Bayesian Inference of Autoregressive Models

Dler Kadir

Thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy

Supervisor: Dr. K Triantafyllopoulos

School of Mathematics and Statistics,

University of Sheffield

Sheffield, U.K.

January 2018

Acknowledgments

I would like to express my gratitude to the following people for their support throughout conducting my PhD study.

First, I would like to express my sincere gratitude to my supervisor Dr.Kostas Triantafyllopoulos for the permanent support of my PhD study, his patience, motivation, and immense knowledge. His guidance helped me think critically throughout the process of writing this thesis. I could not have imagined having a better supervisor for my PhD study.

A special thanks goes to Dr. Keith J Harris, Lecturer at School of Mathematics and Statistics at Sheffi eld University, who proofread the entire thesis.

Many thanks go to Ministry of Higher Education and Scientific Research, Kurdistan Regional Government-Iraq, for giving me the chance to do my PhD abroad.

I would like to thank staff and my colleagues in School of Mathematics and Statistics, for their continuous support and help during my PhD. I am using this opportunity to express my gratitude to a number of my friends (Karzan, Hewir and Yassen) who supported me throughout doing my PhD study. I will be grateful forever for your love.

A special thank goes to my parents, brothers and sisters for being helpful and supportive during my tough times. Finally, I would like to thank my family, Shino, the love of my life who always helps me regain hope after despair, my children, Darya and Deema. They have supported me spiritually throughout writing this thesis and my life in general. This project would not have been possible without the support of them.

Abstract

The principles, models and steps of Bayesian time series analysis and forecasting have been developed extensively during the past forty years. In order to estimate parameters of an autoregressive (AR) model we develop Markov chain Monte Carlo (MCMC) schemes for inference of AR model. It is our interest to propose a new prior distribution placed directly on the AR parameters of the model. Thus, we revisit the stationarity conditions to determine a flexible prior for AR model parameters.

A MCMC procedure is proposed to estimate coefficients of AR(p) model. In order to set Bayesian steps, we determined prior distribution with the purpose of applying MCMC. We advocate the use of prior distribution placed directly on parameters. We have proposed a set of sufficient stationarity conditions for autoregressive models of any lag order.

In this thesis, a set of new stationarity conditions have been proposed for the AR model. We motivated the new methodology by considering the autoregressive model of AR(2) and AR(3). Additionally, through simulation we studied sufficiency and necessity of the proposed conditions of stationarity. We also draw parameter space of AR(3) model for stationary region of Barndorff-Nielsen and Schou (1973) and our new suggested condition. A new prior distribution has been proposed placed directly on the parameters of the AR(p) model. This is motivated by priors proposed for the AR(1), AR(2),..., AR(6), which take advantage of the range of the AR parameters. We then develop a Metropolis step within Gibbs sampling for estimation. This scheme is illustrated using simulated data, for the AR(2), AR(3) and AR(4) models and extended to models with higher lag order.

The thesis compared the new proposed prior distribution with the prior distributions obtained from the correspondence relationship between partial autocorrelations and parameters discussed by Barndorff-Nielsen and Schou (1973). It discusses the study by Jones (1987) in which the author generalized a Jacobian transformation based on the expressions for parameters in terms of partial autocorrelations. We have pointed out the limitations of Jones (1987). The proposed methodology is illustrated using simulated and real data.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and motivation

1.1 Introduction and aim of this thesis

Time series are met frequently in many subject areas when observations are collected over time. They are common throughout social science, economics and the humanities. For instance, a time series dataset can consist of stock prices collected daily, measurements made by sensors on an aircraft, or some from monitoring the vital signs of a patient in a hospital. One of the objectives of time series analysis is to fit an autoregressive model of lag order p. Many approaches have been proposed to estimate parameters of the AR(p), but amongst the most successful mechanisms of recent years have been Bayesian inference methods.

In recent years Bayesian inference has received a huge amount of attention, especially with regards to the use of computational methods, particularly Markov chain Monte Carlo (MCMC) algorithms. Whilst several ways exist to construct these chains, the most popular include the Gibbs sampler, Geman and Geman (1984), a subset of the general MCMC technique of Metropolis et al. (1953) and Hastings (1970).

The principles, models and method of Bayesian time series analysis and forecasting have been developed extensively during the last forty years. This development has involved thorough investigation of statistical and mathematical aspects of forecasting models and related techniques. This thesis concerns estimation of the autoregressive (AR) model of order p.

One of the biggest challenges in developing MCMC schemes for time series analysis is stationary, because it imposes restrictions on the parameter space. Although there are some limited studies on stationarity for autoregressive models which give information about the space of the AR parameters, previous studies were primely limited to the AR(3) model. The issue of deriving stationarity conditions directly on the parameters of the autoregressive model has been of interest for the last 60 years. However, the current solutions depend on iterative-based conditions, which are quite difficult to interpret and do not lead to general conditions for higher lag order.

The aim of this thesis is to estimate parameters of the AR model of time series. A Markov Chain Monte Carlo (MCMC) procedure is proposed to estimate coefficients of such a model. In order to use the Bayesian procedure, we need to determine the prior distributions of the parameters. It seems that it is desirable that one places prior distribution on the parameters, as opposed to placing priors indirectly on functions of the parameters such as roots of the AR polynomial or the partial autocorrelations. The reason behind this is that we believe in Bayesian statistics we should put the prior distribution on the parameters and should get the posterior on the parameters of the interest rather than placing prior distributions on the roots of the polynomial or some function of them. Therefore, in this thesis, we advocate the use of prior distributions placed directly on the parameters. In order to achieve this goal we need to revisit stationarity as it restricts the space of the AR parameters and hence has an important impact on the priors.

In this thesis, we develop stationarity conditions for autoregressive models of any lag order p. We motivate the new methodology by considering the autoregressive models of AR(2) and AR(3). Additionally, through simulation we study sufficiency and necessary of the proposed stationarity conditions for the AR(p) model. We also draw the parameter space of the AR(3) model for the stationary region of Barndorff-Nielsen and Schou (1973) and our new suggested condition. We propose a new prior distribution placed directly on the AR parameters of the AR(p) model. This is motivated by priors proposed for the AR(1), AR(2), ... , AR(6) models, which take advantage of the range of the AR parameters. We then develop a Metropolis within Gibbs algorithm for estimation. This scheme is illustrated using simulated and real data for the AR(2), AR(3) and AR(4) models and was extended to models with higher lag order.

We compare the proposed prior distribution with the prior distributions obtained from the correspondence relationship between partial autocorrelations and parameters discussed by Barndorff-Nielsen and Schou (1973). We discuss the study by Jones (1987) in which the author generalized a Jacobian transformation based on the expressions for parameters in terms of partial autocorrelations. We highlight some of the limitations of Jones (1987) in which the prior distribution of parameters cannot be obtained by using his Jacobian transformation in the case of high order. Simulation and real data are used to illustrate the proposed methodology.

1.2 Structure of the thesis

The layout of the thesis is organized as follows. Chapter 2 briefly outlines the literature review of the different topics present in the thesis. It gives a short background of the application and models of time series. It offers the literature of stationary time series models, it describes the stationary process, real and complex roots, and discusses the classical root criterion to determine whether a time series is stationary or not. This chapter also discusses an overview of Bayesian inference including Markov Chain Monte Carlo, namely Gibbs sampling, the Metropolis- Hastings algorithm and the related topic of convergence of Markov chains. Furthermore, the autoregressive moving average (ARMA) model and arguments for the proposed framework are discussed.

Chapter 3 embraces a study on stationary processes for AR(p) as a vector AR(1) model. This chapter covers the study of sufficiency and necessary for the stationarity conditions of autoregressive models. We propose a new set of stationarity conditions which consist of linear inequalities.

We then derive necessary and sufficient conditions for the AR(2) and AR(3) models, and we provide some results for the AR(4). Moreover, we provide a recursive mechanism to derive sufficient conditions for the AR(p) model, for any $p \geq 2$. A simulation study is conducted for the AR(3) and AR(4) models in order to check sufficiency and necessary for our stationary conditions. We explore the accuracy of the proposed stationarity conditions for the AR(3) and the AR(4) via simulation. Finally, we give an alternative proof for the stationarity conditions of AR(3) (Barndorff-Nielsen and Schou, 1973).

Chapter 4 extends the work of Chapter 3 to formulate prior distributions and to implement MCMC steps. This chapter proposes suitable MCMC schemes consisting of Metropolis within Gibbs sampling. We propose a new prior distribution placed directly on the AR parameters of the AR(p) model. This is motivated by priors proposed for the AR(1), AR(2), ... , AR(6) models, which take advantage of the range of the AR parameters. We then develop a Metropolis within Gibbs algorithm for estimation. This estimation scheme is demonstrated by using simulated data, for the AR(2), AR(3) and AR(4) models and is extended to models for higher lag order p. This chapter also outlines the generalized posterior distribution for the AR(p) model.

Chapter 5 is concerned primarily with two sections in order to study and compare the proposed MCMC scheme with some previous studies relevant to our study. The purpose of the first section is to compare the proposed prior distribution with the prior distributions obtained from the correspondence relationship between partial autocorrelations and parameters as discussed in Barndorff-Nielsen and Schou (1973). It discusses the study by Jones (1987) in which the author generalized a Jacobian transformation based on the expressions for parameters in terms of partial autocorrelations. This chapter illustrates some of the limitations of Jones (1987) in which the prior distribution of parameters cannot be obtained using his Jacobian transformation in the case of high order models. This comparison relies upon some methodological steps and practical results when applying the prior distribution to obtain parameter estimates of the AR(p) model. The objective of the second section is to apply our proposed MCMC scheme, to both real data and simulated data. This section compares the results obtained by our method to results of Box et al. (1976). This comparison is performed on different orders of AR(p) models.

Chapter 6 summarizes the achievements of the thesis and provides overall conclusions. Finally, some future work recommendations are provided. The programming language R has been used throughout the thesis.

Chapter 2

Literature review

2.1 Introduction

This chapter provides a short background on the application and models of time series. It provides the literature of stationary time series models, and in particular it describes the order of stationary, real and complex roots, when models of times series are stationary and when roots lie outside the unit circle. This chapter also gives on the overview of MCMC, Gibbs sampling, the Metropolis Hastings algorithm and related topics. Furthermore, the Auto-Regressive Moving Average (ARMA) model, Auto-Covariance Function (ACF), difference equation and partial auto-correlation in time series are presented in order to discuss the general ideas about them.

2.2 Time series

There are many studies for parameter estimation in time series (Wold (1939); Box et al. (1976); Shumway and Stoffer (2011); Kulahci and Bisgaard (2011)). The availability of computer power has enabled us to perform complex computational Bayesian algorithms via simulation (Geman and Geman (1984); McCulloch and Tsay (1993); Cowpertwait and Metcalfe (2009); Petris et al. (2009); Robert and Casella (2010)).

A time series is a set of observations made regularly in time (Chatfield, 2003). Time series observations are either discrete or continuous. When data are made only at certain times, usually equally spaced, a time series is said to be discrete. When data are taken constantly in time, a time series is said to be continuous (Shumway and Stoffer, 2011).

The objective of time series analysis is to learn from the past and to predict future values of time series, assisting managers or policy makers in making well informed decisions. A time series analysis assesses the main properties of data and its variation. These factors, coupled with advanced computing power, have led to time series methods being widely applicable in government, industry and trade (Cowpertwait and Metcalfe, 2009).

Time series is applied in a wide range of areas. For instance, many known time series appear frequently in economics. Examples include monthly total exports in successive months, or stock markets recorded daily or monthly unemployment rates. Much work has been done regarding the implementation of time series in economics (Anderson (1971) and Enders (2008)). Social scientists focus on inhabitant data, like birth rates, or school registrations. A demographer aims at foretelling changes in inhabitants for as long as 10 or 20 years in coming years (e.g. (Brass, 1974)). Epidemiologists may focus on the number of influenza cases noticed over some time period. In medicine, blood pressure measurements traced over time might be helpful for the evaluation of drugs employed in treating hypertension. A range of time series methods are routinely applied in the physical and environmental areas (Shumway and Stoffer, 2011).

2.3 Stationary process

The fundamental concept for any time series analysis is stationary time series. In practice, however, many economic, financial and climate time series are non-stationary and more difficult to deal with. A common practice way to transform the non-stationary time series to stationary is the technique of differentiation. Furthermore, Box et al. (1976) developed this approach and many refer to their approach as Box-Jenkins analysis. It is substantially for stationary time series that models can be developed and we can forecast future values. However, in many applications, especially in economics and business, it is the non-stationary time series that are the most interesting (Kulahci and Bisgaard, 2011). One of the main characteristics that differentiate time series analysis from classical methods is that observed data taken over time can be dependent. However, in order to do any sort of analysis of time series some type of invariance in the time series has to be assumed. For instance, the mean or variance of the time series is not changing through time. One assumption that can be made is that a time series is stationary. The concept of stationarity is that it is an invariance property that means statistical features of the time series do not alter over time. Any sort of inference of time series would not be applicable, if the marginal distribution were completely different Rao (2008) and Gilgen (2006).

There are two types of stationarity, weak stationarity that only considers the covariance of a process and strict stationarity which is a much stronger condition and imposes distributions that are invariant through time.

Strict stationarity: a time series y_t is said to be strictly stationary if the joint distribution of $y_{t_1}, \ldots, y_{t_n}$ is identical to the joint distribution of $y_{t_{1+h}}, \ldots, y_{t_{n+h}}$ for all $t_1, \ldots, t_n$ and for all h, so that it remains the same after any arbitrary time shift (Cowpertwait and Metcalfe, 2009).

Weak Stationarity: the time series is weak stationary if the mean and the variance are constants for all time t and the auto covariance function between y_t and y_{t+h} only depends on the lag difference (Shumway and Stoffer, 2011). We can say that a time series y_t is defined as weakly stationary if and only if

1 . $E(y_t) = \mu < \infty$, that is, the expectation of y_t is finite and does not depend on t.

2 . $\gamma(y_{t+h}, y_t) = \gamma_h$, that is, for each h , the auto covariance of the random variables y_{t+h} and y_t does not rely on t.

2.4 The ARMA model

In this section the autoregressive and moving average models are briefly discussed. The feature of the MA(q) model is that after h lags there is no correlation between two random variables. On the other hand, at all lags for the AR(p) model, there are correlations that need to be considered. It can be noted that estimating the parameters of an AR(p) model is much easier than an MA. Thus, several ways exist in fitting an AR model to the observed data (it can be noted that if the roots of the feature polynomial lie inside the unit circle, then the AR model could be written as an MA, since it is causal) (Cowpertwait and Metcalfe, 2009). In order to fit an AR model to the data, the order of the model needs to be chosen; often one utilizes AIC, BIC or a similar criterion to define the order. However, in a real application, the chosen order tends to be relatively large, order 14 for instance. The large order is selected due to a complex auto-correlation structure and/or when correlations tend to decay slowly (Shumway and Stoffer, 2011). Therefore, a very helpful generalization that can be more elastic is the ARMA(p , q) model. In this case

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

where ϕ_i and θ_j are parameters of AR and MA respectively, and ε_t is white noise (i.e. ε_t is iid with zero mean and some variance σ^2).

2.4.1 Auto covariance function (ACF)

The ACF is determined as the series of covariance of a stationary process. The linear dependence amongst two points on the same sequence calculated at different times is measured by the auto covariance. It can be said that different time series give an increase to various features in the

ACF (Rao, 2008). In order to evaluate the degree of dependence in the data set and to choose a suitable model for the data, the ACF is one of the essential tools that can be used. If the data are values of a stationary time series (y_t), then the ACF can provide us an estimate of the ACF of (y_t). This suggests that among many possible time series models which model is an appropriate candidate for representing the dependence of the data (Brockwell and Davis, 2006). Suppose that y_t is a stationary time series with zero mean, then the auto covariance function of y_t at lag h is as follows:

$$\gamma(h) = Cov(y_{t+h}, y_t) = E[y_{t+h}y_t]$$

Note that the variance ($\gamma(h=0)$) of the time series is obtained when $h = 0$.

2.4.2 The auto covariance of an auto regressive process (AR)

Let us suppose the AR(p) process of y_t with zero mean:

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t$$

where ϕ_j are the autoregressive parameters of the AR model and ε_t is white noise (ε_t is iid with zero mean and some variance σ^2). It can be assumed that y_t is causal (the roots of $\phi(z)$ lie outside the unit circle). Given that y_t is causal, the recursion for the covariance can be derived. By multiplying both sides by y_{t-h} for ($h \geq 0$) and taking expectations, we have

$$E(y_t y_{t-h}) = \sum_{j=1}^{p} \phi_j E(y_{t-j}y_{t-h}) + \underbrace{E(\varepsilon_t y_{t-h})}_{=0} = \sum_{j=1}^{p} \phi_j E(y_{t-j}y_{t-h})$$

It can be observed that this equation would not hold if the process were not causal, since y_{t-h} and ε_t are not necessarily independent. Let us suppose $c(h) = E(y_0 y_h)$ and utilizing the above equation we can see that the auto covariance satisfies a homogeneous difference equation as follows:

$$c(h) - \sum_{j=1}^{p} \phi_j c(h-j) = 0 \tag{2.1}$$

for $h \leq 0$. Recall the feature polynomial of an AR process $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$ that has the roots $\lambda_1, \ldots, \lambda_p$. The characteristic roots are used to obtain the solution for equation (2.1); it can be seen that if the roots are all distinct, the solution of equation (2.1) satisfies:

$$c(h) = \sum_{j=1}^{p} C_j \lambda_j^{-h} = 0 \tag{2.2}$$

where C_j are constants and are selected depending on the initial values $(c(h) : 1 \leqslant h \leqslant p)$. The straightforward way to prove equation (2.2) is to utilize a plug-in approach.This can be done by plugging $c(h) = \sum_{j=1}^{p} C_j \lambda_j^{-h}$ into equation (2.1) as follows:

$$\begin{aligned} c(h) - \sum_{j=1}^{p} \phi_j c(h-j) &= \sum_{j=1}^{p} C_j (\lambda_j^{-h} - \sum_{i=1}^{p} \phi_i \lambda j^{(-h+i)}) \\ &= \sum_{j=1}^{p} C_j \lambda_j^{-h} \underbrace{(1 - \sum_{i=1}^{p} \phi_i \lambda_j^{i})}_{\phi(\lambda_i)} = 0 \end{aligned} \tag{2.3}$$

When we have a case where the roots of $\phi(z)$ are not distinct, let us suppose the roots are $\lambda_1, \ldots, \lambda_p$ with multiplicity $m_1, \ldots, m_p$ $(\sum_{h=1}^{p} m_h = p)$. Thus, the solution of this satisfies:

$$c(h) = \sum_{j=1}^{p} \lambda_j^{-h} P_{mj}(h) \tag{2.4}$$

where P_{mj} is m_j^{th} order of polynomial and the 'hidden' parameters C_j are in $P_{mj}(h)$ (Shumway and Stoffer (2011) and Rao (2008)).

2.5 Stationarity conditions for ARMA type time series models

Most of the time series literature, which is reviewed below, is particular to ARMA models. As mentioned above, a stationary time series consists of observations fluctuating about a fixed mean level with constant variance over the observational period. That is, the overall behavior of the series remains the same with the passage of time (Peña et al., 2011). Schur (1918) - Cohn (1922) (Schur-Cohn) established a non-linear dynamical system in order to calculate the roots of a certain characteristic polynomial. The theory of stationary processes was developed in the thirties and forties of the last century. Wold (1939)'s work was one of the first studies on stationary time series. In the economics literature the classical roots criterion to check for stationarity was first discussed by Samuelson (1941) and further discussed in Chipman (1950). Stationary time series methods can be used for inference and prediction in a great number of time series (Kulahci and Bisgaard, 2011). Engineering and natural and social sciences are some of the examples of areas that use stationary time series.

Below we discuss some key developments of stationary time series. The inception of work of stationary time series dates back to Wold's decomposition theorem. According to Wold (1939) any stationary time series is likely to be decomposed into two different elements: self-deterministic and stochastic. This decomposition is operated by stochastic models which are employed to evaluate the probability that the next observation of the variable is placed between two specified limits, which define the confidence or the reliability of the forecast value. Wold established first that the autoregressive (AR) model is stationary if the roots of its characteristic polynomial lie outside the unit circle in the complex plane. Several other authors have provided detailed proofs of this important result. Samuelson (1941) derived a transforming dynamic system for solving polynomial equations. This system has been expressed through a set of conditions for finding the roots of the characteristic polynomial. Anderson (1971) shows that the so called "root" criterion of stationarity is required in order to write an AR process as an infinite order moving average model (see also Wold (1939)). Brockwell and Davis (2001) study in detail the stationarity of AR processes. Wise (1956) conducted a study on the stationarity conditions for the autoregressive and moving average time series models. He aimed at deriving stationarity conditions on the space of parameters as opposed to the space of the characteristic polynomial (Brockwell and Davis, 2002). However, these conditions are given iteratively from one to another, there is no general formula and they are highly non-linear. They are very long and they do not seem to have a pattern particularly useful for higher dimensions. In particular, all determinant elements for polynomials are up to the fourth degree. Even for low dimensions there is no insight on how a prior distribution could be formed based on Wise's conditions. Gargantini (1971) modify the Schur-Cohn criterion by reformulating its sufficient condition in the presence of roundoff error for floating-point arithmetic. He discovered the explicit expression of the coefficients of the increasing polynomials up to degree 4.

A groundbreaking study on the understanding of autoregressive models is Barndorff-Nielsen and Schou (1973). The authors show that under stationarity there is a one to one corresponding relationship between the parameters of the AR(p) model ϕ_i and partial autocorrelations π_i . This suggests a one to one transformation because it re-parameterizes Φ in terms of the partial autocorrelation Π of the AR(p) process. This corresponding relationship was proved by the authors for orders 1, 2, 3 and 4. Based upon this transformation, a number of studies (Barnett et al. (1996); McLeod and Zhang (2008)) have been conducted to estimate the parameters of autoregressive models. However, the recommended matrix equation for the relationship between the roots of characteristic polynomials and the parameters of the autoregressive model is very complicated. It becomes even more complex when the order is higher than four (Marriott et al., 1995).

One of the effective studies on the stationarity of time series has been done by Priestley (1981). The core of his study established a fixed domain for the AR(2) model. His purpose was to achieve the conditions of stationarity in the autoregressive model by limiting the domain of complex roots and real roots in a triangular region, the above results were limited to the AR(2) model.

Monahan (1984) propose a simple re-parameterization of the parameters of the ARMA model. He conducted this by drawing values from a uniform distribution which is bounded to ensure the stationarity and invertability regions of the process.

Okuguchi and Irie (1990) derived a set of determinant inequalities in cubic equations, which described an alternative computation of the Schur-Cohn condition. Three conditions were known for the quadratic equation, and the authors above extended them to higher degrees. They found that the conditions stated by Schur-Cohn without using determinant inequalities were different from the original as well as from the simplified Samuelson conditions, see Farebrother (1973). They provided a direct demonstration of the equivalence of the Farebrother (1973) and Schur-Cohn condition: using diagram matrix arguments.

Farebrother (1992) applies the Schur-Cohn criterion for establishing the stationarity region to the AR(2), AR(3) and AR(4) models. He then looks at simplifying the inequalities and provides some alternative expressions. These are all non-linear for the AR(3) model and equivalent to the corresponding conditions of Samuelson (1941).

Marmol (1995) establishes a constant domain for the AR(2) model based on Schur-Cohn theory. In order to make the AR(2) process stationary, Marmol (1995) relies on the use of two positive determinants for parameters of the AR(2) model. Najim (2010) has provided the analysis of stability domain of a second order transfer function. His aim was to obtain the region of stationarity conditions in the autoregressive model by restricting the domain of complex roots and real roots in a triangular region. He also presented the Schur-Cohn stability algorithm based on a transfer function of all pass filter. A transfer function is defined based on the roots of the AR(2) model when the roots lie outside the unit circle. He first expressed a condition required for the stability in terms of k_p as follows,

$$|k_p| = \left|(-1)^p \prod_{i=1}^{p} p_i\right| < 1 \text{ since } |p_i| < 1 \; \forall \; i = 1, \ldots, p$$

Then a transfer function of a p^{th} order all pass filter is developed. After, Najim (2010) has looked at the correspondence relationship between the Schur-Cohn algorithm and his transfer function using polynomial order of p^{th} order. This polynomial vector is used in both the Schur-Cohn algorithm and Levinson algorithm. The Schur-Cohn algorithm is written as follows:

$$\begin{bmatrix} \phi_{p-1}(x^{-1}) \\ x^{p-1}\phi_{p-1}(x) \end{bmatrix} = \begin{bmatrix} \frac{1}{(1-k_p^2)} & \frac{-k_p}{(1-k_p^2)} \\ \frac{-k_p}{(1-k_p^2)}x^{-1} & \frac{1}{(1-k_p^2)}x^{-1} \end{bmatrix} \begin{bmatrix} \phi_p(x^{-1}) \\ x^p\phi_p(x) \end{bmatrix} \tag{2.5}$$

while the expression for the Levinson algorithm satisfies

$$\begin{bmatrix} \phi_p(x^{-1}) \\ x^p\phi_p(x) \end{bmatrix} = \begin{bmatrix} 1 & k_p(x) \\ k_p & x \end{bmatrix} \begin{bmatrix} \phi_p(x^{-1}) \\ x^{p-1}\phi_{p-1}(x) \end{bmatrix} \tag{2.6}$$

where $\phi_p(x^{-1})$ is the AR characteristic polynomial, $\phi_p(x)$ is the transfer function of the AR model and ϕ_p is the p^{th} coefficient of the AR model. The Levinson-Durbin algorithm is a procedure in linear algebra in order to provide the solution of on equation involving a diagonal-constant matrix. It can be noted that the matrices in equations (2.5) and (2.6) are inverse of one another. This means the coefficients of equation (2.6) are the reflection of the Schur-Cohn coefficients. Although this restriction is used as a criterion to determine coefficients of the AR(2) model, it cannot be applied to transfer functions of an order greater than 2.

2.6 Bayesian inference

2.6.1 Introduction to the basic principle

Bayesian statistics requires a significantly different method of considering statistical inference when it is compared to the traditional school such as confidence intervals, p-values, hypothesis testing, etc. A major difference between the Bayesian framework and the frequentist way lies in introducing the prior information through the framework of probability distributions (Dunson, 2001). The prior distribution gives a summary of everything that is obtained about parameter θ, except the data; moreover, in the Bayesian approach, conclusions are normally reached using probability statements.

The use of Bayesian approaches has significantly increased and these approaches have been implemented to a wide range of study areas and scientific research. Bayesian data analysis includes analyzing statistical models by integrating prior information about parameters (Spiegelhalter et al., 2002). In Bayesian inference, the model for the observed quantity $y = (y_1, y_2, ..., y_n)^T$ is defined via a vector of unknown parameters ϕ using a probability distribution $p(y|\phi)$ where it is assumed that ϕ is a random quantity having a prior distribution $p(\phi)$. Thus, inference about ϕ is based on its posterior density $p(\phi|y)$, given by:

$$p(\phi|y) = \frac{p(\phi)p(y|\phi)}{p(y)} \tag{2.7}$$

where $p(y) = \sum p(\phi)p(y|\phi)$ if ϕ is a discrete random variable and $p(y) = \int p(\phi)p(y|\phi)d\phi$ in the continuous case. Equation (2.7) may then be stated in a proportional form:

$$p(\phi|y) \propto p(\phi)p(y|\phi) \tag{2.8}$$

All Bayesian inferences follow from the posterior distribution because it contains all the related knowledge about the parameter of interest ϕ.

2.6.2 MCMC methods

MCMC methods focus on simulating direct draws from some distribution of interest, usually a posterior distribution. Through the MCMC approach one can employ the previous sample to randomly produce the next sample, producing a Markov chain.

Initiated in the early 1990s (Gelfand and Smith, 1990) a particular the MCMC method, the

Gibbs sampler, is very extensively applicable to a broad class of Bayesian problems and it has sparked a great rise in the application of Bayesian analysis. MCMC methods have their origins in the Metropolis algorithm (Metropolis et al., 1953); an attempt made by physicists to compute complicated integrals by considering them as expectations for some distribution and then evaluating this expectation by drawing samples from that distribution. The Gibbs sampler, mentioned above, (Geman and Geman, 1984) has its roots in image processing.

To utilize the MCMC iterations to obtain a satisfactory representation of the real posterior distribution, the chain must converge. Obtaining convergence poses a major application challenge connected to any MCMC approach. This means that the MCMC results no longer depend on the initial condition of the chain. This method for utilizing MCMC iterations as samples from a posterior distribution $f(.)$ can be summed up as follows:

1. Produce a Markov chain $(\phi_1, \ldots, \phi_k)$ employing Gibbs sampling

2. Wait until the Markov chain reaches equilibrium; assume this happens at time T

3. The samples $(\phi_{T+1}, \ldots, \phi_{T+k})$ in the Markov chain can then be taken a sample drawn from $f(.)$

The time before T is defined as the burn-in period. It should be the time the Markov chain really obtained convergence, in other words, the way T is calculated. Regrettably, T cannot be calculated theoretically. Instead informal diagnostics must be utilized to determine if the chain has obtained convergence. Various diagnostic approaches have been presented in order to monitor the convergence of MCMC chains. The commonest and most usable diagnostic method is the trace plot. "The trace plot" is a basic tool which is utilized for observing the mixing behavior of the chain. Samples which are drawn from parameters of interest can be plotted as a function of iteration number from the duplicated chains started from over-dispersed values. Naturally, a wavy pattern suggests strong autocorrelations within the chain, at the same time zigzag patterns demonstrate that the parameter proceeds more freely (Sorensen and Gianola, 2002).

2.6.3 Gibbs sampling

Gibbs sampling proposed by Geman and Geman (1984) is a specific case of the Metropolis-Hastings algorithm. Within Gibbs sampling, a sample (a random number or draw) is generated from all full conditional distributions, hence sampling the joint posterior distribution. This is simple if we can ascertain that each of these distributions is one of the well-known distributions that we can easily sample from. An algorithm description is provided below.

The aim of Gibbs sampling is to iteratively apply a distribution for a random variable conditioning on impermanent (temporary) initial values of the others on a permanent cycle till samples from the process empirically approximate the required marginal distribution (Gill, 2014). Thus, random samples can be drawn from a conditional distribution for ϕ_1 and then ϕ_1 is used to draw ϕ_2 from its conditional distribution and so forth. The iterative nature of the Gibbs sampling algorithm can be simplified by the requirement that it cycles throughout these full conditionals drawing coefficient values on the basis of the most current version of all of the parameters already in the list. The order does not matter, however the use of the most recent draw from the other samples is necessary.

Let us consider that it is intended to draw samples for the set of random variables for the marginal posteriors of $\phi_1, \phi_2, \ldots, \phi_p$, where $(\Phi = \phi_1, \phi_2, \ldots, \phi_p)$ is the parameter vector of interest, but the marginal distribution cannot be obtained from the joint posterior analytically. However, the full conditional distributions, which can easily be sampled from, are available. The procedure can be outlined as follows:

1. Determine initial values for the parameters ϕ, i.e., $\phi^{(0)} = (\phi_1^{(0)}, \phi_2^{(0)}, \ldots, \phi_p^{0})$;

2. Following this, draw from each complete conditional distribution consecutively, employing the current updated values achieved from the earlier and most recent steps. That is, for $i = 1, \ldots, N : y = (y_1, \ldots, y_n)$

(a) draw $\phi_1^{(i)} \sim \pi(\phi_1|\phi_2^{(i-1)}, \phi_3^{(i-1)}, \ldots, \phi_p^{(i-1)}, y)$
(b) draw $\phi_2^{(i)} \sim \pi(\phi_2|\phi_1^{(i)}, \phi_3^{(i-1)}, \ldots, \phi_p^{(i-1)}, y)$

$\vdots$

(c) draw $\phi_p^{(i)} \sim \pi(\phi_p|\phi_1^{(i)}, \phi_2^{(i)}, \ldots, y)$.

