

**EFFECTIVENESS OF RECURSIVE
ESTIMATION OF TIME SERIES ANALYSIS
AND FORECASTING**



T.M.J.A. Cooray
University of Moratuwa, Sri Lanka
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Department of Mathematics

Faculty of Engineering

University of Moratuwa

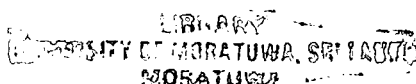
Sri Lanka

2003

**EFFECTIVENESS OF RECURSIVE ESTIMATION
OF
TIME SERIES ANALYSIS AND FORECASTING**

by

T.M.J.A.Cooray



**A thesis submitted to University of Moratuwa for the Degree of Master of
Philosophy**

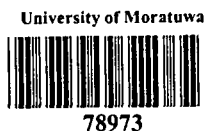


51 " 03 "

519.246.8

**Research work supervised
by**

Dr: M. Indralingam



**DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MORATUWA
MORATUWA
SRI LANKA
JULY 2003**

um Thesis coll.

78973

78973

Declaration

I hereby certify that the work done in this Dissertation is a result of my own effort where, reference is made to the work of authors and this is acknowledged in the text. This Dissertation has not been submitted for another degree.

UOM Verified Signature

T.M.J.A.Cooray

Name of the Candidate

.....
31/07/2003
Signature



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Acknowledgements

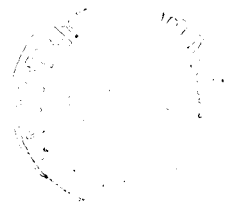
Most of the work in this thesis is the result of joint work with my supervisor Dr M. Indralingam, Senior Lecturer, Department of Mathematics, University of Moratuwa, It is my pleasure to express my gratitude for his invaluable advice and encouragement during my days as a research student.

I would like to take this opportunity to thank Post Graduate unit of University of Moratuwa granting me permission to conduct this research. and The Head of the Department and all the staff members of the Department of Mathematics, University of Moratuwa

Thank all of you who had given their kind hands to complete this task and gather the information ,especially to Mr. Sunil Hemasiri from Central Bank of Sri Lanka who had helped me in obtaining the Statistical data.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk



ABSTRACT

This study is about practical forecasting and analysis of time series, to investigate the effectiveness of recursive estimation of time series analysis and forecasting performance for real data sets. It addresses the question of how to analyze time series data, identify structure, explain observed behavior, modeling those structure and how to use insight gained from the analysis to make informed forecasts. For the purpose of the study total production of paddy and total demand of electricity in Sri Lanka were used. Those values were obtained from the Annual Bulletin, published by the Central Bank of Sri Lanka.

The thesis is organised into two parts. The first part is a course of methods and theory. Time series modelling concepts are described with 'abstract' definitions related to actual time series to give empirical meaning and facilitate understanding. Formal algorithms are developed and methods are applied to analyze data. Two detailed case studies are presented, illustrating the practicalities that arise in time series analysis forecasting. The second part is a course of applied time series analysis and forecasting. It shows how to build the models and perform the analyses shown in the first part using the our own software called "Space" and another downloadable software called the "BATS" application program

The first few chapters are concerned with sing theoretical aspects of en-bloc time series models such as the seasonal decomposition method exponential smoothing method, Winter's seasonal method, and the ARIMA methodology to describe the behaviour of the data series. Even though fairly general, these model do not account for the uncertainties due to the specific choice of trend / seasonal/ level. The main drawbacks in this study are its lack of accessing model uncertainties, when choosing the recursive estimation of time series models based on the Kalman filter. Therefore we used an approach that incorporates all uncertainties involved in the time series modelling simultaneously.

Dynamic state space models provided an excellent basis for constructing and forecasting models for a number of reasons. In particular recursive estimation of time series based on the use of discounting techniques proved to be extremely useful in practice. Many practitioners have a natural feel for the discounting

concept, and furthermore when one discounting factor has been specified, the standard technique may be utilised. In addition to that the Kalman filter based on state space form and Bayesian models can be used to analyse the incomplete data set using EM algorithms.

The last two chapters were devoted for empirical evaluation of data series in order to investigate the effectiveness of recursive estimation of time series. According to the forecast performance of recursive time series models are much more accurate than the en-bloc models. This means that the mean percentage error (MAPE) recursive estimation of time series model is relatively small (nearly 0.5%) so that this method gives higher degrees of accuracy. The recursive estimation of time series models can play an important role of time series modelling. However, these procedures are based on the predictor-corrector type algorithms. Hence without identifying the appropriate structure the variation of parameters could be implemented in contrast to "en-bloc" procedures which could be used only after assuming the specific type of parameter variation.



