BAYESIAN FORECASTING MODEL FOR INFLATION DATA

ZUL AMRY

UNIVERSITI SAINS MALAYSIA

2015

BAYESIAN FORECASTING MODEL FOR INFLATION DATA

by

ZUL AMRY

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

September 2015

ACKNOWLEDGEMENTS

In the nama of ALLAH, Most Gracious, Most Merciful.

First of all, I would like to thank ALLAH SWT for granting me health, patience and spirit to complete this doctoral study. I wish to thank all the people that contributed, supported, inspired, corrected this work, and the people with whom I shared my time during my Ph.D. For sure I will not be able to mention them all, but everyone has my sincere gratitude.

Especially, I would like to thank my main supervisor, Assoc. Prof. Adam Bin Baharum, for his guidance, the time that he dedicated to me, supporting my work, solving problems related with this work, and reviewing the thesis, encouragement and friendship during my study at the Universiti Sains Malaysia. I wish to express my gratitude to Dr. Mohd Tahir Bin Ismail, co-supervisor, for his guidance, discussions and friendship.

Finally, I would like to thank my children, Sisti Nadya Amelia and Marwan Faiz Hilmi for their help and praying to me. A very special thanks to my wife Susi Setiawati, S.Sos, M.AP, for her love, moral support, and encouragement over the years.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

AR	:	Autoregressive
MA	:	Moving Average
ARX	:	Autoregressive with Exogenous
ARMA	:	Autoregressive Moving Average
ARIMA	:	Autoregressive Integrated Moving Average
ARMAX	:	Autoregressive Moving Average with Exogenous
TARMA	:	Treshold Autoregressive Moving Average
AIC	:	Akaike Information Criterion
ACVF	:	Autocovariance function
ACF	:	Autocorrelation Function
PACF	:	Partial Autocorrelation Function
Pdf	:	Probability density function
RMSE	:	Root Mean Square Error
X	:	Random variable
$\underline{X}$	:	Random vector

MODEL PERAMALAN BAYESIAN PADA DATA INFLASI

ABSTRAK

Banyak data dalam masalah ekonomi boleh dibincangkan dengan model siri masa, seperti perbincangan data inflasi dengan menggunakan model ARMA. Tesis ini memberi tumpuan kepada perbincangan mengenai model ARMA dengan pendekatan Bayesian pada data inflasi. Masalah utama dalam analisis peramalan model ARMA adalah bagaimana untuk menganggar parameter dalam model. Di dalam statistik *Bayes* parameter dipandang sebagai kuantiti daripada pemboleh ubah rawak dengan variasi yang digambarkan sebagai taburan kebarangkalian dan dikenali sebagai taburan pendahulu. Tujuan tesis ini adalah untuk menentukan penganggar Bayes, peramalan titik dan selang peramalan dengan pendekatan Bayesian pada data inflasi menggunakan model ARMA di bawah andaian pendahulu normal-gamma dan pendahulu Jeffrey dengan fungsi kerugian kuadratik. Metode penyelidikan yang digunakan adalah model peramalan time series dengan pendekatan Bayesian. untuk menyimpulkan model yang didapati sudah memadai, dilakukan penyelidikan terhadap autokorelasi sisa dengan menerapkan statistik Q-Ljung Box, manakala untuk melihat ketepatan model ramalan digunakan RMSE, MAE, MAPE, and U-Statistics. Hasil penyelidikan adalah penganggar Bayes, peramalan titik, dan selang peramalan dalam ekspresi matematik. Selanjutnya, hasil penyelidikan diterapkan pada data inflasi dan dibandingkan dengan metode tradisional. Hasilnya menunjukkan bahwa metode Bayesian lebih baik dari pada metode tradisional. Simulasi untuk beberapa ukuran data juga menunjukkan bahwa metode Bayesian lebih baik dari pada metode tradisional.

BAYESIAN FORECASTING MODEL FOR INFLATION DATA

ABSTRACT

Many of the data in the economic problem can be addressed by the model time series, such discussions inflation data using ARMA model. This thesis focuses on the discussion of the ARMA model with a Bayesian approach on the inflation data. The major problem in the analysis of ARMA model is the estimation of the parameters in the model. In Bayes statistics, the parameters estimated will be viewed as a quantity of random variable which the variation is described by probability distribution and known as the prior distribution. The objective of this thesis is to determine the Bayes estimator, point forecast, and forecast interval with Bayesian approach for inflation data using ARMA model under normal-gamma prior and Jeffrey's prior assumption with quadratic loss function. The method of research which used is time series forecasting model with Bayesian approach. To conclude whether the result of the model is adequate, an investigation was conducted to test the autocorrelation of the residuals by using the Q- Ljung Box statistic, while looking at the accuracy of forecasting model used the RMSE, MAE, MAPE, and U-Statistics. The results of research are the Bayes estimator, point foecast, and forecast interval in mathematical expression. Furthermore, the result of research is applied to inflation data and compared to traditional method. The results shows that the Bayesian method is better than the traditional method. The simulation for some of the data size shows also that the Bayesian method is better than the traditional method.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Many of the data in the economic problem can be addressed by the model time series, such discussions inflation data using ARMA model. Inflation is the indicator of price developments of goods and service that are consumed by society. In economics the inflation rate is a measure of inflation, or the rate of increase of a price index such as consumer price index. Inflation forecasts play an important role to effectively implement an inflation targeting regim (Svensson, 1997). Moreover, many economic decisions, whether made by policymakers, firms, investors, or consumers, are often based on inflation forecasts, and the accuracy of these forecasts can thus have important repercussions in the economy (Ramirez, 2010). Forecasting inflation with a time series model can be seen in some countries, as in Austrian (Virkun and Sedliacik, 2007), Slovenian (Stovicek, 2007), Pakistan (Salam and Feridun, 2007), China (Mehrotra and Fung, 2008), Sudan (Moriyama and Naseer 2009), Mexico (Ramirez, 2010), Turkey (Saz, 2011), Nigeria (Olajide, 2012), Bangladesh (Faisal, 2012), and Angola (Barros and Alana, 2012).