3. Repeat step 2 until the chain converges.

When convergence is obtained, the entire set of simulation values are from the target posterior distribution and a sufficient number has to be drawn so as to analyse all areas of the posterior. The significant feature throughout every iteration of the cycling of the vector ϕ, conditioning on the values of ϕ that have already been sampled for that cycle; otherwise the last cycle can take the current ϕ values. As a result, the sample value for the pth parameter has to be examined on all the i-step values in the final stage for a given cycle (Gill, 2014). The above-mentioned statements distinctly show that having the complete set of conditional distributions is needed to run the Gibbs sampling algorithm. Sometimes running Gibbs sampling could be very inadequate, however it reduces multidimensional issues to a series of univariate conclusions by considering parameters one by one and as a result it will be uncomplicated to program (Casella and George, 1992). It can be observed that if the Gibbs sampler has been put to run for a relatively long period, a whole sample of the values in the ϕ vector will be generated by a complete cycle of the algorithm.

2.6.4 The Metropolis-Hastings Algorithm

One of the issues of implementing the Monte Carlo integration lies in achieving samples from a complicated probability distribution $p(y)$. Making an attempt to tackle this issue is at the roots of MCMC approaches. Specifically, these trace to efforts by mathematical physicists to combine complicated functions by drawing samples randomly ((Metropolis et al., 1953) and (Hastings, 1970)). (Chib and Greenberg, 1995) provide a clear and distinct review of this method. Our aim is to extract samples from $p(\phi)$ where $p(\phi) = f(\phi)/k$, and the normalizing constant k is unlikely to be realized, and is not easy to calculate (because it involves intractable integrals). The Metropolis algorithm (Metropolis et al., 1953) results in a series of draws from this distribution as follows:

1. Use any starting value of ϕ_0 satisfying $f(\phi_0) > 0$.

2. Employing the most recent ϕ value, sample a proposed point ϕ^* from some distribution $q(\phi^*, \phi_{t-1})$ (Robert and Casella, 2010), which is the probability of returning a value of ϕ^* given a former value of ϕ_1, ϕ_t. The algorithm precedes by computing the acceptance probability which is the probability is defined that the candidate value ϕ^* will be accepted as the next value in the sequence. If ϕ^* is accepted, then $\phi_t = \phi^*$, otherwise $\phi_t = \phi_{t-1}$. This distribution is denoted as the proposal or candidate-generating distribution. Various Metropolis-Hastings algorithms are constructed based on the choice of proposal density. If the proposal density is independent of the current value in the sequence, $q(\phi^*|\phi_{t-1}) = q(\phi^*)$, then the resulting algorithm is called an independence chain. There are other proposal densities that can be defined by letting the density have the form $q(\phi^*|\phi_{t-1}) = h(\phi^* - \phi_{t-1})$, where h is a symmetric density about the origin.

3. With the proposed point being ϕ^*, compute the density ratio at the candidate ϕ^* and the recent ϕ_{t-1}

$$\alpha = \frac{p(\phi^*)}{p(\phi_{t-1})} = \frac{f(\phi^*)}{f(\phi_{t-1})}.$$

Since we are looking at the ratio of $p(\phi)$ under two different values, the normalizing constant k cancels out.

4. When the jump causes a rise in the density $p(\alpha > 1)$, accept the candidate point $(\phi_t = \phi^*)$ and go back to step 2. When the jump causes a fall in the density $p(\alpha < 1)$ then with probability α accept the proposed point, else reject it and go back to step 2.

We can outline the Metropolis sampling algorithm as first computing and then accepting a candidate point with probability α (the probability of a move). This produces a Markov chain $(\phi_0, \phi_1, \ldots, \phi_t, \ldots)$, as the transition probabilities from ϕ_t to ϕ_{t+1} relies only on θ_t and not $(\theta_0, \ldots, \theta_{t-1})$. Following an adequate burn-in period (of say, k steps) the chain reaches its stationary distribution and samples from the vector $(\theta_{k+1}, \ldots, \theta_{k+n})$ are samples from $p(y)$.

2.7 MCMC approaches for stationary time series

A considerable amount of computationally intensive statistical analysis becomes feasible by employing the currently developed techniques of repeated stochastic simulation. Recently, the analysis of stationary processes in time series has been considerably influenced by computational approaches, particularly MCMC. In the AR models, for instance, this is depicted by MCMC schemes for predictive and posterior inference proposed initially by Monahan (1983), Albert and Chib (1993) and McCulloch and Tsay (1993).

In the work of Monahan (1983), the Akaike information criterion (AIC) is utilised in order to determine the autoregressive order. By linking prior information with the parameter structure of the model, Bayesian order-determination methods produce terms that penalize an over parametrization of the model. Monahan (1983) developed a practical method of prior specification for ARMA models. He utilized the normal and inverse gamma distributions and applied his methods to a number of examples. He re-parameterized coefficients $\{\phi_1, \ldots, \phi_p = \Phi\}$ and $\{\theta_1, \ldots, \theta_q = \Theta\}$ in terms of the partial autocorrelations $\pi_1, \ldots, \pi_p = \Pi$ which was used to impose the stationarity of the autoregressive polynomial. Monahan (1983) yielded correct Bayes factors for five substitute models and obtained four step ahead predictive means and standard deviations for each model. However he used a normal prior which was restricted to the stationarity region implied by the partial autocorrelation domain. Such a prior may not be suitable for high order p, q and indeed he only provided examples with $p + q \leq 2$, where p is the order of the AR part and q is the order of the MA part. McCulloch and Tsay (1993) were one of the first to suggest the use of MCMC for AR models. They illustrate their methods on the monthly retail price of systematic unleaded gas in America. Regarding the prior distribution, these authors propose normal priors for the AR parameters and a gamma prior for the precision (inverse of variance). The disadvantage of the proposal methods is that both the prior and the posterior distribution of the AR coefficients have support on the real-line (normal distribution), as a result, there is a positive probability assigned to values of AR parameters that result in non-stationary time series. Thus the methods fail to take the stationary region into account. Albert and Chib (1993) criticized McCulloch and Tsay (1993) and proposed different MCMC simulation methods for estimating the parameters. They suggested a Gibbs sampler for an autoregressive model which has intercept and variance shifts ruled by a Markov structure. Although rejection sampling was considered as an improvement, this did not seem to work well for a general AR order.

Barnett et al. (1996) consider a multiplicative seasonal AR model. They developed McCulloch and Tsay (1993)'s idea and demonstrated how to concurrently select the model order of the regular and seasonal polynomials. They tried to avoid outliers, impose stationarity, and evaluate missing observations. Barnett et al. (1996) made some new developments related to the prior structure in AR and ARMA models with a main emphasis on a prior devoted for partial autocorrelation coefficients, which shows the capability of complex statistical analysis of prior and posterior distribution through MCMC models. This approach gives a degree of flexibility in progressing structured prior distributions in applied situations.

Barnett et al. (1996) concentrate on the priors specified for partial autocorrelation coefficients π_i rather than autoregressive parameters ϕ_i. The authors enforce the stationarity conditions by using a sufficient Metropolis within Gibbs algorithm in order to generate partial autocorrelation. They illustrate how to carry out Gibbs sampling when the AR order is unknown. They also show a way to combine different aspects of fitting AR models giving a more efficient and comprehensive treatment. Barnett et al. (1996) consider a model as follows,

$$y_t = w_t + o_t$$

where y_t is the t^{th} observation, w_t in the integrated AR process and o_t is an addictive outlier. The ARMA(p , q) are

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \quad \theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

with orders p and q respectively. Barnett et al. (1996) provides details on the relationship between partial autocorrelation π_i and parameters ϕ_i . The authors have provided five assumptions for the ARMA(p , q) model.

Assumption 1: The roots of both $\phi(B)$ and $\theta(B)$ lie outside the unit circle.
Assumption 2: Both errors o_t and ε_t follow normal distributions. These errors are modeled as a finite mixture of normals in order to allow for additive and innovation outliers. Therefore, $o_t \sim N(0, K_{1t}\sigma^2)$ and $\varepsilon_t \sim N(0, K_{2t}\sigma^2)$ and $K_t = (K_{1t}, K_{2t})$ has a bivariate multinomial distribution with $K_{1t} \geq 0$ and $K_{2t} \geq 1$. When $K_{1t} > 0$, it indicates that there is an additive outlier at time t. If $K_{2t} > 1$, it means that there is an innovation outlier at time t. Furthermore, in case there is no additive outlier and no innovation outlier then $K_t = (0, 0)$ so that $o_t = 0$ and $\varepsilon_t \sim N(0, \sigma^2)$.

Barnett et al. (1996) assume that $\pi = (\pi_1, \ldots, \pi_p)'$ is the first p partial autocorrelations of a zero mean stationary AR process with AR polynomial $\phi(B)$ and they assume that $\Pi = (\Pi_1, \ldots, \Pi_p)'$ to be the first q partial autocorrelations of an AR process with AR polynomial $\theta(B)$. By re-parameterizing ϕ and θ in terms of π and Π, the stationarity constraints become $-1 < \pi_i < 1$ for $i = 1, \ldots, p$ and $-1 < \Pi_i < 1$ for $i = 1, \ldots, q$. The authors indicated that in order to allow some of the π_j and Π_j to be zero, they let $J_{1j} = 0$ if $\pi_j = 0$ and let $J_{1j} = 1$ otherwise, $j = 1, \ldots, p$. They also let J_{2j} be defined similarly with respect to the Π_j

for $j = 1, \ldots, q$.

Assumption 3: The authors priors incorporate indicator variables to assert a positive prior probability that each partial autocorrelation coefficient is zero. For $i = 1, \ldots, p$ and $j = 1, \ldots, q$ the indicators J_{1j} and J_{2j} are independently distributed. When $J_{1j=1}$, then the prior for π_j has a uniform distribution on (-1, 1), and when $J_{2j=1}$, then the prior for Π_j has a uniform distribution on (-1, 1).

Assumption 4: The prior for μ is flat and the prior for σ^2 is the standard reference prior $f(\sigma^2) \propto \frac{1}{\sigma^2}$.

Assumption 5: There are no additive or innovation outliers when $t \leq 0$.

The advantage of Barnett et al. (1996)'s approach is that one for one draws of each partial autocorrelation can be obtained in a more complicated algorithm. The AR coefficients are re-parametrized in terms of the reflection coefficients. Model order selection is performed by associating a binary indicator variable with each parameter. Regarding sampling their posterior distribution, their MCMC method appeals to stochastic variable selection and reversible jump ideas. They have applied a MCMC method using Metropolis-Hastings. The proposal density is as follows,

$$g(\pi_i) \propto f(y_t|y_{t-1}, \pi_{i\neq j}, J_1, \Pi, J_2, K, O, \sigma^2, \mu)$$

π_j and J_{1j} are generated jointly by first generating J_{1j} from a binomial distribution with the π_j integrated out; next π_j is generated conditional on J_{1j}. O_t and K_t are generated from a multinomial distribution, σ^2 and μ are generated from an inverse gamma and a normal distribution, respectively. A disadvantage of this method is its complexity, in particular regarding high order AR models. The model is more complicated and does not deal with the issue of forecasting .

Barnett et al. (1997) employ MCMC methodology to predict the multiplicative seasonal and autoregressive moving average model. The proposed MCMC places a uniform prior on the partial autocorrelations as well as the standard inverse gamma prior on the variance of the innovations. This work makes use of the correspondence of the partial autocorrelation and the parameters of the AR model, first proven in Barndorff-Nielsen and Schou (1973). However, the algorithm of Barndorff-Nielsen and Schou (1973) is complex and does not focus on the forecasting problem, and this limits the work of Barnett et al. (1997) as well.

Huerta and West (1999) highlight the development, specification and analysis of autoregressive time series models, and expand the work of Barnett et al. (1996). The point that distinguishes the work of Huerta and West (1999) from the work of Barnett et al. (1996) is the development of classes of priors. Huerta and West (1999) include model order uncertainty into the linear AR framework focusing on prior specification for latent structure. This results in a novel class of prior distributions on the characteristic reciprocal roots of the process.

Huerta and West (1999) considered a standard AR model of order p as follows;

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t$$

where ϕ_j are parameters of the AR model and ε_t is white noise (ε_t is iid with zero mean and some variance σ^2). The characteristic polynomial is as follows:

$$\phi(x) = 1 - \sum_{i=0}^{p} \phi_i x^i = \prod_{i=1}^{n} (1 - z_i x)$$

where $\{z_1, \ldots, z_p\}$ are the reciprocals of the roots of $\phi(x) = 0$. By assuming a stationary process, the roots of $\phi(x)$ lie outside the unit circle $|z_i| < 1$ for each i. Authors have often used the back-shift operator as a traditional representation $\phi(B)y_t = \varepsilon_t$

$$\prod_{i=1}^{n} (1 - z_i B) y_t = \varepsilon_t$$

Suppose the roots of z_i are distinct and occur as two conjugate pairs which are real roots and complex roots. Huerta and West (1999) include model order uncertainty into the linear AR framework concentrating on prior specification. This results in a novel class of prior distribution. The authors have selected the marginal prior for both real roots and complex roots. The class prior assumes that both real roots and complex roots are exchangeable. The marginal prior for real roots is as follows.

Huerta and West (1999) selected a prior for autoregressive real roots (r_i) for each $i = 1, \ldots, R_i$ the real roots r_i have a prior over support $|r_i| \leq 1$ with density function as follows:

$$r_i \sim \psi_{r,-1} I_{(-1)}(r_i) + \psi_{r0} I_0(r_i) + \psi_{r_1} I_1(r_i) + (1 - \psi_{r_0} - \psi_{r,-1} - \psi_{r_1}) g_r(r_i)$$

where $I(.)$ is the indicator function, $I_0(r) = 1$ if $r = 0$ and $I_0(r) = 0$ otherwise, and $g_r(.)$ is a continuous density function over (-1, 1). Furthermore, $\psi_{r,-1}$, ψ_{r0} and ψ_{r_1} are the prior probabilities that $r_i = -1, 0, 1$ respectively. Therefore, authors state that the marginal prior for r_i allows for roots on the stationary boundary $r_i = \pm 1$. According to the authors, the continuous part of the prior, $g_r(.)$ specifies the conditional prior pdf over the stationary region $-1 < r_i < 1$.

Huerta and West (1999) also adopted the uniform Dirichlet prior distribution as the default prior for the three probabilities as follows,

$$(\psi_{r_1}, \psi_{r0}, \psi_{r,-1}) \sim Dir(\psi_{r_1}, \psi_{r0}, \psi_{r,-1} | 1, 1, 1)$$

Prior specification for each pair of complex roots are provided with some qualitative features. They structured the prior for complex roots as follows,

$$r_i \sim \psi_{c0} I_0(r_i) + \psi_{c_1} I_1(r_i) + (1 - \psi_{c_1} - \psi_{c0}) g_c(r_i)$$

where $I(.)$ is the indicator function. We will now discuss the components of the above equation.

ψ_{c_0} is the prior probability for the first component which corresponds to the zero root. This prior is coupled with a prior on real roots, a full prior expression of uncertainty about model order is provided. The prior for r_i has non-zero prior probability ψ_{c_0} on the stationary region. The authors have used a uniform Dirichlet hyper prior distribution on the selection probabilities for the complex roots as follows,

$$(\psi_{c_0}, \psi_{c_1}) \sim Dir(\psi_{c_0}, \psi_{c_1}|1, 1)$$

It can be noted that with the real roots, the complex roots are not identified. As the authors stated, under this prior and assuming independence across i, the model is unchanged under arbitrary permutations of the root index i.

Huerta and West (1999) have applied the method of MCMC in order to calculate posterior probabilities. Their MCMC is based on the standard Gibbs sampling algorithm embracing direct simulation of the truncated normal distribution by using the quantile function. Their model parameters are as follows,

$$\eta = \{z_i = 1, \ldots, p; (\psi_{r,-1}, \psi_{r_0}, \psi_{r_1}); (\psi_{c_0}, \psi_{c_1}; \sigma^2\}.$$

Thus, posterior inferences are based on summarizing the full posterior distribution. It can be noted that Huerta and West (1999) have provided posterior conditional distributions for both real roots and complex roots.

Their MCMC appeals to stochastic variable selection and reversible jump ideas. These references show how model uncertainty might be embedded into MCMC methods in the context of linear and stationary time series models. This MCMC is based on a standard Gibbs sampling algorithm embracing direct simulation of the truncated normal by using the quantile function. They selected the marginal prior for real roots and complex roots separately and also select the stationary boundary for each of them. However, there is not a one to one relationship between the roots of the characteristic polynomial and the AR coefficients. As p and q increase, the roots are too many and these lead to prior identification problems. More importantly, the modeler will be interested in the posterior distributions of the coefficients and not of the roots.

2.8 Arguments for the proposed framework

We aim to estimate parameters of the AR model for time series. In this thesis we are placing prior distributions directly on the parameters of the AR model not on functions of the parameters. The reason behind placing priors on the parameters is that we believe in Bayesian statistics we should put the prior distribution on the parameters and should get the posterior on the parameters of the interest. However, there are other possibilities in order to do this.

Albert and Chib (1993) provide a model in which they put a normal prior distribution on the parameters. They obtained a posterior distribution of all unknown parameters and functions by simulating a prior distribution that is essentially a standard normal distribution on the AR coefficients. However, we do not want to use Albert and Chib (1993)‘s approach because we will end up with many rejection steps which is not an appropriate prior distribution to choose or we have to put very appropriate prior distributions around zero.

A novel prior model has been developed by Barnett et al. (1996) for the AR model with the main focus on priors specified for partial autocorrelation coefficients rather than raw AR parameters. They enforce stationarity conditions by utilizing a very effective algorithm, which is Metropolis within Gibbs, in order to generate partial auto-correlations. Basically, they place prior distributions on the roots of the polynomial. We believe that is a good approach for placing a prior on the partial autocorrelation function. However, it is difficult to obtain the posterior distribution because their model is too complicated, especially when the order of the AR model is high. Additionally, their model does not deal with the forecasting problem.

Huerta and West (1999) have developed a model which concentrates on defining classes of prior distributions for latent variables and parameters related to the latent components of the AR model. They proposed a novel class of prior distributions for model order and model parameters, such priors characterizing the number and structure of latent underlying components in AR processes. However, this approach is too complicated to use prior information on periodicity. The authors MCMC algorithm is considerably more complex. Our proposed algorithms are much easier than these three alternatives.

The above discussions are some alternative approaches that could have been used. However, our proposed algorithms are much easier than these alternative approaches. We believe that prior distributions should be directly placed on the parameter rather than the roots of the polynomial or some function of them. Therefore, we will develop stationarity conditions for the AR model in order to define flexible prior distributions which are directly placed on the AR parameters.

Chapter 3

Stationary AR processes

3.1 Introduction

This chapter studies stationarity conditions of AR(p) model. It also highlights the sufficiency and necessity for the stationary conditions of auto-regression models. We apply simulation studies to the AR(3) and AR(4) models in order to check sufficiency and necessity for stationarity conditions of the AR model.

It has previously been mentioned that our primary aim is to estimate autoregressive parameters of Autoregressive for the AR(p) model through MCMC. The parameter estimations for an AR(p) model are obtained directly by putting prior information on parameters instead of its characteristic roots (Barnett et al., 1996) and (McLeod and Zhang, 2008). Therefore, relevant information of parameters are needed in order to satisfy stationarity conditions of the AR(p) model. Consequently, this chapter highlights the stationarity processes for the AR model in which we are focusing on the study of the stationarity of the AR(1), AR(2) and AR(3) models. It also covers how to derive the stationary process for the AR(4) model and how to derive stationarity conditions for high order polynomial models. Furthermore, a simulation study is used in order to know the sufficiency and necessity of recommended stationarity conditions. The results will be used in order to formulate a prior distribution and to implement MCMC algorithms in the next chapter.

3.2 Representing a stationary AR(p) process as a vector AR(1) model.

In this section we provide a simple proof of the root criterion for stationarity of AR(p) models, that is y_t is defined to be stationary if the roots of the characteristic polynomial lie outside the unit circle. Let y_t be generated by the AR(p) model:

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

where ϕ_i are the parameters of the AR(p) model and ε_t is white noise (ε_t is iid with 0 mean and some variance). Then, y_t is a stationary process if and only if the roots of $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ lie outside the unit circle. This proof is different from the previous proofs that have been done through the polynomials of $\phi(z)$ for ARMA(p, q) as can be seen in (Brockwell and Davis, 2001). Let X_t be a vector of time series following the vector AR model of first order;

$$X_t = \Phi X_{t-1} + \varepsilon_t, \tag{3.1}$$

where

$$X_t = \begin{bmatrix} X_{1t} \\ \vdots \\ X_{kt} \end{bmatrix}, \qquad \Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1k} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2k} \\ \vdots & \vdots & & \vdots \\ \phi_{k1} & \phi_{k2} & \dots & \phi_{kk} \end{bmatrix},$$

and ε_t is a multivariate white noise with mean vector $\underline{\mathbf{0}}$ and covariance matrix Σ. Firstly, it can be proven that equation (3.1) is stationary if the eigenvalues of Φ lie inside the unit circle.

Thus, the process can be proven using k times backward iteration as follows:

$$X_t = \Phi^t X_0 + \sum_{i=0}^{\infty} \Phi^i \varepsilon_{t-i}$$

The mean of X_t is $E(X_t) = 0$, and $\lim_{t\to\infty} \Phi^t = 0$ if the eigenvalues of Φ lie inside the unit circle and $Var(X_t)$ is constant if the eigenvalues of Φ also lie inside the unit circle as:

$$Var(X_t) = \sum_{i=1}^{\infty} \Phi^i Var(\varepsilon_{t-i})(\Phi^i)^T = \sum_{i=1}^{\infty} \Phi^i \Sigma (\Phi^i)^T.$$

Since the eigenvalues of Φ lie inside the unit circle, $Var(X_t)$ is finite ($\sum_{i=1}^{\infty} \Phi^i \Sigma (\Phi^i)^T$ is a convergent series). Secondly, by presenting the AR(p) model as a vector of the AR(1) model. Let $\lambda(I)$ be a matrix that has eigenvalues in its diagonal and zeros elsewhere. Therefore, the eigenvalues of $\lambda(I)$ can be applied for to coefficients of the companion matrix of the autoregressive model as follows:

$$X_t = \Phi X_{t-1} + \varepsilon_t \tag{3.2}$$

Equation (3.2) can be written in terms of vectors as follows:

$$X_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where X_{t-1} is a vector of size $(k \times 1)$ and Φ is called the companion matrix of the characteristic polynomial. The companion matrix can be used in order to find the upper and lower bound on the roots.

The variance matrix is as follows:

$$Var(\varepsilon_t) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \sum \quad \text{and} \quad \Phi = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

It can be proven that the eigenvalues of Φ are the inverse of the vector of $\phi(z)$ as follows. Let p=2, the companion matrix is now

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$$

In order to find the eigenvalues of Φ, the solution of $|\phi - \lambda I| = 0$ is needed (see (Hoffman and Kunze, 1971)). It can be shown that the eigenvalues satisfy the inverse of polynomial $\phi(Z)$ as follows:

$$\left|\Phi^{(2)} - \lambda I\right| = \left| \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = \begin{vmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{vmatrix} = 0;$$

then

$$\left|\Phi^{(2)} - \lambda I\right| = (\phi_1 - \lambda)(-\lambda) - \phi_2 = \lambda^2 - \phi_1\lambda - \phi_2 = 0$$

and by dividing by λ^2 we get

$$\left|\Phi^{(2)} - \lambda I\right| = 1 - \frac{\phi_1}{\lambda} - \frac{\phi_2}{\lambda^2} = 0 \tag{3.3}$$

Therefore, by letting $z = \lambda^{-1}$, the equation (3.3) is $\phi(z)$ =0; this can be written as follows:

$$\left|\Phi^{(2)} - \lambda I\right| = 1 - \phi_1 z - \phi_2 z^2 = 0$$

It can be observed that if the roots of $\phi(z)$ lie outside the unit circle, then it can be noted that $\lambda = \frac{1}{z}$ lies inside the unit circle when $|z| > 1$. It is clear that $\{y_t\}$ is stationary if and only if the root λ lies inside the unit circle. The determination between time series of the AR(p) models and the eigenvalue solutions of $\left|\Phi^{(p)} - \lambda I\right| = 0$. Thus, if we take the order p=3, the companion matrix of Φ is as follows:

$$\Phi^{(3)} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

The solution of the characteristic roots of the $\Phi^{(3)}$ matrix is as follows:

$$\left|\Phi^{(3)} - \lambda I\right| = \left|\begin{pmatrix} \phi_1 & \phi_2 & \phi_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}\right| = \begin{vmatrix} \phi_1 - \lambda & \phi_2 & \phi_3 \\ 1 & -\lambda & 0 \\ 0 & 1 & -\lambda \end{vmatrix} = 0,$$

and

$$\begin{aligned} \left|\Phi^{(3)} - \lambda I\right| = 0 \quad &\Rightarrow \quad (-\lambda)(-\phi_1\lambda + \lambda^2 - \phi_2) - (0 - \phi_3) + 0 = 0 \\ &\Rightarrow \quad \phi_1\lambda^2 - \lambda^3 + \phi_2\lambda + \phi_3 = 0 \\ &\Rightarrow \quad -\lambda^3 + \phi_1\lambda^2 + \phi_2\lambda + \phi_3 = 0 \end{aligned}$$

By dividing the above equation by $-\lambda^3$, we can obtain:

$$\left|\Phi^{(3)} - \lambda I\right| \Rightarrow 1 - \frac{\phi_1}{\lambda} - \frac{\phi_2}{\lambda^2} - \frac{\phi_3}{\lambda^3} = 0 \tag{3.4}$$

Therefore, by letting $z = \lambda^{-1}$, the equation (3.4) is $|\phi - \lambda I| = 0$, this can be written as follows:

$$\left|\Phi^{(3)} - \lambda I\right| \Rightarrow 1 - \phi_1 z - \phi_2 z^2 - \phi_3 z^3 = 0$$

It can be seen that if the solution of $\phi(z) = 0$ lie outside the unit circle, then we have reached the point that $\lambda = \frac{1}{z}$ lies inside the unit circle when $|z| < 1$. It is clear that $\{y_t\}$ is stationary. Thus, the eigenvalues of the corresponding ϕ_1, ϕ_2 and ϕ_3 lie inside the unit circle.

For $p \geq 3$,

$$\Phi^{(p)} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

Now we can write $\left|\Phi^{(p)} - \lambda I\right|$ as follows

$$\begin{aligned} \left|\Phi^{(p)} - \lambda I\right| &= \begin{vmatrix} \phi_1 - \lambda & \phi_2 & \dots & \dots & \phi_p \\ 1 & -\lambda & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -\lambda \end{vmatrix} \\ &= (-\lambda) \begin{vmatrix} \phi_1 - \lambda & \phi_2 & \dots & \phi_{p-1} \\ 1 & -\lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & -\lambda \end{vmatrix} - (1) \begin{vmatrix} \phi_1 - \lambda & \phi_2 & \dots & \phi_p \\ 1 & -\lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{vmatrix} \end{aligned}$$

We can write $\left|\Phi^{(p)} - \lambda I\right| = 0$ as follows:

$$\begin{aligned}
\left|\Phi^{(p)} - \lambda I\right| &= (-\lambda)\left|\Phi^{(p-1)} - \lambda I\right| + \phi_p \\
&= (-\lambda)(\lambda^{p-1} - \sum_{i=1}^{p-1} \phi_i \ \lambda^{p-i-1}) + \phi_p \\
&= (-\lambda)(\lambda^{p-1} - (\phi_1 \lambda^{p-2} + \phi_2 \lambda^{p-3} + \cdots + \phi_{p-1}\lambda^0) + \phi_p \\
&= -\lambda^p + \phi_1 \lambda^{p-1} + \phi_2 \lambda^{p-2} + \cdots + \phi_{p-1}\lambda + \phi_p = 0 \qquad (3.5)
\end{aligned}$$

By dividing the above equation (3.5) by($-\lambda^p$), we can obtain:

$$\left|\Phi^{(p)} - \lambda I\right| = 1 - \frac{\phi_1}{\lambda} - \cdots - \frac{\phi_{p-1}}{\lambda^{p-1}} - \frac{\phi_p}{\lambda^p} = 0$$

Therefore, by letting $z = \lambda^{-1}$, we can get;

$$\left|\Phi^{(p)} - \lambda I\right| = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_{p-1} z^{p-1} - \phi_p z^p = 0$$

Thus, it can be noted that the eigenvalues of $\lambda(I)$ for the coefficients of the AR(p) model are less than one when the polynomial $\phi(z)$ lies outside the unit circle because of the fact that $|z| > 1$. Therefore, it can be said that X_t is the series of a stationary process because the characteristic roots lie within the unit circle.

3.3 Difference equation and back-shift operators

The AR model can be defined with regards to an inhomogeneous difference equation. This can be represented with a backshift operator. The time series y_t is AR(p) if it satisfies the following equation

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \varepsilon_t \quad t \in Z$$

where ε_t is white noise (ε_t is iid with 0 mean and some variance). The autoregressive model is a difference equation that could be an infinite number of simultaneous equations. Thus, to get a general solution for this problem, the AR model is written in terms of the backshift operator.

$$y_t - \phi_1 B y_t - \cdots - \phi_p B^p y_t = \varepsilon_t \quad \Rightarrow \quad \phi(B) y_t = \varepsilon_t$$

where $\phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$, and B is the backshift operator. This can be defined as $\phi(B)Y_t = \varepsilon_t$ which provides the solution of the AR difference equation to be $y_t = \phi(B)^{-1}\varepsilon_t$ ((Kulahci and Bisgaard, 2011), (Hipel and McLeod, 1994) and (Box et al., 2008)).

3.4 Autoregressive model of order p

Autoregressive models abbreviated AR(p) are based on the thought that the current value of the series $\{y_t\}$, can be explained as a function of its p past values, $(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$, according to

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{3.6}$$

where $\{y_t\}$ is stationary, $\phi_1, \phi_2, \ldots, \phi_p$ are constants and $\phi_p \neq 0$. Unless otherwise stated, we assume that $\{\varepsilon_t\}$ is a *Gaussian white noise* series with mean zero $(\mu = 0)$ and variance σ_ε^2 . The mean of y_t in equation (3.6) is zero. If the mean, μ , of y_t is not zero, replace (y_t) by $y_t - \mu$ in (3.6), i.e.,

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t,$$

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{3.7}$$

where $\alpha = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$. We note that (3.7) is similar to the regression model, and hence the term auto (or self) regression. Some technical difficulties, however, develop from applying that model since the regressors, $y_{t-1}, \ldots, y_{t-p}$, are random components, whereas ε_t

is fixed. A suitable form of the backward shift operator is as follows:

$$\phi(B)y_t = \varepsilon_t, \tag{3.8}$$

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

The simplest autoregressive model is

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \tag{3.9}$$

where the ε_t is white noise, $WN(0, \sigma^2)$, and y_t is stationary if only if $|\phi_1| < 1$ (Brockwell and Davis, 2001) and (Shumway and Stoffer, 2011).

