect due to the previously realized values $X_1, \cdots, X_{t-1}$. So the function h_t in (4.9) is (4.10) which gives the (predictive) residual. Property (4.9) implies that

$$E(h_t h_j) = 0,\ t \neq j,\ \text{for all } F \in \mathcal{F} \tag{4.11}$$

and hence the functions $h_t, t = 1, \cdots, n$ are uncorrelated. Assume that h_t is differentiable w.r.t. θ for $t = 1, \cdots, n$. The property of uncorrelatedness has been explored for a long time for asymptotics. We define and establish the finite sample "optimum estimating function" or "optimum estimation" for θ given by the distribution F satisfying (4.9) for specified functions $h_t, t = 1, \cdots, n$. This estimation depends only upon the value

$$\{E_{t-1}(\partial h_t / \partial \theta)\} / E_{t-1}(h_t^2) \tag{4.12}$$

which together with (4.9) corresponds to the two moments underlying the *Gauss-Markov* set-up.

Consider the estimating function g of the form

$$g = \sum_{t=1}^{n} a_{t-1} h_t, \tag{4.13}$$

where the functions $h_t;\ t = 1, \cdots, n$ are as defined in (4.9), a_{t-1} is a function of of the random variates $X_1, \cdots, X_{t-1}$ (with $a_0 = 0$) and the parameter θ for $t = 1, \cdots, n$ and assume that a_{t-1} is differentiable w.r.t. θ for $t = 1, \ldots, n$. We may define various functions a_{t-1} in (4.13) to generate the class $\mathcal{G}$ of estimating functions.

Note that Theorem 4.1 implies that if an optimum estimating function g^* exists, then the estimating equation $g^* = 0$ is unique up to a constant multiple.

Theorem 4.2. If an estimating function $g^* \in \mathcal{G}$ with $E_F\{(g_s^*)^2\} \leq E_F\{(g_s)^2\}$, for all $g \in \mathcal{G}$ and $F \in \mathcal{F}$ (i.e. g^* is optimum estimating function), then the function g^* is given by

$$g^* = \sum_{t=1}^{n} a_{t-1}^* h_t, \tag{4.14}$$

where

$$a_{t-1}^* = \{E_{t-1}(\partial h_t/\partial\theta)\}/E_{t-1}(h_t^2).$$

Proof: From (4.13) and (4.11) we have

$$E(g^2) = E\left\{\sum_{t=1}^{n} a_{t-1}^2 E_{t-1}(h_t^2)\right\} = E(A^2)\ (say) \tag{4.15}$$

and

$$\{E(\partial g/\partial\theta)\}^2 = \left[E\sum_{t=1}^{n}\{a_{t-1}E_{t-1}(\partial h_t/\partial\theta) + (\partial a_{t-1}/\partial\theta)E_{t-1}(h_t)\}\right]^2, \tag{4.16}$$

where $\partial a_{t-1}/\partial \theta$ is exist and $E_{t-1}(h_t) = 0$.

Therefore

$$\{E(\partial g/\partial \theta)\}^2 = \left[E\sum_{t=1}^{n} a_{t-1}E_{t-1}(\partial h_t/\partial \theta)\right]^2 = \{E(B)\}^2 \ (say) \quad (4.17)$$

and from(4.15) and (4.17) we have

$$E(g^2)/\{E(\partial g/\partial \theta)\}^2 = E(A^2)/\{E(B)\}^2. \quad (4.18)$$

By Cauchy-Schwarz inequality, (4.18) gives

$$E(g^2)/\{E(\partial g/\partial \theta)\}^2 \geq \frac{1}{E(B^2/A^2)}. \quad (4.19)$$

For $a_{t-1} = a^*_{t-1}$, B^2/A^2 is maximized and $E(B^2/A^2) = \{E(B)\}^2/E(A^2)$

Hence the theorem.

Example 4.1. Consider an AR(1) model given by

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t, t = 2, \cdots, n, \quad (4.20)$$

where μ is a constant and ε_t are uncorrelated with 0 mean and constant variance σ^2.

Define

$$h_t = (X_t - \mu) - \phi(X_{t-1} - \mu),\ t = 2, \cdots, n.$$

Then $E_{t-1}(h_t) = 0$ and $E_{t-1}(h_t^2) = \sigma^2$.

Now $E_{t-1}(\partial h_t/\partial \phi) = -(X_{t-1} - \mu)$ and $E_{t-1}(\partial h_t/\partial \mu) = -(1 - \phi)$.

Further, from Theorem(4.2),

$$\begin{aligned} a^*_{t-1,\phi} &= \{E_{t-1}(\partial h_t/\partial \phi)\}/E_{t-1}(h_t^2) \\ &= -(X_{t-1} - \mu)/\sigma^2 \end{aligned}$$

and

$$\begin{aligned} a^*_{t-1,\mu} &= \{E_{t-1}(\partial h_t/\partial\mu)\}/E_{t-1}(h_t^2) \\ &= -(1-\phi)/\sigma^2. \end{aligned}$$

Therefore the optimum estimating functions for ϕ and μ are

$$g^*_\phi = \frac{-1}{\sigma^2}\sum_{t=2}^{n}(X_{t-1}-\mu)[(X_t-\mu)-\phi(X_{t-1}-\mu)], \tag{4.21}$$

$$g^*_\mu = \frac{-1(1-\phi)}{\sigma^2}\sum_{t=2}^{n}[(X_t-\mu)-\phi(X_{t-1}-\mu)]. \tag{4.22}$$

From the equations (4.21) and (4.22), the optimal estimates for ϕ and μ will be obtained by solving $g^*_\phi = 0$ and $g^*_\mu = 0$. These estimates are

$$\hat{\phi} = \frac{\sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1} - (n-1)\sum_{t=2}^{n} X_t X_{t-1}}{(\sum_{t=2}^{n} X_{t-1})^2 - (n-1)\sum_{t=2}^{n} X_{t-1}^2}, \tag{4.23}$$

$$\hat{\mu} = \frac{\sum_{t=2}^{n} X_t \sum_{t=2}^{n} X_{t-1}^2 - \sum_{t=2}^{n} X_{t-1}\sum_{t=2}^{n} X_t X_{t-1}}{((n-1)\sum_{t=2}^{n} X_{t-1}^2 - (\sum_{t=2}^{n} X_{t-1})^2)(1-\hat{\phi})}. \tag{4.24}$$

The estimates in equations (4.23) and (4.24) are the same as the conditional likelihood estimate for AR(1) model with Gaussian noise (see Peiris, Mellor and Ainkaran (2003) and Ainkaran, Peiris and Mellor (2003)).

Example 4.2. Consider the general AR(p) model with zero mean given by the equation(2.6). Define

$$h_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p},\ t = p+1, \cdots, n.$$

Since $E_{t-1}(X_t) = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p}$ we have $E_{t-1}(h_t) = 0$. Further $E_{t-1}(h_t^2) = \sigma^2$ and $E_{t-1}(\partial h_t / \partial \phi_j) = -X_{t-j}$ and therefore

$$a^*_{t-1,\phi_j} = -X_{t-j}/\sigma^2.$$

In this case, the optimum estimating functions for ϕ_j $(j = 1, \cdots, p)$ are

$$g^*_{\phi_j} = \frac{-1}{\sigma^2} \sum_{t=p+1}^{n} X_{t-j}(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})$$

and the optimal estimate for $\Phi' = (\phi_1, \cdots, \phi_p)$ can be obtained by solving the equations $g^*_{\phi_j} = 0$, $j = 1, \cdots, p$. Clearly, the optimal estimate $\hat{\phi}$ can be written as

$$\hat{\Phi} = \left(\sum_{t=p+1}^{n} \mathbf{X_{t-1}} \mathbf{X'_{t-1}} \right)^{-1} \left(\sum_{t=p+1}^{n} \mathbf{X_{t-1}} X_t \right), \tag{4.25}$$

where $\mathbf{X'_{t-1}} = (X_{t-1}, \cdots, X_{t-p})$ (see Thavaneswaran and Abraham (1988)).

Next we consider a recursive algorithm to estimate the parameters via the estimating function approach.

4.2. Recursive Estimation using Estimating Functions

In this section we develop a recursive algorithm based on the optimal estimating function approach. In recursive estimation, the parameter estimate ϕ_t can be obtained as a function of the previous estimate ϕ_{t-1} and of the new measurements. Now we define the function $f(t-1, X) = f(\{X_1, \cdots, X_{t-1}; \phi(F)\})$ as a measurable function w.r.t.

the σ-algebra generated by $X_1, \cdots, X_{t-1}$ and let

$$h_t = X_t - \phi f(t-1, X), \; t = 1, \cdots, n.$$

Then the optimal estimating function (from Theorem 4.2) is

$$\begin{aligned} g^* &= \sum_{t=1}^{n} a_{t-1}^* h_t \\ &= \sum_{t=1}^{n} a_{t-1}^* (X_t - \phi f(t-1, X)), \end{aligned} \tag{4.26}$$

where a_{t-1}^* is as given in (4.12).

To get the optimal estimate for the parameter ϕ based on a finite sample of size n, we solve the equation $g^* = 0$ in (4.26). Therefore, the optimal estimate based on n observations is given by,

$$\hat{\phi}_n = \sum_{t=1}^{n} a_{t-1}^* X_t \Bigg/ \sum_{t=1}^{n} a_{t-1}^* f(t-1, X) \, . \tag{4.27}$$

To estimate all parameters $\phi_1, \phi_2, \cdots$, we derive a recursive algorithm with an initial value of $\hat{\phi}_0 = 0$ as follows:

Let $\kappa_t = \sum_{j=2}^{t} a_{j-1}^* f_{j-1}$. Using the equation(4.27), we have

$$\begin{aligned} \hat{\phi}_t - \hat{\phi}_{t-1} &= \sum_{j=2}^{t} a_{j-1}^* X_j \Bigg/ \sum_{j=2}^{t} a_{j-1}^* f_{j-1} - \hat{\phi}_{t-1} \\ &= \frac{\kappa_{t-1} a_{t-1}^*}{1 + f(t-1, X) a_{t-1}^* \kappa_{t-1}} (X_t - \hat{\phi}_{t-1} f(t-1, X)). \end{aligned} \tag{4.28}$$

Equation (4.28) equivalent to

$$\hat{\phi}_t = \hat{\phi}_{t-1} + \kappa_t a_{t-1}^* (X_t - \hat{\phi}_{t-1} f(t-1, X)) \, , \text{ for all } t = 2, 3, \cdots . \tag{4.29}$$

and

$$\kappa_t = \frac{\kappa_{t-1}}{1 + f(t-1, X) a^*_{t-1} \kappa_{t-1}}. \qquad (4.30)$$

Given starting values of ϕ_1 and κ_1, we compute the estimate recursively using (4.29) and (4.30). The adjustment, $\hat{\phi}_t - \hat{\phi}_{t-1}$, given in the equation (4.28) is called the prediction error. Note that the term $\hat{\phi}_{t-1} f(t-1, X) = E_{t-1}(X_t)$ can be considered as an estimated forecast of X_t given $X_{t-1}, \cdots, X_1$ (See Thavaneswaran and Abraham (1988) and Tong (1990), p.317).

Example 4.3. Consider the simple AR(1) model. In this case $f(t-1, X) = X_{t-1}$ and $a^*_{t-1} = -X_{t-1}/\sigma^2$.

From (4.30), $\kappa_t = \kappa_{t-1}/\{1 - X^2_{t-1}\kappa_{t-1}/\sigma^2\}$ and from (4.29)

$$\hat{\phi}_t = \hat{\phi}_{t-1} - \kappa_t X_{t-1}(X_t - \hat{\phi}_{t-1} X_{t-1})/\sigma^2 \ , \text{ for all } t = 1, 2, \cdots .$$

Starting with a suitable initial value of $\hat{\phi}_1$ and κ_1 we can estimate $\hat{\phi}_n$ recursively for a given series $X_1, \cdots, X_n$.

The next section considers applications of estimating functions in nonlinear time series.

4.3. Applications of Estimating Functions

This section reports some applications of estimating function (4.1) to five different nonlinear time series models in practice.

4.3.1. Bilinear model. Consider the bilinear model (3.2) and let

$$h_t = X_t - E[X_t|\mathcal{F}_{t-1}^X], \tag{4.31}$$

where

$$E[X_t|\mathcal{F}_{t-1}^X] = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j m_{t-j} + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i} m_{t-j}$$

and $E[\varepsilon_{t-j}|\mathcal{F}_{t-1}^X] = m_{t-j}; j \geq 1$.