LIST OF PUBLICATIONS

1. Cooray, T.M.J.A. and M. Indralingam (2001), Missing value Estimation of Time Series Data Using a Spread Sheet, Sri Lankan Journal of Applied statistics volume 2 (published)
2. Cooray, T.M.J.A. and M. Indralingam (2002) Auto regressive Modelling Approach to Forecasting Paddy Yield, Sri Lankan Journal of Applied statistics volume 3 (to be appeared)
3. Cooray, T.M.J.A. and M. Indralingam (2002) Evaluation of some Techniques for Forecasting of Electricity Demand in Sri Lanka, Sri Lankan Journal of Applied statistics volume 3 (to be appeared)
4. Cooray, T.M.J.A. and M. Indralingam (2003), Modelling Sector wise Demand for Sri Lanka using Bayesian Techniques, Journal of Science, Eastern University Sri Lanka.(to be appeared)



CONTENTS

List of Figures	i-iv
List of Tables	v-viii
List publications	ix
Abbreviations	x-xi

Chapter 1

1.1. Introduction	1
1.2. Nature of Time Series	1
1.3. Analysis of Time Series Using En-bloc Methods	3
1.3.1. Box-Jenkins Approach for ARIMA Methodology	4
1.4. Recursive Time Series Models	7
1.4.1. The State Space and Kalman Filter Models	7
1.4.2. Bayesian Approach	8
1.4.3. Auto Regressive model	10
1.5. Aim of Research	11
1.6. Forecasting Procedures	12
1.7. Out Line of Research	14

Chapter 2

The Traditional en-bloc Time Series Models	15
2.1. Introduction	15
2.2. Decomposition Time Series method	15
2.2.1. Introduction	15
2.2.2. Additive and Multiplicative models	15
2.2.3. The Seasonal and Cyclical Components	17
2.2.4. Test for Seasonality	17
2.2.5. Advantages and Disadvantages of the Decomposition	18
2.3. Exponential Smoothing Method	18
2.3.1. Introduction	18
2.3.2. The Methodology of Exponential Smoothing	18

2.3.3. Determination of an Appropriate Factor	19
2.4. Double Exponential Smoothing Method	20
2.4.1. Advantages and Disadvantages of Exponential Smoothing	21
2.5. Winters' Seasonal Exponential Smoothing	21
2.5.1. Introduction	21
2.5.2. The Additive Winters' Method	22
2.5.3. Updating the Decomposition Results	22
2.5.4. Obtaining the Optimal Weights	23
2.5.5. Advantages and Disadvantages of the Winters' Methodology	24
2.6. Box-Jenkins Methodology	24
2.6.1. Introduction	24
2.6.2. ARIMA Models	25
2.6.3. ARMA (p,q) Models	26
2.6.4. ARIMA (p,d,q) Models	26
2.6.5. Seasonal ARIMA Models	27
2.6.6. Autocorrelation	27
2.6.7. Partial Autocorrelation functions	28
2.6.8. Estimates of Parameters	28
2.6.9. Yule-Walker estimates	28
2.6.10. Model Identification	29
2.6.11. Steps for model identification	29
2.6.12. Model Selection Criteria	32
2.7. Forecast Performance of Time Series	33

Chapter 3

Recursive Estimation of Time series models	38
3.1. Introduction	38
3.2. State Space Form	39
3.2.1. Introduction	39
3.3. The Multivariate State-space Model	40
3.4. Computation of State Space form using the Kalman Filter	41



3.5.	Derivation of the Kalman Filter	42
3.5.1.	Kalman Smoothing	43
3.6.	Estimation of Parameters of the State Space Model	44
3.7.	Applications of State Space form to Different Time Series Models	47

Chapter 4

	Kalman Filter based on Bayesian Forecasting	50
4.1.	Introduction	50
4.2.	Bayesian Dynamic Models	51
4.3.	Definitions and Notation of Model	54
4.4.	Updating equations for Univariate Linear Models	56
4.5.	Posterior Information	57
4.6.	Smoothing or Filtered Distribution	58
4.7.	Sequential Analysis	59
4.8.	Monitoring the Model Forecast	59
4.9.	Variance Analysis	60
4.10.	Discount factor as an aid to choosing W_t	62
4.11.	Monitoring Forecasting Performance	62
4.12.	Intervention Facilities	63
4.13.	Bayes Factor	64
4.14.	Implementation of Model	65