3.4.1 Stationarity conditions of the AR(2) model

In this section we discuss the AR(2) model in more detail and we derive the stationarity region for this model; we propose an alternative and simpler proof. Our derivation is not based on the Schur-Cohn criterion which is adopted in the proof of Marmol (1995) and Najim (2010). Consider the AR(2) process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{3.10}$$

where y_t is stationary and ϕ_1, ϕ_2 are the AR coefficients. Unless otherwise stated, we assume that ε_t is a Gaussian white noise process, i.e., ε_t is *iid with mean zero and variance* σ_ε^2. We can define B to be the backshift operator, so that $B^i y_t = y_{t-i}$; we can then write (3.10) as

$$(1 - \phi_1 B - \phi_2 B^2)y_t = \varepsilon_t, \tag{3.11}$$

or

$$\phi(B)y_t = \varepsilon_t.$$

It can be noticed that (3.10) is stationary when the roots of $(1-\phi_1 z-\phi_2 z^2)=\phi(z)$ lie outside the unit circle. Next, we derive the stationary condition for AR(2). The roots of $\phi(z)$ are

$$z=\frac{\phi_1\pm\sqrt{\phi_1^2+4\phi_2}}{-2\phi_2}$$

The roots of $\phi(z)$ might be real and distinct, real and equal or a complex conjugate pair. Let

$$z_1=\frac{\phi_1+\sqrt{\phi_1^2+4\phi_2}}{-2\phi_2} \text{ and } z_2=\frac{\phi_1-\sqrt{\phi_1^2+4\phi_2}}{-2\phi_2} \tag{3.12}$$

Then,

$$\begin{aligned} z_1 &= \frac{\phi_1+\sqrt{\phi_1^2+4\phi_2}}{-2\phi_2}.\frac{\phi_1-\sqrt{\phi_1^2+4\phi_2}}{\phi_1-\sqrt{\phi_1^2+4\phi_2}} \\ &= \frac{\phi_1^2-(\phi_1^2+4\phi_2)}{-2\phi_2.[\phi_1-\sqrt{\phi_1^2+4\phi_2}]} \end{aligned}$$

So

$$z_1=\frac{2}{\phi_1-\sqrt{\phi_1^2+4\phi_2}} \text{ and } z_2=\frac{2}{\phi_1+\sqrt{\phi_1^2+4\phi_2}},$$

Then from (3.12), z_1^{-1}, z_2^{-1} are

$$z_1^{-1}=\frac{\phi_1-\sqrt{\phi_1^2+4\phi_2}}{2}, \tag{3.13}$$

$$z_2^{-1}=\frac{\phi_1+\sqrt{\phi_1^2+4\phi_2}}{2}. \tag{3.14}$$

We can write $\phi(z) = (1 - z_1^{-1}z)(1 - z_2^{-1}z)$ and so in model (3.10) we can write

$$\varepsilon_t = (1 - z_1^{-1}B)(1 - z_2^{-1}B)y_t$$

From this representation, it follows that $\phi_1 = (z_1^{-1} + z_2^{-1})$ by adding equations (3.13) and (3.14) as follows

$$z_1^{-1} + z_2^{-1} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} + \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} = \phi_1 \tag{3.15}$$

Also

$$\phi_2 = -(z_1 z_2)^{-1}, \tag{3.16}$$

since

$$\Rightarrow z_1^{-1} z_2^{-1} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} = -\phi_2$$

From the definition of stationarity of (3.10) implies $\mid z_i \mid > 1$ thus $\mid \phi_2 \mid = \mid z_1^{-1} \mid\mid z_2^{-1} \mid < 1$.

Case(1): if $\phi(z) = 0$ has two real roots: $(\phi_1^2 + 4\phi_2) \geq 0$

$$-1 < z^{-1} < 1,$$

$$\Rightarrow -1 < z_1^{-1} < z_2^{-1} < 1.$$

$$-1 < z_1^{-1} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} < z_2^{-1} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} < 1. \tag{3.17}$$

From the right-hand side of (3.17):

$$\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} < 1,$$

$$\Rightarrow \sqrt{\phi_1^2 + 4\phi_2} < 2 - \phi_1,$$

$$\Rightarrow \phi_1^2 + 4\phi_2 < 4 - 4\phi_1 + {\phi_1}^2,$$

$$\phi_2 + \phi_1 < 1. \tag{3.18}$$

From the left-hand side of (3.17).

$$\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} > -1,$$

$$\Rightarrow 2 + \phi_1 > \sqrt{\phi_1^2 + 4\phi_2},$$

$$\Rightarrow 4 + 4\phi_1 + \phi_1^2 > \phi_1^2 + 4\phi_2,$$

$$\phi_2 - \phi_1 < 1 \tag{3.19}$$

Case(2): if $\phi(z) = 0$ has two complex conjugate roots: $(\phi_1^2 + 4\phi_2) < 0$

$$\phi_2 < \frac{-\phi_1^2}{4},$$

because $|\ \phi_2\ | < 1$ this implies,

$$-1 < \phi_2 < \frac{-\phi_1^2}{4}. \tag{3.20}$$

A triangle region in the parameter space can be specified by these stationarity conditions. The equations of (3.18), (3.19) and (3.20) allow us to describe the stationary and non-stationary regions of the parameter space in Figure 3.1.

It can be noticed from Figure 3.1 that the triangle shape is the stationary region of real roots and the curved shape is the stationary region of complex roots.

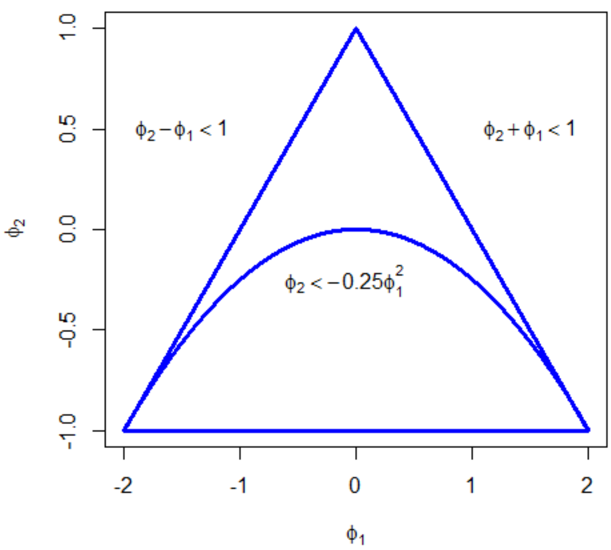


Figure 3.1: The stationarity regions of real roots and complex roots of the AR(2)

3.4.2 Simulation of the stationary region AR(2)

As mentioned before, the stationary region of the AR(2) is a triangle. To ensure the result of the theory, see section 3.4.1, the region of stationarity for complex roots and real roots is a triangle according to the selected poles, as shown in Figure 3.2. We simulate data from an AR(2) and then for each sample we estimate the parameters of ϕ_1 and ϕ_2 , but it is clear that the estimated parameters can result in stationary or non-stationary time series. We plot all parameters that satisfy the stationarity conditions. For this purpose, we increase sample sizes to ensure, through plotting, the whole region of real roots and complex roots is covered which we theoretically proved (see Figure 3.2). It can be noticed that as sample sizes increased, the whole region of real roots is covered for the AR(2) model. However, for the complex roots the whole region is not covered. When sample sizes increased, the shape of the plot approaches its real triangle shape as mentioned in section 3.4.1.

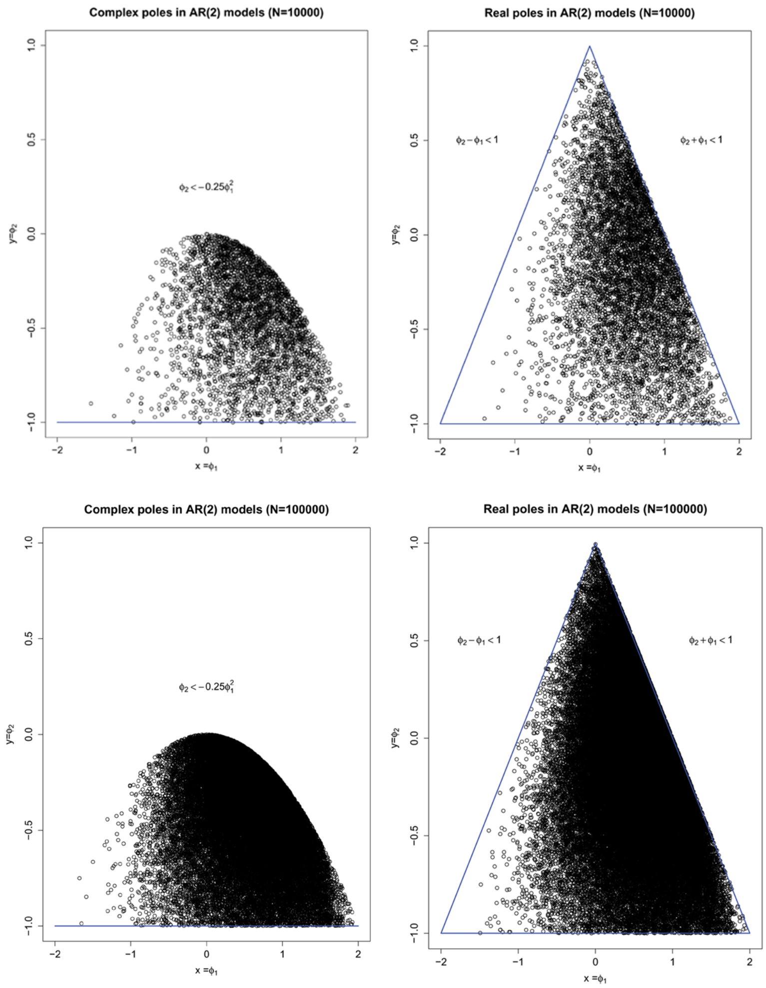


Figure 3.2: Stationarity region of real roots and complex roots in the AR(2) for the different sample sizes obtained from simulated parameters.

3.5 Stationary conditions of autoregressive models

In this section we propose a new set of stationarity conditions for autoregressive models of any order p. We motivate the new methodology by considering first the autoregressive model of order three, AR(3).

3.5.1 Autoregressive model of order $p = 3$

Existing stationarity conditions

Consider that the time series $\{y_t\}$ is generated by the AR(3) autoregressive model defined by

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t, \tag{3.21}$$

where $\{\varepsilon_t\}$ is a white noise process. In the sequel we derive inequalities involving ϕ_1, ϕ_2, ϕ_3 that are sufficient and necessary for the stationarity of $\{y_t\}$. These inequalities have been known since Barndorff-Nielsen and Schou (1973) and their proof is provided by Okuguchi and Irie (1990) and Farebrother (1992), both of which make use of the Schur-Cohn criterion. Our derivation is more direct and not based on the Schur-Cohn criterion. Then we propose a new set of inequalities, which overcome several of the problems encountered in the existing stationarity conditions.

First, we establish a relationship of the roots and the coefficients ϕ_1, ϕ_2 and ϕ_3 with the backward shift operator B, defined as usual by $B^j y_t = y_{t-j}$ (shifting the time series y_t j time points backwards). We can write compactly the model (3.21) as

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3) y_t = \varepsilon_t \quad \text{or} \quad \phi(B) y_t = \varepsilon_t,$$

where $\phi(B)$ is the autoregressive characteristic polynomial in B (here of order $p = 3$). Let $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \phi_3 x^3$ be the characteristic polynomial (in the complex valued x). The classic root criterion states that the time series $\{y_t\}$ is stationary if and only if the roots of $\phi(x)$ lie outside the unit circle. Dividing $\phi(x) = 0$ by x^3, we have that

$$\frac{1}{x^3} - \frac{\phi_1}{x^2} - \frac{\phi_2}{x} - \phi_3 = 0$$

Thus, by letting $z = x^{-1}$, the equation $\phi(x) = 0$ can be written as

$$z^3 - \phi_1 z^2 - \phi_2 z - \phi_3 = 0 \tag{3.22}$$

Obviously, $\{y_t\}$ is stationary if and only if the roots of (3.22) lie inside the unit circle. We now give the correspondence of the roots ρ_1, ρ_2, ρ_3 of (3.22) and the coefficients ϕ_1, ϕ_2, ϕ_3. We write (3.22) as

$$(z - \rho_1)(z - \rho_2)(z - \rho_3) = 0$$

and expand this to get

$$(z^2 - \rho_1 z - \rho_2 z + \rho_1\rho_2)(z - \rho_3) = 0$$

$$z^3 - \rho_1 z^2 - \rho_2 z^2 + \rho_1\rho_2 z - \rho_3 z^2 + \rho_1\rho_3 z + \rho_2\rho_3 z - \rho_1\rho_2\rho_3 = 0$$

$$z^3 - (\rho_1 + \rho_2 + \rho_3)z^2 + (\rho_1\rho_2 + \rho_1\rho_3 + \rho_2\rho_3)z - \rho_1\rho_2\rho_3 = 0 \tag{3.23}$$

If we compare equations (3.22) and (3.23) we obtain

$$\phi_1 = \rho_1 + \rho_2 + \rho_3 \tag{3.24}$$

$$\phi_2 = -\rho_1\rho_2 - \rho_1\rho_3 - \rho_2\rho_3 \tag{3.25}$$

$$\phi_3 = \rho_1\rho_2\rho_3. \tag{3.26}$$

In the next pages we will show that a necessary and sufficient set of conditions for the stationarity of $\{y_t\}$ is

$$\phi_1 + \phi_2 + \phi_3 < 1 \tag{3.27}$$

$$-\phi_1 + \phi_2 - \phi_3 < 1 \tag{3.28}$$

$$\phi_3(\phi_3 - \phi_1) - \phi_2 < 1 \tag{3.29}$$

$$|\phi_3| < 1 \tag{3.30}$$

Necessity: Under the assumption of stationarity we have $|\rho_i| < 1$ for all $i = 1, 2, 3$, which from (3.26) immediately implies condition (3.30). Assume first ρ_1, ρ_2 and ρ_3 are real.

Next we prove (3.27)-(3.29). We start with (3.27).

$$\begin{aligned} \phi_1 + \phi_2 + \phi_3 &= \rho_1 + \rho_2 + \rho_3 - \rho_1\rho_2 - \rho_1\rho_3 - \rho_2\rho_3 + \rho_1\rho_2\rho_3 \\ &= \rho_1(1-\rho_2) + \rho_3(1-\rho_2) - \rho_1\rho_3(1-\rho_2) + \rho_2 \\ &= (1-\rho_2)(\rho_1 + \rho_3 - \rho_1\rho_3) + \rho_2 \qquad (3.31) \\ &< 1 - \rho_2 + \rho_2 = 1, \end{aligned}$$

since

$$\rho_1 + \rho_3 - \rho_1\rho_3 = \rho_1(1-\rho_3) + \rho_3 < 1, \quad \text{as} \quad |\rho_1| < 1.$$

Similarly for (3.28) we have

$$\begin{aligned} -\phi_1 + \phi_2 - \phi_3 &= -\rho_1 - \rho_2 - \rho_3 - \rho_1\rho_2 - \rho_2\rho_3 - \rho_1\rho_3 - \rho_1\rho_2\rho_3 \\ &= -\rho_1(1+\rho_2) - \rho_3(1+\rho_2) - \rho_1\rho_3(1+\rho_2) - \rho_2 \\ &= (1+\rho_2)(-\rho_1 - \rho_3 - \rho_1\rho_3) - \rho_2 \qquad (3.32) \\ &< 1 + \rho_2 - \rho_2 = 1, \end{aligned}$$

since

$$-\rho_1 - \rho_3 - \rho_1\rho_3 = -\rho_1(1+\rho_3) - \rho_3 < 1, \quad \text{as} \quad |\rho_1| < 1.$$

Finally, for (3.28) we have

$$\begin{aligned}\phi_3(\phi_3 - \phi_1) - \phi_2 &= \rho_1\rho_2\rho_3(\rho_1\rho_2\rho_3 - \rho_1 - \rho_2 - \rho_3) + \rho_1\rho_2 + \rho_1\rho_3 + \rho_2\rho_3 \\ &= (1 - \rho_1\rho_3)(\rho_1\rho_2 - \rho_1\rho_2^2\rho_3 + \rho_2\rho_3) + \rho_1\rho_3 \\ &= (1 - \rho_1\rho_3)[\rho_1\rho_2(1 - \rho_2\rho_3) + \rho_2\rho_3] + \rho_1\rho_3 \qquad (3.33) \\ &< 1 - \rho_1\rho_3 + \rho_1\rho_3 = 1,\end{aligned}$$

since

$$\rho_1\rho_2(1 - \rho_2\rho_3) + \rho_2\rho_3 < 1, \quad \text{as} \quad |\rho_1\rho_2| < 1.$$

Now suppose that there are two complex conjugate roots and one real root. Without loss of generality suppose that ρ_2 is real and ρ_1, ρ_3 are the two complex roots. Write $\rho_1 = a + bi$ and $\rho_3 = a - bi$, for some real a and b, where i denotes the imaginary unit. We have $\rho_1 + \rho_3 - \rho_1\rho_3 = 2a - a^2 - b^2 < 1$, since $|\rho_1| = \sqrt{a^2 + b^2} < 1$, hence from (3.31) we have $\phi_1 + \phi_2 + \phi_3 < 1$. Similarly, from $-\rho_1 - \rho_3 - \rho_1\rho_3 = -2a - a^2 - b^2 < 1$ and (3.32), we obtain $-\phi_1 + \phi_2 - \phi_3 < 1$. We also have

$$\begin{aligned}\rho_1\rho_2(1 - \rho_2\rho_3) + \rho_2\rho_3 &= \rho_2(a + bi)[(1 - \rho_2(a - bi)] + \rho_2(a - bi) \\ &= 2\rho_2 a - (\rho_2 a)^2 - (\rho_2 b)^2 < 1,\end{aligned}$$

since ρ_2 is real and $(\rho_2 a)^2 + (\rho_2 b)^2 < 1$. Thus, from (3.33) we have $\phi_3(\phi_3 - \phi_1) - \phi_2 < 1$. This proves (3.27)-(3.30) for complex roots.

Sufficiency: Now we prove that if conditions (3.27)-(3.30) are satisfied, then $\{y_t\}$ is stationary. From condition (3.30) and equation (3.26) at least one of $|\rho_1|, |\rho_2|, |\rho_3|$ must be strictly less than one. Without loss of generality suppose $|\rho_2| < 1$.

Consider first the case of real roots ρ_1, ρ_2, ρ_3. Assume that $\{y_t\}$ were not stationary, that is $|\rho_1| \geq 1$ or $|\rho_3| \geq 1$. If $\rho_3 > 1$, then $\rho_1 + \rho_3(1-\rho_1) \geq \rho_1 + 1 - \rho_1 = 1$, hence from (3.31) we have $\phi_1 + \phi_2 + \phi_3 \geq 1$, which contradicts condition (3.27). If $\rho_3 < -1$, then $-\rho_1 - \rho_3(1+\rho_1) \geq -\rho_1 + 1 + \rho_1 = 1$, hence from (3.32) we have $-\phi_1 + \phi_2 - \phi_3 \geq 1$, which contradicts condition (3.28). By interchanging the roles of ρ_1 and ρ_3 we obtain that $|\rho_1| \geq 1$ contradicts either (3.27) or (3.28). Hence, it is necessary that $|\rho_1| < 1$, $|\rho_2| < 1$ and $|\rho_3| < 1$, i.e., $\{y_t\}$ is stationary.

Consider now the case of two complex conjugate roots. As before and without loss of generality we assume that ρ_1, ρ_3 are complex, while ρ_2 is real. We write as before $\rho_1 = a + bi$ and $\rho_3 = a - bi$. As before, from condition (3.30) and $|\phi_3| = |\rho_1||\rho_2||\rho_3| < 1$ we have that at least one of ρ_1, ρ_2, ρ_3 has modulus less than one; without loss of generality assume $|\rho_2| < 1$. Suppose that we have $|\rho_1| = |\rho_3| = \sqrt{a^2+b^2} \geq 1$. From (3.33) we have

$$\phi_3(\phi_3 - \phi_1) - \phi_2 = (1 - a^2 - b^2)[2\rho_2 a - (\rho_2 a)^2 - (\rho_2 b)^2] + a^2 + b^2. \tag{3.34}$$

Put $u = a^2+b^2$, $A = 2\rho_2 a - (\rho_2 a)^2 - (\rho_2 b)^2$ and $B = (1-u)A + u$. Since $u \geq 1$, if $A \leq 1$, we have $B \geq 1$. We note that for $A \leq 0$, $(1-u)A \geq 0$ and so $B = (1-u)A + u \geq 1$, since $u \geq 1$. If $0 < A \leq 1$, then $(1-A)u \leq 1 - A$ or $B = (1-u)A + u \geq 1$. We can see that $A > 1$ is not possible. Indeed with the definition of A as above we have

$$(\rho_2 a - 1)^2 = 1 + \rho_2^2 a^2 - 2\rho_2 a > -\rho_2^2 b^2 \quad \text{or} \quad A = 2\rho_2 a - \rho_2^2 a^2 - \rho_2^2 b^2 < 1.$$

Thus, from (3.34) we have $\phi_3(\phi_3 - \phi_1) - \phi_2 \geq 1$, which contradicts (3.29). Hence $|\rho_1| < 1$, $|\rho_2| < 1$ and $|\rho_3| < 1$, i.e., $\{y_t\}$ is stationary. We note that condition (3.29) introduces non-linear terms in the left hand side of the inequality and hence it is this condition that prevents generalisation to higher orders $p > 3$. Even if interest is restricted to AR(3), it is not easy to place a prior distribution on (ϕ_1, ϕ_2, ϕ_3) satisfying (3.29).

The new stationarity conditions

We propose that condition (3.29) can be replaced by two linear inequalities, which are simple to interpret (like the inequalities (3.27) and (3.28)) and can be generalised for any order p. We show that sufficient conditions for the stationarity of AR(3) are

$$\phi_1 + \phi_2 + \phi_3 < 1 \tag{3.35}$$

$$-\phi_1 + \phi_2 - \phi_3 < 1 \tag{3.36}$$

$$-\phi_1 - \phi_2 + \phi_3 < 1 \tag{3.37}$$

$$\phi_1 - \phi_2 - \phi_3 < 1 \tag{3.38}$$

$$|\phi_3| < 1. \tag{3.39}$$

Inequalities (3.35), (3.36), (3.39) are the same as before, but now (3.29) is replaced by (3.37) and (3.38). We prove that (3.35)-(3.39) are sufficient for stationarity. First we prove that if conditions (3.35)-(3.39) are satisfied, then $\{y_t\}$ is stationary. It suffices to prove that conditions (3.35)-(3.39) imply conditions (3.27)-(3.30). To this end, we need to prove that the two conditions (3.37) and (3.38) imply (3.29). Indeed,

$$\begin{aligned} \phi_3(\phi_3 - \phi_1) &\leq |\phi_3(\phi_3 - \phi_1)| \\ &= |\phi_3||\phi_3 - \phi_1| \\ &< |\phi_3 - \phi_1|, \end{aligned} \tag{3.40}$$

since $|\phi_3| < 1$.

- If $\phi_3 \geq \phi_1$, from $-\phi_1 - \phi_2 + \phi_3 < 1$ we have $|\phi_3 - \phi_1| = \phi_3 - \phi_1 < 1 + \phi_2$, thus from (3.40) we obtain $\phi_3(\phi_3 - \phi_1) - \phi_2 < 1$.

- If $\phi_3 < \phi_1$, from $\phi_1 - \phi_2 - \phi_3 < 1$, we have $|\phi_3 - \phi_1| = \phi_1 - \phi_3 < 1 + \phi_2$, thus from

(3.40) we obtain $\phi_3(\phi_3 - \phi_1) - \phi_2 < 1$.

Thus, in every case condition (3.27) is satisfied, hence $\{y_t\}$ is stationary.

3.5.2 Checking sufficiency for the stationary conditions of the AR(3) model

In order to illustrate sufficiency of the stationary conditions for the AR(3) model, we discuss some examples in order to know whether the parameters satisfy both recommended stationary conditions and Barndorff-Nielsen and Schou (1973)'s conditions. This can be done by assigning a set of parameters randomly for the AR(3) model that satisfy both our recommended stationary conditions and Barndorff-Nielsen and Schou (1973)'s conditions as mentioned in equations (3.27)-(3.29). Suppose we assign the values of the parameters of the AR(3) model as follows:

$$y_t = -0.4y_{t-1} - 0.8y_{t-2} - 0.6y_{t-3} + \varepsilon_t$$

It can be noted that the parameters of the above AR(3) model satisfy all four inequality stationary conditions of (3.35)-(3.38) and Barndorff-Nielsen and Schou (1973)'s stationary conditions of (3.27)-(3.29). Additionally, we assigned different values for the parameters of the AR(3) model as follows:

$$y_t = -0.38y_{t-1} - 0.42y_{t-2} + 0.36y_{t-3} + \varepsilon_t$$

It can be seen in the second example that the parameters also satisfy the stationarity conditions of (3.27)-(3.29) proposed by Barndorff-Nielsen and Schou (1973). However, the parameters only satisfy three conditions of inequality stationary conditions. The fourth condition (3.37) is not satisfied with $-\phi_1 - \phi_2 + \phi_3 = 1.16$. Therefore, our recommended stationary conditions are not sufficient in this example. For this reason, in order to understand better the sufficiency of our recommended stationarity conditions in comparison with existing conditions a simulation study is conducted.

Because our aim is to generalize the stationarity conditions, there might be other conditions that are sufficient but not necessary and indeed much simpler conditions might be possible. However, our proposed conditions are likely to be specific to a particular lag. Thus, it will be difficult to propose a general formula. Because of the fact that we want to generalize, we propose our stationarity conditions. Later in this chapter we provide a general theorem to which our stationarity conditions are a particular case.

3.5.3 Grouping conditions of the AR(3) model

After realizing the sufficient stationarity conditions (3.35)-(3.39) are not necessary too, we put these stationarity conditions into groups. Essentially, we put the proposing stationarity conditions that we have into groups for presentation purposes. The stationarity conditions of group A are as follows:

$$\phi_1 + \phi_2 + \phi_3 < 1$$
$$-\phi_1 + \phi_2 - \phi_3 < 1$$
$$-\phi_1 - \phi_2 + \phi_3 < 1$$

Furthermore, the stationarity conditions of group B are as follows:

$$\phi_1 + \phi_2 + \phi_3 < 1$$
$$-\phi_1 + \phi_2 - \phi_3 < 1$$
$$\phi_1 - \phi_2 - \phi_3 < 1$$

If we put all the conditions from groups A and B together, we end up with another group which we call group AB as follows:

$\phi_1 + \phi_2 + \phi_3 < 1$

$-\phi_1 + \phi_2 - \phi_3 < 1$

$-\phi_1 - \phi_2 + \phi_3 < 1$

$\phi_1 - \phi_2 - \phi_3 < 1$

As we have seen the sufficient and necessary group conditions are either group A or B. We also saw that if we put all these conditions together, we get the union of group A and B which is group AB. The group AB is a subgroup of one of the other two groups. Based on the stationary conditions of the AR(3) model, stationary conditions of AR(p) models are generalized for $p > 3$

3.5.4 Autoregressive model of order $p \geq 4$.

We turn our attention to sufficient conditions and derive a general set of sufficient stationarity conditions for AR(p). Later on in chapter 4 we will use these sufficient conditions to build a prior for MCMC schemes. Consider the general autoregressive model of order $p \geq 1$, defined by

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \tag{3.41}$$

where as before $\{\varepsilon_t\}$ is a white noise process. The following theorem provides a means of building up the stationarity conditions. In fact, this is a general theorem which is not a particular for the AR(4). Here, we use this theorem in order to build up the sufficient conditions for the (AR) model.

Theorem 3.1. *Consider a time series* $\{y_t\}$ *generated by* (3.41) *and define two new time series* $\{x_t\}$ *and* $\{z_t\}$ *of lag order* $p-1$ *as:*

$$x_t = (\phi_1 + \phi_2)x_{t-1} + \phi_3 x_{t-2} + \cdots + \phi_p x_{t-p+1} + \varepsilon_{xt},$$

$$z_t = (\phi_1 - \phi_2)z_{t-1} - \phi_3 z_{t-2} - \cdots - \phi_p z_{t-p+1} + \varepsilon_{zt},$$

If $\{x_t\}$ and $\{z_t\}$ are stationary then $\{y_t\}$ is stationary too.

Proof. We establish the following representations of y_t:

$$\begin{aligned} y_t &= (\phi_1+\phi_2)y_{t-1}+\phi_3 y_{t-2}+\cdots+\phi_p y_{t-p+1}+\varepsilon_t \\ &\quad -\phi_2 y_{t-1}+\phi_2 y_{t-2}+\phi_3 y_{t-3}+\cdots+\phi_p y_{t-p} \\ &\quad -\phi_3 y_{t-2}-\cdots-\phi_p y_{t-p+1} \\ &= (\phi_1+\phi_2)y_{t-1}+\phi_3 y_{t-2}+\cdots+\phi_p y_{t-p+1}+\varepsilon_t \\ &\quad -\phi_2(y_{t-1}-y_{t-2})-\phi_3(y_{t-2}-y_{t-3})-\cdots-\phi_p(y_{t-p+1}-y_{t-p}) \\ &= (\phi_1+\phi_2)y_{t-1}+\sum_{i=3}^{p}\phi_i y_{t-i+1}+\varepsilon_t-\sum_{i=2}^{p}\phi_i(y_{t-i+1}-y_{t-i}) \qquad (3.42) \end{aligned}$$

and

$$\begin{aligned} y_t &= (\phi_1-\phi_2)y_{t-1}-\phi_3 y_{t-2}-\cdots-\phi_p y_{t-p+1}+\varepsilon_t \\ &\quad +\phi_2 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p} \\ &\quad +\phi_3 y_{t-2}+\cdots+\phi_p y_{t-p+1} \\ &= (\phi_1-\phi_2)y_{t-1}-\sum_{i=3}^{p}\phi_i y_{t-i+1}+\varepsilon_t+\sum_{i=2}^{p}\phi_i(y_{t-i+1}+y_{t-i}) \qquad (3.43) \end{aligned}$$

Since $\{x_t\}$ is stationary, from (3.42), it follows that

$$y_t+\sum_{i=2}^{p}\phi_i(y_{t-i+1}-y_{t-i}) \qquad (3.44)$$

is a stationary process. Also, since $\{z_t\}$ is stationary, from (3.43), it follows that

$$y_t-\sum_{i=2}^{p}\phi_i(y_{t-i+1}+y_{t-i}) \qquad (3.45)$$

If we add (3.44) and (3.45) we have that

$$y_t - \sum_{i=2}^{p} \phi_i y_{t-i}$$

is stationary and by using y_t as in (3.41) we have found $\phi_1 y_{t-1}$ to be stationary. Thus $\{y_t\}$ is stationary. If x_t and z_t are stationary then y_t is stationary too. However, if y_t is stationary, then x_t and z_t might not be stationary. Therefore, the converse of the theorem is not true. We give an example later in order to illustrate this. □

For $p=3$ we can see that the sufficient stationarity conditions (3.35) - (3.39) of the AR(3) model (3.21) can be obtained by the stationarity conditions of the two AR(2) models

$$x_t = (\phi_1 + \phi_2)x_{t-1} + \phi_3 x_{t-2} + \varepsilon_{xt} = \phi_1^* x_{t-1} + \phi_2^* x_{t-2} + \varepsilon_{xt}, \tag{3.46}$$

$$z_t = (\phi_1 - \phi_2)z_{t-1} - \phi_3 z_{t-2} + \varepsilon_{zt} = \phi_1^{**} z_{t-1} + \phi_2^{**} z_{t-2} + \varepsilon_{zt}. \tag{3.47}$$

From the stationarity conditions of AR(2) we have

$$\phi_1^* + \phi_2^* < 1 \Rightarrow \phi_1 + \phi_2 + \phi_3 < 1$$

$$-\phi_1^* + \phi_2^* < 1 \Rightarrow -\phi_1 - \phi_2 + \phi_3 < 1$$

$$\phi_1^{**} + \phi_2^{**} < 1 \Rightarrow \phi_1 - \phi_2 - \phi_3 < 1$$

$$-\phi_1^{**} + \phi_2^{**} < 1 \Rightarrow -\phi_1 + \phi_2 - \phi_3 < 1$$

$$|\phi_2^*| < 1, \quad |\phi_2^{**}| < 1 \quad \Rightarrow |\phi_3| < 1,$$

which are exactly conditions (3.35) - (3.39).

Moving to $p=4$ we can see that the sufficient stationarity conditions (3.48) - (3.56) of the AR(4) model ($y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t$) can be obtained by the stationarity conditions of the following two AR(3) models.

$$x_t = (\phi_1 + \phi_2)x_{t-1} + \phi_3 x_{t-2} + \phi_4 x_{t-3} + \varepsilon_{xt}$$
$$z_t = (\phi_1 - \phi_2)z_{t-1} - \phi_3 z_{t-2} - \phi_4 z_{t-3} + \varepsilon_{zt}$$

In order to proceed further, we need the necessary conditions of x_t and z_t being stationary.