Then it follows from Theorem (4.2) that the optimal estimating function

$$g_i^* = \sum_{t=p+1}^{n} a_{t-1,i}^* h_t,$$

where

$$a_{t-1,i}^* = E_{t-1}\left(\frac{\partial h_t}{\partial \phi_i}\right) / E_{t-1}(h_t^2)$$

and

$$E_{t-1}(h_t^2) = \sigma_\varepsilon^2 \{1 + \sum_{j=1}^{q} \theta_j^2 + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij}^2 X_{t-i}^2\}.$$

Now it can be shown that

$$a_{t-1,\phi_i}^* = \frac{-X_{t-i}}{\sigma_\varepsilon^2 \{1 + \sum_{j=1}^{q} \theta_j^2 + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij}^2 X_{t-i}^2\}}, \tag{4.32}$$

$$a_{t-1,\theta_j}^* = \frac{-m_{t-j} - (\theta_j + \sum_{i=1}^{r} \beta_{ij} X_{t-i})(\partial m_{t-j}/\partial \theta_j)}{\sigma_\varepsilon^2 \{1 + \sum_{j=1}^{q} \theta_j^2 + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij}^2 X_{t-i}^2\}} \tag{4.33}$$

and

$$a_{t-1,\beta_{ij}}^* = \frac{-X_{t-i} m_{t-j} - (\theta_j + \beta_{ij} X_{t-i})(\partial m_{t-j}/\partial \beta_{ij})}{\sigma_\varepsilon^2 \{1 + \sum_{j=1}^{q} \theta_j^2 + \sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij}^2 X_{t-i}^2\}}. \tag{4.34}$$

Using (4.31)-(4.34), we have the following $(p + q + rs)$ equations

$$g_{\phi_i}^* = \sum_{t=max(p,r)+1}^{n} a_{t-1,\phi_i}^* h_t = 0,$$

$$g_{\theta_j}^* = \sum_{t=max(p,r)+1}^{n} a_{t-1,\theta_j}^* h_t = 0$$

and

$$g^*_{\beta_{ij}} = \sum_{t=max(p,r)+1}^{n} a^*_{t-1,\beta_{ij}} h_t = 0.$$

Solving these equations we get the estimates for all the parameters $\phi_i, i = 1, \cdots, p$; $\theta_j, j = 1, \cdots, q$; $\beta_{ij}, i = 1, \cdots, r; j = 1, \cdots, s$.

4.3.2. Random Coefficient Autoregressive (RCA) Model. Consider the RCA models in (3.11). Let

$$\begin{aligned} h_t &= X_t - E[X_t | \mathcal{F}^X_{t-1}] \\ &= X_t - \sum_{i=1}^{k} \phi_i X_{t-i} \end{aligned}$$

and $\Phi' = (\phi_1, \cdots, \phi_k)$. Then it follows from Theorem (4.2) that the optimal estimating function for ϕ_i is

$$g^*_i = \sum_{t=k+1}^{n} \frac{-X_{t-i}}{\sigma^2_t} \left(X_t - \sum_{i=1}^{k} \phi_i X_{t-i} \right), \tag{4.35}$$

where

$$a^*_{t-1,i} = \frac{-X_{t-i}}{\{\sigma^2_\varepsilon + \sigma^2_\beta \sum_{i=1}^{k} X^2_{t-i}\}}, \tag{4.36}$$

since $E(h^2_t) = \sigma^2_\varepsilon + \sigma^2_\beta \sum_{i=1}^{k} X^2_{t-i} = \sigma^2_t$(say).

The optimal estimate for Φ can be obtained by solving the set of equations $g^*_i = 0, i = 1, \cdots, k$.

Clearly, the optimal estimate can be written as

$$\hat{\Phi} = \left(\sum_{t=k+1}^{n} \mathbf{X_{t-1}} \mathbf{X'_{t-1}} / \sigma^2_t \right)^{-1} \left(\sum_{t=k+1}^{n} \mathbf{X_{t-1}} X_t / \sigma^2_t \right), \tag{4.37}$$

where $\mathbf{X'_{t-1}} = (X_{t-1}, \cdots, X_{t-k})$.

Nicholls and Quinn (1980) and Tjϕstheim (1986) derived the maximum likelihood estimate for Φ given by

$$\tilde{\Phi} = \left(\sum_{t=k+1}^{n} \mathbf{X_{t-1}} \mathbf{X'_{t-1}} \right)^{-1} \left(\sum_{t=k+1}^{n} \mathbf{X_{t-1}} X_t \right), \tag{4.38}$$

which is not efficient but strongly consistent and asymptotically normal. However, $\hat{\Phi}$ given in (4.37) has the consistency and normality property (Thavaneswaran and Abraham (1988)). The optimal estimator $\hat{\Phi}$ depends on σ_ε^2 and σ_β^2 which are not known in practice and we estimate $\tilde{\Phi}$, $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\beta^2$ using least square method.
Let

$$\begin{aligned} v_t &= h_t^2 - E_{t-1}(h_t^2) \\ &= h_t^2 - \sigma_\varepsilon^2 - \sigma_\beta^2 \sum_{i=1}^{k} X_{t-i}^2 \end{aligned}$$

To get the least square estimates $\hat{\sigma}_\varepsilon^2$ of σ_ε^2 and $\hat{\sigma}_\beta^2$ of σ_β^2, minimize $\sum_{t=k+1}^{n} v_t^2$ with respect to σ_ε^2 and σ_β^2 respectively.
Normal equations are

$$\sum_{t=k+1}^{n} (\hat{h}_t^2 - \sigma_\varepsilon^2 - \sigma_\beta^2 Y_{t-k}) = 0$$

and

$$\sum_{t=k+1}^{n} (\hat{h}_t^2 - \sigma_\varepsilon^2 - \sigma_\beta^2 Y_{t-k}) Y_{t-k} = 0,$$

where $Y_{t-k} = \sum_{i=1}^{k} X_{t-i}^2$. Thus the least square estimates are

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=k+1}^{n} \hat{h}_t^2 Y_{t-k} \sum_{t=k+1}^{n} Y_{t-k} - \sum_{t=k+1}^{n} \hat{h}_t^2 \sum_{t=k+1}^{n} Y_{t-k}^2}{(\sum_{t=k+1}^{n} Y_{t-k})^2 - (n-k) \sum_{t=k+1}^{n} Y_{t-k}^2} \tag{4.39}$$

$$\hat{\sigma}_\beta^2 = \frac{(n-k) \sum_{t=k+1}^{n} \hat{h}_t^2 Y_{t-k} - \sum_{t=k+1}^{n} \hat{h}_t^2 \sum_{t=k+1}^{n} Y_{t-k}}{(n-k) \sum_{t=k+1}^{n} Y_{t-k}^2 - (\sum_{t=k+1}^{n} Y_{t-k})^2} \tag{4.40}$$

(see Nicholls and Quinn (1982), p.42). $\tilde{\Phi}$ could be used initially to estimate h_t and then $\hat{\Phi}$ may be calculated with the estimated values $\hat{\sigma}_{\varepsilon}^2$ and $\hat{\sigma}_{\beta}^2$.

Suppose $k = 1$ then the model in (3.11) bacome RCA(1) model ,

$$X_t - (\phi + \beta(t))X_{t-1} = \varepsilon_t. \tag{4.41}$$

The optimal estimate for this model from (4.37) is

$$\phi^* = \sum_{t=2}^{n} \frac{X_t X_{t-1}}{\sigma_{\varepsilon}^2 + \sigma_{\beta}^2 X_{t-1}^2} / \sum_{t=2}^{n} \frac{X_{t-1}^2}{\sigma_{\varepsilon}^2 + \sigma_{\beta}^2 X_{t-1}^2} \tag{4.42}$$

(see Thavaneswaran and Abraham (1988) or Thavaneswaran and Peiris (1998) for details).

4.3.3. Doubly Stochastic model. Consider the doubly stochastic model in (3.19) and let

$$h_t = X_t - \left(\sum_{i=1}^{p} \phi_i f_i(\mathcal{F}_{t-1}^X) X_{t-i} + \sum_{j=1}^{q} \theta_j g_j(\mathcal{F}_{t-1}^X) m_{t-j} \right), \tag{4.43}$$

where $E[\varepsilon_{t-j}|\mathcal{F}_{t-1}^X] = m_{t-j}$ and $E[(\varepsilon_{t-j} - m_{t-j})^2|\mathcal{F}_{t-1}^X] = \gamma_{t-j}$; for all $j \geq 1$. For the evaluation of m_{t-j} and γ_{t-j}, we can use a Kalman-like recursive algorithm (see Thavaneswaran and Abraham (1988), p.102, Shiryayev (1984), p.439).

$$E(h_t^2) = \sigma_{\varepsilon}^2 + \sum_{j=1}^{q} \theta_j^2 g_j^2(\mathcal{F}_{t-1}^X)\gamma_{t-j}.$$

Then

$$a_{t-1,\phi_i}^* = \frac{-f_i(\mathcal{F}_{t-1}^X) X_{t-i}}{\sigma_{\varepsilon}^2 + \sum_{j=1}^{q} \theta_j^2 g_j^2(\mathcal{F}_{t-1}^X)\gamma_{t-j}} \tag{4.44}$$

and

$$a^*_{t-1,\theta_j} = \frac{-g_j(\mathcal{F}^X_{t-1})m_{t-j} - \theta_j g_j(\mathcal{F}^X_{t-1})(\partial m_{t-j}/\partial\theta_j)}{\sigma^2_\varepsilon + \sum_{j=1}^q \theta_j^2 g_j^2(\mathcal{F}^X_{t-1})\gamma_{t-j}}. \tag{4.45}$$

Using (4.43)-(4.45), we have the following $(p+q)$ equations

$$g^*_{\phi_i} = \sum_{t=p+1}^n a^*_{t-1,\phi_i} h_t = 0,$$

and

$$g^*_{\theta_j} = \sum_{t=p+1}^n a^*_{t-1,\theta_j} h_t = 0.$$

Solving these equations we can get the estimates for all the parameters $\phi_i,\ i = 1, \cdots, p;\ \theta_j,\ j = 1, \cdots, q$.

Example 4.4. Consider a doubly stochastic model with RCA sequence (Tjøstheim (1986))

$$X_t = \phi_t f(t, \mathcal{F}^X_{t-1}) + \varepsilon_t, \tag{4.46}$$

where the more general stochastic sequence $\{\phi_t\}$ and the function of past values $f(t, \mathcal{F}^X_{t-1})$ are replaced by the sequence $\phi + \beta(t)$ and X_{t-1} respectively. When $\beta(t)$ is a MA(1) sequence of the form

$$\beta(t) = e_t + e_{t-1},$$

where $\{\varepsilon_t\}$, $\{\beta(t)\}$ and $\{e_t\}$ are zero mean square integrable independent processes with constant variances σ^2_ε, σ^2_β and σ^2_e.
In this case $E[X_t|\mathcal{F}^X_{t-1}]$ depends on the posterior mean $m_t = E[e_t|\mathcal{F}^X_{t-1}]$ and variance $\gamma_t = E[(e_t - m_t)^2|\mathcal{F}^X_{t-1}]$ of e_t. For the evaluation of m_t and γ_t, we further assume that $\{\varepsilon_t\}$ and $\{e_t\}$ are Gaussian and $X_0 = 0$.

