STATE SPACE MODELS AND ESTIMATION OF MISSING OBSERVATIONS IN TIME SERIES

By

Biwott Daniel Kiprotich

A project submitted for the partial fulfillment of the degree of **Master of Science** in mathematical statistics, of

Kenyatta University

July, 2000

DECLARATION

This is my original work and has not been presented for a degree award in any other University in part or whole.

Sign---

DANIEL KIPROTICH BIWOTT

This work has been presented with my approval as the University Supervisor.

Sign---

DR. LEO O. ODONGO

Department of mathematics

Kenyatta University.

ABSTRACT

In this project we have considered a non-linear time series model which encompasses several standard non-linear models for time series as special cases. It also offers two methods for estimating missing observations, one using prediction and fixed point smoothing algorithm and the other using optimal estimating equation theory.

Recursive estimation of missing observations in an *Autoregressive Conditionally Heteroscedastic* (ARCH) model and the estimation of missing observations in a linear time series model are shown to be special cases.

For the case of prediction and fixed point smoothing algorithm, we have generalised the formula developed by Abraham and Thavaneswaran (1991) for estimating missing observations to a case when there are more than two missing observations. Simulation studies have been carried out on AR(1) data to illustrate the application of the formula.

DEDICATION

I dedicate this work to my parents, Mr. Samwel Kibiwot Samoei and Mrs. Tomdila Jepkoech Samoei, for their tireless effort in sponsoring my studies at the university up to this level.

ACKNOWLEDGEMENTS

My sincere thanks first goes to the Almighty GOD for His unfailing provision, protection, sustenance and the unmerited mercy, He gave me during my course of study and helping my parents go through hectic experience of meeting the huge burden of my tuition throughout my education.

Secondly, I owe my gratitude to my parents, my wife Nancy and our Son Dennis. Not forgetting the entire Samoei`s family for their love, patience and encouragement they gave me throughout my education.

I also thank the entire university for offering me a chance to pursue university education.

I am greatly indebted to Dr. Odongo, who apart from being my dedicated Supervisor, has been a constant source of inspiration as a great teacher and researcher. I am happy to testify that he helped me in everything except making mistakes. Any errors and omissions that remain are entirely my responsibility.

My appreciation also go to my University lecturers; Dr. C. Wafula, Dr. Njenga, Dr. Jairu and the rest of the university lecturers for their encouragement and for freely sharing their wisdom whenever I sought their assistance. I cannot forget to give further appreciation to Dr. C. Wafula for the way he taught me Time Series analysis course which aroused my interest in the choice of topic. Not forgetting Mr. Benson Atichi Alati who despite the Computer Labs being closed made sure that we got access to a computer screen, to enable us complete our typing work in time and Mr. Wesley Kibet for allowing me print this project using his computer.

I am extremely grateful to my colleagues in the struggle, Otwombe, Mulati, Mumbua and Kisaka for their company, assistance and suggestions throughout this course. I regret that there is no way except a general expression of gratitude in which I can give thanks to friends, siblings and relatives who encouraged me, kindly made inquiries about my progress, and voiced their hopes for successful completion of this course. To you all, may GOD bless you.

Table of Contents

	Page
Declaration	ii
Abstract	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vii
List of Abbreviations and Notations	x
List of Tables	xii
List of Figures	xiii
Overview	xiv
CHAPTER ONE: FUNDAMENTAL CONCEPTS AND DEFINITIONS	**1**
1.0: Introduction	1
1.1: Stationary Time Series	1
1.2: The ACVF and ACF	2
1.3: Sample Autocorrelation Function (SACF)	5
1.4: Sample Partial Autocorrelation Function (SPACF)	7
1.5: Some Useful Stochastic Processes	7
1.6: ARMA process	12

1.7: Integrated ARIMA Model 14

1.8: Estimation of parameters 15

CHAPTER TWO: STATE SPACE MODELS AND KALMAN FILTER 19

2.0: Introduction 19

2.1: State Space representation 20

2.2: The Kalman Filter 23

2.3: Recursive estimation of state and Likelihood computation for state space models 25

2.4: State Space model for Time Series with non-stationarity and irregularity 27

2.5: Missing and outlying observations in Time Series 28

2.6: State Space representation to Maximum Likelihood fitting of Bilinear models with missing observations 29

CHAPTER THREE: ESTIMATION OF MISSING OBSERVATIONS 33

3.0: Introduction 33

3.1: Nonlinear State Space models 34

3.2: Application to missing data 41

3.3: (ARCH) type model with One missing observation 41

3.4: Two consecutive missing observations 45

3.5: Three consecutive missing observations 51

3.6: Optimal estimation of missing observations 62

CHAPTER FOUR: AN EMPIRICAL STUDY 67

4.0: Introduction 67

4.1: Estimation of missing observations on simulated
AR(1) data 67

4.2: Discussion of tables 85

CHAPTER FIVE: CONCLUSION AND ACCOMPLISHMENTS 86

5.0: Introduction 86

5.1: Achievements 86

5.2: Suggestions 87

5.3: Conclusion 88

REFERENCES 89

Abbreviations and Notations

ACF	Autocorrelation Function
ACVF	Autocovariance Function
AIC	Akaike Information Criteria
AR	Autoregressive
ARCH	Autoregressive conditionally Heteroscedatic Model
ARIMA	Autoregressive Integrated Moving Average
B	The Backward Shift Operator
BL	Bilinear Models
E	Expected Value
MA	Moving Average
M.L.E.	Maximum Likelihood Estimate
op	Optimal Estimate
PACF	Partial Autocorrelation Function
RCA	Random Coefficient Autoregression
SACF	Sample Autocorrelation Function
SPACF	Sample Partial Autocorrelation Function
∇	Difference Operator

F_t^y	The σ-field generated by observations y up to time t
$\mathrm{E}(\theta_t \mid F_{t-1}^y)$	The expectation of θ_t given σ-field generated by observations y up to time $t-1$
$X(t)$	Continuous parameter Time Series.
X_t	Discrete parameter Time Series
$\widetilde{X}_{m+j\|m+i}$	The estimate of $(m+j)th$ observation based on X_{m+i}

List of Tables

Page

Table 4.1: Simulated AR (1) data when $\phi = 0.86$ and $X_0 = 100$ 69

Table 4.2: One missing value at position 5,10, and 15 with their estimates 70

Table 4.3: Two consecutive missing values from positions 4 to 5, 11 to 12 and 15 to16 with their estimates 71

Table 4.4: Three consecutive missing values from positions 5 to 7, 10 to 12 and 15 to17 with their estimates 72

Table 4.5: Five consecutive missing values from positions 5 to 9, and 15 to19 with their estimates 73

Table 4.6: Ten consecutive missing values from positions 5 to 14 with their estimates 74

Table 4.7: Eighteen consecutive missing values from positions 2 to 19 with their estimates 75

Table 4.8: Simulated AR(1) data when $\phi = 0.5$ and $X_0 = 1000000$ 76

Table 4.9: One missing value at position 5,10, and 15 with their estimates 77

Table 4.10: Two consecutive missing values from positions 5 to 6, 10 to 11 and 15 to16 with their estimates 78

Table 4.11: Three consecutive missing values from positions 5 to 7, 10 to 12 and 15 to17 with their estimates 79

Table 4.12: Five consecutive missing values from positions 5 to 9, and 15 to19 with their estimates 80

Table 4.13: Ten consecutive missing values from positions 5 to 14 with their estimates 81

Table 4.14: Eighteen consecutive missing values from positions 2 to 19 with their estimates 82

List of figures

Fig. 1: Time plot for actual and estimated data for eighteen missing observations when $\phi = 0.86$ and $X_0 = 100$ 83

Fig.2: Time plot for actual and estimated data for eighteen missing observations when $\phi = 0.5$ and $X_0 = 1000000$ 84

OVERVIEW

One of the unfortunate facts facing data analysts is missing data. Data that are known to have been observed erroneously can fairly safely be categorized as missing. Erroneous data can also wreak havoc with the estimation and forecasting of time series models.

In the past, estimation of missing observations has been considered among others by Ansley & Kohn (1983), Jones (1985), Abraham & Thavaneswaran (1991), Penza & Shea (1997) and Nassiuma (1999). In particular, Palma & Chan (1997) and Chan & Palma (1998) develop state space method for dealing with missing observations in the Long – memory context. More recently Palma & Del Pino (1999) have given statistical analysis of incomplete Long-range dependent data and the application of these procedures to the analysis of the annual minimum water levels of Nile River.

This project addresses both theoretical and methodological issues related to the estimation of missing observations. Estimates are calculated by means of state space models and Kaman filter. These techniques are illustrated with statistical analysis of simulated data.

In this project we have extended the formula derived in Abraham & Thavaneswaran (1991) to estimate two or more missing observations. The formula is then applied on simulated data on AR (1) model.

Chapter one introduces the fundamental concepts and definitions required in understanding the rest of the work in this project. Chapter two discusses state space representation and Kalman filter of some Time Series models. Estimation of missing observations by means of state space models and Kalman filter are indicated.

In chapter three we have extended the methods of Abraham & Thavaneswaran (1991) to estimate more than two consecutive missing observations. Optimal estimation method is also considered. In chapter four we carry out an empirical study based on results of chapter three. Here we estimate the missing observations artificially created on an AR (1) simulated data.

CHAPTER ONE

FUNDAMENTAL CONCEPTS AND DEFINITIONS IN TIME SERIES ANALYSIS

1.0: Introduction

A Time Series is a collection of observations made sequentially in time. Examples occur in a variety of fields, ranging from economics to engineering. Methods of analyzing time series constitute an important area of statistics .

The theory and practice is very different from other branches of statistics in that;

(i) In a time series the observations are dependent (or correlated) with time and the order of the observation is therefore important. Hence, classical statistical procedures and techniques that rely on independence assumptions are no longer applicable.

(ii) Asymptotic theory generally tends to be meaningful in practice. This is not the case with time series. In fact the temporal or spatial dependence in time series is what is of interest and importance Anderson (1982).

There are several objectives of analyzing a time series, which include; description, explanation, prediction and control.

1.1: Stationary time series

A time series is said to be stationary if there is no systematic change in mean (i.e. no trend), if there is no systematic change in variance, and if strictly periodic variations have been removed.

We now give a mathematical definition of a stationary time series;

Definition I: A time series is said to be stationary in the *strict* sense or *strong* sense iff;

$$F\left(x_{t_1}, x_{t_2}, \ldots, x_{t_k}\right) = F\left(x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\right) \ldots\ldots(1.1.1)$$

for all $(t_1, t_2, \ldots, t_k)$ and h. That is the probability structure of the process should not evolve with time, that is shifting the time origin by an amount h has no effect on the joint distributions, which must therefore depend only on the time intervals $t_1, t_2, \ldots, t_k$.

Definition II: A process is said to be stationary in the *weak* sense (second order stationary) if its mean is a constant independent of time but depends only on the lag h between the variables. The mean and autocovariance functions must be finite.

$$\mathrm{E}(X_t) = \mu \ldots\ldots(1.1.2)$$

and $Cov(X_t X_{t+h}) = \sigma(h) \ldots\ldots(1.1.3)$.

Since the mean of a stationary process is a constant, we can conveniently take it to be zero. Hence, in the case $\sigma(h) = \mathrm{E}(X_t X_{t+h})$ we note that $\sigma(h)$ is influence by unit of measurement but to compare properties of different time series. We need a function that does not depend on unit of measurement and such a function is the theoretical autocorrelation function (ACF) given by

$$\rho(h) = \frac{\sigma(h)}{\sigma(0)} = \frac{Cov(X_t X_{t+h})}{Var(X_t)} \ldots\ldots(1.1.4)$$

1.2: The Autocovariance (ACVF) and Autocorrelation (ACF) Function.

For a stationary process X_t we have the mean $\mathrm{E}(X_t) = \mu$ and variance $Var(X_t) = \mathrm{E}(X_t - \mu)^2 = \sigma^2$ which are constants and the covariance $Cov(X_t X_s)$ which are functions only of the time difference $|t - s|$. Hence in this case, we write the covariance between X_t and X_{t+h} as

$$\sigma(h) = Cov(X_t, X_{t+h}) = \mathrm{E}(X_t - \mu)(X_{t+h} - \mu) \ldots\ldots(1.2.1)$$

and the autocorrelation between X_t and X_{t+h} as

$$\rho(h)=\frac{Cov(X_t,X_{t+h})}{\sqrt{Var(X_t)Var(X_{t+h})}}=\frac{\sigma(h)}{\sigma(0)}\ldots\ldots(1.2.2),$$

where we note that $Var(X_t)=Var(X_{t+h})=\sigma(0), h=0$. As functions of *h*, $\sigma(h)$ is called the Autocovariance Function (ACVF) and $\rho(h)$ is called the Autocorrelation Function (ACF) in time series analysis, since they represent the covariance and correlation between X_t and X_{t+h} from the same process, separated only by *h* time lags.

It is easy to see that, for a stationary process that the autocovariance function $\sigma(h)$ and the autocorrelation function $\rho(h)$ have the following properties Wei (1990);

1. $\sigma(0)=Var(X_t);\ldots\ \rho(0)=1$
2. $|\sigma(h)|\le\sigma(0);\ldots\ |\rho(h)|\le 1$ (1.2.3)
3. $\sigma(h)=\sigma(-h)$ and $\rho(h)=\rho(-h)$ for all h

i.e. $\sigma(h)$ and $\rho(h)$ are even functions and hence symmetric about the time origin $h=0$. This follows from the fact that the time difference between X_t and X_{t+h} and X_t and X_{t-h} are the same. Therefore the ACF is often plotted only for the non-negative lags such a plot of $\rho(h)$ against *h* is sometimes called a correlogram.

Another important property of the $\sigma(h)$ and $\rho(h)$ is that they are positive semi-definite in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\sigma|t_i-t_j|\ge 0\ldots\ldots(1.2.4)$$

and $$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho|t_i-t_j|\ge 0\ldots\ldots(1.2.5)$$

for any set of time points $t_1,t_2,\ldots,t_n$ and any real numbers $\alpha_1,\alpha_2,\ldots,\alpha_n$. A necessary condition for a function to be the ACVF or ACF function of a process is that, it be positive semi-definite.

In addition to ACF between X_t and X_{t+h} we may want to investigate the correlation between X_t and X_{t+h} after their mutual linear dependency on the intervening variables $X_{t+1}, X_{t+2}, \ldots, X_{t+h-1}$ has been removed. This is the following conditional correlation

$$Corr(X_t X_{t+h} \mid X_{t+1}, X_{t+2}, \ldots, X_{t+h-1}) \qquad (1.2.6)$$

referred to as *Partial Autocorrelation Function* (PACF) in time series analysis.

PACF can be derived by considering the regression model, where the dependent variable X_{t+h} from a zero mean stationary process is regressed on h lagged variables $X_{t+h-1}, X_{t+h-2}, \ldots, X_t$ i.e.

$$X_{t+h} = \phi_{h1} X_{t+h-1} + \phi_{h2} X_{t+h-2} + \ldots + \phi_{hh} X_t + e_{t+h} \qquad (1.2.7),$$

where ϕ_{hi} denote the *ith* regression parameter and e_{t+h} is a normal error term uncorrelated with $X_{t+h-j}, j \geq 1$. Multiplying X_{t+h-j} on both sides of the above regression equation and taking the expectation, we get

$$\sigma_i = \phi_{h1}\sigma_{i-1} + \phi_{h2}\sigma_{i-2} + \ldots + \phi_{hh}\sigma_{i-h} \qquad (1.2.8)$$

and hence

$$\rho_j = \phi_{hi}\rho_{j-1} + \phi_{h2}\rho_{j-2} + \ldots + \phi_{hh}\rho_{j-h} \qquad (1.2.9)$$

for $j = 1,2,\ldots,h$ we have the following system of equations

$$\left.\begin{array}{l} \rho_1 = \phi_{h0}\rho_0 + \phi_{h1}\rho_1 + \ldots + \phi_{hh}\rho_{h-1} \\ \rho_2 = \phi_{h1}\rho_1 + \phi_{h2}\rho_0 + \ldots + \phi_{hh}\rho_{h-2} \\ . \\ . \\ . \\ \rho_h = \phi_{h1}\rho_{h-1} + \phi_{h2}\rho_{h-2} + \ldots + \phi_{hh}\rho_0 \end{array}\right\} \qquad (1.2.10).$$

Using cramers rule successively for $h = 1,2,\ldots,$

we have

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

.

.

.

$$\phi_{hh} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{h-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{h-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \rho_{h-3} & \cdots & \rho_1 & \rho_h \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{h-2} & \rho_{h-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{h-3} & \rho_{h-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \rho_{h-3} & \cdots & \rho_1 & 1 \end{vmatrix}} \quad \text{.....................................}(1.2.11)$$

For alternative definition see Wei (1990: sect.2.3)

The practical autocorrelation between X_t and X_{t+h} can be obtained as the regression coefficient associated with X_t, when regressing X_{t+h} on its h lagged variables $X_{t+h-1}, X_{t+h-2}, ..., X_t$ as in (1.2.7). Because ϕ_{hh} has become a standard notation for the partial autocorrelation between X_t and X_{t+h} in time series literature. As a function of h , ϕ_{hh} is equally referred to as the partial autocorrelation function (PACF).

1.3: Sample Autocorrelation Function (SACF).

For a given observed time series $X_1, X_2, ..., X_n$ the sample ACF is defined as

$$\hat{\rho}_h = \frac{\hat{\sigma}(h)}{\hat{\sigma}(0)} = \frac{\sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}, h = 0,1,2,...,n, \text{.................................}(1.3.1),$$

where $\bar{X}=\sum_{t=1}^{n}\frac{X_t}{n}$, the sample mean of the series. A plot of $\hat{\rho}_h$ versus h is sometimes called a sample correlogram. For a stationary Gaussian process, Bartlett (1946) has shown that for $h>0$, and $h+j>0$

$$Cov(\hat{\rho}_h,\hat{\rho}_{h+j})\approx\frac{1}{n}\sum_{i=-\infty}^{\infty}\begin{pmatrix}\rho_i\rho_{i+j}+\rho_{i+h}\rho_{i-h}-2\rho_h\rho_i\rho_{i-h-j}- \\ 2\rho_{h+j}\rho_i\rho_{i-h}+2\rho_h\rho_{h+1}\rho_i^2\end{pmatrix}\text{..........................}(1.3.2).$$

For large n, $\hat{\rho}_h$ is approximately normally distributed with mean ρ_h and variance

$$Var(\hat{\rho}_h)\approx\frac{1}{n}\sum_{i=-\infty}^{\infty}\left(\rho_i^2+\rho_{i+h}\rho_{i-h}-4\rho_h\rho_i\rho_{i-h}+2\rho_h^2\rho_i^2\right)\text{..........................}(1.3.3)$$

for a process in which $\rho_h=0$ for h>m Bartlett`s approximation of (1.3.3) becomes

$$Var(\hat{\rho}_h)\approx\frac{1}{n}\left(1+2\rho_1^2+2\rho_2^2+\ldots+2\rho_m^2\right)\text{..........................}(1.3.4)$$

In practice, $\rho_i(i=1,2,\ldots,m)$ are unknown and are replaced by their sample estimates $\hat{\rho}_i$ and we have the following large-lag standard error of $\hat{\rho}_h$,

$$S_{\hat{\rho}_h}=\sqrt{\frac{1}{n}\left(1+2\hat{\rho}_1^2+\ldots+2\hat{\rho}_m^2\right)}\text{..........................}(1.3.5)$$

To test a white noise process we use

$$S_{\hat{\rho}_h}=\sqrt{\frac{1}{n}}\text{..........................}(1.3.6).$$

Note that ,

$$\hat{\rho}_h=\frac{\sum_{t=1}^{n-h}(X_t-\bar{X})(X_{t+h}-\bar{X})}{\sum_{t=1}^{n}(X_t-\bar{X})^2}$$

$$=\frac{\sum_{t=h+1}^{n}(X_t-\bar{X})(X_{t-h}-\bar{X})}{\sum_{t=1}^{n}(X_t-\bar{X})}=\hat{\rho}_{-h}\text{..........................}(1.3.7).$$

We see that, the sample ACF is also symmetric about the origin h=0.

1.4: Sample Partial Autocorrelation Function (SPACF).

The sample (PACF) $\hat{\phi}_{hh}$ is obtained by substituting ρ_i by $\hat{\rho}_i$ in equation (1.3.2). Instead of calculating the complicated determinants for large h in (1.3.2) a recursive method starting with $\hat{\phi}_{11} = \hat{\rho}_1$ for computing $\hat{\phi}_{hh}$ has been given by Durbin (1960) as follows;

$$\hat{\phi}_{h+1,h+1} = \frac{\hat{\rho}_{h+1} - \sum_{j=1}^{h} \hat{\phi}_{hj} \hat{\rho}_{h+1-j}}{1 - \sum_{j=1}^{h} \hat{\phi}_{hj} \hat{\rho}_j} \quad \ldots\ldots(1.4.1)$$

and $\hat{\phi}_{h+j,j} = \hat{\phi}_{hj} - \hat{\phi}_{h+1,h+1} \hat{\phi}_{h,h+1-j}, j = 1,2,\ldots,h,$ ……(1.4.2)

The method hold also for calculating the theoretical PACF ϕ_{hh}. Under the hypothesis that the underlying process is a white noise sequence the variance of $\hat{\phi}_{hh}$ can be approximated by

$$Var\left(\hat{\phi}_{hh}\right) \approx \frac{1}{n} \quad \ldots\ldots(1.4.3)$$

Hence , $\pm 2/\sqrt{n}$ can be used as a critical limits on ϕ_{hh} to test the hypothesis of a white noise process.

1.5: Some useful stochastic processes.

This section describe several different types of stochastic processes (see Chatfield (1989)) which are sometimes useful in setting up a model for a time series.

(i) A purely random process (White noise)

A discrete-time process is called a purely random process if it consists of a sequence of random variable $\{X_t\}$ which are mutually independent and identically distributed. From the definition it follows that the process has a constant mean and variance and that

$$\sigma(h)=\begin{cases}Cov(X_t,X_{t+h}) & \\ 0 & h=\pm1,\pm2,.........\end{cases}$$

As the mean and ACVF do not depend on time, the process is second-order stationary. The ACF is given by

$$\rho(h)=\begin{cases}1 & h=0\\ 0 & h=\pm1,\pm2,.......\end{cases}$$

A purely random process is sometimes called a white noise, particularly by engineers. Process of this type are useful in many situations as building blocks for more complicated process such as MA process.

(ii) Random Walk

Suppose that (e_t) is a discrete purely random process with mean μ and variance σ^2. A process $\{X_t\}$ is said to be a random walk if

$$X_t = X_{t-1}+e_t, X_0 = 0 \qquad \text{.........(1.5.1)}$$

The process is customarily started at zero when t = 0 so that

$$X_1 = X_0 + e_1 = e_1$$
$$X_2 = X_1 + e_2 = e_1 + e_2$$

.

.

.

$$X_t = X_{t-1} + e_t = e_1 + e_2 + ... + e_t$$

$$X_t = \sum_{i=1}^{t} e_i \qquad \text{.........(1.5.2)}$$

$E(X_t)=t\mu$ and $Var(X_t)=t\sigma^2$, hence as the mean and variance changes with time t, the process is non-stationary.