Case 1: If both x_t and z_t satisfy the AB conditions, then we obtain the results in eight conditions for the AR(4) model.

Case 2: If one of x_t and z_t satisfy the conditions of groups A or B, then we obtain a new group of conditions including seven conditions for the AR(4) model.

Case 3: If both x_t and z_t satisfy the conditions of either groups A or B (but not AB), then we obtain another new group of conditions containing six conditions for the AR(4) model.

In all three cases we have the extra condition $|\phi_4| < 1$. In the application of Theorem 3.1 we ignore cases 2 or 3 and go with case 1. For $p \geq 4$ some justifications for this choice are as follows:

- We do not know the necessary stationarity conditions and as p increases the structure of the groups gets more complicated. As groups vary from one to another the conditions change within the same AR(p) model.

- AB group conditions include all four inequality conditions to be satisfied. We show in section 3.5.7 that based on simulated values, AB covers around 70% of the stationarity region of the AR(3) model. However, A group conditions, which include three inequality conditions, covers around 15% of the stationarity region of the AR(3) model and group B conditions covers the remaining 15% of the stationarity region. For a given time series, we have either that the group conditions of AB which include all conditions are satisfied or three conditions are satisfied in which case we have only group conditions A or B but not AB.

- Adopting group AB we impose stronger conditions (rather than, e.g., A or B) and as p increases this has the effect to shrink the resulting stationarity region towards the middle of the axes (see Figure 3.5). This suggests that it is unlikely in general for these conditions to be necessary as well, in particular for large p. However, as we aim to use these conditions in order to construct weakly informative priors (see chapter 4), the shrinking of the stationarity region is not a big concern. In fact for the AR(3) model we show in chapter 4 that we get similar posterior samples whether we operate with a true A or B or AB.

Operating as above by adopting case 1 enables us to go from $p-1$ to p for $p \geq 4$ using only the proposed sufficient conditions.

As a result we can use conditions (3.35) - (3.39) of AR(3), in order to derive the stationarity conditions of the AR(4) model, defined by

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t.$$

It turns out that the conditions are

$$\phi_1 + \phi_2 + \phi_3 + \phi_4 < 1 \tag{3.48}$$

$$-\phi_1 - \phi_2 + \phi_3 - \phi_4 < 1 \tag{3.49}$$

$$-\phi_1 - \phi_2 - \phi_3 + \phi_4 < 1 \tag{3.50}$$

$$\phi_1 + \phi_2 - \phi_3 - \phi_4 < 1 \tag{3.51}$$

$$\phi_1 - \phi_2 - \phi_3 - \phi_4 < 1 \tag{3.52}$$

$$-\phi_1 + \phi_2 - \phi_3 + \phi_4 < 1 \tag{3.53}$$

$$-\phi_1 + \phi_2 + \phi_3 - \phi_4 < 1 \tag{3.54}$$

$$\phi_1 - \phi_2 + \phi_3 + \phi_4 < 1 \tag{3.55}$$

$$|\phi_4| < 1. \tag{3.56}$$

There are nine conditions; the first four come from the stationarity conditions of the AR(3) time series x_t, the next four come from the conditions of z_t and the last condition comes from both.

Next we give the number of inequalities involved for any $p \geq 1$. The following table shows the number of inequalities for orders $p = 1, 2, 3, 4$.

Order p	No. of inequalities n_p
1	1
2	3
3	5
4	9

Let n_p be the number of inequalities of order p. From Theorem 3.1 we observe that for each $p \geq 3$ we have inequalities coming from two AR $(p-1)$ models, so we have

$$n_p = 2(n_{p-1} - 1) + 1 = 2n_{p-1} - 1.$$

For example, we can trivially verify this from the table.

Now writing n_p recursively we have

$$\begin{aligned} n_p &= 2n_{p-1} - 1 \\ &= 2(2n_{p-2} - 1) - 1 = 2^2 n_{p-2} - 2^1 - 2^0 = \cdots \\ &= 2^{p-2} n_2 - \sum_{i=0}^{p-3} 2^i \\ &= 3 \times 2^{p-2} - \frac{2^{p-2} - 1}{2 - 1} \\ &= 3 \times 2^{p-2} - 2^{p-2} + 1 \\ &= 2 \times 2^{p-2} + 1 \\ &= 2^{p-1} + 1, \end{aligned} \tag{3.57}$$

for any $p \geq 3$.

Note that $n_p = 2^{p-1} + 1$ also works for $p = 2$, as it is trivial to verify, and hence (3.57) is true for any $p > 1$.

Finally, we provide an efficient way to compute the stationarity conditions of any p. The proposed sufficient conditions for the stationarity of the time series $\{y_t\}$ generated by the AR (p) model (3.41) can be compactly written in matrix form as

$$A_p \phi \prec \mathbf{1} \quad \text{and} \quad |\phi_p| < 1, \tag{3.58}$$

where $\prec$ denotes element wise inequality $(<)$, A_p is a $p \times p$ matrix of elements 1 and -1 (see below),

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \text{and} \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

We define $A_1 = 1$ and for any $p \geq 2$, the matrix A_p is given by

$$A_p = \begin{bmatrix} A_{p-1} & 1^*_{2^{p-2}} \\ -A_{p-1} & 1^*_{2^{p-2}} \end{bmatrix},$$

where for any even positive integer $n \geq 2$, 1^*_n is an n-dimensional column vector with successive elements 1 and -1, i.e.,

$$1^*_n = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \\ 1 \\ -1 \end{bmatrix},$$

while for $n = 1$ we set $1^*_1 = 1$.

For example, we have

$$A_2 = \begin{bmatrix} A_1 & 1^*_1 \\ -A_1 & 1^*_1 \end{bmatrix}$$

and (3.58) yields

$$A_2\phi = \begin{bmatrix} A_1 & 1^*_1 \\ -A_1 & 1^*_1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \prec \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

which together with $|\phi_2| < 1$ (given in (3.58)) gives the stationarity conditions of AR(2), i.e., $\phi_1 + \phi_2 < 1$, $-\phi_1 + \phi_2 < 1$, $|\phi_2| < 1$.

For $p = 3$ we have

$$A_3 = \begin{bmatrix} A_2 & 1_2^* \\ -A_2 & 1_2^* \end{bmatrix} = \left[\begin{array}{cc|c} 1 & 1 & 1 \\ -1 & 1 & -1 \\ \hline -1 & -1 & 1 \\ 1 & -1 & -1 \end{array}\right]$$

Thus, (3.58) yields

$$A_3\phi = \left[\begin{array}{cc|c} 1 & 1 & 1 \\ -1 & 1 & -1 \\ \hline -1 & -1 & 1 \\ 1 & -1 & -1 \end{array}\right] \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} \prec \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

which together with $|\phi_3| < 1$ yields the stationarity conditions (3.35) - (3.39).

For $p = 4$ we have

$$A_4 = \begin{bmatrix} A_3 & 1_4^* \\ -A_3 & 1_4^* \end{bmatrix} = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ \hline -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{array}\right].$$

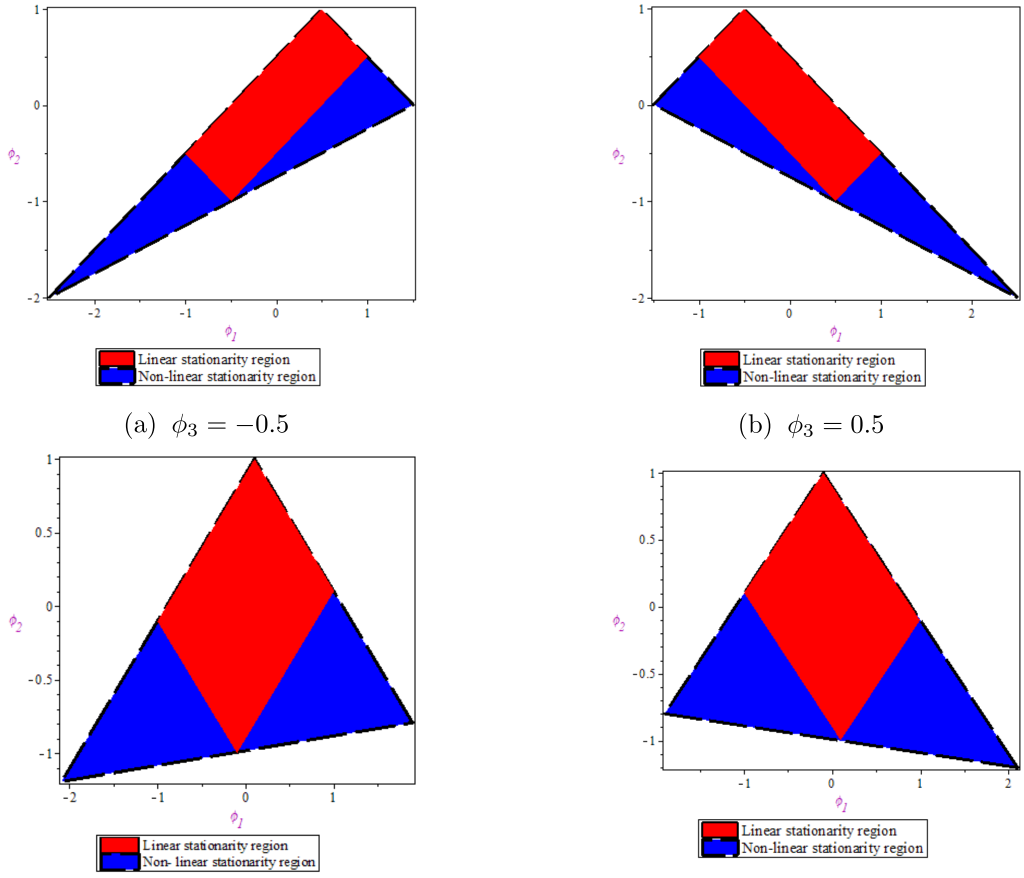
Thus, (3.58) yields

$$A_4\phi = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ \hline -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{array}\right] \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{bmatrix} \prec \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

which together with $|\phi_3| < 1$ yields the stationarity conditions given in page 57.

3.5.5 Comparing stationarity regions for linear and non-linear condition of the AR(3) model

In this section we aim to compare the stationary region of our linear inequality conditions of (3.35) - (3.39) with the stationary region of the non-linear inequality condition of (3.27) - (3.30). In general, it is hard to compare our linear conditions with non-linear equations by visualizing 3-D graphs. Thus, in order to see how the our stationary region would look like with regard to the stationary region of non-linear conditions of Barndorff-Nielsen and Schou (1973), we construct 2-D graphs which are easier to interpret. We first fixed ϕ_3 using different values between (-1, 1) and then we constructed 2-D graphs of linear and non-linear inequality conditions. We used the following four different values of ϕ_3: -0.5, -0.1, 0.1 and 0.5.

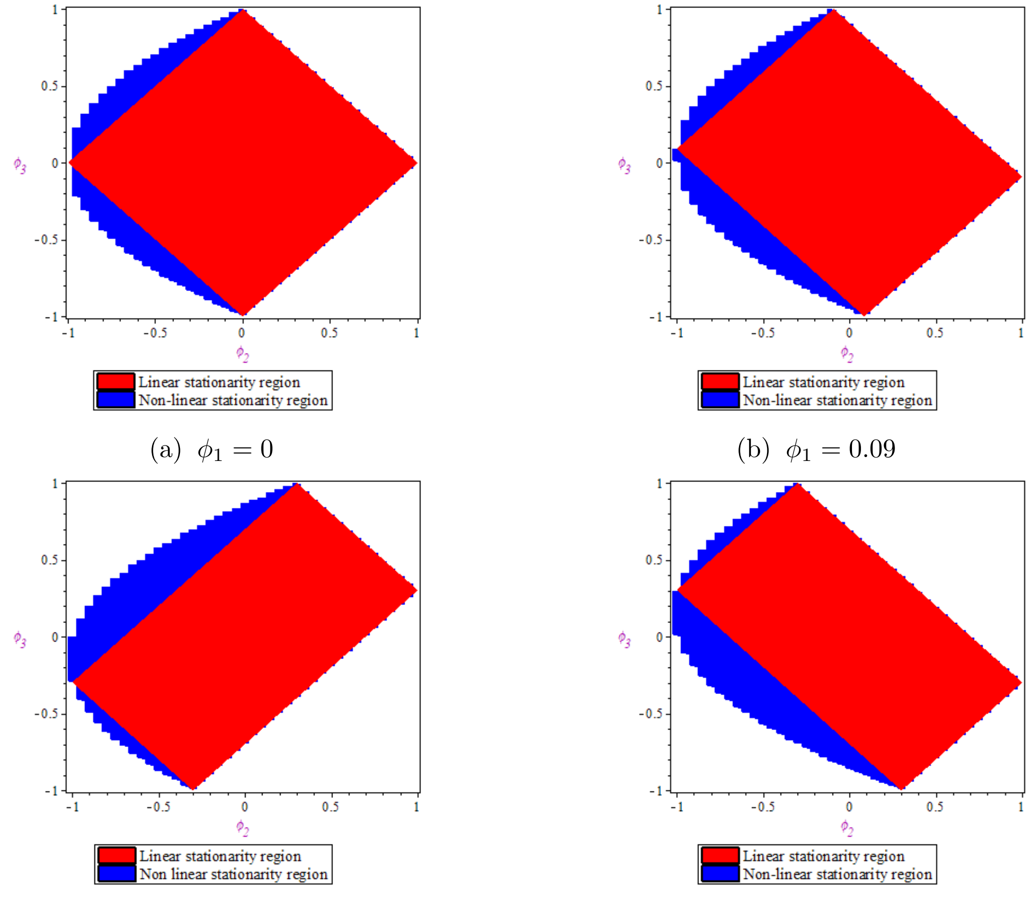


(a) $\phi_3 = -0.5$ (b) $\phi_3 = 0.5$

(c) $\phi_3 = -0.1$ (d) $\phi_3 = 0.1$

Figure 3.3: Show the stationarity regions for the fixed value of ϕ_3 of the two group conditions. The blue shaped area shows the difference between the two group conditions and the red shaped area shows the stationarity region from our proposed conditions. The big triangle shows the overall stationarity conditions.

We fixed ϕ_3 in Figure 3.3 in order to compare the stationarity regions of linear with non-linear conditions in 2-D plot. It can be noticed from Figure 3.3 that the stationary region of linear inequality conditions are not the same as the stationary region of non-linear inequality conditions by Barndorff-Nielsen and Schou (1973). However, the stationary region of our linear inequality conditions is within the non-linear stationary region.



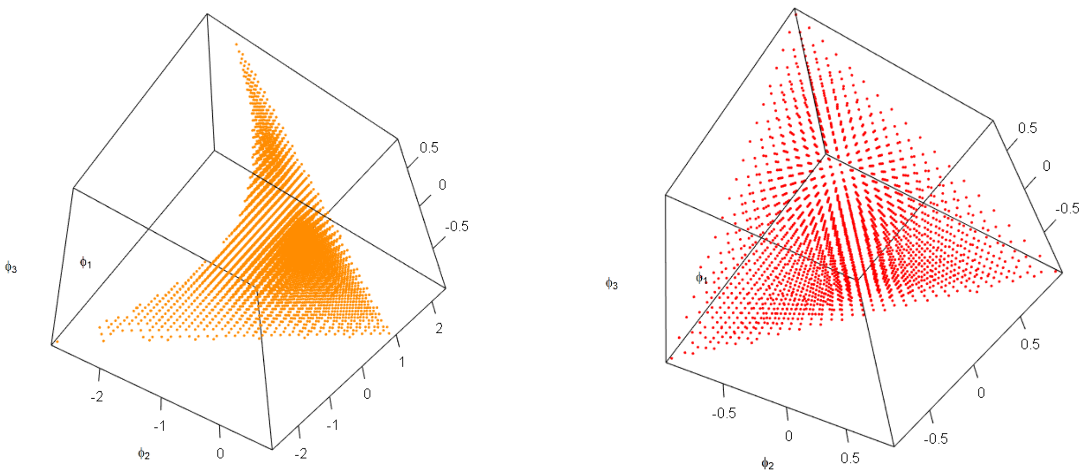
(a) $\phi_1 = 0$ (b) $\phi_1 = 0.09$

(c) $\phi_1 = -0.3$ (d) $\phi_1 = 0.3$

Figure 3.4: Show the stationarity regions for the fixed value of ϕ_1 of the two group conditions. The blue shaped area shows the difference between the two group conditions and the red shaped area shows the stationarity region from our proposed conditions. The big triangle shows the overall stationarity conditions.

Additionally, we need to see the effect of quadratic term in non-linear conditions when we compare them with our linear conditions. This can be done by fixing either ϕ_1 or ϕ_2. Figure 3.4 shows the stationarity regions for the fixed value of ϕ_1 of the two group conditions. There are some differences between the two group conditions. However, our linear conditions, which is a red shaped area, cover a wide areas of non-linear conditions which is a blue shaped area. Obviously, from figures of 3.3 and 3.4, we observed that there are significant differences between linear and non-linear conditions. But, we use our proposed linear conditions in order to identify prior distribution and we will show that later for MCMC.

Moreover, we can simulate values in order to compare the stationarity regions of both inequality conditions by visualizing 3-D graphs. Thus, we simulated values of ϕ_1, ϕ_2 and ϕ_3 in order to see how the stationary region of our linear conditions would look like based on the stationary region of non-linear conditions. Then, we perform a rejection sampling on the simulated parameters in order to save only those values which satisfy the inequality conditions of Barndorff-Nielsen and Schou (1973).



(a) Non-linear (b) Linear

Figure 3.5: A 3D plot of linear and non-linear stationarity conditions for the AR(3) model using simulated values

Despite the fact that side of the 3-D plot has a similar triangular shape, the two shapes do not coincide. The ranges of all three parameters of the linear inequality conditions are between(-1, 1). However, the ranges of ϕ_1 , ϕ_2 and ϕ_3 were between -2.9 and 2.3, -2.9 and 1, and -1 and 1 respectively. We have seen from both the 2-D figure 3.3 and the 3-D figure 3.5 that the stationary region of our linear inequality conditions is not the same as the stationary region of the non-linear conditions, but our stationary region is within the stationary region of the non-linear conditions. Therefore, in order to cover the stationary region completely as Barndorff-Nielsen and Schou (1973) has covered, the idea of grouping stationarity conditions of the AR(3) model can be used. This is presented in the next section.

3.5.6 Comparing stationarity regions for linear grouping conditions and non-linear conditions of the AR(3) model

We have already seen that AB is a set of sufficient conditions for the AR(3) model. In this section, we aim to find necessary conditions for the AR(3) model. When we have all four conditions, we have sufficient conditions. However, when we are trying to look at necessary conditions, they are not always AB. The necessary conditions can be AB but they can be three out of the four conditions

Figures 3.3, 3.4 and 3.5 suggest that our linear conditions do not completely cover the stationarity conditions of Barndorff-Nielsen and Schou (1973). Therefore, we use the idea of sub setting conditions. Group AB contains all four inequality conditions of (3.35) - (3.38). Group B contains three inequality conditions from equation (3.35)- (3.37), and we take out the equation of (3.37) in order to obtain the group conditions B.

Regarding the grouping of stationarity conditions, for a given time series, either we have AB conditions for which all conditions are satisfied or a subset of three conditions are satisfied in which we have only A or B but not A and B. It can be seen from Figure 3.6 that switching conditions for the AR(3) model is adequate, when we compare the proposed conditions with the non-linear stationarity conditions of Barndorff-Nielsen and Schou (1973).

Note that the right hand-side of Figure 3.6 is based on the new conditions of stationarity whereas the left side is Barndorff-Nielsen and Schou (1973)'s conditions. It is noted that as long as the interval becomes wider, the range of ϕ_1 and ϕ_2 are not changed and the same plots are obtained. So that the intervals for the parameters are not exceeded from -2.9, 2.3 for ϕ_1 and -2.9, 1 for ϕ_2 .

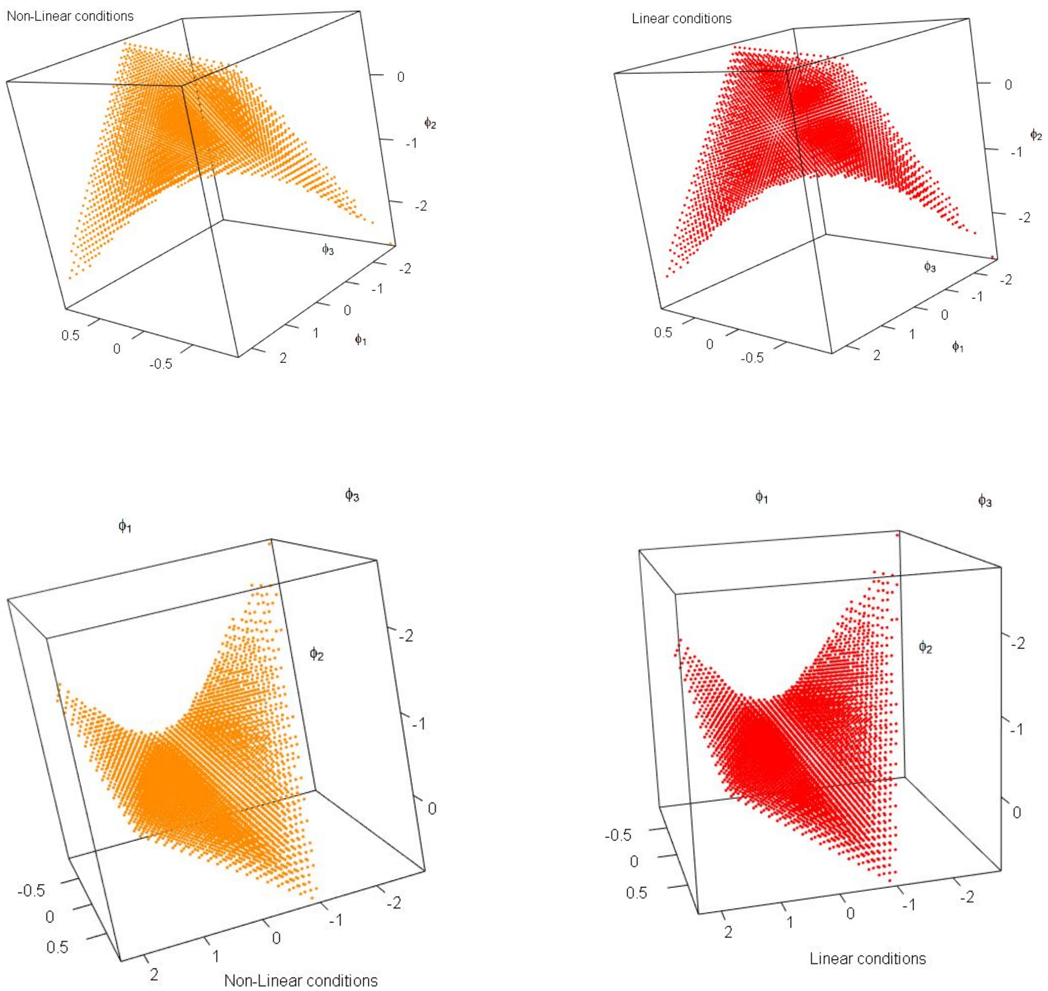


Figure 3.6: 3D plots of grouping of linear stationarity conditions and non-linear stationarity conditions for the AR(3) model using simulated values presenting different angles.

It would seem that the obtained result from Figure 3.6, the results obtained from groups A, B and AB cover the same stationarity region of non-linear conditions.

3.5.7 Explanation of the stationary conditions of the AR(3) model

It was mentioned earlier that stationary conditions for the AR(3) model can be achieved based on the non-linear conditions available from the study of Barndorff-Nielsen and Schou (1973). In order to determine the precision of the recommended linear stationary conditions for the AR(3) model, the corresponding relationship between the partial auto-correlations π_i and the parameters ϕ_i from the study of Barndorff-Nielsen and Schou (1973) are employed. The reason behind this is to confirm that the stationary conditions are totally 100% achieved for parameters in the AR(3) model. As has been mentioned before, the corresponding relationship between partial auto-correlations and parameters is as follows:

$$\phi_1 = \pi_1 - \pi_1\pi_2 - \pi_2\pi_3 \tag{3.59}$$

$$\phi_2 = \pi_2 - \pi_1\pi_3 + \pi_1\pi_2\pi_3 \tag{3.60}$$

$$\phi_3 = \pi_3 \tag{3.61}$$

According to the study of Barndorff-Nielsen and Schou (1973), if the values of partial auto-correlations are between -1 and 1, then the group conditions can be satisfied. Based on this information, n observations are simulated for partial auto-correlations that are uniformly distributed on [-1, 1]. Parameter estimations of ϕ_1, ϕ_2 and ϕ_3 can be calculated for simulated π's based on equations (3.59)-(3.61). After obtaining on $n \times 3$ matrix of parameters ϕ_s, the inequality conditions are checked for each of the stationary group conditions of A, B and AB (see sections 3.5.5 and 3.5.6). Figure 3.7 illustrates the percentage of satisfaction of parameters that satisfy the stationarity conditions for the linear group conditions and the non-linear conditions of the Barndorff-Nielsen and Schou (1973) study. The result indicates that the parameters that satisfy the stationarity conditions of the Barndorff-Nielsen and Schou (1973) study can 100% satisfy the stationary linear

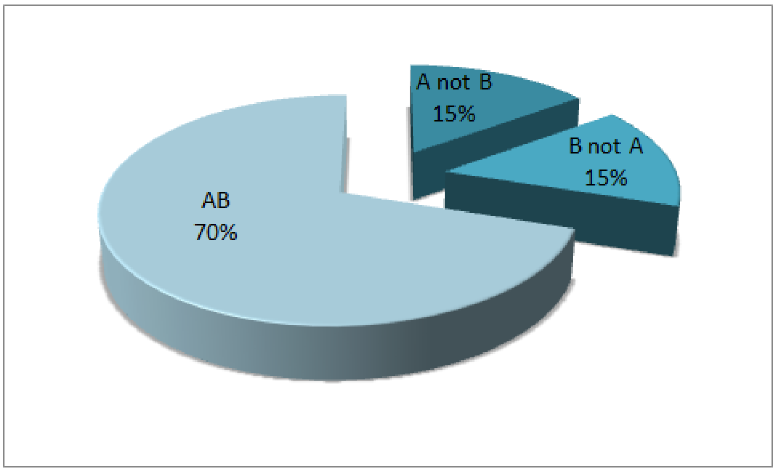


Figure 3.7: The percentage of satisfaction for the stationary conditions from non-linear to linear conditions.

conditions, but in different percentages for the three group conditions as shown in Figure 3.7. For a given time series, we have either that the group conditions of AB which contain all conditions are satisfied or that three conditions are satisfied in which case we have only group conditions A or B but not AB. Thus, the parameters can be satisfied for at least one of the groups of A, B or AB for the AR(3) model.

3.6 Simulation study for the stationarity conditions of the AR model

In this section, a simulation study is conducted in order to check and illustrate the sufficiency and necessity of conditions of the AR(3) and AR(4) models. One of the issues that arises in the statistical analysis of autoregressive models is the very complex nature of the domain of the regression parameters (Barndorff-Nielsen and Schou, 1973). These authors discovered a corresponding relationship between the parameters of the autocorrelation function (ϕ_s) and the partial autocorrelation function (π_s). The idea is that the correspondence relationship between π_s and ϕ_s can be used to simulate from and to partial auto-correlations. In order to confirm the relationships by simulation, we want to simulate π_s and ϕ_s for checking the sufficiency and necessity for the AR(3) and AR(4) models.

By simulating partial auto-correlations between (-1,1), AR coefficients can be obtained and then stationarity conditions can be checked. When parameters ϕ_s are simulated, the partial autocorrelation can be calculated in order to check sufficiency. Additionally, the stationarity regions are detected for both linear inequality and linearity conditions using rejection sampling on the simulated bounds. Although the corresponding relationship of π_s and ϕ_s is one-to-one, there is no general formula for obtaining an expression of π_s as a function of ϕ_s. In this and the next sections we derive this correspondence relationship.