Autogeressive Moving Average (ARMA) model is the particular form of Autoregressive Integrated Moving Average (ARIMA) models which are developed by Box & Jenkins and have been extensively used in many fields as statistics models, particular in connection with problem of forecast. Forecast of time series model is a forecast using the observations of past data, it performed investigations to extrapolate future values of series, whereas the major problem in the analysis of ARMA model is how to estimate the parameters in the model.

In the estimation theory, there are two popular approaches, namely the classical statistics approach and Bayesian statistics approach. Classical statistics is fully determined by the inferential process based on sample data from the population. In contrast, Bayesian statistics are not only used the sample from population but also employ an initial knowledge of each parameter. The issue of estimation theory is the problem that continues to grow. In classical statistics, the parameters of population estimated will be viewed as an unknown fixed quantity, while in Bayesian statistics, the estimated parameters will be viewed as a quantity of random variable in which the variation is described by the probability of distribution and known as the prior distribution. This distribution represent the initial knowledge about the parameters before the observation of a sample is taken and it can be used as a mathematical tool to obtain the best estimator. A prior distribution supposed to represent what we know about unknown parameters before the available data, plays an important role in Bayesian analysis to get a final decision, by multiplying the prior distribution to information of data in the form of the likelihood function is obtained the posterior distribution which will be used in the inference. In practice, it is always not easy to obtain the posterior distribution of any likelihood function with prior distribution appropriate; sometimes there is a difficult mathematical form to be completed analytically. To overcome this problem, statisticians have limited the prior distribution in families of the specific distribution based on the likelihood function that also known as conjugate prior whereas Jeffrey suggested a prior distribution which is mathematically constructed based on the likelihood function that known as Jeffrey's prior. The estimator of the parameters of a distribution model obtained by Bayesian analysis is known as Bayes estimator.

1.2 Problem Statement

The classical forecasting have been depeloved by Box and Jenkins (1976). There are three steps accomplished in the process of fitting the ARMA(p,q) model to a time series: identification of the model, estimation of the parameters, and model checking to conclude whether the models obtained are adequate for forecasting. The main difference between the Bayesian approach and the classical approach is that in the Bayesian approach, the parameters supposed as random variables, which are described by their probability density function, whereas the classical approach considers the parameters to be fixed but unknown.

Bayesian forecasting encompasses statistical theory and methods in time series analysis and time series forecasting. Main idea of Bayesian forecasting is the predictive distribution of the future given the past data follows directly from the joint probabilistic model. Predictive distribution is derived from the sampling predictive density, weighted by the posterior distribution. For Bayesian forecasting problems, Bayesian analysis generates point and interval forecasts by combining all the information and sources of uncertainty into a predictive distribution for the future values.

This thesis focuses on the application of mathematical and statistics methods for Bayesian forecasting in the ARMA model, how the rules of mathematics statistics using to determine the formula of point and interval Bayesian forecasts in the ARMA model. The problem in this thesis is "how the mathematic formula of Bayesian multiperiod forecasting for ARMA model under normal-gamma prior and Jeffrey's prior with quadratic loss function".

1.3 Objective

The objective of this research is to determine Bayes estimator of parameters, point estimate and interval estimate for ARMA multiperiod forecast model by using normal-gamma prior and Jeffrey's prior with quadratic loss function to be applied to the inflation data.

1.4 Scope of the study

This research is a study by applying a set of scientific works such as journals, text books, research results and other scientific works in mathematics and statistics related to the research. This research is discussed based on theories of mathematics and statistics in the form of definition, theorem, lemma and its properties. The ARMA(p,q) model used is

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + e_t \tag{1.1}$$

where ϕ_i and θ_j parameters,$\{e_t\}$ is a sequence of *i i d* normal random variables with $e_t \sim N(0, \tau^{-1})$ and $\tau > 0$ unknown. Likelihood function that is used refers to what has been done by Box and Jenkins. Prior distribution that is used are the normal gamma prior and Jeffrey's prior with quadratic loss function.

1.5 Significance of the study

Time series forecasting is a forecasting model which is founded based on past data, to investigate the pattern and extrapolated into the future and Bayesian data analysis can be defined as a method for summarizing uncertainty and making estimates and predictions using probability statements conditional on observed data and an assumed model (Gelman, 2008).

Bayesian method of inference and forecasting all derive from two simple principles, namely *principle of explicit formulation* and *principle of relevant* *conditioning* (Geweke and Whiteman, 2004). In principle of explicit formulation, express all assumption using formal probability statements about the joint distribution of future events of interest and relevant events observed at the time decisions, including forecast, must be made. In principle of relevant conditioning, the forecasting use the distribution of future events conditional on observed relevant events and an explicit loss function.

There have been a lot of works relating to Bayesian analysis and time series forecasting. In this thesis, we combines Bayesian analysis and time series forecasting. This research focuses on the application of mathematical and statistical rules in Bayesian multiperiod forecasting. This problem have discussed by Liu (1995) on the ARX model, whereas in this thesis will be developed by using ARMA model. Others the paper related to ARMA model, normal-gamma prior and Jeffrey's prior are Kleibergen and Hoek (1996) using ARMA model and Jeffrey's prior. Mohamed et al. (2002), also using ARMA model with Jeffrey's prior Fan & Yao (2008) using ARMA model and normal-gamma prior and Uturbey (2006) using ARMA model and inverse gamma prior.

