LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
Faculty of Technology
Computational Engeneering

Horahenage Dixon Vimalajeewa

PARAMETER IDENTIFICATION IN TIME SERIES MODELS

Examiners: Professor Heikki Haario
Dr. (Adj. Prof.) Marko Laine

ABSTRACT

Lappeenranta University of Technology

Faculty of Technology

Computational Engineering

Horahenage Dixon Vimalajeewa

Parameter Identification in Time Series Models

Master's thesis

2015

68 pages, 27 figures, 13 tables

Examiners: Professor Heikki Haario

Dr. (Adj. Prof.) Marko Laine

Keywords: State space, Kalman filter, Auto Regressive Moving Average (ARMA), Dynamic Linear Model (DLM), Markov Chain Monte Carlo (MCMC),

Time series analysis can be categorized into three different approaches: classical, Box-Jenkins, and State space. Classical approach makes a basement for the analysis and Box-Jenkins approach is an improvement of the classical approach and deals with stationary time series. State space approach allows time variant factors and covers up a broader area of time series analysis.
This thesis focuses on parameter identifiablity of different parameter estimation methods such as LSQ, Yule-Walker, MLE which are used in the above time series analysis approaches. Also the Kalman filter method and smoothing techniques are integrated with the state space approach and MLE method to estimate parameters allowing them to change over time.
Parameter estimation is carried out by repeating estimation and integrating with MCMC and inspect how well different estimation methods can identify the optimal model parameters. Identification is performed in probabilistic and general senses and compare the results in order to study and represent identifiability more informative way.

Acknowledgements

I would like to express my gratitude to the Department of Mathematics and Physics giving me a chance to study my master's degree and also providing the scholarship during my studies. Also my heartfelt thanks go to Professor Heikki Haario and Dr. (Adj. Prof.) Marko Laine for giving continuous guidance, remarks and engagement through the whole study.

I also would like to appreciate and acknowledge Matylda Jablonska's contribution, ideas and comments. Especially, My thanks to Isambi Sailon Mbalawata for his continuous assistance during this study.

Finally, I am grateful my bachelor supervisor Dr. Abeyratne and all other teachers for their guidance and support to continue my studies. Also my thanks to my family, and all my love ones, who helped me throughout my studies.

Lappeenranta, April 20, 2015.

Dixon Vimalajeewa

Contents

List of Symbols and Abbreviations **7**

1 INTRODUCTION **8**

1.1 Purpose of the Thesis 8

1.2 Structure of the Thesis 9

2 TIME SERIES ANALYSIS APPROACHES **10**

3 CLASSICAL APPROACH **11**

3.1 Trend 11

3.2 Seasonality 12

3.3 Cycle 13

3.4 Irregular fluctuations 13

4 BOX-JENKINS APPROACH **14**

4.1 Stationary Data 14

4.2 Auto-correlation and Partial Auto-correlation 15

4.3 ARMA Models 17

4.4 Model Identification 18

4.5 Model Estimation 19

4.6 Model Validation 19

4.6.1 Residual Analysis 20

4.7 Forecasting 20

4.8 Discussion 21

5 BOX-JENKINS APPROACH MODEL PARAMETER ESTIMATION METHODS 22

5.1 Yule-Walker Estimation 22

5.2 Least Square Estimation (LSQ) 23

5.3 Maximum Likelihood Estimation (MLE) 24

5.3.1 Conditional Likelihood Function 25

5.3.2 Evaluation of Conditional Likelihood 26

5.3.3 MLE Estimates 26

6 DYNAMIC LINEAR MODEL APPROACH 28

6.1 State Space representation of Time Series 28

6.2 Dynamic Linear Models 28

6.3 DLM Representation of ARMA Models 29

6.4 DLM Representation of Linear Regression Model 31

6.5 Kalman Filter Estimation Method 32

6.5.1 Kalman Algorithm 33

6.5.2 Parameter Estimation 34

6.6 Smoothing 34

7 TIME SERIES ANALYSIS WITH MARKOV CHAIN MONTE CARLO 36

7.1 Bayesian Inference Method 36

7.2 Markov Chain Monte Carlo (MCMC) Method 37

7.2.1 Metropolis Algorithm 38

8 SIMULATION EXAMPLES AND REAL WORLD APPLICATIONS 39

8.1 Simulation Examples 39

8.1.1 Pre-analysis 39

8.1.2 Matlab *garchfit* Gstimation Method 40

8.1.3 ARMA Parameter Estimation 44

8.1.4 DLM Estimation 47

8.1.5 MCMC Analysis 47

8.2 Real World Examples 52

8.2.1 Internet Server Logging Example 53

8.2.2 Airline Passenger Example 55

8.2.3 Seat Belt Example 59

9 DISCUSSION 62

10 CONCLUSIONS 63

REFERENCES 64

List of Tables 66

List of Figures 67

List of Symbols and Abbreviations

iid	independently and identically distributed
PDF	Probability Density Function
MLE	Maximum Likelihood Estimation
DLM	Dynamic Linear Model
MCMC	Markov Chain Monte Carlo
LSQ	Least Square
ARMA	Auto Regressive Moving Average
AR	Auto Regressive
MA	Moving Average
SSE	Sum of Squares Error
MSE	Mean Squares Error
PDF	Probability Density Funaction

1 INTRODUCTION

Time series analysis is an important tool in vast range of fields such as medicine, engineering, economic, social. Global climate studies, dynamical system behaviour, and stock market fluctuations are some examples where time series analysis is commonly used. More examples can be found in *Harvay*(1989), *West Harrison*(1997), *Durbin and Koopman*(2001), *Kunsch*, and [1]. Therefore, its interesting and important to know how time series analysis is applied in these kind of applications.
Considering the development of time series analysis, only few impotent stages are mentioned here. *Yule* (1926) proposed a method to observe correlation between time series variables. *Box* and *Jenkins* (1970) proposed improved version of *Yule*'s method later called *Box-Jenkins* approach, but *Commandeur and Koopman* (2007) argued that *Box-Jenkins* is problematic in some cases. Also, *Kalman* introduced Kalman filter method in 1960. Later, DLMs, Kalman filter with *Bayesian* framework offered much more flexibility in wide range of time series applications.
Basically, model identification and model building, parameter estimation, and forecasting are the main stages of all time series analysis applications [2]. In order to achieve desired goals from time series analysis, strong knowledge about these stages is very helpful.

1.1 Purpose of the Thesis

Among the above mentioned stages, model parameter estimation has a considerable impact on final results and conclusions because they are based on the estimated model. Since the availability of many different estimation methods, parameter identifiabillity may differ from method to method. Therefore, the basic purpose of this thesis is to study the parameter identifiability of different time series model parameter estimation methods.
Basically, ARMA models and DLMs parameter estimation is studied using common estimation methods such as LSQ, Yule-Walker, MLE, Kalman filter with MLE, and Smoothing techniques. Then, parameter identifiability of each estimation method is studied in two ways: repeating estimation process and Bayesian inference with MCMC approach.

1.2 Structure of the Thesis

To fulfil the objectives, this study is organized in a way that first, we precisely explain the basic background of time series analysis. Second, based on that background, the main time series analysis approaches and their different parameter estimation methods are discussed. Third, simulation exercises and different real world cases are taken into consideration to study how parameter estimation methods contribute to parameter identification process.
The theoretical part consists of the main time series analysis approaches; Classical, Box-Jenkins, State space with DLM and their parameter estimation methods. Also, a short discussion about Bayesian framework with MCMC in parameter estimation is included in the first five sections.
The practical part includes simulations and real world examples based on ARMA models and DLMs in order to study the parameter identifiability. Then identifiability is studied in two ways; repeating the estimation and performing MCMC analysis and then comparing how estimations close to their true values, relation with other parameters as well.
Matlab software with GARCH and *dlmtbx* Matlab toolboxes are used to perform the simulation examples and the real data problems in the practical part.

2 TIME SERIES ANALYSIS APPROACHES

All analysis methods have their own objectives and they depend on the purpose of analysis. According to [2], there are three main objectives in time series analysis and they explain how the analysis is conducted and what are the intended outcomes:

(1) **Description:** The first step is to have a look at the given time series and get rough idea of the behaviour of data by using simple statistical plots and measures. Then extract basic properties of time series such as trend, seasonality, turning points, extreme behaviours.

(2) **Explanation:** Identifying the dynamic mechanism that generates the process of which sequence of observations are available.

(3) **Prediction:** The typical aim of time series analysis is to make predictions for future observations.

To conduct a time series analysis, it is necessary to have some statistical models which should be able to explain the given process well. Basically, there are two statistical models called ***Error model*** and ***Stochastic model***. According to [2], they are

(1) **Error Model**

$$X_t = f(t) + \epsilon_t, \qquad t = 1, 2, \cdots, N, \tag{1}$$

where $\epsilon_t \sim iid(0, \sigma^2)$, f is an explicit mathematical function can describe the mean behaviour of the observations .

(2) **Stochastic Model**

$$X_t = g(\epsilon_t, \epsilon_{t-1}, \cdots), \tag{2}$$

where g is a function of stochastic variables.
This model is totally stochastic since the mechanism that generates ϵ is stochastic

According to time series analysis methods in literature, there are three main approaches called ***classical approach***, ***Box Jenkins approach***, and ***State Space approach*** to achieve these objectives. First, the classical approach is explained following Section. The basic concept of this approach is used in the two other approaches are discussed in Sections (4), (5), and (6).

3 CLASSICAL APPROACH

In the classical approach, under the Equation (1), every time series is considered as a combination of unobserved factors called *components* of time series. Basically, there are four components: *Trend (T), Cycle (C), seasonality(S), Irregular fluctuations(I)* and the main purpose of this decomposition is to analyse them component wise. The decompositions can be *Additive* or *Multiplicative*, but this decomposition is not unique since number of assumptions have to be made based on the behaviour of the given time series.

$$\begin{aligned} x_t &= T_t + C_t + S_t + I_t \quad \text{Additive,} \\ x_t &= T_t C_t S_t I_t \qquad \text{Multiplicative.} \end{aligned} \tag{3}$$

3.1 Trend

Trend is the long term change in the mean level of time series. In practice, trend is not directly estimated and the analysis focuses on *trend-cycle*. Trend-cycle is the variation low frequency in time series when medium and high frequency fluctuations have been filtered out and it is estimated by removing seasonal and irregular components from the original data [19].
The error model is used to estimate trend and it may take linear or polynomial form such as *linear, quadratic, exponential*. The trend estimation process consists of three different cases:

(a) If the functional form, f is known, then we only need to estimate the parameters. For example, Figure 1 represents a linear case and regression method can be used to estimate f.

(b) If the functional form of f is unknown, first, approximate a good functional form for f and then follow the part (a).

(c) If (a) or (b) is not possible, use smoothing techniques such as *exponential smoothing, or moving average* to remove the random behaviour of time series. These techniques are non-parametric and have the same functional behaviour over the entire time period.

3.2 Seasonality

The same pattern in fixed time interval is repeating in the entire time period, as in Figure 1. Mathematically this is expressed $x(t) = x(t+s) = x(t+2s) = \cdots$, where x- data, t - time, s-seasonal time period. If $s = 4, 6, 12$, then it can be called as quarter, semi-annual or annual seasonal effect respectively.
The regression method can also be used to estimate the seasonal component of a given time series. Suppose a periodic function is denoted by $g(t)$ and the model is given by Equation (1). We can use g to express the data generation process instead of f in Equation (1). There are two ways to estimate g:

(a) Representing g as a sum of dummy variables

$$g(t) = \sum_{j=1}^{S} \gamma_j d_{jt},$$

where $d_{jt} = 1$ at j^{th} period and zero otherwise, γ_j is the level of phenomenon at j^{th} period of the entire period.

(b) Seasonal components as a sum of harmonic components.
For more details of these two methods follow [4].

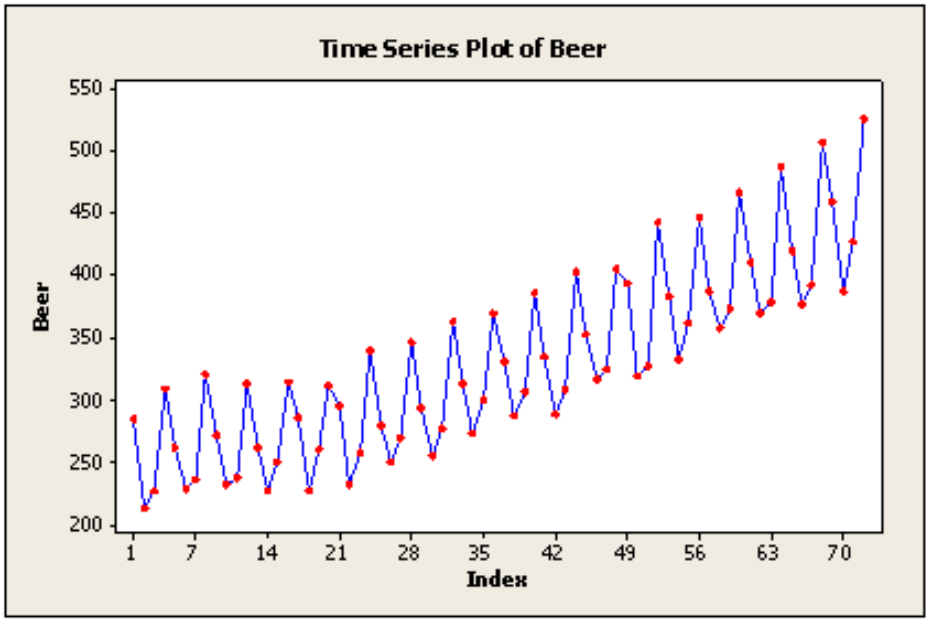


Figure 1: Trend and seasonal behaviours of beer selling process

3.3 Cycle

Cyclic component contain other cyclic behaviours besides seasonality. This is calculated in different ways, such as by elimination or averaging out trend effect by *residual method* [19]. Examples can be found in Chapter 8, [3].

3.4 Irregular fluctuations

The symmetric or random fluctuations except seasonal and cyclic components are considered as irregular fluctuations. If the process is extremely irregular, then it is difficult to apply any analytic methodologies. The ***moving average*** and other non-parametric method can be used for this type of time series. This method does not need to define any functional form since it is non-parametric approach and it captures the dynamic behavior well of a given time series. More details about this method can be found in [19, 3]

Component wise analysis is more informative but different methods have to be used to analyze each component. Therefore, this approach is coupled with other approaches and makes good basis for other approaches to start the analysis with well defined mathematical model to carry out better analysis.

4 BOX-JENKINS APPROACH

George Box and *Gwilym Jenkins* (1979) introduced this method to analyze stationary time series data. This is a collection of statistical concepts and principles in order to find best fit model and make forecast on time series data.
Here, time series is considered as a finite realization of a stochastic process and the method focuses on the stochastic behaviour of the data. Moreover, the error term is considered simultaneously with other components since it has a considerable impact on the behaviour of time series. Therefore, the stochastic model given by Equation (2) is used to describe the process in a stochastic point of view. This approach also helps to recognize the dynamic behaviour as well, but the problem is that the classical approach overcome here [2].
Basically, the procedure consists of four stages: *identification, estimation, validation and forecasting.* These stages give clear path to express time series by a mathematical model and do prediction and forecasting. Mathematical concepts such as seasonality, correlation, ARMA models are used commonly to explain these stages. Therefore, first, we will briefly discuss these concepts.

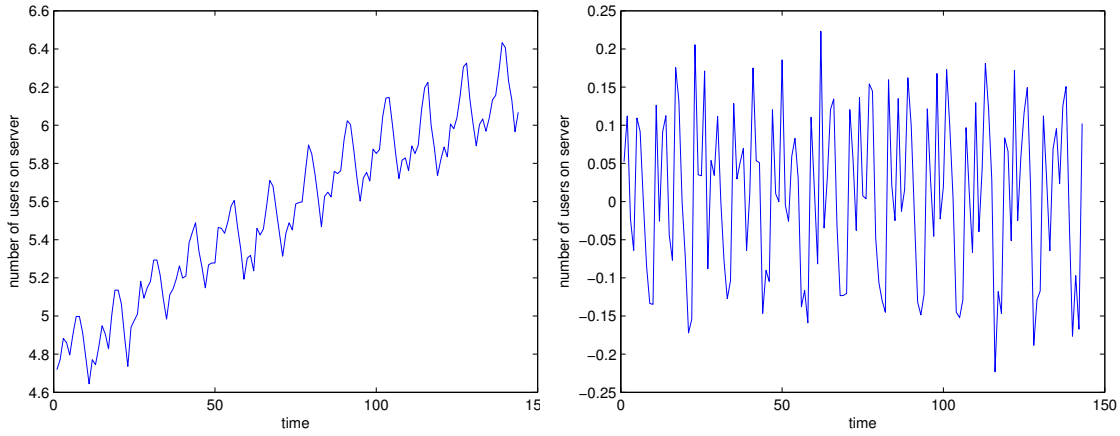
4.1 Stationary Data

If the statistical properties such as mean, variance, auto correlation structure of a time varying process do not change over time a process is called stationary. If the given data is not stationary, **regular differencing, transformations** can be used to make the data stationary. Moreover, statistical tests such as *Dickey-Fuller, Phillips-Perron* can be used to check the stationary condition [15].