Chapter 5

	Recursive Estimation of Time-Varying Parameter Models	69
5.1.	Introduction	69
5.2.	Derivation of $\hat{\theta}_{t s}$ By Recursively Regression of $\hat{\theta}_t$	69
5.3.	Discounted Weighted Regression & Forecasting	73
5.3.1.	Introduction	73
5.3.2.	Recursive Ordinary Least Square Estimation (ROLS)	74
5.3.3.	Parameter Variation & Recursive Estimation	75

5.4.	Auto Regression Model	76
	Akaike's Information Criterion (AIC)	78

Chapter 6

	Other Usage of Recursive Estimation of Time Series Models	80
6.1.	Introduction	80
6.2.	Analysis of Missing Data and EM Algorithm	80
6.2.1.	Introduction	80
6.3.	Univariate Sample with missing data	82
6.4.	Analysis of Incomplete data (with missing values)	83
	Based on likelihood estimations	
6.5.	Maximizing over the parameters and Missing data	85
6.6.	Application of EM algorithm in the State space form	87
6.7.	Modeling with Missing Value in Bayesian Technique	89
6.8.	Communicating Missing Values Using BAT soft ware	89

Chapter 7

	Implementation of Recursive Methods in EXCEL Spread Sheet	91
7.1.	Introduction	91
7.2.	Structure of Kalman Filter Based on State Space Model	91
7.3.	Modification of program for incomplete data sets (Missing Values)	94
7.4.	Implementation of Computer Codes for Method of Auto-regression	94

Chapter 8

	Interactive Time Series Analysis and Forecasting	97
8.1.	Introduction	97
8.2.	Data Used for Study	98
8.3.	Initial Analysis	100
8.3.1	Analysis of Electricity Demand data set	100

8.3.2. Analysis of Paddy data set	113
-----------------------------------	-----

Chapter 9

Empirical Evaluation of Specific Time Series Models	121
9.1. Introduction	121
9.2. Case of Paddy Data	121
9.2.1. Analysis of Paddy Data Using Seasonal Decomposition Model	123
9.2.2 Analysis of Paddy Data Using Winter's Seasonal Model	125
9.2.3 Box Jenkins ARIMA Methodology	127
9.3 Analysis of Paddy Data Using Recursive Estimation Models	130
9.3.1 Introduction	130
9.3.3 Model specifications for proposed method	130
9.3.2 Recursive Estimation of State Space Model	130
9.4 Bayesian Forecasting Techniques for Paddy data Series	134
9.4.1 Retrospective Analysis	138
9.4.2 Monitoring Forecast performance	140
9.4.3 Analysis with Monitoring	141
9.5 Analysis of Paddy data set using Method of Autoregression	143
9.6 Summary of Forecast Performance for case of paddy data	149
9.7 Analysis in Case of Electricity Demand Data	151
9.7.1 Empirical Evaluation of Electricity data using Exponential Smoothing method	152
9.7.2 Box Jenkins ARIMA Methodology	154
9.8 Analysis of Electricity Demand Data Using Recursive Estimation Models	155
9.8.1 Introduction	155
9.8.2 Recursive Estimation of State Space Model	156
9.8.3 Model specifications for proposed method	156
9.9 Analysis of Electricity Demand Using Auto regression Method	159
9.8 Analysis of Electricity data using Bayesian Technique	162

9.9 Summary of Forecast Performance for case of paddy data	168
Chapter 10	
Modeling With Incomplete Data Set	170
10.1 Introduction	170
10.2 Analysis of missing values using state space model	170
10.3 Analysis Incomplete data using Bayesian Approach	174
10.3.1 Analysis with Missing Values in Case of Electricity Data	176
Chapter 11	
11.1 Discussion	181
1.1 Performance of forecast summary	183
11.2 Scope or Further Research	185
11.3 Conclusions	185



LIST OF FIGURES

Figure 2.1: Behavior of additive model

Figure 2.2 behavior of multiplicative model

Fig: 4.1 The Dynamic linear model conditional structure

Fig: 8.1a plot of original graph of . electricity data

Fig: 8.2a Quarterly Indices of Electricity of Sri Lanka

Fig:8.2b Original series and smoothed of quarterly electricity demand of Sri Lanka

Fig: 8.3a SSE values for different values of α for electricity demand series

Fig: 8.4 Plot of smoothed and original values, estimated from exponential smoothing method

Fig: 8.5 original and smoothed graphs based on Kalman filter recursions

Figure 8.6 Time plot of quarterly electricity demand in Sri Lanka

Figure 8.7 One step forecasts

Figure 8.78 Intervention menu

Figure 8.9 Intervention analysis.