However, it is interesting to note that the first difference of a random walk, given by

$\Delta X_t = X_t - X_{t-1} = e_t$ for a purely random process, which is therefore stationary.

(iii) Moving Average process

Suppose that $\{e_t\}$ is a white noise with mean zero and variance σ^2. Then the process $\{X_t\}$ is said to be a moving average process of order q (MA(q)) if

$$X_t = \beta_0 e_t + \beta_1 e_{t-1} + \ldots + \beta_q e_{t-q} \qquad (1.5.3)$$

which is written as

$$X_t = \sum_{j=0}^{q} \beta_j e_{t-j} \quad \text{where } \{\beta_j\} \text{ are constants therefore } \mathrm{E}(X_t) = 0$$

$$Var(X_t) = \sigma^2 \sum_{j=0}^{q} \beta^2{}_j \qquad (1.5.4)$$

since, the e`s are independent.

We have

$$\sigma(h) = \mathrm{cov}(X_t, X_{t+h}) = \mathrm{E}(X_t, X_{t+h})$$

$$= \mathrm{E}\left[\sum_{j=0}^{q} \beta_j e_{t-j}, \sum_{k=0}^{q} \beta_k e_{t+h-k}\right]$$

$$= \sum_{j=0}^{q}\sum_{k=0}^{q} \beta_j \beta_k \mathrm{E}\left[e_{t-j} e_{t-(k-h)}\right]$$

when $j = k\text{-}h$ i.e. $k = j+h$ then

$$\mathrm{E}(e_{t-j} e_{t-(k-h)}) = \begin{cases} \sigma^2, \ldots\ldots & j = k-h \\ 0, \ldots\ldots & j \neq k-h \end{cases}$$

so that

$$\sigma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \beta_j \beta_{j+h}, \ldots\ldots & 0 \le h \le q \\ 0, \ldots\ldots & h > q \\ \sigma(-h), \ldots\ldots & h < 0 \end{cases}$$

$$\text{since } Cov(e_s, e_t) = \begin{cases} \sigma^2, \ldots\ldots & s = t \\ 0, \ldots\ldots & s \neq t. \end{cases}$$

As $\sigma(h)$ does not depend on t, the mean is constant, the process is second order stationary for all values of the $\{\beta_t\}$. Furthermore, if the e`s are normally distributed, then so are the X`s and we have a strictly stationary normal process.

The ACF of the MA(q) process is given by

$$\rho(h)=\begin{cases}1, & h=0\\ \dfrac{\sum_{j=0}^{q-1}\beta_j\beta_{j+k}}{\sum_{j=1}^{q}\beta_j^2}, & h=1,2,\dots,q \\ \rho(-h), & h<0\end{cases} \quad \text{.........(1.5.6).}$$

Note that the ACF 'cuts off' at lag q, which is a special feature of MA process.

(iv) Autoregressive Processes

Let $\{e_t\}$ be a purely random process with mean zero and variance σ^2. Then a process $\{X_t\}$ is said to be an autoregressive process of order p (AR(p)) if

$$X_t=\alpha_1X_{t-1}+\alpha_2X_{t-2}+\dots+\alpha_pX_{t-p}+e_t \quad \text{.........(1.5.7)}$$

(a) First order process AR(1)

$X_t=\alpha_1X_{t-1}+e_t, |\alpha|<1$, successively replacing the X_t`s we obtain

$$\begin{aligned}X_t &= e_t+\alpha X_{t-1}\\ &= e_t+\alpha(\alpha X_{t-2}+e_{t-1})=e_t+\alpha e_{t-1}+\alpha^2X_{t-2}\\ &= e_t+\alpha e_{t-1}+\alpha^2e_{t-2}+\alpha^3X_{t-3}\\ &.\\ &.\\ &.\\ &= e_t+\alpha e_{t-1}+\alpha^2e_{t-2}+\dots+\alpha^se_{t-s}+\alpha^{s+1}X_{t-(s+1)} \quad \text{.........(1.5.8).}\end{aligned}$$

Hence, we have

$$\mathrm{E}\left(X_t-\sum_{j=0}^{s}\alpha^je_{t-j}\right)^2=\alpha^{2s+2}\mathrm{E}\left(X^2_{t-(s+1)}\right) \quad \text{.........(1.5.9).}$$

Note that, $\mathrm{E}(X_t)=\alpha\mathrm{E}(X_{t-1})+\mathrm{E}(e_t)$, but $\mathrm{E}(e_t)=0$

$\mathrm{E}(X_t)-\alpha\mathrm{E}(X_{t-1})=0$.

Assuming the process is stationary $E(X_t) = E(X_{t-1})$ then, $(1-\alpha)E(X_t) = 0$. But $\alpha \neq 1,_{....}$ *hence,*$_{.....}$ $E(X_t) = 0$.

Therefore,

$$E\left(X_t - \sum_{j=0}^{s} \alpha^j e_{t-j} \right)^2 = \alpha^{2s+2} Var\left(X_{t-(s+1)}\right)$$
$$= \alpha^{2s+2} Var(X_t) \qquad (1.5.10)$$

since, X_t is stationary $Var(X_t)$ is a constant and because $|\alpha| < 1$ RHS of (1.5.10) tends to zero as $s \to \infty$. This implies that when $|\alpha| < 1,_{...} X_t$ converges in probability to $\sum_{j=0}^{\infty} \alpha^j e_{t-j}$.

Hence ,

$$X_t = \sum_{j=0}^{\infty} \alpha^j e_{t-j} \qquad (1.5.11).$$

From this $E(X_t) = 0$ and $Var(X_t) = \sum_{j=0}^{\infty} \alpha^{2j} Var(e_{t-j}) = \sigma^2 \sum_{j=0}^{\infty} \alpha^{2j}$.

Hence, $Var(X_t) = \dfrac{\sigma^2}{1-\alpha^2} = \sigma(0) \qquad (1.5.12).$

Also ,

$$\sigma(h) = E(X_t, X_{t+h})$$
$$= E\left(\sum_{j=0}^{\infty} \alpha^j e_{t-j} \sum_{k=0}^{\infty} \alpha^k e_{t+h-k} \right)$$
$$= \sigma^2 \alpha^h \sum_{j=0}^{\infty} \alpha^{2j}$$

which gives

$$\sigma(h) = \frac{\sigma^2 \alpha^h}{1-\alpha^2} \qquad (1.5.13).$$

Therefore the ACF is given by dividing (1.5.13) by (1.5.12) which gives

$$\rho(h) = \alpha^h \qquad (1.5.14).$$

(b) The general order AR(p) process

The *pth* order AR(p) process is written as

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + e_t \qquad (1.5.15).$$

Using backward shift operator equation (1.5.15) gives

$$\left(1-\alpha_1 B-\alpha_2 B^2-\ldots-\alpha_p B^p\right)X_t=e_t \qquad (1.5.16).$$

Hence ,

$$\sigma(h)=\alpha_1\sigma(h-1)+\alpha_2\sigma(h-2)+\ldots+\alpha_p\sigma(h-p), \quad h>0 \qquad (1.5.17)$$

where $E(e_t, X_{t-h})=0, \quad h>0$. Hence we have the following recursive relationship for ACF.

$$\rho(h)=\alpha_1\rho(h-1)+\alpha_2\rho(h-2)+\ldots+\alpha_p\rho(h-p), \quad h>0 \qquad (1.5.18).$$

From (1.5.18) we see that the ACF $\rho(h)$ is determine by the difference equation

$\theta_p(B)=\left(1-\alpha_1 B-\alpha_2 B^2-\ldots-\alpha_p B^p\right)\rho(h)=0$ or using the Yule-Walker equation which gives the general form of obtaining $\rho(h)$ as

$$\rho(h)=A_1\pi_1^h+A_2\pi_2^h+\ldots+A_p\pi_p^h \qquad (1.5.19)$$

where $\pi_1,\pi_2,\ldots,\pi_p$ are the roots of the auxiliary equation. The constants $A_1, A_2,\ldots, A_p$ can be solved using the initial conditions.

1.6: Autoregressive moving average ARMA process.

A stationary and invertible processes can be represented either in a MA form or in an AR form. However, a problem with either representation is that it may contain too many parameters.

Mixed AR and MA process leads to the following useful ARMA process.

$$\theta_p(B)X_t=\phi_q(B)e_t \qquad (1.6.1)$$

where

$$\theta_p(B)=1-\alpha_1 B-\alpha_2 B^2\ldots-\alpha_p B^p$$

and $\phi_q(B)=1-\beta_1 B-\beta_2 B^2-\ldots-\beta_q B^q$.

For this process to be invertible, we require that the roots of $\phi_q(B)=0$ lie outside the unit circle. To be stationary we require that the roots of

$\theta_p(B)=0$ lie outside the unit circle. Also, we assume that $\theta_p(B)=0$ and $\phi_q(B)=0$ share no common roots. We have in general

$$\pi(B)X_t = e_t \quad \text{.........(1.6.2)}$$

where

$$\pi(B)=\frac{\theta_p(B)}{\phi_q(B)}=\left[1-\pi_1 B-\pi_2 B^2-........\right]$$

This process can also be written as a purely moving average representations.

$$X_t = \Psi(B)e_t \quad \text{.........(1.6.3)}$$

where ,

$$\Psi(B)=\frac{\phi_q(B)}{\theta_p(B)}=\left[1+\Psi_1 B+\Psi_2 B^2+........\right].$$

To derive the ACF of ARMA(p,q) we write (1.6.3) as

$$X_t=\alpha_1 X_{t-1}+\alpha_2 X_{t-2}+...+\alpha_p X_{t-p}+e_t-\beta_1 e_{t-1}-\beta_2 e_{t-2}-...-\beta_q e_{t-q} \quad \text{........(1.6.4)}$$

multiplying this by X_{t-h} on both sides of equation (1.6.4) and taking expectation we obtain

$$\sigma(h)=\alpha_1\sigma(h-1)+\alpha_2\sigma(h-2)+...+\alpha_p\sigma(h-p)+\mathrm{E}(X_{t-h}e_t)-$$
$$\beta_1\mathrm{E}(X_{t-h}e_{t-1})-...-\beta_q\mathrm{E}(X_{t-h}e_{t-q})$$

because , $\mathrm{E}(X_{t-h}e_{t-i})=0,\ldots\ h>i$.

We have

$$\sigma(h)=\alpha_1\sigma(h-1)+...+\alpha_p\sigma(h-p),\ldots\ h\ge(q+1) \quad \text{.........(1.6.5)}$$

and hence, the ACF is

$$\rho(h)=\alpha_1\rho(h-1)+\alpha_2\rho(h-2)+...+\alpha_p\rho(h-p),\ldots\ h\ge(q+1) \quad \text{.........(1.6.6)}.$$

Equation (1.6.6) satisfy the *pth* order homogeneous difference equation, as in the AR(p) process. Therefore the ACF of an ARMA(p,q) model tails off after lag q just like AR(p) process. However, the first q AR, $\rho(q),\rho(q-1),...,\rho(1)$ depends on both AR and MA parameters in the model and serve as initial values for the pattern. This distinction is useful in model identification.

1.7: Integrated ARIMA models.

In practice most time series are non-stationary and in order to fit a stationary model it is necessary to remove non-stationarity source of variation. If the observed series is non stationary in the mean (trend) then we can difference the series to make it stationary. The ARMA(p,q) model is given by equation (1.6.4), hence X_t in the equation (1.6.4) is replaced by $\Delta^d X_t = w_t$ then we have a model capable of describing a certain type of non-stationary series. Such a model is called an integrated ARIMA model Chatfield (1989). We write

$$w_t = \Delta^d X_t = (1-B)^d X_t \qquad (1.7.1)$$

Hence, the general autoregressive moving average process (ARIMA) is of the form

$$w_t = \alpha_1 w_{t-1} + \alpha_2 w_{t-2} + \ldots + \alpha_p w_{t-p} + e_t - \beta_1 e_{t-1} - \beta_2 e_{t-2} - \ldots - \beta_q e_{t-q} \qquad (1.7.2)$$

Hence ,

$$\phi(B) w_t = \theta(B) e_t \qquad (1.7.3)$$

$$\phi(B(1-B))^d X_t = \theta(B) e_t \qquad (1.7.4).$$

Thus we have an ARMA(p,q) model for w_t. The model (1.7.4) describe the *dth* difference of X_t is called an ARIMA process of order (p,d,q) i.e. ARIMA(p,d,q). the model for X_t is clearly non-stationary as the AR operator $\phi(B)(1-B)^d$ has d roots on the unit circle.

Example 1.7 A random walk is given by

$$X_t = X_{t-1} + e_t$$
$$X_t - X_{t-1} = e_t$$
$$(1-B)X_t = e_t.$$

Hence , $w_t = e_t$ this is an ARIMA(0,1,0) process.

1.8: Estimation of parameters.

We need to estimate the parameters of the models we have mention in the previous sections.

1.8.1: Estimating the parameters of an AR process

Consider an AR process of order, p with mean μ given by

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + \ldots + \alpha_p(X_{t-p} - \mu) + e_t \quad \text{.............(1.8.1)}$$

Given N observations $x_1, x_2, \ldots, x_N$, the parameters $\mu, \alpha_1, \alpha_2, \ldots, \alpha_p$ may be estimated by the least square method by minimizing

$$S = \sum_{t=p+1}^{N} \left[x_t - \mu - \alpha_1(x_{t-1} - \mu) - \ldots - \alpha_p(x_{t-p} - \mu)\right]^2$$

with respect to $\mu, \alpha_1, \alpha_2, \ldots, \alpha_p$. If the e_t process is normal, then the least square estimates are in addition the maximum likelihood estimates conditional on the first p values in the series being fixed.

Consider the AR(1) process given by

$$X_t - \mu = \alpha(X_{t-1} - \mu) + e_t \quad \text{.............(1.8.2)}$$

for $t = 2,3,4,\ldots$

Hence,

$$S = \sum\left[x_t - \mu - \alpha(x_{t-1} - \mu)\right]^2$$

$$\frac{\partial S}{\partial \mu} = -\sum_{t=2}^{N}(1-\alpha)\{x_t - \mu - \alpha(x_{t-1} - \mu)\} = 0 \quad \text{.............(1.8.3).}$$

This leads to

$$\hat{\mu} = \frac{\bar{x}_{(2)} - \hat{\alpha}\bar{x}_{(1)}}{1 - \hat{\alpha}} \quad \text{.............(1.8.4)}$$

where $\bar{x}_{(1)}$ and $\bar{x}_{(2)}$ are the means of the first and the last (N-1) observations (see Chatfield (1989: sect. 4.2)).

Also

$$\frac{\partial S}{\partial \alpha} = \sum -\left(x_t - \mu - \alpha(x_{t-1} - \mu)\right)(x_{t-1} - \mu) = 0 \quad \text{.............(1.8.5).}$$

Hence , (1.8.5) leads to

$$\hat{\alpha} = \frac{\sum_{t=2}^{N}(x_t - \hat{\mu})(x_{t-1} - \hat{\mu})}{\sum_{t=2}^{N}(x_{t-1} - \hat{\mu})^2} \text{..........} (1.8.6).$$

Let $\bar{x}_{(1)} \approx \bar{x}_{(2)} \approx \bar{x}$, we have

$$\hat{\mu} = \bar{x} \text{..........} (1.8.7)$$

substituting this value in equation (1.8.6) with further approximation on the denominator we obtain

$$\hat{\alpha} = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2} = \frac{\sigma(1)}{\sigma(0)} = r_1 = \rho(1) \text{..........} (1.8.8).$$

This approximate estimation of α is also intuitively appealing since r_1 is an estimate of $\rho(1)$ and $\rho(1) = \alpha$ for an AR(1) process. Higher order AR processes may be fitted by least square method in the same way.

1.8.2: Estimating parameters of a MA process.

This is more difficult than in the case of an AR process because efficient explicit estimators cannot be found. Instead some form of numerical iteration must be performed.

Consider a MA(1) process

$$X_t = \mu + e_t + \beta e_{t-1} \text{..........} (1.8.9).$$

We would like to write the residual sum of squares $\sum e_t^2$ solely in terms of observed X_t and parameters μ and β as we did for the AR process so that we can differentiate with respect to μ and β and hence find the least square estimates by equating the derivatives to zero.

Unfortunately the residual sum of squares is not a quadratic function of the parameters and so explicit least squares estimate cannot be found.

We can equate sample and theoretical first order autocorrelation coefficients by

$$r_1 = \frac{\hat{\beta}}{1-\hat{\beta}^2} \quad \text{...} (1.8.10)$$

and choose the solution such that $|\hat{\beta}| < 1$. But this gives rise to inefficient estimators.

An appropriate suggestion given by Box and Jenkins (1970, chapter 7) is to select a suitable starting values for μ and β such that $\hat{\mu} = \bar{x}$ and $\hat{\beta}$ given by the solution of (1.8.10) recursively in the form

$$e_t = X_t - \mu - \beta e_{t-1}, \ldots \text{define}, \ldots e_0 = 0$$

$$\left.\begin{aligned} e_1 &= X_1 - \hat{\mu} \\ e_2 &= X_2 - \hat{\mu} - \hat{\beta} e_1 \\ &. \\ &. \\ &. \\ e_N &= X_N - \hat{\mu} - \hat{\beta} e_{N-1} \end{aligned}\right\} \text{...} (1.8.11)$$

Then $\sum_{t=1}^{N} e_t^2$ may be calculated. This procedure is then repeated for other values of μ, β and the sum of squares $\sum e_t^2$ computed for a grid of points in the (μ, β) plane. We may then determine by inspection, the least square estimates of μ and β that minimizes $\sum e_t^2$. A similar type of iterative procedure may be used for higher order processes.

1.8.3: Estimating the parameters of an ARMA model.

Suppose now that a mixed autoregressive/ moving average (ARMA) model is found to be appropriate for a given time series, then the estimation problems for an ARMA model are similar to those for MA mode in that an iterative procedure must be used to estimate the parameters. The residual sum of squares can be calculated at every point on a suitable grid of parameter values which gives the minimum sum of squares may the be assessed. Alternatively some sort of optimization procedure may be used.

As an example, consider the ARMA(1,1) process whose ACF decrease exponentially after lag 1. This model may be recognized as appropriate if the sample ACF has a similar form. The models given by

$$X_t - \mu = \alpha(X_{t-1} - \mu) + e_t - \beta e_{t-1} \quad \text{..........(1.8.12)}$$

Given N observations $x_1, x_2, \ldots, x_N$, we guess values for μ, α, β, set $e_0 = 0$ and $x_0 = \mu$, and calculate the residuals recursively by

$$\left.\begin{aligned} e_1 &= x_1 - \mu \\ e_2 &= x_2 - \mu - \alpha(x_1 - \mu) + \beta e_1 \\ &. \\ &. \\ &. \\ e_N &= x_N - \mu - \alpha(x_{N-1} - \mu) + \beta e_{N-1} \end{aligned}\right\} \quad \text{..........(1.8.13)}$$

The residual sum of squares $\sum_{t=1}^{N} e_t^2$ may be calculated. Then other values of μ, α and β may be tried until the minimum residual sum of squares is found. Details of this may be found in Box and Jenkin (1970) and Priestley (1981, chapter 5).

1.8.4: Estimating the parameters of an ARIMA model.

In practice most time series are non-stationary, and the stationary models are not applicable. We can difference an observed time series until it is stationary. Once the stationarity is achieved the estimation strategy adapted for the ARMA process is employed to estimate the parameters.

CHAPTER TWO

STATE SPACE MODELS AND THE KALMAN FILTER

2.0: Introduction

A general class of models, which has generated a lot of interest, are the state space models. They were originally developed by control engineers, particularly for applications in navigation systems such as controlling the position of a space rocket. However , they have also been found to be useful in other type of time series problems such as short term forecasting.

This chapter introduces state space models for the time series analysts as well as describing the Kalman filter, which is an important general method of handling state space models.

The state of a process is the minimum set of information from the past and present necessary to predict the future. Thus the state space representation is based on the markov property which implies that given the present state, the future of a system is independent of its past. Consequently, the state space representation of a system is also called the markovian representation of a system.

For a linear time-invariant system, its state space representation is described by the state equation

$$x(t+1) = Fx(t) + G\varepsilon_{t+1} \quad \text{.......} (2.0.1)$$

and output equation

$$X_t = Hx(t) \quad \text{.......} (2.0.2)$$

where $x(t)$ is a state vector of dimension k, F is a kxk transition matrix, G is a kxn input matrix, ε_{t+1} is an nx1 vector of the input, to the system, X_t is an mx1 vector of the output, and H is an mxk output or observation matrix. If both the ε_{t+1} and X_t are stochastic processes, then the state space representation is given by

$$x(t+1) = Fx(t) + Gu_{t+1} \qquad (2.0.3)$$
$$X_t = Hx(t) + v_t \qquad (2.0.4)$$

where $u_{t+1} = \varepsilon_{t+1} - \mathrm{E}(\varepsilon_{t+1} \mid \varepsilon_t, t \le n)$ is the nx1 vector of one- step-ahead forecast error of the process ε_t and v_t is an mxs vector of disturbances assume to be independent of u_t. The vector u_{t+1} is also known as the innovation of the input ε_t at time $(t+1)$. When $X_t = \varepsilon_t$, v_t vanishes from (2.0.4) and the state space representation of a stationary stochastic process ε_t becomes

$$\left.\begin{array}{l} x(t+1) = Fx(t) + G\varepsilon_t \\ X_t = Hx(t) \end{array}\right\} \qquad (2.0.5)$$

Thus, the process X_t is the output of the time invariant linear stochastic system driven by a white noise input ε_t. The $x(t)$ is known as the state of the process.

The state equation and the output equation are also referred to as the measurement equation and the observation equation respectively. The state space representation of a system is relate to the Kalman filter and was originally developed by control engineers, (See Kalman (1960)).

2.1: state space representation.

Priestley (1988) has given an alternative way of representing the general ARMA process in the form

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = e_t + \beta_1 e_{t-1} + \ldots + \beta_q e_{t-q} \qquad (2.1.1)$$

based on the so- called 'state- space' (or 'markovian') representation which provides a compact description of any finite- parameter linear model. The basic idea rests simply on the well-known result that any finite-order linear differential or difference equation can be expressed as a vector first-order equation. For example, if we take the AR(2) model.

$$e_t = X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} \qquad (2.1.2)$$

we have, $x_t^{(2)} = X_t, x_t^{(1)} = -\alpha_2 X_{t-1} \left[= -\alpha_2 x_{t-1}^{(2)}\right]$, then (2.1.2) may be written as

$$X_t = -\alpha_1 X_{t-1} - \alpha_2 X_{t-2} + e_t \qquad (2.1.3)$$

and write in vector form as

$$\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & -\alpha_2 \\ 1 & -\alpha_1 \end{bmatrix} \begin{bmatrix} x_{t-1}^{(1)} \\ x_{t-1}^{(2)} \end{bmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} e_t \quad \text{.........(2.1.4)}$$

where, $X_t = [0 \quad 1] \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix}$.