3.6.1 Mapping of partial correlation into parameters for the AR(3) model

The corresponding relationships between parameters of the AR(3) model and partial autocorrelations as described by Barndorff-Nielsen and Schou (1973) are

$$\phi_1 = \pi_1 - \pi_1\pi_2 - \pi_2\pi_3 \tag{3.62}$$

$$\phi_2 = \pi_2 - \pi_1\pi_3 + \pi_1\pi_2\pi_3 \tag{3.63}$$

$$\phi_3 = \pi_3 \tag{3.64}$$

where $|\pi_i| < 1$. The objective here is to find an expression of π_i as a function of ϕ_i. From equation (3.64) we know that

$$\pi_3 = \phi_3,$$

and from (3.62) we also know that

$$\phi_1 + \pi_2\pi_3 = \pi_1 - \pi_1\pi_2$$

$$\phi_1 + \pi_2\pi_3 = \pi_1(1 - \pi_2)$$

Therefore, the expression of π_1 is given by

$$\pi_1 = \frac{\phi_1 + \phi_3\pi_2}{1 - \pi_2} \tag{3.65}$$

Also from (3.63), we obtain

$$\phi_2 + \pi_1\phi_3 = \pi_2 + \pi_1\pi_2\phi_3$$

thus, the expression of π_2 is given by

$$\pi_2 = \frac{\phi_2 + \phi_3\pi_1}{1 + \pi_1\phi_3} \tag{3.66}$$

It can be seen that π_2 in the equation of (3.66) depends only on ϕ_2, ϕ_3 and π_1. By substituting equation(3.65) into (3.66), the expression of π_1 can be obtained as follows:

$$\begin{aligned}\pi_1 &= \frac{\phi_1 + \phi_3(\frac{\phi_2+\phi_3\pi_1}{1+\pi_1\phi_3})}{1 - (\frac{\phi_2+\phi_3\pi_1}{1+\pi_1\phi_3})} \\ &= \frac{\phi_1 + \phi_2\phi_3}{1 - \phi_2 - \phi_1\phi_3 - \phi_3^2}\end{aligned} \tag{3.67}$$

Now by adding equation (3.67) to equation (3.66), we can get the expression of π_2 as follows

$$\pi_2 = \frac{\phi_2 + \phi_3(\frac{\phi_1+\phi_2\phi_3}{1-\phi_2-\phi_1\phi_3-\phi_3^2})}{1 + \phi_3(\frac{\phi_1+\phi_2\phi_3}{1-\phi_2-\phi_1\phi_3-\phi_3^2})} \tag{3.68}$$

Therefore, the expression of π_2 using the corresponding relationship is given by

$$\begin{aligned}\pi_2 &= \frac{\phi_1\phi_3 - \phi_1\phi_2\phi_3 + \phi_2 - \phi_2^2}{1 - \phi_2 - \phi_3^2 + \phi_2\phi_3^2} \\ &= \frac{(\phi_1\phi_3 + \phi_2)(1 - \phi_2)}{(1 - \phi_3^2)(1 - \phi_2)}\end{aligned}$$

Thus,

$$\pi_2 = \frac{\phi_1\phi_3 + \phi_2}{1 - \phi_3^2} \tag{3.69}$$

In conclusion, an expression of π_i as a function of ϕ_i can be given as follows:

$$\pi_1 = \frac{\phi_1 + \phi_2\phi_3}{1 - \phi_2 - \phi_1\phi_3 - \phi_3^2} \tag{3.70}$$

$$\pi_2 = \frac{\phi_1\phi_3 + \phi_2}{1 - \phi_3^2} \tag{3.71}$$

$$\pi_3 = \phi_3 \tag{3.72}$$

3.6.2 Mapping of partial correlations into parameters for the AR(4) model

Moving on to AR(4) we derive the relationship of π_i as a function of ϕ_i. From Barndorff-Nielsen and Schou (1973), we get

$$\phi_1 = \pi_1 - \pi_1\,\pi_2 - \pi_2\,\pi_3 - \pi_3\,\pi_4 \tag{3.73}$$

$$\phi_2 = \pi_2 - \pi_1\,\pi_3 - \pi_2\,\pi_4 + \pi_1\,\pi_2\,\pi_3 + \pi_1\,\pi_3\,\pi_4 - \pi_1\,\pi_2\,\pi_3\,\pi_4 \tag{3.74}$$

$$\phi_3 = \pi_3 - \pi_1\,\pi_4 + \pi_1\,\pi_2\,\pi_4 + \pi_2\,\pi_3\,\pi_4 \tag{3.75}$$

$$\phi_4 = \pi_4 \tag{3.76}$$

For the purpose of both sufficiency and necessity, the study will depend on the conditions above, i.e., equations (3.73)-(3.76). However, to confirm necessity for the provided conditions from the AR(4) model, we will use the information that we have in equations (3.73)-(3.76) in order to obtain the mapping of partial correlations into parameters for the AR(4) model. Therefore, the following equations (3.77)-(3.80) can be obtained from the above equations (3.73)-(3.76):

$$\pi_1 = \frac{-\pi_2\,\pi_3 - \pi_3\,\pi_4 - \phi_1}{\pi_2 - 1} \tag{3.77}$$

$$\pi_2 = \frac{\pi_1\pi_3\pi_4 - \pi_1\pi_3 - \phi_2}{-\pi_1\pi_3 + \pi_1\pi_3\pi_4 + \pi_4 - 1} \tag{3.78}$$

$$\pi_3 = \frac{\pi_1\pi_2\pi_4 - \pi_1\pi_4 - \phi_3}{-\pi_2\pi_4 - 1} \tag{3.79}$$

$$\pi_4 = \phi_4 \tag{3.80}$$

We need to reach the point that the right hand side of equations (3.77)-(3.80) do not include π_s. By substituting (3.80) into (3.77) we obtain,

$$\pi_1 = \frac{-\pi_2\pi_3 - \pi_3\phi_4 - \phi_1}{\pi_2 - 1} \tag{3.81}$$

Then by substituting (3.81) into (3.79) we obtain the following equation for π_3

$$\begin{aligned}\pi_3 &= \left(\frac{(-\pi_2\pi_3 - \pi_3\phi_4 - \phi_1)\,\pi_2\phi_4}{\pi_2 - 1} - \frac{(-\pi_2\pi_3 - \pi_3\phi_4 - \phi_1)\,\phi_4}{\pi_2 - 1} - \phi_3\right)(-\pi_2\phi_4 - 1)^{-1} \\ &= \frac{-\phi_1\phi_4 - \phi_3}{{\phi_4}^2 - 1} \end{aligned}\tag{3.82}$$

Then by substituting (3.82) into (3.81) we obtain

$$\begin{aligned}\pi_1 &= \left(-\frac{\pi_2\,(-\phi_1\phi_4 - \phi_3)}{{\phi_4}^2 - 1} - \frac{(-\phi_1\phi_4 - \phi_3)\,\phi_4}{{\phi_4}^2 - 1} - \phi_1\right)(\pi_2 - 1)^{-1} \\ &= \frac{\pi_2\phi_1\phi_4 + \phi_3\pi_2 + \phi_3\phi_4 + \phi_1}{\left({\phi_4}^2 - 1\right)(\pi_2 - 1)} \end{aligned}\tag{3.83}$$

Then substituting (3.83) and (3.82) into (3.74) we obtain

$$\pi_2 = \left(\frac{(\pi_2\phi_1\phi_4 + \phi_3\pi_2 + \phi_3\phi_4 + \phi_1)(-\phi_1\phi_4 - \phi_3)\phi_4}{\left({\phi_4}^2 - 1\right)^2(\pi_2 - 1)} - \frac{(\pi_2\phi_1\phi_4 + \phi_3\pi_2 + \phi_3\phi_4 + \phi_1)(-\phi_1\phi_4 - \phi_3)}{\left({\phi_4}^2 - 1\right)^2(\pi_2 - 1)} - \phi_2 \right) \left(-\frac{(\pi_2\phi_1\phi_4 + \phi_3\pi_2 + \phi_3\phi_4 + \phi_1)(-\phi_1\phi_4 - \phi_3)}{\left({\phi_4}^2 - 1\right)^2(\pi_2 - 1)} + \frac{(\pi_2\phi_1\phi_4 + \phi_3\pi_2 + \phi_3\phi_4 + \phi_1)(-\phi_1\phi_4 - \phi_3)\phi_4}{\left({\phi_4}^2 - 1\right)^2(\pi_2 - 1)} + \phi_4 - 1 \right)^{-1}$$

Then by simplifying the above equation, we obtain the following equation for π_2:

$$\pi_2 = -\frac{\phi_1\phi_3{\phi_4}^2 - \phi_2{\phi_4}^3 + {\phi_1}^2\phi_4 - \phi_2{\phi_4}^2 + {\phi_3}^2\phi_4 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_2}{{\phi_1}^2{\phi_4}^2 - {\phi_4}^4 + 2\phi_1\phi_3\phi_4 + {\phi_3}^2 + 2{\phi_4}^2 - 1} \tag{3.84}$$

It can be seen that after reaching the point above that there is no more π_5 on the right-hand side of (3.84). We now obtain the following equation for π_1 by substituting (3.84) into (3.83)

$$\pi_1 = \left(-\frac{\left(\phi_1\phi_3{\phi_4}^2 - \phi_2{\phi_4}^3 + {\phi_1}^2\phi_4 - \phi_2{\phi_4}^2 + {\phi_3}^2\phi_4 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_2\right)\phi_1\phi_4}{{\phi_1}^2{\phi_4}^2 - {\phi_4}^4 + 2\phi_1\phi_3\phi_4 + {\phi_3}^2 + 2{\phi_4}^2 - 1} - \frac{\left(\phi_1\phi_3{\phi_4}^2 - \phi_2{\phi_4}^3 + {\phi_1}^2\phi_4 - \phi_2{\phi_4}^2 + {\phi_3}^2\phi_4 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_2\right)\phi_3}{{\phi_1}^2{\phi_4}^2 - {\phi_4}^4 + 2\phi_1\phi_3\phi_4 + {\phi_3}^2 + 2{\phi_4}^2 - 1} + \phi_3\phi_4 + \phi_1 \right) \left({\phi_4}^2 - 1\right)^{-1} \left(-\frac{\phi_1\phi_3{\phi_4}^2 - \phi_2{\phi_4}^3 + {\phi_1}^2\phi_4 - \phi_2{\phi_4}^2 + {\phi_3}^2\phi_4 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_2}{{\phi_1}^2{\phi_4}^2 - {\phi_4}^4 + 2\phi_1\phi_3\phi_4 + {\phi_3}^2 + 2{\phi_4}^2 - 1} - 1 \right)^{-1} \tag{3.85}$$

Clearly, it can be seen that equation (3.85) is mathematically very complicated in order to determine the mapping of partial correlation into parameters for π_1 and π_2. Therefore, the computer software Maple (version 18) is used to overcome this problem. After using the Maple software, the result below is obtained.

$$\pi_1 = -\frac{\phi_1\phi_2\phi_4 - \phi_3{\phi_4}^2 - \phi_1\phi_4 + \phi_2\phi_3 + \phi_3\phi_4 + \phi_1}{{\phi_1}^2\phi_4 + \phi_1\phi_3\phi_4 - \phi_2{\phi_4}^2 - {\phi_4}^3 + \phi_1\phi_3 + {\phi_3}^2 + {\phi_4}^2 + \phi_2 + \phi_4 - 1} \tag{3.86}$$

The final mapping of partial correlations Π into parameters(Φ) for the AR(4) model is as follows:

$$\pi_1 = -\frac{\phi_1\phi_2\phi_4 - \phi_3{\phi_4}^2 - \phi_1\phi_4 + \phi_2\phi_3 + \phi_3\phi_4 + \phi_1}{{\phi_1}^2\phi_4 + \phi_1\phi_3\phi_4 - \phi_2{\phi_4}^2 - {\phi_4}^3 + \phi_1\phi_3 + {\phi_3}^2 + {\phi_4}^2 + \phi_2 + \phi_4 - 1} \tag{3.87}$$

$$\pi_2 = -\frac{\phi_1\phi_3{\phi_4}^2 - \phi_2{\phi_4}^3 + {\phi_1}^2\phi_4 - \phi_2{\phi_4}^2 + {\phi_3}^2\phi_4 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_2}{{\phi_1}^2{\phi_4}^2 - {\phi_4}^4 + 2\phi_1\phi_3\phi_4 + {\phi_3}^2 + 2{\phi_4}^2 - 1} \tag{3.88}$$

$$\pi_3 = \frac{-\phi_1\phi_4 - \phi_3}{{\phi_4}^2 - 1} \tag{3.89}$$

$$\pi_4 = \phi_4 \tag{3.90}$$

In order to confirm that our mapping of partial correlations (π) into parameters (ϕ) for AR(p) models is correct, two approaches are mentioned here.

1 . Simulating π_s .

One way to confirm this is by simulating π_s conditioning on the fact that the values of π_s should be between (-1, 1). Then the values of ϕ_s can be calculated using equations (3.73)-(3.76). After, by substituting ϕ_s into equations (3.87)-(3.90), the new values of π_s can be calculated. Then, it can be noted that the values of simulated π_s are the same as the values that the obtained from equations (3.87)-(3.90). This confirms that mapping partial correlation into parameters in AR(p) models is perfect.

2 . Cancelling ϕ_4 .

The second way to confirm that the mapping of partial correlations into parameters of AR(p) models is applicable on a wider level is that, if we cancel out ϕ_4 (ϕ_4 = 0) in equations (3.87)-(3.90), it would give the same equations (3.70) - (3.72) of mapping partial correlations into parameters in the AR(3) model. In general, in order to map partial correlations into parameters for AR(p) models, if we set $\phi_p = 0$, it would give the same equations for mapping partial correlations into parameters for the AR($p-1$) model.

3.6.3 Checking necessity of the AR(3) stationarity conditions via simulation

In this section we demonstrate the necessity of the stationarity conditions of the AR(3) model. The following steps, which are derived from the mapping equations between partial autocorrelations and parameters, will be performed (see Section 3.6). However, here we only show the latest version of our trial and put them into practice to show whether the parameters are necessary or not. In order to do that, we first simulate 300000 values for all π's via the Uniform distribution between (-1, 1); see Table 3.1. Secondly, the parameters ϕ's are computed based on the simulated π's, those parameters that satisfy stationary conditions for the AR(3) model according to equations (3.59)-(3.61).

$$\pi_1 = \frac{\phi_1 + \phi_2\phi_3}{1 - \phi_2 - \phi_1\phi_3 - \phi_3^2} \tag{3.91}$$

$$\pi_2 = \frac{\phi_1\phi_3 + \phi_2}{1 - \phi_3^2} \tag{3.92}$$

$$\pi_3 = \phi_3 \tag{3.93}$$

Finally, the simulated parameters are checked to classify the parameters that satisfy the stationary linear conditions (3.35)-(3.38). Table 3.1 shows that conditions one and two are 100% satisfied and conditions three and four are 85% satisfied, where C1, C2, C3 and C4 represent the stationarity conditions of equations (3.35)-(3.38), respectively. We narrow down these conditions into two group conditions which we call group condition A and group condition B. The former means conditions one, two, and three are all satisfied

Table 3.1: Necessity for the AR(3) stationary conditions.

n	Simulation			Calculate			C1	C2	C3	C4	A	B	Final
	π_1	π_2	π_3	ϕ_1	ϕ_2	ϕ_3							
1	-0.395	0.884	0.540	-0.523	0.909	0.540	1	1	1	1	1	1	1
2	-0.700	0.826	0.847	-0.821	0.929	0.847	1	1	1	1	1	1	1
3	0.178	-0.810	0.482	0.712	-0.965	0.482	1	1	1	0	1	0	1
4	-0.064	-0.772	-0.026	-0.134	-0.775	-0.026	1	1	1	1	1	1	1
5	-0.482	-0.670	0.017	-0.794	-0.657	0.017	1	1	0	1	0	1	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
299997	-0.691	-0.124	-0.451	-0.833	-0.474	-0.451	1	1	1	1	1	1	1
299998	-0.575	0.023	-0.490	-0.550	-0.252	-0.490	1	1	1	1	1	1	1
299999	0.542	-0.671	-0.027	0.889	-0.647	-0.027	1	1	1	0	1	0	1
300000	-0.813	-0.961	-0.147	-1.735	-1.195	-0.147	1	1	0	1	0	1	1

together, whereas, the latter means conditions one, two, and four are all satisfied at the same time. Both groups A and B are satisfied by 85% of the observations. Finally, we realized that for all the simulated data at least one of the two group conditions A or B has been met.

3.6.4 Checking sufficiency for the AR(3) stationary conditions via simulation

As we have already discussed the AR(3) model parameters do not go beyond -3 and 3 for determining the stationary region of the AR(3) model. In accordance with the study of Barndorff-Nielsen and Schou (1973), parameters ϕ_s have a relationship with the partial auto-correlations π_s (see equations (3.59)-(3.61)). Barndorff-Nielsen and Schou (1973) proved that if the values of π are between [-1,1], the attained parameters must satisfy the stationary conditions of the AR model. We use the simulation method to verify these conditions (see Table 3.2). First, 1000000 observations of parameters ϕ_s are simulated via the Uniform distribution to ensure that they have met all our four conditions explained

Table 3.2: Sufficiency for AR(3) stationary condition

n	Similation			Check condition				Filter	Calculate			Check
	ϕ_1	ϕ_2	ϕ_3	C1	C2	C3	C4		π_1	π_2	π_3	π
1	-0.155	-0.446	-0.151	1	1	1	1	1	-0.062	-0.432	-0.151	1
2	0.58	-1.892	0.662	1	1	0	0	0	-0.325	-2.685	0.662	0
3	-0.986	2.091	0.809	0	0	1	1	0	-0.745	3.752	0.809	0
4	-2.957	2.198	0.593	1	0	0	1	0	-8.106	0.685	0.593	0
5	-0.61	0.381	0.763	1	1	1	1	1	-0.635	-0.202	0.763	1
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
999998	-1.11	-0.487	-0.403	1	0	0	1	0	-1.042	-0.047	-0.403	0
999999	-0.218	0.053	-0.429	1	1	1	1	1	-0.36	0.18	-0.429	1
1000000	2.462	0.641	0.471	0	1	1	0	0	-2.705	2.313	0.471	0

in Section 3.5.1. Then, each of our conditions is checked in order to know whether the four inequality conditions are satisfied or not. It can be seen from Table 3.2 columns named C1, C2, C3, and C4 that "1" means the inequality condition is satisfied and zero indicates the inequality condition is not satisfied. Then we used the rejection sampling method in order to accept observations that satisfy the four conditions of (3.35)-(3.38) together. Afterwards, the partial auto-correlations π_s are calculated by using equations (3.91)-(3.93). Finally, from the accepted values we verified that 100% of the parameter values that satisfied the inequality conditions have met the π_s assumptions that its values are between [-1, 1]. This means that our conditions are sufficient 100% of the time.

3.6.5 Grouping conditions for the AR(4) model

The non-linear conditions of the AR(4) model are unknown as we had for the AR(3) model. Thus, we do not know which subsets we can can have for the AR(4) model. There are many possible subsets we can have by considering the linear conditions. Therefore, the simulation method can be used in order to discover the subsets of A, B, C and D. In order to attain sufficiency and necessity for stationarity conditions of the AR(4) model, four group conditions are built which are groups A, B, C and D. Table 3.2 will be used in order to discover the different subsets. In the AR(3) model, we know that the subsets have three inequalities and $\phi_i < 1$. Thus, there were three dimensions and there were three inequalities in the subgroups. In the AR(4) model, we think that there will be four inequalities. Therefore, we can consider that there are four inequalities in each subgroup. Consequently, Table 3.2 is used in order to find subsets in the AR(4) form.

Group A consists of equations (3.48), (3.50), (3.53) and (3.55), group B consists of equations (3.49), (3.51), (3.52) and (3.54), group C consists of equations (3.48), (3.49), (3.54) and (3.55) and group D consists of equations (3.49), (3.50), (3.51) and (3.52)

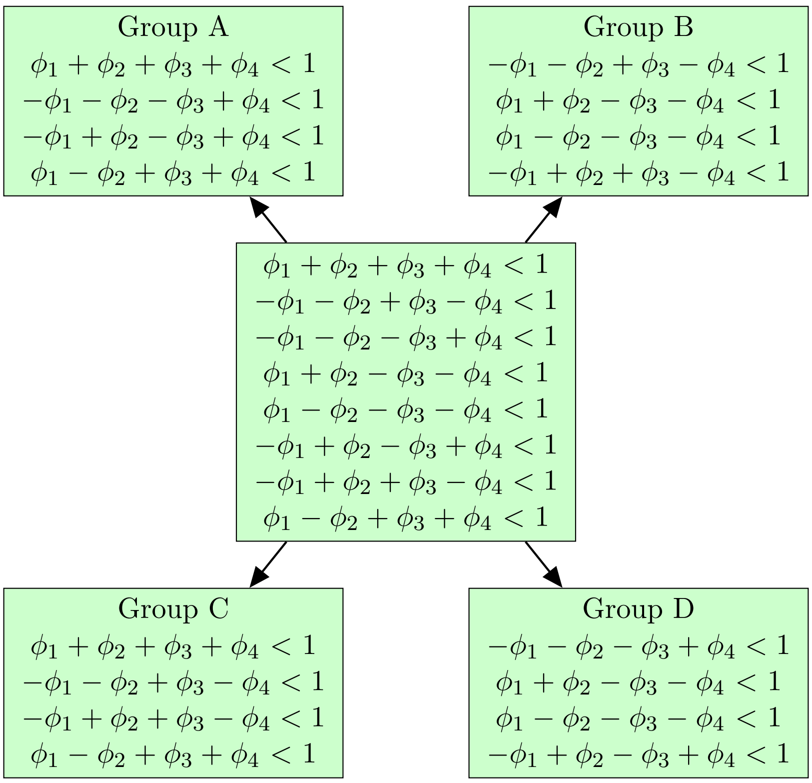


Figure 3.8: Group conditions for the AR(4) model.

3.6.6 Necessity of the AR(4) stationary conditions

In Section 3.5.1, we have proposed that the AR(4) model has eight inequality equations which restrict the values of the four parameters. To check necessity of the inequality

Table 3.3: Necessity for AR(4) stationary conditions

n	Simulate partial autocorrelation				Calculate parameters				Conditions for AR(4)								Group condtions				Satisfied
	π_1	π_2	π_3	π_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4	C1	C2	C3	C4	C5	C6	C7	C8	A1368	B2457	C1278	D2345	atleast one
1	-0.095	-0.179	-0.249	0.196	-0.108	-0.166	-0.218	0.196	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0.138	0.624	-0.864	0.418	0.951	0.389	-1.11	0.418	1	1	1	0	1	0	1	1	0	0	0	0	0
3	-0.784	-0.002	-0.695	-0.646	-1.236	-0.902	-1.204	-0.646	1	1	1	1	1	1	1	1	1	1	1	1	1
4	0.937	0.134	-0.544	0.192	0.989	0.465	-0.714	0.192	1	1	1	0	0	1	1	1	1	0	0	0	1
5	-0.345	-0.44	0.213	0.269	-0.461	-0.244	0.322	0.269	1	1	1	1	1	1	1	1	1	1	1	1	1
6	-0.142	0.174	0.47	0.699	-0.528	0.069	0.609	0.699	0	1	1	1	1	1	1	1	0	1	0	1	1
7	-0.106	0.456	0.575	0.905	-0.84	0.046	0.865	0.905	0	1	1	1	1	1	1	1	0	1	0	1	1
8	0.156	0.017	-0.498	0.615	0.468	0.036	-0.597	0.615	1	1	1	1	1	1	1	1	1	1	1	1	1
9	-0.318	-0.468	0.816	0.983	-0.886	-0.002	0.899	0.983	0	1	1	1	1	1	1	0	0	1	0	1	1
10	-0.523	-0.404	-0.257	-0.871	-1.062	-1.109	-0.987	-0.871	1	0	1	1	0	1	1	1	1	0	0	0	1
11	-0.515	-0.346	-0.051	0.017	-0.71	-0.375	-0.039	0.017	1	1	1	1	1	1	1	1	1	1	1	1	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
99993	-0.745	0.117	0.265	0.203	-0.743	0.232	0.405	0.203	1	1	1	1	1	1	1	1	1	1	1	1	1
99994	-0.573	-0.589	0.453	-0.599	-0.373	-0.282	0.067	-0.599	1	0	1	1	1	1	0	1	1	0	0	1	1
99995	-0.309	-0.476	0.128	-0.183	-0.372	-0.494	0.055	-0.183	1	0	1	1	1	1	1	1	1	0	1	1	1
99996	-0.632	0.645	0.169	-0.716	-0.212	1.171	-0.07	-0.716	1	1	1	1	1	1	0	1	1	0	0	1	1
99997	0.161	0.02	-0.108	-0.28	0.129	0.047	-0.063	-0.28	1	1	1	1	1	1	1	1	1	1	1	1	1
99998	0.063	0.65	-0.527	0.903	0.841	0.064	-0.856	0.903	0	1	1	1	1	0	1	1	0	1	0	0	1
99999	0.909	0.621	0.975	-0.878	0.594	0.537	0.745	-0.878	0	1	1	0	1	1	0	1	0	0	0	0	0
100000	0.968	0.161	0.036	-0.658	0.83	0.218	0.566	-0.658	1	1	1	0	0	1	1	1	1	0	0	0	1

conditions of the parameters, we take advantage of Barndorff-Nielsen and Schou (1973) which showed that the π's have the following relationship with the ϕ's as written below:

$$\phi_1 = \pi_1 - \pi_1 \pi_2 - \pi_2 \pi_3 - \pi_3 \pi_4 \quad (3.94)$$

$$\phi_2 = \pi_2 - \pi_1 \pi_3 - \pi_2 \pi_4 + \pi_1 \pi_2 \pi_3 + \pi_1 \pi_3 \pi_4 - \pi_1 \pi_2 \pi_3 \pi_4 \quad (3.95)$$

$$\phi_3 = \pi_3 - \pi_1 \pi_4 + \pi_1 \pi_2 \pi_4 + \pi_2 \pi_3 \pi_4 \quad (3.96)$$

$$\phi_4 = \pi_4 \quad (3.97)$$

C1 to C8 represent the stationarity conditions of equations (3.48) to (3.55), respectively. The π's are between [-1,1] in order to obtain parameters that satisfy the stationary conditions of the AR model. Therefore, 100000 observations of partial autocorrelations are simulated between (-1,1) and then the parameters are calculated using equations (3.94) to (3.97). After checking the stationary conditions for each of the achieved set of parameters, the percentage that, at least one of the group conditions is satisfied, is calculated. It can be noted that 85% of the simulated data satisfies at least one of the group conditions of Section 3.6.5. This can be another important aspect of the data. Moreover, regarding to all conditions which have been met at once, there is a very small percentage and it is only 8%. This means that if all conditions are together, they are very unlikely to be necessary, they are sufficient conditions but unlikely to be necessary conditions. Because of the fact that 8% is a small proportion, this proves that we need to create subsets. Also, there is one more thing that needs to be shown, and it is the result of Group A or Group B being met. Due to the fact that all eight basic inequality equations can be seen in both groups, we thought it might be important to state the result. Therefore, 77% of the simulated data met either the Group A or B condition but not both together.

3.6.7 Sufficiency for AR(4) stationary conditions

The process we follow is the same as that for the AR(3) with some small changes. The following steps, hence, are used that involve some expressions that come from rather complicated equations, see Section 3.5.1. However, here we only show the latest version of our trial. We simulate a number of values for all ϕ's between (a, b); we assume that the range of a and b is not less than (-4) and (4). This assumption is based on the information on the stationary regions for the AR(2) and AR(3) models available from Figures (3.2) and (3.5). We simulated a series of sets of ϕ's by taking the stationary region for the AR(4) model into account. Then, the simulated parameters are checked via rejection sampling in order to classify the parameters that satisfy the stationary linear conditions for the AR(4) model and those that do not satisfy them. The π's are computed for those parameters that satisfy the stationary conditions for the AR(4) model according to

Table 3.4: Sufficiency for AR(4) stationary conditions.

n	Simulate parameters				Check stationary conditions								Satisfied	Calculate				Satisfied
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	C1	C2	C3	C4	C5	C6	C7	C8	all C	π_1	π_2	π_3	π_4	π C
1	1.887	0.999	0.923	-0.797	0	1	1	0	1	1	1	0	0	0.774	-2.495	-1.594	-0.797	0
2	-0.708	3.088	-1.494	-0.684	1	1	1	0	1	0	0	1	0	0.872	0.062	-1.898	-0.684	0
3	0.002	-0.104	0.195	-0.669	1	1	1	1	1	1	1	1	1	-0.249	-0.144	0.351	-0.669	1
4	2.422	-3.138	-3.042	0.433	1	1	0	0	0	1	1	0	0	3.874	-0.100	-2.453	0.433	0
5	0.362	-0.307	0.080	0.137	1	1	1	1	1	1	1	1	1	0.258	-0.306	0.132	0.137	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
99998	-2.257	0.019	-2.803	0.970	1	1	0	1	1	0	1	1	0	0.010	-0.246	-85.651	0.970	0
99999	0.348	-0.039	0.089	0.242	1	1	1	1	1	1	1	1	1	0.404	0.020	0.183	0.242	1
100000	-3.386	-1.165	2.661	0.532	1	0	0	1	1	1	0	1	0	-1.072	-0.046	1.198	0.532	0

the equations below

$$
\begin{aligned}
\pi_1 &= -\frac{\phi_1\phi_2\phi_4 - \phi_3{\phi_4}^2 - \phi_1\phi_4 + \phi_2\phi_3 + \phi_3\phi_4 + \phi_1}{{\phi_1}^2\phi_4 + \phi_1\phi_3\phi_4 - \phi_2{\phi_4}^2 - {\phi_4}^3 + \phi_1\phi_3 + {\phi_3}^2 + {\phi_4}^2 + \phi_2 + \phi_4 - 1} \\
\pi_2 &= -\frac{\phi_1\phi_3{\phi_4}^2 - \phi_2{\phi_4}^3 + {\phi_1}^2\phi_4 - \phi_2{\phi_4}^2 + {\phi_3}^2\phi_4 + \phi_1\phi_3 + \phi_2\phi_4 + \phi_2}{{\phi_1}^2{\phi_4}^2 - {\phi_4}^4 + 2\phi_1\phi_3\phi_4 + {\phi_3}^2 + 2{\phi_4}^2 - 1} \\
\pi_3 &= \frac{-\phi_1\phi_4 - \phi_3}{{\phi_4}^2 - 1} \\
\pi_4 &= \phi_4
\end{aligned}
$$

Since in Barndorff-Nielsen and Schou (1973) ϕ 's are computed from π 's, we here convert the situation and as shown we derive π 's from ϕ 's. The purpose of doing this is to know whether any values of π 's are outside the range of [-1, 1], hence indicating non-stationarity. Table 3.4 displays that 100% of the values of the simulated ϕ satisfy the stationary conditions, resulting in π 's in [-1,1]. We are using Table 3.4 in order to confirm the results we already presented mathematically. We know from this table that since these stationarity conditions are sufficient, the PACF gives us values between (-1, 1). However, it is our interest to see what are the range of values of the PACF.

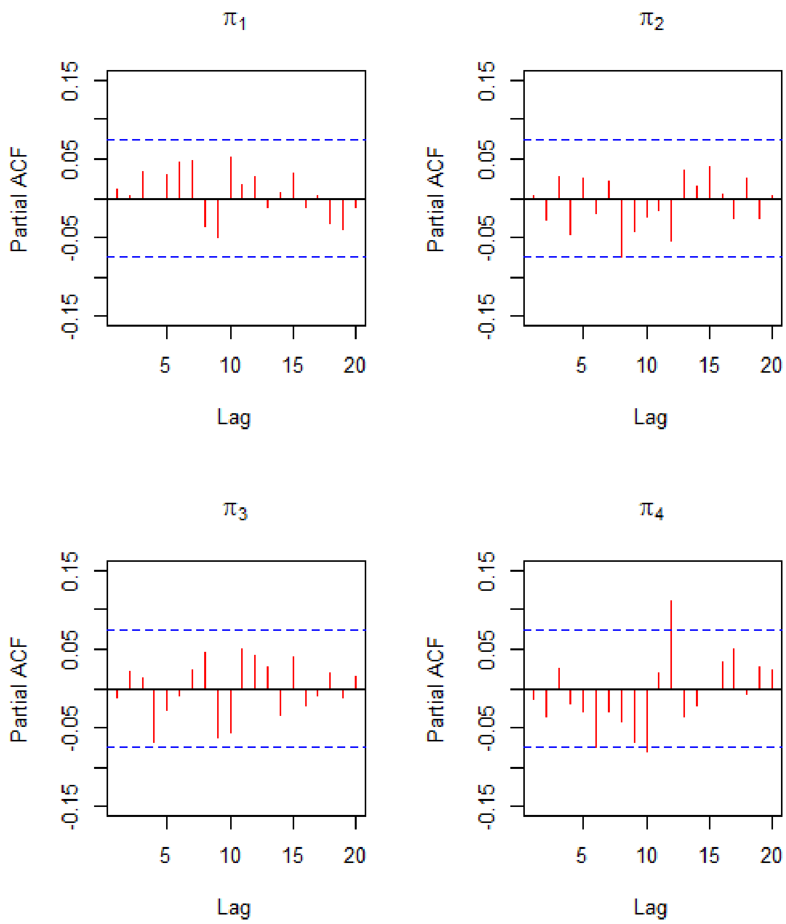


Figure 3.9: Partial ACF for partial auto-correlations of the AR(4) model.

Figure 3.9 illustrates the PACF ranges of the four π's. The ranges of all π's are between (-0.1, 0.1). This means that there is no significant correlation. Additionally, we want to look whether the π's are independent to each other by visualizing graphs. Figure 3.10 indicates that there are no relationships between the partial autocorrelations. Thus, we conclude that π's are independent to each other.

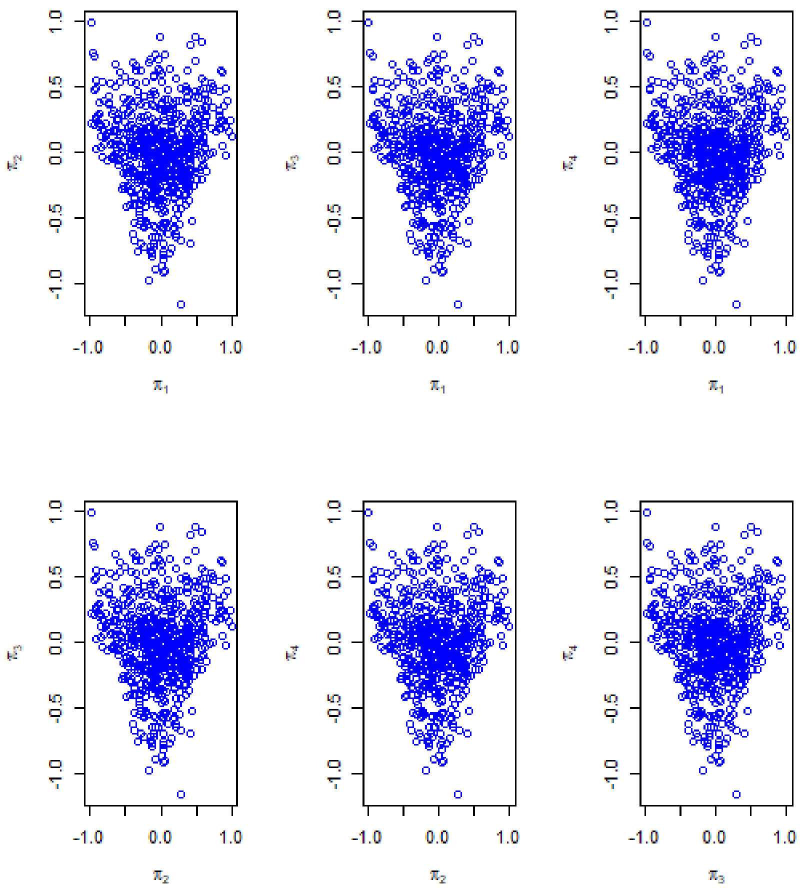


Figure 3.10: shows the relationship between partial autocorrelation function π 's

Chapter 4

MCMC methods for autoregressive models

4.1 Introduction

As stated earlier a major focus of our project is on the estimation of autoregressive parameters through MCMC methods. As is established in the literature (Barndorff-Nielsen and Schou (1973) and Huerta and West (1999)) a prior distribution on parameters or transformations of them, must respect the requirement of stationarity which imposes conditions on those parameters. Thus, in this chapter, we define a new prior distribution placed directly on the AR parameters. We go on to propose suitable MCMC schemes for estimation. This is achieved through information obtained on the stationary conditions for the AR(p) model, for relatively low lag order $(p \leq 6)$.