Then m_t and γ_t satisfy the following Kalman-like recursive algorithm (see Shiryayev,(1984), p.439):

$$m_t = \sigma_e^2 X_{t-1}[X_t - (\phi + m_{t-1})X_{t-1}]/\sigma_{t-1}^2 \tag{4.47}$$

and

$$\gamma_t = \sigma_e^2 - [\sigma_e^4 X_{t-1}^2/\sigma_{t-1}^2], \tag{4.48}$$

where $\sigma_{t-1}^2 = \sigma_\varepsilon^2 + (\sigma_\varepsilon^2 + \gamma_{t-1})X_{t-1}^2$. Starting with initial values $\gamma_1 = \sigma_e^2$ and $m_1 = 0$, m_{t-1} and γ_{t-1} can be obtained recursively and hence

$$E[X_t|\mathcal{F}_{t-1}^X] = (\phi + m_{t-1})X_{t-1} \tag{4.49}$$

and

$$\begin{aligned} E_{t-1}(h_t^2)] &= E\{[X_t - E_{t-1}(X_t)]^2|\mathcal{F}_{t-1}^X\} \\ &= \sigma_\varepsilon^2 + (\sigma_\varepsilon^2 + \gamma_{t-1})X_{t-1}^2 \qquad (4.50) \\ &= \sigma_{t-1}^2. \end{aligned}$$

To obtain the optimal estimating function, notice that

$$E_{t-1}(\partial h_t/\partial\phi) = [1 + (\partial m_{t-1}/\partial\phi)]X_{t-1}.$$

Since γ_t is independent of ϕ, the derivative

$$\frac{\partial m_t}{\partial \phi} = \frac{-\sigma_e^2 X_{t-1}^2(1 + (\partial m_{t-1}/\partial\phi))}{\sigma_\varepsilon^2 + (\sigma_\varepsilon^2 + \gamma_{t-1})X_{t-1}^2} \tag{4.51}$$

can be calculated recursively. Therefore the optimal estimating function is

$$g_\phi^* = \sum_{t=2}^{n} (X_t - \phi X_{t-1})[1 + (\partial m_{t-1}/\partial\phi)]X_{t-1}/\sigma_{t-1}^2$$

and the optimal estimate is

$$\hat{\phi} = \frac{\sum_{t=2}^{n}\{[1 + (\partial m_{t-1}/\partial\phi)]X_t X_{t-1}/\sigma_{t-1}^2\}}{\sum_{t=2}^{n}\{[1 + (\partial m_{t-1}/\partial\phi)]X_{t-1}^2/\sigma_{t-1}^2\}} \tag{4.52}$$

(see Thavaneswaran and Abraham (1988), p.103).

4.3.4. Threshold Autoregressive (TAR) model. Consider the TAR model given in (3.25) with $\phi_{0j} = 0$ and

$$h_t = X_t - \sum_{j=1}^{l}\sum_{i=1}^{k_j} \phi_{ij} X_{t-i} I(X_{t-d} \in D_j), \tag{4.53}$$

then

$$E(h_t^2) = E(\varepsilon_t^2) = \sigma_\varepsilon^2.$$

The estimating function for ϕ_{ij} is

$$g_{\phi_{ij}}^* = -\sum_{t=k+1}^{n} X_{t-i} I(X_{t-d} \in D_j) \left\{ X_t - \sum_{j=1}^{l}\sum_{i=1}^{k_j} \phi_{ij} X_{t-i} I(X_{t-d} \in D_j) \right\} /\sigma_\varepsilon^2, \tag{4.54}$$

where $k = \max_{1 \le j \le l}(k_j)$.

Solving the equations $g_{\phi_{ij}}^* = 0$, one gets the solutions for the estimates

$$\Phi = \{\phi_{ij}; i = 1, \cdots, k_j; j = 1, \cdots, l\}.$$

i.e.

$$\Phi = \left(\sum_{t=k+1}^{n} \mathbf{X_{t-1}} \mathbf{X'_{t-1}} I(X_{t-d} \in D_j) \right)^{-1} \left(\sum_{t=k+1}^{n} \mathbf{X_{t-1}} X_t I(X_{t-d} \in D_j) \right), \tag{4.55}$$

where $\mathbf{X'_{t-1}} = (X_{t-1}, \cdots, X_{t-k})$.

4.3.5. ARCH model. Consider the ARCH(p) model given in (3.30) and let

$$h_t = X_t^2 - \phi_0 - \sum_{i=1}^{p} \phi_i X_{t-i}^2. \tag{4.56}$$

Then

$$\begin{aligned} E_{t-1}(h_t^2) &= E_{t-1}(v_t^2) \\ &= E_{t-1}(X_t^4 - \sigma_t^4) \\ &= E_{t-1}(X_t^4) - E_{t-1}(\sigma_t^4) \\ &= E_{t-1}(\sigma_t^4 \varepsilon_t^4) - E_{t-1}(\sigma_t^4) \\ &= E_{t-1}(\sigma_t^4)[E_{t-1}(\varepsilon_t^4) - 1]. \end{aligned} \tag{4.57}$$

Using $\sigma_t^2 = \phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2$ from (3.29) we have

$$E_{t-1}(\sigma_t^4) = [\phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2]^2$$

and if ε_t is standard normal then $E(\varepsilon_t^4) = 3$ and

$$E_{t-1}(h_t^2) = 2[\phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2]^2,$$

$$E_{t-1}(\frac{\partial h_t}{\partial \phi_i}) = -X_{t-i}^2,$$

$$a_{t-1,\phi_i}^* = \frac{-X_{t-i}^2}{2[\phi_0 + \sum_{i=1}^{p} \phi_i X_{t-i}^2]^2}.$$

In this case

$$\begin{aligned} g_{\phi_i}^* &= \sum_{t=p+1}^{n} a_{t-1,\phi_i}^* h_t \\ &= \sum_{t=p+1}^{n} \frac{-X_{t-i}^2}{2[\phi_0 + \sum_{j=1}^{p} \phi_j X_{t-j}^2]^2} [X_t^2 - \phi_0 - \sum_{j=1}^{p} \phi_j X_{t-j}^2], \end{aligned}$$

$i = 1, \cdots, p.$

For a given value of ϕ_0, solve these p equations, $g^*_{\phi_i} = 0$ and get the estimates for the parameters $\phi_i, i = 1, \cdots, p$.

If ε_t are not standard normal, use the kurtosis formula in Theorem 3.1 (a) to calculate $E(\varepsilon_t^4)$ and then substitute in (4.57).
Furthermore, if ε_t are not normal, Thavaneswaran, Appadoo and Samanta (2004) used an approach to calculate the kurtosis of the GARCH models with non-normal innovations.
The next section considers the Smoothed Estimating Functions for stochastic volatility models and the applications on some nonlinear time series models.

4.4. Smoothed Estimating functions

Recently, Novak (2000) discussed the problem of obtaining a generalized kernel density estimator with independent observations. The kernel function smoothing approach is a useful tool to estimate the conditional mean for nonlinear time series. Further work on the smoothing methodology in various directions can be found in Thavaneswaran (1988), Thavaneswaran and Singh (1993) and Thavaneswaran and Peiris (1996). The purpose of this work is to explore the implications of the result of Novak (2000) for estimating the conditional mean of a nonlinear time series.
We shall attempt to develop a more systematic approach and discuss a generalized kernel smoothed estimate for nonlinear time series models. Our approach yields the most recent estimation results for nonlinear

time series as special cases and, in fact, we are able to weaken the conditions in the maximum likelihood procedure.

Let $X_1, \cdots, X_n$ be a random sample from a distribution with nonzero density $f(x)$. The classical kernel density estimator of $f(x)$ at any given x was proposed by Rosenblatt (1956) and studied by Parzen (1962) and Nadaraja (1974). The corresponding kernal density estimator $\hat{f}_n(x)$ of $f(x)$is given by

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} f_\gamma\left((X_i - x)b^{-1}\right) b^{-1}, n \in \mathcal{N}, \tag{4.58}$$

where the kernel $f_\gamma(.)$ is the density of some symmetric random variable γ, and the smoothing parameters $b = b(n)$ (band width) are determined by a statistician. It is assumed that the density f is continuous function of x, f_γ vanishes outside the interval $[-T, T]$ and has at most a finite number of discontinuity points. The generalized kernel estimator in Novak (2000) is

$$f_{n,\alpha}(x) = \frac{1}{n}\sum_{i=1}^{n} f_{b\gamma}\left((X_i - x)f^\alpha(X_i)\right) f^\alpha(X_i)\, I_i, \tag{4.59}$$

where $\alpha \in \mathcal{R}$, $I_i = I\{|x - X_i| f^\alpha(x) < bT_+\}$, and T_+ is a constant greater than T .

We propose the following extended version of the Novak's estimator in (4.59) for any continuous function g of f:

$$f_{n,g}(x) = \frac{1}{n}\sum_{i=1}^{n} f_{b\gamma}\left((X_i - x)g(f(\,(X_i)))\right) g(f\,(X_i))I_i, \tag{4.60}$$

where $I_i = I\{|x - X_i|g(f(\,(x)) < bT_+\}$. If $g(x) = f^\alpha(x)$, then estimator (4.60) reduces to the estimator (4.59) and for $\alpha = \frac{1}{2}$ it coincides with the estimator given in Abramson (1982). Thus, the class of estimators $\{f_{n,g}(x)\}$ in (4.60) generalizes the kernel density estimators in (4.59). Now we recall the theorem 4.2 on stochastic processes and apply it to obtain generalized kernel smoothed estimates for some nonlinear time series models. Suppose we want to estimate the time varying parameter $\alpha(t)$ for a process having the first two conditional moments $E[X_t|\mathcal{F}_{t-1}^X] = \alpha(t)h(\mathcal{F}_{t-1}^X)$ and $\text{var}[X_t|\mathcal{F}_{t-1}^X] = \sigma^2(\mathcal{F}_{t-1}^X)$. Let

$$\epsilon_t = X_t - E(X_t|\mathcal{F}_{t-1}^X). \tag{4.61}$$

The smoothed version of the least squares estimating function for estimating $\theta = \alpha(t_0)$ can be written as

$$S_n^{ls}(t_0) = \sum_{t=1}^{n} w(\frac{t_0 - t}{b})h(\mathcal{F}_{t-1}^X)(X_t - \theta h(\mathcal{F}_{t-1}^X)), \tag{4.62}$$

where $w(\frac{t_0-t}{b})$ is a suitably chosen kernel with bandwidth b.

The corresponding smoothed version of the optimal estimating equation studied in Thavaneswaran and Peiris (1996) is

$$S_n^{opt}(t_0) = \sum_{t=1}^{n} w(\frac{t_0 - t}{b})a_{t-1}^* \epsilon_t = 0, \tag{4.63}$$

where $a_{t-1}^* = \frac{\partial \epsilon_t}{\partial \theta}/\sigma^2 h(\mathcal{F}_{t-1}^X)$ is the optimal value as in Theorem 4.2.

Notes:

1. If $\epsilon_t' s$ are independent and have the density $f(\cdot)$, then it follows from Godambe (1960) that the optimal estimating function for $\theta = f(x_t)$ for fixed t in $x_t = \theta + \epsilon_t$ is the score function

$$\sum_{t=1}^{n} \frac{\partial}{\partial \theta} \log f(x_t - \theta) = 0 \tag{4.64}$$

and the corresponding smoothed optimal estimating function is

$$\sum_{t=1}^{n} w(\frac{x_0 - x_t}{b})\frac{\partial}{\partial\theta}\log f(x_t - \theta) = 0 \tag{4.65}$$

which is the same as the one considered in Staniswalis (1989).

2. When $h(\cdot) = 1$, the model in (4.61) reduces to a time series model with a time varying parameter $\alpha(t)$.

3. Furthermore, if $\epsilon_t^{\prime}s$ are independent and $h(\cdot) = 1$ with $\sigma^2 =$ constant, then the above model in (4.61) corresponds to a regression model considered in Stainswalis (1989).

Now in analogy with (4.60), we propose a generalized kernal smoothed optimal estimating equation for estimating the conditional mean parameter $\theta = \alpha(t)$ at t_0 as:

$$\sum_{t=1}^{n} f_{b\gamma}\left((t_0 - t)g(\alpha(x_t))\right) g(\alpha(x_t))a_{t-1}^{*}\epsilon_t = 0, \tag{4.66}$$

where g is a continuous function of $\alpha(t)$ through the observations as in (4.60). When $g(x) = 1$, (4.66) reduces to (4.63). Using the notation $g(\alpha(.)) = g(.)$, the explicit form of the resulting estimator from the estimating equation (4.66) can be written as:

$$\theta^{gen}(t) = \frac{\sum_{t=1}^{n} f_{b\gamma}\left((t_0 - t)g(x_t)\right) g(x_t)a_{t-1}^{*}x_t}{\sum_{t=1}^{n} f_{b\gamma}\left((t_0 - t)g(x_t)\right) g(x_t)a_{t-1}^{*}h(\mathcal{F}_{t-1}^{x})}. \tag{4.67}$$

We now consider potential applications of (4.67) for some stochastic volatility models in time series analysis.

4.5. Applications of Smoothed Estimating Functions

This section reports some applications of (4.67) to time series models discussed before.