In this case

$$x_t^{(1)} = -\alpha x_{t-1}^{(2)}$$
$$x_t^{(2)} = x_{t-1}^{(1)} - \alpha_1 x_{t-1}^{(2)} + e_t$$

But ,

$$x_t^{(1)} = -\alpha_2 x_{t-1}^{(2)}$$
$$x_{t-1}^{(1)} = -\alpha_2 x_{t-2}^{(2)}.$$

Hence ,

$$x_t^{(2)} = -\alpha_2 x_{t-2}^{(2)} - \alpha_1 x_{t-1}^{(2)} + e_t \quad \text{.........(2.1.5)}$$

and since $X_t = x_t^{(2)}, X_{t-1} = x_{t-1}^{(2)}$ and $X_{t-2} = x_{t-2}^{(2)}$ thus (2.1.5) becomes

$X_t = -\alpha_1 X_{t-1} - \alpha_2 X_{t-2} + e_t$ which is equation of the form (2.1.3).

AR(3) process can be written as

$$X_t = -\alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \alpha_3 X_{t-3} + e_t \quad \text{.........(2.1.6)},$$

which can be written in state space form as

$$\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\alpha_3 \\ 1 & 0 & -\alpha_2 \\ 0 & 1 & -\alpha_1 \end{bmatrix} \begin{bmatrix} x_{t-1}^{(1)} \\ x_{t-1}^{(2)} \\ x_{t-1}^{(3)} \end{bmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} e_t \quad \text{.........(2.1.7)}$$

where,

$$X_t = [0 \quad 0 \quad 1] \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \end{bmatrix}.$$

Note, $x_t^{(3)} = X_t$ and

$$x_t^{(1)} = -\alpha_3 x_{t-1}^{(3)} \quad \text{.........(2.1.8)}$$
$$x_t^{(2)} = x_{t-1}^{(1)} - \alpha_2 x_{t-1}^{(3)} \quad \text{.........(2.1.9)}$$
$$x_t^{(3)} = x_{t-2}^{(2)} - \alpha_1 x_{t-1}^{(3)} + e_t \quad \text{.........(2.1.10)},$$

where

and
$$x_{t-1}^{(2)} = x_{t-2}^{(1)} - \alpha_2 x_{t-2}^{(3)} \qquad (2.1.11)$$
$$x_{t-2}^{(1)} = -\alpha_3 x_{t-3}^{(3)}$$

from (2.1.9) and (2.1.8) respectively, hence (2.1.10) becomes,

$$x_t^{(3)} = -\alpha_3 x_{t-3}^{(3)} - \alpha_2 x_{t-2}^{(3)} - \alpha_1 x_{t-1}^{(3)} + e_t \qquad (2.1.12).$$

By noting that, $x_t^{(3)} = X_t, x_{t-1}^{(3)} = X_{t-1}, x_{t-2}^{(3)} = X_{t-2}$ and $x_{t-3}^{(3)} = X_{t-3}$, equation (2.1.12) reduces to (2.1.6).

By the same argument the general AR(p) process.

$$X_t = -\alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \ldots - \alpha_p X_{t-p} + e_t \qquad (2.1.13)$$

which can be written in state space form as

$$\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \\ . \\ . \\ . \\ x_t^{(p)} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & . & . & . & 0 & -\alpha_p \\ 1 & 0 & 0 & . & . & . & 0 & -\alpha_{p-1} \\ 0 & 1 & 0 & . & . & . & 0 & -\alpha_{p-2} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & 1 & -\alpha_1 \end{bmatrix} \begin{pmatrix} x_{t-1}^{(1)} \\ x_{t-1}^{(2)} \\ x_{t-1}^{(3)} \\ . \\ . \\ . \\ x_{t-1}^{(p)} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 1 \end{pmatrix} e_t \qquad (2.1.14).$$

To recover X_t from the vector $\left(x_t^{(1)} \quad x_t^{(2)} \quad . \; . \; . \quad x_t^{(p)}\right)$, we write

$$X_t = (0 \quad 0 \quad . \; . \; . \quad 1) \begin{pmatrix} x_t^{(1)} \\ x_t^{(2)} \\ . \\ . \\ . \\ x_t^{(p)} \end{pmatrix} \qquad (2.1.15)$$

The pair (2.1.14) and (2.1.15) is completely equivalent to (2.1.13) but whereas (2.1.13) involved p-stage dependence so that X_t is non-markovian (2.1.14) involve only a one-stage dependence so that $\left(x_t^{(1)} \quad x_t^{(2)} \quad . \; . \; . \quad x_t^{(p)}\right)$ is a vector markov process.

Exactly the same approach may be use to write the general ARMA model in the state space form. Thus, assuming that $q \le (p-1)$ in ARMA (p,q)

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = e_t + \beta_1 e_{t-1} + \ldots + \beta_q e_{t-q} \qquad (2.1.16).$$

We first write (2.1.16) as

$$X_t + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} = e_t + \beta_1 e_{t-1} + \ldots + \beta_{p-1} e_{t-p+1} \qquad (2.1.17)$$

where $\beta_j = 0$ $(j > q)$ and it is then easy to verify that (2.1.17) can be written in state space form as

$$x(t+1) = Fx(t) + Ge_t \qquad (2.1.18)$$
$$X_t = Hx(t) \qquad (2.1.19).$$

The *pxp* matrix F is given by

$$F = \begin{bmatrix} 0 & 0 & . & . & . & 0 & -\alpha_p \\ 1 & 0 & . & . & . & 0 & -\alpha_{p-1} \\ 0 & 1 & . & . & . & 0 & -\alpha_{p-2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 & -\alpha_1 \end{bmatrix}$$

the px1 matrix G by

$G' = \{\beta_{p-1} \quad \beta_{p-2} \quad . \quad . \quad . \quad \beta_1 \quad 1\}$ then 1xp matrix H by

$H' = (0 \quad 0 \quad . \quad . \quad . \quad 1)$ e_t is px1 matrix and the px1 vector $x(t)$ is $x'(t) = (x_1(t) \quad x_2(t) \quad . \quad . \quad . \quad x_p(t))$. (In fact, $x_p(t)$ is simply the variable X_t in (2.1.7) and $x_{p-1}(t), x_{p-2}(t), \ldots\ldots$ are defined recursively in terms of $x_p(t)$). In this formulation $x(t)$ is called the state vector, F is called the system matrix, G the input matrix and H is the observation matrix.

2.2: The Kalman Filter.

If the state space model is given by

$$X_t = h_t^T \theta_t + n_t \qquad (2.2.1)$$
$$\theta_t = G_t \theta_{t-1} + w_t \qquad (2.2.2)$$

where equation (2.2.1) is the observation equation and (2.2.2) is the transition equation. The 'errors' in the observation and transition equations are generally assumed to be uncorrelated with each other at all time periods. We may further assumed that n_t is $N(0, \sigma_n^2)$ while

w_t is multivariate normal with zero mean vector and known variance-covariance matrix denoted by Σ_t.

In the state space modeling, the prime objective is to estimate the signal in the presence of noise. In other words we want to estimate the state vector θ_t. The Kalman Filter provides a general methods of doing this. It consist of a set of equations which allow as to update the estimate of θ_t when anew observation becomes available. This updating procedure has two stages, called the prediction stage and the update stage.

Suppose we have observed a time series up to time t-1, and that $\hat{\theta}_{t-1}$ is the 'best' estimator for θ_{t-1} based on information up to this time. By 'best' we mean that it is the minimum mean square error estimation. Further, suppose that we have evaluated the variance-covariance matrix of $\hat{\theta}_{t-1}$ which we denote by p_{t-1}. The first stage, called the **prediction** stage, is concerned with forecasting θ_t from time t-1, and we denote the resulting estimation in an obvious notation by $\hat{\theta}_{t|t-1}$. Considering the updating equation (2.2.2) where w_t is still unknown at time t-1, the obvious estimator for θ_t is given by

$$\hat{\theta}_{t|t-1} = G_t\hat{\theta}_{t-1} \qquad (2.2.3)$$

with variance-covariance matrix

$$p_{t|t-1} = G_t p_{t-1} G_t^T + \Sigma_t \qquad (2.2.4)$$

equation (2.2.3) and (2.2.4) are the prediction equations. Equation (2.2.4) follows from standard results on variance-covariance matrices for vector variable Chatfield and Collins (1980).

When the new observation at time t, X_t becomes available, the estimator of θ_t can be modified to take account of this extra information. The prediction error is given by

$$e_t = X_t - h_t^T\hat{\theta}_{t|t-1} \qquad (2.2.5)$$

and it can be shown that the updating equations are given by

$$\hat{\theta}_t = \hat{\theta}_{t|t-1} + k_t e_t \quad \text{.......(2.2.6)}$$

and $p_t = p_{t|t-1} - k_t h_t^T p_{t|t-1}$ ....(2.2.7)

where

$$k_t = \frac{p_{t|t-1} h_t}{\left[h_t^T p_{t|t-1} h_t + \sigma_n^2\right]} \quad \text{.......(2.2.8)}$$

is called the Kalman gain matrix, which in the univariate case is just a vector. Equation (2.2.4) and (2.2.5) constitute the second stage of the Kalman Filter and are called the updating equations.

A major practical advantage of the Kalman Filter is that;

1. The calculations are recursive, so that although the current estimates are based on the whole past History of measurements, there is no need for an ever-expanding memory. Recursive methods, such as exponential smoothing, are increasingly becoming popular in many areas of statistics.
2. Kalman Filter converges fairly quickly, but can also follow the movement of a system where the underlying model is evolving through time.

The Kalman Filter is applied to state space models which are linear in the parameters. In practice many time series models, such as multiplicative seasonal models, are non-linear. Then it may be possible to apply a Filter, called the **extended Kalman Filter** , by making a locally linear approximation to the model. Application to data where the noise is not necessarily normally distributed are also possible Kitagawa (1987) and Nassiuma (1994)

2.3: Recursive estimation of state and likelihood computation for state space models.

Let a state space model be given by

$$\left.\begin{aligned} x(n) &= F(n)X(n-1) + G(n)v(n) \\ y(n) &= H(n)X(n) + w(n) \end{aligned}\right\} \quad \text{.......(2.3.1)}$$

where $w(n) \sim N(0, R(n))$ and $v(n) \sim N(0, Q(n))$. Given the observations y(1), y(2),...,y(N) and the initial conditions $X(0 \mid 0), v(0 \mid 0)$, the one step ahead predictor and the Filter are obtained from the Kalman Filter algorithm; (Kitagawa (1983)).

Time update (prediction)

$$\left.\begin{array}{l} x(n \mid n-1) = F(n)X(n-1 \mid n-1) \\ v(n \mid n-1) = F(n)v(n-1 \mid n-1)F(n)^t + G(n)Q(n)G(n)^t \end{array}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.3.2)$$

Observation update (Filtering)

$$\left.\begin{array}{l} k(n) = v(n \mid n-1)H(n)^t\left[H(n)v(n \mid n-1)H(n)^t + R(n)\right]^{-1} \\ X(n \mid n) = X(n \mid n-1) + k(n)[y(n) - H(n)X(n \mid n-1)] \\ v(n \mid n) = (I - k(n)H(n))v(n \mid n-1). \end{array}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.3.3)$$

Using this estimates, the smoothed value of the state $X(n)$ given the entire observations, y(1), y(2),...,y(N), is obtained by the fixed interval smoothing algorithm Kitagawa (1983).

$$\left.\begin{array}{l} A(n) = v(n \mid n)F(n)^t v(n+1 \mid n)^{-1} \\ x(n \mid N) = X(n \mid n) + A(n)[X(n+1 \mid N) - X(n+1 \mid n)] \\ v(n \mid N) = v(n \mid n) + A(n)v(n+1 \mid N) - v(n+1 \mid n)A(n)^t \end{array}\right\} \ldots\ldots\ldots\ldots\ldots(2.3.4)$$

The state space representation and the Kalman Filter yield an efficient algorithm for the likelihood of a time series model. Since

$$f(y(1), y(2), \ldots, y(N)) = \prod_{n=1}^{N} f(y(n) \mid y(1), y(2), \ldots, y(n-1)) \ldots\ldots\ldots\ldots\ldots\ldots(2.3.5)$$

with

$$\begin{array}{l} f(y(n) \mid y(1), y(2), \ldots, y(n-1)) = \\ (2\pi v(n))^{-1/2} \exp\left\{-(y(n) - H(n)X(n \mid n-1))^2 \mid 2v(n)\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.3.6) \\ v(n) = H(n)v(n \mid n-1)H(n)^t + R(n), \end{array}$$

The log likelihood of the model is obtained by

$$\ell = -\frac{1}{2}\left\{N \log 2\pi + \sum_{n=1}^{N} \log v(n) + \sum_{n=1}^{N}(y(n) - H(n)X(n \mid n-1))^2 \mid 2v(n)\right\} \ldots(2.3.7)$$

The maximum likelihood estimate of the parameters are obtained by maximizing (2.3.7) with respect to those parameters. The Akaike Information Criterion (AIC) is then defined by

AIC=-2(maximum log likelihood) +

2 (number of parameters) ....(2.3.8)

According to the minimum AIC procedure, the model with the smallest value of AIC statistics is selected as the best model.

We have an interpretation of the log likelihood obtained by estimating the initial state vector as a natural estimate of the expected log likelihood of the predictive distribution Kitagawa (1983).

2.4: State space model for Time Series with non-stationarity and irregularity.

2.4.1: Model for stationary Time Series.

We first consider an autoregressive (AR) model,

$$Y(n)=\sum_{i=1}^{M}a(i)\,Y(n-i)+v(n)\qquad(2.4.1)$$

This model has a state space representation of the form (2.3.1) where the constant matrices F, G and H are defined by

$$F=F(a,m)=\begin{bmatrix}0 & 0 & . & . & . & a(m)\\ 1 & 0 & . & . & . & a(m-1)\\ 0 & 1 & . & . & . & a(m-2)\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ 0 & 0 & . & . & 1 & a(1)\end{bmatrix},G=G_m=\begin{bmatrix}0\\0\\.\\.\\.\\1\end{bmatrix}\qquad(2.4.2)$$

$H=H_m=(0\ \ 0\ \ .\ \ .\ \ .\ \ 1)$ and the state vector $x(n)$ is defined by

$$x(n)=[Y(n),Y(n-1),\ldots,Y(n-m+1)]'\qquad(2.4.3)$$

2.4.2:Model for non stationarity in the mean.

We consider a nonstationary time series in the mean which is of the form,

$$Y(n)=f(n)+z(n)+w(n)\qquad(2.4.4)$$

where $f(n)$ is a smooth mean value function z(n) is an AR process of order M, w(n) is an observational noise. The state space representation for z(n) is already given in (2.4.2). Thus, by defining an argumented state vector $x(n)$ and matrices F,G and H by

$$X(n)=\begin{bmatrix} x_1(n) \\ \ldots\ldots \\ x_2(n) \end{bmatrix}, F=\begin{bmatrix} F(a_m) & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & F_k \end{bmatrix} \quad \ldots\ldots(2.4.5)$$

$$H=\begin{bmatrix} H_m & \vdots & H_k \end{bmatrix}$$

We obtained a state space representation in actual modeling, the mean value function may be expressed as the sum of several components. In that case the state space model becomes more complicated. Models for trend estimation, seasonal adjustment and trading day adjustment are given in Kitagawa (1981).

2.5: Missing and Outlying observation in Time Series.

We are going to deal with estimation of consecutively missing observation in the next chapter, but, in this section we need to show how Kalman Filter allows the state space model to be time varying which will enable us to handle outlying observation in Time series.

The Kalman Filter given in (2.3.2) and (2.3.3) is composed of two steps. If an observation is missed, we have only to skip the observation update step. For details see Jones (1980).

It is well known that there are two types of outliers in Time series. The first type of outliers affects only one observation (additive outliers); the second type has long influence in its future value (innovation type). The state space representation (2.3.1) is well suited to treat these outliers. In this model an additive outlier is included in the observational noise sequence w(n), and the innovation outlier is included in v(n) and thus has a long effect through the system dynamics. Now our model for outlier is (for scalar case) (2.3.1) with

$$R(n) = \begin{cases} \sigma^2(n), & n = s_1, s_2, \ldots, s_l \\ \sigma^2, & otherwise \end{cases}$$

$$Q(n) = \begin{cases} \tau^2(n), & n = t_1, t_2, \ldots, t_m \\ \tau^2, & otherwise \end{cases}$$

where $\sigma^2(n) >> \sigma^2$ and $\tau^2(n) >> \tau^2$. $v(n), \; n = t_1, t_2, \ldots, t_m$ and $w(n), \; n = s_1, s_2, \ldots, s_l$ are assumed to be outliers. The number of each outliers, m and l, the time of occurrence, $t_1, t_2, \ldots, t_m$ and $s_1, s_2, \ldots, s_l$, are unknown as well as the augmented variances $\sigma^2(n)$ and $\tau^2(n)$. Our procedure is as follows;

We first estimate the model (2.3.1) by assuming that there are no outliers. This means that we use constant values for the variances σ^2 and τ^2. Then the system noise and observational noise sequences are obtained from the Kalman Filter algorithm. A better model will be obtained by refitting the model by enlarging the variances using the estimated ratios.

2.6: State Space representation to maximum likelihood fitting of bilinear models with missing observations.

There are different state space representation of the BL model (See e.g. Pham-Dinh (1985), Guegan (1987) and Gabr (1991)). For simplicity we restrict ourselves to the stationary and invertible BL (p,0,p,1) model, namely

$$X_t + \sum_{i=1}^{p} a_i X_{t-i} = e_t + \sum_{j=1}^{p} b_j X_{t-j} e_{t-1} \qquad (2.6.1)$$

The conditions for stationarity and invertibility for the above model are given in Subba Rao and Gabr (1984) and Liu (1989). Gabr (1981) has used the following state space representation

$$X(t) = F(t) X(t-1) + Ce_t \qquad (2.6.2)$$

$$y_t = HX(t) \qquad (2.6.3)$$

The state space formulation of the BL model allows us to compute the likelihood function of observation when some of these observations are missing and to estimate the missing values.

Define $\alpha(t \mid t-1)$ and p(t|t-1) as the mean vector and covariance matrix of $X(t)$, conditional on the information at time t-1

Subba Rao (1993). Here, $\alpha(t \mid t-1)$ is viewed as an estimator for $X(t)$ and p(t|t-1) is regarded as its conditional error covariance, or mean square error (MSE) matrix. Given $\hat{X}(t-1)$, the optimal estimator of the state vector at time t-1, together with its MSE matrix p(t-1), defined by

$$p(t-1) = \mathrm{E}\left[\{\hat{X}(t-1) - X(t-1)\}\{\hat{X}(t-1) - X(t-1)\}^T\right],$$

the optimal estimator of $X(t)$ is given by

$$\alpha(t \mid t-1) = F(t)\hat{X}(t-1), \qquad (2.6.4)$$

where the covariance matrix of the estimation error is given by

$$p(t \mid t-1) = F(t)p(t-1)F^T(t) + \sigma^2 CC^T \qquad (2.6.5)$$

The updating equations, given a new observation y_t, are

$$\hat{X}(t) = \alpha(t \mid t-1) + p(t \mid t-1)H^T\left[y_t - H\alpha(t \mid t-1)\right]/S_t, \qquad (2.6.6)$$
$$p(t) = p(t \mid t-1) - p(t \mid t-1)H^T Hp(t \mid t-1)/S_t, \qquad (2.6.7)$$

where

$$S_t = p_{11}(t \mid t-1) = Hp(t \mid t-1)H^T \qquad (2.6.8).$$

Note that $p_{11}(t \mid t-1)$ is the upper left-hand element of the $p(t+1 \mid t)$. The prediction error is given by

$$V_t = y_t - H\alpha(t \mid t-1), \qquad (2.6.9)$$

(for details see, e.g. Ljung and Söderström (1983) and Harvey (1989)).

Given N observations $(y_1, y_2, \ldots, y_N)$, one seeks the MLE of the parameters. $\theta^T = [-a_1 \ldots -a_p b_1 \ldots -b_p]$ in (2.6.1). Let $Z(t-1) = \{y_{t-1}, y_{t-2}, \ldots, y_1\}$, then the conditional probability density function of y_t, conditional on Z(t-1) is normal for t = 1,2,...,N. Therefore, the likelihood function

$$L(y_1, y_2, \ldots, y_N, \theta, \sigma^2) = \prod_{t=1}^{N} f(y_t \mid z(t-1))$$

can be constructed by the production error decomposition, yielding

$$Log = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma^2 - \frac{1}{2}\sum_{t=1}^{N}\log S_t - \frac{1}{2\sigma^2}\sum_{t=1}^{N} V_t^2 / S_t \qquad (2.6.10).$$

The parameter σ^2 cannot be remove completely from (2.6.10) as in the linear ARMA models case (see Jones (1980) and Harvey (1989)). The reason is that initial values of the elements of the variance-covariance matrix p(t) contain different orders of σ^2 which makes it impossible to remove σ^2 completely.

It is very often the case in practice that the values of the Time Series are recorded at unequally space times, through failure to observe one or more values of the series. As in linear model (Jones (1980) and Harvey and Pierse (1984)), the prediction errors associated with non-missing observations can be obtained simply by skipping the Kalman Filter update equation at the points where the observations are missing. Thus, when an observation $y(t)$ is missing, the Kalman recursion skips equations (2.6.4), (2.6.5), (2.6.8) and (2.6.9) are simply replaced by

$$\hat{X}(t) = \alpha(t \mid t-1)$$
$$p(t) = p(t \mid t-1),$$

respectively.

We have to replace the missing values X(t) in the expression for F(t+1) by its estimate $\hat{X}(t)$. Thus, the model now is no longer exactly conditionally Gaussian but approximately so. Hence , the corresponding term in log likelihood given by (2.6.10), is omitted from the likelihood. Thus, the approximate likelihood function is of the form (2.6.10) with the summations covering only those values of t for which the variable is actually observed.

Ones the parameters of the BL model have been estimated, the approximate MSE estimate of missing observations can be calculated by smoothing.

The most straightforward of the smoothing algorithms, known as the fixed-point smoother, can be applied by augmenting the state-space model and applying the Kalman Filter. Full details can be found in Anderson and Moore (1979) and Harvey (1989).

CHAPER THREE

ESTIMATION OF MISSING OBSERVATIONS

3.0: Introduction

In chapter one we have given definitions and necessary fundamental concepts in time series analysis. In chapter two we have discussed procedures for estimation of missing observations using state space representations.

In this chapter we are going to deal with linear and nonlinear time series models which encompass several standard linear and nonlinear models for time series as special cases. We will outline two methods for estimating missing observations as given in Abraham and Thavaneswaran (1991), based on;

1. *prediction and fixed point smoothing algorithms* and
2. *optimal estimating equation theory.*

We have extended the first method to estimate more than two missing observations. Abraham and Thavaneswaran (1991) gave an estimate of up to two missing observations.

Recursive estimation of missing observations in an autoregressive conditionally heteroscedastic (ARCH) model and the estimation of missing observation in a linear time series model are shown to be special cases. Construction of optimal estimates of missing observations using estimating equation theory is discussed and applied to some nonlinear models. These two methods use state space representations described in chapter two.