Fig: 8.10a plot of original graph of paddy data

Fig: 8.10b Sample ACF of paddy data

Fig: 8.10c Partial auto correlation of paddy data

Fig: 8.11a MSD values for different values of paddy data set

Fig: 8.11b Estimated forecast and original values for α , β and γ values

Fig: 8.11c Estimated seasonal indices for paddy data series from the Winter's seasonal (phaseII) program

Figure(9.1): Original series for paddy data. The total values for the paddy production for "YALA" and "MAHA" seasons in Sri Lanka, from 1958- 2000, $T= 96$

Figure(9.2): Box-Cox transformation for the paddy data.

Figure(9.3) Transformed series for paddy data

Figure 9.4 showed the output graph having original and estimated smoothed series use of Seasonal decompose method.

Fig: 9.4 Smoothed and Actual values of Log(paddy) data series, $T=94$

Fig: 9.5 Seasonal indicis

Fig: 9.6 Mean squared deviation of paddy data series for different discounts

Figure9.7 Forecast values and original values of log(paddy) data series, when MSD was minimum



Figure 9.8 Transformed series for paddy data

Figure 9.9 Estimated Sample ACF and PACF of residuals of paddy data series

Figure 9.11 Graph of actual and forecasted values for Ln(paddy) data in State space model

Figure 9.12 Graph of transformed paddy data series

Figure 9.13 Graph of ln(paddy) data series, as considering steady model (On-line analysis) or forward filtering

Fig: 9.14 On-line estimated factor for paddy data series.

Figure 9.15 Retrospective (backward filtered) graph of level, growth. and seasonal factor for Ln(paddy) data series

Table 9.17 Forecast series for Ln(paddy) data set

Figure 9.16 Monitor setting

Figure 9.17 Forecast horizon for Ln(paddy) data series using Bayesian Approach

Figure 9.18 Estimated forecast value and actual values from the appropriate model for Ln(paddy) data series, using Autoegression methodology

Figure 9.20 Total quarterly electricity demand in Sri Lanka for the period of 1988-2000, Source: Values are taken from annual bulletin published by Central Bank of Sri Lanka.

Figure 9.21 showed the output graph having Mean square root deviation use of Exponential smoothing model.

Figure 9.22 graph of final forecast and original values for electricity demand series when $\alpha(= .36)$ is minimized.

Figure 9.23 Graph of actual and forecasted values for electricity demand data in State space model



Figure 9.24 On line growth of electricity data

Figure 9.25 Predictions at quarter 3, 1992 for electricity data

Figure 9.26 prediction after 1995/Q1

Figure 9.27 shows the fitted values from this intervention analysis.

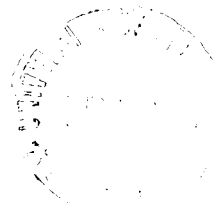
Figure 10.1 Actual and estimated values of incomplete data set for electricity demand series

Figure 10.2 Estimated values for missing observations using EM algorithm on the electricity data

Figure 10.3 Graph of electricity data in case of certain values considered as missing

Fig: 8.4a MSD values for different values of α, β, γ for paddy series

Fig: 8.4c estimated seasonal indices for paddy series from the Winter's seasonal (phase I) program