Main idea in this thesis is to form a conditional posterior predictive distribution based on the posterior predictive distribution and conditional distribution. By Integrating the conditional posterior predictive distribution can be obtained the marginal conditional posterior predictive distribution, whereas the result of point forecasts and forecast interval can obtained via the marginal conditional posterior predictive distribution

The importance of this research is to show how the rules of mathematics and statistics are useful in Bayesian forecasting model, especially on the Bayesian multiperiod forecasting for ARMA model using normal-gamma prior and Jeffrey's prior with quadratic loss function.

1.6 Organization of the thesis

This thesis contains five chapters as follows :

Chapter 1 contains the the introduction, problem statement, objective, scope of the study, significance of the study and organization of the thesis.

Chapter 2 contains the study of literature that includes introduction, time series analysis, Bayesian time series analysis, and critical review.

Chapter 3 contains the methodology of research that includes the basics of mathematical statistics, some important distributions, some rules of matrix, Bayesian forecasting, multiperiod forecasting, the forecast accuracy criteria, stages of research, and stages of analysis to determine : likelihood function, normal gamma prior, Jeffrey's prior, posterior distribution, Bayes estimator, Bayes variance, conditional predictive density, conditional posterior predictive density, marginal conditional posterior predictive density, posterior mean and posterior variance, point forecast, interval forecast.

Chapter 4 contains the result of analysis that includes of result of point forecast dan forecast interval, computational procedure, application to one set of real data and simulation to compare of forecasting between the Bayesian method with the traditional methods.

The final chapter contains the conclusions and recommendation that includes the conclusion, summary, contributions, and suggestions for further research.

CHAPTER 2

STUDY OF LITERATURE

2.1 Introduction

This chapter presents study of literature that includes time series modeling, Bayesian time series modeling, literature review related to research in this thesis, and inflation.

In time series modeling are presented the mean and variance, the autocovariance function, stationarity, the autocorrelation function, the partial autocorrelation function, testing of stationarity, the ARMA model, and the Box-Jenkins modeling, In Bayesian time series analysis are presented the Bayesian inference, the likelihood function, the prior distribution, Bayes theorem, and the Bayes estimator with decision theory. In the literature review described summary of the several papers related to research in this thesis.

2.2 The time series analysis

We assume that the time series values we observe are the realisations of random variables $Y_1, Y_2, \ldots, Y_T$, which are in turn part of a larger stochastics process $\{Y_t; t \in Z\}$.

2.2.1 The mean and variance

<u>*Definition 2.2.1*</u> (Wei, 1994)

For a given real-valued process $\{Y_t; t \in Z\}$, the *mean* of the process is

$$\mu = E(Y_t) \tag{2.1}$$

and the *variance* of the process is

$$\sigma^2 = E[(Y_t - \mu_t)^2] \tag{2.2}$$

μ and σ^2 can be estimated from sample data by

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t \tag{2.3}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{t=1}^{n}\left(Y_t - \bar{Y}\right)^2 \tag{2.4}$$

2.2.2 The autocovariance function (ACVF)

Definition 2.2.2 (Wei, 1994)

The *autocovariance* between Y_t and Y_{t+k} is defined as

$$\gamma_k = Cov(Y_t, Y_{t+k}) = E\left[(Y_t - \mu)(Y_{t+k} - \mu)\right] \tag{2.5}$$

that can be estimated from sample data by

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y}) \tag{2.6}$$

and the set $\{\hat{\gamma}_k, k = 0, 1, 2, \ldots\}$ is known as the *autocovariance function.*

2.2.3 Stationarity

A time series $\{Y_t; t \in Z\}$ is said to be stationary, if its behavior does not change over time. This means that the values always tend to vary about the same level and that their variability is constant over time. There are two common definitions of stationary.

Definition 2.2.3 (Ihaka, 2005)

A time series $\{Y_t; t \in Z\}$ is said to be *strictly stationary* if $k>0$ and any $t_1, \ldots, t_k \in Z$, the distribution of $\left(Y_{t_1}, \ldots, Y_{t_k}\right)$ is the same as that for $\left(Y_{t_1+u}, \ldots, Y_{t_k+u}\right)$ for every value of u.

This definition states that if Y_t is stationary then $\mu(t) = \mu(0)$ and

$\gamma(s,t)=\gamma(s-t,0)$, where $\mu(t)=E(Y_t)$ and $\gamma(s,t)=Cov(Y_s,Y_t)$.

Definition 2.2.4 (Ihaka, 2005)

A time series $\{Y_t; t \in Z\}$ is said to be *weakly stationary* if $E|Y_t|^2 < \infty$, $\mu(t)=\mu$ and $\gamma(t+u,t)=\gamma(u,0)$. In the case of Gaussian time series, the two definitions of stationarity are equivalent. For the non stationary time series**,** Box - Jenkins recommends the differencing approach to achieve stationarity.