We propose a new prior distribution placed directly on the AR parameters of the AR(p) model. This is motivated by priors proposed for AR(1), AR(2), ... , AR(6), which take advantage of the range of the AR parameters. We then develop a Metropolis within Gibbs algorithm for estimation. This scheme is illustrated using simulated data for the AR(2), AR(3) and AR(4) models and then we extend to models with higher lag order. MCMC has been applied on a set of simulated data; the data have been simulated on the basis of an AR model.

4.2 Using the Gibbs sampler for AR(1)

Assume n observations are available, say $y_1, y_2, \ldots, y_n$. The aim is to estimate the unknown parameters of ϕ and σ^2. We use the AR(1) model $y_t = \phi y_{t-1} + \varepsilon_t$ where ε_t is white noise and $\varepsilon_t \sim N(0, \sigma^2)$. To compute with the Gibbs sampler, we need to derive the conditional posterior distribution of the parameters. We assume that the prior distribution of ϕ is a uniform distribution, and the prior of $\frac{1}{\sigma^2}$ (precision) is a gamma distribution, i.e.,

$$\phi \sim U(-1, 1),$$

$$\sigma^2 \sim IG(a, b) \quad \text{or} \quad \frac{1}{\sigma^2} \sim G(a, b).$$

The aim of employing the Gibbs sampler is to discover the posterior distribution of the unknown parameters (ϕ, σ^2). This requires taking samples from the two distributions below:

$$p(\phi \mid y, \sigma^2) \text{ and } p(\sigma^2 \mid y, \phi)$$

From Bayes' theorem we have

$$\begin{aligned} p(\phi \mid y, \sigma^2) &\propto p(y \mid \phi, \sigma^2) p(\phi) \\ &\propto \prod_{t=1}^{n} e^{-\frac{1}{2\sigma^2}(y_t - \phi y_{t-1})^2} p(\phi) \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi y_{t-1})^2} I_{[-1,1]}, \end{aligned} \tag{4.1}$$

where $I_{[-1,1]}$ is the indicator function on [-1,1]. We can extend the part of $\sum_{t=1}^{n}(y_t - \phi y_{t-1})^2$ from equation (4.1) in order to complete the square and obtain a truncated normal posterior. Therefore

$$\begin{aligned}
\sum_{t=1}^{n}(y_t-\phi y_{t-1})^2 &= \sum_{t=1}^{n}(y_t^2+\phi^2 y_{t-1}^2-2y_t\phi y_{t-1}) = \sum_{t=1}^{n} y_t^2+\phi^2\sum_{t=1}^{n} y_{t-1}^2-2\phi\sum_{t=1}^{n} y_t y_{t-1}\\
&= \sum_{t=1}^{n} y_t^2+\phi^2\sum_{t=1}^{n} y_{t-1}^2-2\phi\sum_{t=1}^{n} y_t y_{t-1}+\frac{(\sum_{t=1}^{n} y_t y_{t-1})^2}{\sum_{t=1}^{n} y_{t-1}^2}-\frac{(\sum_{t=1}^{n} y_t y_{t-1})^2}{\sum_{t=1}^{n} y_{t-1}^2}\\
&= \sum_{t=1}^{n} y_{t-1}^2\left(\phi^2-2\phi\frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}+\left(\frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}\right)^2\right)+\sum_{t=1}^{n} y_t^2-\frac{(\sum y_t y_{t-1})^2}{\sum_{t=1}^{n} y_{t-1}^2}\\
&= \frac{\left(\phi-\frac{\sum_{t=1}^{n} y_t y_{t-1}}{\sum_{t=1}^{n} y_{t-1}^2}\right)^2}{\frac{1}{\sum y_{t-1}^2}}+\sum_{t=1}^{n} y_t^2-\frac{(\sum_{t=1}^{n} y_t y_{t-1})^2}{\sum_{t=1}^{n} y_{t-1}^2} \qquad (4.2)
\end{aligned}$$

Now by using (4.1) and (4.2) we see that

$$p(\phi \mid \sigma^2, y) \propto \exp\left\{-\frac{\sum_{t=1}^{n} y_{t-1}^2}{2\sigma^2}\left(\phi-\frac{\sum_{t=1}^{n} y_t y_{t-1}}{\sum_{t=1}^{n} y_{t-1}^2}\right)^2\right\} I_{[-1,1]}, \qquad (4.3)$$

i.e., $\phi \mid \sigma^2, y \sim N\left(\frac{\sum_{t=1}^{n} y_t y_{t-1}}{\sum_{t=1}^{n} y_{t-1}^2}, \frac{\sigma^2}{\sum_{t=1}^{n} y_{t-1}^2}\right) I_{[-1,1]}$ is a truncated normal distribution. Likewise, the posterior distribution of $1/\sigma^2$ is

$$\begin{aligned}
p(1/\sigma^2 \mid y,\phi) &= p(y \mid 1/\sigma^2,\phi)p(1/\sigma^2 \mid \phi)\\
&= \prod_{t=1}^{n} p(y_t \mid y_{t-1},\phi,1/\sigma^2)p(1/\sigma^2 \mid \phi)\\
&\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t-\phi y_{t-1})^2}\left(\frac{1}{\sigma^2}\right)^{a-1} e^{-\frac{b}{\sigma^2}}\\
&\propto \left(\frac{1}{\sigma^2}\right)^{a+\frac{n}{2}-1} e^{-\left(b+\frac{1}{2}\sum_{t=1}^{n}(y_t-\phi y_{t-1})^2\right)\frac{1}{\sigma^2}} \qquad (4.4)
\end{aligned}$$

Comparing equation (4.4) with the gamma distribution we see that

$$\frac{1}{\sigma^2}|y,\phi \sim G\left(a+\frac{n}{2}, b+\frac{1}{2}\sum_{t=1}^{n}(y_t-\phi y_{t-1})^2\right). \qquad (4.5)$$

We can now apply the above Gibbs sampler to simulated data. The AR coefficient ϕ is simulated by a truncated normal distribution, which ensures the sampled $\phi^{(i)}$ are in the stationarity region [-1,1], and the variance σ^2 of the white noise process is simulated from an inverse gamma distribution. First, we simulate $n = 500$ observations from the AR(1) model with $\phi = 0.3$ and $\sigma^2 = 1$, then we estimate the parameters of ϕ and σ^2 in terms of the unknown parameters of the conditional distributions. Second, we use these observations and we employ the Gibbs sampler to estimate ϕ and σ^2 when the number of iterations is $k = 10000$. Following this, by using the mode of the simulated ϕ and σ^2, we estimate the parameters ϕ and σ^2, for iteration $k = 10000$. Figure 4.1 shows that ϕ has converged to 0.3 and σ^2 has converged to 1. To assess the adequacy of the MCMC estimates a Monte Carlo experiment has been used, i.e., the above procedure of the Gibbs sampler is repeated N=100 times. Next, the mean of the N=100 MCMC estimators has been taken for both parameters ϕ and σ^2. As a result, the convergence of the parameters can be observed as the means were calculated to be $\phi = 0.29338$ and $\sigma^2 = 1.06$. We repeated this process for different ϕ and number of observations under $\sigma^2 = 1$. Regarding the prior distribution of σ^2, as mentioned before the prior of $\frac{1}{\sigma^2}$ (precision) is a gamma distribution $\frac{1}{\sigma^2} \sim G(a,b)$. The prior of $\frac{1}{\sigma^2}$ is weakly informative, since the variance of σ^2 is large. Therefore, we set the parameters of the prior of the gamma distribution to a=3 and b=10; see Table 4.1.

Table 4.1: Illustration of different results obtained from simulation study for ϕ and number of observations when assuming $a = 3$, b= 10 and $\sigma^2 = 1$.

The estimation of parameters of AR(1) by MCMC			
Simulation		Estimation by MCMC	
n	ϕ	ϕ	σ^2
500	0.3	0.29439(0.05297)	1.00253(0.07694)
	0.5	0.49954(0.04220)	1.07256(0.08681)
	0.8	0.79536(0.03625)	0.98625(0.12002)
1000	0.3	0.29956(0.02792)	0.98933(0.05563)
	0.5	0.49915(0.02449)	0.83844(0.06108)
	0.8	0.80027(0.01909)	1.05825(0.08836)

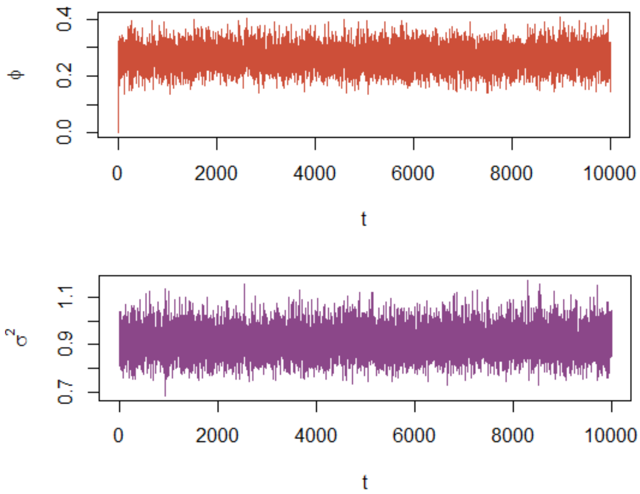


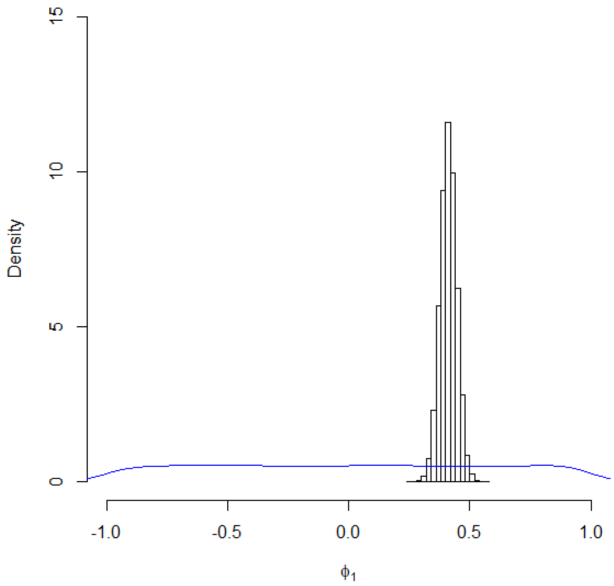
Figure 4.1: Trace plots of ϕ and σ^2 of the Gibbs sampler for an AR(1) model (k=10000, $\phi = 0.3$ and $\sigma^2 = 1$).

In Table 4.1, the a and b parameters chosen above have clearly worked well, because we know that the true value of σ^2 is 1. But, in reality we will not know σ^2. Hence a weakly informative prior on σ^2 will be explored (Ando, 2010). So we consider σ^2, if $\sigma^2 > 1$ (unknown) we assume that σ^2 is large therefore we try to take some different σ^2 which is greater than one. Thus, the estimation of ϕ and σ^2 converges adequately to each ϕ and σ^2 assumed for simulating the AR(1) model. We know that if $\sigma^2 > 1$ then $\frac{1}{\sigma^2} < 1$. Thus, we assumed that $a = 3$ and $b = 10$ in order to always obtain a result for which $E(\frac{1}{\sigma^2}) < 1$.

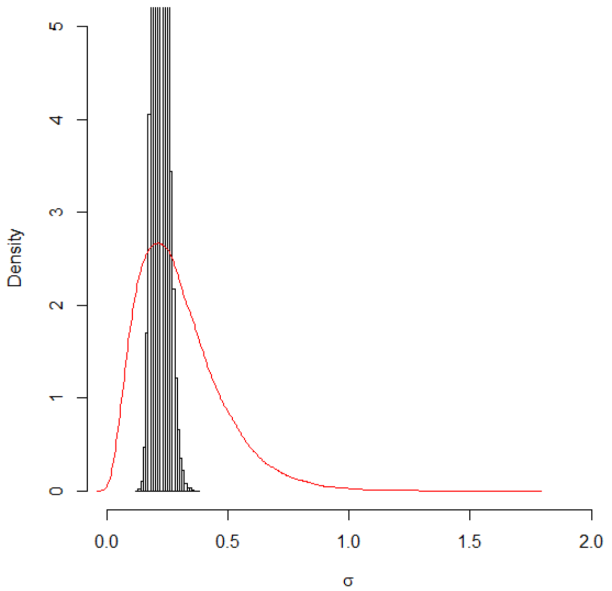
The Gibbs sampler is applied in order to estimate the parameters ϕ and σ^2 (see Table 4.2).

4.3 Sampling the parameters of the AR(1) to AR(4) models

In order to assess the performance of the MCMC we shall simulate data from an AR(p) model. We need to have information about the parameters in order to simulate data.



(a) ϕ



(b) σ^2

Figure 4.2: Illustration of the prior and posterior densities of the parameters ϕ and σ^2 for the AR(1) model.

Table 4.2: Illustration of different results obtained from simulation study for ϕ, σ^2 and number of observations through mode.

The estimation of parameters of AR(1) by MCMC				
Simulation			Estimation by MCMC	
n	ϕ	σ^2	ϕ	σ^2
500	0.3	9	0.30462 (0.04241)	10.9244 (0.01714)
	0.3	16	0.29776 (0.04089)	19.1466 (0.01319)
	0.3	100	0.30231 (0.03675)	118.7674 (0.00541)
500	0.5	9	0.49936 (0.03542)	9.126308 (0.02035)
	0.5	16	0.49861 (0.03408)	15.23834 (0.01553)
	0.5	100	0.50226 (0.03325)	101.4965 (0.00512)
500	0.8	9	0.79439 (0.02782)	10.86549 (0.02828)
	0.8	16	0.79861 (0.02294)	13.35446 (0.02166)
	0.8	100	0.79911 (0.02439)	108.3137 (0.00839)
1000	0.3	9	0.29782 (0.02871)	10.18267 (0.01233)
	0.3	16	0.29801 (0.02833)	18.96022 (0.00961)
	0.3	100	0.29921 (0.02195)	82.87487 (0.00454)
1000	0.5	9	0.50365 (0.02648)	10.27576 (0.01331)
	0.5	16	0.50408 (0.02445)	18.30929 (0.01098)
	0.5	100	0.49576 (0.02334)	104.8193 (0.00406)
1000	0.8	9	0.80022 (0.01704)	8.664881 (0.01888)
	0.8	16	0.79801 (0.01839)	17.74708 (0.01331)
	0.8	100	0.80157 (0.01686)	115.5993 (0.00559)

There are several ways of doing this, but here we use correspondence of the partial autocorrelations (π_i) with the AR parameters. In particular, as these autocorrelations lie in $[-1, 1]$ we simulate values π_i between (-1, 1).

After simulating π_i, the values of ϕ_i in the AR(p) model can be calculated based on the corresponding relationship between π_i and ϕ_i as discussed in Section 3.6 based on equations (3.73)-(3.75). The AR(2) model is used as the example for clarifying how n observations of the AR(2) model are simulated in order to apply MCMC to estimate parameters of the AR(2) model. It is well-known that the corresponding relationship between π_i and ϕ_i for the AR(2) model stated by Barndorff-Nielsen and Schou (1973) as follows

$$\phi_1 = \pi_1(1 - \pi_2) \tag{4.6}$$

$$\phi_2 = \pi_2 \tag{4.7}$$

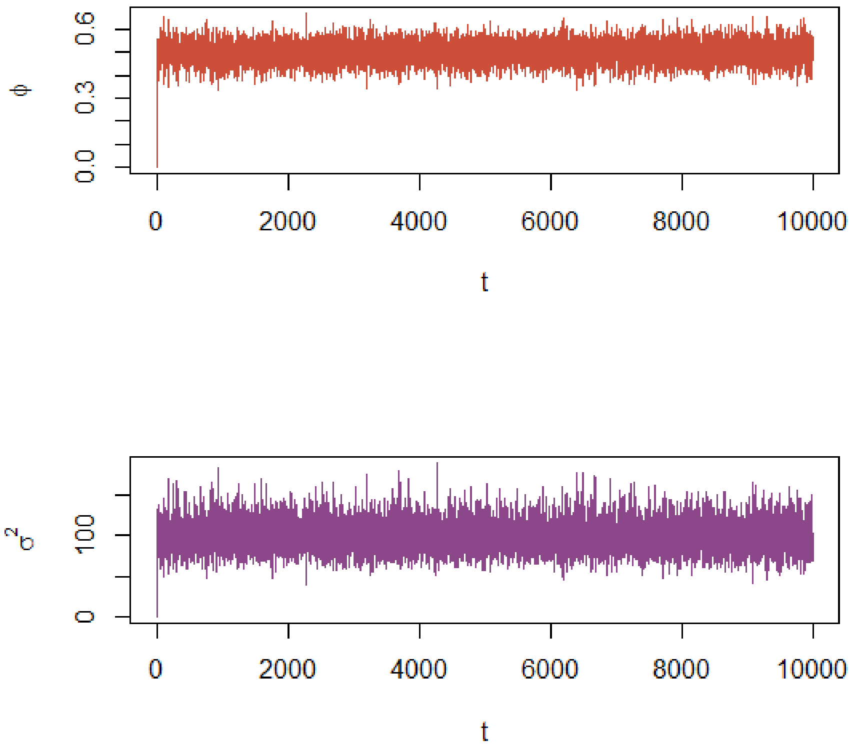


Figure 4.3: Trace plots of ϕ and σ^2 from one iteration of the Gibbs sampler for an autoregressive model AR(1) with K=10000, $\phi = 0.5$ and $\sigma^2 = 100$.

The purpose behind this is to simulate n observations by using ϕ_1 and ϕ_2, which are obtained from equations (4.6) and (4.7).

4.4 Prior distribution of the AR(2) model

To determine the prior distribution of the AR(2) model, based on the available information, we consider the parameters of the AR(2) model in more detail. The stationarity conditions in (3.18) and (3.19) restrict the range of parameters ϕ_1 and ϕ_2 as follows:

$$\phi_1 + \phi_2 < 1 \Rightarrow \phi_1 < 1 - \phi_2 \tag{4.8}$$

$$\phi_2 - \phi_1 < 1 \Rightarrow \phi_1 > \phi_2 - 1 \tag{4.9}$$

$$|\phi_2| < 1 \Rightarrow -1 < \phi_2 < 1 \tag{4.10}$$

Based on the information in equations (4.8)-(4.10), it can be seen that ϕ_1 and ϕ_2 lie in the following ranges $\phi_2 - 1 < \phi_1 < 1 - \phi_2$ and $-1 < \phi_2 < 1$. We want to place prior distributions directly on the parameters. Thus, we choose uniform distributions for the priors because we want to have uninformative prior distributions. We propose that $\phi_1|\phi_2$ has a uniform distribution and ϕ_2 has a uniform distribution, i.e., $\phi_1 \mid \phi_2 \sim U(\phi_2 - 1, 1 - \phi_2)$ and $\phi_2 \sim U(-1, 1)$.

From this we can propose the joint prior distribution of ϕ_1 and ϕ_2 :

$$p(\phi_1, \phi_2) = p(\phi_1 \mid \phi_2).p(\phi_2) = \begin{cases} \frac{1}{4(1-\phi_2)} & \text{if } \phi_2 - 1 < \phi_1 < 1 - \phi_2 \text{ and } -1 < \phi_2 < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

There might be some interest in the marginal prior distributions of ϕ_1 and ϕ_2 :

$$p(\phi_1) = \int_{\phi_2} p(\phi_1, \phi_2) d\phi_2 = \int_{-1}^{1+|\phi_1|} \frac{1}{4(1-\phi_2)} d\phi_2 = \frac{\log(2) - \log(|\phi_1|)}{4} \tag{4.12}$$

By using equations (4.11) and (4.12), $p(\phi_2 \mid \phi_1)$ is obtained as follows:

$$p(\phi_2 \mid \phi_1) = \frac{p(\phi_1, \phi_2)}{p(\phi_1)} = \frac{1}{(1-\phi_2)(\log(2) - \log(|\phi_1|))}. \tag{4.13}$$

In order to estimate parameters through MCMC, the conditional posterior distributions need to be derived. This is discussed in detail in the next section.

4.5 Mean and variance of the prior distribution of parameters for the AR(2) model

This section adopts the joint prior distribution (4.11) proposed in the previous section. It then derives the mean and the variance of ϕ_1 and ϕ_2. This might be useful when considering what are the effects placed on ϕ_1 and ϕ_2 when we set the prior to be (4.11). According to equation (4.12), we get

$$\begin{aligned} E(\phi_1) &= \int_{-2}^{2} \phi_1 p(\phi_1) d\phi_1 = \int_{-2}^{2} \phi_1 \frac{\log(2) - \log|\phi_1|}{4} d\phi_1 \\ &= \frac{\log(2)}{4} \int_{-2}^{2} \phi_1 d\phi_1 - \frac{1}{4} \int_{-2}^{2} \phi_1 \log|\phi_1| d\phi_1 = \frac{\log(2)}{8}\phi_1^2 - \frac{1}{4}\phi_1 \\ &= \log|\phi_1| + \frac{1}{8}\phi_1^2 \Big|_{-2}^{2} = 0 \end{aligned}$$

Furthermore, the variance of ϕ_1 is

$$\begin{aligned} \mathrm{var}(\phi_1) &= E(\phi_1^2) = \int_{-2}^{2} \phi_1^2 \frac{\log(2) - \log|\phi_1|}{4} d\phi_1 = -\frac{\phi_1^3 \left(3\log\left(|\phi_1|\right) - 3\log\left(2\right) - 1\right)}{36} \Big|_{-2}^{2} \\ &= \frac{4}{9} \approx 0.444 \end{aligned}$$

From the joint prior distribution (4.11) it is easy to see that the marginal prior of ϕ_2 is the uniform $U(-1, 1)$ distribution with mean 0 and variance $\frac{1}{3}$.

Figure 4.4 shows the PDF of the prior distribution of ϕ_1. It can be noticed that it a high mode plot and the range of ϕ_1 is between (-2, 2). It clear that the mean is zero in which we have already proved mathematically.

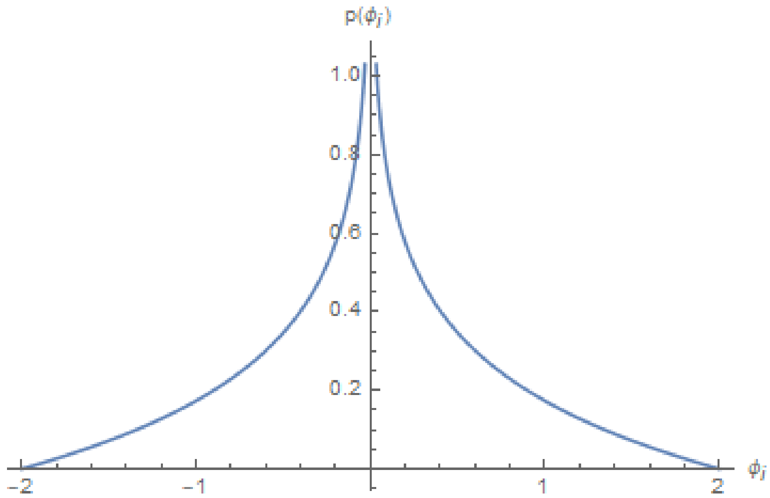


Figure 4.4: The marginal prior distribution of ϕ_1 .

4.6 Posterior distribution for the AR(2) model ϕ_1

In order to apply Bayes' theorem, we consider n observations, $y_1, y_2, \ldots, y_n$, from the AR(2) model:

$$y_t = \phi_1 y_{t-1} - \phi_2 y_{t-2} + \varepsilon_{t-2},$$

where ε_t is white noise and $\varepsilon_t \sim N(0, \sigma^2)$. The unknown parameters here are ϕ_1, ϕ_2 and σ^2 . Thus, the posterior distribution of $\phi_1 | \phi_2$ is as follows:

$$\begin{aligned} p(\phi_1 \mid y, \sigma^2, \phi_2) &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2} . p(\phi_1 \mid \phi_2) \\ &= e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2} . \frac{1}{2(1-\phi_2)} \end{aligned}$$

In order to expand the part of $\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2$, initially we set $z_t = y_t - \phi_2 y_{t-2}$,

$$p(\phi_1 \mid y, \sigma^2, \phi_2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum(z_t-\phi_1 y_{t-1})^2}.I_{[a]}, \tag{4.14}$$

where

$$I_{[a]} = \begin{cases} 1 & a \in A \\ 0 & a \notin A \quad \text{where A=}\{\ (\phi_1, \phi_2) : \phi_2 - 1 < \phi_1 < 1 - \phi_2 \text{ and } -1 < \phi_2 < 1\}. \end{cases}$$

Now, we can expand the part of $\sum(z_t - \phi_1 y_{t-1})^2$ from equation (4.14) by following the same steps of equation (4.2). Then the following equation can be obtained

$$p(\phi_1 \mid y, \sigma^2, \phi_2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum y_{t-1}^2\left[\phi_1 - \frac{\sum z_t y_{t-1}}{\sum y_{t-1}^2}\right]^2} I_{[a]}$$

Now, it can be seen that the posterior distribution for $\phi_1 \mid \phi_2$ is truncated normally distributed with specific mean and variance as follows

$$\phi_1 \mid \phi_2 \sim N_{[\phi_2-1, 1-\phi_2]}\left(\frac{\sum y_t y_{t-1} - \phi_2 \sum y_{t-1} y_{t-2}}{\sum y_{t-1}^2}, \frac{\sigma^2}{\sum y_{t-1}^2}\right) \tag{4.15}$$

However, the posterior distribution of $\phi_2 \mid \phi_1, y, \sigma^2$ is:

$$\begin{aligned} p(\phi_2 \mid y, \sigma^2, \phi_1) &\propto p(y \mid \phi_1, \phi_2, \sigma^2).p(\phi_2 \mid \phi_1) \\ &= e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2})^2}.\frac{1}{(1-\phi_2)(\log(2) - \log(|\phi_1|))} \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(z_t-\phi_2 y_{t-2})^2}\frac{1}{(1-\phi_2)} I_{[a]} \end{aligned} \tag{4.16}$$

where

$$I_{[a]} = \begin{cases} 1 & a \in A \\ 0 & a \notin A \quad \text{where } A\text{=}\{\ (\phi_1, \phi_2) : \phi_2 - 1 < \phi_1 < 1 - \phi_2 \text{ and } -1 < \phi_2 < 1\}. \end{cases}$$

By following the same steps of equation 4.5, the posterior distribution of $\frac{1}{\sigma^2}$ is:

$$\frac{1}{\sigma^2}|y, \phi_1, \phi_2 \sim G\left(a + \frac{n}{2}, b + \frac{1}{2}\sum_{t=1}^{n}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2\right). \quad (4.17)$$

It can be noted that the posterior distribution of $\phi_1|\phi_2$ has a known distribution which is a truncated normal distribution. Therefore, a Gibbs sampler step can be used to simulate ϕ_1. However, the posterior distribution of $\phi_2 \mid \phi_1, y, \sigma^2$ has an unknown distribution. Thus, a Metropolis step is used in order to simulate ϕ_2. In the next section we propose a Metropolis step in order to sample from, within a Gibbs sampler, $\phi_1, \phi_2 \mid y, \sigma^2$.

4.7 MCMC application for the AR(2) model

In order to perform MCMC for estimating the parameters of the AR(2) model, we first simulate data from that model with different values of ϕ_1 and ϕ_2 where $\sigma^2 = 1$. In order to apply Gibbs for ϕ_1, the full conditional posterior distribution is required. the distribution of $\phi_1 \mid y, \phi_2$ is truncated normal distribution with range $\phi_2 - 1, 1 - \phi_2$, mean $\frac{\sum y_t y_{t-1} - \phi_2 \sum y_{t-1} y_{t-2}}{\sum y_{t-1}^2}$ and variance $\frac{\sigma^2}{\sum y_{t-1}^2}$ see (4.15).

However, for ϕ_2 we are not able to use the Gibbs sampling because the distribution of $\phi_2 \mid \phi_1, y$ is not known owing to the last part shown in equation (4.16). One idea is to abandon Gibbs sampling altogether and to use a Metropolis step for both ϕ_1 and ϕ_2. This is a poor choice because it is hard to find suitable proposals; as we will see later the random walk proposal does not do a good job. A more appealing proposal is to adopt a Metropolis within Gibbs approach whereby ϕ_1 is updated by Gibbs while ϕ_2 is updated by a Metropolis step. The details of this approach are discussed in this and the next section.

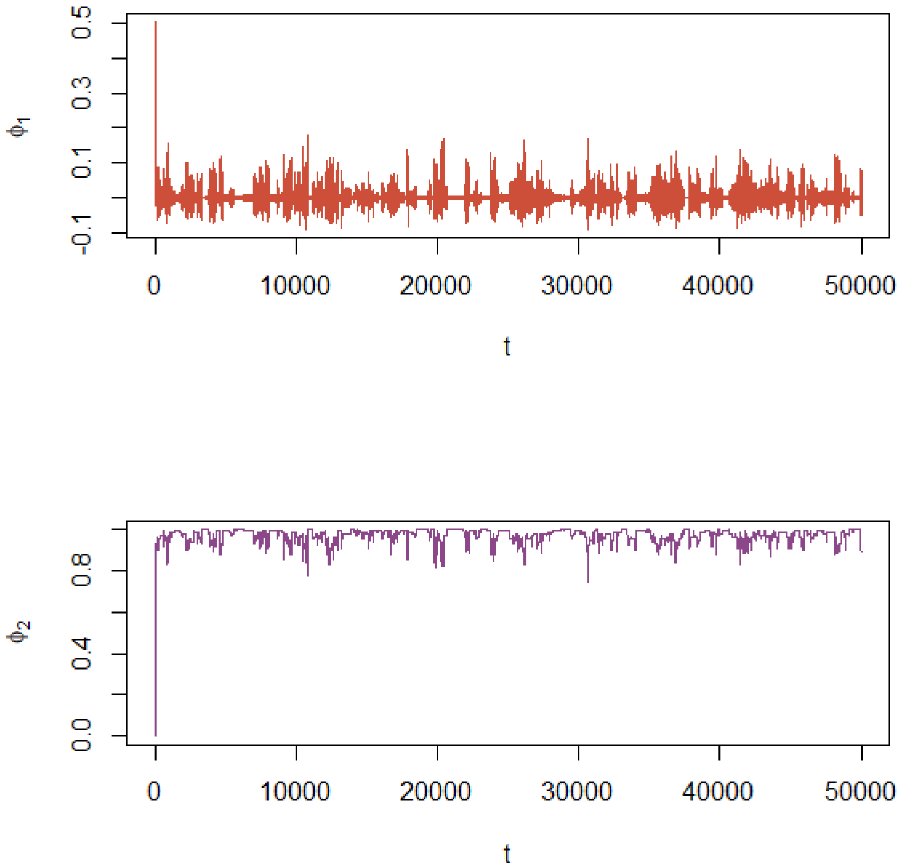


Figure 4.5: Trace plots of the estimated parameters ϕ_1 and ϕ_2 via MCMC of the AR(2) model with K=50000, $\phi_1 = 0.7$ and $\phi_2 = -0.7$.