LIST OF TABLES

- Table 2.1 Appropriate values for Box -Cox Transformation
- Table 2.2 Characteristic of theoretical ACF and PACF for stationary process.
- Table 4.1 Summary of recursive estimation of Bayesian approach for Univariate time series models
- Table 8.1 displays SS E values and corresponding α values
- Table 8.2 Part of the calculated values for α , β , and γ using the Winter's Seasonal(phase1) program
- Table 8.4 estimated log likelihood values together with transition matrix Φ
- Table 9.1 Forecast values for last two values in paddy data series
- Table 9.2 Forecast values for last two values in paddy data series using Winter's seasonal model
- Table 9.3(a) Calculated Box-Pierce values, degrees of freedom, sample variance of error term and AIC values for all possible models applied to the Ln(paddy) data
- Table 9.3(b) Estimated parameters for the final model $(1-B)(1-B^2)Ln(x_t) = z_t - \alpha_1 z_{t-1} - \alpha_2 z_{t-2}$
- Table 9.3(C) Forecast values from fitted model for the Ln(paddy) data
- Table 9.4 Log likelihood and transition matrix of State Space model after each iteration is completed for the Ln(paddy) data series
- Table 9.6 Forecast values for last two values in paddy data series using State Space model
- Table 9.7 Possible discounts for paddy data set
- Table 9.7 Forecast values of last two observations for Ln(paddy) data series
- Table 9.9 Estimated Chi-squared values based on 1 degrees of freedom and BIC(K) values of order determination criterion for Ln(paddy) data series
- Table 9.10 Covariance Structure of lag terms of simulated data series

Table 9.11: Final Estimate of Parameters Fitted Model for Ln(paady) series, applying Autoregression approach for appropriate model

Table 9.12 Analysis of Variance for Appropriate model for Ln(paddy) data

Table 9.13 Estimated Forecast values from fitted model for the Ln(paddy) data

Table 9.14 Summary of Forecast performance of specific models in the case of paddy data

Table 9.14 Forecast performance of electricity demand using Exponential smoothing model

Table 9.15 Calculated Box-Pierce values, degrees of freedom, sample variance of error term for all possible models applied to the electricity demand data

Table 9.16 Estimated parameters for the final model $(1 - \phi_1 B - \phi_2 B^2)(1 - \Phi_1 B^4)X_t = z_t(1-B)(1-B^4)(x_t) = z_t$



Table 9.17 Forecast values from fitted model for the Ln(paddy) data

Table 9.18 Log likelihood and transition matrix of State Space model after each iteration is completed for the electricity demand data set

Table 9.19 Forecast values for last two values in electricity demand data set using State Space model

Table 9.20 Estimated Chi-squared values based on 1 degrees of freedom and BIC(K) values of order determination criterion for electricity demand data series

Table 9.21 Covariance Structure of lag terms of simulated data series

Table 9.22: Final Estimate of Parameters Fitted Model for Ln(paady) series, applying Autoregression approach for appropriate model

Table 9.23 Analysis of Variance for Appropriate model for Ln(paddy) data

Table 9.24 Estimated Forecast values from fitted model for the electricity data

Table 9.25 Estimated Forecast values from fitted model for the electricity data

Table 9.26 Forecast performance of case of electricity data

**Table 10.1 Successive parameter estimates using EM algorithm on the electricity demand data
(with missing values)**

Table 10.2 Estimated values by EM algorithm in the case of electricity demand data

Table 10.3 Analysis summary statistics for electricity data

Table 10.4 Estimated missing values after backward filtering in Bayesian Techniq

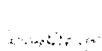


Table 10.5 Summary estimated missing values based on State Space model and Bayesian model

series Table 11.1 Forecast performance of case of electricity data

Table 11.2 Forecast performance of case of paddy data



ABBREVIATIONS

ACF	Autocorrelation function
AIC	Akaike's Information Criterion
ARIMA(p,d,q)	Autoregressive Integrated moving averages of order p, d and q, where p order of autoregressive terms and q order of moving averages terms and d number of differenced required
ARMA(p,q)	Autoregressive of order p and moving averages of order q
AR	Autogressive components
MA	Moving averages components
BATS	Bayesian Applied Time Series Soft ware
BIC	Schwarz Information Criterion
B	Backward operator
∇	Difference operator
ρ_k	Autocorrelation function of lag k
$\hat{\rho}_k$	Sample autocorrelation at lag k
φ_{kk}	Partial autocorrelation function of lag k
$\hat{\varphi}_{kk}$	Sample autocorrelation at lag k
Eq ⁿ	Equation
ME	Mean error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MSS	Mean sum of squares
MS _E (P)	Error mean sum squares

C_p	Mallows C_p statistic
$R(K)$	Autocovariance function of lag k
$ARIMA(p,d,q)(P,D,Q)_s$	Seasonal autoregressivs integrated moving model of normal components p,q and seasonal components P and Q and differenced d for normal components and D for seasonal components respectively
OLS	Ordinary least squares
CHI	Chi-squared statistic
GDP	Gross Domestic Products
RSS	Regression sum of squares
RSP	Residual Sum of Product
SSE	Sum of Squares of Error