> **Differencing:** A simple mathematical concept to make a new data series is taking the difference between each two consecutive data points. Suppose $\{X_t\}_{t=1}^{N}$ is a non-stationary time series and stationary series $\{Y_t\}_{t=1}^{N}$ is obtained by $Y_t = \bigtriangledown X_t = X_t - X_{t-1}$ for $t = 2, \cdots, N$. Usually, two differencing are enough to obtain stationary but it is not a rule.

Note: Trend and irregularities can be adjusted by applying differencing and log or inverse transformations. Moreover, seasonal non-stationarity is adjusted by seasonal differencing [2]. For instance, Figure 2a shows a strong trend and seasonality and data is transformed into a stationary form by applying log transformation to

achieve seasonal stationary and by differencing for trend stationary. Figure 2b shows stationary version of the original data.



(a) Non-stationary time series (b) Stationary time series

Figure 2: Stationary version of non-stationary time series is obtained by log transformation and differencing

4.2 Auto-correlation and Partial Auto-correlation

Auto-correlation and partial auto correlation describe dependency or relationship between random variables based on covariance. Covariance measures how random variables change/behave together and it is defined for random variables X and Y as

$$cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)], \tag{4}$$

where μ_x, μ_y are means of X and Y respectively and E denotes the expectation. Correlation is an extended version of Equation (4) to quantify the strength of the relationship and it is defined for arbitrary two random variables X and Y as

$$\rho_{xy} = \frac{cov(X, Y)}{\sigma_x \sigma_y}, \tag{5}$$

where σ_x and σ_y are the standard deviations of X and Y respectively.

Auto-correlation Function (ACF)

ACF represents the dependency between time series observations. Sometimes the terms *lagged correlation* and *serial correlation* are used. Mathematically, ACF at time lag k is defined as

$$\gamma_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma_x^2} = \frac{cov(X_t, X_{t+k})}{\sigma_x^2}, \tag{6}$$

where X_t are time series observations, μ and $\sigma_x^2$are the mean and variance of series $\{X_t\}_{t=1}^{N}$.
Usually, ACF is expressed in auto-covariance terms as

$$\rho_k = \frac{\gamma_k}{\gamma_0}. \tag{7}$$

Partial Auto-correlation Function (PACF)
PACF measures the correlation between two lags when the mutual dependency of lags between those two lags has been removed. For instance, the partial correlation between X_t and X_{t+k} is the correlation between these two lags when the mutual linear dependency between $X_{t+1}, \cdots, X_{t-(k-1)}$ has been removed. This is also called as the conditional correlation ($cov(X_t, X_{t+k}|X_{t+1}, \cdots, X_{t-(k-1)})$ and usually denoted by r_{kk} and defined as

$$r_{kk} = \frac{\gamma_k - \gamma_{k-1}^2}{1 - \gamma_{k-1}^2}, \tag{8}$$

where γ_k is the auto correlation at lag k.
Generally, these two statistical measures measure the correlation between time lags but, in partial autocorrelation procedure the correlation of time lags between the two times is being removed. One example where these functions are used to identify model order of ARMA are explained later. Figure 3 presents ACF and PACF of ARMA(3,5) process. The order of the process can be identified by the number of significant lags in each graph. Details about this will be given later in Subsection 4.4

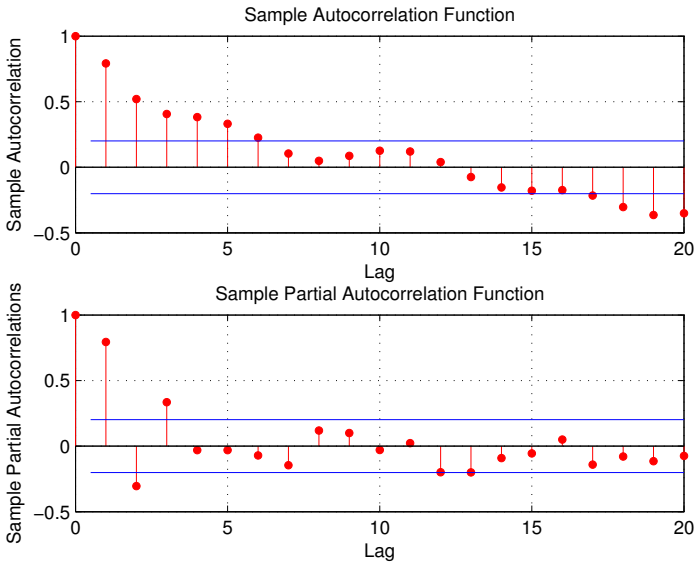


Figure 3: ACF and PACF of ARMA(3,5)

4.3 ARMA Models

The mathematical models which are used in Box-Jenkins approach are called as ARMA models and they are used for stationary data only. There are two basic models: AR and MA and the other models such as ARAM, ARIMA, ARMAX are just combinations of them. Here, we discuss only ARMA model out them of and details about other models can be found in [20].

(1) **Auto Regressive (AR) model:** AR process of order p, AR(p), is defined by

$$X_t = C + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2), \tag{9}$$

where X_t is the observation at time t, C is a constant, $\phi_i, i = 1 : p$ are the auto regressive parameters that describe the effect of unit change in two consecutive observations, ϵ_t is error disturbance, and σ_ϵ^2 is called innovative variance.

(2) **Moving Average (MA) model:** The MA process of order q, MA(q), is defined by

$$X_t = C + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2), \tag{10}$$

where X_t is observation at time t, C is a constant, θ_i's are moving average parameters which describe the unit change of two consecutive time series observations, ϵ_t is error disturbance.
Furthermore, a MA(q) process has a property called *invertibility* if the roots of the characteristic equation of MA(q) process in Equation (11) lie inside a unit circle [15]. Every invertible MA(q) processes can be written as a infinite-order AR process [7].

$$1 = \sum_{i=1}^{\infty} = \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q + \cdots . \tag{11}$$

For example, consider a MA(1) process. According to Equation (10), MA(1) can be written as $Y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$. Then $\epsilon_t = Y_t - \theta_1 \epsilon_{t-1}$ and replacing t by $t-1, t-2, \cdots$ it can be written as $\epsilon_{t-1} = Y_{t-1} - \theta_1 \epsilon_{t-2}$, $\epsilon_{t-2} = Y_{t-2} - \theta_1 \epsilon_{t-3} = \cdots$
Finally, combining all these values and substituting them to $\epsilon_t = Y_t - \theta_1 \epsilon_{t-1}$,

$$\begin{aligned}
\epsilon_t &= Y_t - \theta 1(Y_{t-1} - \theta_1 \epsilon_{t-2}) = Y_t - \theta 1 Y_{t-1} - \theta_1^2 \epsilon_{t-2} \\
&= Y_t - \theta 1 Y_{t-1} - \theta_1^2 (Y_{t-2} - \theta_1 \epsilon_{t-3}) = Y_t - \theta_1 Y_{t-1} - \theta_1^2 Y_{t-2} - \theta_1^3 \epsilon_{t-3} \\
&\vdots \\
&= Y_t - \theta_1 Y_{t-1} - \theta_1^2 Y_{t-2} - \theta_1^3 Y_{t-3} - \cdots .
\end{aligned}$$

If $|\theta_1| < 1$, this can continue further and finally the result becomes

$$Y_t = \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \theta_1^3 Y_{t-3} + \cdots + \epsilon_t.$$

It is clear that this is of the form of AR(∞).

(3) **Auto Regressive Moving Average (ARAM) model:** The ARMA process is a combination of AR(p) and MA(q) processes it is denoted by ARMA(p, q) and defined as

$$X_t = C + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \tag{12}$$

where C is a constant, ϕ_i's and θ_i's are AR and MA parameters respectively, ϵ_t are the iid error terms and p and q refer to the AR part order and MA part order respectively.
This model is applied for well behaved time series and explains how the current observation linearly depends on the past observations and on current and past error disturbances.

Now, based on above discussed concepts, the stages of Box-Jenkins are discussed.

4.4 Model Identification

Model selection is the first step in Box-Jenkins procedure. Since there might be many possible models for one data set, there should be a formal way to find most suitable model. There are different methods such as graphical methods, *Bayesian information criteria (BIC), Akaike information criteria (AIC), and Reversible Jump MCMC* (see [3, 15]). Here, we used the first two, only.

(a) **Bayesian information criteria (BIC)**
BIC was introduced by *Schwartz*(1978) to select a suitable model for a given data set. Even though BIC is related with the Bayesian and MLE estimation methods, it is possible to use in ARMA model selections as well. The model which has highest posterior probability (i.e minimum BIC) is selected as the most suitable model from set of models. BIC is calculated as

$$BIC = -\ln L(\hat{\theta}|y) + k\ln(n), \tag{13}$$

where y are observations, L is likelihood function for y based on model M_k, $\hat{\theta}$ is the MLE estimate of model parameter vector θ_k of the model M_k, obtained by maximizing $L(\theta_k|y)$ over $\Theta(k)$, where $\Theta(k)$ is the parameter space and $\theta_k \in \Theta(k)$, $k \in \{k_1, k_2, \cdots, k_L\}$ for $L > 1$, n is sample size.
Here we only represents the BIC formula by Equation (13). More details about AIC and BIC can be found in [14] and [3].

(b) **Graphical method**
Plot data and extract rough idea about the components and check whether the stationary condition holds. Apply necessary modification to transform to stationary if needed. Then, use ACF and PACF of stationary data to identify model order. PACF is used to identify order of AR processes, while the ACF is used to identify MA order. Table 1 presents summarized information how this can be done more precisely .

Table 1: ARMA model order identification statements by ACF and PACF

Model	ACF	PACF
AR(p) (ARMA(p,0))	decreasing towards zero	significant util p^{th} lag
MA(p) (ARMA(0,q))	significant until q^{th} lag	decreasing towards zero
ARMA(p,q)	decreasing towards zero	decreasing towards zero

4.5 Model Estimation

Model estimation is very impotent because fitted model is obtained via estimating the model parameters. There are different methods to the estimate parameters such as Yule-Walker, LSQ, and MLE and they can be used according to the type of ARMA model. The selection of estimation method depends on many factors such as the model, computational efficiency, accuracy, intended final outcomes. Next Section (5) explains more precisely the parameter estimation methods.

4.6 Model Validation

After model identification and estimation, adequacy of the estimated model or ability to explains the given process and the model assumptions are checked. Further

modifications to the current model are then applied if necessary. This is performed by *diagnosis (residual)* analysis.

4.6.1 Residual Analysis

Basically, in residual analysis we check some conditions such as whether the fitted model explains the entire process well, are the model assumptions fulfilled, accuracy of the predictions and forecasts. If they seem to be on satisfactory levels, then it is confirmed that an adequate model has been selected for the given data. Statistical tools such as scatter plot of residuals, residual standard error, coefficient of determination (R^2), normal qq plot, and some statistical tests (t test, z test) are commonly used here. Residual analysis and some diagnosis are listed below.

(a) Check model assumptions: Plot estimated residuals $\{r_t^*(\hat{\theta})\}$ to see if the residuals are stationary and also check that they are iid with mean zero and variance one. The independence of auto correlations $\hat{\rho}_\epsilon$'s can be checked by using following test statistic. More details are in [2].

$$Q^2 = \sum_{k=1}^{m} \left(\sqrt{N}\hat{\rho}_{\epsilon(\hat{\theta})}(k)\right)^2 \sim \chi^2_{m-n}. \tag{14}$$

(b) The significant of the estimated parameters can be checked by t **tests**.

(c) R^2 value tells that how much estimated model explains original data and should be close to one if the best fit model has been selected.

(d) Linear behaviour of normal qq-plot of residuals checks the normality assumption of the errors.

If all these criteria verified, it can be said that a good, or adequate model has been selected. On the other hand, if some of these are not satisfied, it means that some modification is needed. We start again from the first stage and repeat until get the best model.

4.7 Forecasting

The last stage of the Box-Jenkins procedure is forecasting. When the best fit model has been identified, it can be used for forecasting future observation. One way to

check predictive ability ts to divide the data into two sets called training and testing sets, and then confirm that the current model perform well with the training data. Finally, use the testing set to forecast future data. There different ways to carry out this, with more details, can be found in [15].

4.8 Discussion

The Box-Jenkins procedure gives methodology to analyze stationary time series. On the other hand, *Durbin and Koopman (2001, p.53)* pointed out that, it is not possible to achieve the stationary behaviour to all time series. The reasons are time varying factors that are considered as nuisance factors and removed (or assume as constants) before the start of analysis. Therefore, it is not possible to analyse the dynamic behaviours of time series by ARMA methods. Section 6 explains how to solve these problems and gives a comprehensive approach for time series analysis.

5 BOX-JENKINS APPROACH MODEL PARAMETER ESTIMATION METHODS

The purpose of discussing the Box-Jenkins approach for model parameter estimation is to discuss the basic concepts of different estimation methods in order to select the most suitable method. Here, three main methods Yule-Walker, LSQ and MLE are discussed

5.1 Yule-Walker Estimation

In Yule-Walker method, the population moments are equated with the sample moments to obtain a set of equations whose solution gives the estimator. Therefore it is also known as *Method of Moments Estimation.* This method is an efficient estimator for AR model, but not for MA and ARMA process [7]. Consider the general AR(p) model given in Equation (9) and multiply both sides by X_{t-k} and then take the expectations:

$$x_t x_{t-k} = \phi_1 x_{t-1} x_{t-k} + \phi_2 x_{t-2} x_{t-k} + \cdots + \phi_p x_{p-k} x_{t-k} + \epsilon_t x_{t-k}$$
$$E[x_t x_{t-k}] = \phi_1 E[x_{t-1} x_{t-k}] + \phi_2 E[x_{t-2} x_{t-k}] + \cdots + \phi_p E[x_{t-p} x_{t-k}],$$

where $E[(x_t - \bar{x})(x_{t-k} - \bar{x})] = cov(x_t, x_{t-k}) = c_{t-k}$ and $E[x_{(t-p)} x_{(t-k)}] = 0$, $\bar{x} = 0$ (stationarity of AR). Dividing the above expression by $N - 1$, we get

$$c_k = \phi_1 c_{k-1} + \phi_2 c_{k-2} + \cdots + \phi_p c_{p-k} \ (\ cov = c, c_{-l} = c_l).$$

Dividing this by c_0, according to Equation (6), we get

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{p-k}.$$

Finally, writing the above expression for all time points $t = 1, \cdots, N$, it can be expressed in matrix form

$$\gamma_p = \Gamma_p \Phi, \tag{15}$$

where $\gamma_p = (\gamma_1, \gamma_2, \cdots, \gamma_p)^T$, $\Gamma_p = \{\gamma_{i-j}\}_{i,j=1,2,\cdots,p}$ (variance covariance matrix of X), $\Phi = (\phi_1, \gamma_2, \cdots, \phi_{p-1}, \gamma_p)^T$.

Next, replace the population covariance γ by sample covariances $\hat{\gamma}(k)$. Then parameters ϕ can be estimated by solving Equation 15.

$$\hat{\phi} = \hat{\Gamma}_p^{-1}\hat{\gamma}_p, \quad \text{where} \quad \hat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(x_{t+k} - \bar{x})(x_t - \bar{x}). \tag{16}$$

This set of equations is called *Yule-Walker* equations. They are often expressed in terms of the auto correlation function rather than the auto covariance function as $\hat{\phi} = \hat{\Gamma}_p^{-1}\hat{\rho}_p$. *Durbin-Levinson* method can be used to solve this set of equations [2].