Finding a good proposal can be hard and there is no general rule to obtain one. We, hence, first start with a random walk where ϕ_2 is updated at iteration i using

$$\phi_2^{(i)} = \phi_2^{(i-1)} + \varepsilon_t, \quad \text{with} \quad \varepsilon_t \sim N(0, V^2)$$

and

$$V = \sqrt{\frac{1-s}{s}} \qquad \text{for some} \quad 0 < s < 1.$$

A rejection sampling step is performed for the conditions of $|\phi_2^{(i)}| \geq 1$. Referencing to our result gained from practical part there are many trials as we used in the basic step simulation. We did simulation several times with various values of ϕ_1 and ϕ_2. According to Table 4.3 the simulated data was based on $\phi_1 = 0.7$ and $\phi_2 = -0.7$ and the outcome is far from the true parameter for ϕ_1 with mean=0.0523 and sd=0.0214, and mean=0.9681 and sd=0.0281 for ϕ_2 when we had s=0.1 and σ=0.1; this result has 5.3% acceptance rate. This result can be seen clearly from Figure 4.5.

Moreover, despite having poor results from the first trial, we had the second simulation trial $\phi_1 = 0.4$ and $\phi_2 = 0.5$. Referring to Table 4.3 we can show that the result is poor because it is clear that both the Gibbs sampling and Metropolis cannot estimate the parameters ϕ_1 and ϕ_2 properly since the obtained results are far from the true values. The estimated mean for $\phi_1 = 1.0941$ with its sd=0.0955, and the estimated mean for $\phi_2 = -0.3477$ with its sd=0.1040.

Finally we tested the method for a third set of simulated data with $\phi_1 = 0.2$ and $\phi_2 = 0.2$. Here, we still faced the same issue as we had for the previous simulated data. We repeated the same procedures as the other trials and a set of values for s and σ were used which can be seen in Table 4.3.

Additionally, we make some changes to s and σ in order to see different outputs. Several different values of s and σ were used for the simulation data and the random walk proposal. From the same table (Table 4.3), many trials were conducted and their acceptance rates were recorded.

The conclusion is that the random walk proposal does not seem to work well, leading to poor estimation. Thus, in the next section we develop a new proposal for the Metropolis step.

Table 4.3: Shows the parameter estimation of the AR(2) model via MCMC application with K=50000, b=1000 and n =150.

s	σ	$\phi_1(sd)$	$\phi_2(sd)$	Acc .Rate
True parameters		$\phi_1 = 0.7$	$\phi_2 = -0.7$	Trial 1
0.1	0.1	0.0523 (0.0214)	0.9681 (0.0281)	5.3%
	1	0.3251 (0.1977)	0.2371 (0.4473)	42.2%
	10	0.4383 (0.2163)	0.0260 (0.4937)	59.1%
0.5	0.1	0.0641 (0.0239)	0.9640 (0.0321)	7.2%
	1	0.3233 (0.1971)	0.2419 (0.4466)	42.2%
	10	0.4298 (0.2119)	0.0731 (0.4823)	62.2%
0.9	0.1	0.0073 (0.0242)	0.9629 (0.0330)	11.5%
	1	0.2763 (0.1877)	0.3524 (0.4233)	51.2%
	10	0.3510 (0.2176)	0.1769 (0.4956)	65.3%
True parameters		$\phi_1 = 0.4$	$\phi_2 = 0.5$	Trial 2
0.1	0.1	1.0941 (0.0955)	-0.3477 (0.1040)	22.1%
	1	1.3155 (0.1968)	-0.6193 (0.2321)	45.3%
	10	1.3375 (0.1893)	-0.6491 (0.2270)	59.3%
0.5	0.1	1.1027 (0.1004)	-0.3573 (0.1111)	22.85%
	1	1.3481 (0.1976)	-0.6118 (0.2374)	47.2%
	10	1.3312 (0.1929)	-0.6395 (0.2366)	61.2%
0.9	0.1	1.1125 (0.0981)	-0.3687 (0.1076)	30.69%
	1	1.2777 (0.2128)	- 0.5732 (0.2567)	54.2%
	10	1.2944 (0.2138)	- 0.5942 (0.2575)	61.3%
True parameters		$\phi_1 = 0.2$	$\phi_2 = 0.2$	Trial 3
0.1	0.1	0.0220 (0.0621)	0.8423 (0.1742)	15.33%
	1	0.1001 (0.1112)	0.3791 (0.5299)	54.2%
	10	0.1010 (0.1113)	0.3726 (0.5376)	55.1%
0.5	0.1	0.0244 (0.0637)	0.8279 (0.1909)	17.3%
	1	0.1032 (0.1107)	0.3617 (0.5243)	59.6%
	10	0.1029 (0.1104)	0.3634 (0.5232)	60.1%
0.9	0.1	0.0289 (0.0673)	0.8082 (0.2101)	37.8%
	1	0.1006 (0.1093)	0.3766 (0.5148)	62.1%
	10	0.0960 (0.1075)	0.4008 (0.5045)	62.3%

4.8 MCMC application with a new proposal for the AR(2) model

It has been mentioned previously that estimating the parameters using the Metropolis algorithm provided poor estimation. It can be seen that results from parameter estimates using the Gibbs sampler were more accurate than using Metropolis. Therefore, the proposal of the random walk seems not to be able to estimate parameters of ϕ_1 and

ϕ_2 accurately at the same time. For instance, if ϕ_1 is estimated well, then ϕ_2 is not estimated as well.

Another problem that can be faced using the Metropolis approach is that we cannot provide informative priors for the AR(3) and AR(4) models. This means that information about the priors of $p(\phi_j/\phi_{(-j)})$ ($j \neq 1$) cannot be obtained. This is because an informative prior cannot be obtained and the random walks proposal cannot estimate parameters precisely. We sample from (4.16) and so we use Metropolis for $\phi_1|\phi_2, y$ with (4.17) as proposed. Thus, our revised recommended proposal has a truncated normal distribution with respect of the stationarity conditions as follows:

$$\phi_2 \sim N_{[\phi_1-1,1-\phi_1]}\left(\frac{\sum y_t y_{t-2} - \phi_1 \sum y_{t-1}y_{t-2}}{\sum y_{t-2}^2}, \frac{\sigma^2}{\sum y_{t-2}^2}\right) \tag{4.18}$$

The same posterior distributions of (4.15) and (4.16) are used to sample from $p(\phi_1|\phi_2, y, \sigma^2)$ and $p(\phi_2 \mid \phi_1, y, \sigma^2)$, hence the Gibbs sampler and Metropolis are applied using the new proposal in (4.18).

In order to evaluate the performance of our new MCMC approach, the same three simulated data sets of the previous section are used. Firstly, MCMC is applied using the first simulated data set of 150 observations when initial information of $\phi_1 = 0.8$, $\phi_2 = -0.8$ and $\sigma^2 = 1$ with a weakly informative prior on σ^2 will be explored (Ando, 2010). So we consider $\alpha = 3$ and $\beta = 10$. After applying MCMC using the new recommended proposal, the obtained parameter estimates are $\phi_1 = 0.796$ and $\phi_2 = -0.801$. This indicates that parameter estimation using the new proposal is more accurate than that using the random walk proposal. The difference between the true value and estimated value for ϕ_1 using the random walk proposal is 0.14. However, the difference between the true value and estimated value for ϕ_1 using our recommended proposal is 0.0032. Moreover, the difference between the true value and estimated value for ϕ_2 using the random walk

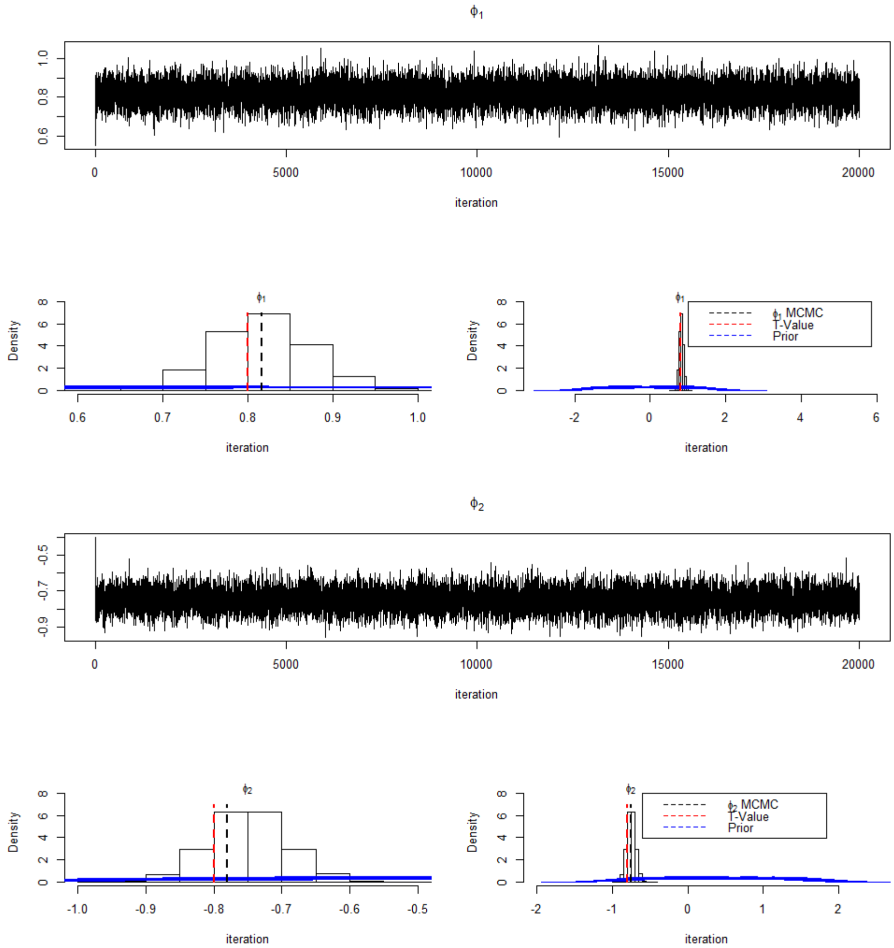


Figure 4.6: Trace plots and histogram of estimated parameters ϕ_1 and ϕ_2 via MCMC of the AR(2) model with K=20000, $\phi_1 = 0.8$ and $\phi_2 = -0.8$.

proposal is 0.4. However, the difference between the true value and estimated value for ϕ_2 using our recommended proposal is 0.001. This means that the accuracy of estimating parameters of the AR(2) model using the recommended proposal is considerably higher than using a random walk proposal. Figure 4.6 shows trace plots and histograms and illustrates the convergence of the MCMC for the AR(2) model.

Secondly, MCMC is again applied to the second simulated data set using initial information of $\phi_1 = -0.4$, $\phi_2 = -0.5$ and $\sigma^2 = 1$. Table 4.4 indicates the parameter estimates of ϕ_1 and ϕ_2. It can be seen that using the recommended proposal gives us more precise

Table 4.4: Shows the parameter estimation of the AR(2) model via MCMC application using the recommended proposal.

Parameters	ϕ_1 Mean	ϕ_1 Mode	ϕ_1 SD	ϕ_2 Mean	ϕ_2 Mode	ϕ_2 SD
real ϕ's	ϕ_1=0.8 and ϕ_2=-0.8					
Mean	0.7976	0.7971	0.0512	-0.8018	-0.8028	0.0505
real ϕ's	ϕ_1=-0.4 and ϕ_2=-0.5					
Mean	-0.4048	-0.4046	0.0938	-0.5068	-0.5084	0.0941
real ϕ's	ϕ_1=-0.3 and ϕ_2=-0.3					
Mean	-0.3004	-0.3052	0.1050	-0.3039	-0.3048	0.1046

results than using a random walk proposal, and errors have reduced to 0.02% and 0.02% for ϕ_1 and ϕ_2, respectively.

4.9 A prior distribution for the AR(3) model

For a given time series with AR coefficients ϕ_1, ϕ_2 and ϕ_3, we have corresponding stationarity conditions that are either in group A, B or AB. If we assume the true values of ϕ_1, ϕ_2 and ϕ_3 are corresponding to group A, then we need to provide a prior that is constrained on the stationarity region which is imposed by the conditions of group A. And, if the true values of ϕ_1, ϕ_2 and ϕ_3 are corresponding to group B, then we should choose a prior that reflects the stationary region of group B.

To cover the stationary conditions of the AR(3) model, conditions are divided into three groups in order to determine the stationary region of the AR(3) model as described in equations (3.35) to (3.39) in Section 3.6.1. For the purpose of determining the prior distribution for the parameters of the AR(3) model, the inequality stationary conditions are separated into three groups which are groups A, B and AB. The information has been mentioned in detail in Section 3.5.6 in order to know how to switch a new inequality stationary condition from equations (3.35) to (3.38) for two different stationary condition groups which are groups A and B. Details about the prior distribution of each of the stationary condition groups of the AR(3) model are discussed in the following sections.

4.9.1 Prior distribution for group A

One of the most important parts in conducting MCMC method is choosing the right prior in order to achieve the best result because MCMC is based on Posterior and Prior. As it is known that in every order of autoregressive model there is a number of conditions which have to be satisfied as they are already mentioned in section 3.5.6. Therefore, in AR(3) model, the conditions of equations of (3.35) to (3.38) have been reached.

These conditions are used in simulation study for the stationary assumption purpose and they are divided into two main groups named (A) and (B). Likewise, group (A) is based on (3.35),(3.36) and (3.37) conditions as well as group (B) is based on ((3.35),(3.36) and (3.38)) conditions. The purpose of making these two groups is to cover the stationary area for AR(3) model. From section 3.5.3 we know that a time series is being stationary if at least one of the group conditions are satisfied. Prior condition for AR(3) model as mentioned above we have three group constraints of inequality which are;

Group A conditions:

$$\phi_1 + \phi_2 + \phi_3 < 1 \tag{4.19}$$

$$-\phi_1 + \phi_2 - \phi_3 < 1 \tag{4.20}$$

$$-\phi_1 - \phi_2 + \phi_3 < 1 \tag{4.21}$$

$$|\phi_3| < 1. \tag{4.22}$$

From the conditions of (4.19)-(4.22), we have the following inequalities:

$$\phi_1 < 1 - \phi_2 - \phi_3 \tag{4.23}$$

$$\phi_1 > \phi_2 - \phi_3 - 1 \tag{4.24}$$

$$\phi_1 > -\phi_2 + \phi_3 - 1. \tag{4.25}$$

From equations (4.24) and (4.25), we can get that $\phi_1 > |\phi_2 - \phi_3| - 1$. Additionally, from equations (4.23), (4.24) and (4.25), we see that ϕ_1 lies in the range $|\phi_2 - \phi_3| - 1 < \phi_1 < 1 - \phi_2 - \phi_3$. Therefore, we propose that the conditional prior distribution of $\phi_1 \mid \phi_2, \phi_3$ is uniform, i.e., $\phi_1 \mid \phi_2, \phi_3 \sim U(|\phi_2 - \phi_3| - 1, 1 - \phi_2 - \phi_3)$. So, the prior distribution of ϕ_1 given ϕ_2, ϕ_3 is:

$$p(\phi_1 \mid \phi_2, \phi_3) = \frac{1}{2 - \phi_2 - \phi_3 - |\phi_2 - \phi_3|} I_{[a]}. \tag{4.26}$$

where

$$I_{[a]} = \begin{cases} 1 & a \in A \\ 0 & a \notin A \end{cases} \tag{4.27}$$

where A={(ϕ_1, ϕ_2, ϕ_3): $|\phi_2 - \phi_3| - 1 < \phi_1 < 1 - \phi_2 - \phi_3$, $|\phi_2| < 1$ and $|\phi_3| < 1$ }.

By adding inequalities (4.19) and (4.20), we have $\phi_2 < 1$, and from equation (4.21), we obtain $\phi_2 > \phi_3 - \phi_1 - 1$. Hence, ϕ_2 lies in the range $\phi_3 - \phi_1 - 1 < \phi_2 < 1$. Thus, we propose that the prior of distribution $\phi_2 \mid \phi_1, \phi_3$ is uniform, i.e., $\phi_2 \mid \phi_1, \phi_3 \sim U(\phi_3 - \phi_1 - 1, 1)$.This prior distribution has density function

$$p(\phi_2 \mid \phi_1, \phi_3) = \frac{1}{2 + \phi_1 - \phi_3} I_{[-2.9,1]} \tag{4.28}$$

We have chosen the range of ϕ_2 to be (-2.9, 1) based on the 3-D plot of Figure 3.5 which was already shown in Section 3.5.5. Regarding ϕ_3, we propose that the marginal prior distribution of ϕ_3 is uniform, i.e., $\phi_3 \sim U(-1,1)$, with density

$$p(\phi_3) = \frac{1}{2} I_{[-1,1]} \tag{4.29}$$

Now, from equations (4.26) (4.28) and (4.29), the joint prior distribution group A is:

$$\begin{aligned} p(\phi_1, \phi_2, \phi_3) &\propto p(\phi_1 \mid \phi_2, \phi_3).p(\phi_2 \mid \phi_1, \phi_3).p(\phi_3) \\ &\propto \frac{1}{(2 + \phi_1 - \phi_3)(2 - \phi_2 - \phi_3 - |\phi_2 - \phi_3|)} \end{aligned} \tag{4.30}$$

4.9.2 Prior distribution for group B

Similarly, in order to choose the prior distribution for the stationary conditions for group B, from equations (3.35), (3.36) and (3.38), we can see that ϕ_1 satisfies

$$\phi_1 < 1 - \phi_2 - \phi_3 \tag{4.31}$$

$$\phi_1 > \phi_2 - \phi_3 - 1 \tag{4.32}$$

$$\phi_1 < 1 + \phi_2 + \phi_3 \tag{4.33}$$

From equations (4.31) and (4.33), the range of ϕ_1 is

$$\phi_1 < 1 - |\phi_2 + \phi_3| \tag{4.34}$$

With reference to equations (4.32) and (4.34), we end up with ϕ_1 having the uniform distribution on the interval

$\phi_2 - \phi_3 - 1 < \phi_1 < 1 - |\phi_2 + \phi_3|$.

Therefore, the prior distribution of $\phi_1 \mid \phi_2, \phi_3$ is

$$p(\phi_1 \mid \phi_2, \phi_3) \quad = \quad \frac{1}{2 - |\phi_2 + \phi_3| - \phi_2 + \phi_3} I_{[-1,1]}. \tag{4.35}$$

Also, regarding the prior distribution of $\phi_2 \mid \phi_1, \phi_3$, we follow the same strategy as we did for group A. By adding equation (4.31) and (4.32) we can get that ϕ_2 is less than one, i.e., $(\phi_2 < 1)$, and from equations (4.33), we can obtain

$$\phi_2 > \phi_1 - \phi_3 - 1, \tag{4.36}$$

so that $\phi_1 - \phi_3 - 1 < \phi_2 < 1$. Thus, the prior distribution for ϕ_2 is:

$$p(\phi_2 \mid \phi_1, \phi_3) = \frac{1}{2 - \phi_1 + \phi_3} \tag{4.37}$$

By multiplying equations (4.29), (4.35) and (4.37) together, the joint prior distribution for group B is:

$$p(\phi_1, \phi_2, \phi_3) \quad \propto \quad \frac{1}{(2 - |\phi_2 + \phi_3| - \phi_2 + \phi_3)(2 - \phi_1 + \phi_3)} \tag{4.38}$$

4.9.3 Prior distribution for Group AB

In Section 3.5.1, we have proposed a prior for each group separately. Here, we combine both groups into a group that involves all four conditions. Next we work out the range of ϕ_1 in order to propose a suitable prior distribution. From equations (3.35) to (3.38) we have

$$\phi_1 < 1 - \phi_2 - \phi_3 \tag{4.39}$$

$$\phi_1 > \phi_2 - \phi_3 - 1 \tag{4.40}$$

$$\phi_1 > -\phi_2 + \phi_3 - 1 \tag{4.41}$$

$$\phi_1 < 1 + \phi_2 + \phi_3 \tag{4.42}$$

Thus, from equations (4.39) and (4.42) we obtain that $\phi_1 < 1 - |\phi_2 + \phi_3|$. Likewise from equations (4.40) and (4.41) we can also obtain that $\phi_1 > |\phi_2 - \phi_3| - 1$. Hence, the range of ϕ_1 is $|\phi_2 - \phi_3| - 1 < \phi_1 < 1 - |\phi_2 + \phi_3|$. Therefore, we propose that $\phi_1 \mid \phi_2, \phi_3$ is uniformly distributed, i.e., $\phi_1 \mid \phi_2, \phi_3 \sim U(|\phi_2 - \phi_3| - 1, 1 - |\phi_2 + \phi_3|)$. So, the prior density for $\phi_1 \mid \phi_2, \phi_3$ is:

$$p(\phi_1 \mid \phi_2, \phi_3) = \frac{1}{2 - |\phi_2 + \phi_3| - |\phi_2 - \phi_3|} I_{[a]} \tag{4.43}$$

where

$$I_{[a]} = \begin{cases} 1 & a \in A \\ 0 & a \notin A \end{cases}$$

with $A = \{ (\phi_1, \phi_2, \phi_3) : |\phi_2 - \phi_3| - 1 < \phi_1 < 1 - |\phi_2 + \phi_3| \,,\ |\phi_2| < 1$ and $|\phi_3| < 1)\}$. By adding equations (4.39) and (4.40), we see that ϕ_2 is less than minus one, i.e., $\phi_2 < 1$. And, by adding equations (4.41) and (4.42) we can also obtain that ϕ_2 is greater than one, i.e., $\phi_2 > -1$. Thus, the range of ϕ_2 is $-1 < \phi_2 < 1$. It can be noticed that ϕ_2 is independent from ϕ_3 . Therefore, the prior distribution for $\phi_2 | \phi_3$ is uniform, i.e., $\phi_2 | \phi_3 \sim U(-1, 1)$.

$$p(\phi_2 | \phi_3) = \frac{1}{2} \tag{4.44}$$

Therefore, the joint prior distribution for the AB group is:

$$p(\phi_1, \phi_2, \phi_3) \quad \propto \quad \frac{1}{4(2 - |\phi_2 + \phi_3| - |\phi_2 - \phi_3|)} \tag{4.45}$$

4.10 Posterior inference of the AR(3) model

After proposing a prior distribution for the parameters of the AR(3) model, the posterior distribution can be routinely obtained. To this end, we assume that n observations are available, say $y_1, y_2, \ldots, y_n$, from the AR(3) model.

$$y_t = \phi_1 y_{t-1} - \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t,$$

where ε_t is white noise and $\varepsilon_t \sim N(0, \sigma^2)$. The unknown parameters here are ϕ_1, ϕ_2, ϕ_3 and σ^2. As has been stated in the previous sections, inequality stationary conditions for the AR(3) model are divided into three different groups. Therefore, in the calculation of the posterior distribution of a parameter restricted to lie within a particular group (A, B, or AB) knowledge of that group is required. Here we only consider the case when ϕ_1, ϕ_2 and ϕ_3 lie in group AB of the previous Section, but Section 4.14 discusses this in more detail. The conditional posterior distribution of ϕ_1 is:

$$\begin{aligned} p(\phi_1 \mid \phi_2, \phi_3, y, \sigma^2) &\propto p(y \mid \phi_1, \phi_2, \phi_3, \sigma^2).p(\phi_1|\phi_2, \phi_3) \\ &= e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2} \times \frac{1}{2 - |\phi_2 + \phi_3| - |\phi_2 - \phi_3|} \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2}.I_{[a]}, \end{aligned} \tag{4.46}$$

where

$$I_{[a]} = \begin{cases} 1 & a \in A \\ 0 & a \notin A \end{cases} \tag{4.47}$$

with A=$\{ (\phi_1, \phi_2, \phi_3) : |\phi_2 - \phi_3| - 1 < \phi_1 < 1 - |\phi_2 + \phi_3|, \ |\phi_2| < 1$ and $|\phi_3| < 1\}$. So, in order to obtain the posterior distribution of $\phi_1 \mid \phi_2, \phi_3, y, \sigma^2$, we write $z_t = y_t - \phi_2 y_{t-2} - \phi_3 y_{t-3}$ and then

$$p(\phi_1 \mid \phi_2, \phi_3, y, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum(z_t - \phi_1 y_{t-1})^2}.I_{[a]} \quad (4.48)$$

Now, the part of $\sum(z_t - \phi_1 y_{t-1})^2$ from equation (4.48) can be expanded by following the same steps as in equation (4.2), hence

$$p(\phi_1 \mid \phi_2, \phi_3, y, \sigma^2) \propto e^{-\frac{\sum y_{t-1}^2}{2\sigma^2}\left(\phi_1 - \frac{\sum z_t y_{t-1}}{\sum y_{t-1}^2}\right)^2} I_{[a]}.$$

It can be observed that the posterior distribution for $\phi_1|\phi_2, \phi_3 y, \sigma^2$ is a truncated normal distribution with mean and variance specified as follows:

$$\phi_1 \mid \phi_2, \phi_3, y, \sigma^2 \sim N_{[a,b]}\left(\frac{\sum y_t y_{t-1} - \phi_2 \sum y_{t-1} y_{t-2} - \phi_3 \sum y_{t-1} y_{t-3}}{\sum y_{t-1}^2}, \frac{\sigma^2}{\sum y_{t-1}^2}\right). \quad (4.49)$$

where $a = 1 - |\phi_2 + \phi_3|$ and $b = |\phi_2 - \phi_3| - 1$. In order to derive the conditional posterior distribution of $\phi_2|\phi_1, \phi_3, y, \sigma^2$, we follow a similar argument as before, i.e.,

$$\begin{aligned} p(\phi_2 \mid \phi_1, \phi_3, y, \sigma^2) &\propto p(y \mid \phi_1, \phi_2, \phi_3, \sigma^2) p(\phi_2 \mid \phi_1, \phi_3) \\ &\propto p(y \mid \phi_1, \phi_2, \phi_3, \sigma^2) p(\phi_1, \phi_2, \phi_3) \quad (4.50) \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2} \times \frac{1}{2 - |\phi_2 + \phi_3| - |\phi_2 - \phi_3|}.I_{[-1,1]} \end{aligned}$$

For the conditional posterior distribution of $\phi_3|\phi_1, \phi_2, y, \sigma^2$, we get

$$\begin{aligned} p(\phi_3 \mid \phi_1, \phi_2, y, \sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2} p(\phi_3 \mid \phi_1, \phi_2) \\ &\propto p(y \mid \phi_1, \phi_2, \phi_3, \sigma^2) p(\phi_1, \phi_2, \phi_3) \quad (4.51) \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2} \times \frac{1}{2 - |\phi_2 + \phi_3| - |\phi_2 - \phi_3|}.I_{[-1,1]} \end{aligned}$$

Now we turn our attention to the posterior distribution for both groups A and B. We can apply similar steps to the posterior distribution for the AB group.

First, the posterior distribution of group A is:

$$\phi_1 \mid \phi_2, \phi_3, y, \sigma^2 \sim N_{[a,b]}\left(\frac{\sum y_t y_{t-1} - \phi_2 \sum y_{t-1}y_{t-2} - \phi_3 \sum y_{t-1}y_{t-3}}{\sum y_{t-1}^2}, \frac{\sigma^2}{\sum y_{t-1}^2}\right) I_{[a]}. \tag{4.52}$$

$I_{[a]}$. is the same structure as has been shown in equation (4.47) using different A which is

A={ $(\phi_1, \phi_2, \phi_3) : |\phi_2 - \phi_3| - 1 < \phi_1 < 1 - \phi_2 - \phi_3\,,\ \phi_3 - \phi_1 - 1 < \phi_2 < 1$ and $|\phi_3| < 1$ }.

$$p(\phi_2 \mid \phi_1, \phi_3, y, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2} \times \frac{1}{2 - \phi_2 - \phi_3 - |\phi_2 - \phi_3|} I_{[-1,1]}. \tag{4.53}$$

$$p(\phi_3 \mid \phi_1, \phi_2, y, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2} \times \frac{1}{2 - |\phi_2 + \phi_3| - |\phi_2 - \phi_3|} I_{[a]} \tag{4.54}$$

Regarding the posterior distribution of B, we have

$$\phi_1 \mid \phi_2, \phi_3, y, \sigma^2 \sim N_{[a,b]}\left(\frac{\sum y_t y_{t-1} - \phi_2 \sum y_{t-1}y_{t-2} - \phi_3 \sum y_{t-1}y_{t-3}}{\sum y_{t-1}^2}, \frac{\sigma^2}{\sum y_{t-1}^2}\right) I_{[a]}. \tag{4.55}$$

where

A={ $(\phi_1, \phi_2, \phi_3) : \phi_2 - \phi_3 - 1 < \phi_1 < 1 - |\phi_2 - \phi_3|\,,\ \phi_3 - \phi_1 - 1 < \phi_2 < 1$ and $|\phi_3| < 1$ }.

$$p(\phi_1 \mid \phi_2, \phi_3, y, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3})^2} \times \frac{1}{2 - \phi_2 + \phi_3 - |\phi_2 + \phi_3|} I_{[a]} \tag{4.56}$$

$$p(\phi_3 \mid \phi_1, \phi_2, y, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3})^2} \times \frac{1}{2-|\phi_2+\phi_3|-|\phi_2-\phi_3|} \tag{4.57}$$

By following the same steps of equation (4.5), the posterior distribution of $\frac{1}{\sigma^2}$ for each group is:

$$\frac{1}{\sigma^2}|y,\phi_1,\phi_2,\phi_3 \sim G\left(a+\frac{n}{2}, b+\frac{1}{2}\sum_{t=1}^{n}(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3})^2\right). \tag{4.58}$$

As has been indicated, we have been able to assign a posterior distribution for the AR(3) model; we also could assign one for all group conditions by following the same steps as for group AB. In the next section how MCMC can be applied for the AR(3) model will be discussed. In the conclusion of this section, we derive conditional distributions assuming the parameters lie in groups A, B and AB. In the next section we will implement these MCMC schemes in R.

4.11 An MCMC application for the AR(3) model

To estimate parameters of the AR(3) model, 150 observations are simulated from the AR(3) model where $\sigma = 1$ in order to obtain estimates of the parameters ϕ_1 , ϕ_2 and ϕ_3 for the group conditions of A, B and AB as mentioned previously. The Gibbs sampling approach is used to estimate ϕ_1 for each equation (4.49), (4.52) and (4.55) of the group conditions assuming that each ϕ_1 has the truncated normal distribution. A Metropolis step is used to estimate the parameters ϕ_2 and ϕ_3 of all group conditions using a random walk as a proposal. It can be seen from Table 4.5 that two simulated data sets are employed to estimate the parameters of the AR(3) model using different values of the variance of the proposal s. The variance σ^2 has an Inverse Gamma distribution, hence the precision $\frac{1}{\sigma^2} \sim Gamma(\alpha, \beta)$ assuming $\alpha = 3$ and $\beta = 10$.