3.1 Nonlinear state space models.

The linear state space system is given by,

$$\theta_{t+1} = \alpha_t \theta_t + \beta_t u_{t+1}, \; y_t = A_{t-1}\theta_t + B_{t-1}v_t \qquad (3.1.1)$$

where, θ_t and u_t are p x1 vectors, y_t and v_t are qx1 vectors, α_t and β_t are pxp matrices, and A_t and B_t are matrices of dimensions pxp and qxq, respectively $\{y_t\}$ represents the observed time series, whereas $\alpha_t, A_t, \beta_t, B_t$ are known matrices of nonrandom functions. The vectors $\{u_t\}, \{v_t\}$ are independent each being a sequence of independent normal random vectors, having components with zero mean and unit variances. In order to handle various deviations which may occur in practice several generalization of (3.1.1) have been suggested.

In this Chapter we consider the model in (3.1.1) with random coefficients. We allow the coefficients in (3.1.1) to depend on past observations as follows;

$$\alpha_t = \alpha\left(t, F_t^y\right), \; \beta_t = \beta\left(t, F_t^y\right), \; A_{t-1} = A\left(t-1, F_{t-1}^y\right) \text{ and } B_{t-1} = B\left(t-1, F_{t-1}^y\right),$$

where F_t^y denotes the σ-field generated by observation y up to time t. We refer to (3.1.1) under this settings as the generalized model. This generalized model encompasses some of the non linear time series models that have been proposed in the literature.

(i) ARCH models: supposed that $\alpha_t = \alpha, A_{t-1} = A$ and $B_t \equiv 0$ so that,

$$\theta_{t+1} = \alpha\theta_t + \beta_t u_{t+1}, y_t = A\theta_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.1.2).$$

This is the ARCH model describe in Engle (1982)

(ii) Dynamic linear state space models: When $\{\alpha_t\}, \{\beta_t\}$ and $\{B_t\}$ are constant matrices and A_{t-1} is a matrix of known functions at *t-1* (i.e. A_{t-1} is F_{t-1}^{y} measurable) the generalized model (3.1.1) becomes,

$$\theta_t = \alpha\theta_t + u_t$$

$$y_t = A_{t-1}\theta_t + v_t \text{ ....(3.1.3)}$$

Which is the state space model described in Harrison and stevens (1976).

(iii) Doubly stochastic time series model (c.f. Tjostheim (1986)):

When $\alpha_t = 1, \beta_t = 1$ $u_{t+1} = \varepsilon_{t+1} + \varepsilon_{t-1}$ and $B_t = 1$ then (3.1.1)becomes

$$\theta_{t+1} = \theta_t + u_{t+1}, y_t = A_{t-1} + v_t.$$

This corresponds to the doubly stochastic time series model,

$$\theta_{t+1} = \theta_t + \varepsilon_{t+1} + \varepsilon_{t-1}, y_t = \theta_t f\left(t, F_{t-1}^{y}\right) + e_t \text{ ....(3.1.5)}$$

considered in Thavaneswaran and Abraham (1988). When $f\left(t, F_{t-1}^{y}\right) = y_{t-1}$, this turns out to be a special case of the Random Coefficient Autoregression (RCA) model of Nicholls and Quin (1982). Moreover if we take,

$$\theta_t = \alpha_{t-1}\theta_{t-1} + \varepsilon_t, y_t = \theta_t \text{ with } \theta_t = \phi + \pi\exp\left(-\gamma y_{t-1}^2\right) \text{ ....(3.1.6).}$$

Then the generalized model (3.1.1) describes the exponential autoregressive model of Ozaki (1985).

Theorem 3.1.1. This theorem gives prediction and fixed point smoothing algorithms for the generalized model (3.1.1).

Let ,

$$\hat{\theta}_t = E\left(\theta_t \setminus F_{t-1}^{y}\right), \Sigma_t = \left[\left(\theta_t - \hat{\theta}_t\right)\left(\theta_t - \hat{\theta}_t\right)^T \setminus F_{t-1}^{y}\right].$$

Then,

$$\hat{\theta}_{t+1} = \alpha_t\hat{\theta}_t + k_t\left[y_t - \hat{y}_t\right],$$

$$\Sigma_{t+1} = \beta_t\beta_t^T + (\alpha_t - kA_{t-1})\Sigma_t(\alpha_t - k_tA_{t-1})^T + k_tB_{t-1}B_{t-1}^Tk_t^T$$

where,

$$k_t = \alpha_t \Sigma_t A_{t-1}^T \left(A_{t-1} \Sigma_t A_{t-1}^T + B_{t-1} B_{t-1}^T \right)^+ \quad \text{and} \quad \hat{y}_t = E\left(y_t \setminus F_{t-1}^y\right)$$

M^T and M^+ denote the transpose and the pseudoinverse respectively of a matrix M.

<u>Proof</u>: Let $w_t = y_t - \hat{y}_t$ where from the model (3.1.1).

$$y_t = A_{t-1}\theta_t + B_{t-1}v_t \quad \text{and} \quad \theta_{t+1} = \alpha_t \theta_t + \beta_t u_{t+1}$$

$$\mathrm{E}\left(y_t \setminus F_{t-1}^y\right) = A_{t-1}\mathrm{E}\left(\theta_t \setminus F_{t-1}^y\right) + B_{t-1}\mathrm{E}\left(v_t \setminus F_{t-1}^y\right).$$

but $v_t \sim \mathrm{N}(0,1)$, so that we get

$$\hat{y}_t = A_{t-1}\hat{\theta}_t \ .$$

Hence,

$$w_t = y_t - \hat{y}_t = A_{t-1}\theta_t + B_{t-1}v_t - A_{t-1}\hat{\theta}_t$$

that is,

$$w_t = A_{t-1}\left(\theta_t - \hat{\theta}_t\right) + B_{t-1}v_t \quad \text{......} (3.1.7).$$

$$\mathrm{E}\left[w_t w_t^T \setminus F_{t-1}^y\right] = \mathrm{E}\left[\left(A_{t-1}\left(\theta_t - \hat{\theta}_t\right) + B_{t-1}v_t\right)\left(A_{t-1}\left(\theta_t - \hat{\theta}_t\right) + B_{t-1}v_t\right)^T \setminus F_{t-1}^y\right]$$

$$= A_{t-1}\mathrm{E}\left[\left(\theta_t - \hat{\theta}_t\right)\left(\theta_t - \hat{\theta}_t\right)^T \setminus F_{t-1}^y\right]A_{t-1}^T + B_{t-1}\mathrm{E}\left(v_t v_t^T \setminus F_{t-1}^y\right)B_{t-1}^T .$$

but, $\mathrm{E}\left(v_t v_t^T \setminus F_{t-1}^y\right) = I$

and $\mathrm{E}\left[\left(\theta_t - \hat{\theta}_t\right)\left(\theta_t - \hat{\theta}_t\right)^T \setminus F_{t-1}^y\right] = \Sigma_t$.

Thus,

$$\mathrm{E}\left(w_t w_t^T \setminus F_{t-1}^y\right) = A_{t-1}\Sigma_t A_{t-1}^T + B_{t-1}B_{t-1}^T \quad \text{......} (3.1.8).$$

Also,

$$\mathrm{E}\left(\theta_t w_t^T \setminus F_{t-1}^y\right) = \mathrm{E}\left[\left(\theta_t - \hat{\theta}_t\right)w_t^T \setminus F_{t-1}^y\right]$$

since ,

$$\mathrm{E}\left(\hat{\theta}_t w_t^T \setminus F_{t-1}^y\right) = 0 .$$

Moreover it follows from the definition of Σ_t that ,

$$\mathrm{E}\left[\left(\theta_t-\hat{\theta}_t\right)w_t^T \setminus F_{t-1}^y\right]=\mathrm{E}\left[\left(\theta_t-\hat{\theta}_t\right)\left\{A_{t-1}\left(\theta_t-\hat{\theta}_t\right)+B_{t-1}v_t\right\}^T \setminus F_{t-1}^y\right],$$

$$=\mathrm{E}\left[\left(\theta_t-\hat{\theta}_t\right)\left(\theta_t-\hat{\theta}_t\right)^T \setminus F_{t-1}^y\right]A_{t-1}^T+\mathrm{E}\left[\left(\theta_t-\hat{\theta}_t\right)v_t^T \setminus F_{t-1}^y\right]B_{t-1}^T$$

but $\mathrm{E}\left[\left(\theta_t-\hat{\theta}_t\right)v_t \setminus F_{t-1}^y\right]=0$, then

$$\mathrm{E}\left[\left(\theta_t-\hat{\theta}_t\right)w_t \setminus F_{t-1}^y\right]=\Sigma_t A_{t-1}^T \qquad (3.1.9).$$

Using the fact that the σ-field generated by the observations up to time t , namely F_t^y is the same as the σ-field generated by w_t, F_t^w, we have,

$$\hat{\theta}_{t+1}=\mathrm{E}\left(\theta_{t+1} \setminus F_t^y\right)=\alpha_t \mathrm{E}\left(\theta_t \setminus F_t^y\right)$$
$$=\alpha_t \mathrm{E}\left(\theta_t \setminus F_{t-1}^y, w_t\right).$$

Using the properties of normal random vectors i.e.

$$\mathrm{E}(X \setminus Y)=\mathrm{E}(X)+\frac{\operatorname{cov}(XY)}{\operatorname{var}(Y)}(Y-\mathrm{E}(Y))$$

$$Var(X \mid Y)=Var(X)-\frac{Cov^2(XY)}{Var(Y)}.$$

Hence,

$$\hat{\theta}_{t+1}=\alpha_t \mathrm{E}\left(\theta_t \setminus F_{t-1}^y, w_t\right)$$

$$=\alpha_t \mathrm{E}\left[\theta_t \setminus F_{t-1}^y\right]+k_t\left(y_t-\hat{y}_t\right),$$

$$\hat{\theta}_{t+1}=\alpha_t \hat{\theta}_t+k_t\left(y_t-\hat{y}_t\right) \qquad (3.1.10)$$

where,

$$k_t=\alpha_t \mathrm{E}\left[\theta_t w_t^T \setminus F_{t-1}^y\right]\left\{\mathrm{E}\left[w_t w_t^T \setminus F_{t-1}^y\right]\right\}^+.$$

But from (3.1.8) and (3.1.9) k_t becomes,

$$k_t=\alpha_t \Sigma_t A_{t-1}^T\left[A_{t-1}\Sigma_t A_{t-1}^T+B_{t-1}B_{t-1}^T\right]^+ \qquad (3.1.11).$$

Moreover,

$$\theta_{t+1}-\hat{\theta}_{t+1}=\alpha_t\theta_t+\beta_{t-1}u_{t+1}-\alpha_t\hat{\theta}_t-k_t\left(y_t-\hat{y}_t\right)$$

$$= \alpha_t\left(\theta_t - \hat{\theta}_t\right) + \beta_t u_{t+1} - k_t\left[A_{t-1}\left(\theta_t - \hat{\theta}_t\right) + B_{t-1}v_t\right]$$

$$= \left(\theta_t - \hat{\theta}_t\right)\left(\alpha_t - k_t A_{t-1}\right) + \beta_t u_{t-1} - B_{t-1}v_t k_t \ .$$

But,

$$\Sigma_{t+1} = \mathrm{E}\left[\left(\theta_{t+1} - \hat{\theta}_{t+1}\right)\left(\theta_{t+1} - \hat{\theta}_{t+1}\right)^T \setminus F_{t-1}^y\right] \text{..} (3.1.12)$$

$$= \mathrm{E}\left[\left\{\left(\theta_t - \hat{\theta}_t\right)\left(\alpha_t - k_t A_{t-1}\right) + \beta_t u_{t+1} - B_{t-1}v_t k_t\right\}\left\{\left(\theta_t - \hat{\theta}_t\right)\left(\alpha_t - k_t A_{t-1}\right) + \beta_t u_{t+1} - B_{t-1}v_t k_t\right\}^T \setminus F_{t-1}^y\right]$$

since $u_{t+1} \sim \mathrm{N}(0,1)$ and $v_t \sim \mathrm{N}(0,1)$

then,

$$\Sigma_{t+1} = \left(\alpha_t - k_t A_{t-1}\right)\mathrm{E}\left[\left(\theta_t - \hat{\theta}_t\right)\left(\theta_t - \hat{\theta}_t\right) \setminus F_{t-1}^y\right]\left[\alpha_t - k_t A_{t-1}\right]^T + \beta_t \mathrm{E}\left(u_{t+1}u_{t+1}^T \setminus F_{t-1}^y\right)\beta_t^T + B_{t-1}k_t \mathrm{E}\left(v_t v_t^T \setminus F_{t-1}^y\right)B_{t-1}^T k_t^T$$

so that,

$$\Sigma_{t+1} = \left(\alpha_t - k_t A_{t-1}\right)\Sigma_t\left(\alpha_t - k_t A_{t-1}\right)^T + \beta_t\beta_t^T + k_t B_{t-1}B_{t-1}^T k_t^T \text{...........} (3.1.13).$$

Hence the proof of theorem (3.1.1).

We now introduce another theorem on fixed-point smoother to obtain recursive estimates of m missing observations say , $y_m = \left(y_{t_1}, y_{t_2}, ..., y_{t_j}, ..., y_{t_m}\right)$. The basic idea here is the same as that in the derivation of the recursive estimate of a parameter $\theta_{t_j}\,(j = 1,2,.....,m)$,base on the observations up to time $t\left(t > t_j\right)$, as a function of the estimate base on $t-1\left(t > t_j + 1\right)$ and the observation at time t. This will also enable us to get an idea of how the estimate of the parameter (missing value) changes when a new observation becomes available.

Theorem 3.1.2.

For $t > t_j$, let $\tilde{\theta}_{t_j \setminus t} = \mathrm{E}\left(\theta_{t_j} \setminus F_t^y\right)$ be the estimate of θ_{t_j} base on the observations up to time t, Σ_t be the covariance matrix ,

$$\widetilde{\Sigma}_t = \mathrm{E}\left[\left(\theta_{t_j} - \hat{\theta}_{t_j}\right)\left(\theta_{t_j} - \hat{\theta}_{t_j}\right)^T \setminus F_{t-1}^y\right] \text{ and } \Sigma_t^* = \mathrm{E}\left[\left(\theta_{t_j} - \hat{\theta}_{t_j \setminus t}\right)\left(\theta_{t_j} - \hat{\theta}_{t_j \setminus t}\right)^T \setminus F_{t-1}^y\right].$$

Then ,

$$\widetilde{\theta}_{t_j \setminus t} = \widetilde{\theta}_{t_j \setminus t-1} + \widetilde{k}_t\left(y_t - A_{t-1}\hat{\theta}_t\right), \quad t > t_j ,$$

where

$$\widetilde{k}_t = \widetilde{\Sigma}_t A_{t-1}^T\left[A_{t-1}\Sigma_t A_{t-1}^T - B_{t-1}B_{t-1}^T\right]^+ ,$$
$$\widetilde{\Sigma}_{t+1} = \widetilde{\Sigma}_t\left[\alpha_t - k_t A_{t-1}\right]^T , \qquad \Sigma_{t-1}^* = \Sigma_t \quad \text{for } t < t_j \text{ and}$$
$$\Sigma_t^* = \Sigma_{t-1}^* - \widetilde{\Sigma}_t A_{t-1}^T \widetilde{k}_t^T ,\ldots\ldots t \geq t_j .$$

Proof: Given $t > t_j$,

$$\widetilde{\theta}_{t_j \setminus t} = \mathrm{E}\left(\theta_{t_j} \setminus F_t^y\right) = \mathrm{E}\left(\theta_{t_j} \setminus F_{t-1}^y, w_t\right) \ldots\ldots (3.1.15),$$

where ,

$$w_t = y_t - \hat{y}_t .$$

Applying the results from the Normal theory we get

$$\widetilde{\theta}_{t_j \setminus t} = \mathrm{E}\left(\theta_{t_j} \setminus F_{t-1}^y\right) + \widetilde{k}_t w_t \ldots\ldots (3.1.15).$$

Equation (3.1.1) gives

$$y_t = A_{t-1}\theta_t + B_{t-1}v_t \quad \text{where} \quad v \sim \mathrm{N}(0,1).$$

But ,

$$\hat{y}_t = \mathrm{E}\left(y_t \setminus F_{t-1}^y\right) = A_{t-1}\hat{\theta}_t$$

so that ,

$$\widetilde{\theta}_{t_j \setminus t} = \widetilde{\theta}_{t_j \setminus t-1} + \widetilde{k}_t\left(y_t - A_{t-1}\hat{\theta}_t\right) \ldots\ldots (3.1.16),$$

where

$$\widetilde{k}_t = \mathrm{E}\left(\theta_{t_j} w_t^T \setminus F_{t-1}^y\right)\left[\mathrm{E}\left(w_t w_t^T \setminus F_{t-1}^y\right)\right]^+ .$$

The second factor remains the same as the "innovation" variance as in k_t.

As in theorem (3.1.1) we have

$$\mathrm{E}\left(\theta_t w_t^T \setminus F_{t-1}^y\right) = \mathrm{E}\left[\left(\theta_{t_j} - \hat{\theta}_{t_j}\right) w_t^T \setminus F_{t-1}^y\right], \sin ce \mathrm{E}\left(\hat{\theta}_{t_j} \setminus F_{t-1}^y\right) = 0$$

but ,

$$w_t^T = (y_t - \hat{y}_t)^T = \left[A_{t-1}\theta_{t_j} + B_{t-1}v_t - A_{t-1}\hat{\theta}_{t_j}\right]^T$$
$$= \left[A_{t-1}\left(\theta_{t_j} - \hat{\theta}_{t_j}\right) + B_{t-1}v_t\right]^T.$$

Now,

$$\mathrm{E}\left[\left(\theta_{t_j} - \hat{\theta}_{t_j}\right) w_t^T \setminus F_{t-1}^y\right] = \mathrm{E}\left[\left(\theta_{t_j} - \hat{\theta}_{t_j}\right)\left\{A_{t-1}\left(\theta_{t_j} - \hat{\theta}_{t_j}\right) + B_{t-1}v_t\right\}^T \setminus F_{t-1}^y\right]$$
$$= \left[\left(\theta_{t_j} - \hat{\theta}_{t_j}\right)\left(\theta_{t_j} - \hat{\theta}_{t_j}\right)^T \setminus F_{t-1}^y\right] A_{t-1} + \mathrm{E}\left[\left(\theta_{t_j} - \hat{\theta}_{t_j}\right) v_t^T \setminus F_{t-1}^y\right] B_{t-1}^T$$
$$= \widetilde{\Sigma}_t A_{t-1}$$
$$Since, \mathrm{E}\left[\left(\theta_{t_j} - \hat{\theta}_{t_j}\right) v_t^T \mid F_{t-1}^y\right] = 0.$$

Also

$$\mathrm{E}\left(w_t w_t^T \mid F_{t-1}^y\right) = \mathrm{E}\left[\left(A_{t-1}\left(\theta_t - \hat{\theta}_t\right) + B_{t-1}v_t\right)\left(A_{t-1}\left(\theta_t - \hat{\theta}_t\right) + B_{t-1}v_t\right)^T \mid F_{t-1}^y\right]$$

which leads to

$$\mathrm{E}\left(w_t w_t^T \mid F_{t-1}^y\right) = A_{t-1}\widetilde{\Sigma}_t A_{t-1}^T + B_{t-1}B_{t-1}^T \quad \text{.......................................}(3.1.18).$$

Then,

$$\widetilde{k}_t = \widetilde{\Sigma}_t A_{t-1}^T \left[A_{t-1}\Sigma_t A_{t-1}^T + B_t B_{t-1}^T\right]^+ \quad \text{.......................................}(3.1.19)$$

and

$$\widetilde{\Sigma}_{t+1} = \widetilde{\Sigma}_t\left(\alpha_t - k_t A_{t-1}\right)^T \quad \text{.......................................}(3.1.20),$$

$\Sigma_t^* = \Sigma_t$ for $t < t_j$ also

$$\Sigma_t^* = \Sigma_{t-1}^* - \widetilde{\Sigma}_t A_{t-1}^T \widetilde{k}_t^T, t \geq t_j$$

3.2 Application to missing data

There are two different approaches of estimating missing values in time series;

1. *A Bayesian approach* : Which uses Kalman filtering technique.
2. *A non-Bayesian approach* : where the missing values are treated as parameters (fixed). In section 3.3 to 3.5 we follow the Kalman type recursive approach to estimate the missing values by replacing them with normal random variables. This type of approach may be viewed as one which uses a prior for the parameter which replaces the missing value.

3.3: Autoregressive Conditionally Heteroscedastic (ARCH) type models with one missing observation.

Now we indicate an appropriate way to modify a given non linear time series to reflect the fact that the observation at time m is missing . Let $\{X_t\}$ be a time series in which X_m is missing and $X'_N = \{X_1, X_2, ..., X_{m-1}, X_{m+1}, ..., X_N\}$. If we know the first two conditional moments $E[X_{t+1} \setminus F_t^x]$ and $Var[X_{t+1} \setminus F_t^x]$,then X_{t+1} can be written as

$$X_{t+1} = E[X_{t+1} \setminus F_t^x] + X_{t+1} - E[X_{t+1} \setminus F_t^x] \quad \text{.........(3.3.1)}.$$

Suppose that the time series X_t satisfies,

$$E[X_{t+1} \setminus F_t^x] = \alpha_{t-1} X_t \text{ and } X_{t+1} - E[X_{t+1} \setminus F_t^x] = \beta_{t-1} u_{t+1} \quad (3.3.2).$$

where α_{t-1} and β_{t-1} are F_{t-1}^x measurable and $\{u_t\}$is an i.i.d. $\sim N(0,1)$ sequence. Then X_{t+1} has the ARCH representation .

$$X_{t+1} = \alpha_{t-1} X_t + \beta_{t-1} u_{t+1} \quad \text{.........(3.3.3)}$$

The restriction in (3.3.2) is introduced to apply the recursive approach . However, the method of Section 3.6. can be applied in the more general set up in which the coefficient of X_t and u_{t+1} are F_t^x measurable.

Now we consider the estimation of a missing observation as a parameter estimation problem in a particular formulation of the generalized model (3.1.1).

$$\left.\begin{aligned} \theta_{t+1} &= \alpha_{t-1}\theta_t + \beta_{t-1}u_{t+1} \\ X_t &= A_{t-1}\theta_t \\ y_t &= A_{t-1}\theta_t + B_{t-1}v_t \end{aligned}\right\} \qquad (3.3.4)$$

with $A_{m-1} = 0, B_{m-1} = 1, A_t = 1, t \neq m; B_t = 0, t \neq m;$

Then $Y = (X_1, X_2, \ldots, X_{m-1}, v_m, X_{m+1}, X_{m+2}, \ldots, X_N)$ is the extended observed series. Here v_m is a random variable replacing the missing observation. Such a formulation was also considered in Abraham and Thavaneswaran (1991).