5.2 Least Square Estimation (LSQ)

The basic idea behind this method is to estimate parameters which minimize the sum of squares of errors; where errors are the difference between predicted model values and observed values. Here, the general procedure is explained using AR(2) process. According to Equation (9),

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \theta_p Y_{t-p} + \epsilon_t \text{ for } t = 1, 2, 3, \cdots, N.$$

Since this process is linear, the residual sum of squares is calculated by

$$SS(\hat{r_t^*}) = \sum_{t=1}^{N}(\hat{r_t^*})^2, \tag{17}$$

where $\hat{r_t^*} = Y_t - (\theta_1 Y_{t-1} + \theta_2 Y_{t-2})$.
If ϵ_t's are truly white noises, then their ACF has no spikes and PACF values would be small. It would help to obtain the most reasonable estimators for θ_1 and θ_2 [15]. Now find the parameter values which minimize Equation (17) using some suitable numerical minimization method. Similar procedure can be followed in matrix form and this is computationally quite easy and fast. First, write the AR(2) model for each time step for $t = 1, 2, \cdots, N$ $Y_3 = \theta_1 Y_2 + \theta_2 Y_1, Y_4 = \theta_1 Y_3 + \theta_2 Y_2, \cdots, Y_N = \theta_1 Y_{N-1} + \theta_2 Y_{N-2}$. Second, put above set of linear equations into matrix form

$$Y = X\Theta, \tag{18}$$

where $Y = \begin{pmatrix} Y_3 & Y_4 & \cdots & Y_N \end{pmatrix}^T, X = \begin{pmatrix} Y_2 & Y_3 & \cdots & Y_{N-1} \\ Y_1 & Y_2 & \cdots & Y_{N-1} \end{pmatrix}^T, \Theta = \begin{pmatrix} \theta_1 & \theta_2 \end{pmatrix}^T.$

Third, since Equation (18) of the form $Y = Xb$, it can be solved by applying simple matrix operations. The solution is given by

$$\hat{b} = (X^T X)^{-1} X^T Y. \tag{19}$$

In literature, the *Direct Inversion method* is also used to estimate AR parameters, but mathematically it is same as the LSQ [5]. Also, LSQ is not able to apply on ARMA models because then the Equation (17) has stochasticity since the random error term is included and error term is not known in practice. The *Psuedo-likelihood* is a good alternative method that can be used in this case [2].

5.3 Maximum Likelihood Estimation (MLE)

MLE is a general estimation method and it can also be used to estimate model parameters of all ARMA models. Moreover, this method can be considered as a special case of Bayesian approach [11]. Even though MLE is comparatively complicated method compared to earlier methods, it allows analyzer to go beyond the simple problems and offers comprehensive methodology to deal with stochastic time series processes. Main steps of the general MLE process are given bellow.

(a) If data set is iid with marginal PDF $f(y;\theta)$ then the joint PDF of the sample of data $Y_N = (Y_1, Y_2, \cdots, Y_N)$ is simply the product of marginal PDFs of each observation.

$$f(y;\theta) = f(y_1, y_2, \cdots, y_N) = \Pi_{i=1}^{N} f(y_i;\theta). \tag{20}$$

(b) The *likelihood function (L)* is the joint density function of Y given θ

$$L(y|\theta) = \Pi_{i=1}^{N} f(y_i;\theta). \tag{21}$$

(c) The *log-likelihood (l)* is the log value of Equation (21)

$$l(\theta) = \log(L(y|\theta)) = \sum_{i=1}^{N} \log f(y_i;\theta). \tag{22}$$

(d) Finally, the estimated θ which maximizes L is usually denoted by $\hat{\theta}$ and defined as

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{N} \log f(y_i;\theta). \tag{23}$$

Consider the general ARMA process is given by Equation (12). A random variable $Y_N|Y_{N-1}$ contains only ϵ_N as a random component at time N and does not depend on anything since its a white noise. Therefore $y_N|Y_{N-1}$ and Y_N are independent, hence the PDF of Y_N is

$$f(y_N|\theta,\sigma_\epsilon^2) = f(y_N|y_{N-1},\theta,\sigma_\epsilon^2) f(y_{N-1}|\theta,\sigma_\epsilon^2), \tag{24}$$

where θ is set of all ARMA model parameters.
The joint PDF of ARMA model, given in Equation (12), can be calculated by factorizing the joint PDF into conditional PDFs and PDF of initial values as follows:

$$\begin{aligned} f(y_1, y_2, \cdots, y_N, \Theta) &= f(y_1|\Theta) f(y_2|y_1, \Theta) \cdots f(y_t|y_{t-1}, \cdots, y_1, \Theta) \cdots \\ &\quad f(y_N|y_{t-N}, \cdots, y_1, \Theta) \\ &= f(y_1|\theta) f(y_2|Y_1, \Theta) \cdots f(y_t|Y_{t-1}, \Theta) \cdots f(y_N|Y_{N-1}, \Theta) \\ &= \Pi_{t=p+1}^{N} f(y_t|Y_{t-1}, \Theta) f(Y_p|\theta, \Theta), \end{aligned} \tag{25}$$

where p is AR model order of ARMA(p, q), and $\Theta = (\theta, \sigma_\epsilon^2)$.
The likelihood function is calculated from Equation (25) as

$$L(Y_N; \Theta) = \left(\Pi_{t=p+1}^{N} f(y_t|Y_{t-1}, \Theta)\right) f(Y_p|\Theta). \tag{26}$$

The Equation (26) is called the **exact likelihood** function of ARAM(p, q) and without the part $f(Y_p|\theta, \sigma_\epsilon^2)$ it is called the ***conditional likelihood***. The log-likelihood function can be calculated following the Equations (22) and (26) as

$$l(\theta|Y) = ln(f(Y_p|\Theta)) + \sum_{t=p+1}^{N} f(y_t|Y_{t-1}; \Theta), \tag{27}$$

where $Y_t = (y_{t-1}, \cdots, y_1)$.
These two likelihoods can be used to calculate MLE of given model, but according to [13], both functions give same answer for stationary models. It may differ in finite samples because of non-stationary and non-invertibility of time series. In practice, the conditional likelihood is used more often than the exact likelihood because computationally it is easier to calculate.

5.3.1 Conditional Likelihood Function

Since special attention is needed to estimate $f(Y_p|\theta, \sigma_\epsilon^2)$, the likelihood given by Equation (26) is not used commonly. The *conditional likelihood function*:

$$L(y_N; \Theta) = \left(\Pi_{t=p+1}^{N} f(y_t|Y_{t-1}, \Theta)\right) \tag{28}$$

is used for estimation processes. When the size of data set is large enough, there is no much difference between MLE estimates which come from likelihood given by Equations (26) or (28).

5.3.2 Evaluation of Conditional Likelihood

Let θ be all model parameters as θ and let σ_ϵ^2 be the error variance of ARMA model which is assumed to be known. Then, one step ahead forecast $\hat{y}_{t|t-1}$ is the mean of $y_t|Y_{t-1}$. According to the Equation (12), the prediction error is $r_t^* = y_t - \hat{y}_{t|t-1}$(residuals) and its variance is σ_ϵ^2. Since ϵ is assumed as white noise, the PDF of Y_t is

$$f(y_t|y_{t-1}, \theta, \sigma_\epsilon^2) = \frac{1}{\sigma_\epsilon\sqrt{2\pi}} \exp\left(\frac{((y_t - \hat{y}_{t|t-1})(\theta))^2}{2\sigma_\epsilon^2}\right). \tag{29}$$

The conditional likelihood of Y_N can be written following Equations (28) and (29),

$$L(Y_N, \theta, \sigma_\epsilon^2) = (\sigma_\epsilon^2 2\pi)^{(\frac{N-p}{2})} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=p+1}^{N}((r_t^*(\theta))^2\right). \tag{30}$$

5.3.3 MLE Estimates

The MLE estimate $\hat{\theta}$ is the parameter value which maximizes (30). Equation (30) gets its maximum when $\sum_{t=p+1}^{N}(r_t^*(\theta))^2$ is minimum. Therefore, we find $\hat{\theta}$ which minimizes

$$SSE(\theta) = \sum_{t=p+1}^{N}(r_t^*(\theta))^2. \tag{31}$$

This is the same as what we discussed in LSQ method. It reveals that in this sense LSQ and MLE are same. Moreover, if σ_ϵ^2 is assumed as unknown parameter in the beginning, it can also be estimated by differentiating the Equation (30) with respect to σ_ϵ^2 which gives an unbiased estimator

$$\hat{\sigma}_\epsilon^2 = \frac{SSE(\hat{\theta})}{N-p}. \tag{32}$$

In order to understand this method more clearly, two examples are given below how MLE is carried out for AR and MA models

MLE for AR(1) processes: The model can be written using Equation (9)

$$y_t = c + \phi y_{t-1} + \epsilon_t \text{ , where } \epsilon_t \sim N(0, \sigma_{\epsilon_t}^2), |\phi| < 1, t = 1, 2, 3, \cdots, N.$$

First, calculate the mean and variance of each y_t. According to Equation (9), it can be shown that y_t is also Gaussian as given in Equation (33)

$$\begin{aligned} y_1|\theta &\sim N\left(\frac{c}{1-\phi}, \frac{1-\phi^2}{2\sigma^2}\right), \\ y_t|(y_{t-1}, \cdots, y_1; \theta) &\sim N(c + \phi y_{t-1}, \sigma^2), t = 2, 3, \cdots, N. \end{aligned} \tag{33}$$

Let $\theta = (c, \phi, \sigma^2)$ be the set of all parameters that has to be estimated. According to the Equation (33), the PDFs of y_1 and y_t for $t = 2, \cdots, N$ are

$$\begin{aligned} f(y_1|\theta) &= \left(2\pi\frac{\sigma^2}{1-\phi^2}\right)^{-\frac{1}{2}} \exp\left(-\frac{1-\phi^2}{2\sigma^2}(y_1 - \frac{c}{1-\phi})^2\right), \\ f(y_t|y_{t-1}, \cdots, y_1; \theta) &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - c - \phi y_{t-1})^2\right). \end{aligned} \tag{34}$$

Finally, the exact log likelihood is calculated using Equations (27) and (34) as

$$\begin{aligned} l(\theta|y) &= \left[-\left(\frac{N-1}{2}\right)\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=2}^{N}(y_t - c - \phi y_{t-1})^2\right] \\ &\quad + \left[-\frac{1}{2}\log(\frac{2\pi\sigma^2}{(1-\phi^2)}) - \frac{1-\phi^2}{2\sigma^2}(y_1 - \frac{c}{1-\phi})^2\right]. \end{aligned} \tag{35}$$

Since the exact log-likelihood in (35) is not a linear function in of θ, numerical maximization methods, such as *Newton-Raphson*, have to be used to estimate MLE of θ as mentioned in [11]. When the second part of (35) is ignored, then (35) becomes the conditional log-likelihood of AR(1):

$$l(\theta|y) = -\frac{N-1}{2}\log(2\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{N}(r_t^*)^2, \tag{36}$$

where $r_t^* = y_t - (c + \hat{\phi} y_{t-1})$ is residual, $\hat{\phi}$ MLE estimete of ϕ.
It is clear that $\sum_{t=2}^{N}(r_t^*)^2$ has to minimize to maximize (36). This error minimization is same as in the LSQ procedure in Section 5.2

MLE for MA(1) processes: As in AR(1) case, same procedure is followed to get MLE for MA parameters. Therefore, only the final result is given bellow

$$\begin{aligned} l(\theta|y) &= \log\left(f_{y_N,\cdots,y_1|\epsilon_0=0}(y_N, y_{t-1}, \cdots, |\epsilon_0 = 0; \theta)\right) \\ &= -\frac{N}{2}\log(2\pi) - \frac{N}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{N} e_t^2. \end{aligned} \tag{37}$$

It is clear that the Equations (37) and (36) are different. Also, this is not same as LSQ method and an iterative process has to be used [12].

MLE is the most general estimation method. In ARMA model case, the exact probabilities of the first p observations of an AR(p) or the first q observations of MA(q) have to be included explicitly. In conditional MLE, all first p or q observations are assumed to be known and used as inputs to the estimation process [5]. Next, state space approach with ***Kalman filter*** method will give more flexible way of time series analysis in state space approach, MLE is used to estimate the parameters, also.

6 DYNAMIC LINEAR MODEL APPROACH

The dynamic linear model (DLM) approach gives comprehensive explanation of the behavior of time variant processes. This approach is based on *state space* methodology (because DLM is a special case of state space method).

6.1 State Space representation of Time Series

The unobserved components of time series at a certain time point are called *states*. Analysis of dynamics of components, as well as other time variant factors, can be carried out by the state space method. Therefore, this method gives an explicit explanation of the dynamic behavior of time series as well as solutions for the challenges and problems that have been arisen in Box-Jenkins procedure.
The state space approach gives an explicit structure for decomposition of time series. It allows to change time variant factors such as components of time series over time in order to capture their dynamic in uni-variate and multivariate cases. Also it allows for missing data, and can tackle with stationary and non stationary problems. DLMs are used to model the components and *Kalman filter* recursion with MLE and smoothing techniques are used to estimate model parameters, compute predictions of states and to reconstruct the behavior and forecast.

6.2 Dynamic Linear Models

The general DLM is specified by the state space representation with *Gaussian error* assumptions, but these extensions are not always necessary [3]. Here we consider the *Gaussian* case only. *Gaussian linear state space model* is expressed by Equations (38a), (38b) called *observation equation* and *state equation*, respectively, with the assumption of independence of v_t and w_t:

$$Y_t = F_t\phi_t + v_t, \qquad v_t \sim N(0, V_t), \tag{38a}$$

$$\phi_t = G_t\phi_{t-1} + w_t, \qquad w_t \sim N(0, W_t), \tag{38b}$$

where Y_t are observations, ϕ_t is state at time t. We assume that $\phi_0 \sim N(\mu_0, C_0)$, G_t and F_t are known system matrices, v_t and w_t are mutually independent random vectors for $t \geq 1$ with mean zero and known covariances V_t and W_t, respectively.
The Equation (38) facilitates a structural framework to model time variant processes

and analyze them together. The following two examples demonstrate how to decompose time series its time variant factors and model them to study their dynamic behaviors.

Examples for DLMs

(1) **Local level model:** This is a simple DLM where the local level is allowed to change over time:

$$\begin{aligned} Y_t &= \mu_t + \epsilon_{t(obs)}, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2), \\ \mu_t &= \mu_{t-1} + \xi_{t(level)}, \qquad \xi_t \sim N(0, \sigma_\xi^2), \end{aligned} \tag{39}$$

where μ_t is the local level and Y_t are the observations, and ϵ, ξ are the random disturbances with mean zero and variances $\sigma_\epsilon^2, \sigma_\xi^2$ respectively.
The Equations (39) and (38) are equal when $G_t = F_t = 1$, $V_t = \sigma_\epsilon^2$, and $W_t = \sigma_\xi^2$.

(2) **Trend model:** This model allows to change the local level (μ_t) and the trend (λ_t) over time and is very useful in trend analysis problems:

$$\begin{aligned} Y_t &= \mu_t + \epsilon_{t(obs)}, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2), \\ \mu_t &= \mu_{t-1} + \lambda_{t-1} + \xi_{t(level)}, \qquad \xi_t \sim N(0, \sigma_\xi^2), \\ \lambda_t &= \lambda_{t-1} + \gamma_{t(trend)}, \qquad \gamma_t \sim N(0, \sigma_\gamma^2). \end{aligned} \tag{40}$$

This model can be expressed in the form of Equation (38), when
$G_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $F_t = [1 \quad 0]$, $\phi_t = [\mu_t \quad \lambda_t]^T$, $W = \begin{bmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\xi^2 \end{bmatrix}$, $V_t = \sigma^2$.

Similarly, it is possible to build DLMs for the other components seasonal, irregularities, and cycles, examples can be found in [1, 3, 8].
In the above examples, the entries of the system matrices F_t, G_t and the error covariances matrices V_t, W_t are constants. These kind of DLMs are known as the *time invariant*. Special case of time invariant DLMs are the ARMA models [1].

6.3 DLM Representation of ARMA Models

Since ARMA models can be expressed in DLM from without changing the distribution of measurement process (Y_t), the DLM form of ARMA can apply for any

non-stationary case without applying any modification on data. The DLM representations of ARMA is not unique. The following representation is used over this thesis. Consider the general ARMA(p, q) model,

$$y_t = C + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t. \tag{41}$$

Let $r = \max(p, q+1), \phi_i = 0$ for $i > p$, and $\theta_i = 0$ for $i > q$ and then rewrite the above equation as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \phi_r y_{t-r} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_{r-1} \epsilon_{t-(r-1)}. \tag{42}$$

According to [1], the DLM form of the Equation (42) is expressed by the measurement equation: $y_t = F\Theta_t + v_t$ and the state equation: $\Theta_t = G\Theta_{t-1} + R\epsilon_t$,

where $\Theta_t = \begin{pmatrix} y_t \\ \phi_2 y_t + \cdots + \phi_r y_{t-r+1} + \theta_1 \epsilon_t + \cdots + \theta_{r-1} \epsilon_{t-r+2} \\ \vdots \\ \phi_{r-1} y_{t-1} + \phi_r y_{t-2} + \theta_{r-2} \epsilon_t + \theta_{r-1} \epsilon_{t-1} \\ \phi_r y_{t-1} + \theta_{r-1} \epsilon_t \end{pmatrix}$,

$F_t = \begin{pmatrix} 1 & [0]_{1\times r-1} \end{pmatrix}$, $G_t = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{r-1} & 0 & 0 & \cdots & 1 \\ \phi_r & 0 & 0 & \cdots & 0 \end{pmatrix}$, $W = R \times R' \times \sigma^2$, $V = 0$ (since ARMA is a stationary model $v_t = 0$), $R = \begin{pmatrix} 1 & \theta_1 & \cdots & \theta_{r-1} \end{pmatrix}^T$, $\epsilon_t \sim \text{iid}(0, \sigma^2)$

Example: Consider the ARMA(2,2) model.