2.2.4 The autocorrelation function (ACF)

Definition 2.2.5 (Wei, 1994)

The *autocorrelation* between Y_t and Y_{t+k} is defined as

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)}\,\sqrt{Var(Y_{t+k})}} \tag{2.7}$$

that can be estimated from sample data by

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}\left(Y_t - \bar{Y}\right)\left(Y_{t+k} - \bar{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \bar{Y}\right)^2} \tag{2.8}$$

and the set $\{\hat{\rho}_k,\ k = 0, 1, 2, \ldots\}$ is called the *autocorrelation function.*

2.2.5 The partial autocorrelation function (PACF)

Definition 2.2.6 (Wei, 1994)

The *partial autocorrelation* between Y_t and Y_{t+k} is defined as

$$\phi_{kk} = \frac{Cov\left[\left(Y_t - \hat{Y}_t\right), \left(Y_{t+k} - \hat{Y}_{t+k}\right)\right]}{\sqrt{Var\left(Y_t - \hat{Y}_t\right)}\sqrt{Var\left(Y_{t+k} - \hat{Y}_{t+k}\right)}} \tag{2.9}$$

$$\hat{Y}_{t+k} = \alpha_1 Y_{t+k-1} + \alpha_2 Y_{t+k-2} + \cdots + \alpha_{k-1} Y_{t+1} \tag{2.10}$$

is the best linear estimate of Y_{t+k}.

where $\alpha_i (1 \le i \le k-1)$ is the mean squared linear regression coefficients obtained from minimizing

$$E\left(Y_{t+k} - \hat{Y}_{t+k}\right)^2 = E\left(Y_{t+k} - \alpha_1 Y_{t+k-1} - \cdots - \alpha_{k-1} Y_{t+1}\right)^2 \qquad (2.11)$$

ϕ_{kk} can be estimated from sample data by

$$\hat{\phi}_{kk} = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \cdots & \hat{\rho}_1 & \hat{\rho}_k \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \cdots & \hat{\rho}_1 & 1 \end{vmatrix}} \qquad (2.12)$$

and the set $\left\{\hat{\phi}_{kk}, k = 0, 1, 2, \ldots\right\}$ is called the *partial autocorrelation function*

2.2.6 Testing of stationarity

One of the ways to observe the stationary a time series is through graph of ACF. If a graph of ACF of time series values either *cuts off fairly quickly* or *dies down fairly quickly*, then the time series is known d be considered *stationary.* If a graph of ACF *dies down extremely slowly* then the time series values should be considered *non-stationary.*

2.2.7 Autoregressive Moving Average (ARMA) model

The ARMA model is a model of stationary time series that have a form of regression of past values and past its residual. ARMA model are a combination of Autoregressive (AR) model and Moving-Average (MA) model.

Definition 2.2.7 (Enders, 1995)

The sequence $\{\varepsilon_t\}$ is *white noise process* if for each time period t,

(i) $E(\varepsilon_t)=E(\varepsilon_{t-1})=\ldots=0$

(ii) $Var(\varepsilon_t)=Var(\varepsilon_{t-1})=\ldots=\sigma_\varepsilon^2$ (2.13)

(iii) For all j $Cov(\varepsilon_t,\ \varepsilon_{t-s})=Cov(\varepsilon_{t-j},\ \varepsilon_{t-j-s})=0$ for all s

Definition 2.2.8 (Ihaka, 2005)

If Y_t satisfies

$$Y_t=\phi_1 Y_{t-1}+\cdots+\phi_P Y_{t-P}+\varepsilon_t \qquad (2.14)$$

where ε_t is white –noise and the ϕ_u are constants, then Y_t is called *an auto regressive series of order p*, denoted by AR(p)

Rewrite AR(*p*) series by using *backshift operator* $B^k Y_t = Y_{t-k}$

$$\phi(B)Y_t=\varepsilon_t \qquad (2.15)$$

where $\phi(B)=1-\phi_1 B-\ldots-\phi_p B^p$

The AR(1) series is defined by

$$Y_t=\phi Y_{t-1}+\varepsilon_t \qquad (2.16)$$

or

$$(1-\phi B)Y_t=\varepsilon_t \qquad (2.17)$$

Definition 2.2.9 (Ihaka, 2005)

A time series $\{Y_t\}$ which satisfies

$$Y_t=\varepsilon_t+\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q} \qquad (2.18)$$

(whit $\{\varepsilon_t\}$ white-noise) is said to be *a moving average process of order q* or MA(*q*) process.

Rewrite MA(*q*) series by using *backshift operator* $B^k\varepsilon_t=\varepsilon_{t-k}$

$$Y_t = \theta(B)\varepsilon_t \tag{2.19}$$

where $\theta(B) = 1 + \theta_1 B + \dots \theta_q B^q$

The MA(1) series defined by

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{2.20}$$

or

$$Y_t = (1 + \theta B)\varepsilon_t \tag{2.21}$$

<u>*Definition 2.2.10*</u> (Ihaka, 2005)

If a series satisfies

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \tag{2.22}$$

(whit $\{\varepsilon_t\}$ white-noise), it is called an *autoregressive moving average process of order (p, q)* or ARMA(p,q) series.

Rewrite ARMA(p,q) series

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \tag{2.23}$$

or

$$\phi(B)Y_t = \theta(B)\varepsilon_t \tag{2.24}$$

The ARMA(1,1) series is defined by

$$Y_t = \phi Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1} \tag{2.25}$$

or

$$(1 - \phi B)Y_t = (1 + \theta B)\varepsilon_t \tag{2.26}$$

Refers to Wei (1994) and Enders (1995), the ACVF, ACF and PACF are determined by the following formulas :

$$\gamma_k = \begin{cases} \dfrac{1+\theta^2+2\phi\beta}{1-\phi^2}\sigma^2, & k=0 \\ \dfrac{(1+\phi\theta)(\phi+\theta)}{1-\phi^2}\sigma^2, & k=1 \\ \phi\gamma_{k-1}, & k \geq 2 \end{cases} \tag{2.27}$$

$$\rho_k = \begin{cases} 1, & k=0 \\ \dfrac{(1+\phi\theta)(\phi+\theta)}{1+\theta^2+2\phi\theta}, & k=1 \\ \phi\,\rho_{k-1}, & k \geq 2 \end{cases} \tag{2.28}$$

$$\phi_{kk} = \begin{cases} \rho_1, & k=1 \\ \dfrac{\rho_2-\rho_1^2}{1-\rho_1^2}, & k=2 \\ \dfrac{\rho_k - \sum_{j=1}^{k-1}\phi_{k-1.j}\rho_{k-j}}{1-\sum_{j=1}^{k-1}\phi_{k-1.j}\rho_{k-j}}, & k \geq 3 \end{cases} \tag{2.29}$$

where $\phi_{kj} = \phi_{k-1.j} - \phi_{kk}\,\phi_{k-1.k-j}$, $j=1, 2, 3, ..., k-1$