4.5.1. Random Coefficient Autoregressive (RCA) Model. Consider the RCA(1) model defined by allowing random additive perturbations $\{c(t)\}$ for autoregressive coefficients of ordinary AR models. That is, a random process $\{X_t\}$ is an RCA model if it is of the form

$$X_t - \{\phi(t) + \beta(t)\}X_{t-1} = \varepsilon_t, \tag{4.68}$$

where the parameters $\phi(t)$ are to be estimated, $\{\epsilon_t\}$ and $\{\beta(t)\}$ are zero mean square integrable independent processes with constant variances σ_ϕ^2 and σ_β^2 and $\{\beta(t)\}$ is independent of $\{\varepsilon_t\}$ and $\{X_{t-i}\}; i \geq 1$.

Write

$$\epsilon_t = X_t - E[X_t|\mathcal{F}_{t-1}^X] = X_t - \phi(t)X_{t-1}. \tag{4.69}$$

Let $\theta = \phi(t_0)$ be the value of $\phi(t)$ for a given value of t_0. Then the generalized kernel smoothed optimal estimating function for θ is given by

$$\sum_{t=1}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t)a_{t-1}^*\varepsilon_t = 0, \tag{4.70}$$

where $a_{t-1}^* = \dfrac{E(\frac{\partial\varepsilon_t}{\partial\theta}|\mathcal{F}_{t-1}^X)}{E(\varepsilon_t^2|\mathcal{F}_{t-1}^X)}$.

Now it can be shown that $a_{t-1}^* = -\dfrac{X_{t-1}}{\{\sigma_\varepsilon^2 + X_{t-1}^2\sigma_\beta^2\}}$
and the optimal generalized kernel smoothed estimate is given by

$$\theta^{gen}(t_0) = \frac{\sum_{t=2}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t)a_{t-1}^*X_t}{\sum_{t=2}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t)a_{t-1}^*X_{t-1}}\ . \tag{4.71}$$

This simplifies to

$$\theta^{gen}(t_0) = \frac{\sum_{t=2}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t) \dfrac{X_t X_{t-1}}{\sigma_\varepsilon^2 + \sigma_\beta^2 X_{t-1}^2}}{\sum_{t=2}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t) \dfrac{X_{t-1}^2}{\sigma_\varepsilon^2 + \sigma_\beta^2 + X_{t-1}^2}} . \tag{4.72}$$

However, Nicholls and Quinn (1980) obtained the least squares estimate of θ (a fixed parameter) and the smoothed version of their estimate is given by

$$\theta^{LS}(t_0) = \frac{\sum_{t=2}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t) X_t X_{t-1}}{\sum_{t=2}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t) X_{t-1}^2} . \tag{4.73}$$

Clearly, (4.73) is quite different from (4.72). (See Nicholls and Quinn (1980) and Tjϕstheim (1986) for more details). In fact, using the theory of estimating function one can argue that the generalized kernel smoothed optimal estimating function is more informative than the least squares one.

4.5.2. Doubly stochastic time series. Consider the class of nonlinear models given by

$$X_t - \theta_t h(t, F_{t-1}^X) = \varepsilon_t, \tag{4.74}$$

where $\{\theta_t\}$ is a general stochastic process.These are called doubly stochastic time series models. Note that $\{\alpha(t) + c(t)\}$ in (4.68) is now replaced by a more general stochastic sequence $\{\theta_t\}$ and X_{t-1} is replaced by a function of the past, $h(t, F_{t-1}^X)$. It is clear that the random coefficient autoregressive processes given in (4.68) are special cases of what Tjøstheim (1986) refers to (4.74) as doubly stochastic time series models. When $\{\theta_t\}$ is a moving average (MA) sequence of the form

$$\theta_t = \alpha(t) + e_t + e_{t-1}, \tag{4.75}$$

where $\{\theta_t\}, \{\varepsilon_t\}$ are square integrable-independent random variables and $\{e_t\}$ consists of zero mean square integrable random variables independent of $\{\varepsilon_t\}$. In this case $E(X_t|F_{i-1}^X)$ depends on the posterior mean, $m_t = E(e_t|F_t^X)$, and variance $\nu_t = E[e_t - m_t)^2|F_t^X]$ of e_t. Thus, for the evaluation of m_t and ν_t we further assume that $\{\varepsilon_t\}$ and $\{e_t\}$ are Gaussian and that $X_0 = 0$. Then m_t and ν_t satisfy the following Kalman-like recursive algorithms (see Shiryayev, 1984, p.439):

$$m_t = \frac{\sigma_e^2 h(t, F_{t-1}^X)[X_t - (\alpha(t) + m_{t-1})h(t, F_{t-1}^X)]}{\sigma_\varepsilon^2 + h^2(t, F_{t-1}^X)(\sigma_\varepsilon^2 + \nu_{t-1})} \tag{4.76}$$

and

$$\nu_t = \sigma_t^2 - \frac{h^2(t, F_{t-1}^X)\sigma_e^4}{\sigma_\varepsilon^2 + h^2(t, F_{t-1}^X)(\sigma_\varepsilon^2 + \nu_{t-1})} \ ,$$

where $\nu_0 = \sigma_e^2$ and $m_0 = 0$. Hence

$$E(X_t|F_{t-1}^X) = (\alpha(t) + m_{t-1})h(t, F_{t-1}^X)$$

and

$$\begin{aligned} E(h_t^2|F_t^X) &= E\{[X_t - E(X_t|F_{t-1}^X)]^2|F_{t-1}^X\} \\ &= \sigma_\varepsilon^2 + h^2(t, F_{t-1}^X)(\sigma_\varepsilon^2 + \nu_{t-1}) \end{aligned}$$

can be calculated recursively.

Then the generalized smoothed optimal estimating function for $\theta = \alpha(t_0)$ turns out to be

$$G_n^{gen} = \sum_{t=1}^{n} f_{b\gamma}\left((t_0 - t)g(X_t)\right) g(X_t) a_{t-1}^* e_t = 0, \tag{4.77}$$

where g is a continuous function of $\alpha(t)$ through the observations as in (4.60) and $a^*_{t-1} = E[(\partial e_t/\partial\theta)|F^X_{t-1}]/E[e^2_t|F^X_{t-1}]$. The explicit form of the resulting estimator from the estimating equation (4.77) can be written as

$$\theta^{gen}_n = \frac{\sum_{t=2}^n f_{b\gamma}\left((t_0-t)g(X_t)\right)g(X_t)a^*_{t-1}X_t}{\sum_{t=2}^n f_{b\gamma}\left((t_0-t)g(X_t)\right)g(X_t)a^*_{t-1}h(F^X_{t-1})}, \quad (4.78)$$

where

$$a^*_{t-1} = h(t, F^X_{t-1})(1+(\partial m_{t-1}/\partial\theta))/\{\sigma^2_\varepsilon + h^2(t, F^X_{t-1})(\sigma^2_\varepsilon + \nu_{t-1})\}. \quad (4.79)$$

Since ν_t is independent of $\theta = \alpha(t_0)$, the relation

$$\partial m_t/\partial\theta = -\{\sigma^2_\varepsilon h^2(t, F^X_{t-1})(1+\partial m_{t-1}/\partial\theta)\}/\{\sigma^2_\varepsilon + h^2(t, F^X_{t-1})(\sigma^2_\varepsilon + \nu_{t-1})\}$$

can be used to calculate this derivative recursively.

The conditional least-squares approach of Tjϕstheim (1986) leads to an estimator

$$\theta^{LS}_n = \frac{\sum_{t=2}^n f_{b\gamma}\left((t_0-t)g(X_t)\right)g(X_t)h(t, F^X_{t-1})(1+(\partial m_{t-1}/\partial\theta))X_t}{\sum_{t=2}^n f_{b\gamma}\left((t_0-t)g(X_t)\right)g(X_t)[h(t, F^X_{t-1})(1+(\partial m_{t-1}\,/\,\partial\theta))]} \quad (4.80)$$

which does not take into account the variances σ^2_ε. However, as can be seen from (4.72) and (4.73), the optimal estimate θ^{gen}_n adopts a weighting scheme based on σ^2_ε and σ^2_e. In practice, these quantities may be obtained using some non-linear optimization techniques. (See Thavaneswaran and Abraham (1988), Section 4).

4.5.3. Threshold autoregressive process. We now consider an application of the theory of smoothed estimating equations in the context of the threshold autoregressive model with only one residual process given by Tjϕstheim (1986) as follows:

$$X_t - \sum_{j=1}^{n} \alpha_j(t) X_{t-1} H_j(X_{t-1}) = \varepsilon_t, \tag{4.81}$$

where $H_j(X_{t-1}) = I(X_{t-1} \in D_j), I(\cdot)$ being the indicator function and $D_1, D_2, \cdots D_m$ are disjoint regions of R such that $\cup D_j = R$. Then we have

$$\varepsilon_t = X_t - E(X_t | F_{t-1}^X) = X_t - \sum_{j=1}^{p} \alpha_j(t) X_t H_j(X_{t-1})$$

and

$$E(h_t^2 | F_{t-1}^X) = E(\varepsilon_t^2) = \sigma_\varepsilon^2 \ .$$

Hence the optimal generalized kernel estimate for $\alpha_j(t)$ based on the n observations is

$$\theta_{n,j}^{gen} = \frac{\sum_{t=1}^{n} f_{b\gamma}\left((t_0 - t) g(X_t)\right) g(X_t) a_{t-1}^* X_t X_{t-1} H_j(X_{t-1})}{\sum_{t=2}^{n} f_{b\gamma}\left((t_0 - t) g(X_t)\right) g(X_t) X_{t-1}^2 H_j(X_{t-1})} \tag{4.82}$$

which reduces to the one obtained by Thavaneswaran and Peiris(1996) for $g(\cdot) = 1$.

4.5.4. ARCH model. Consider an ARCH(1) model given by

$$X_t = \sigma_t \varepsilon_t, \tag{4.83}$$

$$\sigma_t^2 = \theta_0 + \theta_1 X_{t-1}^2, \tag{4.84}$$

where $\{\varepsilon_t\}$ is an independent sequence of random variables having mean zero and variance σ_ε^2. This basic univariate ARCH model has been extended in a number of directions. For example, some are dictated by economic insight and others by (broadly) statistical ideas. The most important generalization of this class is the extension to include moving average parts and is called the generalized ARCH (GARCH) model. The simplest example of the GARCH class is GARCH(1,1) given by (see, Shephard(1996))

$$X_t = \varepsilon_t \sigma_t \tag{4.85}$$

$$\sigma_t^2 = \theta_0 + \theta_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{4.86}$$

Clearly, this model can be written as a non-Gaussian linear ARMA model for X_t^2 as:

$$X_t^2 = \theta_0 + \theta_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + v_t = \theta_0 + (\theta_1 + \beta_1) X_{t-1}^2 + v_t - \beta_1 v_{t-1}, \tag{4.87}$$

where $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$ is a martingale difference. Using the martingale estimating functions given in equation (4.14), one can easily obtain the recursive estimates of the parameters. In practice, the initial values of the parameters are obtained by first fitting an ARMA(1,1) model for the non-Gaussian process X_t^2.

By writing the model in the following state space form for X_t^2

$$X_t^2 = \theta_0 + (\theta_1 + \beta_1) X_{t-1}^2 + v_t - \beta_1 v_{t-1} \tag{4.88}$$

$$v_t = \beta_1 v_{t-1} + e_t \tag{4.89}$$

one can easily obtain the recursive estimates of σ_t^2.

Now suppose that we have the first two conditional moments

$$E[X_t^2 | F_{t-1}^X] = \alpha(t) h(F_{t-1}^X)$$

and

$$\text{var}[X_t^2|F_{t-1}^X] = \sigma^2(F_{t-1}^X)$$

and we want to estimate the time varying parameter $\alpha(t)$.

Let

$$\varepsilon_t = X_t^2 - E(X_t^2|F_{t-1}^X). \qquad (4.90)$$

The smoothed version of the least squares estimating function for estimating $\theta = \alpha(t_0)$ can be written as

$$S_n^{ls}(t_0) = \sum_{t=1}^{n} w(\frac{t_0 - t}{b})h(F_{t-1}^X)(X_t^2 - \theta h(F_{t-1}^X)), \qquad (4.91)$$

where $w(\frac{t_0-t}{b})$ is a suitably chosen kernel with bandwidth b.