Classical model: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta_{t-1} \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$.

DLM representation:

$$y_t = F_t \theta_t, \qquad \theta_t = G_t \theta_{t-1} + R\epsilon_t,$$

where $G_t = \begin{pmatrix} \phi_1 & 1 & 0 \\ \phi_2 & 0 & 1 \\ \phi_3 & 0 & 0 \end{pmatrix}$, $F_t = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$, $V_t = 0$, $W_t = RR'\sigma^2$, $R = \begin{pmatrix} 1 & \theta_1 & \theta_2 \end{pmatrix}^T$, $\epsilon_t \sim \text{N}(0, \sigma^2)$, and $\phi_3 = 0$.

The two different DLM representations of the AR(2) process, $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$, where $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ given bellow is an example to prove that there is no unique DLM representation for ARMA models.

(a) The transition equation is $\theta_t = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \theta_{t-1} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}$, where $\theta_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}$.
Then the evolution equation system matrices are

$$G_t = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \; w_t = \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}, \; W = \begin{pmatrix} \sigma_t^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

The measurement equation is $y_t = F_t \theta_t$, where $F_t = \begin{pmatrix} 1 & 0 \end{pmatrix}$, $v_t = 0$.

(b) Another representation is $y_t = F_t \theta_t$,
where $F_t = \begin{pmatrix} 1 & 0 \end{pmatrix}, \theta_t = \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \theta_t = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix}$.

Different examples can be found in [8].

6.4 DLM Representation of Linear Regression Model

A linear regression model, with dynamic regression parameters, i.e. those that can depend on time, can be put in DLM form. A simple dynamic linear regression model is

$$y_t = \beta_{t,0} + \beta_{t,1} x_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma^2), \tag{43}$$

where x_t is the independent variable, y is the explanatory variable, and ϵ_t is the error term. The system matrices DLM corresponding to the Equation (38) are

$$\phi_t = [\beta_{t,0}, \quad \beta_{t,1}]^T, \; F_t = [1, \quad x_t], \; G_t = I_{2\times 2}, V = \sigma^2, \text{ and } W_t = \mathrm{diag}(\sigma^2_{\beta_{t,0}}, \sigma^2_{\beta_{t,1}}).$$

On the other hand, when $V = \sigma^2$ and $W_t = 0$, the DLM of this regression model becomes equal to its classical form. Moreover, when the error terms $\xi_t = 0$ and $\gamma_t = 0$ of the Equations (39) and (40), they also represent linear regression models.

According to the Subsections 6.2, 6.3, and 6.4, it can be said that the state space approach has great flexibility of applying in many applications. Next, the challenge is to estimate DLM states and their model parameters. The states are estimated recursively by the *Kalman filter* method, given the data and MLE is used to estimate the other parameters such as error variances. Also smoothing techniques are used to reconstruct time series when the current states of a given process are known.

6.5 Kalman Filter Estimation Method

This recursive estimation method was invented by *R. E. Kalman, 1960* and can be used to study the dynamics of systems such as control of complex dynamic problems and analysis of measurements and estimation problems [10]. This method gives an efficient way to study many time variant process as well.
The main idea behind this method is to estimate the optimal current state of a dynamic system based on the past and current observations. The next state is predicted based on the previous observations and together with new observations, the prediction is used to update the next state to an optimal estimate. Therefore, this is also called *predictor-corrector* method. According to [1], three main filtering steps in the whole process are *recursion, forecasting observation, updating state*, and *forecasting next state.*
Consider the general filtering steps with the state space given by Equation (38). The aim of this filtering is to estimate the optimal state value ϕ_t at every time step $(t = 1, 2, 3, ..., N)$. The iteration can be used to estimate the model parameters via MLE at the same time, also [1].

(a) The predictive distribution of next state $p(\phi_t|Y_{t=1:t-1})$ is computed from the *filtering distribution* $p(\phi_{t-1}|Y_{1:t-1})$ and the conditional distribution of $p(\phi_t|\phi_{t-1})$

$$p(\phi_t|Y_{t=1:t-1}) = \int p(\phi_t|\phi_{t-1})p(\phi_{t-1}|Y_{1:t-1})d_{\phi_{t-1}}. \tag{44}$$

In the linear Gaussian case, suppose that the estimated state and its covariance at $(t-1)^{th}$ step are ϕ_{t-1}^{est} and C_{t-1}^{est}. Then the predicted state ϕ_t^P at the t^{th} step has Gaussian distribution and whose mean and covariance are given below.

$$\phi_t^p = G_t\phi_{t-1}^{est}, \qquad C_t^p = \mathrm{cov}(G_k\phi_{t-1}^{est} + w_k^p) = G_tC_{t-1}^{est}G_t^T + W_t. \tag{45}$$

(The recursion starts assuming the initial state $\phi_0 \sim \mathrm{N}(\phi_0^{est}, C_0^{est})$)

(b) One-step ahead predictive distribution of observations is computed.

$$p(y_t|y_{1:t-1}) = \int p(y_t|\phi_t)p(\phi_t|y_{1:t-1})d_{\phi_t}. \tag{46}$$

The results ϕ_t^p and C_t^p, that come from the part a, are used to update the new observation Y_t^p. According to the Equation (38a), Y_t^p has Gaussian distribution and according to the Equation (38a), its mean and variance are

$$Y_t^p = F_t\phi_t^p, \qquad \text{var}(Y_t^p) = F_tC_t^pF_t^T + V_t. \tag{47}$$

(c) The filtered distribution $p(\phi_t|Y_{1:t})$ is computed by using the *Bayesian* theorem, using $p(\phi_t|y_{1:t-1})$, and likelihood $L(y_t|\phi_t)$.

$$p(\phi_t|y_{1:t}) = \frac{p(y_t|\phi_t)p(\phi_t|y_{1:t-1})}{L(y_t|\phi_t)}. \tag{48}$$

Now the filtered state is computed, the distribution $\phi_t^{est}|Y_{1:t}$ is Gaussian with mean and variance

$$\phi_t^{est} = \phi_t^p + R_t(Y_t - F_t\phi_t^p), \quad C_t^{est} = C_t^p - R_tF_tC_t^p, \tag{49}$$

where the matrix $R_t = C_t^pF_t^T(F_tC_t^pF_t^T + W_t)^{-1}$ is known as *Kalman Gain.*

According to the above discussion, a summarized version of the recursive computation of the Kalman filter state estimation is given below.

6.5.1 Kalman Algorithm

According to the results from Section 6.5, **Kalman filter algorithm** can be summarized as follows

(1) **Prediction:** use ϕ_{t-1}^{est} and C_{t-1}^{est} to compute

(a) $\phi_t^p = G_t\phi_{t-1}^{est}$

(b) $cov(G_t\phi_{t-1}^{est} + w_t^p) = G_tC_{t-1}^{est} + W_t$

(2) **Updating:** Combine the prior current state (ϕ_t^p) with observation Y_k to compute

(a) $R_t = C_k^pF_t^T(F_tC_t^pF_t^T + V_t)^{-1}$

(b) $\phi_t^{est} = \phi_t^p + R_t(Y_t - F_t\phi_t^p)$

(c) $C_t^{est} = C_t^p - R_t F_t C_t^p$

Increase time step t to go to step 1. Follow [1] and [6] for more details.
Moreover, only the states can be estimated by assuming that all the other parameters such as error variances (W_t, V_t) are known. It is not possible always and MLE method can be used to estimate them assuming the *Markov properties* of given system [16].

6.5.2 Parameter Estimation

According to [16], there are three ways to estimate parameters

(a) Subject level knowledge with trial and error to fix parameters without any algorithmic tuning.

(b) Maximum likelihood function with a numerical optimization routine to find MLE of the parameters and plug the estimations back to the equation and re-fit the DLM model.

(c) MCMC sampling from the posterior distribution of the parameters to estimate the parameters or to integrate out their uncertainty.

Here, we use (b) and (c) only and part (c) is explained in Section (7). Using the marginal likelihood function $l(y_{1:N}|\Theta)$ and assuming the Markov properties of the system, the likelihood function can be obtained sequentially by Kalman filter method:

$$-2\log(p(y_{1:N}|\Theta)) = c + \sum_{t=1}^{N} \left[(Y_t - F_t \phi_t^{est})^T C_t^Y (Y_t - F_t \phi_t^{est}) + \log(|C_t^Y|) \right], \quad (50)$$

where Θ is set of all parameters which have to be estimated, c is a constant. It depends on the model prediction covariance C_t^Y. With this likelihood, a numerical optimization method can be used to estimate the parameters. More details and proofs of these formulas can be found in [1, 6, 16].

6.6 Smoothing

Smoothing is used to reconstruct the underlying behaviour of time dependent process when data is given. For instance, this can be used in economic studies when

somebody wants to understand the behavior of consumption of a consumer product backward from certain day to its history within certain past few years. This is carried out by computing the conditional distribution of states, given data ($p(\phi_{1:N}|y_{1:N})$) and estimate all states backward in history. The general steps of smoothing are summarized below as given in [1].

(1) Calculate the conditional distribution of states, given data $y_{1:N}$ and the state sequence $\{\phi_0, \phi_1, \cdots, \phi_N\}$. The Bayesian formula is used here.

$$p(\phi_t|\phi_{t+1}, y_{1:N}) = \frac{p(\phi_{t+1}|\phi_t)p(\phi_t|y_{1:t})}{p(\phi_{t+1}|y_{1:t})}. \tag{51}$$

(2) The smoothing distribution of states given data is computed according to the following backward recursion at t, starting from $p(\phi_t|y_{1:N})$, where we remove the dependency of ϕ_t on ϕ_{t+1} by integrating it out:

$$p(\phi_t|y_{1:N}) = p(\phi_t|y_{1:t}) \int \frac{p(\phi_{t+1}|\phi_t)}{p(\phi_{t+1}|y_{1:t})} p(\phi_{t+1}|y_{1:N}) d\phi_{t+1}. \tag{52}$$

Now, the smoothing steps for the state space given by Equation (38) are stated below following the above two steps. Since it has been assumed that states have Gaussian distribution, the distribution of states is Gaussian, also. That is, if $\phi_{t+1}|y_{1:N} \sim N(\mu_{t+1}, C_{t+1})$, then $\phi_t|y_{1:N} \sim N(\mu_t, C_t)$ for $t = N, N-1, \cdots, 2, 1$, where

$$\begin{aligned} \mu_t &= \phi_t^{est} + C_t^{est} G_{t+1}' R_{t+1}^{-1} (\mu_t - \phi_t^p), \\ C_t &= C_t^{est} - C_t^{est} G_{t+1} R_{t+1}^{-1} (R_{t+1} - C_{t+1}) R_{t+1}^{-1} G_{t+1} C_t^{est}. \end{aligned}$$

Proofs and further information about methods can be found in [1]. Here, we only use these results to estimate states. As it has mentioned above, the state space approach has great advantages such as solving the problems that we faced in Box-Jenkins approach, as well as looking back and reconstructing time series behavior, and flexibility of applying it in wide variety of applications. In some cases, such as *censored (binary) data*, these general formulas cannot be used, because the distribution of errors is not Gaussian, but maybe *Binormial* or *Poisson*. Details and examples can be seen in [17]

7 TIME SERIES ANALYSIS WITH MARKOV CHAIN MONTE CARLO

The estimation methods that have been discussed so far do not give much probabilistic sense for the parameters. Probabilistic explanation helps to understand behavior of the parameters precisely. It can be used to improve the reliability of outcomes of analysis. In *Bayesian inference* framework, the parameters are treated as random variables. Combined with MCMC methodology we can give derive fully probabilistic analysis for the parameters.

7.1 Bayesian Inference Method

The Bayesian inference method facilitates to study the revised belief when the actual belief is given with prior information. Suppose a parameter θ, estimated by data $Y = \{y_1, y_2, \cdots, y_n\}$, is described by a PDF $p(y|\theta)$. The Bayesian philosophy states that the θ can not be determined exactly but uncertainty about the parameter is expressed through probability statements and distributions. Following steps describe the essential elements of the Bayesian inference.

(1) The probability distribution of θ, $p(\theta)$ is known as the prior belief (existing statistical information such as mean, variance, skewness).

(2) Given the observations Y, choose a suitable statistical model $p(y|\theta)$ to describe the distribution of Y given θ, $p(y|\theta)$. This defines the likelihood function.

(3) Update the belief of θ combining information from the prior distribution and data and calculating *posterior distribution* $p(\theta|y)$ using the **Bayes' theorem** as follows

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}, \tag{53}$$

where $\int p(y|\theta)p(\theta)d\theta$ is a normalizing constant. According to the Equation (21), the likelihood function ($L(\theta|y)$) of θ is proportional to $p(y|\theta)$, Equation (53) can be written as $p(\theta|y) = \frac{L(\theta|y)p(\theta)}{\int L(\theta|y)p(\theta)d\theta}$.

Example: Consider the model $Y = g(x, \theta) + \epsilon$ expresses a certain process with $\epsilon \sim N(0, \sigma^2)$. According to the Equation(21),

$$l(y|\theta) \propto \exp\left(-\tfrac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - g(x_t,\theta))^2)\right).$$

Now this result is integrated with the prior information $p(\theta)$ to calculate the posterior of θ, $p(\theta|y)$. The Equation (53) can be used to do it. It is clear that the Bayesian inference methods gives comprehensive explanation for parameters and some advantages of this method are:

(1) It provides a way of combining prior information within a solid decision theoretical framework.

(2) It incorporates past information about the parameters by a prior distribution for future analysis, when new data become available.

(3) The posterior, $p(\theta|y)$ gives fully probabilistic descriptions about θ and it can be used as a prior in future analysis.

All inferences in the Bayesian method utilize the posterior distribution $p(\theta|y)$. In practice, MCMC is used to obtain the posterior distribution and the following Section explains integration of Bayesian inference with the MCMC methodology.

7.2 Markov Chain Monte Carlo (MCMC) Method

Simulation on random variables from *Markov chain* based on *Monte Carlo* methods is called MCMC. The simulations created by MCMC depend only on the previous stage of simulation, so the series has the *Markov property*. This method is very popular and has great flexibility and advantages in many applications.
The MCMC method quite a successful way of calculating the distribution of the model parameters, i.e. the posterior distribution. Sometimes it is difficult by analytically, but it is possible to generate samples from an arbitrary posterior density by MCMC. Several other aspects of Markov chain method also contribute to its success. When the simulation algorithms have been implemented correctly, MCMC guarantees to converge to the target distribution regardless of where the chain has been initialized [6].
Simple and most commonly used algorithm is the *Metropolis algorithm*, but there are other improved versions such as *Metropolis-Hasting algorithm, Gibbs sampling*. Details can be found in [6, 15].

7.2.1 Metropolis Algorithm

American physicist and computer scientist *Nicolas C. Metropolis* invented this method which is simple and practical. It is used to obtain samples from any complicated target distribution of even with high dimension. Suppose we want to obtain N samples from uni-variate distribution with PDF $p(\theta|y)$ and θ^n is the n^{th} sample, already obtained. To use this algorithm, let the initial value be θ^0 and let $q(\theta^{n+1}|\theta^n)$ be a symmetric proposal density. According to [6], the main steps are:

(1) set $n = 0$, choose a starting point θ^0, this can be arbitrary any point as long as $p(\theta^0|y) > 0$.

(2) generate a new sample θ^{new} using a symmetric proposal distribution $q(: |\theta^n)$ (in many cases a Gaussian distribution).

(3) calculate $r = \min\{\frac{p(\theta^{new}|y)}{p(\theta^n|y)}, 1\}$

(4) sample u from the uniform distribution $U(0, 1)$.

(5) set $\theta^{n+1} = \theta^{new}$ if $u < r$; otherwise set $\theta^{n+1} = \theta^n$

(6) set $n = n + 1$ if $n < N$, the number of desired samples, return to step 2 or otherwise stop

Selecting the proposal distribution q is important because if it is not suitable, it would lead to inefficient implementations. In most cases, a Gaussian distribution is used. Because it is symmetric, the probability of moving from the current point θ^n to the proposed point θ^{new} is the same as moving backwards from the proposed point θ^{new} to the current point θ^n. More details can be found in [6, 15].
Finally, it is clear that MCMC can be used to give a comprehensive analysis in the parameter estimation process because the posterior of parameters obtained by MCMC gives full probabilistic representation of the parameters. Therefore, when this is cooperated with time series analysis, it would yield better analysis results such as in study of states of DLMs, reconstruction of underlying behaviour, model fitting, parameter estimation, and forecasting with DLMs as well as ARMA models.