2.2.8 Box-Jenkins model

The ARIMA approach was first popularized by Box and Jenkins (1976), and ARIMA model are often reffered to as Box-Jenkins model. The ARMA model is a special model of the ARIMA model. There are three primary stages in building a Box-Jenkins time series model, that is the model *identification, estimation and diagnostic checking* and *application.*

Step 1 : Identification

The first step in developing a Box-Jenkins model is to see the stationarity of time series by considering the graph of ACF. If the series is not stationary, it can often be converted to a stationary series by differencing, that is the origin series is replaced by a series of differences.

To identify the models that will be selected can be done by looking the characteristics of the ACF and PACF.

The following table summarizes how to identify the model of the data by using of characteristics for the ACF and PACF (Madsen, 2008).

Table 2.1: Characteristics for the ACF and PACF

	ACF, ρ_k	PACF, ϕ_{kk}
AR(*p*)	Damped exponential and/or sine functions	ϕ_{kk} =0 for k>p
MA(*q*)	ρ_k = 0 for k >q	Dominated by damped exponential and/or sine function
ARMA(*p*,*q*)	Damped exponential and/or sine functions after lag (q-p)	Dominated by damped exponential and/or sine function after lag (p-q)

After model identification, there may be several adequate models that can be used to represent a given data set. One method of model selection is based on is Akaike's Information Criterion (AIC). The AIC is defined are as follows (Wei, 2006) :

$$AIC = n \ln \hat{\sigma}_a^2 + 2M \qquad (2.30)$$

where n is the number of observations, $\hat{\sigma}_a^2$ is the maximum likelihood estimate of σ_a^2 and M is the number of parameters. The best model is given by the model with the lowest AIC value.

Step 2 : Estimation and diagnostic checking

After identifying a tentative model, the next step is to estimate the parameters in the model, One way for a method to estimate the parameters of ARMA model refer to Wei (1994) is the maximum likelihood method as follows :

Consider the series $\{X_t,\ t = 1,2,\ldots,n\}$, ARMA(p,q) model and parameters : $\phi = (\phi_1, \phi_2, \ldots, \phi_p)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_q)$, $\mu = E(X_t)$ and $\sigma_a^2 = E(a_t^2)$ in the model :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q} \tag{2.31}$$

where $Y_t = X_t - \mu$, X_t are stationary or transformed stationary series and $\{a_t\}$ are *i.i.d.* $N(0, \sigma_a^2)$

The equation (2.31) can be written as :

$$a_t = y_t - \sum_{j=1}^{q} \theta_j a_{t-j} - \sum_{i=1}^{p} \phi_i y_{t-i} \tag{2.32}$$

Because $a_t \sim N(0, \sigma_a^2)$, then the joint probability density of $a = (a_1, a_2, \ldots, a_n)$

$$P(a \mid \phi,\ \theta,\ \sigma_a^2) = \prod_{i=1}^{n} \left[(2\pi\sigma_a^2)^{-\frac{1}{2}} \exp\left(-\frac{a_t^2}{2\sigma_a^2} \right) \right]$$

$$= (2\pi\sigma_a^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_a^2} \sum_{t=1}^{n} a_t^2 \right) \tag{2.33}$$

The conditional log likelihood function of parameter ϕ, μ, θ, and σ_a^2 is

$$\ln L(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln 2\pi\,\sigma_a^2 - \frac{S(\phi, \mu, \theta)}{2\sigma_a^2} \tag{2.34}$$

where $S(\phi, \mu, \theta) = \sum_{t=1}^{n} a_t^2 (\phi, \mu, \theta \mid X_n, a_n, X)$, $X = (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$, $X_n = (X_{1-p}, \cdots X_{-1},, X_0)^{\mathrm{T}}$ and $a_n = (a_{1-q}, \cdots, a_{-1}, a_0)^{\mathrm{T}}$

The quantities of $\hat{\phi}$, $\hat{\mu}$ and $\hat{\theta}$ which maximize equation (2.33) are called the conditional maximum likelihood estimators (MLE).

Significance test for parameter estimates indicate wheter some terms in the model may be unnecessary. The model must be checked for adequacy by considering the properties of the residuals wheter the model assumption are satisfied. The basic assumption is that the $\{a_t\}$ are white noise, that is a_t's are uncorrelated random shock with zero mean and constant variance. If the residuals satisfy these assumption then the Box-Jenkins model is chosen for the data. An overall check of model adequacy is provided by the Ljung-Box Q statistics. The test statistic Q (Wei, 1994) is :

$$Q = n(n+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{n-k} \sim \chi^2(K-p-q) \qquad (2.35)$$

If $Q > \chi^2_{1-\alpha}(K-p-q)$ the adequacy of the model is rejected at the level α.

where n is the sample size, $\hat{\rho}_k^2$ is the autocorrelation of residuals at lag k and K is the number of lags being tested.