The corresponding smoothed version of the optimal estimating equation studied in Thavaneswaran and Peiris (1996) is

$$S_n^{opt}(t_0) = \sum_{t=1}^{n} w(\frac{t_0 - t}{b})a_{t-1}^* \varepsilon_t = 0, \qquad (4.92)$$

where $a_{t-1}^* = \frac{\partial \varepsilon_t}{\partial \theta}/\sigma^2 h(F_{t-1}^X)$ is the optimal value as in Theorem 4.2.

Chapter 5 considers a simulation study based on four popular models consisting of

(i) RCA model

(ii) Doubly Stochastic model

(iii) TAR model

(iv) ARCH model.

We investigate and report the bias and mse of the estimates using estimating function approach.

Chapter 5

Simulation Study for the Estimates Via Estimating Functions

This chapter considers a simulation study and investigates the finite sample properties of estimates based on the estimating function approach. We first generate a sample of 1500 observations from a known nonlinear model with a given set of parameters using S-plus. We now pick the last 1000 values as our sample. The estimating equations in Chapter 4 are used to compute the parameter estimates based on the estimating functions. Four popular nonlinear models are considered for illustration and parameter estimates are obtained by the Newton-Raphson algorithm. We repeat the simulation and estimation k times and at the end of each estimation, we compute the mean, variance, bias and mean square error (mse) of the estimates (rounded to four decimal points).

Let $\hat{\phi}_i$ be the estimate of the parameter ϕ at the i^{th} iteration then

$$\begin{aligned}\text{mean} = \bar{\phi} &= \frac{1}{k}\sum_{i=1}^{k}\hat{\phi}_i, \quad \text{variance} = \frac{1}{k}\sum_{i=1}^{k}(\hat{\phi}_i - \bar{\phi})^2, \\ \text{bias} &= \frac{1}{k}\sum_{i=1}^{k}(\hat{\phi}_i - \phi) \quad \text{and} \quad \text{mse} = \frac{1}{k}\sum_{i=1}^{k}(\hat{\phi}_i - \phi)^2.\end{aligned}$$

Below we tabulate these results for $k = 10000$ and illustrate the bias and mse graphically for the four nonlinear models discussed in Chapter 4.

5.1. RCA Model

Consider the RCA(1) model given by

$$X_t = (\phi + \beta(t))X_{t-1} + \varepsilon_t, \tag{5.1}$$

where

(i) $\{\varepsilon_t\}$ and $\{\beta(t)\}$ are zero mean, square integrable, independent processes with constant variances σ_ε^2 and σ_β^2,

(ii) $\{\beta(t)\}$ is independent of $\{\varepsilon_t\}$ and $\{X_{t-1}\}$,

(iii) ϕ is the parameter to be estimated.

For this simulation study, we take $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$ and $\beta(t) \sim \text{NID}(0, \sigma_\beta^2)$ with $\sigma_\varepsilon^2 = \sigma_\beta^2 = 1$. We simulate a sample of 1500 values from this RCA(1) model.

The optimal estimating function for ϕ given in (4.35) is

$$g_\phi^* = \sum_{t=2}^{n} \frac{-X_{t-1}}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2} \left(X_t - \phi X_{t-1}\right). \tag{5.2}$$

and the corresponding optimal estimate for ϕ is

$$\hat{\phi} = \sum_{t=2}^{n} \frac{X_t X_{t-1}}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2} \Bigg/ \sum_{t=2}^{n} \frac{X_{t-1}^2}{\sigma_\epsilon^2 + \sigma_\beta^2 X_{t-1}^2}. \tag{5.3}$$

We take the last 1000 values of the sample and use the equation (5.3) to estimate $\hat{\phi}$. Now repeat the simulation and estimation 1000 times to calculate the mean, variance, bias and mse for the different values of $\hat{\phi}$. These values are given in Table 13 with the corresponding true values of ϕ. The bias and the mse of this estimation method are given in Figure 21 and Figure 22 respectively.

ϕ	mean of $\hat{\phi}$	variance of $\hat{\phi}$	Bias of $\hat{\phi}$	mse of $\hat{\phi}$
0.05	0.0511	0.0041	0.0011	0.0041
0.10	0.1020	0.0041	0.0020	0.0041
0.15	0.1490	0.0043	-0.0010	0.0043
0.20	0.1975	0.0042	-0.0025	0.0042
0.25	0.2522	0.0041	0.0022	0.0042
0.30	0.2991	0.0041	-0.0009	0.0041
0.35	0.3519	0.0042	0.0019	0.0042
0.40	0.4000	0.0042	0.0000	0.0042
0.45	0.4548	0.0041	0.0048	0.0041
0.50	0.5040	0.0039	0.0040	0.0039
0.55	0.5498	0.0038	-0.0002	0.0038
0.60	0.6000	0.0035	0.0000	0.0035
0.65	0.6479	0.0040	-0.0021	0.0040
0.70	0.7028	0.0041	0.0028	0.0041
0.75	0.7483	0.0037	-0.0017	0.0037
0.80	0.7999	0.0037	-0.0001	0.0037
0.85	0.8487	0.0042	-0.0013	0.0042
0.90	0.9026	0.0039	0.0026	0.0039
0.95	0.9516	0.0037	0.0016	0.0037

TABLE 13. Mean, variance, bias and mse of $\hat{\phi}$ for the RCA(1) model.

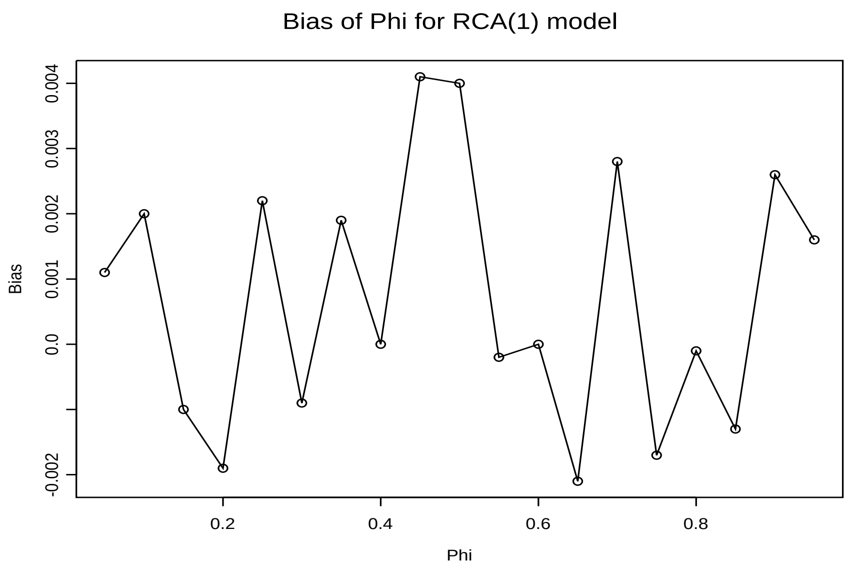


FIGURE 21. Plot of the bias of $\hat{\phi}$ in RCA(1)

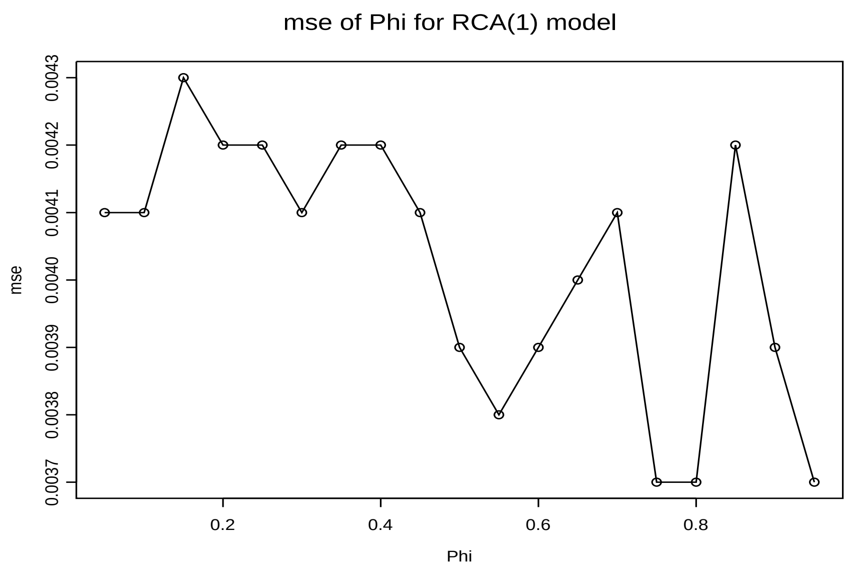


FIGURE 22. Plot of the mse of $\hat{\phi}$ in RCA(1)

In practice, σ_ε^2 and σ_β^2 are unknown and we proceed as follows. First find the least square estimate $\tilde{\phi}$ of ϕ using

$$\tilde{\phi} = \sum_{t=2}^{n} X_t X_{t-1} \Bigg/ \sum_{t=2}^{n} X_{t-1}^2 \tag{5.4}$$

and use this $\tilde{\phi}$ to estimate the least square estimates $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_\beta^2$ using

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=2}^{n} \hat{h}_t^2 X_{t-1}^2 \sum_{t=2}^{n} X_{t-1}^2 - \sum_{t=2}^{n} \hat{h}_t^2 \sum_{t=2}^{n} X_{t-1}^4}{(\sum_{t=2}^{n} X_{t-1}^2)^2 - (n-1)\sum_{t=2}^{n} X_{t-1}^4} \tag{5.5}$$

and

$$\hat{\sigma}_\beta^2 = \frac{(n-1)\sum_{t=2}^{n} \hat{h}_t^2 X_{t-1}^2 - \sum_{t=2}^{n} \hat{h}_t^2 \sum_{t=2}^{n} X_{t-1}^2}{(n-1)\sum_{t=2}^{n} X_{t-1}^4 - (\sum_{t=2}^{n} X_{t-1}^2)^2}, \tag{5.6}$$

where

$$\hat{h}_t^2 = X_t - \tilde{\phi} X_{t-1}.$$

We now use these least square estimates (5.5) and (5.6) in (5.3) to estimate $\hat{\phi}$. Let $\delta = (\phi, \sigma_\varepsilon^2, \sigma_\beta^2)$ and $\hat{\delta} = (\hat{\phi}, \tilde{\phi}, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_\beta^2)$, then the mean, variance, bias and mse of the estimate $\hat{\delta}$ for different values of δ are tabulated in Tables 14 and 15 for comparison.

δ	mean of $\hat{\delta}$	variance of $\hat{\delta}$
(0.1,1,0.8)	(0.1033,0.1014,1.4690,0.4505)	(0.0017,0.0048,0.1315,0.0127)
(0.1,0.5,0.8)	(0.1005,0.0870,0.3748,0.4566)	(0.0026,0.0069,0.0204,0.0116)
(0.2,0.6,0.2)	(0.1964,0.1935,0.0425,0.3123)	(0.0011,0.0017,0.0000,0.0054)
(0.3,0.2,1)	(0.3012,0.2629,0.2994,0.4609)	(0.0035,0.0216,0.0931,0.0318)
(0.4,1,0.3)	(0.4005,0.3989,1.0083,0.0796)	(0.0009,0.0010,0.0051,0.0017)
(0.5,1.5,0.4)	(0.4966,0.4961,2.2599,0.1559)	(0.0011,0.0012,0.0722,0.0047)
(0.5,0.2,0.6)	(0.4992,0.4951,0.0458,0.2872)	(0.0015,0.0023,0.0001,0.0113)
(0.6,0.8,0.4)	(0.6004,0.5977,0.6560,0.1476)	(0.0008,0.0011,0.0045,0.0025)
(0.7,0.1,0.3)	(0.6986,0.6990,0.0101,0.0809)	(0.0007,0.0008,0.0000,0.0010)
(0.7,0.8,0.4)	(0.6978,0.6913,0.6728,0.1407)	(0.0008,0.0010,0.0043,0.0018)
(0.8,0.1,0.8)	(0.7987,0.7988,0.6360,0.0089)	(0.0004,0.0004,0.0012,0.0001)
(0.9,0.4,0.2)	(0.9001,0.8984,0.1635,0.0365)	(0.0002,0.0003,0.0002,0.0002)