8 SIMULATION EXAMPLES AND REAL WORLD APPLICATIONS

ARMA models and DLMs parameter identifiablility of different estimation methods are discussed with synthetic and real examples. The identification of the parameters is studied by two separate approaches. First, we generate data sets with know parameter values and perform the parameter estimation for each set. The collection of estimated parameters defines the overall behavior of the method and we can study how close the estimated parameters are to the known values. Second, we analyze one fixed data set with MCMC. This provides us with the posterior distribution given one data set and we can check whether the Bayesian analysis of one data set corresponds to that obtained by repeated simulation in the first approach. MATLAB is used to implement simulations with *GARCH* and *dlmtbx* toolboxes.

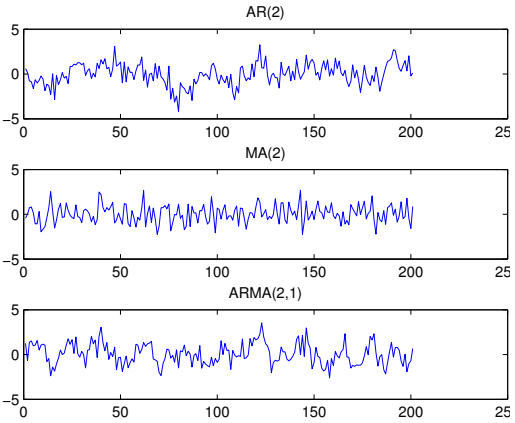
8.1 Simulation Examples

Three different ARMA models are used: AR(2), $Y_t = 0.2Y_{t-1} + 0.3Y_{t-2} + \epsilon_t$, MA(2), $Y_t = \epsilon_t + 0.2\epsilon_{t-1} - 0.2\epsilon_{t-2}$, and ARMA(2,1), $Y_t = 0.2Y_{t-1} + 0.3Y_{t-2} + \epsilon_t + 0.2\epsilon_{t-1}$. We assume that the initial data points for these models are zero and $\epsilon \sim N(0, 1)$. First, pre-analysis is carried out by simulating data samples from these models to understand the model identification. Second, we discuss the contribution of each estimation method in parameter identification of these ARMA models.

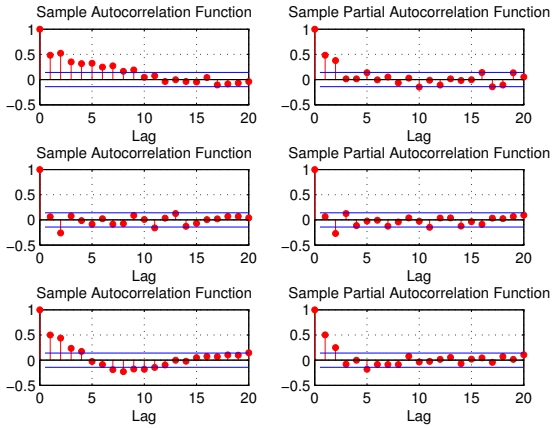
8.1.1 Pre-analysis

Data samples are generated from the above ARMA models. Therefore, the stationarity of data and initial ARMA model order identification have to be studied. Graphical methods are used to check the stationary. Since the model orders are already known, we can confirm the ARMA model order identification statements in Table 1 are true and BIC works well (see Subsection 4.4).

(1) **Graphical method:** Plot the data samples and check the stationary condition. Also check the significant number of lags of ACF and PACF plots given in Figure 4b corresponding to the each data sample. The results are summarized in Table 2.



(a) Data samples



(b) ACFs and PACFs

Table 2: AR(2), MA(2), and ARMA(2,1) models ACF and PACF properties

Model	ACF	PACF
AR(2)	decreasing towards zero	significant until 2^{nd} lag
MA(2)	significant until 2^{nd} lag	decreasing towards zero
ARMA(2,1)	decreasing towards zero	decreasing towards zero

(2) **Calculation of BIC:** BIC is calculated only from MA(2) model for different p and q values and the outcomes are tabulated in Table 3. It is clear that the minimum BIC 2.8417 occurs at $p = 0, q = 2$. It confirms that BIC can also be used to select most suitable model for a given data set.

8.1.2 Matlab *garchfit* Gstimation Method

This function in MATLAB GARCH toolbox is used in time series analysis for *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) processes. Especially, GARCH is used for the time-varying variance (volatility) processes in econometric applications, but it also gives a platform to deal with ARMA models [20]. *Garchfit* estimates ARMA parameters with the initial model specification structure is made by the *garchset* function. Then very precise output specification structure of the estimations is returned with their significance, variance, as well as total adequacy of the fitted model.

First, generate data samples of size 200 from the above ARMA models and then estimate the parameters and the error variance σ^2 by the *garchfit*. Second, the same

Table 3: Estimated BIC from MA(2) data set for 25 different models

(p,q)	0	1	2	3	4	5
0	2.9327	2.8714	2.8417	2.8479	2.85178	2.8552
1	2.8542	2.8586	2.8481	2.8517	2.8563	2.8618
2	2.8564	2.8557	2.8494	2.8551	2.8580	2.8676
3	2.8439	2.8492	28549	2.8613	2.8677	2.8689
4	2.8509	2.8559	2.8616	2.8726	2.8743	2.8797
5	2.85230	2.8633	2.8722	2.8782	2.8809	2.8798

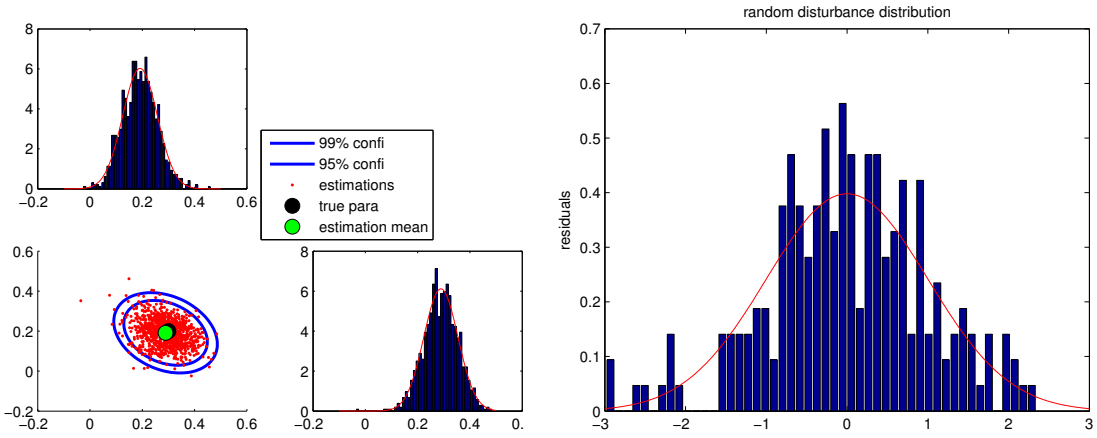
procedure is repeated for 1000 times to check how the estimations are concentrated around their true values. The means of outcomes are given in Table 4 and they are approximately close to their true values in one realization case as well as 1000 realization case.

Table 4: Estimated parameter values of AR(2), MA(2), and ARMA(2,1) models by *garch* function

Model	parameter	One realization	mean of 1000 realizations
AR(2)	$\theta_1 = 0.2000$	0.1698	0.1915
	$\theta_2 = 0.3000$	0.2869	0.2879
MA(2)	$\phi_1 = 0.2000$	0.1991	0.1935
	$\phi_2 = -0.2000$	-0.2347	-0.2017
ARMA(2,1)	$\theta_1 = 0.2000$	0.1770	0.2071
	$\theta_2 = 0.3000$	0.2779	0.2921
	$\phi_1 = 0.2000$	0.2063	0.1880

Pairwise scatter plots of all parameter combinations and a normalized histogram corresponding to the each parameter are shown by Figures 5a, 6a, 7a respectively. The true values of the parameters are close to the means of estimates. Also the normalized histograms of the estimated parameters are distributed around their true values. Moreover, Figures 5b, 6b, and 7b depict that the normalized error estimates of each model has a normal distribution with mean zero and variance one. Therefore, the parameter identification performances are quite good in this approach.

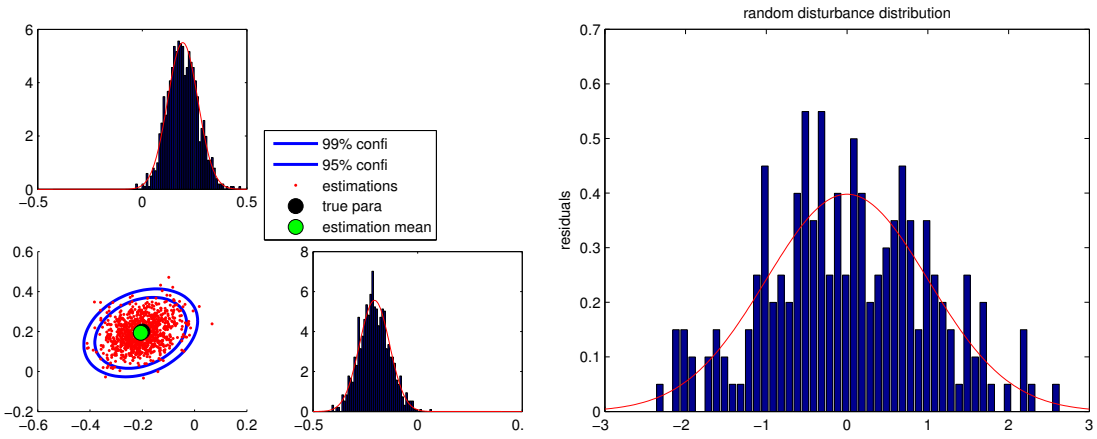
(a) AR(2) estimation



(a) Normalized histograms and scatter plot (b) Normalized histogram of σ^2

Figure 5: AR(2) parameter estimation results from *garchfit* function

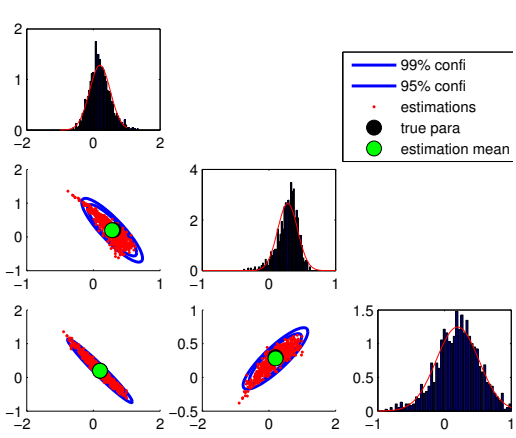
(b) MA(2) estimation

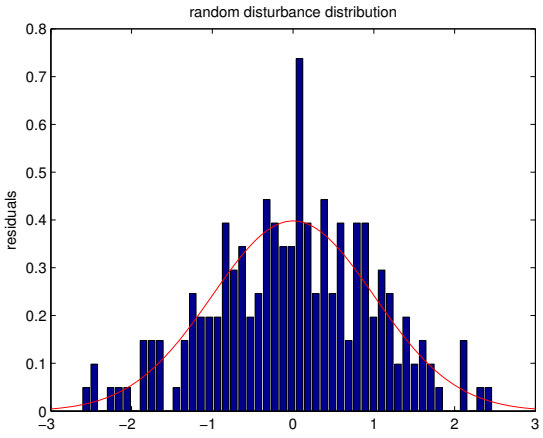


(a) Normalized histograms and scatter plot (b) Normalized histogram of σ^2

Figure 6: MA(2) parameter estimation results from *garchfit* function

(c) ARMA(2,1) estimation





(a) Normalized histograms and scatter plots (b) Normalized histogram of σ^2

Figure 7: ARMA(2,1) parameter estimation results by *garchfit*function

The MSE of the parameters are in the Table (5).

Table 5: MSE of *garchfit* estimated parameters of AR(2), MA(2), and ARMA(2,1) models

Model	parameter	One realization
AR(2)	$\theta_1 = 0.2000$	0.03916
	$\theta_2 = 0.3000$	0.0309
MA(2)	$\phi_1 = 0.2000$	0.0302
	$\phi_2 = -0.2000$	0.0316
ARMA(2,3)	$\theta_1 = 0.2000$	0.1549
	$\theta_2 = 0.3000$	0.0731
	$\phi_1 = 0.2000$	0.1623

8.1.3 ARMA Parameter Estimation

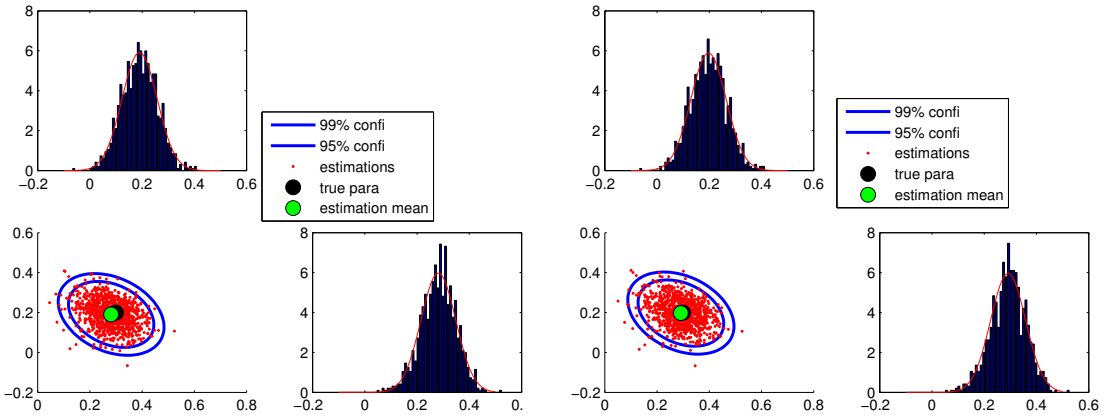
The same procedure of repeated simulation is continued as in earlier section. Since the LSQ and MLE methods are computationally same, we use only MLE and Yule-Walker. Table 6 presents the means of estimated parameters by Yule-walker and MLE methods. When closely looking at the results, estimates are close to their true values. Also the estimated error variance σ^2 close to its true value.

Table 6: Means of estimated parameters of AR(2), MA(2), and ARMA(2,1) from 1000 realizations

Model \ Estimation Method	parameter with true value	Yule-Walker method	MLE method
AR(2)	$\theta_1 = 0.2000$	0.1907	0.2004
	$\theta_2 = 0.3000$	0.2885	0.2985
	$\sigma^2 = 1.000$	0.9789	0.9921
MA(2)	$\phi_1 = 0.2000$	-	0.2010
	$\phi_2 = -0.2000$	-	-0.2040
	$\sigma^2 = 1.000$	-	0.9603
ARMA(2,1)	$\theta_1 = 0.2000$	-	0.2029
	$\theta_2 = 0.3000$	-	0.2905
	$\phi_1 = 0.2000$	-	0.1917
	$\sigma^2 = 1.000$	-	0.9926

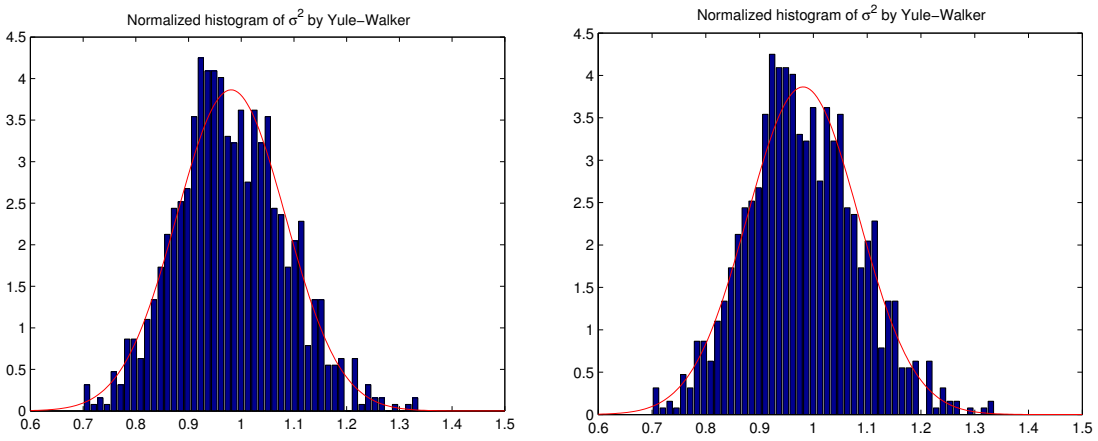
Pairwise scatter plots of estimates given in Figures 8, 10, and 11. Actual parameter values close to the means of their estimates. Also the normalized histograms of the corresponding parameters are also distributed around their true parameter values. Moreover, the normalized histograms of estimated σ^2 from each estimation method are shown in Figures 9, 10b, and 11b are almost close to its true value and hold the normality assumption.