Using theorem 3.1.1 and 3.1.2 we have equations

$$k_t = \alpha_{t-1}\Sigma_t A_{t-1}\left(A_{t-1}^2\Sigma_t + B_{t-1}^2\right)^+,$$
$$\Sigma_{t+1} = \beta_{t-1}^2 + (\alpha_{t-1} - k_t A_{t-1})\Sigma_t + k_t^2 B_{t-1}^2.$$

Substituting the values of A_t, A_{t-1}, B_t and B_{t-1} gives

$$k_t = \alpha_{t-1}\Sigma_t[\Sigma_t]^+$$
$$k_t = \alpha_{t-1} \qquad (3.3.5)$$

for $t \neq m$

and

$$\Sigma_{t+1} = \beta_{t-1}^2 \qquad (3.3.6).$$

Also for $t = m$

$$k_m = \alpha_{m-1}\Sigma_m A_{m-1}\left[A_{m-1}^2\Sigma_m + B_{m-1}^2\right]^+,$$

but $A_{m-1} = 0$, hence

$$k_m = 0 \qquad (3.3.7)$$

and $\Sigma_{m+1} = \beta_{m-1}^2 + (\alpha_{m-1} - k_m A_{m-1})^2 \Sigma_m + k_t^2 B_{m-1}^2$,

but $k_m = 0$ to give

$$\Sigma_{m+1} = \beta_{m-1}^2 + \alpha_{m-1}^2 \Sigma_m$$

where from (3.3.6) $\Sigma_{t+1} = \beta_{t-1}^2$, so that at t = m

$$\Sigma_{m+1} = \beta_{m-1}^2$$
$$\Sigma_m = \beta_{m-2}^2.$$

Substituting for Σ_m we have

$$\Sigma_{m+1} = \beta_{m-1}^2 + \alpha_{m-1}^2 \beta_{m-2}^2 \quad \text{.....(3.3.8)}.$$

Also from theorem 3.1.2 we have

$$\widetilde{\Sigma}_{t+1} = \widetilde{\Sigma}_t (\alpha_{t-1} - k_t A_{t-1}) \text{ for } t \neq m.$$

But from (3.3.5) $k_t = \alpha_{t-1}$ and $A_{t-1} = 1$

so that ,

$$\widetilde{\Sigma}_{t-1} = 0 \quad \text{.....(3.3.9)},$$

and $\widetilde{\Sigma}_{m+1} = \widetilde{\Sigma}_m (\alpha_{m-1} - k_m A_{m-1})$

but $k_m = 0$

so that

$$\widetilde{\Sigma}_{m+1} = \widetilde{\Sigma}_m \alpha_{m-1}, \text{ also } \widetilde{\Sigma}_m = \Sigma_m = \beta_{m-2}^2.$$

Hence,

$$\widetilde{\Sigma}_{m+1} = \alpha_{m-1}^2 \beta_{m-2}^2 \quad \text{.....(3.3.10)}.$$

We also have from (3.1.16)

$$\widetilde{\theta}_{t_j|t} = \widetilde{\theta}_{t_j|t} + \widetilde{k}_t (y_t - A_{t-1}\hat{\theta}_t),$$

but in this case $t_j = m$ and $A_t = 1$ gives,

$$\widetilde{\theta}_{m|t} = \widetilde{\theta}_{m|t-1} + \widetilde{k}_t (y_t - \hat{\theta}_t) \text{ and at } t = m+1$$

$$\widetilde{\theta}_{m|m+1} = \widetilde{\theta}_{m|m} + \widetilde{k}_{m+1} (y_{m+1} - \hat{\theta}_{m+1}) \quad \text{.....(3.3.11)}$$

since, $\theta_t = y_t = X_t$

To find $\tilde{\theta}_{m|m}$ we have that

$$\tilde{\theta}_{m|m} = \mathrm{E}\left[\theta_m \mid F_m^y\right] = \mathrm{E}\left[\left(\alpha_{m-2}\theta_{m-1} + \beta_{m-2}u_m\right) \mid F_m^y\right]$$

$$= \alpha_{m-2}\hat{\theta}_{m-1} = \alpha_{m-2}X_{m-1}.$$

Next,

$$\hat{\theta}_{m+1} = \mathrm{E}\left[\theta_{m+1} \mid F_{m-1}^y\right] = \mathrm{E}\left[\left(\alpha_{m-1}\theta_m + \beta_{m-1}u_{m+1}\right) \mid F_{m-1}^y\right]$$

which gives $\hat{\theta}_{m+1} = \alpha_{m-1}\hat{\theta}_m$ and $\hat{\theta}_m = \alpha_{m-2}\hat{\theta}_{m-1} = \alpha_{m-2}X_{m-1}$ combining these gives $\hat{\theta}_{m+1} = \alpha_{m-1}\alpha_{m-2}X_{m-1}$.

Moreover,

$$\tilde{k}_t = \tilde{\Sigma}_t A_{t-1}^T\left[A_{t-1}\Sigma_t A_{t-1}^T + B_{t-1}B_{t-1}^T\right]^+$$

$$\tilde{k}_t = \frac{\tilde{\Sigma}_t}{\Sigma_t} \qquad (3.3.12)$$

since $A_{t-1} = 1, B_{t-1} = 0$ and at t = m+1 (3.3.12) leads to

$$\tilde{k}_{m+1} = \frac{\tilde{\Sigma}_{m+1}}{\Sigma_{m+1}} = \frac{\alpha_{m-1}\beta_{m-2}^2}{\beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2} \qquad (3.3.13)$$

from (3.3.8) and (3.3.10).

Thus the estimate of the *m-th* observation based on X_{m+1} is

$$\tilde{X}_{m|m+1} = \alpha_{m-2}X_{m-1} + \frac{\alpha_{m-1}\beta_{m-2}^2}{\beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2}\left[X_{m+1} - \alpha_{m-1}\alpha_{m-2}X_{m-1}\right] \qquad (3.3.14),$$

which simplifies to

$$\tilde{X}_{m|m+1} = \frac{\alpha_{m-2}\left[\beta_{m-1}^2\right]X_{m-1} + \alpha_{m-1}\beta_{m-2}^2X_{m+1}}{\beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2} \qquad (3.3.15).$$

Moreover, for a nonlinear model of the form

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

in which the *m-th* observation X_m is missing, the estimate of X_m based on F_{m+1}^y is given by

$$\widetilde{X}_{m|m+1} = \frac{\phi X_{m-1}}{1+\phi^2 X_{m-1}^2}\left[X_{m-2} + X_{m+1}\right]$$

Autoregressive models with deterministic time varying coefficient:

Model of the form

$$X_t - \alpha(t,\phi)X_{t-1} = u_t$$

have been found to be quite useful, as in 3.3.15 it can be shown that the estimate $\widetilde{X}_{m|m+1}$ of the missing observation based on F_{m+1}^{y} is given by

$$\widetilde{X}_{m|m+1} = \frac{\alpha(m+1,\phi)X_{m+1} + \alpha(m,\phi)X_{m-1}}{1+\alpha^2(m+1,\phi)}$$

Bilinear models:

Consider the model

$$X - \phi X = cu_t + \beta X_{t-2}u_t$$

The estimation of a missing observation, X_m, can be obtained by writing the model as

$$X_t = \alpha_{t-2}X_{t-1} + \beta_{t-2}u_t$$

where

$$\alpha_{t-2} = \phi \quad \text{and} \quad \beta_{t-2} = c + \beta X_{t-2}$$

Hence the estimate of $X_m, \widetilde{X}_{m|m+1} = \mathrm{E}\left[X_m \mid F_{m+1}^{y}\right]$ and is given by

$$\widetilde{X}_{m|m+1} = \frac{\phi\beta_{m-2}^2 X_{m+1} + \phi\beta_{m-1}^2 X_{m-1}}{\beta_{m-1}^2 + \phi^2\beta_{m-2}^2}.$$

3.4: Two consecutive missing observations.

(a) The estimate of *mth* observation based on X_{m+2}.

We consider a slightly modified form of the model (3.3.4) in which we let

$$X_{t+1} = \alpha_{t-2}X_t + \beta_{t-2}u_{t+1}$$

Here X_m and X_{m+1} missing and α_t is F_t^x measurable. The problem is to estimate X_m based on the available data

$$X'_N = (X_1, X_2, \dots, X_{m-1}, X_{m+2}, \dots X_N).$$

The corresponding state space model may be written as

$$\left.\begin{aligned}\theta_{t+1} &= \alpha_{t-2}\theta_t + \beta_{t-2}u_{t+1}\\ X_t &= A_{t-1}\theta_t\\ y_t &= A_{t-1}\theta_t + B_{t-1}v_t\end{aligned}\right\}\dots\dots(3.4.1)$$

where $A_m = A_{m-1} = 0, B_m = B_{m-1} = 1$ and $B_t = 0, A_t = 1$ for $t \neq m, m+1$.

Then $Y = (X_1, X_2, \dots, X_{m-1}, v_m, X_{m+2}, \dots X_N)$ is the extended observed series Here v_m is a normal random variable replacing the missing observation. By theorem (3.1.1) we need to show that for $t \neq m, m+1, k_t = \alpha_{t-2}$ and $k_m = k_{m+1} = 0$.

Hence,

$$\left.\begin{aligned}k_t &= \alpha_{t-2}\Sigma_t A_{t-1}\left[A_{t-1}^2\Sigma_t + B_{t-1}^2\right]^+,\\ \Sigma_{t+1} &= \beta_{t-2}^2 + (\alpha_{t-2} - k_t A_{t-1})^2\Sigma_t + k_t^2 B_{t-1}^2\end{aligned}\right\}\dots\dots(3.4.2).$$

This implies that,

$$k_t = \alpha_{t-2}\dots\dots(3.4.3)$$

$$\Sigma_{t+1} = \beta_{t-2}^2\dots\dots(3.4.4),$$

and at $t = m, m+1, A_{m-1} = A_m = 0$ and $B_{m-1} = B_m = 1$ we obtain

$$k_m = k_{m+2} = 0.$$

Next equation (3.4.2) gives

$$\Sigma_{m+1} = \beta_{m-2}^2 + \alpha_{m-2}^2\Sigma_m\dots\dots(3.4.5)$$

$$\Sigma_{m+2} = \beta_{m-1}^2 + \alpha_{m-1}^2\Sigma_{m+1}\dots\dots(3.4.6).$$

Substituting (3.4.5) in (3.4.6) we get

$$\begin{aligned}\Sigma_{m+2} &= \beta_{m-1}^2 + \alpha_{m-1}^2(\beta_{m-2}^2 + \alpha_{m-2}^2\Sigma_m),\\ &= \beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\Sigma_m.\end{aligned}$$

Since from (3.4.4) $\Sigma_{t+1} = \beta_{t-2}^2$ then at t = m

$$\Sigma_{m+1} = \beta_{m-2}^2,$$
$$\Sigma_m = \beta_{m-3}^2 \qquad (3.4.7).$$

Thus,

$$\Sigma_{m+2} = \beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\beta_{m-3}^2 \qquad (3.4.8)$$

Also from theorem 3.1.2 we have from equations (3.1.19) and (3.1.20)

$$\tilde{k}_t = \tilde{\Sigma}_t A_{t-1}\left[A_{t-1}^2\Sigma_t + \beta_{t-1}^2\right]^+,$$
$$\tilde{\Sigma}_{t+1} = \tilde{\Sigma}_t(\alpha_{t-1} - k_t A_{t-1}).$$

Hence we have

$$\tilde{k}_t = \tilde{\Sigma}_t[\Sigma_t]^+ = \frac{\tilde{\Sigma}_t}{\Sigma_t} \qquad (3.4.9)$$

and

$$\tilde{\Sigma}_{t+1} = \tilde{\Sigma}_t\alpha_{t-1} \qquad (3.4.10)$$

since at

$A_m = A_{m-1} = 0; B_m = B_{m-1} = 1; A_t = 1, B_t = 0$ at $t \neq m, m+1$ also we get

$$\tilde{k}_m = \tilde{k}_{m+1} = 0 \qquad (3.4.11).$$

Next we have

$$\tilde{\Sigma}_{m+1} = \tilde{\Sigma}_m\alpha_{m-2} \qquad (3.4.12)$$
$$\tilde{\Sigma}_{m+2} = \tilde{\Sigma}_{m+1}\alpha_{m-1} \qquad (3.4.13).$$

Substituting (3.4.12) in (3.4.13) gives

$$\tilde{\Sigma}_{m+2} = \alpha_{m-1}\alpha_{m-2}\tilde{\Sigma}_m = \alpha_{m-1}\alpha_{m-2}\Sigma_m.$$

But from (3.4.7) $\Sigma_m = \beta_{m-3}^2$

so that,

$$\tilde{\Sigma}_{m+2} = \alpha_{m-1}\alpha_{m-2}\beta_{m-3}^2 \qquad (3.4.14).$$

We have from equation (3.1.16) as

$$\tilde{\theta}_{m|t} = \tilde{\theta}_{m|t-1} + \tilde{k}_t\left(y_t - A_{t-1}\hat{\theta}_t\right),$$

but $A_t = 1$ so that this equation becomes

$$\tilde{\theta}_{m|t} = \tilde{\theta}_{m|t-1} + \tilde{k}_t\left(y_t - \hat{\theta}_t\right)$$

and at t = m+2 we obtain

$$\tilde{\theta}_{m|m+2} = \tilde{\theta}_{m|m+1} + \tilde{k}_{m+2}\left(y_{m+2} - \hat{\theta}_{m+2}\right).$$

But,

$$\tilde{\theta}_{m|m+1} = \mathrm{E}\left[\theta_m \mid F^y_{m+1}\right] = \mathrm{E}\left[\left(\alpha_{m-3}\theta_{m-1} + \beta_{m-3}u_m\right) \mid F^y_{m+1}\right]$$

$$= \mathrm{E}\left[\alpha_{m-3}\theta_{m-1} \mid F^y_{m+1}\right] = \alpha_{m-3}\hat{\theta}_{m-1} = \alpha_{m-3}X_{m-1}.$$

Next,

$$\hat{\theta}_{m+2} = \mathrm{E}\left[\theta_{m+2} \mid F^y_{m+1}\right] = \mathrm{E}\left[\left(\alpha_{m-1}\theta_{m+1} + \beta_{m-1}u_{m+1}\right) \mid F^y_{m+1}\right]$$

$$= \alpha_{m-1}\mathrm{E}\left[\theta_{m+1} \mid F^y_{m+1}\right] = \alpha_{m-1}\hat{\theta}_{m+1}$$

$\hat{\theta}_{m+2} = \alpha_{m-1}\hat{\theta}_{m+1}$, $\hat{\theta}_{m+1} = \alpha_{m-2}\hat{\theta}_m$ and $\hat{\theta}_m = \alpha_{m-3}\hat{\theta}_{m-1} = \alpha_{m-3}X_{m-1}$ which combines to give $\hat{\theta}_{m+2} = \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}X_{m-1}$.

Thus the estimate of m*th* observation based on X_{m+2} is

$$\tilde{X}_{m|m+2} = \alpha_{m-3}X_{m-1} + \tilde{k}_{m+2}\left(X_{m+2} - \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}X_{m-1}\right) \qquad (3.4.15).$$

But

$$\tilde{k}_{m+2} = \frac{\alpha_{m-1}\alpha_{m-2}\beta^2_{m-3}}{\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2} + \alpha^2_{m-1}\alpha^2_{m-2}\beta^2_{m-3}} \qquad (3.4.16).$$

Hence (3.4.15) becomes,

$$\tilde{X}_{m|m+2} = \alpha_{m-3}X_{m-1} + \frac{\alpha_{m-1}\alpha_{m-2}\beta^2_{m-3}}{\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2} + \alpha^2_{m-1}\alpha^2_{m-2}\beta^2_{m-3}}\left(X_{m+2} - \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}X_{m-1}\right),$$

which simplifies to give

$$\tilde{X}_{m|m+2} = \frac{\alpha_{m-3}\left(\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2}\right)X_{m-1} + \alpha_{m-1}\alpha_{m-2}\beta^2_{m-3}X_{m+2}}{\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2} + \alpha^2_{m-1}\alpha^2_{m-2}\beta^2_{m-3}} \qquad (3.4.17).$$

Which is the required estimate of *mth* observation based on X_{m+2}.

(b) The estimate of the (*m+1)th* observation based on X_{m+2}.

We estimate X_{m+1} observation based on the available data $X'_N = (X_1, X_2, ..., X_{m-1}, \hat{X}_m, X_{m+2}, ..., X_N)$, where $\hat{X}_m$ is the estimate of X_m from the previous step.

The corresponding state space models becomes,

$$\left.\begin{aligned} \theta_{t+1} &= \alpha_{t-3}\theta_t + \beta_{t-3}u_{t+1} \\ X_t &= A_{t-1}\theta_t \\ y_t &= A_{t-1}\theta_t + B_{t-1}v_t \end{aligned}\right\} \quad (3.4.18).$$

In this case we have $A_m = 0, B_m = 1;$ and $B_t = 0, A_t = 1$ for $t \neq m+1$. The extended data will be, $Y = (X_1, X_2, ..., X_{m-1}, \hat{X}_m, v_{m+1}, X_{m+2}, ... X_N)$, where v_{m+1} is a normal random variable replacing the missing observation. This estimate is now treated as if only one observation is missing in the data.

Using theorems 3.1.1 we have the modified form as,

$\Sigma_{t+1} = \beta_{t-3}^2 + (\alpha_{t-3} - k_t A_{t-1})^2 \Sigma_t + k_t^2 B_{t-1}^2$ and

$$k_t = \alpha_{t-3}\Sigma_t A_{t-1}\left[A_{t-1}^2\Sigma_t - \beta_{t-1}^2\right]^+ \text{, then}$$

$$k_t = \alpha_{t-3}\Sigma_t[\Sigma_t]^+ = \alpha_{t-3} \quad (3.4.19)$$

and $\Sigma_{t+1} = \beta_{t-3}^2 + (\alpha_{t-3} - k_t)^2 \Sigma_t$

But from (3.4.19) $k_t = \alpha_{t-3}$, hence

$$\Sigma_{t+1} = \beta_{t-3}^2 \quad (3.4.20).$$

Also from theorem 3.1.2 we have

$$\tilde{k}_t = \tilde{\Sigma}_t A_{t-1}\left[A_{t-1}^2\Sigma_t + B_{t-1}^2\right]^+ \text{ and } \tilde{\Sigma}_{t+1} = \tilde{\Sigma}_t(\alpha_{t-3} - k_t A_{t-1})^T$$

since,

$A_t = 1, B_t = 0; t \neq m+1$ then this becomes

$$\tilde{k}_t = \Sigma_t[\Sigma_t]^+ = \frac{\tilde{\Sigma}_t}{\Sigma_t} \text{ as in } (3.4.9)$$

and

$$\tilde{\Sigma}_{t+1} = 0 \qquad (3.4.21)$$

since $A_{t-1} = 1$ and $k_t = \alpha_{t-3}$.

Setting $t = m+1$ $A_m = 0, B_m = 1$ we obtain

$$k_{m+1} = 0 \qquad (3.4.22)$$

and $\Sigma_{m+1} = \beta^2_{m-2} + \alpha^2_{m-2}\Sigma_{m+1}$.

But from (3.4.20) at $t = m$

$$\Sigma_{m+2} = \beta^2_{m-3}$$

hence,

$$\Sigma_{m+2} = \beta^2_{m-2} + \alpha^2_{m-2}\beta^2_{m-3} \qquad (3.4.23)$$

and $\tilde{\Sigma}_{m+2} = \tilde{\Sigma}_{m+2}(\alpha_{m-2} - k_{m+1}A_m)$

at $A_m = 0, k_{m+1} = 0$ we have

$$\tilde{\Sigma}_{m+2} = \alpha_{m-2}\tilde{\Sigma}_{m+1} \text{ and } \tilde{k}_{m+1} = 0.$$

Also,

$$\tilde{\Sigma}_{m+2} = \alpha_{m-2}\tilde{\Sigma}_{m+1} = \alpha_{m-2}\Sigma_{m+1} = \alpha_{m-2}\beta^2_{m-3}$$

$$\tilde{\Sigma}_{m+2} = \alpha_{m-2}\beta^2_{m-3} \qquad (3.4.24).$$

But at $t \geq t_j$, $\tilde{\theta}_{t_j|t} = \tilde{\theta}_{t_j|t-1} + \tilde{k}_t(y_t - \hat{\theta}_t)$

since $A_t = 1$.In this case if we let $t_j = m+1$ and $t = m+2$ we have

$$\tilde{\theta}_{m+1|m+2} = \tilde{\theta}_{m+1|m+1} + \tilde{k}_{m+2}(y_{m+2} - \hat{\theta}_{m+2})$$

where,

$$\hat{\theta}_{m+1|m+1} = \mathrm{E}[\theta_{m+1} \mid F^y_{m+1}] = \mathrm{E}[(\alpha_{m-3}\theta_m + \beta_{m-3}u_{m+1}) \mid F^y_{m+1}]$$

which gives

$$\hat{\theta}_{m+1|m+1} = \alpha_{m-3}\hat{\theta}_m = \alpha_{m-3}\hat{X}_m \,.$$

Next

$$\hat{\theta}_{m+2} = \mathrm{E}\left[\theta_{m+2} \mid F^y_{m+1}\right] = \mathrm{E}\left[(\alpha_{m-2}\theta_{m+1} + \beta_{m-2}u_{m+2}) \mid F^y_{m+1}\right]$$

$$= \alpha_{m-2}\mathrm{E}\left[\theta_{m+1} \mid F^y_{m+2}\right] = \alpha_{m-2}\hat{\theta}_{m+1}$$

and $\hat{\theta}_{m+1} = \alpha_{m-3}\hat{\theta}_m$ which combines to give

$\hat{\theta}_{m+2} = \alpha_{m-2}\alpha_{m-3}\hat{\theta}_m = \alpha_{m-2}\alpha_{m-3}\hat{X}_m$, $\hat{\theta}_m = \hat{X}_m$ since the data has been observed up to time $t = m$.

Hence,

$$\widetilde{X}_{m+1|m+2} = \alpha_{m-3}\hat{X}_m + \widetilde{k}_{m+2}\left(X_{m+2} - \alpha_{m-2}\alpha_{m-3}\hat{X}_m\right) \quad \text{(3.4.25)}.$$

but,

$$\widetilde{k}_{m+2} = \frac{\widetilde{\Sigma}_{m+2}}{\Sigma_{m+2}} = \frac{\alpha_{m-2}\beta^2_{m-3}}{\beta^2_{m-2} + \alpha^2_{m-2}\beta^2_{m-3}} \quad \text{(3.4.26)},$$

replacing for $\widetilde{k}_{m+2}$ in (3.4.25) we obtain

$$\widetilde{X}_{m+1|m+2} = \alpha_{m-3}\hat{X}_m + \frac{\alpha_{m-2}\beta^2_{m-3}}{\beta^2_{m-2} + \alpha^2_{m-2}\beta^2_{m-3}}\left(X_{m+2} - \alpha_{m-2}\alpha_{m-3}\hat{X}_m\right).$$

Simplifying this gives

$$\widetilde{X}_{m+1|m+2} = \frac{\alpha_{m-3}\left(\beta^2_{m-2}\right)\hat{X}_m + \alpha_{m-2}\beta^2_{m-3}X_{m+2}}{\beta^2_{m-2} + \alpha^2_{m-2}\beta^2_{m-3}} \quad \text{(3.4.27)}.$$

3.5: Three consecutive missing observations.

(a) The estimate of *mth* observation based on X_{m+3}.