Step 3 : Application

In this step, use model to the forecast.

2.3 Bayesian approach

There are two popular approach in the statistics, namely the classical or traditional statistics approach and Bayesian statistics approach. estimation theory. The main difference between the Bayesian approach and the Traditional approach is that in the Bayesian approach the parameters is viewed as random variables, whereas the traditional approach considers the parameters to be fixed but unknown.

The Bayesian approach seeks to optimally merge information from two source: knowledge that is known from theory or opinion formed at the beginning of the research in the form of a prior distribution and information contained in the data in the form of likelihood functions. In the Bayesian approach, we combine any new information that is available with the prior information we have, to form the posterior distribution as basis for the statistical procedure. Figure 2.1 presents the statistical procedure for Bayesian approach (Bijak, 2010).

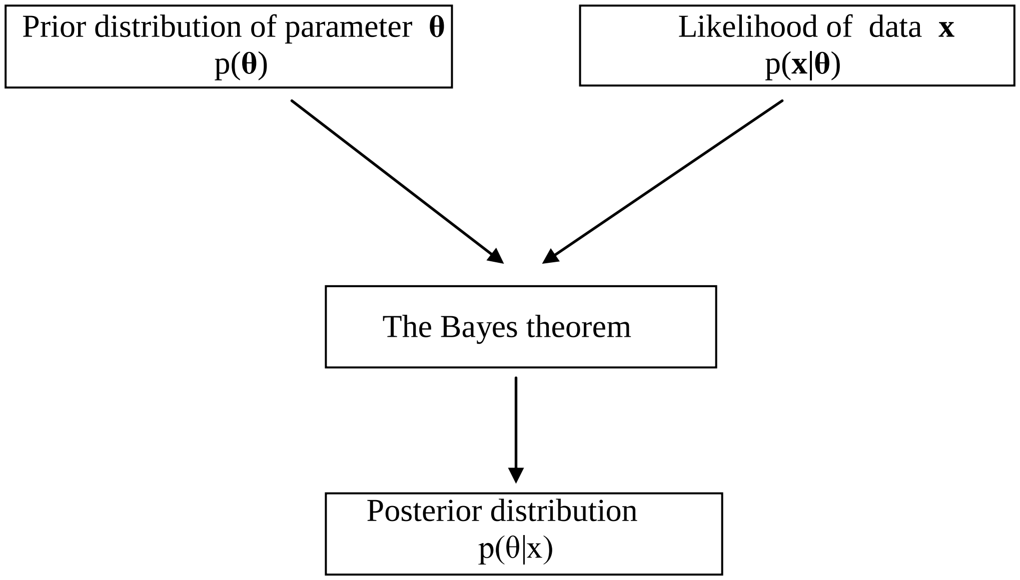

Figure 2.1 : Statistical procedure for Bayesian approach.

Bayesian inference is a method of analysis that combines information collected from experimental data with the knowledge one has prior to performing the experiment (Ramachandran and Tsokos, 2009)

In the Bayesian inference, parameters are supposed to be random variables, which are described by their probability density function. A complete probabilistic model of the observed data and the unobserved parameters is developed through thespecification of their joint probability density function. The previous information about the parameters, that is, before the observation of the data, is expressed through a prior probability density function. By applying the Bayes theorem, this prior information is combined with the observed data to give the posterior probability density function, which represents knowledge about the parameters after the observation of the data. When a Bayesian analysis is conducted, inferences about the unknown parameters are derived from the posterior distribution.

Bayesian forecasting is a natural product of a Bayesian approach to inference (Geweke and Whitemann, 2004).

The Bayesian approach to inference, as well as decision-making and forecasting, involves conditioning on what is known to make statements about what is notknown. Thus Bayesian forecasting is a mild redundancy, because forecastng is at the core of the Bayesain approach to just about anything.

The difference of approaches to probabilistic forecasting between the Traditional approach and the Bayesian approach are presented in the Table 2.2 are as follows:

Tabel 2.2 : Difference between Traditional approach and Bayesian approach.

No	Traditional approach	Bayesian approach
1.	Probability related to the frequency of events.	Probability is a subjective measure of belief.
2	Model parameters are constant, yet unknown.	Model parameters are random and have probability distribution.
3	Inferential process based on sample data from population.	Inferential process used not only the sample from population but also employ the initial knowledge about parameter of population.
4	In the statistics analysis need to discuss sampling distribution	There is no need to discuss sampling distribution for statistic of analysis.

2.3.1 Likelihood function

The likelihood function is a function of parameter based a random sample

Definition 2.3.1 (Bain & Engelhardt, 1992)

The joint density function of n random variables $X_1, X_2, \ldots, X_n$ evaluated at $x_1, x_2, \ldots, x_n$ say $f(x_1, x_2, \ldots, x_n; \theta)$, is referred to as a likelihood function. For fixed $x_1, x_2, \ldots, x_n$ the likelihood function is a function of θ and often denoted by $L(\theta)$.

If $X_1, X_2, \ldots, X_n$ represents a random sample from $f(x;\theta)$, then the likelihood function is :

$$L(\theta) = f(x_1; \theta) \ldots f(x_n; \theta) \tag{2.36}$$

2.3.2 Prior distribution

The prior probability distribution can be used to state initial beliefs about the population of interest before any data collected is used.

Normal-gamma prior are combines normal distribution and gamma distribution.

Definition 2.3.2 (Pole et al. 1984)

Let X be a conditionally normally distributed random quantity so that $X \mid \phi \sim N(m, C\phi^{-1})$, where m and C are know constants. Let ϕ a gamma random quantity, $\phi \sim G\left(\frac{n}{2}, \frac{d}{2}\right)$ for any n>0 and d>0. The joint distribution of X and ϕ is called the *normal-gamma distribution*.