Table 14. Mean and variance of $\hat{\delta}$ for the RCA(1) model

δ	bias of $\hat{\delta}$	mse of $\hat{\delta}$
(0.1,1,0.8)	(0.0033,0.0014,0.4690,-0.3495)	(0.0017,0.0048,0.3501,0.1348)
(0.1,0.5,0.8)	(0.0005,-0.0130,-0.1252,-0.3434)	(0.0025,0.0070,0.0359,0.1294)
(0.2,0.6,0.2)	(-0.0036,-0.0065,-0.1575,-0.2877)	(0.0011,0.0017,0.0248,0.0882)
(0.3,0.2,1)	(0.0012,-0.0371,0.0994,-0.5391)	(0.0034,0.0228,0.1021,0.3221)
(0.4,1,0.3)	(0.0005,-0.0011,0.0083,-0.2204)	(0.0008,0.0010,0.0051,0.0503)
(0.5,1.5,0.4)	(-0.0034,-0.0039,0.7599,-0.2441)	(0.0011,0.0012,0.6490,0.0642)
(0.5,0.2,0.6)	(-0.0008,-0.0049,-0.1542,-0.3128)	(0.0015,0.0023,0.0239,0.1090)
(0.6,0.8,0.4)	(0.0004,-0.0023,-0.1440,-0.2524)	(0.0008,0.0011,0.0252,0.0662)
(0.7,0.1,0.3)	(-0.0014,-0.0010,-0.0899,-0.2191)	(0.0007,0.0008,0.0081,0.0490)
(0.7,0.8,0.4)	(-0.0022,-0.0087,-0.1272,-0.2593)	(0.0008,0.0011,0.0205,0.0691)
(0.8,0.1,0.8)	(-0.0013,-0.0012,-0.1640,-0.0911)	(0.0004,0.0004,0.0281,0.0084)
(0.9,0.4,0.2)	(0.0001,-0.0016,-0.2365,-0.1635)	(0.0002,0.0003,0.0561,0.0269)

TABLE 15. Simulated bias and mse of $\hat{\delta}$ for the RCA(1) model

5.2. Doubly Stochastic Model

Consider a doubly stochastic model of order (1,0) given by,

$$X_t = (\phi + \beta(t))X_{t-1} + \varepsilon_t, \tag{5.7}$$

where $\beta(t)$ is an MA(1) process satisfying

$$\beta(t) = e_t + e_{t-1}.$$

Assume that $\{\varepsilon_t\}$, $\{\beta(t)\}$ and $\{e_t\}$ are zero mean, square integrable, independent processes with constant variances σ_ε^2, σ_β^2 and σ_e^2. For convenience assume the normality of $\{\varepsilon_t\}$ and $\{e_t\}$.

Simulation study: Simulate a sample of 1500 from the doubly stochastic model given in (5.7) and evaluate the posterior mean $m_t = E_{t-1}[e_t]$ and variance $\gamma_t = E_{t-1}[(e_t - m_t)^2]$ using the algorithms:

$$m_t = \sigma_e^2 X_{t-1}[X_t - (\phi + m_{t-1})X_{t-1}]/\sigma_{t-1}^2 \tag{5.8}$$

and

$$\gamma_t = \sigma_e^2 - [\sigma_e^4 X_{t-1}^2/\sigma_{t-1}^2], \tag{5.9}$$

where $\sigma_{t-1}^2 = \sigma_\varepsilon^2 + (\sigma_\varepsilon^2 + \gamma_{t-1})X_{t-1}^2$.

Starting with initial values $\gamma_1 = \sigma_e^2$ and $m_1 = 0$, we obtain m_t and γ_t and the derivative

$$\frac{\partial m_t}{\partial \phi} = -\sigma_e^2 X_{t-1}^2(1 + (\partial m_{t-1}/\partial\phi))/\sigma_{t-1}^2 \tag{5.10}$$

recursively. Using these recursive estimates (5.8), (5.9) and (5.10), we can obtain the estimate $\hat{\phi}$ of ϕ as

$$\hat{\phi} = \frac{\sum_{t=2}^{n}\{[1 + (\partial m_{t-1}/\partial\phi)]X_t X_{t-1}/\sigma_{t-1}^2\}}{\sum_{t=2}^{n}\{[1 + (\partial m_{t-1}/\partial\phi)]X_{t-1}^2/\sigma_{t-1}^2\}}. \tag{5.11}$$

Repeat this calculation 10000 times to obtain the mean, variance, bias and mse for the different values of $\hat{\phi}$. These values are given in Table 16 with the corresponding true values of ϕ. The bias and the mse are given in Figures 23 and 24 respectively.

ϕ	mean of $\hat{\phi}$	variance of $\hat{\phi}$	Bias of $\hat{\phi}$	mse of $\hat{\phi}$
0.05	0.0694	0.0073	0.0194	0.0076
0.10	0.0946	0.0097	-0.0054	0.0096
0.15	0.1388	0.0096	-0.0112	0.0099
0.20	0.1961	0.0111	-0.0039	0.0110
0.25	0.2602	0.0095	0.0102	0.0095
0.30	0.3194	0.0078	0.0194	0.0081
0.35	0.3500	0.0100	0.0000	0.0099
0.40	0.4007	0.0072	0.0007	0.0071
0.45	0.4745	0.0087	0.0245	0.0092
0.50	0.5039	0.0072	0.0039	0.0072
0.55	0.5469	0.0080	-0.0031	0.0080
0.60	0.6092	0.0048	0.0092	0.0048
0.65	0.6533	0.0057	0.0033	0.0057
0.70	0.7052	0.0075	0.0052	0.0074
0.75	0.7482	0.0067	-0.0018	0.0066
0.80	0.7955	0.0045	-0.0045	0.0045
0.85	0.8496	0.0037	-0.0004	0.0037
0.90	0.9100	0.0071	0.0100	0.0071
0.95	0.9470	0.0023	-0.0030	0.0023

TABLE 16. Mean, variance , bias and mse of $\hat{\phi}$ for the doubly stochastic (1,0) model in (5.7)

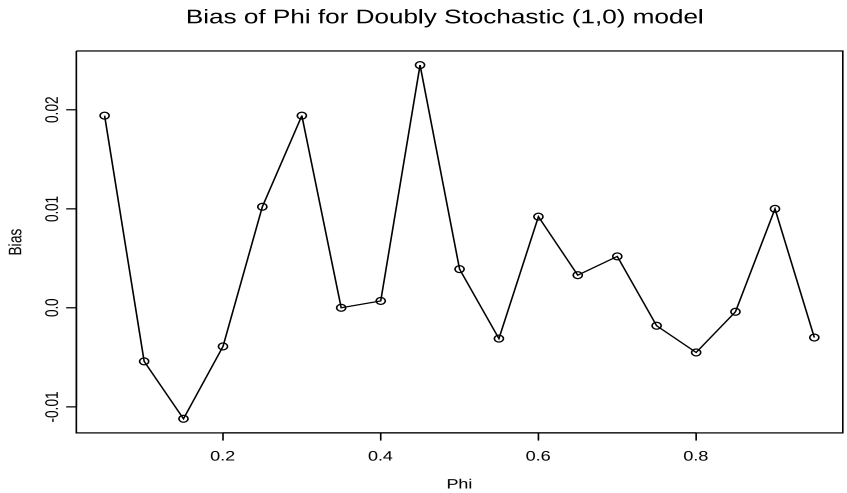


FIGURE 23. Plot of the bias of $\hat{\phi}$ in doubly stochastic (1,0) model

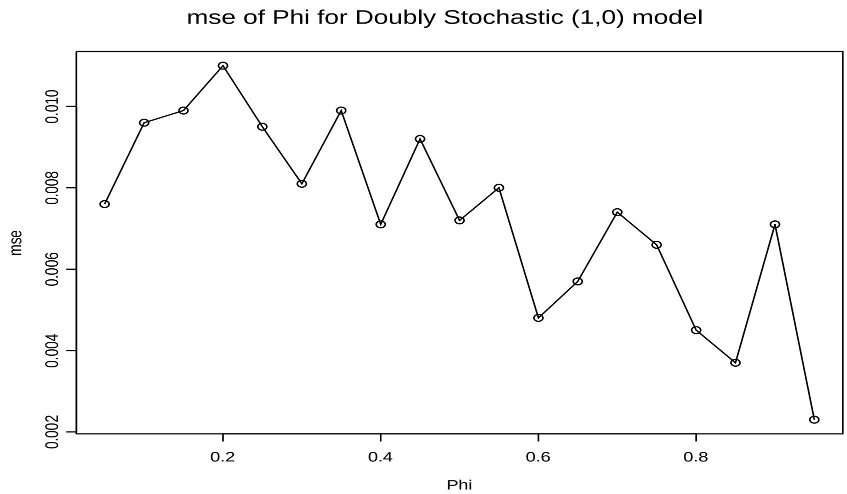


FIGURE 24. Plot of the mse of $\hat{\phi}$ in doubly stochastic (1,0) model

5.3. TAR Model

Consider a TAR(2;1,1) model given by

$$X_t = \begin{cases} \phi_0 + \phi_1 X_{t-1} + \varepsilon_t & \text{if } X_{t-1} \geq 0, \\ \phi_0 - \phi_1 X_{t-1} + \varepsilon_t & \text{if } X_{t-1} < 0. \end{cases}$$

This can be written as

$$X_t = \phi_0 + \phi_1 |X_{t-1}| + \varepsilon_t, \tag{5.12}$$

where the sequence $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$.

In this simulation study, simulate a sample of 1500 from this TAR(2;1,1) model. For given values of ϕ_0 and σ_ε^2, the corresponding optimal estimating function for ϕ_1 is

$$g_{\phi_1}^* = \sum_{t=2}^{n} -|X_{t-1}|[X_t - \phi_0 - \phi_1 |X_{t-1}|]/\sigma_\varepsilon^2$$

and the optimal estimate of ϕ_1 is

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n} X_t |X_{t-1}| - \phi_0 \sum_{t=2}^{n} |X_{t-1}|}{\sum_{t=2}^{n} X_{t-1}^2}. \tag{5.13}$$

We estimate $\hat{\phi}_1$ in equation (5.13) and repeat this 10000 times to calculate the mean, variance, bias and mse for the different values of ϕ_1. These values are given in Table 17 with the corresponding true values of ϕ_1. The bias and the mse are given in Figure 25 and Figure 26 respectively.

ϕ_1	mean of $\hat{\phi}_1$	variance of $\hat{\phi}_1$	Bias of $\hat{\phi}_1$	mse of $\hat{\phi}_1$
0.05	0.0501	0.0005	0.0001	0.0005
0.10	0.0992	0.0004	-0.0008	0.0005
0.15	0.1503	0.0004	0.0003	0.0004
0.20	0.1997	0.0004	-0.0003	0.0004
0.25	0.2495	0.0004	-0.0005	0.0004
0.30	0.2992	0.0003	-0.0008	0.0004
0.35	0.3497	0.0003	-0.0003	0.0004
0.40	0.3993	0.0003	-0.0007	0.0003
0.45	0.4495	0.0002	-0.0005	0.0003
0.50	0.4988	0.0002	-0.0012	0.0003
0.55	0.5492	0.0002	-0.0008	0.0002
0.60	0.5997	0.0001	-0.0003	0.0001
0.65	0.6490	0.0001	-0.0010	0.0001
0.70	0.6996	0.0001	-0.0004	0.0001
0.75	0.7497	0.0001	-0.0003	0.0002
0.80	0.7996	0.0000	-0.0004	0.0002
0.85	0.8501	0.0000	0.0001	0.0001
0.90	0.8997	0.0000	-0.0003	0.0000
0.95	0.9499	0.0000	-0.0001	0.0001

TABLE 17. Mean, variance, bias and mse of $\hat{\phi}$ for the TAR(2;1,1) model in (5.12)($\phi_0 = 1$)