The modified state space model is given by

$$\left.\begin{aligned} \theta_{t+1} &= \alpha_{t-3}\theta_t + \beta_{t-3}u_{t+1} \\ X_t &= A_{t-1}\theta_t \\ y_t &= A_{t-1}\theta_t + B_{t-1}v_t \end{aligned}\right\} \quad \text{(3.5.1)}.$$

In this case we have three consecutive missing observations X_m, X_{m+1}, and X_{m+2} for which we need to estimate X_m before we estimate the remaining two observations respectively. The initial observation set is $X'_N = (X_1, X_2, ..., X_{m-1}, X_{m+3}, ...X_N)$. The extended observed series is $Y = (X_1, X_2, ..., X_{m-1}, v_m, X_{m+3}, ..., X_N)$ where v_m is a normal random variable replacing the missing observation.

Using theorems 3.1.1 and 3.1.2 we have the modified set of equations

$$k_t = \alpha_{t-3}\Sigma_t A_{t-1}\left[A_{t-1}^2\Sigma_t + B_{t-1}^2\right]^+ \text{.......}(3.5.2)$$

and $\Sigma_{t+1} = \beta_{t-3}^2 + (\alpha_{t-3} - k_t A_{t-1})^2\Sigma_t + k_t^2 B_{t-1}^2$.......(3.5.3).

In this case $A_{m-1} = A_m = A_{m+1} = 0, B_{m-1} = B_m = B_{m+1} = 1; A_t = 1, B_t = 0; t \neq m, m+1, m+2$.

This implies that,

$$k_t = \alpha_{t-3} \text{ and } \Sigma_{t+1} = \beta_{t-3}^2 \text{.......}(3.5.4).$$

Also

$$k_m = k_{m+1} = k_{m+2} = 0.$$

Next we have

$$\Sigma_{m+1} = \beta_{m-3}^2 + \alpha_{m-3}^2\Sigma_m \text{.......}(3.5.5)$$
$$\Sigma_{m+2} = \beta_{m-2}^2 + \alpha_{m-2}^2\Sigma_{m+1} \text{.......}(3.5.6)$$
$$\Sigma_{m+3} = \beta_{m-1}^2 + \alpha_{m-1}^2\Sigma_{m+2} \text{.......}(3.5.7)$$

we combine (3.5.5), (3.5.6) and (3.5.7) to obtain

$$\Sigma_{m+2} = \beta_{m-2}^2 + \alpha_{m-2}^2\left(\beta_{m-3}^2 + \alpha_{m-3}^2\Sigma_m\right)$$
$$\Sigma_{m+2} = \beta_{m-1}^2 + \alpha_{m-2}^2\beta_{m-3}^2 + \alpha_{m-2}^2\alpha_{m-3}^2\Sigma_m.$$

Then,

$$\Sigma_{m+3} = \beta_{m-1}^2 + \alpha_{m-1}^2\left(\beta_{m-2}^2 + \alpha_{m-2}^2\beta_{m-3}^2 + \alpha_{m-2}^2\alpha_{m-3}^2\Sigma_m\right)$$
$$= \beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\beta_{m-3}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\alpha_{m-3}^2\Sigma_m \text{.......}(3.5.8)$$

but from (3.5.3) $\Sigma_{t+1} = \beta_{t-3}^2$ so that at $t = m$ we have

$$\Sigma_{m+1} = \beta_{m-3}^2$$
$$\Sigma_m = \beta_{m-4}^2.$$

Hence, (3.5.8) becomes

$$\Sigma_{m+3} = \beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\beta_{m-3} + \alpha_{m-1}^2\alpha_{m-2}^2\alpha_{m-3}^2\beta_{m-4}^2 \ldots\ldots\ldots\ldots\ldots(3.5.9).$$

From theorem 3.1.2

$$\widetilde{k}_t = \widetilde{\Sigma}_t A_{t-1}\left[A_{t-1}^2\Sigma_t + B_{t-1}^2\right]^+$$
$$\widetilde{\Sigma}_{t+1} = \widetilde{\Sigma}_t(\alpha_{t-3} - k_t A_{t-1})$$

so that we get

$$\widetilde{k}_t = \frac{\widetilde{\Sigma}_t}{\Sigma_t}$$

and $\widetilde{\Sigma}_{t+1} = 0$ since $k_t = \alpha_{t-3}$.

$$\left.\begin{aligned} \widetilde{\Sigma}_{m+1} &= \alpha_{m-3}\widetilde{\Sigma}_m \\ \widetilde{\Sigma}_{m+2} &= \alpha_{m-2}\widetilde{\Sigma}_{m+1} \\ \widetilde{\Sigma}_{m+3} &= \alpha_{m-1}\Sigma_{m+2} \end{aligned}\right\}\ldots\ldots\ldots\ldots\ldots(3.5.10a).$$

Combining (3.5.10a) we obtain

$$\widetilde{\Sigma}_{m+3} = \alpha_{m-1}\alpha_{m-2}\widetilde{\Sigma}_{m+1}$$
$$\widetilde{\Sigma}_{m+3} = \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\widetilde{\Sigma}_m \ldots\ldots\ldots\ldots\ldots(3.5.10b)$$

but $\widetilde{\Sigma}_m = \Sigma_m = \beta_{m-4}^2$

Hence, (3.5.10b) becomes

$$\widetilde{\Sigma}_{m+3} = \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\beta_{m-4}^2 \ldots\ldots\ldots\ldots\ldots(3.5.11).$$

From (3.3.10)

$\widetilde{\theta}_{m|t} = \widetilde{\theta}_{m|t-1} + \widetilde{k}_t\left(y_t - \hat{\theta}_t\right)$ and at $t = m+3$, we have

$$\widetilde{\theta}_{m|m+3} = \widetilde{\theta}_{m|m+2} + \widetilde{k}_{m+3}\left(y_{m+3} - \hat{\theta}_{m+3}\right) \ldots\ldots\ldots\ldots\ldots(3.5.12).$$

To find $\widetilde{\theta}_{m|m+2}$ we have

$$\widetilde{\theta}_{m|m+2} = \mathrm{E}\left[\theta_m \mid F_{m+2}^y\right] = \mathrm{E}\left[(\alpha_{m-4}\theta_{m-1} + \beta_{m-4}u_m) \mid F_{m+2}^y\right]$$

$$= \alpha_{m-4}\mathrm{E}\left[\theta_{m-1} \mid F^{y}_{m+2}\right] = \alpha_{m-4}\hat{\theta}_{m-1} = \alpha_{m-4}X_{m-1}.$$

Next,

$$\hat{\theta}_{m+3} = \mathrm{E}\left[\theta_{m+3} \mid F^{y}_{m+2}\right] = \mathrm{E}\left[\left(\alpha_{m-1}\theta_{m+2} + \beta_{m-1}u_{m+3}\right) \mid F^{y}_{m+2}\right] = \alpha_{m-1}\hat{\theta}_{m+2},$$

$\hat{\theta}_{m+2} = \alpha_{m-2}\hat{\theta}_{m+1}$, $\hat{\theta}_{m+1} = \alpha_{m-3}\hat{\theta}_{m}$ and $\hat{\theta}_{m} = \alpha_{m-4}\hat{\theta}_{m-1}$ which combines to give

$$\begin{aligned}\hat{\theta}_{m+3} &= \alpha_{m-1}\left(\alpha_{m-2}\hat{\theta}_{m+1}\right)\\ &= \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\hat{\theta}_{m}\\ &= \alpha_{m-}\alpha_{m-2}\alpha_{m-3}\alpha_{m-4}\hat{\theta}_{m-1}.\end{aligned}$$

Also we let $\hat{\theta}_{m-1} = X_{m-1}$ so that

$$\hat{\theta}_{m+3} = \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\alpha_{m-4}X_{m-1} \ldots\ldots(3.5.13)$$

$$\widetilde{k}_{m+3} = \frac{\widetilde{\Sigma}_{m+3}}{\Sigma_{m+3}} = \frac{\alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\beta^2_{m-4}}{\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2} + \alpha^2_{m-1}\alpha^2_{m-2}\beta^2_{m-3} + \alpha^2_{m-1}\alpha^2_{m-2}\alpha^2_{m-3}\beta^2_{m-4}} \ldots\ldots(3.5.14).$$

Hence (3.5.12) becomes

$$\widetilde{X}_{m|m+3} = \alpha_{m-4}X_{m-1} + \frac{\alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\beta^2_{m-4}}{\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2} + \alpha^2_{m-1}\alpha^2_{m-2}\beta^2_{m-3} + \alpha^2_{m-1}\alpha^2_{m-2}\alpha^2_{m-3}\beta^2_{m-4}}\begin{pmatrix} X_{m+3} - \\ \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\alpha_{m-4}X_{m-1}\end{pmatrix}.$$

Simplifying this gives

$$\widetilde{X}_{m|m+3} = \frac{\alpha_{m-4}\left(\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2} + \alpha^2_{m-1}\alpha^2_{m-2}\beta^2_{m-3}\right)X_{m-1} + \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\beta^2_{m-4}X_{m+3}}{\beta^2_{m-1} + \alpha^2_{m-1}\beta^2_{m-2} + \alpha^2_{m-1}\alpha^2_{m-2}\beta^3_{m-3} + \alpha^2_{m-1}\alpha^2_{m-2}\alpha^2_{m-3}\beta^2_{m-4}} \ldots\ldots(3.5.15).$$

(b) The estimate of $(m+1)th$ observation based on X_{m+3}

This is treated as two missing observations as in section 3.4. the modified state space representations are;

$$\left.\begin{aligned}\theta_{t+1} &= \alpha_{t-4}\theta_t + \beta_{t-4}u_{t+1}\\ X_t &= A_{t-1}\theta_t\\ y_t &= A_{t-1}\theta_t + B_{t-1}v_t\end{aligned}\right\} \ldots\ldots(3.5.16).$$

The initial observation is $X'_N = \left(X_1, X_2, \ldots, X_{m-1}, \hat{X}_m, X_{m+3}, \ldots, X_N\right)$, we see that X_{m+1} and X_{m+2} are missing hence, we need to estimate X_{m+1}. Then the extended observed series is $Y=\left(X_1,X_2,\ldots,X_{m-1},\hat{X}_m,v_{m+1},X_{m+3},\ldots,X_N\right)$, here $\hat{X}_m$ is the estimate of X_m base on X_{m+3} and v_{m+1} is a normal random variable replacing the missing observation.

Using theorem 3.1.1 and 3.1.2 as before we have

$$A_m = A_{m+1} = 0, B_m = B_{m+1} = 1; A_t = 1, B_t = 0; t \neq m+1, m+2$$

$k_t = \alpha_{t-4}$ and

$$\Sigma_{t+1} = \beta_{t-4}^2 \qquad (3.5.17).$$

Also $k_{m+1} = k_{m+2} = 0$.

Next we have

$$\Sigma_{m+2} = \beta_{m-2}^2 + \alpha_{m-3}^2 \Sigma_{m+1}$$
$$\Sigma_{m+3} = \beta_{m-2}^2 + \alpha_{m-2}^2 \Sigma_{m+2}.$$

Hence

$$\Sigma_{m+3} = \beta_{m-2}^2 + \alpha_{m-2}^2\left(\beta_{m-3}^2 + \alpha_{m-3}^2 \Sigma_{m+1}\right)$$
$$= \beta_{m-2}^2 + \alpha_{m-2}^2 \beta_{m-3}^2 + \alpha_{m-2}^2 \alpha_{m-3}^2 \Sigma_{m+1} \qquad (3.5.18).$$

From (3.5.17) at $t = m$ $\Sigma_{m+1} = \beta_{m-4}^2$

hence (3.5.18) becomes

$$\Sigma_{m+3} = \beta_{m-2}^2 + \alpha_{m-2}^2 \beta_{m-3}^2 + \alpha_{m-2}^2 \alpha_{m-3}^2 \beta_{m-4}^2 \qquad (3.5.19)$$

and from theorem 3.1.2 we have

$$\tilde{k}_t = \frac{\tilde{\Sigma}_t}{\Sigma_t}$$

$$\tilde{\Sigma}_{t+1} = 0.$$

Also

$$\tilde{\Sigma}_{m+2} = \alpha_{m-3} \tilde{\Sigma}_{m+1}$$
$$\tilde{\Sigma}_{m+3} = \alpha_{m-2} \tilde{\Sigma}_{m+2}$$

so that,

$$\tilde{\Sigma}_{m+3} = \alpha_{m-2} \alpha_{m-3} \tilde{\Sigma}_{m+1} = \alpha_{m-2} \alpha_{m-3} \Sigma_{m+1},$$

where $\Sigma_{m+1} = \beta_{m-4}^2$.

Hence,

$$\widetilde{\Sigma}_{m+3} = \alpha_{m-2}\alpha_{m-3}\beta_{m-4}^2 \ldots\ldots(3.5.20).$$

Also,

$$\widetilde{k}_t = \frac{\widetilde{\Sigma}_t}{\Sigma_t}$$

so that ,

$$\widetilde{k}_{m+3} = \frac{\widetilde{\Sigma}_{m+3}}{\Sigma_{m+3}}$$

where from equations (3.5.19) and (3.5.20) we get

$$\widetilde{k}_{m+3} = \frac{\alpha_{m-2}\alpha_{m-3}\beta_{m-4}^2}{\beta_{m-2}^2 + \alpha_{m-2}^2\beta_{m-3}^2 + \alpha_{m-2}^2\alpha_{m-3}^2\beta_{m-4}^2} \ldots\ldots(3.5.21).$$

Then

$$\widetilde{\theta}_{t_j|t} = \widetilde{\theta}_{t_j|t-1} + \widetilde{k}_t\left(y_t - A_{t-1}\hat{\theta}_t\right), t > t_j \ldots\ldots(3.5.22)$$

but $A_t = 1$ and at $t_j = m+1$ and $t = m+3$

so that,

$$\widetilde{\theta}_{m+1|m+3} = \widetilde{\theta}_{m+1|m+2} + \widetilde{k}_{m+3}\left(y_{m+3} - \hat{\theta}_{m+1}\right) \ldots\ldots(3.5.23).$$

We have

$$\widetilde{\theta}_{m+1|m+2} = \mathrm{E}\left[\theta_{m+1} \mid F_{m+2}^y\right] = \mathrm{E}\left[\left(\alpha_{m-4}\theta_m + \beta_{m-4}u_{m+1}\right) \mid F_{m+2}^y\right]$$

$$= \alpha_{m-4}\mathrm{E}\left(\theta_m \mid F_{m+2}^y\right)$$

$$= \alpha_{m-4}\hat{\theta}_m,$$

but $\hat{\theta}_m = \hat{X}_m$ since we have made an observation up to time $t = m$.

Hence, $\widetilde{\theta}_{m+1|m+2} = \alpha_{m-4}\hat{X}_m$.

Next,

$$\hat{\theta}_{m+3} = \mathrm{E}\left[\theta_{m+3} \mid F_{m+2}^y\right] = \mathrm{E}\left[\left(\alpha_{m-2}\theta_{m+2} + \beta_{m-2}u_{m+3}\right) \mid F_{m+2}^y\right] = \alpha_{m-2}\hat{\theta}_{m+2}$$

$\theta_{m+2} = \alpha_{m-3}\hat{\theta}_{m+1}$, $\hat{\theta}_{m+1} = \alpha_{m-4}\hat{\theta}_m = \alpha_{m-4}\hat{X}_m$ and these combines to give

$$\hat{\theta}_{m+3} = \alpha_{m-2}\alpha_{m-3}\hat{\theta}_{m+1} = \alpha_{m-2}\alpha_{m-3}\alpha_{m-4}\hat{\theta}_m,$$

$$\hat{\theta}_{m+3} = \alpha_{m-2}\alpha_{m-3}\alpha_{m-4}\hat{X}_m \quad \text{.......... (3.5.24)}$$

where $\hat{X}_m$ is the estimate of X_m obtained earlier in the same section. Hence (3.5.23) becomes,

$$\tilde{X}_{m+1|m+3} = \alpha_{m-4}\hat{X}_m + \tilde{k}_{m+3}\left(X_{m+3} - \alpha_{m-2}\alpha_{m-3}\alpha_{m-4}\hat{X}_m\right)$$

$$= \alpha_{m-4}\hat{X}_m + \frac{\alpha_{m-2}\alpha_{m-3}\beta^2_{m-4}}{\beta^2_{m-2} + \alpha^2_{m-2}\beta^2_{m-3} + \alpha^2_{m-2}\alpha^2_{m-3}\beta^2_{m-4}}\left(X_{m+3} - \alpha_{m-2}\alpha_{m-3}\alpha_{m-4}\hat{X}_m\right)$$

Simplifying this gives

$$\tilde{X}_{m+1|m+3} = \frac{\alpha_{m-4}\left(\beta^2_{m-2} + \alpha^2_{m-2}\beta^2_{m-3}\right)\hat{X}_m + \alpha_{m-2}\alpha_{m-3}\beta^2_{m-4}X_{m+3}}{\beta^2_{m-2} + \alpha^2_{m-2}\beta^2_{m-3} + \alpha^2_{m-2}\alpha^2_{m-3}\beta^2_{m-4}} \quad \text{.......... (3.5.25).}$$

(c) The estimate of $(m+2)th$ observation based on X_{m+3}.

This is treated as one missing observation since X_m and X_{m+1} have been estimated. Then we have the new set of observations as $X'_N = \left(X_1, X_2, \ldots, X_{m-1}, \hat{X}_m, \hat{X}_{m+1}, X_{m+3}, \ldots, X_N\right)$, the new state space representations are;

$$\left.\begin{aligned} \theta_{t+1} &= \alpha_{t-5}\theta_t + \beta_{t-5}u_{t+1} \\ X_t &= A_{t-1}\theta_t \\ y_t &= A_{t-1}\theta_t + B_{t-1}v_t \end{aligned}\right\} \quad \text{.......... (5.5.26).}$$

Then $X'_N = \left(X_1, X_2, \ldots X_{m-1}, \hat{X}_m, \hat{X}_{m+1}, v_{m+2}, X_{m+3}, \ldots, X_N\right)$ is the extended observed series. Here $\hat{X}_m$ and $\hat{X}_{m+1}$ are the estimates of X_m and X_{m+1} from the previous steps. v_{m+2} is a normal random variable replacing the missing observation. In this case $A_{m+2} = 0, B_{m+2} = 1; A_t = 1, B_t = 0; t \neq m+2$

Again from theorems 3.1.1 and 3.1.2 we have

$$k_t = \alpha_{t-5}, \Sigma_{t+1} = \beta^2_{t-5} \quad \text{.......... (3.5.27)}$$

$$k_{m+2} = 0 \;, \Sigma_{m+3} = \beta_{m-3}^2 + \alpha_{m-3}^2 \Sigma_{m+2} \qquad (3.5.28).$$

Setting $t = m$ in (3.5.28) we obtain

$$\Sigma_{m+1} = \beta_{m-5}^2$$
$$\Sigma_{m+2} = \beta_{m-4}^2.$$

Hence, (3.5.28) now becomes

$$\Sigma_{m+3} = \beta_{m-3}^2 + \alpha_{m-3}^2 \beta_{m-4}^2 \qquad (3.5.29).$$

Also

$$\widetilde{k}_t = \frac{\widetilde{\Sigma}_t}{\Sigma_t}$$

so that,

$$\widetilde{k}_{m+3} = \frac{\widetilde{\Sigma}_{m+3}}{\Sigma_{m+3}} \qquad (3.5.30),$$

$$\widetilde{k}_{m+2} = 0 \,, \; \widetilde{\Sigma}_{m+3} = \alpha_{m-3} \widetilde{\Sigma}_{m+2} = \alpha_{m-3} \Sigma_{m+2} \,.$$

Thus,

$$\widetilde{\Sigma}_{m+3} = \alpha_{m-3} \beta_{m-4} \qquad (3.5.31)$$

so that,

$$\widetilde{k}_{m+3} = \frac{\alpha_{m-3} \beta_{m-4}^2}{\beta_{m-3}^2 + \alpha_{m-3}^2 \beta_{m-4}^2} \qquad (3.5.32).$$

Then ,

$$\widetilde{\theta}_{t_j|t} = \widetilde{\theta}_{t_j|t-1} + \widetilde{k}_t \left(y_t - A_{t-1} \hat{\theta}_t \right) t > t_j \qquad (3.5.33)$$

but $A_t = 1$ at $t_j = m+2$ and $t = m+3$ so that (3.5.33) becomes

$$\widetilde{\theta}_{m+2|m+3} = \widetilde{\theta}_{m+2|m+2} + \widetilde{k}_{m+3} \left(y_{m+3} - \hat{\theta}_{m+3} \right) \qquad (3.5.34).$$

Thus ,

$$\hat{\theta}_{m+2|m+2} = \mathrm{E}\left[\theta_{m+2} \mid F_{m+2}^y \right] = \mathrm{E}\left[\left(\alpha_{m-4} \theta_{m+1} + \beta_{m-4} u_{m+2} \right) \mid F_{m+2}^y \right]$$
$$= \alpha_{m-4} \mathrm{E}\left[\theta_{m+1} \mid F_{m+2}^y \right] = \alpha_{m-4} \hat{\theta}_{m+1}$$

Since we have made an observation up to time $t = m+1$ we have

$$\hat{\theta}_{m+1} = \hat{X}_{m+1}$$

so that,

$$\hat{\theta}_{m+2|m+2} = \alpha_{m-4}\hat{X}_{m+1}.$$

Next,

$$\hat{\theta}_{m+3} = \mathrm{E}\left[\theta_{m+3} \mid F^{y}_{m+2}\right] = \mathrm{E}\left[\left(\alpha_{m-3}\theta_{m+2} + \beta_{m-3}u_{m+3}\right) \mid F^{y}_{m+2}\right]$$

$$= \alpha_{m-3}\mathrm{E}\left[\theta_{m+2} \mid F^{y}_{m+2}\right] = \alpha_{m-3}\hat{\theta}_{m+2}$$

that is,

$\hat{\theta}_{m+3} = \alpha_{m-3}\hat{\theta}_{m+2}$ and $\hat{\theta}_{m+2} = \alpha_{m-4}\hat{\theta}_{m+1}$ combines to give

$$\hat{\theta}_{m+3} = \alpha_{m-3}\left(\alpha_{m-4}\hat{\theta}_{m+1}\right) = \alpha_{m-3}\alpha_{m-4}\hat{X}_{m+2}.$$

Hence (3.5.34) gives

$$\hat{X}_{m+2/m+3} = \alpha_{m-4}\hat{X}_{m+1} + \hat{k}_{m+3}\left(X_{m+3} - \alpha_{m-3}\alpha_{m-4}\hat{X}_{m+1}\right).$$

Substituting for $\hat{k}_{m+3}$ we have

$$\widetilde{X}_{m+2|m+3} = \alpha_{m-4}\hat{X}_{m+1} + \frac{\alpha_{m-3}\beta^{2}_{m-4}}{\beta^{2}_{m-3} + \alpha^{2}_{m-3}\beta^{2}_{m-4}}\left(X_{m+3} - \alpha_{m-3}\alpha_{m-4}\hat{X}_{m+1}\right)$$

which simplifies to give

$$\widetilde{X}_{m+2|m+3} = \frac{\alpha_{m-4}\left(\beta^{2}_{m-3}\right)\hat{X}_{m+1} + \alpha_{m-3}\beta^{2}_{m-4}X_{m+3}}{\beta^{2}_{m-3} + \alpha^{2}_{m-3}\beta^{2}_{m-4}} \quad \text{.....................................(3.5.35)}.$$

We now list these models in order, to obtained the general pattern of obtaining estimate of missing observation at any stage. Hence we have;

1. $$\widetilde{X}_{m|m+1} = \frac{\alpha_{m-2}\left(\beta^{2}_{m-1}\right)X_{m-1} + \alpha_{m-1}\beta^{2}_{m-2}X_{m+1}}{\beta^{2}_{m-1} + \alpha^{2}_{m-1}\beta^{2}_{m-2}}.$$

2. $$\widetilde{X}_{m|m+2} = \frac{\alpha_{m-3}\left(\beta^{2}_{m-1} + \alpha^{2}_{m-1}\beta^{2}_{m-2}\right)X_{m-1} + \alpha_{m-1}\alpha_{m-2}\beta^{2}_{m-3}X_{m+2}}{\beta^{2}_{m-1} + \alpha^{2}_{m-1}\beta^{2}_{m-2} + \alpha^{2}_{m-1}\alpha^{2}_{m-2}\beta^{2}_{m-3}}$$

$$\widetilde{X}_{m+1|m+2} = \frac{\alpha_{m-3}\left(\beta^{2}_{m-2}\right)\hat{X}_{m} + \alpha_{m-2}\beta^{2}_{m-3}X_{m+2}}{\beta^{2}_{m-2} + \alpha^{2}_{m-2}\beta^{2}_{m-3}}.$$

3. $$\widetilde{X}_{m|m+3} = \frac{\alpha_{m-4}\left(\beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\beta_{m-3}^2\right)X_{m-1} + \alpha_{m-1}\alpha_{m-2}\alpha_{m-3}\beta_{m-4}^2 X_{m+3}}{\beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\beta_{m-3}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\alpha_{m-3}^2\beta_{m-4}^2}$$

$$\widetilde{X}_{m+1|m+3} = \frac{\alpha_{m-4}\left(\beta_{m-2}^2 + \alpha_{m-2}^2\beta_{m-3}^2\right)\hat{X}_m + \alpha_{m-2}\alpha_{m-3}\beta_{m-4}^2 X_{m+3}}{\beta_{m-2}^2 + \alpha_{m-2}^2\beta_{m-3}^2 + \alpha_{m-2}^2\alpha_{m-3}^2\beta_{m-4}^2}$$

$$\widetilde{X}_{m+2|m+3} = \frac{\alpha_{m-4}\left(\beta_{m-3}^2\right)\hat{X}_{m+1} + \alpha_{m-3}\beta_{m-4}^2 X_{m+3}}{\beta_{m-3}^2 + \alpha_{m-3}^2\beta_{m-4}^2}.$$

.