The joint probability density function of X and ϕ is just product $p(X|\phi)\, p(\phi)$ from which it is easily deduced that the conditional distribution $\phi|X$ has density :

$$p(\phi \mid X) \propto \left(\frac{\phi}{2\pi C}\right)^{\frac{1}{2}} \exp\left[-\frac{\phi(x-m)^2}{2C}\right] \frac{d^{\frac{n}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \phi^{\frac{n}{2}-1} \exp\left[-\frac{\phi d}{2}\right] \tag{2.37}$$

$$\propto \phi^{\frac{n+1}{2}-1} \exp\left[-\frac{\phi}{2}\left\{\frac{(x-m)^2}{C}-d\right\}\right] \tag{2.38}$$

This is the form of a gamma distribution with parameter $\frac{n+1}{2}$ and

$\frac{C^{-1}(x-m)^2+d}{2}$

Jeffrey suggested a prior distribution which is mathematically constructed based on the likelihood function and known as *Jeffrey's prior.*

Let $L(\theta|X) \propto f(X|\theta)$ be the likelihood function for θ based on observation X and $\theta=(\theta_1, \theta_2, \ldots, \theta_p)$ is a vector, Jeffrey (Berger, 1985) suggests a prior :

$$\pi(\theta) = \left[\det I_{ij}(\theta)\right]^{1/2} \tag{2.39}$$

where $I(\theta)$ is the (pxp) Fisher information matrix with (i,j) element.

$$I_{ij}\ \theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j}\log\ f\left(X \mid \theta\right)\right] \tag{2.40}$$

2.3. 3 Bayes theorem

Bayesian inference is a method of analysis that combines information collected from experimental data with the knowledge one has prior to performing the experiment. Bayesian and classical methods take basically different outlooks toward statistical inference. In the Bayesian approach, we combine any new information that is available with the prior information we have to form the basis for the statistical procedure. The classical approach to statistical inference that we have studied so far is based on the random sample alone. Bayes theorem calculates the posterior distribution as proportional to product of a prior distribution and the likelihood function. The prior distrbution is a probability model describing the knowledge about the parameter before observing the currently available data, whereas the likelihood function is simply the name given to the model applied to the data. Formally, Bayes theorem presented are following:

Suppose there are two discrete random variables *X* and *Y*, then the joint probability function can be written $p(x,y) = p(x|y)\,p_Y(y)$ and the marginal probability density function of X (Ramachandran & Tsokos, 2009) is:

$$p_X(x) = \sum_y p(x,y) = \sum_y p(x|y)\,p_Y(y) \tag{2.41}$$

Bayes' rule for the conditional $p(y|x)$ is

$$p(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{p(x|y)\,p_Y(y)}{p_X(x)} = \frac{p(x|y)\,p_Y(y)}{\sum_y p(x|y)\,p_Y(y)} \tag{2.42}$$

If Y is continuous, the Bayes theorem can be stated as :

$$p(y|x) = \frac{p(x|y)\,p_Y(y)}{\int p(x|y)\,p_Y(y)\,dy} \tag{2.43}$$

by using the proporsional notation ($\propto$) which can be expressed by

$$p(y|x) \propto p(x|y)\,p_Y(y) \tag{2.44}$$

where $p(y|x)$ is posterior distribution, $p(x|y)$ is likelihood function and $p_Y(y)$ is prior distribution.

2.3. 4 Bayes estimator with decision theory

The loss function, the risk function, the Bayes risk and Bayes estimator are the elements of Bayes estimator with decision theory are defined (Bain & Engelhardt, 1992) as follows

Definition 2.3.3

If T is an estimator of $\tau(\theta)$, then a *loss function* is any real-valued function $L(t,\theta)$ such as for $L(t;\theta) \geq 0$ every t and $L(t;\theta) = 0$ when $t = \tau(\theta)$

Definition 2.3.4

The *risk function* is defined to be the expected loss.

$$R_T(\theta) = E\left[L(T;\theta)\right] \tag{2.45}$$

Definition 2.3.5

For a random sample from $f(x;\theta)$, the *Bayes risk* of an estimator T relative to a risk function $R_T(\theta)$ and pdf $p(\theta)$ is the estimator average risk with respect to $p(\theta)$,

$$A_T = E_\theta\left[R_t(\theta)\right] = \int_\Omega R_T(\theta)\, p(\theta)\, d\theta \tag{2.46}$$

Definition 2.3.6

For a random sample from $f(x;\theta)$, the *Bayes estimator* T^* relative to the risk function $R_T(\theta)$ and pdf $p(\theta)$ is the estimator with minimum expected risk,

$$E_\theta\left[R_{T^*}(\theta)\right] \le E_\theta\left[R_T(\theta)\right] \quad \text{for every estimator T} \tag{2.47}$$

Theorem 2.3.7

The Bayes estimator T of $\tau(\theta)$ under the squared error loss function, $L(T;\theta) = \left[T - \tau(\theta)\right]^2$ is the conditional mean of $\tau(\theta)$ relative to the posterior distribution,

$$T = E_{\theta|X}\left[\tau(\theta)\right] = \int \tau(\theta)\; f_{\theta|X}(\theta)\, d\theta \tag{2.48}$$

Proof (Appendix 1)

2.4 Literature review

The model, the likelihood function or approximate likelihood function, the prior distribution and the joint posterior distribution or the posterior predictive density from the papers related to the Bayesian analysis and Bayesian forecasting in ARMA models are as follows :

Uturbey (2006) in to discuss the ARMA model by Bayesian method applied to streamflow data using ARMA (*p*,*q*) model :

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{q=1}^{q} b_j e_{t-j} + e_t \tag{2.49}$$

where e_t is Gaussian process, independent and identically distributed (*i i d*) with zero mean and variance σ_e^2 . a_i and b_j are parameters.