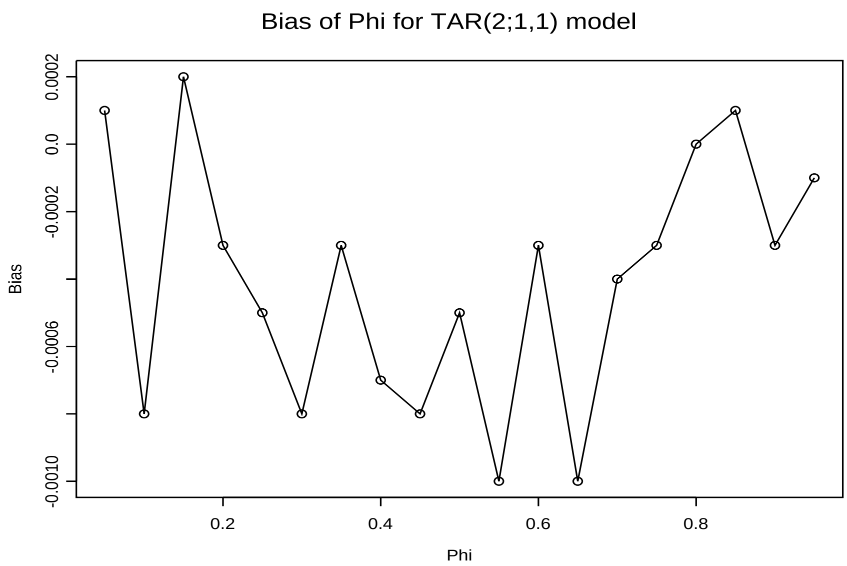


FIGURE 25. Plot of the bias of $\hat{\phi}$ in TAR(2;1,1) model

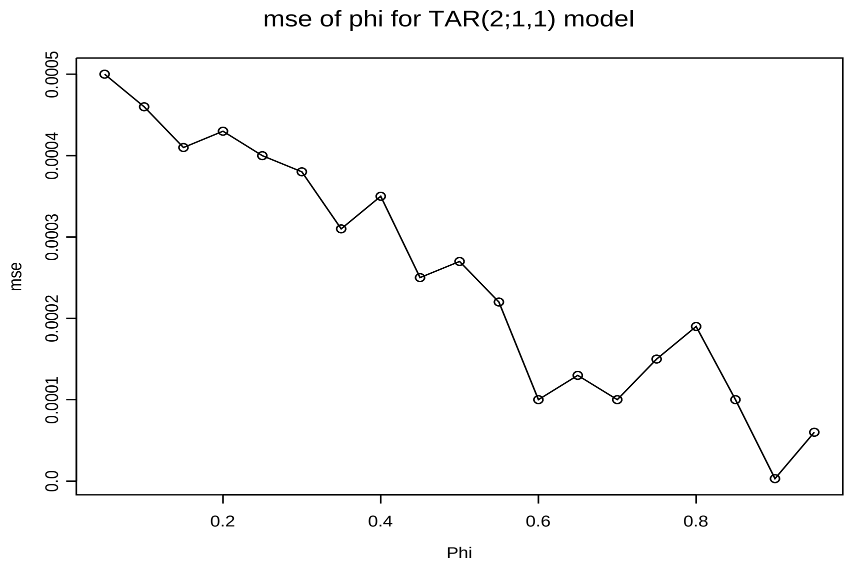


FIGURE 26. Plot of the mse of $\hat{\phi}$ in TAR(2;1,1) model

5.4. ARCH model

Consider an ARCH(1) model

$$
\begin{aligned}
X_t &= \sigma_t \varepsilon_t, \\
\sigma_t^2 &= \phi_0 + \phi_1 X_{t-1}^2, \qquad (5.14)
\end{aligned}
$$

where the sequence $\varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$. Here we simulate a sample of 1500 from the ARCH(1) model in (5.14). For given values of ϕ_0 and σ_ε^2, the optimal estimating function for ϕ_1 is

$$
g_{\phi_1}^* = \sum_{t=2}^{n} \frac{-X_{t-1}^2}{2[\phi_0 + \phi_1 X_{t-1}^2]^2} [X_t^2 - \phi_0 - \phi_1 X_{t-1}^2]
$$

and the optimal estimate $\hat{\phi}_1$ is obtained by solving the nonlinear equation $g_{\phi_1}^* = 0$ using the Newton-Raphson method. We repeat this 10000 times to calculate the mean, variance, bias and mse for the different values of ϕ_1. These values are given in Table 18 with the corresponding true values of ϕ_1 and the graphs for the bias in Figure 27 and the mse in Figure 28 are also given.

ϕ_1	mean of $\hat{\phi}_1$	variance of $\hat{\phi}_1$	Bias of $\hat{\phi}_1$	mse of $\hat{\phi}_1$
0.05	0.0483	0.0002	-0.0017	0.0002
0.10	0.0989	0.0002	-0.0011	0.0002
0.15	0.1474	0.0003	-0.0026	0.0003
0.20	0.1977	0.0003	-0.0023	0.0003
0.25	0.2471	0.0004	-0.0029	0.0004
0.30	0.2971	0.0004	-0.0029	0.0004
0.35	0.3461	0.0004	-0.0039	0.0004
0.40	0.3970	0.0004	-0.0030	0.0004
0.45	0.4478	0.0003	-0.0022	0.0003
0.50	0.4958	0.0004	-0.0042	0.0004
0.55	0.5470	0.0003	-0.0030	0.0003
0.60	0.5971	0.0003	-0.0029	0.0003
0.65	0.6468	0.0002	-0.0032	0.0003
0.70	0.6960	0.0003	-0.0040	0.0003
0.75	0.7469	0.0002	-0.0031	0.0002
0.80	0.7968	0.0002	-0.0032	0.0002
0.85	0.8472	0.0001	-0.0028	0.0001
0.90	0.8967	0.0001	-0.0033	0.0000
0.95	0.9480	0.0001	-0.0020	0.0001

TABLE 18. Mean, variance, bias and mse of $\hat{\phi}$ for the ARCH(1) model in (5.14)($\phi_0 = 5$)

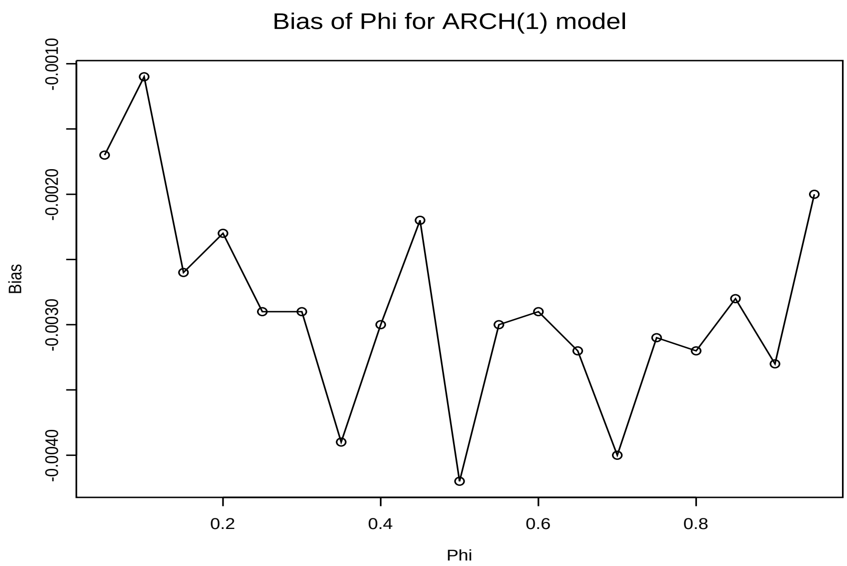


FIGURE 27. Plot of the bias of $\hat{\phi}$ in ARCH(1) model

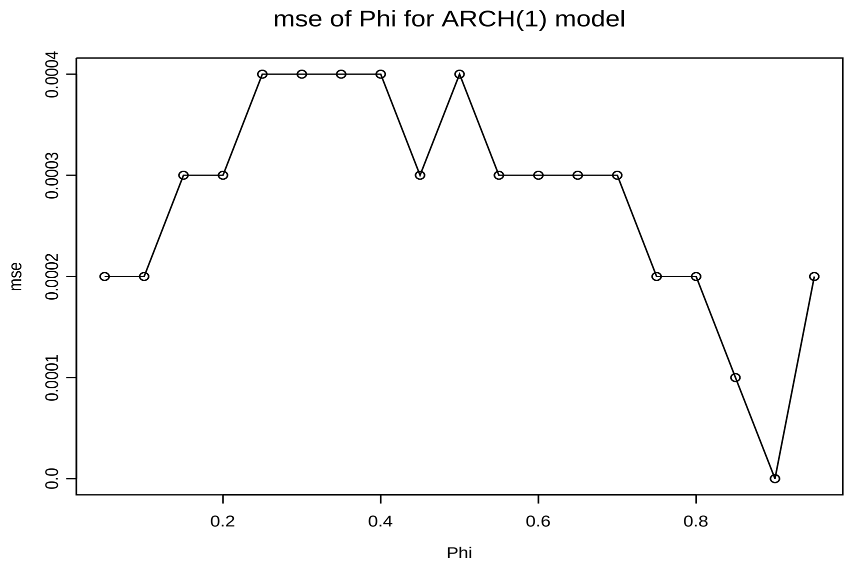


FIGURE 28. Plot of the mse of $\hat{\phi}$ in ARCH(1) model

5.5. Summary and conclusion

Chapters 5 of this thesis mainly investigated the estimating function approach in parameter estimation for four different nonlinear models in practice.

In this chapter we concentrated on the estimation of the parameter ϕ based on the theory on Chapter 4. The properties of $\hat{\phi}$ are intractable for finite samples and one way to investigate it is to use simulation methods. A limited simulation study was conducted for models mentioned on p.108 assuming all other parameters are known except ϕ. From each table it is interesting to note that $\hat{\phi}$ has small mse and bias in each case as expected. Tables and graphs for each model show these properties clearly. Mse appears to decrease with increasing ϕ values from 0 to 1. The theoretical reason for the above behaviour has not been investigated in this thesis and this will be an interesting future research object.

Below we report histograms related to our simulation study to show the finite sample behaviour of $\hat{\phi}$. Note that except in the ARCH(1) case the distribution of $\hat{\phi}$ looks fairly symmetrical. The ARCH(1) distributions appear to be left skewed with the skewness becoming more pronounced as the ϕ value increases.

Fig 29 - all four distributions appear to be centered around the parameter value with a large range of values.

Fig 30 - not as tightly clustered around mode as Fig 29 plots.

Fig 31 - lightly clustered around ϕ values - hint of skewness when ϕ=0.9.

Fig 32 - ϕ=0.1 pattern appears to be different to the other 3 not as strongly skewed.