.

.

s. $$\widetilde{X}_{m|m+s} = \frac{\left[\begin{array}{l}\alpha_{m-(s+1)}\left(\beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\beta_{m-3}^2 + \ldots + \alpha_{m-1}^2\alpha_{m-2}^2\ldots\alpha_{m-(s+1)}^2\beta_{m-s}^2\right)X_{m-1} + \\ \alpha_{m-1}\alpha_{m-2}\ldots\alpha_{m-s}\beta_{m-(s+1)}^2 X_{m+s}\end{array}\right]}{\beta_{m-1}^2 + \alpha_{m-1}^2\beta_{m-2}^2 + \alpha_{m-1}^2\alpha_{m-2}^2\beta_{m-3}^2 + \ldots + \alpha_{m-1}^2\alpha_{m-2}^2\ldots\alpha_{m-s}^2\beta_{m-(s+1)}}$$

.

.

.

$$\widetilde{X}_{m+(s-1)|m+s} = \frac{\alpha_{m-(s+1)}\left(\beta_{m-s}^2\right)\hat{X}_{m+(s-2)} + \alpha_{m-s}\beta_{m-(s+1)}^2 X_{m+s}}{\beta_{m-s}^2 + \alpha_{m-s}^2\beta_{m-(s+1)}^2}$$

The estimate of $\widetilde{X}_{m+1|m+2}$ are the same as those obtained by Abraham (1981) and Miller and Ferreiro (1984). The approach in (3.3.2) presented in section 3.3 can only handle some of the nonlinear models mentioned earlier. Before we consider a more general approach in section 3.6 we need to give an example of the model above.

Lets give an example of AR(1) model with a constant conditional variance , $\beta_t^2 = \sigma^2$ and $\alpha_t = \phi$ (See Abraham and Thavaneswaran (1991)).

In this case the estimates of the missing observations $X_m, X_{m+1}, X_{m+2}, \ldots, X_{m+(s-1)}$ for an AR(1) model up to s missing values will be as follows;

1. $$\widetilde{X}_{m|m+1} = \frac{\phi X_{m-1} + \phi X_{m+1}}{1+\phi^2}$$

2. $$\widetilde{X}_{m|m+2} = \frac{\phi\left(1+\phi^2\right)X_{m-1} + \phi^2 X_{m+2}}{1+\phi^2+\phi^4}$$

$$\widetilde{X}_{m+1|m+2} = \frac{\phi \hat{X}_m + \phi X_{m+2}}{1+\phi^2}$$

3. $$\widetilde{X}_{m|m+3} = \frac{\phi\left(1+\phi^2+\phi^4\right)X_{m-1} + \phi^3 X_{m+3}}{1+\phi^2+\phi^4+\phi^6}$$

$$\widetilde{X}_{m+1|m+3} = \frac{\phi\left(1+\phi^2\right)\hat{X}_m + \phi^2 X_{m+3}}{1+\phi^2+\phi^4}$$

$$X_{m+2|m+3} = \frac{\phi \hat{X}_{m+1} + \phi X_{m+3}}{1+\phi^2}$$

.

.

.

s. $$\widetilde{X}_{m|m+s} = \frac{\phi\left(1+\phi^2+\phi^4+\ldots+\phi^{2(s-1)}\right)X_{m-s} + \phi^s X_{m+s}}{1+\phi^2+\phi^4+\ldots+\phi^{2s}}$$

.

.

.

$$\widetilde{X}_{m+(s-1)|m+s} = \frac{\phi X_{m+(s-2)} + \phi X_{m+s}}{1+\phi^2}.$$

Hence the general form of these sequence of estimates is given by;

$$\widetilde{X}_{m+j|m+i} = \frac{\phi\left(\sum_{r=0}^{i-j-1}\phi^{2r}\right)X_{m+j-1} + \phi^{i-j}X_{m+i}}{\sum_{r=0}^{i-j}\phi^{2r}} \quad \text{.................................(3.5.37)}$$

where $j = 0,1,2,\ldots,s-1$ and $i = 1,2,3,\ldots,s$ for $i > j$.

3.6 Optimal estimation of missing observation.

Following Godambe (1985), the optimal estimation of parameters in adaptive as well as non adaptive nonlinear time series has been discussed in Thavaneswaran and Abraham (1988). In this section, we briefly describe the estimation of missing observations considering them as parameters.

Let $y_1, y_2, \ldots, y_n$ be the observed time series with $y_m (1 < m < n)$ missing and the parameters θ_t be those known from the generalized model 3.1.1.

The following theorem gives the extended version of Godambe`s theorem (1985) to the multi-parameter multi-valued stochastic process.

Theorem 3.6.1. The optimal estimating function L is given by;

$$g^o = \sum_{t=2}^{n} a^o_{t_i-1} h_{t_i} \quad \text{...(3.6.1)}$$

where L is the class of unbiased estimating function g , and $a^o_{t_i-1} = \mathrm{E}\left[\frac{\partial h_{t_i}}{\partial \theta} \setminus F^y_{t_i-1}\right]\left(\mathrm{E}\left[h_{t_i} h^T_{t_i} \mid F^y_{t_i-1}\right]\right)^+$ provided the inverse $\left(\mathrm{E}\left[h_{t_i} h^T_{t_i} \mid F^y_{t_i-1}\right]\right)^+$ exists.

We now give an example to demonstrate the use of this theorem.

3.6.1. Heteroscedasticity model with lagged dependent variables.

White (1980) considered a general specification of a regressive model by allowing the exogenous regressors to be stochastic. The important case when some of the exogenous variables are lagged values of the endogenous variables have been treated in Nicholls and Pagan (1983).

They considered the model,

$$y_t = \sum_{j=1}^{p} \theta_j y_{t-j} + e_t \quad \text{.......(3.6.2)}$$

where the error sequence $\{e_t\}$ satisfies

$$\mathrm{E}(e_t \mid F_{t-1}^y) = 0, \mathrm{E}(e_t^2 \mid F_{t-1}^y) = \sigma_t^2 .$$

Let ,

$$h_t = y_t - \mathrm{E}(y_t \mid F_{t-1}^2) = y_t - \theta^T X_{t-1}, \text{ and } \mathrm{E}(h_t^2 \mid F_{t-1}^y) = \sigma_t^2$$

where $\theta^T = (\theta_1, \theta_2, ..., \theta_p)$, $X_{t-1}^T = (y_{t-1}, y_{t-2}, ..., y_{t-p})$ and θ^T denotes the transpose of θ. Then the optimal estimating function $g_{n,\theta}^o$ is given by

$$\hat{\theta}_n = \left(\sum_{t=p+1}^{n} X_{t-1} X_{t-1}^T \Big/ \sigma_t^2 \right)^{-1} \left(\sum_{t=p+1}^{n} X_{t-1} y_t \Big/ \sigma_{t-1}^2 \right),$$

the least-square estimate of θ is given by

$$\tilde{\theta}_n = \left(\sum_{t=p+1}^{n} X_{t-1} X_{t-1}^T \right)^{-1} \left(\sum_{t=p+1}^{n} X_{t-1} y_t \right),$$

(See Thavaneswaran and Abraham (1991)) and is independent of the conditional variance σ_t^2. Moreover , if e_t is assume to be normal then $\hat{\theta}_n$ become the likelihood estimate (m.l.e) and is more efficient than $\tilde{\theta}_n$. We note here that for any series y_t can be written as

$$y_t = \mathrm{E}(y_t \mid F_{t-1}^y) + y_t - \mathrm{E}(y_t \mid F_{t-1}^y) = \mathrm{E}(y_t \mid F_{t-1}^y) + h_t .$$

If $\mathrm{E}(y_t \mid F_{t-1}^y)$ and $Var(y_t \mid F_{t-1}^y)$ can be written as functions of finite number of parameters, then (3.6.2) correspond to a heteroscedastic model.

Example 3.6.1. (ARCH model). Consider the model $y_{t+1} = \alpha_t y_t + \beta_t u_{t+1}$ where $\alpha_t = \alpha(t, F_t^y, \theta), \beta_t = \beta(t, F_t^y, \theta)$ and $\{u_t\}$ are a sequence of i.i.d. random variables having mean zero and finite variance σ^2. Here it should be noted that we are not making any distributional assumption on the errors. It can be shown that the optimal estimate of y_m satisfies

$$\sum_{t=1}^{n} a_t^*(y_{t+1} - \alpha_t y_t) = 0 \qquad (3.6.3)$$

where

$$a_t^* = \frac{\mathrm{E}\left[\dfrac{\partial}{\partial y_m}(y_{t+1} - \alpha_t y_t) \mid F_t^y\right]}{\beta_t^2 \sigma^2}$$

(See Abrahams and Thavaneswaran (1991))

In special case of an AR(1) model,

$$y_{t+1} = \phi y_t + u_{t+1}$$

we have

$$u_{t+1} = h_t = y_{t+1} - \phi y_t$$

so that

$$a_t^* = \mathrm{E}\left[\frac{\partial}{\partial y_m}(y_{t+1} - \phi y_t) \mid F_t^y\right]\left[\mathrm{E}\left(u_{t+1}^2 \mid F_t^y\right)\right]^{-1}$$ from equation (3.6.1).

Thus,

$$g_n^o = \sum_{t=1}^{n} \mathrm{E}\left[\frac{\partial}{\partial y_m}(y_{t+1} - \phi y_t) \mid F_t^y\right]\left[\mathrm{E}\left(u_{t+1}^2 \mid F_t^y\right)\right]^{-1}(y_{t+1} - \phi y_t)$$

Hence,

$$g_{n,y_m}^o = \frac{(y_m - \phi y_{m-1})}{\sigma^2} - \frac{\phi(y_{m+1} - \phi y_m)}{\sigma^2} = 0$$

Therefore the estimate of y_m turns out to be the solution of the equation

$$(\hat{y}_m - \phi y_{m-1}) - \phi(y_{m+1} - \phi \hat{y}_m) = 0$$

which is given by

$$\hat{y}_m(op) = \frac{\phi(y_{m-1} + y_{m+1})}{1 + \phi^2} \qquad (3.6.4)$$

This is the same as what we obtained in the Section 3.5.

Example 3.6.2. (RCA model). Let $y_t = (\phi + \beta_t)y_{t-1} + u_t$ where $\{u_t\}$ and $\{\beta_t\}$ are zero mean square integrable independent sequence and $Var(u_t) = \sigma_u^2$ and $Var(\beta_t) = \sigma_\beta^2; \beta_t$ is independent of $\{u_t\}$ and $\{y_{t-1}\}$. Then the optimal estimate of y_m (treated as a parameter) can be given as a solution of the nonlinear equation.

Let $u_t + \beta_t y_{t-1} = y_t - \phi y_{t-1}$ then $h_t = u_t + \beta_t y_{t-1} = y_t - \phi y_{t-1}$ and the optimal estimating function is $g^o = \sum_{t=1}^{n} a_{t-1}^* h_t$ where

$$a_{t-1}^* = \mathrm{E}\left[\frac{\partial h_t}{\partial y_m} \mid F_{t-1}^y\right]\left[\mathrm{E}\left(h_t h_t^T \mid F_{t-1}^y\right)\right]^{-1}$$

$$= \mathrm{E}\left[\frac{\partial}{\partial y_m}(y_t - \phi y_{t-1}) \mid F_{t-1}^y\right]\left[\mathrm{E}(u_t + \beta_t y_{t-1})^2 \mid F_{t-1}^y\right]^{-1}$$

$$g_n^o = \sum_{t=1}^{n} \mathrm{E}\left[\frac{\partial}{\partial y_m}(y_t - \phi y_{t-1}) \mid F_{t-1}^y\right]\left[\mathrm{E}(u_t + \beta_t y_{t-1})^2 \mid F_{t-1}^y\right]^{-1}(y_t - \phi y_{t-1})$$

$$g_{n,y_m}^o = \ldots + \mathrm{E}\left[\frac{\partial}{\partial y_m}(y_{m+1} - \phi y_m) \mid F_m^y\right]\left[\mathrm{E}(u_{m+1} + \beta_{m+1} y_m)^2 \mid F_m^y\right]^{-1}(y_{m+1} - \phi y_m) +$$

$$\mathrm{E}\left[\frac{\partial}{\partial y_m}(y_m - \phi y_{m-1}) \mid F_{m-1}^y\right]\left[\mathrm{E}(u_m + \beta_m y_{m-1})^2 \mid F_{m-1}^y\right]^{-1}(y_m - \phi y_{m-1}) + \ldots\ldots$$

Hence,

$$g_{n,y_m}^o = -\phi\left[\sigma_u^2 + \sigma_\beta^2 y_m^2\right]^{-1}(y_{m+1} + \phi y_m) + \left(\sigma_m^2 + \sigma_\beta^2 y_{m-1}^2\right)^{-1}(y_m - \phi y_{m-1}) = 0$$

so that

$$\left[(\hat{y}_m - \phi y_{m-1})/\left(\sigma_u^2 + \sigma_\beta^2 y_{m-1}^2\right)\right] - \left[\phi(y_{m+1} - \phi\hat{y}_m)/\left(\sigma_u^2 + \sigma_\beta^2 y_m^2\right)\right] = 0 \ldots\ldots\ldots\ldots\ldots(3.6.5).$$

It can also be seen that the least square estimate of y_m is the solution of

$$(\hat{y}_m - \phi y_{m-1}) - \phi(y_{m+1} - \phi\hat{y}_m) = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.6.6)$$

and is given by

$$\hat{y}_m(LS) = \phi(y_{m-1} + y_{m+1}) \Big/ (1 + \phi^2).$$

This is the same as the previously obtained for an AR(1) process. However , the optimal estimate will not be the same in both cases.

This estimate depend on the conditional variance of the observed series which in turn depends on the missing value, y_m. Hence we first find the least square estimate $\hat{y}_m(LS)$ of y_m and then use it to obtain the weights $W_1 = \sigma_u^2 + \sigma_\beta^2 y_{m-1}^2$ and $W_2 = \sigma_u^2 + \sigma_\beta^2 \hat{y}_m^2(LS)$ to calculate the optimal estimate which simplifies to

$$\hat{y}_m(op) = \frac{\phi(W_2 y_{m-1} + W_1 y_{m+1})}{\phi^2 W_1 + W_2}$$

Where in this algorithms we assume that the model parameters ϕ, σ_u^2 and σ_β^2 are known.

CHAPTER FOUR

AN EMPIRICAL STUDY

4.0: Introduction.

In this chapter we carry out an empirical study to illustrate the results obtained in chapter three on simulated AR(1) data. Some values are with held and then estimated as though they were missing.

We compare the simulated AR(1) data and corresponding estimated missing values generated using formula (3.5.37) in chapter three.

Tables 4.2 to 4.7 and Tables 4.9 to 4.14 gives actual data and estimates of missing values for different values of the parameter ϕ and initial value X_0.

4.1: Estimation of missing observations on simulated AR(1) data.

AR(1) process is written as $X_t = \phi X_{t-1} + e_t$, where e_t is a purely random process which is normally distributed with mean zero and unit variance (i.e. $e_t \sim \mathrm{N}(0,1)$), ϕ is some constant given within the range $|\phi| < 1$ and X_0 is the initial observation which we choose to determine the size of the data.

Hence we have,

$$\left.\begin{aligned} X_1 &= \phi X_0 + e_1 \\ X_2 &= \phi X_1 + e_2 \\ &. \\ &. \\ &. \\ X_n &= \phi X_{n-1} + e_n \end{aligned}\right\} \qquad (4.1.1)$$

We generated AR(1) data using a computer programme. The general formula (3.5.37) in chapter three was then applied to the simulated AR(1) data at various sections of the data where missing observations where artificially created. The Tables 4.2 to 4.7 illustrates the missing values from a given position and their estimates and Tables 4.9 to 4.14 illustrate the same for different values of ϕ and X_0.

If we set $\phi = 0.86$ and $X_0 = 100$, we obtain the following AR(1) data for the first 20 values.

Table 4.1: Simulated AR(1) data when $\phi = 0.86$ and $X_0 = 100$.

t	X_t
1	86.21399
2	73.03785
3	63.71109
4	53.52442
5	47.14499
6	38.94666
7	35.22709
8	31.41615
9	25.95318
10	21.70792
11	17.5829
12	15.35208
13	12.63766
14	11.30084
15	9.111297
16	8.188405
17	6.679135
18	5.224529
19	3.283825
20	2.94902

Table 4.2: One missing value at position 5 , 10 and 15 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	86.21399	86.21399	
2	73.03785	73.03785	
3	63.71109	63.71109	
4	53.52442	53.52442	
5	47.14499	-	45.71461
6	38.94666	38.94666	
7	35.22709	35.22709	
8	31.41615	31.41615	
9	25.95318	25.95318	
10	21.70792	-	21.52278
11	17.5829	17.5829	
12	15.35208	15.35202	
13	12.63766	12.63766	
14	11.30084	11.30084	
15	9.111297	-	9.634832
16	8.188405	8.188405	
17	6.679135	6.679135	
18	5.224529	5.224529	
19	3.283825	3.283825	
20	2.94902	2.94902	

Position	Maximum percentage deviation
5	3.03%

Table 4.3: Two consecutive missing values from position 4 to 5, 11 to 12 and 15 to 16 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	86.21399	86.21399	
2	73.03785	73.03785	
3	63.71109	63.71109	
4	53.52442	-	54.28141
5	47.14499	-	46.08884
6	38.94666	38.94666	
7	35.22709	35.22709	
8	31.41615	31.41615	
9	25.95318	25.95318	
10	21.70792	21.70792	
11	17.5829	-	18.29044
12	15.35208	-	15.25982
13	12.63766	12.63766	
14	11.30084	11.30084	
15	9.111297	-	9.554145
16	8.188405	-	8.025192
17	6.679135	6.679135	
18	5.224529	5.224529	
19	3.283825	3.283825	
20	2.94902	2.94902	

Positions of missing values	Calculated χ^2 value	Table value $\chi^2_{0.05,1}$
4 - 5	0.20076	3.841
11-12	0.02872	
15-16	0.24778	

Table 4.4: Three consecutive missing values from position 5 to 7, 10 to 12 and 15 to 17 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	86.21399	86.21399	
2	73.03785	73.03785	
3	63.71109	63.71109	
4	53.52442	53.52442	
5	47.14499	-	46.53628
6	38.94666	-	40.60873
7	35.22709	-	35.60669
8	31.41615	31.41615	
9	25.95318	25.95318	
10	21.70792	-	21.95128
11	17.5829	-	18.44966
12	15.35208	-	15.36853
13	12.63766	12.63766	
14	11.30084	11.30084	
15	9.111297	-	9.492511
16	8.188405	-	7.900518
17	6.679135	-	6.488584
18	5.224529	5.224529	
19	3.283825	3.283825	
20	2.94902	2.94902	

Positions of missing values	Calculated χ^2 value	Table value $\chi^2_{0.05,2}$
5-7	0.21350	5.991
10-11	0.04547	
15-17	0.03151	

Table 4.5: Five consecutive missing values from position 5 to 9 and 15 to 19 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	86.21399	86.21399	
2	73.03785	73.03785	
3	63.71109	63.71109	
4	53.52442	53.52442	
5	47.14499	-	46.03886
6	38.94666	-	39.60257
7	35.22709	-	34.06884
8	31.41615	-	29.31157
9	25.95318	-	25.22233
10	21.70792	21.70792	
11	17.5829	17.5829	
12	15.35208	15.35202	
13	12.63766	12.63766	
14	11.30084	11.30084	
15	9.111297	-	9.48101
16	8.188405	-	7.877255
17	6.679135	-	6.453028
18	5.224529	-	5.17587
19	3.283825	-	4.016674
20	2.94902	2.94902	

Positions of missing values	Calculated χ^2 value	Table value $\chi^2_{0.05,4}$
5-9	0.21686	9.488
15-19	0.19848	

Table 4.6: Ten consecutive missing values from position 5 to 14 with their estimates.

t	**Actual AR(1) data (X_t)**	**Data with missing values**	**Estimated values**
1	86.21399	86.21399	
2	73.03785	73.03785	
3	63.71109	63.71109	
4	53.52442	53.52442	
5	47.14499	-	45.96669
6	38.94666	-	39.45658
7	35.22709	-	33.84572
8	31.41615	-	29.00622
9	25.95318	-	24.82522
10	21.70792	-	21.21522
11	17.5829	-	18.08614
12	15.35208	-	15.36927
13	12.63766	-	13.00266
14	11.30084	-	10.9324
15	9.111297	9.111297	
16	8.188405	8.188405	
17	6.679135	6.679135	
18	5.224529	5.224529	
19	3.283825	3.283825	
20	2.94902	2.94902	

Positions of missing values	**Calculated χ^2 value**	**Table value $\chi^2_{0.05,9}$**
5-14	0.37234	16.92

Table 4.7: Eighteen consecutive missing values from position 2 to 19 with their estimates.

t	**Actual AR(1) data (X_t)**	**Data with missing values**	**Estimated values**
1	86.21399	86.21399	
2	73.03785	-	74.11012
3	63.71109	-	63.69526
4	53.52442	-	54.73207
5	47.14499	-	47.01625
6	38.94666	-	40.37197
7	35.22709	-	34.6478
8	31.41615	-	29.71327
9	25.95318	-	25.45593
10	21.70792	-	21.77875
11	17.5829	-	18.59792
12	15.35208	-	15.84095
13	12.63766	-	13.44501
14	11.30084	-	11.35549
15	9.111297	-	9.524767
16	8.188405	-	7.91112
17	6.679135	-	6.477774
18	5.224529	-	5.192006
19	3.283825	-	4.024678
20	2.94902	2.94902	

Positions of missing values	**Calculated χ^2 value**	**Table value $\chi^2_{0.05,17}$**
2-19	0.53467	27.59

We generate another 20 values of AR(1) using different values of ϕ and X_0 as illustrated in Table 4.8 and Tables 4.9 to 4.14 shows the missing values and their estimates for this simulation.