(1) AR(2) parameter estimation



(a) Estimations via Yule-Walker method (b) Estimations via MLE method

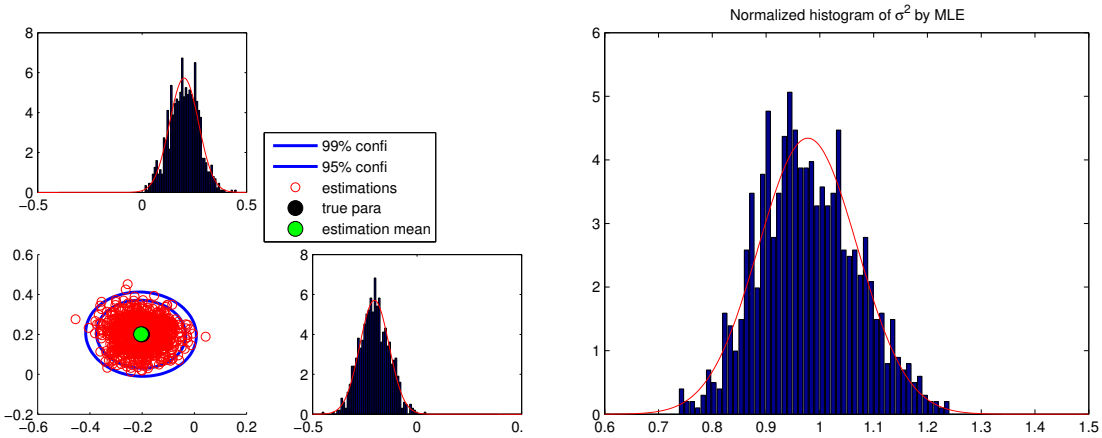
Figure 8: Normalized histograms and scatter plot of AR(2) parameter estimation results via Yule-Walker and MLE methods



(a) σ^2 from Yule-Walker method (b) σ^2 from MLE method

Figure 9: Normalized histograms of σ^2 estimations via Yule-Walker and MLE

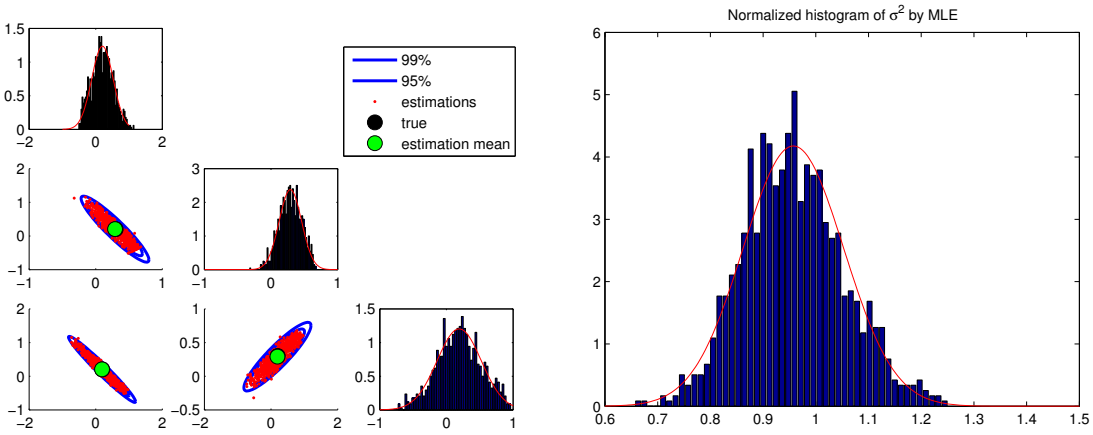
(2) Scatter plots and normalized histograms of MA(2) model parameter estimation



(a) Estimations via MLE method (b) Normalized histogram of σ^2

Figure 10: MA(2) parameter estimation results via MLE estimation results

(3) Scatter plots and normalized histograms of ARMA(2,1) model parameter estimation



(a) Estimations via MLE method (b) σ^2 estimations via MLE method

Figure 11: ARMA(2,1) parameter estimation results from MLE method

Finally, we can say that Yule-Walker, LSQ, and MLE methods performs well in parameter identification by repetition estimation process.

8.1.4 DLM Estimation

First, we transform all the above ARMA model into DLMs and start estimation according to the discussion in the Section 6.3. The estimation is carried out by Kalman filter method once only since the high computational time. In the next subsection, this estimation is integrated with MCMC. Table 7 summarizes the estimates and they are close to the true values. It means that in the DLM approach identifies the parameters quite well.

Table 7: DLM parameter estimations for one sample

Model / parameter	AR(2)	MA(2)	ARMA(2,1)
$\theta_1 = 0.2000$	0.1793	-	0.1969
$\theta_2 = 0.3000$	0.3032	-	0.3001
$\phi_1 = 0.2000$	-	0.2109	0.2126
$\phi_2 = -0.2000$	-	-0.2025	-
$\sigma^2 = 1.000$	0.8959	0.9120	1.0320

8.1.5 MCMC Analysis

MCMC analysis is performed on the DLMs of the above ARMA models and the model parameters are estimated by MLE method. We use MCMC to find the optimal estimates for the parameters. They are obtained by calculating the posterior distributions of each parameter using only one data set. The results are presented as point estimates and in graphical representations to understand how MCMC works with DLM in parameter identification process.
According to the estimation results given in Table 8, posterior means of the parameters are the optimal estimates and they are close to their true values.

Table 8: Means of MCMC chains of ARMA parameters are estimated by DLM estimation method

Model / parameter	AR(2)	MA(2)	ARMA(2,1)
$\theta_1 = 0.2000$	0.17492	-	0.2137
$\theta_2 = 0.3000$	0.30038	-	0.2118
$\phi_1 = 0.2000$	-	0.2014	0.2289
$\phi_2 = -0.2000$	-	-0.1899	-
$\sigma^2 = 1.000$	0.9573	1.0010	1.0275

(a) **AR(2) estimations:** Figure 12 shows the pairwise scatter plot of MCMC estimates and normalized histograms of each parameter from DLM approach. The MCMC means give the optimal parameter values according to the posterior. Also the posterior distributions of the parameters are presented in Figure 13a. Figure 13b presents estimates of the error variance σ^2, which are distributed around its true value.

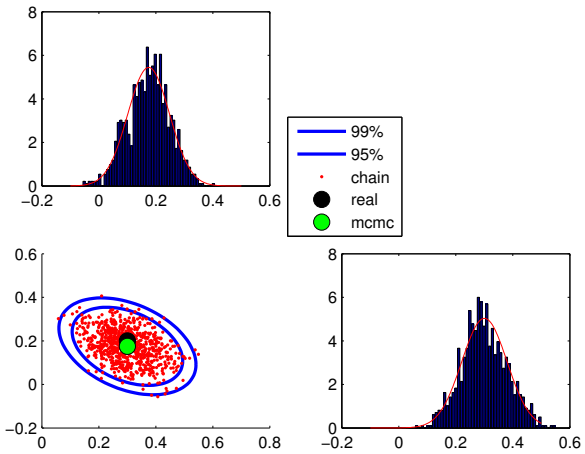
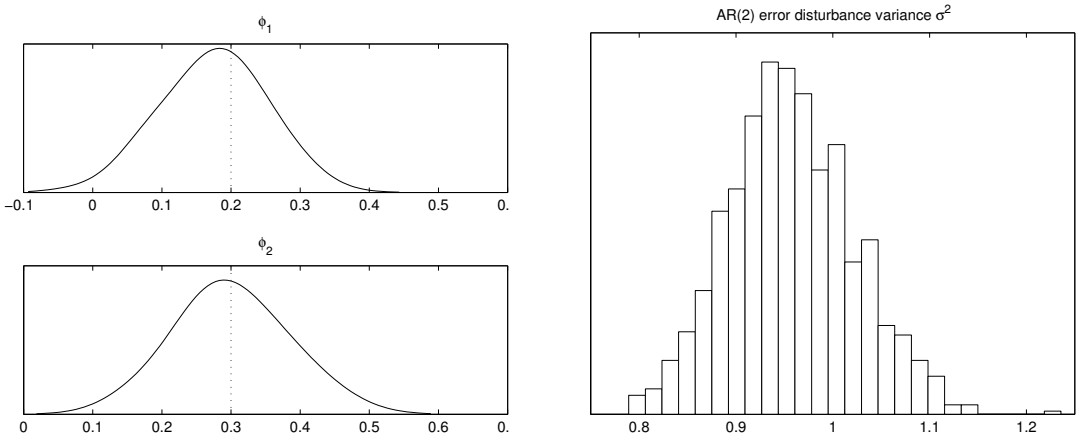


Figure 12: Scatter of MCMC estimates and normalized histograms of θ_1 and θ_2 of AR(2)



(a) posteriors θ_1 and θ_2 (b) histogram of σ^2 estimations

Figure 13: MCMC posterior distribution of θ_1 and θ_2 of AR(2) and histogram of σ^2 estimates by DLM approach

(b) **MA(2) estimations:** Same as in AR(2) estimation case, the MCMC estimates results presented in Figures 14 and 15. The posterior means of the parameters are close to their true values and the uncertainty of estimates are not high.

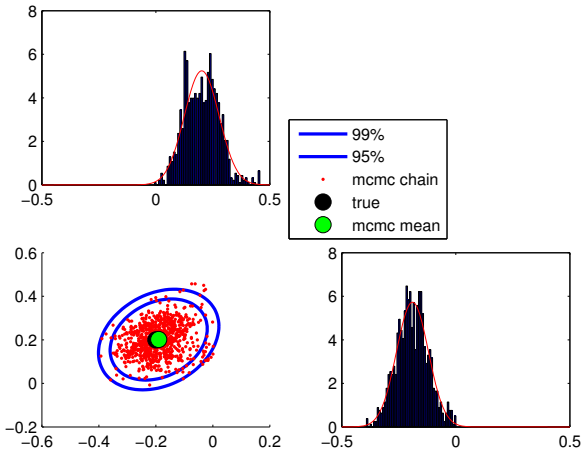
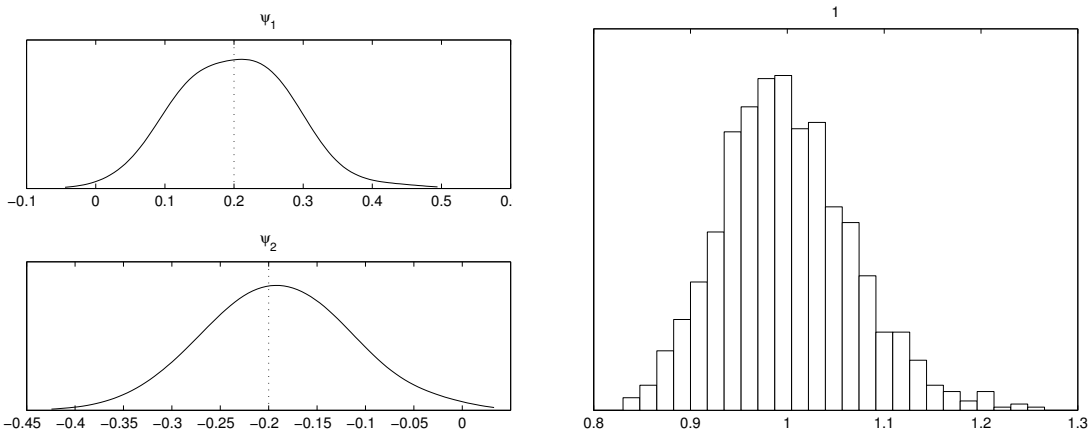


Figure 14: Scatter of MCMC estimates and normalized histograms of ϕ_1 and ϕ_2 of MA(2)

(c) **ARMA(2,1) estimations:** Pairwise scatter plots and normalized histograms of the parameters are presented in Figure 16 with their true values. There is some diverge of MCMC means of estimates from their true values but it is not



(a) posteriors of ϕ_1 and ϕ_2 (b) histogram of σ^2 estimations

Figure 15: MCMC posterior distribution of ϕ_1 and ϕ_2 of MA(2) and histogram of σ^2 estimates by DLM approach

much.

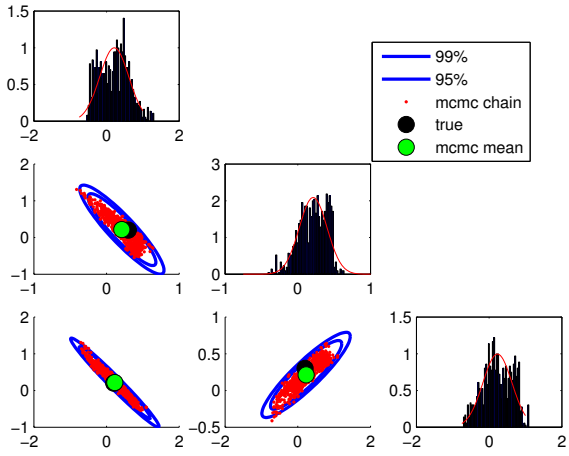
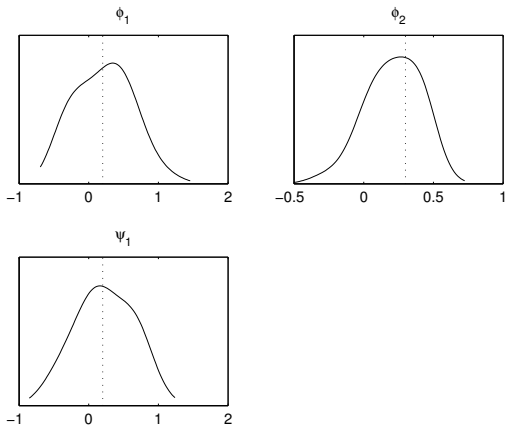
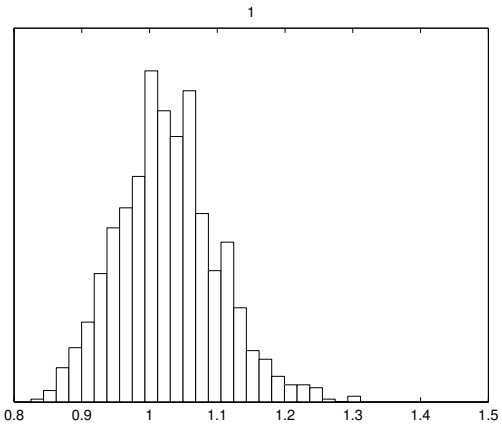


Figure 16: Pairwise of scatter of MCMC estimates and normalized histograms of θ_1, θ_2, and ϕ_1 of ARMA(2,1) with DLM

Figures 17a and 17b show posteriors of the parameters and histogram of the error variance σ^2 estimations respectively and they are closely distributed around their true values. Also the uncertainty of each distribution is not too high.



(a) posteriors of θ_1, θ_2, and ϕ_1



(b) histogram of σ^2 estimations

Figure 17: MCMC results of θ_1, θ_2, ϕ_2, and σ^2 of ARMA(2,1) with DLM

According to the results in Table (8) and Figures from 12 to 17, it seems to be that the parameter identifiability of DLM with MCMC performances are quite good. Also MSE of the each parameter given in Table 9 are small.

Table 9: MCMC yielded MSE of estimated parameters

Model / parameter	AR(2)	MA(2)	ARMA(2,1)
$\theta_1 = 0.2000$	0.0274	-	0.1491
$\theta_2 = 0.3000$	0.0302	-	0.0748
$\phi_1 = 0.2000$	-	0.0283	0.1515
ϕ_2 = -0.2000	-	0.0289	-

Moreover, the other result that can be seen from the MCMC analysis; there is no unique value for a given model parameter and there are many possibilities. For example, the MCMC simulations on the ARMA(2,1) parameters shown in Figure 18, there are many values close to their true values. It means that there could be many possible parameter values to explain same data sample.

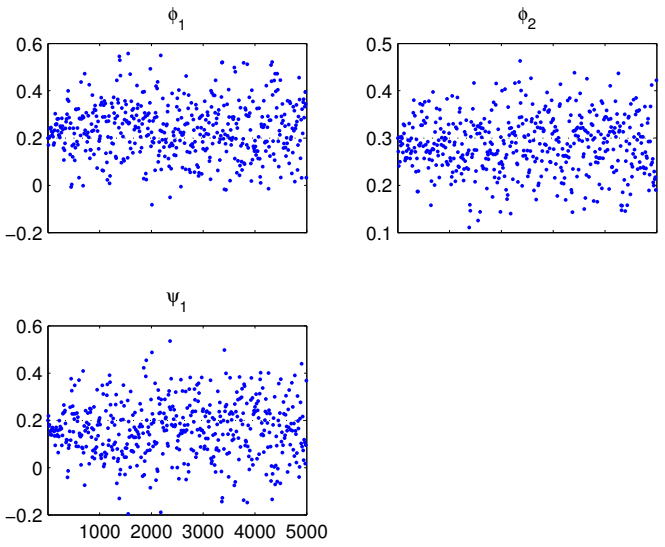


Figure 18: MCMC chains of θ_1, θ_2, and ϕ_2 of ARMA(2,1) with DLM

Finally, all the simulations reveal that the parameter identifiability of these methods differs the way how we carry out the estimation and the estimation method as well. Therefore, the spread of estimates around their true values varies in the scatter plots.