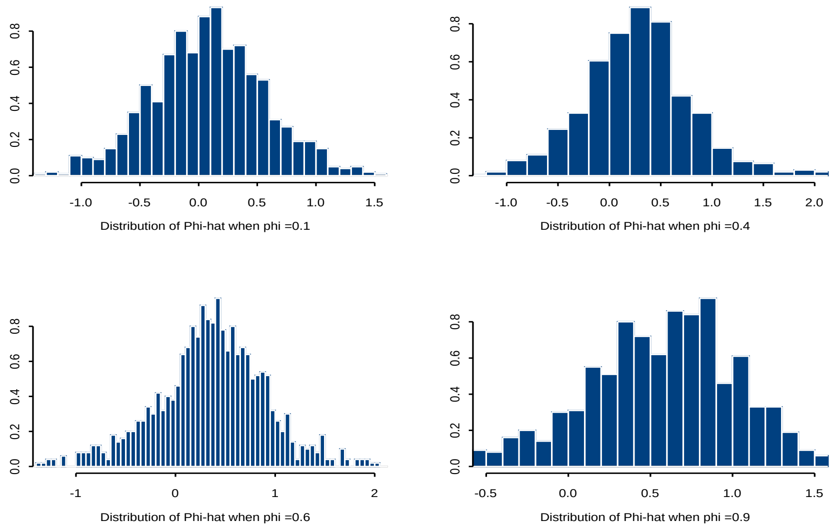


FIGURE 29. Distribution of $\hat{\phi}$ for the RCA(1) model

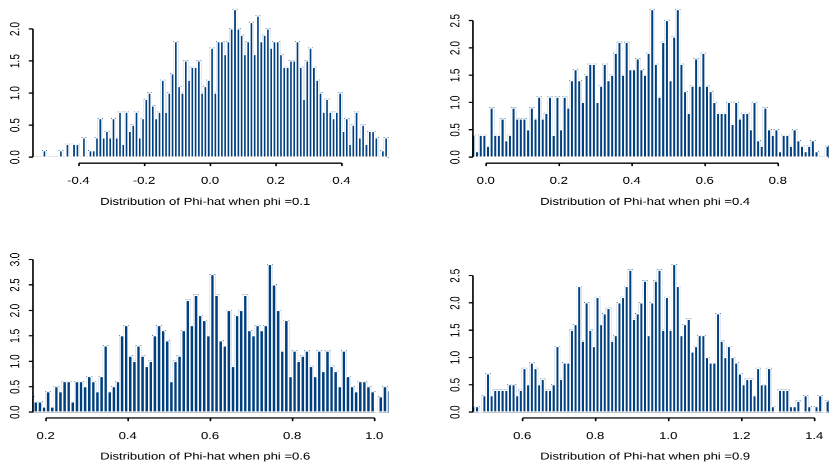


FIGURE 30. Distribution of $\hat{\phi}$ for the doubly stochastic (1,0) model

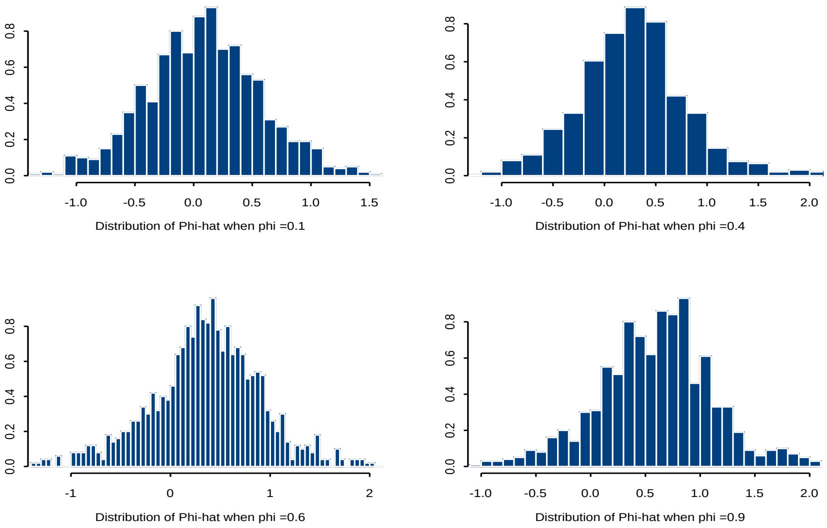


FIGURE 31. Distribution of $\hat{\phi}$ for the TAR((2;1,1) model

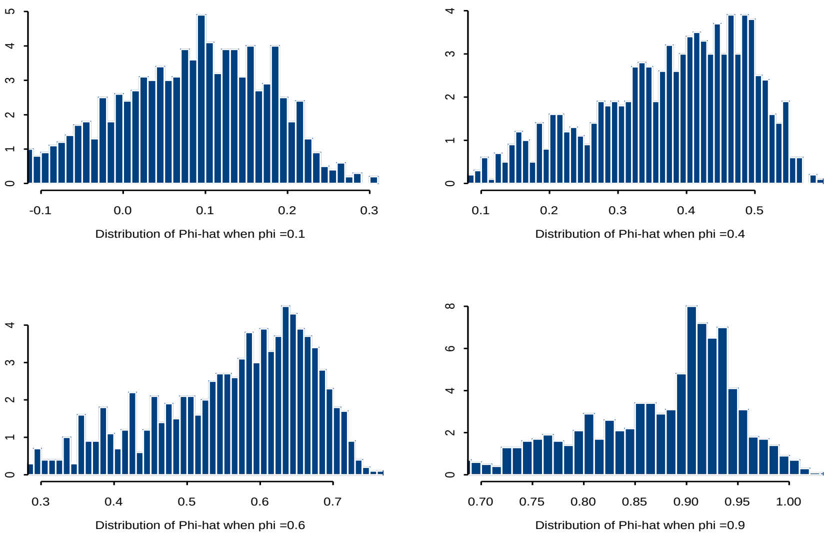


FIGURE 32. Distribution of $\hat{\phi}$ for the ARCH(1) model

References

[1] Abraham, B. and Ledolter, J. (1983). *Statistical Methods for Forecasting*. John Wiley, New York.

[2] Abraham, B., Thavaneswaran, A. and Peiris, S. (1997). On the prediction scheme for some nonlinear time series models using estimating functions, *IMS Lecture notes- Monograph series,* 259-271, I.V. Basawa, V.P.Godambe, R. Taylor (eds.).

[3] Abramson, I.S. (1982). On bandwidth variation in kernel estimates a square root law. *Ann. Math. Statist.,* **10**, 1217-1223.

[4] Ainkaran, P., Peiris, S. and Mellor, R. (2003). A note on the analysis of short AR(1) type time series models with replicated observations, *Current Research in Modelling, Data Mining and Quantitative Techniques*, University of Western Sydney Press, 143-156.

[5] Bai, X., Russell, R. and Tiao, G.C. (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innorvations, *Journal of Econometrics*, **114**, 349-360.

[6] Box, G.E.P and Jenkins, G.M. (1976), *Time Series analysis: Forecasting and Control,* Holden-Day, San Francissco.

[7] Conlisk, J. (1974). Stability in a random coefficient model. *Internat. Econom. Rev.*, **15**, 529-533.

[8] Conlisk, J. (1976). Interactive Markov chains. *J. Mathematical Sociology*, **4(2)**, 157-185.

[9] Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (1996). *Time Series Models: Monographs on Statistics and Applied Probability*, Chapman and Hall, London.

[10] Cox, D.R. and Solomon, P.J. (1988). On testing for serial correlation in a large number of small samples. *Biometrika*, **75(1)**, 145-148.

[11] Dai, Y. and Billard, L. (1998). A space time bilinear model and its identication. *Journal of Time Series Analysis*, **19**, 657-679.

[12] Dai, Y. and Billard, L. (2003). Maximum likelihood estimation in space time bilinear models, *Journal of Time Series Analysis*, **24(1)**, 25–44.

[13] Diebolt, J. and Gugan, D. (1991). The autoregressive time-series model β-ARCH, *C. R. Acad. Sci. Paris Sr. I Math.,* **312(8)**, 625–630.

[14] Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica*, **50**, 987-1008.

[15] Franses, P.H. and Van Dijk, D. (2000), *Non-linear Time Series Models in Empirical Finance*, Cambridge University Press.

[16] Godambe, V.P. (1960). An optimum property of regular maximum likelihood equations, *Ann. Math. Statist.*, **31**, 1208-1211 .

[17] Godambe, V.P.(1976). Conditional likelihood and unconditional optimum Estimating equations, *Biometrika*, **63**, 277-284.

[18] Godambe, V.P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika*, **72**, 419-428.

[19] Granger, C.W.J. and Anderson, A. (1978). *Non-linear Time Series Modelling. Applied Time Series Analysis*. Academic Press, New York.

[20] Ha, J. and Lee, S. (2002). Coefficient constancy test in AR-ARCH models. *Statistics and Probability Letters*, **57(1)**, 6577.

[21] Hili, O. (1999). On the estimate of β-ARCH models, *Statistics and Probability Letters*, **45**, 285-293.

[22] Hjellivik, V. and Tjøsthiem, D. (1999). Modelling panels of intercorrelated autoregressive time series. *Biometrika*, **86(3**), 573-590.

[23] Holst, U., Lindgren, G., Holst, J. and Thuvesholmen, M. (1994). Recursive estimation in switching autoregressions with a Markov regime, *Journal of Time Series Analysis*, **15(5)**, 489-506.

[24] Kalman, R.E. (1960). A new approach to linear filtering. *Journal of Basic Engineering, Trans. ASM*, **D 82**, 35-45.

[25] Lee, O. (1997). Limit theorems for some doubly stochastic processes, *Statistics and Probability Letters*, **32**, 215-221.

[26] Lee, S. (1998). Coefficient consistency test in a random coefficient autoregressive model, *Journal of Statistical Planning and Inference*, **74**, 93-101.

[27] Lu, Z. (1996). A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) models, *Statistics and Probability Letters*, **30**, 305-311.

[28] Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH$(1, 1)$

and covariance stationary GARCH(1, 1) models, *Econometrica*, **64(3)**, 575-596.

[29] Nadaraja, E.A. (1974). On integrated mean square error of some nonparametric estimates for the density function, *Theory Probab. Appl.,* **19**, 133-141.

[30] Nicholls, D.F. and Quinn, B.G. (1982). *Random Coeffcient Autoregressive Models: An Introduction.* Lecture Notes in Statistics, **Vol. 11**. Springer, New York.

[31] Novak, S.Y. (2000). Generalized kernel density estimator, *Theory of Probability and Applications,* **90**, 301-306.

[32] Parzen, E. (1962). On estimation of probability density and mode, *Ann. Math. Statist.,* **33**, 1065-1076.

[33] Peiris, S., Mellor, R. and Ainkaran, P.(2003). Maximum likelihood estimation for short time series with replicated observations: A simulation study, *InterStat*, **9**, 1-16.

[34] Peiris, M. S. and Thavaneswaran, A. (2004). A note on the filtering for some time series models, *Journal of Time Series Analysis*, **25(3)**, 397-407.

[35] Priestley, M.B. (1988). *Non-linear and Non-stationary Time Series Analysis*, Academic Press, New York.

[36] Pourahmadi, M. (1986). On stationarity of the solution of a doubly stochastic model, *Journal of Time Series Analysis*, **7(2)**, 123–131.

[37] Rai, S.N., Abraham, B. and Peiris, S. (1995). Analysis of short time series with an over-dispersion model, *Commun. Statist. Theory. Meth.,* **24(2)**, 335-348.

[38] Ramanathan, T.V. and Rajarshi, M.B. (1994). Rank test for testing the randomness of autoregressive coefficients, *Statistics and Probability Letters*, **21**, 115-120.

[39] Robinson, P.M. (1978). Statistical inference for a random coefficient autoregressive model, *Scandinavian Journal of Statistics*, **5**, 163-168.

[40] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.,* **27**, 832-835.

[41] Searle, S.R. (1971). *Linear Models*, Wiley, New York.

[42] Shephard, N. (1996). Statistical Aspects of ARCH and Stochastic Volatility, *Time Series Models,* 1-55, D.R. Cox, D.V. Hinkley, O.E. Brandorff-Nielson (eds.).

[43] Shiryayev, A.N. (1984). *Probability.* Springer, New York.

[44] Slenholt, B.K. and Tjøstheim, D. (1987). Multiple bilinear time series models, *Journal of Time Series Analysis*, **8(2)**, 221-233.

[45] Staniswalis, J.G. (1989). The kernel estimate of a regression function in likelihood based models, *J. Amer. Statist. Assoc.*, **84**, 276-283.

[46] Subba Rao, T. (1981). On the theory of bilinear series models, *J. of the Royal Stat. Soc. B*, **43**, 244-255.

[47] Thavaneswaran, A. and Abraham, B. (1988). Estimation for nonlinear time series models using estimating equations, *Journal of Time Series Analysis*, **9**, 99-108.

[48] Thavaneswaran, A., Appadoo, S. S. and Samanta, M. (2004). Random coefficient GARCH models, *Mathematical and Computer Modelling*, (to appear).

[49] Thavaneswaran, A. and Peiris, S. (1996). Nonparametric estimation for some nonlinear models, *Statistics and Probability Letters*, **28**, 227-233.

[50] Thavaneswaran, A. and Peiris, S. (2004). Smoothed estimates for models with random coefficients and infinite variance, *Mathematical and Computer Modelling*, **39**, 363-372.

[51] Thavaneswaran, A., Peiris, S. and Ainkaran, P. (2003). Smoothed estimating functions for some stochastic volatility models, submitted.

[52] Thavaneswaran, A. and Peiris, S. (2003). Generalized smoothed estimating functions for nonlinear time series, *Statistics and Probability Letters*, **65** , 51-56.

[53] Thavaneswaran, A. and Peiris, S. (1998). Hypothesis testing for some time-series models: a power comparison. *Statistics and Probability Letters*, **38** , 151-156.

[54] Thavaneswaran, A. and Singh, J. (1993). A note on smoothed estimating functions, *Ann. Inst. Statist. Math.*, **45**, 721-729.

[55] Tjøstheim, D. (1986). Some doubly stochastic time series models. *Journal of Time Series Analysis*, **7(1)**, 51-72.

[56] Tong, H. (1983). *Threshold Models in Nonlinear Time Series Models.* Springer Lecture Notes in Statistics, **21**.

[57] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press, Oxford.

[58] Tong, H. and Lim, K.S. (1980). Threshold autoregression, limits, cycles and cyclical data , *J. Royal Stat. Soc.*, **B, 42**, 245-292.

[59] Tsay, R.S. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.*, **84(405)**, 231-140.

[60] Wiener, N. (1958). *Nonlinear Problems in Random Theory*, John Wiley , New York.

[61] Wu, B. and Chang, C. (2002). Using genetic algorithms to parameters (d, r) estimation for threshold autoregressive models, *Computational Statistics and Data Analysis,* **38(3)**, 315–330.