Table 4.8: Simulated AR(1) data when $\phi = 0.5$ and $X_0 = 1000000$.

`

t	X_t
1	500000
2	249999.8
3	125001.2
4	62500.23
5	31250.01
6	15622.97
7	7809.437
8	3905.664
9	1953.414
10	977.1117
11	488.8349
12	244.358
13	122.5056
14	61.77014
15	33.05295
16	15.78969
17	7.215367
18	3.808213
19	2.473077
20	1.745028

Table 4.9: One missing values at position 5 , 10 and 15 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	500000	500000	
2	249999.8	249999.8	
3	125001.2	125001.2	
4	62500.23	62500.23	
5	31250.01	-	31249.28
6	15622.97	15622.97	
7	7809.437	7809.437	
8	3905.664	3905.664	
9	1953.414	1953.414	
10	977.1117	-	976.8995
11	488.8349	488.8349	
12	244.358	244.358	
13	122.5056	122.5056	
14	61.77014	61.77014	
15	33.05295	-	31.02393
16	15.78969	15.78969	
17	7.215367	7.215367	
18	3.808213	3.808213	
19	2.473077	2.473077	
20	1.745028	1.745028	

Position	Maximum percentage deviation
15	6.14%

Table 4.10: Two consecutive missing values from position 5 to 6, 10 to 11 and 15 to 16 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	500000	500000	
2	249999.8	249999.8	
3	125001.2	125001.2	
4	62500.23	62500.23	
5	31250.01	-	31249.53
6	15622.97	-	15623.59
7	7809.437	7809.437	
8	3905.664	3905.664	
9	1953.414	1953.414	
10	977.1117	-	976.7415
11	488.8349	-	488.4398
12	244.358	244.358	
13	122.5056	122.5056	
14	61.77014	61.77014	
15	33.05295	-	30.78871
16	15.78969	-	15.20163
17	7.215367	7.215367	
18	3.808213	3.808213	
19	2.473077	2.473077	
20	1.745028	1.745028	

Positions of missing values	Calculated χ^2 value	Table value $\chi^2_{0.05,1}$
5-6	0.00003	3.841
10-11	0.00032	
15-16	0.17701	

Table 4.11: Three consecutive missing values from position 5 to 7 10 to 12 and 15 to 17 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	500000	500000	
2	249999.8	249999.8	
3	125001.2	125001.2	
4	62500.23	62500.23	
5	31250.01	-	31250.06
6	15622.97	-	15624.92
7	7809.437	-	7812.232
8	3905.664	3905.664	
9	1953.414	1953.414	
10	977.1117	-	976.7462
11	488.8349	-	488.4517
12	244.358	-	244.3829
13	122.5056	122.5056	
14	61.77014	61.77014	
15	33.05295	-	30.88013
16	15.78969	-	15.4302
17	7.215367	-	7.695365
18	3.808213	3.808213	
19	2.473077	2.473077	
20	1.745028	1.745028	

Positions of missing values	Calculated χ^2 value	Table value $\chi^2_{0.05,2}$
5-7	0.00124	5.991
10-12	0.00044	
15-17	0.18295	

Table 4.12: Five consecutive missing values from position 5 to 9 and 15 to 19 with their estimates.

t	**Actual AR(1) data (X_t)**	**Data with missing values**	**Estimated values**
1	500000	500000	
2	249999.8	249999.8	
3	125001.2	125001.2	
4	62500.23	62500.23	
5	31250.01	-	31250.13
6	15622.97	-	15625.09
7	7809.437	-	7812.596
8	3905.664	-	3906.4
9	1953.414	-	1953.405
10	977.1117	977.1117	
11	488.8349	488.8349	
12	244.358	244.358	
13	122.5056	122.5056	
14	61.77014	61.77014	
15	33.05295	-	30.90335
16	15.78969	-	15.48824
17	7.215367	-	7.817251
18	3.808213	-	4.054887
19	2.473077	-	2.319966
20	1.745028	1.745028	

Positions of missing values	**Calculated χ^2 value**	**Table value $\chi^2_{0.05,4}$**
5-9	0.00170	9.488
15-19	0.21821	

Table 4.13: **Ten consecutive missing values from position 5 to 14 with their estimates.**

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	500000	500000	
2	249999.8	249999.8	
3	125001.2	125001.2	
4	62500.23	62500.23	
5	31250.01	-	31250.12
6	15622.97	-	15625.06
7	7809.437	-	7812.539
8	3905.664	-	3906.284
9	1953.414	-	1953.172
10	977.1117	-	976.6453
11	488.8349	-	488.4415
12	244.358	-	244.4584
13	122.5056	-	122.7046
14	61.77014	-	62.30301
15	33.05295	33.05295	
16	15.78969	15.78969	
17	7.215367	7.215367	
18	3.808213	3.808213	
19	2.473077	2.473077	
20	1.745028	1.745028	

Positions of missing values	Calculated χ^2 value	Table value $\chi^2_{0.05,9}$
5-14	0.00714	16.92

Table 4.14: Eighteen consecutive missing values from position 2 to 19 with their estimates.

t	Actual AR(1) data (X_t)	Data with missing values	Estimated values
1	500000	500000	
2	249999.8	-	250000
3	125001.2	-	125000
4	62500.23	-	62500
5	31250.01	-	31250
6	15622.97	-	15625
7	7809.437	-	7812.5
8	3905.664	-	3906.25
9	1953.414	-	1953.125
10	977.1117	-	976.5632
11	488.8349	-	488.2828
12	244.358	-	244.1437
13	122.5056	-	122.0765
14	61.77014	-	61.04752
15	33.05295	-	30.54231
16	15.78969	-	15.30825
17	7.215367	-	7.728314
18	3.808213	-	4.012536
19	2.473077	-	2.303026
20	1.745028	1.745028	

Positions of missing values	Calculated χ^2 value	Table value $\chi^2_{0.05,17}$
2-19	0.27716	27.59

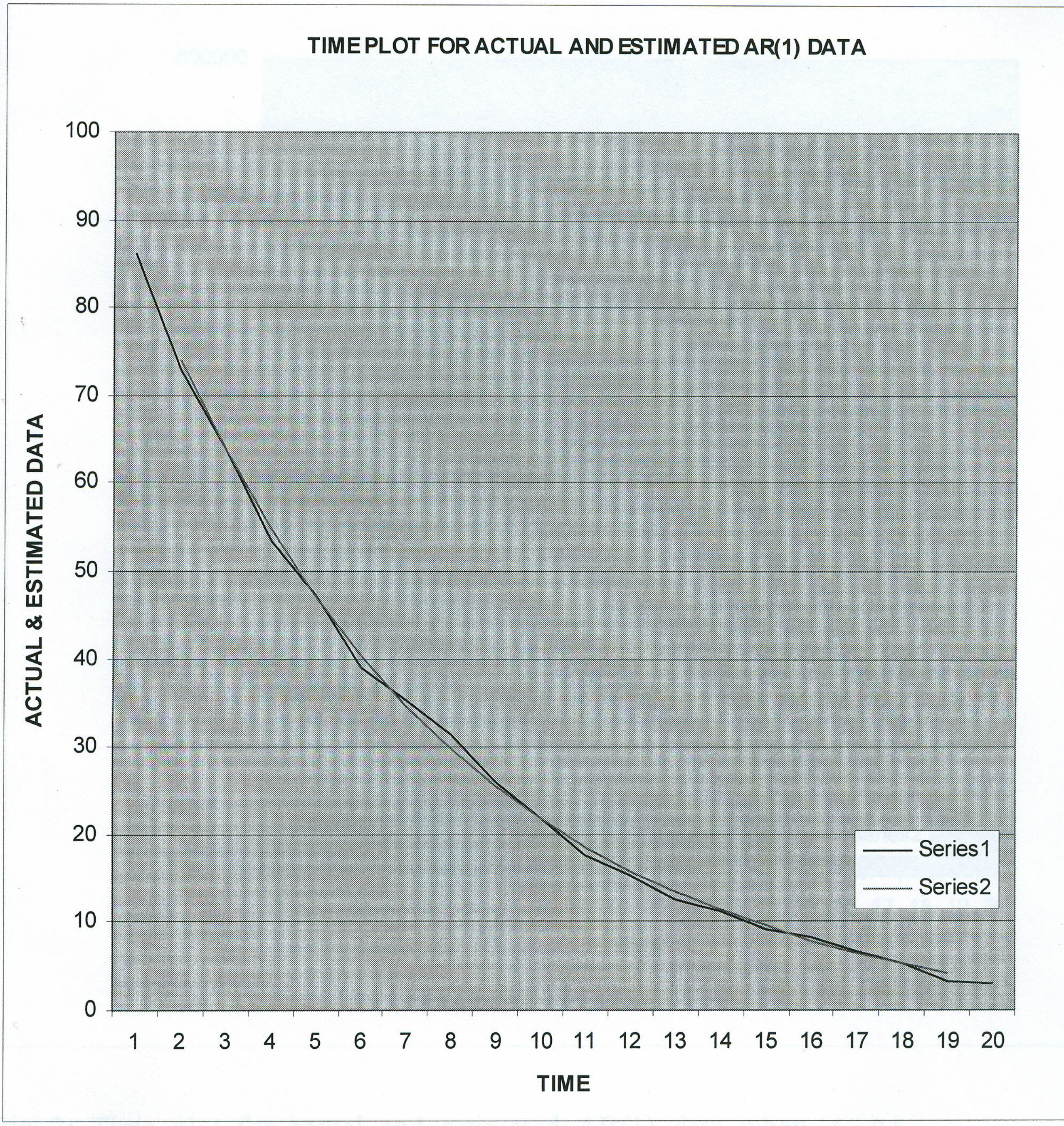

Fig. 1: Time plot for actual and estimated AR(1) data for eighteen missing observations when $\phi = 0.86$ and $X_0 = 100$

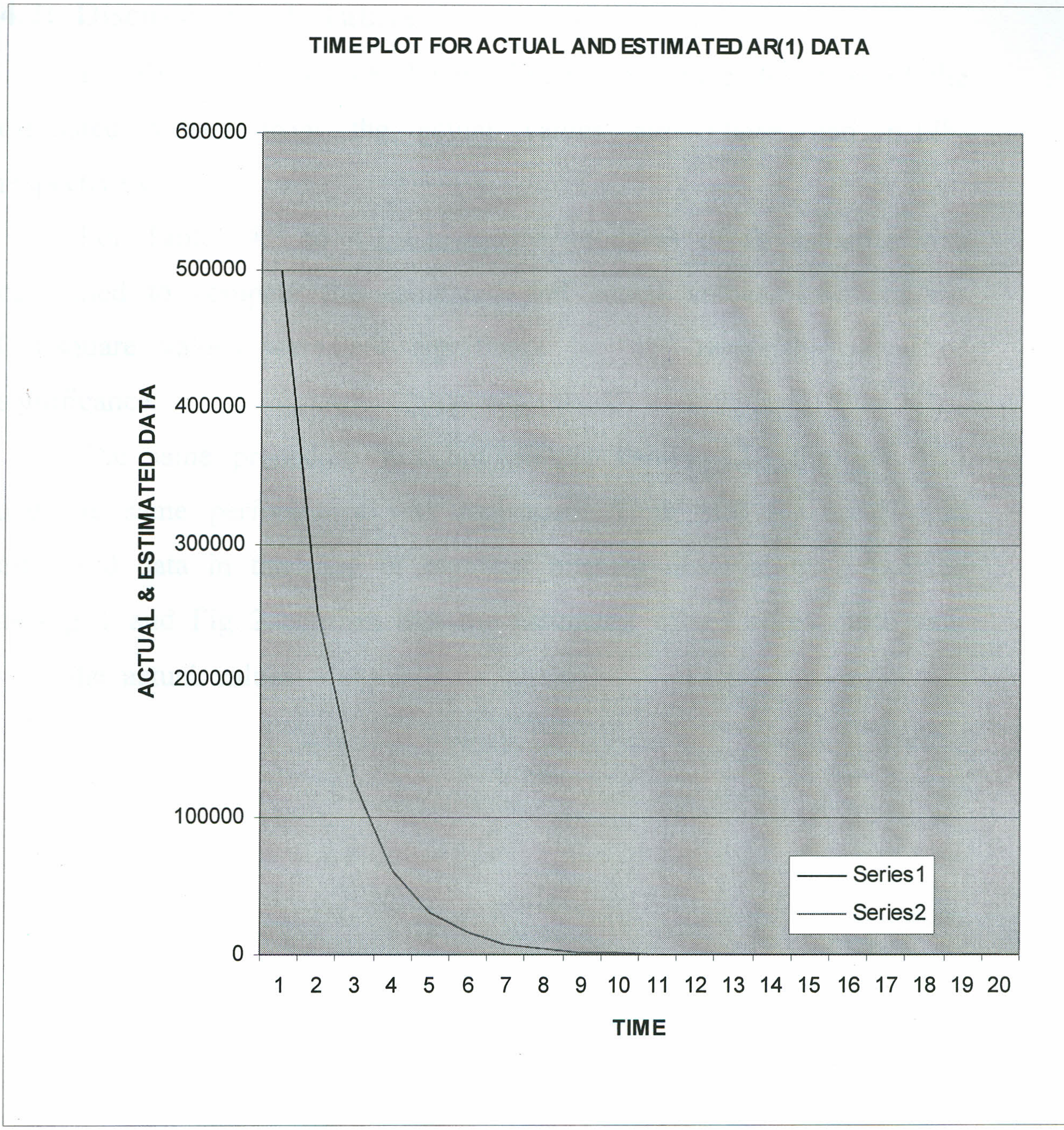

Fig. 2: Time plot for actual and estimated AR(1) data when $\phi = 0.5$ and $X_0 = 1000000$

In both Tables: Series 1 is actual AR(1) data.

Series 2 is estimated AR(1) data.

4.2: Discussions of Tables

In tables 4.2 and 4.8 the maximum percentage deviation of the estimated values from the actual values are 3.03% and 6.14% respectively.

For Tables 4.3 to 4.7 Chi-square goodness of–fit statistics was calculated to compare the estimated and actual values. Each of the Chi-square values were not significant for any reasonable level of significance.

The same procedure was applied on Tables 4.10 through 4.14 and the same performance was registered. Time plot for actual and estimated data in the case of eighteen missing observations presented as Fig. 1 and Fig. 2 confirm that the estimated values agree quite well with the actual values.

CHAPTER FIVE

CONCLUSION AND ACCOMPLISHMENTS

5.0: Introduction

In this chapter we outline the general achievements of the project. We also suggest areas for further research which have emerged during the course of our study.

5.1: Achievements

Most methods developed for estimation of missing observations in Time Series analysis have been limited to the case of one or two consecutive missing observations. In this project we have employed the state space models which can handle irregularly spaced data. Missing observations in a Time Series can safely be treated as special case of such data. In particular we have extended the formula derived in Abraham and Thavaneswaran (1991) to encompass the case where there are more than two missing observations. In a special case we apply the formula on AR(1) simulated data and it seems to perform satisfactorily.

We have given complete proofs to Theorems 2.1 and 2.2 of Abraham and Thavaneswaran (1991), here restated as Theorems 3.1.1 and 3.1.2.

A comprehensive and simple exposition of state space representation for some commonly used Time Series models has been presented. This will go along way in introducing and popularizing this technique to undergraduate students of a Time Series analysis course, and other users of Time Series.

Theory of estimating function following Godambe`s (1985) theorem, is used to obtained optimal estimates of missing observations in some a linear and nonlinear Time series models.

5.2: Suggestions

From this project it should however be noted, that the procedures may not cover all the nonlinear Time Series situations and the methods should be adapted to meet particular needs. Further research is required to develop a unified approach that would cover all conceivable linear and nonlinear models.

Further simulation studies could be conducted to asses the performance of the general formula derived in chapter three on higher order AR, MA, ARMA and ARIMA models.

In the optimal estimation method we used algorithms in which it is assumed that the model parameters ϕ, σ_u^2 and σ_β^2 are known. Such an assumption about the model parameters is not uncommon in the context of estimation of missing observations. However, in practice, model parameters may be estimated using part of the data (see Abraham (1981)). Such an approach for optimal estimation of missing observations in a nonlinear models remains to be an area for further research.

5.3: Conclusion

The occurrence of missing observations is quite common in Time Series and the general model 3.1.1 may be used to characterized such situations. This project offers two alternative techniques for estimating missing observations. The methodology in chapter three section 3.3 to 3.5 can be applied for a restricted class of models whenever the normality assumption is made on the errors.

The optimal estimation method in chapter three section 3.6 is more general and is useful when a practitioner has doubts about specifying a particular distribution for the errors. It should be noted however, that the procedures may not cover all the nonlinear Time Series situations and that other methods should be developed to meet particular needs as they arise.

REFERENCES

Abraham, B. (1981). Missing observation in time series, *comm. Statist.* A-Theory methods, **10**, 1645-1653.

Abraham, B. and Thavaneswaran, A. (1991).A nonlinear Time series model and estimation of missing observation. *Ann. Inst. Stat. Math* Vol 43, 422-528.

Anderson, B.D.O and Moore, J.B. (1979). *Optimal Filtering*, Englewood Cliffs, NJ: Prentice Hall.

Anderson, O.D. (1982). Time series analysis and forecasting: Anderson, O.D and M.R Perryman (eds.), *applied time series analysis.* New York. North-Holland pub. C, 522-528.

Ansley, C. F. and Kohn, R. (1883). Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. *Biometrika* **70,** 275-278

Bartlett, M, S (1946). On the theoretical specification of sampling Properties of autocorrelated time series. Journal of the royal Statistical society, series **B8**, 27-41.

Box, G.E.P and Jenkins, G.M (1970). *Time analysis forecasting and control.* San Francisco, Holden-Day.

Chan, N. H. and Palma, W. (1998). State-space modeling of long-memory process. *Ann. Statist.* **26**, 719-740.

Chatfield, C. (1989). The analysis of time series. An introduction, London: Chapman and Hall.

Chatfield, C. and Collins, A. J. (1980). *Introduction to multivariate Analysis,* London: Chapman and Hall.

Durbin, J. (1960). " The fitting of time series models". Review of the international statistical Institute. **28**, 233-244.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica,* **50**, 987-1007.

Gabr, M. M. (1991*). Recursive Estimation of bilinear Time series models,* Technical Report No. 207, Dept. of math, UMIST.

Godambe, V.P. (1985). The foundation of finite sample estimation in stochastic processes, Biometrika **12**, 419-128.

Guegan, D. (1987). Different representations for bilinear models. *J. Of Time Ser.*, **8(4),** 389-408.

Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting (with discussion), J. *Roy. Statist.* Soc.Ser. B, **38**, 205-248.

Hervey, A.C. (1989). *Forecasting, structural Time series models and Kalman Filter,* Cambridge: Cambridge university press.

Hervey, A.C. and Pierce, R.G. (1984). Estimating missing observations in economic time series. *J. Amer. Statist. Assoc.* **79**, 125-131.

Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to Time series with missing observations, Technometrics; **22,3,** 389-395.

Jones, R. H. (1985). Time series analysis with unequally space data. In: E.J.Hannan, P. R.Krishnaiah and M. M. Rao, Eds., *Handbook of Statistics,* Vol. 5. North-Holland, Amsterdam, 157-177.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Engrg., Series D,* **82**, 35-45.

Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive technique, Journal *of time series Analysis,* **2**, 103-116.

Kitagawa, G. (1983). Changing spectrum estimation, J. *of sound and vibration,* **89**, No. 4.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion) J. *Amer. Statist. Ass.* **82**, *1032-1063.*

Liu , J. (1989). A simple condition for the existence of some stationery bilinear Time Series. *J. of Time Ser. Anal.*, **10(1)**, 33 -39.

Ljung, L. G. and Söderströn, M. (1983). *Theory and practice of Recursive identification,* MIT press, Cambridge.

Nassiuma. D. (1994). Symmetric Stable Sequences with missing observations. *J. Time Ser. Anal.* **15**. 313-323.

Nassiuma, D. (1999). A note on interpolation of stable process. Egerton University, Njoro. *Unpublished manuscript.*

Nichols, D.F. and B.G. Quinn (1982). Random Coefficient Autoregressive models: *An introduction lecture notes in statistics no.* 11. Springer-Verlag, New York.

Nichols, D.F. and A.R. Pagan (1983). Heteroscedasticity in models with lag dependent variables. *Econometrica,* **51**,1233-1242.

Ozaki, T. (1985). Nonlinear time series models and dynamic systems, *handbook of statistics,* Vol. 5 (eds, E.J. Hannan, P.R. Krishnaiah and M.M. Rao), 25-83, North Holland, Amsterdam.

Palma, W. and Chan, N. H. (1997). Estimation and forecasting of long-memory processes. J. *Forecasting* **16**, 395-410.

Palma, W. and Del Pino, G. (1999). Statistical analysis of incomplete long-range dependent data. *Biometrika*, **86**, 4, 965-972

Penza, J. and Shea, B. (1997). The exact likelihood of an autoregressive-moving average model with incomplete data. *Biometrika* **84**, 919-928.

Pham-Dinh, T. (1985). Bilinear Markovian representation and bilinear models. *Stoch. Proc. Appl.,* **20**, 295-306.

Priestley, M.B. (1981). Spectral Analysis and Time Series. Vol. 1 &2. London, Academic Press.

Priestley, M.B. (1988). Current developments in time series modeling. *Journal of Econometrics,* North-Holland.

Subba Rao, T. and Gabr, M, M (1984). The estimation and predication of subset bilinear time series models with applications. *J. of Time Ser. Anal,* ***2(3)****, 153 -171.*

Subba Rao, T. (1993). Development in Time Series analysis. London: Chapman & Hall.

Thavaneswaran, A. and Abraham, B. (1988). Estimation for nonlinear time Series models using estimating equations. *J. Time ser. Anal.* ***9****, 99-108.*

Thavaneswaran, A. and Abraham, B. (1991). Estimation of multivariate nonlinear time series models. *Journal of stst. Plann. & Inference.* **29**, 351-1653.

Tjostheim, D (1986). Estimation in nonlinear time series models, stochastic process. Appl. **21**, 251-273.

Wei, W. W. S. (1990). *Time series analysis.* New York. Addison-Wesley Publishing Company.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, econometrica **48**, 817-838.