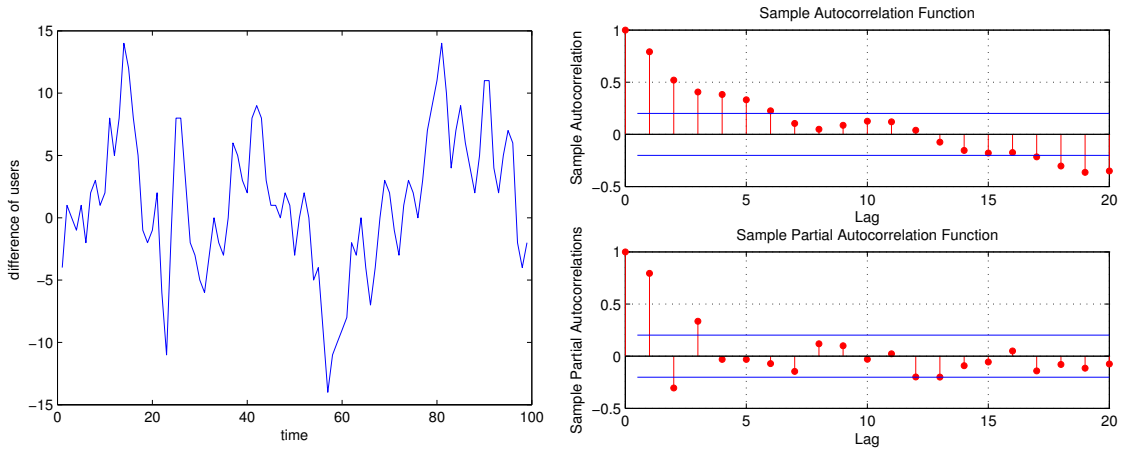
When the estimation is repeated with in new data each time, we get totally independent collection of estimates for each parameters. The means of these estimates are close to their true values. In MCMC analysis, we use only one data sample and generates series of estimates converging towards the most optimum estimator of each parameter. Therefore, the means of MCMC chains give the most suitable estimates and they may differ from their true parameter values. Also true value could be outside from the 95% confidence contour of MCMC points but it happens rarely.

8.2 Real World Examples

Three different real world examples which represent stationary, strong seasonality and trend, and sharp fluctuations respectively, are discussed here. The DLM approach is used to build models and parameter identification is carried out by MCMC.

8.2.1 Internet Server Logging Example

The number of users logged on Internet server were recorded and then the difference of number of users in each minute were calculated. Figure 19a shows the data and it looks stationary. Therefore, ARMA model can be used to analyze this data set and the main steps are given below.



(a) Data

(b) ACF and PACF

Figure 19: ACF and PACF of user logging data on Internet server

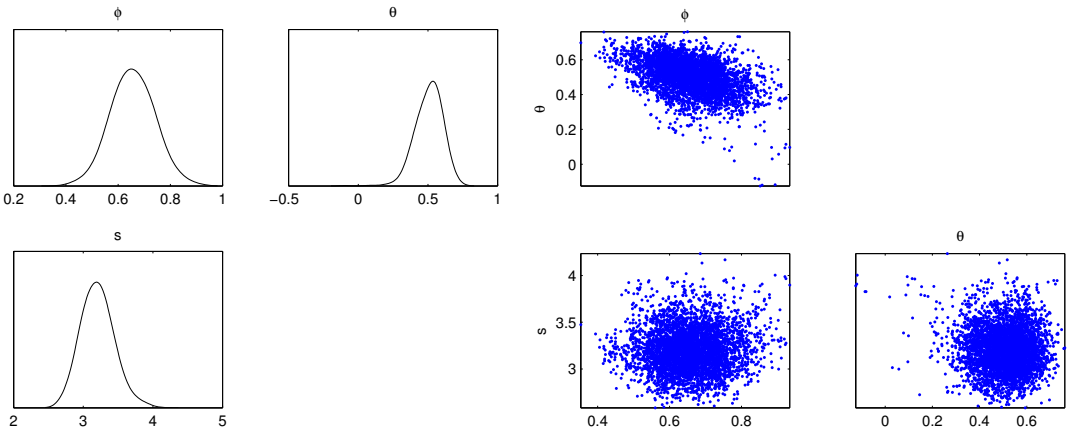
(1) **ARMA model order identification:** First, BIC is calculated for different ARMA(p, q) and the outcomes are tabulated in Table (10). The minimum BIC (5.16302) occurs when $p = 1, q = 1$.

Table 10: BIC for different ARMA models for user logging data on Internet server

(p,q)	0	1	2	3	4	5
0	6.399	5.6060	5.3299	5.3601	5.3773	5.3983
1	5.3983	5.2736	5.3195	5.3288	5.3603	5.3984
2	5.3532	5.3199	5.3569	5.3671	5.3677	5.4089
3	5.2765	5.3224	5.3656	5.4071	5.4039	5.4563
4	5.3223	5.3675	5.4130	5.4529	5.4450	5.4997
5	5.3222	5.3675	5.4130	5.4529	5.4450	5.4997

Also the ACF and PACF plots given in Figure 19b indicates that ACF and PACF are significant till 1^{th} lag. Therefore the best fit model is selected as ARMA(1,1).

(b) **MCMC parameter estimation:** The parameter estimation is carried out by DLM with MCMC. The parameter estimates are $\phi_1 = 0.65623, \theta_1 = 0.50662, \sigma_\epsilon^2 = 3.2044$. Also the posterior distributions and pairwise scatters of estimates of the parameters are shown in Figure 20. Moreover, the scatter plots of θ_1 and ϕ_1 with σ_ϵ^2 do not depict correlation means that σ_ϵ^2 estimates are independent.



(a) Posteriors of $\theta_1, \phi_1, \sigma_\epsilon^2$ (b) pair wise scaters of $\theta_1, \phi_1, \sigma_\epsilon^2$

Figure 20: MCMC parameter estimation results

Moreover, different ARMA models were fitted on the data and checked how DLM works in parameter identification with MCMC. The summarized results are given in Table 11. Of course, when model order is being increased, the likelihood is increasing and model over fit the data.

Table 11: Estimates of four different ARMA model parameters

	ARMA(2,1)	ARMA(3,1)	ARMA(1,2)	ARMA(1,3)	ARMA(2,3)
ϕ_1	0.5961	1.0548	0.6743	0.8136	0.6280
ϕ_2	0.0446	-0.5528	0.4870	-	0.1706
ϕ_3	-	0.2998	-	-	-
θ_1	0.5414	0.1040	-0.0191	0.3848	0.5685
θ_2	-	-	-	-0.2017	-0.1816
θ_3	-	-	-	-0.2143	-0.2741

Also Table 12 presents the correlation between the ARMA(2,3) parameters. We can see that the AR parameters are strongly correlated with the first MA

parameter than the other two. It means that the best fit model order should be less than this model order.

Table 12: Correlation between ARMA(2,3) model parameters

	ϕ_1	ϕ_2	θ_1	θ_2	θ_3
ϕ_1	1.0000	-0.9491	-0.9438	-0.7191	-0.0892
ϕ_2	-0.9491	1.0000	0.8677	0.5247	-0.0730
θ_1	-0.9438	0.8677	1.0000	0.7767	0.0884
θ_2	-0.7191	0.5247	0.7767	1.0000	0.5319
θ_3	-0.0892	-0.0730	0.0884	0.5319	1.0000

(c) **Prediction:** Predictions via MCMC analysis given Figure 21 seems that the prediction lines covers the behavior of the original data well.

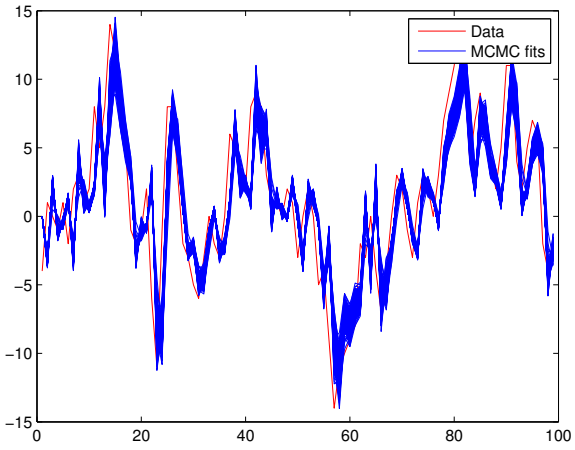


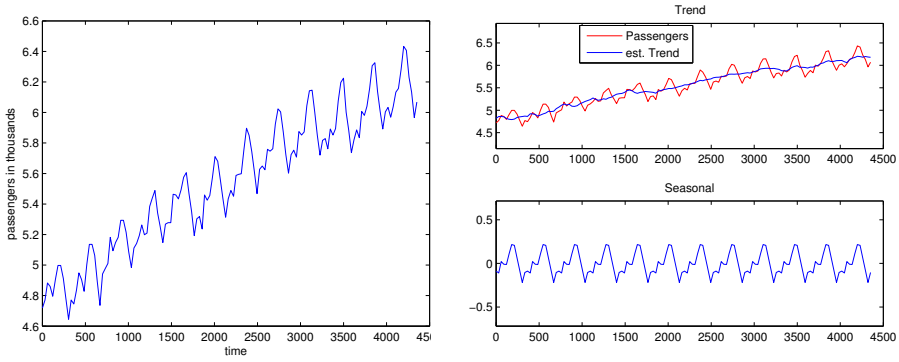
Figure 21: ARMA(1,1) predictions for user logging on Internet server data

Finally, we can say that the estimation method has been identified the model parameters quite well.

8.2.2 Airline Passenger Example

The data set given in Figure 22a is about the airline passengers in thousands from January 1949 to December 1960 was taken from, *Hyndman, R.J., time series data*

library. According to Figure 22b, the data depicts strong trend and seasonal behaviour The basic model for the data set and models for its components are given



(a) Airline passenger data (b) Trend and seasonality

Figure 22: Airline passenger data form January 1949 to December 1960

below.

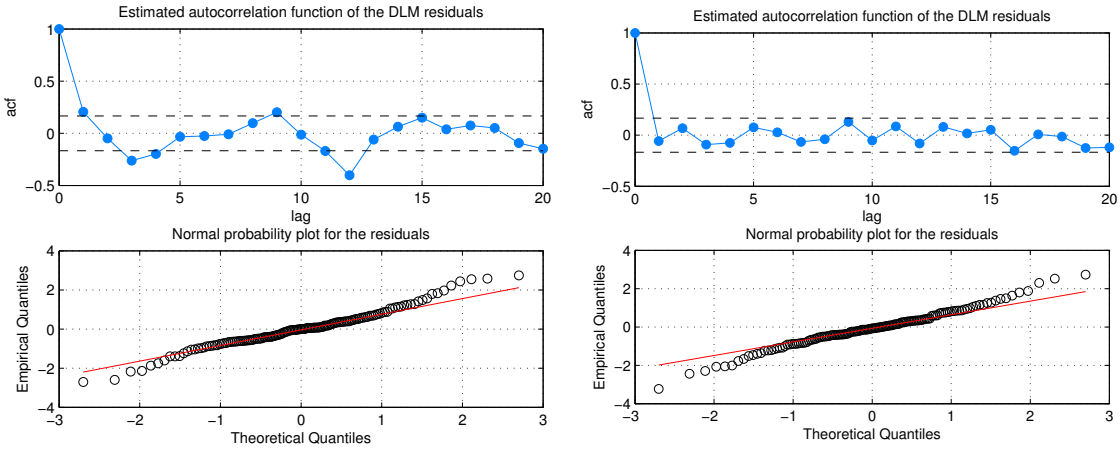
$$y_t = \mu_t + \gamma_t + \epsilon_t, \tag{54}$$

where

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \zeta_t, \\ \gamma_t &= \sum_{j=1}^{s-1} \gamma_{j,t-1} + \omega_t. \end{aligned}$$

indicate the level, trend, and seasonal components, respectively. $\epsilon_t, \eta_t, \zeta_t$, and ω_t are random disturbances with zero mean and variances $\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\zeta^2$, and σ_ω^2 respectively and these are the model parameters have to be estimated, $s = 12$ indicates number of seasons.

First, above model was fitted to the data but the residual diagnosis given in Figure 23a did not look optimal. This may happens because of non-modeled feature has some unknown effect with correlated noise [16]. Therefore, AR(1) component is added into the model. The new model fits well since the residual diagnosis given in Figure 23b seem to be optimum. And the MLE estimates of the parameters are $(\sigma_\epsilon, \sigma_\eta, \sigma_\omega, \sigma_{\phi_1}, \phi_1) = (1.0312, 0.0004, 0.0020, 0.0239, 0.8056)$.

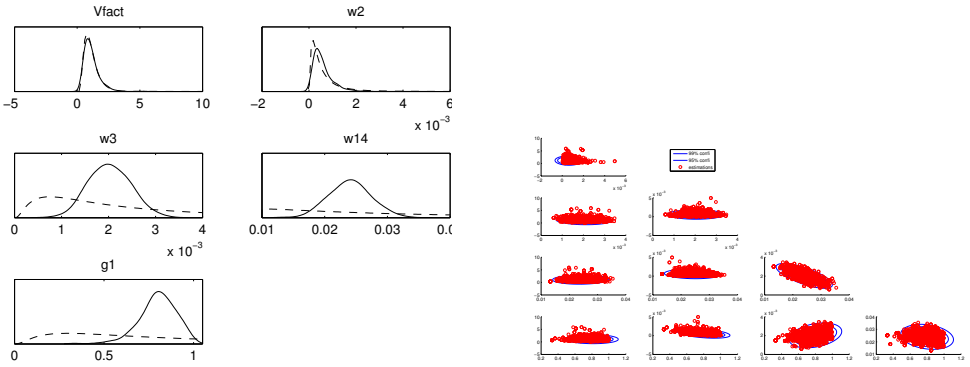


(a) Residual diagnosis without AR

(b) Residual diagnosis with AR(1)

Figure 23: Residual diagnosis of the airline data

Posterior distributions of the parameters given in Figure 24a looks well compared to their priors. It means that there is some relation between the parameter values and the observations. Also pairwise scatter plots given in Figure 24b) show how the parameters are related each other. Table 13 quantifies the correlation between the parameters and we can see that σ_ω and σ_{ϕ_1} are highly correlated compared to all the other parameters. It means that there is an effect of the AR component on the seasonal behaviour of the data.



(a) Prior with posterior of the parameters

(b) Pairwise scatters of parameters

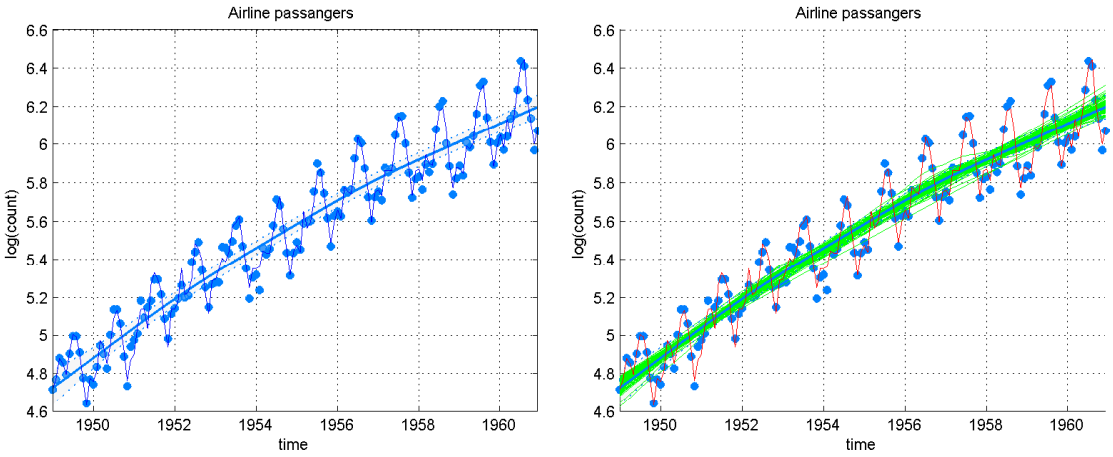
Figure 24: MCMC analysis results of the parameters

In order see the performances of the estimates in the predicting process, data and trend component were predicted from the estimated model. Figures 25a and 25b

Table 13: Correlation between model the parameters

	σ_ϵ	σ_η	σ_ω	σ_{ϕ_1}	ϕ_1
σ_ϵ	1.0000	-0.0190	-0.0191	-0.0107	-0.0390
σ_η	-0.0190	1.0000	-0.0135	-0.0739	-0.3885
σ_ω	-0.0191	-0.0135	1.0000	-0.7060	0.3428
σ_{ϕ_1}	-0.0107	-0.0739	-0.7060	1.0000	-0.2895
ϕ_1	-0.0390	-0.3885	0.3428	-0.2895	1.0000

show the prediction results and we can see that they cover the underlying behaviour of the data well.