The likelihood function of this model is :

$$L\left(y \mid a_i, b_j, p, q, \sigma_e^2\right) \approx \left(2\pi\sigma_e^2\right)^{\frac{-(n-p)}{2}} e^{\frac{-1}{2\sigma_e^2}\sum_{t=p+1}^{n} e_t^2} \tag{2.50}$$

$$\text{with } e_t = y_t - \sum_{i=1}^{p} a_i y_{t-i} - \sum_{q=1}^{q} b_j e_{t-j}$$

The prior distribution are combination of the uniform, normal and inverse gamma density :

$$p \mid p_{\max} \sim Unif\{0, \cdots, p_{\max}\}$$
$$q \mid q_{\max} \sim Unif\{0, \cdots, q_{\max}\} \tag{2.51}$$

$$a_i^{(p)} \mid p, \sigma_a^2 \sim N\left(0, \sigma_a^2\right); \quad \text{i=1,…, p}$$
$$b_j^{(q)} \mid q, \sigma_b^2 \sim N\left(0, \sigma_b^2\right); \quad \text{j=1,…, q} \tag{2.52}$$

$$\sigma_a^2 \mid \alpha_a, \beta_a \sim IG\left(\alpha_a, \beta_a\right)$$
$$\sigma_b^2 \mid \alpha_{ab}, \beta_b \sim IG\left(\alpha_b, \beta_b\right) \quad . \tag{2.53}$$

$$\sigma_e^2 \mid \alpha_e, \beta_e \sim IG\left(\alpha_e, \beta_e\right) \tag{2.54}$$

The joint posterior distribution of the all parameters is :

$$p\left(a_i^{(p)}, b_j^{(q)}, p, q, \sigma_e^2 \mid y\right) \propto p\left(y \mid a_i^{(p)}, b_j^{(q)}, p, q, \sigma_e^2\right).$$

$$p\left(a_i^{(p)} \mid p,\sigma_a^2\right) . p\left(\sigma_a^2 \mid \alpha_a,\beta_a\right) \qquad (2.55)$$

$$p\left(b_j^{(q)} \mid q,\sigma_b^2\right) . p\left(\sigma_b^2 \mid \alpha_b,\beta_b\right) .$$

$$p\left(\sigma_e^2 \mid \alpha_e,\beta_e\right) . \; p\left(p \mid p_{\max}\right) . \; p\left(q \mid q_{\max}\right)$$

and the conditional marginal posterior distribution of the parameters of interest can be obtained by integrating equation (2.55).

Safadi and Morettin (2000) in study of Bayesian analysis for threshold autoregressive moving average (TARMA) model , using the

TARMA(2; p_1, p_2, q_1, q_2) model

$$Y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1}\phi_{1i}Y_{t-1} + a_t^{(1)} + \sum_{j=1}^{q_1}\theta_{1j}a_{t-j}^{(1)}, \; if \; Y_{t-d} \le r \\ \phi_{20} + \sum_{i=1}^{p_2}\phi_{21i}Y_{t-1} + a_t^{(2)} + \sum_{j=1}^{q_2}\theta_{2j}a_{t-j}^{(2)}, \; if \; Y_{t-d} > r \end{cases} \qquad (2.56)$$

where $\{a_t^{(1)}\}$and $\{a_t^{(2)}\}$ is a sequence of $i \; i \; d$ ~ $N(0,\tau_1^{-1})$ and $N(0,\tau_2^{-1})$ respectively. $\gamma_1 = (\phi_{10},\ldots,\phi_{1p_1},\theta_{11},\ldots,\theta_{1q_1})^T$, $\gamma_2 = (\phi_{20},\ldots,\phi_{2p_2},\theta_{21},\ldots,\theta_{2q_2})^T$

τ_1, τ_2, r , d are parameters and r is called the threshold parameter.

The approximate likelihood function of this model is :

$$L^*(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d \mid Y_{p+1},\ldots,Y_n) \propto \; \tau_1^{n_1/2} \; \tau_2^{n_2/2} \; \exp.$$

$$\left\{ -\frac{\tau_1}{2}\left[\sum_1 \left(Y_t - \left(\phi_{10} + \sum_{i=1}^{p_1}\phi_{1i}Y_{t-i} + \sum_{j=1}^{q_1}\theta_1\hat{a}_{t-j}^{(2)} \right)\right)^2 \right.\right.$$

$$\left.\left. - \; \frac{\tau_2}{2}\sum_2 \left(Y_t - \left(\phi_{20} + \sum_{i=2}^{p_2}\phi_{2i}Y_{t-i} + \sum_{j=1}^{q_2}\theta_{2j}\hat{a}_{t-j}^{(2)} \right)\right)^2 \right]\right\} \qquad (2.57)$$

The prior distribution are combination of the uniform, normal and gamma density :

$$P(\gamma_1,\gamma_2,\tau_1,\tau_2,r,d) \propto \; \tau_1^{n_1/2-1} \; \tau_2^{n_2/2-1} \; \exp.$$

$$\left\{ -\frac{1}{2}\sum_{i=1}^{2}\left[r_i\left(\beta_i + (\gamma_i - \gamma_{0i})^T Q_i (\gamma_i - \gamma_{0i}) \right)\right]\right\} \qquad (2.58)$$