(a) DLM fit(blue line) with data (blue dot) and trend (solid blue line)

(b) Simulated samples(green) of the trend component

According to the above results, the estimates perform well in different analysis steps such as model fitting, component analysis, and simulations. Therefore, it can be said that DLM with MCMC has been identified the model parameter well and gives much more flexibility in analysing time series which have strong trend and seasonality without applying any modification on the original data.

8.2.3 Seat Belt Example

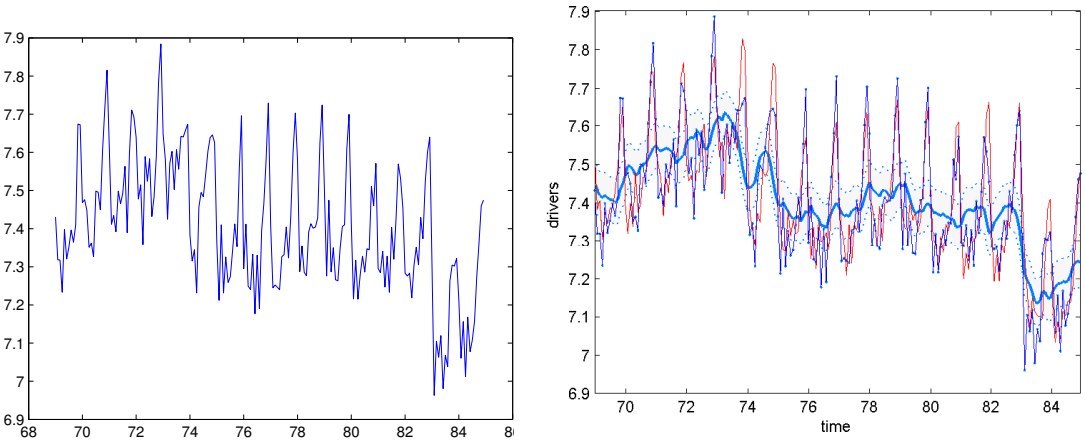
This example is about the road accidents in United Kingdom from January 1969 to December 1984. Many factors were taken into account, but here we consider only two factors. The number of divers who killed or seriously injured in road accidents and the monthly petrol prices. These two series are used to demonstrate the parameter identification of uni-variate structural time series by DLM approach. More details about this example can be found in [3, 17].
Figure 26a shows divers series and the overall trend. The trend looks constant but it breaks around 1974 and 1983 with seasonal pattern as well. The full DLM given by Equation (55) is used to analyse the drivers data with the petrol prices.

$$y_t = \mu_t + \gamma_t + \beta x_t + \epsilon_t, \tag{55}$$

where the level and seasonal parts are modeled as

$$\begin{aligned} \mu_t &= \mu_{t-1} + \eta_t, \\ \gamma_t &= \sum_{j=1}^{s-1} \gamma_{j,t-1} + \omega_t, \end{aligned}$$

respectively, and $\epsilon_t, \eta_t, \omega_t$ are the error components with zero mean and variances $\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\omega^2$ respectively and these are the model parameters. The price of petrol is included as a regression and intervention component βx_t, $s = 12$ is the number of seasons.
First, the analysis is carried out with the seasonal and level components and second, include the petrol price factor to the model to check how well the model covers general behaviour of the data.



(a) Drivers series

(b) DLM fit(red line) with data (blue dot and line) and level (solid blue line)

Parameter estimates are $\sigma_\epsilon, \sigma_\eta, \sigma_\omega = 1.8856, 0.0320, 0.0005$ and Figures 26b and 27 show the DLM predicted data with general behaviour (level), smoothed, and predictive residuals, respectively. Since the estimated level component follows the general pattern well except at the sharp breaks, residuals are quite high at those points only. Also the spread of simulated samples from the level components given in Figure 28a

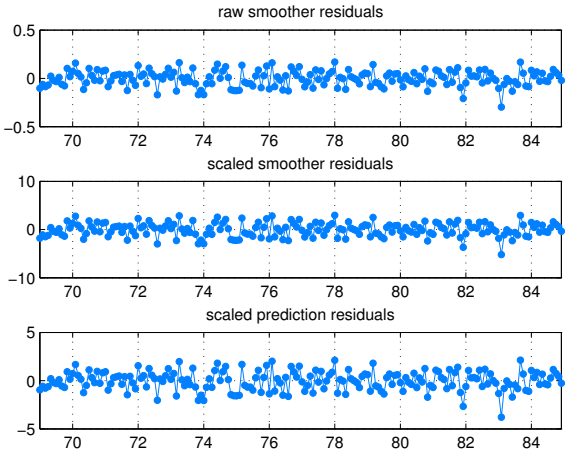
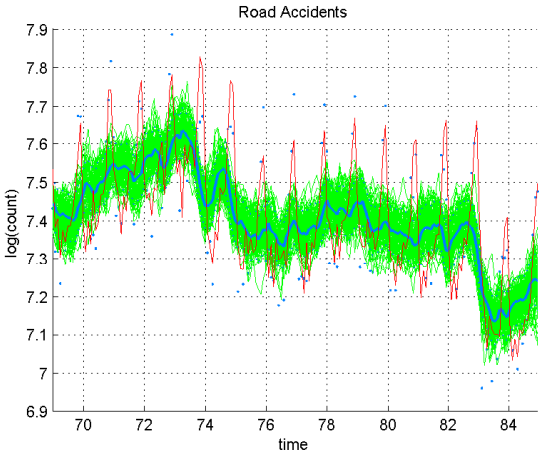
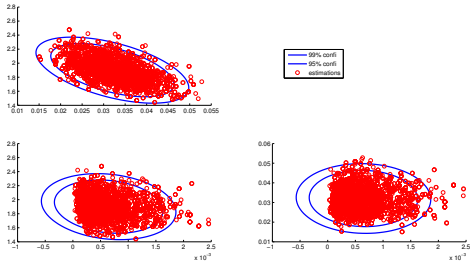


Figure 27: One step ahead residuals from smoothing and prediction

is quite high. It means that the posteriors have not been well identified distribution of the parameters. Moreover, Figure 28b shows pairwise scatters of MCMC chains of the parameters and some estimations are outside from the confidence contours. Therefore the estimated parameters do contribute well in the model prediction process except at the break points. The reasons for these two breaks around mid 1974 and 1984 were the introduction of seat belt laws and the increase of petrol prices



(a) simulated samples of level (green)



(b) Pairwise scatter plots of the parameters

respectively. Therefore, including the petrol price as a regression and intervention variable to the above model, analysis can be continued further to get better result. Since the '*dlmtbx*' toolbox has no direct facility to model intervention variables, analysis is stopped at this stage. The '*ssm-1.0.1*' Matlab toolbox has that facility and detailed description for this example can be found in [4].

9 DISCUSSION

The aim of this thesis was to study the parameter identifiability of different parameter estimation methods which are used in time series analysis. In order to build a sufficient theoretical basis for this study, time series analysis approaches were categorized into three main approaches classical, Box-Jenkins, and state space to study them separately. In addition to that, MCMC analysis concepts in the parameter estimation processes were discussed to present estimation results in probabilistic sense and more informative ways.

In some cases, ARMA models cannot be used because it is difficult to achieve the stationarity for some data. DLMs with Kalman filter method facilitate to analyze many different time series and data is not necessary to be stationary. Furthermore, the basic ARMA model identification methods: ACF, PACF, and BIC have been addressed in this study. These methods returned the expected outcomes from the pre-selected ARMA models. Therefore, these methods can be recommended to use in ARMA model identification.

Basically, ARMA model parameter identifiability was studied with Yule-Walker, LSQ, and MLE methods. Yule-Walker is an efficient method to estimate AR parameters but not for ARMA and MA models whereas LSQ and MLE can use for any ARMA model. Also LSQ and MLE are computationally same under the Gaussian assumption of errors. Moreover, LSQ can only be used in simulation cases because the error term in ARMA models is not observed in practice and it add stochasticity to the LSQ objective function. Pseudo likelihood method is an alternative method which can be used to estimate parameters in real cases. When there are many local maximas, MLE method faces some problems such as flat likelihood. We have to how likelihood values change with different initial parameters.

First, parameter identifiability of the above estimation methods was discussed while repeating the estimation for several times with simulated data sets were generated by known AR(2), MA(2), and ARMA(2,1) models. Since data change in each realization, estimates also change every time. Finally, we get mutually independent series of estimates concentrated around their true values. Even though means of estimates approximate the true parameter values, these estimates suffer with uncertainty. Furthermore, these series of estimates do not provide much probabilistic sense for the parameters.

The parameter identification by DLM with MCMC analysis gives very good outcomes. The difference from the repeating case is that the MCMC means of the parameters are not always close to their true values. Also the posteriors of the parameters represent them probabilistically well and not always symmetric around

their true values. The reason is MCMC returns the most optimum estimations and they are not just a collection of single estimations as in earlier case. Since DLMs allows to include all time variant parameters, the estimation results are more informative and has more precise interpretation to understand clearly the underlying phenomena of time series.
Since DLM with MCMC performs well in parameter identification with simulated data and offers much more flexibility, this estimation approach was applied on three different real data sets. DLM facilitates to combine classical decomposition of time series and characterize the time variant components by allowing them to change over time to study their behaviours well. The posteriors and scatter plots depict the distributions and relationships of the parameters, respectively. Moreover, the predictions of data and simulations on components approximate the original behaviours well. It means that DLM with MCMC has been identified model parameter well.

10 CONCLUSIONS

According to the simulation and real world examples results, we can say that DLM modeling of time series together with MCMC for parameter estimation performs well and offers much flexibility in the parameter identification process while giving fully probabilistic representation for the parameters. On the other hand, since the recursive identification also gives good approximations, but it does not fully contribute explain the given process well. Furthermore, the DLM approach handles the challenges arising in the Box-Jenkins approach and also provides extensive explanation for the time variant factors. In addition, the classical approach makes good basis to start well organized analysis and to reach desired goals more effectively.
There are some future directions that come up from this study. When the Gaussian assumption is not possible, as for binary data, the Kalman filter formulation has to be changed in an appropriate way. The *particle filtering* approach is one way to handle such cases. Since bi-variate cases are common in real world, it is a way to continue this study further, starting with the basic background described here. Also, if time series consists of interval valued data, these estimation methods need to be adjusted to identify parameters to carry out analysis results. Finally, when there is stochasticity in the LSQ estimation cases, the pseudo likelihood estimation method is a popular approach to estimate the parameters. Therefore, these extensions could be studied with the state space approach as well.

References

[1] Petris, G., Petrone, S. and Campagoni, P. (2009) *Dynamic Linear Models with R*, Springer Science and Business Media, New York

[2] Gerolimetto, M., (2010) *Introduction to time series analysis* available at *http://www.dst.unive.it/ margherita/TSLectureNotes.pdf*

[3] Durbin, J., koopman, S. J. (2012) *Time Series Analysis by State Space Methods*, Second edition, Oxford University press, United Kindom.

[4] Harvey, A. C., Shephard, N. (1993) *Structural Time Series Models*, In *Handbook of Statistics*, Vol. 11, Elsevier Science Publishers

[5] Kavalieris, L., Hannan, E. J., and Salalu, M. (2001) *Lecture note for generalized least squares estimation of ARMA models*, University of Otago, Australian National University, and Nigerian Educational Research and Development Council

[6] Solonen, A., Haario, H. and Laine, M. (2009) *Lecture note for Statistical Analysis in Modeling*, Lappeenranta University of Technology and Finnish Meteorological Institute

[7] Plosser, C. I., Schwert, G. W (1977). Estimation of a non-invertible moving average process, *Journal of Econometrics 6*, pp. 199-224

[8] Mathew, J., Patra, P., Pradhan, D.K., and Kuttyamma, A. J. (2102) *Eco-friendly Computing and Communication Systems and Communications in Computer and Information Science*, Springer, International Conference, ICECCS 2012, Kochi, India

[9] Rajan, M.P., Mathew J. (2012) Kalman Filter and Financial Time Series Analysis, *ICECCS 2012, CCIS 305*, pp. 339-359

[10] Kalman, R. E. (1960) A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME-Journal of Basic Engineering*, 82 (Series D), pp. 35-45

[11] Thomas, S. (2009) *Lecture note for estimating AR/MA models*

[12] Zivot, E. (2005) *Lecture note for estimation of ARMA models*

[13] Madsen, H.,(2012) *Lecture note for time series analysis, informatics, and modelling*, Technical University of Denmark, Lyngby

[14] Model Selection (The Bayesian Information Criterion) available at *http://myweb.uiowa.edu/cavaaugh/mslec5ho.pdf*

[15] Sailon, I. (2009) *Lecture note for MCMC analysis of classical time series Algorithms*, Department of mathematics and Physics, Lappeenranta University of Technology

[16] Laine, M. (2014) *Dynamic Linear Model Tutorial* available at *http://helios.fmi.fi/ lainema/dlm/dlmtut.html*

[17] Peng, J., Aston, J. A. D. (2011) The state space models toolbox for Matlab, *Journal of Statistical Softweare*, voulme 41

[18] Laine, M. (2013). MCMC toolbox for Matlab available at *http://helios.fmi.fi/ lainema/mcmc/*

[19] http://stats.oecd.org/glossary/detail.asp?ID=6693

[20] GARCH user guide available at *https://pokrovka11.files.wordpress.com /2014/12/garch matlab toolbox2*

List of Tables

1 ARMA model order identification statements by ACF and PACF . . 19

2 AR(2), MA(2), and ARMA(2,1) models ACF and PACF properties . 40

3 Estimated BIC from MA(2) data set for 25 different models 41

4 Estimated parameter values of AR(2), MA(2), and ARMA(2,1) models by *garch* function 41

5 MSE of *garchfit* estimated parameters of AR(2), MA(2), and ARMA(2,1) models 43

6 Means of estimated parameters of AR(2), MA(2), and ARMA(2,1) from 1000 realizations 44

7 DLM parameter estimations for one sample 47

8 Means of MCMC chains of ARMA parameters are estimated by DLM estimation method 48

9 MCMC yielded MSE of estimated parameters 51

10 BIC for different ARMA models for user logging data on Internet server 53

11 Estimates of four different ARMA model parameters 54

12 Correlation between ARMA(2,3) model parameters 55

13 Correlation between model the parameters 58

List of Figures

1 Trend and seasonal behaviours of beer selling process 12

2 Stationary version of non-stationary time series is obtained by log transformation and differencing 15

3 ACF and PACF of ARMA(3,5) 16

5 AR(2) parameter estimation results from *garchfit* function 42

6 MA(2) parameter estimation results from *garchfit* function 42

7 ARMA(2,1) parameter estimation results by *garchfit*function 43

8 Normalized histograms and scatter plot of AR(2) parameter estimation results via Yule-Walker and MLE methods 45

9 Normalized histograms of σ^2 estimations via Yule-Walker and MLE . 45

10 MA(2) parameter estimation results via MLE estimation results . . . 46

11 ARMA(2,1) parameter estimation results from MLE method 46

12 Scatter of MCMC estimates and normalized histograms of θ_1 and θ_2 of AR(2) 48

13 MCMC posterior distribution of θ_1 and θ_2 of AR(2) and histogram of σ^2 estimates by DLM approach 49

14 Scatter of MCMC estimates and normalized histograms of ϕ_1 and ϕ_2 of MA(2) 49

15 MCMC posterior distribution of ϕ_1 and ϕ_2 of MA(2) and histogram of σ^2 estimates by DLM approach 50

16 Pairwise of scatter of MCMC estimates and normalized histograms of θ_1, θ_2, and ϕ_1 of ARMA(2,1) with DLM 50

17 MCMC results of θ_1, θ_2, ϕ_2, and σ^2 of ARMA(2,1) with DLM 51

18 MCMC chains of θ_1, θ_2, and ϕ_2 of ARMA(2,1) with DLM 52

19 ACF and PACF of user logging data on Internet server 53

20 MCMC parameter estimation results 54

21 ARMA(1,1) predictions for user logging on Internet server data . . . 55

22 Airline passenger data form January 1949 to December 1960 56

23 Residual diagnosis of the airline data 57

24 MCMC analysis results of the parameters 57

27 One step ahead residuals from smoothing and prediction 60