Table 4.5: Shows the parameter estimates of the AR(3) model via MCMC application. We used the Gibbs sampling to obtain ϕ_1 and Metropolis is used to obtain ϕ_2 and ϕ_3 when the proposal is random walk with k=30000, m=1000, $\alpha = 3$ and $\beta = 10$.

s	$\phi_1(Sd)$	$\phi_2(Sd)$	$\phi_3(Sd)$	Acc .Rate
$\phi_1 = 0.1$, $\phi_2 = 0.2$ and $\phi_3 = 0.1$				
0.99	0.153(0.148)	-0.226(0.678)	0.0217(0.675)	45.6%
0.9	0.006(0.112)	0.733(0.609)	0.176(0.415)	7.4%
0.5	0.154(0.122)	-0.086(0.680)	-0.278(0.742)	15.4%
0.1	0.065(0.138)	0.468(0.740)	0.237(0.664)	12.1%
0.01	0.123(0.184)	0.144(0.602)	-0.412(0.801)	9.2%
Different parameter		$\phi_1 = -0.4$, $\phi_2 = -0.8$ and $\phi_3 = -0.6$		
0.99	-0.497(0.272)	-0.085(0.618)	-0.681(0.394)	13.1%
0.9	-0.535(0.200)	-0.287(0.514)	-0.744(0.403)	5.9%
0.5	-0.383(0.188)	-0.677(0.563)	-0.201(0.686)	6.1%
0.1	-0.559(0.127)	-0.052(0.421)	-0.695(0.643)	3.21%
0.01	-0.382(0.112)	-0.368(0.469)	0.779(0.552)	1.15%

Because of the fact that the simulated data sets satisfy both of the groups AB and A, therefore the posterior distributions of both group conditions AB and A are used to estimate the parameters, respectively. After applying MCMC to the two data sets, it can be noted that parameter estimation using both Gibbs sampling and Metropolis steps do not give us precise and consistent results. This suggests that our parameter estimates are poor. However, using the Gibbs sampler to estimate ϕ_1 for the group condition A is more closer to its true value. It is worth mentioning that we have used other different simulated dataset parameters, but poor estimates throughout persist. Furthermore, Figure 4.7 indicates that none of the simulated parameters have reached convergence after 30000 iterations.

In the simulated data considered here, the true values of the parameters are known, hence the true groups A, B, AB within which the parameters lie are also known. In real data these groups will not be known, hence there must be some extra uncertainty associated with the prior distribution. This is highlighted in Section 4.14 by utilizing Bayes factors for model choice.

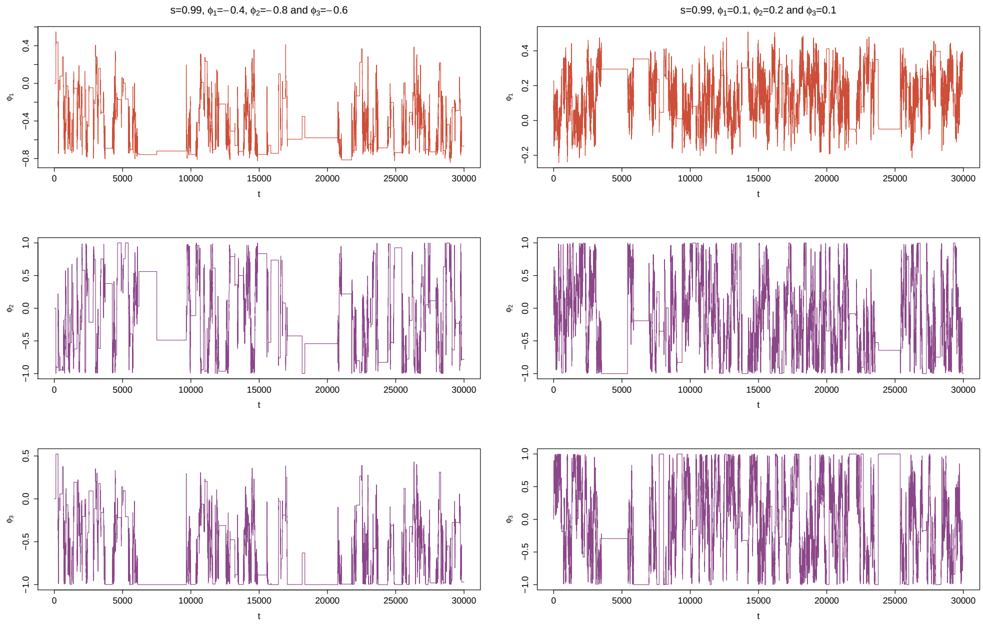


Figure 4.7: Trace plots of the simulated parameters ϕ_1 , ϕ_2 and ϕ_3 for the AR(3) model.

4.12 New MCMC proposal distribution for the AR(3) model

After realizing that the Random walk proposal was unable to precisely estimate the parameters of the AR(3) model, as mentioned in Section (4.11), we thought to use a new proposal, similar to the proposal used for the AR(2) model in Section (4.8). The idea of the new proposal is to assume a proposal for the parameters ϕ_2 and ϕ_3 in the form of a truncated normal distribution as follows:

$$\phi_2 \sim N_{[-1,1]}\left(\frac{\sum y_t y_{t-2} - \phi_1 \sum y_{t-1}y_{t-2} - \phi_3 \sum y_{t-2}y_{t-3}}{\sum y_{t-2}^2}, \frac{\sigma^2}{\sum y_{t-2}^2}\right) \tag{4.59}$$

$$\phi_3 \sim N_{[-1,1]}\left(\frac{\sum y_t y_{t-3} - \phi_1 \sum y_{t-1}y_{t-3} - \phi_2 \sum y_{t-2}y_{t-3}}{\sum y_{t-3}^2}, \frac{\sigma^2}{\sum y_{t-3}^2}\right) \tag{4.60}$$

Regarding the posterior distribution of $p(\phi_1 \mid \phi_2, \phi_3, y, \sigma^2)$, by fixing the right-hand side of equation (4.46) we obtained a good sufficient result for ϕ_1 using the Gibbs sampler.

We obtain a new proposal of equation (4.59) and (4.60) by fixing the right hand sides of equations (4.50) and (4.51). Our new proposal is sufficient since it provides a good estimate.

Regarding the conditional distributions for ϕ_2 and ϕ_3, the same strategies of Section 4.11 are used for $\phi_1|\phi_2, \phi_3, y$. Through applying the Gibbs sampler ϕ_1 parameters are estimated, see Figure 4.8, for each of the group conditions the AB, A and B via equations (4.49), (4.52) and (4.55). Regarding the conditional posteriors of $\phi_2|\phi_1, \phi_3, y$ and $\phi_3|\phi_1, \phi_2, y$, the parameters of ϕ_2 and ϕ_3 are estimated using a Metropolis algorithm for each group condition AB, A and B via equations (4.50), (4.53) and (4.56) to estimate the parameter ϕ_2 and (4.51), (4.54) and (4.57) to estimate the parameter ϕ_3. Figure 4.9 shows the competence and the precision of the MCMC process when estimating the parameters ϕ_2 and ϕ_3 for the AR(3) model, this is when the true parameters in the simulation process are $\phi_1 = -0.4$, $\phi_2 = -0.8$ and $\phi_3 = -0.6$. Regarding σ^2, we use an inverse gamma distribution in parallel with the relevant discussion for the AR(2) model. Based on our prior beliefs of σ^2, $\alpha = 3$ and $\beta = 10$ is used. With regards to the acceptance rate as mentioned for the AR(2) model, its value is high because our proposals are close to the posterior distributions, see Hoff (2009) and Robert and Casella (2010). The blue lines of Figure 4.8 indicates we have used different priors $p(\phi_i)$.

4.13 Bayes Factor

Model selection can be performed via the so-called posterior odds, that is the product of the Bayes factor and the prior odds. The Bayes factor amongst a null model and an alternative model is the ratio of their likelihoods. Given any two models, this leads to the assumption of posterior model probabilities (Steel, 2008). Suppose we desire to compare two models with the same mathematical structure, varying them only via the value of

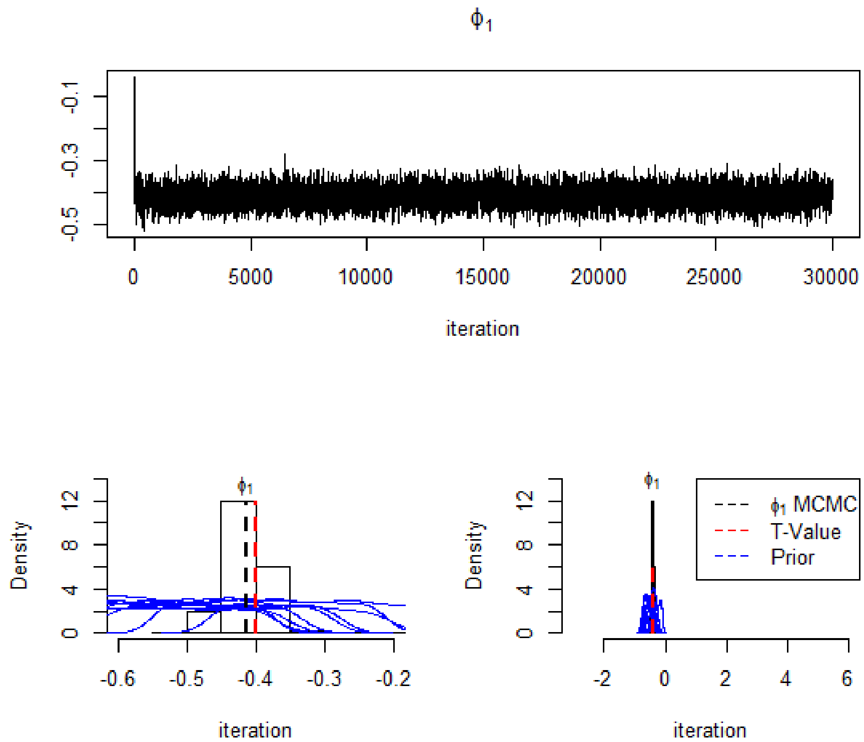


Figure 4.8: Shows the converged results of the parameter estimates for the AR(3) model via the Gibbs sampling with K=30000, $\phi_1 = -0.4$, $\phi_2 = -0.8$, $\phi_3 = -0.6$, $\alpha = 3$ and $\beta = 10$. The blue dashed lines indicate that different priors is used for ϕ_1, and the histogram of the right-hand side is zoomed from the histogram of the left-hand side.

their defining parameters. Let M_0 the null model and let M_1 indicate the alternative one. Furthermore, each model gives a predictive distribution for y_t given ϕ at time t. The densities are:

$$p(y_t|\phi, M_i) \qquad where \;\; i = 0, 1$$

where ϕ is historical knowledge that is common to the two models at the time t , and y_t is an observation from a time series. The involvement of M_i in the above formula differentiates between the two models which are M_0 and M_1. Therefore, the predictive

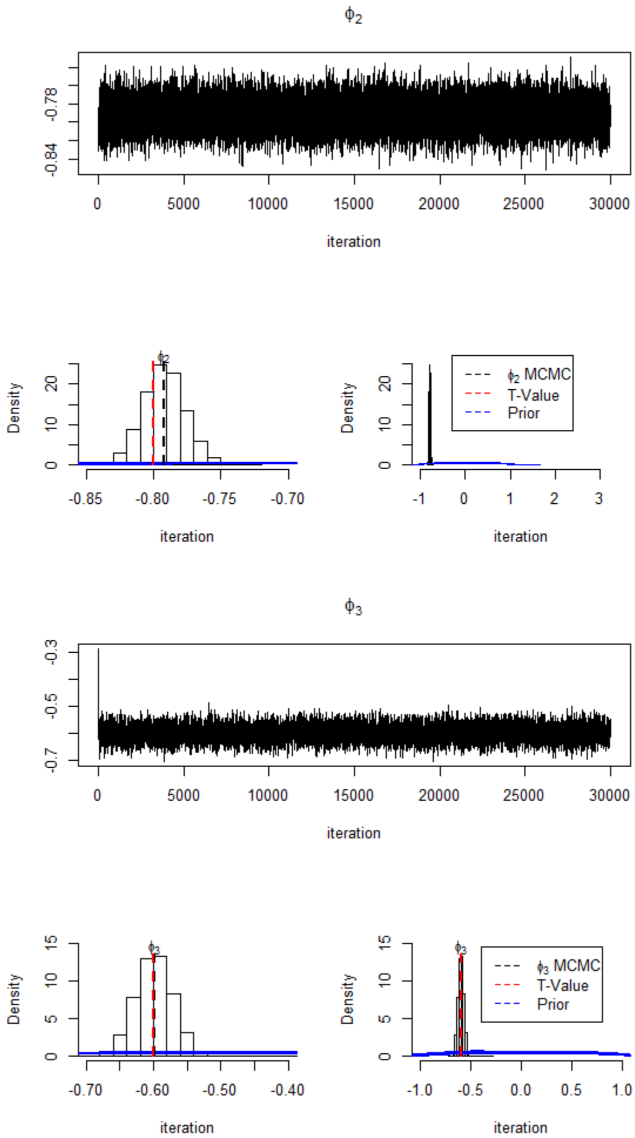


Figure 4.9: Converged results of the parameter estimates for ϕ_2 and ϕ_3 via the Metropolis algorithm with K=30000, $\phi_1 = -0.4$, $\phi_2 = -0.8$, $\phi_3 = -0.6$, $\alpha = 3$ and $\beta = 10$ the blue dashed lines indicate that different priors is used for ϕ_2 and ϕ_3, and the histograms of the right-hand side are zoomed from the histograms of the left-hand side.

densities at time t are (Kass and Raftery (1995), and West and Harrison (1999)):

$$p(y_t|\phi) = p(y_t|\phi, M_i) \qquad where \;\; i = 0, 1$$

Thus, the Bayes factor for M_0 against M_1 based on the observed data of y_t is:

$$K = \frac{p(y|M_1)}{p(y|M_0)} = \frac{\int_{\phi_1} p_1(y|\phi_1)p(\phi_1)d\phi_1}{\int_{\phi_0} p_0(y|\phi_0)p(\phi_0)d\phi_0},$$

K is known as the Bayes factor (BF). If K is large, the data indicate that there is more evidence in support of the null model, and less in favour of the alternative model. Jeffreys (1998) has used some rules for interpreting the BF as follows:

- If 1 $<$ Bayes Factor $\leq$ 3, then there is weak evidence for M_1
- If 3 $<$ Bayes Factor $\leq$ 10, then there is substantial evidence for M_1
- If 10 $<$ Bayes Factor $\leq$ 100, then there is strong evidence for M_1
- If Bayes Factor $>$ 100, then there is decisive evidence for M_1

4.14 MCMC procedure of estimating parameters of the AR(3) model using Bayes factors

There are several conditions in the AR(3) model in order to guarantee stationarity, as discussed in Section 4.9. In our MCMC method we have used both Gibbs sampling and Metropolis Hastings. As given in the theory part, in order to cover stationarity of the AR(3) model, the conditions are divided into three groups A, B and AB. Subsequently, MCMC is applied using the posterior distributions, which were obtained for every group condition, see Section 4.10. Hence, the Bayes factor was used to compare every pair of models from the three models for each group in order to make a distinction between them.

First of all we used the truncated normal distribution since the posterior distribution of ϕ_1 in all groups has to be limited to a particular range. After that we applied the Bayes factor which tells us whether group A works better on a given dataset or group B based on the parameter estimates.

All of the estimated parameters are involved, which means we check every single one of them and use them in the Bayes factor formula as mentioned in Section 4.13. In order to perform our MCMC method, we need to have specified data as the algorithm is based on it. Therefore, we used a simulated dataset generated using particular parameters with values $\phi_1 = 0.5, \phi_2 = -0.4$ and $\phi_3 = -0.4$ that satisfy group condition A as an example. We simulated this data to give priority to group A conditions and then finally we will approve it via the Bayes factor.

MCMC is then applied for group A conditions and as we know from theory (Section 4.9.1) that ϕ_1 in both group conditions is simulated by a Gibbs sampling step. Then, we took another step forward to estimate ϕ_2 and ϕ_3 via Metropolis, see equations (4.52) and (4.55). This means we need a proposal for each parameter and we recommend the truncated normal proposal, see equations (4.59) and (4.60). Table 4.6 shows that the accuracy of parameter estimation for group A is, to a great extent, satisfactory, as the errors between the estimated parameters and the true parameters are between 2% and 6%, which is achieved when the posterior of A (M_o) is applied. More importantly, when the posterior distribution of the AB model which is used as an alternative model (M_1) in this process, the accuracy of estimation is better than the accuracy of parameter estimations when applying MCMC for the group A, in this case, the error does not exceed 2%. Similarly, the same procedure is used for group B conditions (see Table 4.7).

In the Bayes factor we include every estimated value of the parameters and the results of this procedure is presented in Table 4.6. We used the data with parameters satisfying the group A conditions which means that the result can tell us that group A supports the data better than group B.

Figures 4.10 and 4.11 show that the parameter estimates are very close to each other using the

Table 4.6: Results of MCMC and Bayes factor of some of the different parameters that have been used for simulated data. The null model is group A and the alternative model is either group B or AB.

Real ϕ's	0.5	-0.4	-0.4	BF (Sd)
M_o (A)	0.5316(0.076)	-0.445(0.082)	-0.343(0.078)	1.040(0.082)
M_1 (B)	0.1907(0.073)	-0.176(0.066)	-0.602(0.069)	
M_o (A)	0.5316(0.076)	-0.445(0.082)	-0.343(0.078)	1.063(0.067)
M_1 (AB)	0.5207(0.076)	-0.388(0.068)	-0.401(0.069)	
Real ϕ's	0.7	-0.5	-0.2	BF(Sd)
M_o (A)	0.7515(0.069)	-0.611(0.075)	-0.191(0.071)	1.040(0.082)
M_1 (B)	0.5439(0.073)	-0.433(0.066)	-0.330(0.069)	
M_o (A)	0.7538(0.070)	-0.513(0.176)	-0.188(0.069)	1.064(0.066)
M_1 (AB)	0.7661(0.074)	-0.501(0.067)	-0.186(0.067)	

Table 4.7: Results of MCMC and Bayes factor of some of the different parameters that have been used for simulated data. The null model is group B and the alternative model is either group A or AB.

Real ϕ's	-0.6	-0.6	-0.2	BF(Sd)
M_o (B)	-0.611(0.078)	-0.584(0.080)	-0.170(0.079)	1.006(0.006)
M_1 (A)	-0.456(0.082)	-0.431(0.079)	0.0211(0.083)	
M_o (B)	-0.613(0.079)	-0.585(0.080)	-0.174(0.080)	1.006(0.006)
M_1 (AB)	-0.613(0.086)	-0.540(0.081)	0.184(0.084)	
Real ϕ's	-0.68	0.4	0.74	BF(Sd)
M_o (B)	-0.579(0.116)	0.3948(0.136)	0.7487(0.111)	1.019(0.053)
M_1 (A)	-0.537(0.048)	0.4212(0.057)	0.795(0.059)	
M_o (B)	-0.582(0.117)	0.394(0.137)	0.7499(0.110)	1.020(0.053)
M_1 (AB)	-0.540(0.050)	0.4149(0.057)	0.757(0.062)	

Table 4.8: Results of MCMC and Bayes factor of some of the different parameters that have been used for simulated data. The null model is group AB and the alternative model is either group B or B.

Real ϕ's	-0.2	0.3	0.8	BF(Sd)
M_o (AB)	-0.159(0.067)	0.3164(0.054)	0.75(0.064)	1.085(0.421)
M_1 (A)	-0.155(0.067)	0.3149(0.056)	0.7424(0.065)	
M_o (AB)	-0.159(0.078)	0.3075(0.066)	0.7493(0.072)	1.155(0.519)
M_1 (B)	-0.147(0.076)	0.3136(0.063)	0.7253(0.073)	
Real ϕ's	-0.3	0.3	0.6	BF(Sd)
M_o (AB)	-0.403(0.060)	0.2632(0.074)	0.6628(0.070)	1.030(0.268
M_1 (A)	-0.401(0.062)	0.263(0.073)	0.6573(0.069)	
M_o (AB)	-0.403(0.061)	0.2674(0.072)	0.661(0.07)	1.050(0.277)
M_1 (B)	-0.406(0.070)	0.2718(0.078)	0.6411(0.069)	
Real ϕ's	0.7	0.7	-0.8	BF(Sd)
M_o (AB)	0.6252(0.060)	0.7291(0.056)	-0.796(0.062)	1.263(0.417
M_1 (A)	0.654(0.051)	0.7006(0.054)	-0.86(0.075)	
M_o (AB)	0.6642(0.064)	0.7047(0.047)	-0.845(0.064)	1.018(0.306)
M_1 (B)	0.6569(0.051)	0.7058(0.047)	-0.852(0.066)	

prior of group condition A compared to the prior of group AB, and using the prior of group B compared to the prior of group AB. This means that although the Bayes factor could not reject the alternative model, the prior of group conditions AB approximately satisfy both group conditions A and B.

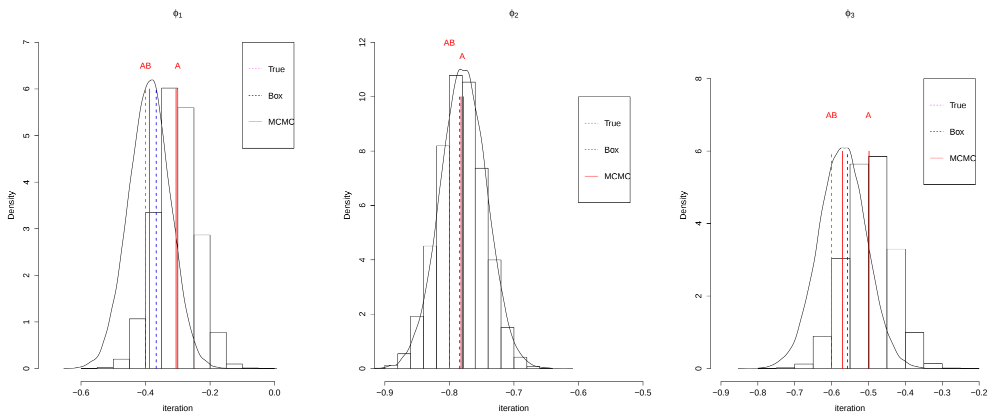


Figure 4.10: Shows the distinction of the parameter estimates of the AR(3) model between null model (AB) and alternative model (A). the red dot line is the true values, the blue dashed line is the parameter estimates using Box-Jenkins and the red line is the parameter estimates using MCMC. The curve shapes are the null models and the histograms are the alternative models.

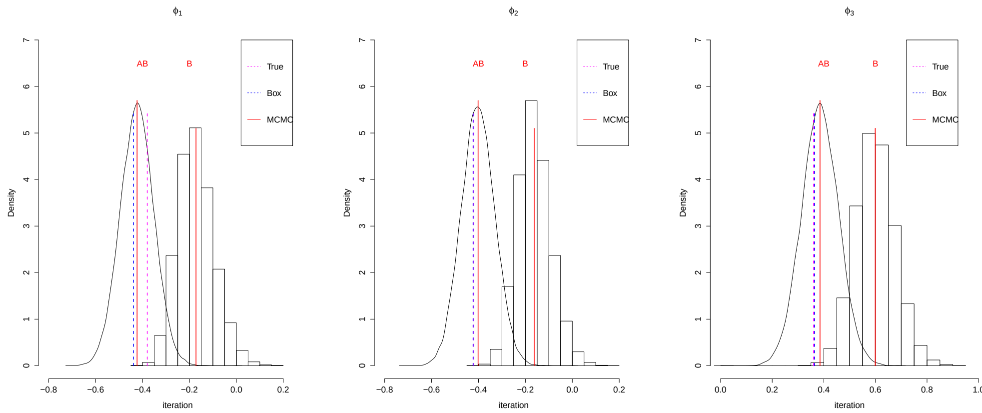


Figure 4.11: Shows the distinction of the parameter estimates of the AR(3) model between null mode (AB)l and alternative model (B). the red dot line is the true values, the blue dashed line is the parameter estimates using Box-Jenkins and the red line is the parameter estimates using MCMC. The curve shapes are the null models and the histograms are the alternative models.

We verify easily this conclusion with our actual values of $\phi_1 = 0.5$, ϕ_2 =-0.4 and ϕ_3 =-0.4. We obtain ϕ_1 =0.532 and ϕ_2 =-0.445 and ϕ_3 =-0.343 based on group A conditions, while we obtain ϕ_1 =0.191, ϕ_2 =-0.177 and ϕ_3 =-0.603 based on group B conditions. We have seen that the results of the Bayes factor were around one (see Table 4.6), we therefore could not reject the alternative model, see Kass and Raftery (1995) and Triantafyllopoulos and Bersimis (2016). We have repeated this process for several different values of the parameters. The results are tabulated in Tables 4.7 and 4.8.

Regarding the intersection of group conditions of A and B ($A \cap B$), we did not consider the true values of parameters of the AR(3) model when they are in the $A \cap B$. This is because when we consider $A \cap B$, they only satisfy linear inequality equations (3.35) and (3.36). These inequality conditions are already rejected when we use rejection sampling because they lie outside outside the overall stationarity conditions

4.15 Performance of the Bayes factor for determining the AR(3) model

In this section we conduct a simulation study in order to understand the performance of the Bayes factor, in particular why it is sometimes unable to reject the alternative model when model B is used as an alternative model for model A, despite the difference noticed in the precision of parameter estimation for the two models. This case is studied via data simulation for the AR(3) model based on the estimated parameters for both models. 150 observations are simulated for the AR(3) model based on the condition that the parameters satisfy group A. For instance, the chosen values of ϕ_1 , ϕ_2 and ϕ_3 were 0.5, -0.4 and -0.4, respectively. Then, MCMC is applied on the simulated data with the purpose of estimating the parameters of the AR(3) model via the prior distributions available for groups A and B (see equations (4.30) and (4.38)). The observed posterior modes of ϕ_1 , ϕ_2 and ϕ_3 obtained by applying MCMC for group A were 0.532, -0.445 and -0.343 , respectively, and the observed posterior modes of ϕ_1 , ϕ_2 and ϕ_3 of group B were 0.191, 0.177 and 0.603, respectively. Therefore, group A priors are to a large extent satisfactory. As can be seen, the estimation of parameters obtained from group A is, to some extent, more precise than those obtained from group B see Table 4.7. We applied the Bayes factor in a way that group A plays the role of the null model (M_o) and group B plays the role of the alternative model (M_1) as can be seen in Table (4.6). This procedure has been repeated for many different parameters in order to be more confident about the performance of the Bayes factor. The conclusion is that the null model using the prior distribution of group AB cannot be rejected against the alternative models using groups A and B, respectively. Therefore, the prior distribution of group AB conditions can be used as an alternative prior for both group conditions A and B in order to estimate the parameters of the AR(3) model, as shown in Table 4.6 and 4.7. In conclusion, the prior distribution of AB captures much of the uncertainty and hence we can obtain precise estimation.

Table 4.9: Pseudo code of the MCMC procedure for the AR(2) model.

Update 1: ϕ_1 initiate values: $\phi_1^{(0)}, \phi_2^{(0)}$ and $\sigma^{2(0)}$ $\phi_1^{(new)}$: sample from the truncated normal distribution of equation(4.15) $\phi_1^{(new)}$: generated based on $\phi_2^{(old)}$ and $\sigma^{2(old)}$
Update 2: ϕ_2 Set a proposal distribution $q(\phi_2^{(new)}\|\phi_2^{(old)})$ as in equation (4.18) $\phi_2^{(new)}$: sample from $q(\phi_2^{(new)}\|\phi_2^{(old)})$ u: sample from U(0,1) $\alpha = min\left[\frac{p(\phi_2^{(new)})q(\phi_2^{(old)}\|\phi_2^{(new)})}{p(\phi_2^{(old)})q(\phi_2^{(new)}\|\phi_2^{(old)})}, 1\right]$. If $\alpha \geq u$, then $\phi_2 = \phi_2^{new}$; otherwise $\phi_2 = \phi_2^{old}$.
Update 3: $\sigma^{2(new)}$: sample the precision$(\frac{1}{\sigma^2})^{(new)}$ from the Gamma distribution of equation (4.17) $\sigma^{2(new)}$: generated based on $\phi_1^{(new)}, \phi_2^{(new)}, a = 3$ and $b = 10$

Table 4.10: Pseudo code of the MCMC procedure for the AR(3) model.

Update 1: ϕ_1 initiate values: $\phi_1^{(0)}, \phi_2^{(0)}, \phi_3^{(0)}$. and $\sigma^{2(0)}$ $\phi_1^{(new)}$: sample from the truncated normal distribution of equation(4.49). $\boldsymbol{\phi_1^{(new)}}$: generated based on $\phi_2^{(old)}$, $\phi_3^{(old)}$ and $\sigma^{2(old)}$.
Update 2: ϕ_2 Set a proposal distribution $q(\phi_2^{(new)}\|\phi_2^{(old)})$ as in equation (4.59). $\phi_2^{(new)}$: sample from $q(\phi_2^{(new)}\|\phi_2^{(old)})$. u: sample from U(0,1) $\alpha = min\left[\frac{p(\phi_2^{(new)})q(\phi_2^{(old)}\|\phi_2^{(new)})}{p(\phi_2^{(old)})q(\phi_2^{(new)}\|\phi_2^{(old)})}, 1\right]$. If $\alpha \geq u$, then $\phi_2 = \phi_2^{new}$; otherwise $\phi_2 = \phi_2^{(old)}$.
Update 3: ϕ_3 Set a proposal distribution $q(\phi_3^{(new)}\|\phi_3^{(old)})$ as in equation (4.60). $\phi_3^{(new)}$: sample from $q(\phi_3^{(new)}\|\phi_3^{(old)})$. u: from U(0,1) $\alpha = min\left[\frac{p(\phi_3^{(new)})q(\phi_3^{(old)}\|\phi_3^{(new)})}{p(\phi_3^{(old)})q(\phi_3^{(new)}\|\phi_3^{(old)})}, 1\right]$. If $\alpha \geq u$, then $\phi_3 = \phi_3^{(new)}$; otherwise $\phi_3 = \phi_3^{(old)}$.
Update 4: $\sigma^{2(new)}$ $\boldsymbol{\sigma^{2(new)}}$: sample the precision$(\frac{1}{\sigma^2})^{(new)}$ from the Gamma distribution of equation (4.58). $\boldsymbol{\sigma^{2(new)}}$: generated based on $\phi_1^{(new)}, \phi_2^{(new)}, \phi_3^{(new)}, a = 3$ and $b = 10$.

4.16 Prior distribution of the AR(4) model

In order to derive a prior distribution for the AR(4) model, we need to derive the equation of the prior distribution as follows:

$$p(\phi_1, \phi_2, \phi_3, \phi_4) \propto p(\phi_4) \cdot p(\phi_3|\phi_4) \cdot p(\phi_2|\phi_3, \phi_4) \cdot p(\phi_1|\phi_2, \phi_3, \phi_4) \tag{4.61}$$

In order to do this, we need to have all of the conditions related to the AR(4) model and from them we are able to find the prior distributions for each parameter. Therefore, we can take advantage of Section 3.6.4 where the conditions of the AR(4) model were found which are shown below:

$$\phi_1 + \phi_2 + \phi_3 + \phi_4 < 1 \tag{4.62}$$

$$-\phi_1 - \phi_2 + \phi_3 - \phi_4 < 1 \tag{4.63}$$

$$-\phi_1 - \phi_2 - \phi_3 + \phi_4 < 1 \tag{4.64}$$

$$\phi_1 + \phi_2 - \phi_3 - \phi_4 < 1 \tag{4.65}$$

$$\phi_1 - \phi_2 - \phi_3 - \phi_4 < 1 \tag{4.66}$$

$$-\phi_1 + \phi_2 - \phi_3 + \phi_4 < 1 \tag{4.67}$$

$$-\phi_1 + \phi_2 + \phi_3 - \phi_4 < 1 \tag{4.68}$$

$$\phi_1 - \phi_2 + \phi_3 + \phi_4 < 1 \tag{4.69}$$

$$|\phi_4| < 1. \tag{4.70}$$

It can be seen that there are nine inequalities and they should satisfy the stationarity conditions. Due to the fact that ϕ_4 is bounded, we easily deduce that an appropriate prior distribution for ϕ_4 is the uniform distribution on (-1, 1). Thus, from (4.70) it can be said that the prior distribution of ϕ_4 can be as follows:

$$p(\phi_4) \propto \frac{1}{2} \tag{4.71}$$

Similarly, finding the prior distribution of $\phi_3|\phi_4$ can be done several ways. First, by adding (4.66) to (4.67) we can obtain $-2\phi_3 < 2 \Rightarrow \phi_3 > -1$, and by adding (4.62) to (4.63), we can get $2\phi_3 < 2 \Rightarrow \phi_3 < 1$. We can say that ϕ_3 is independent of ϕ_4 so $p(\phi_3|\phi_4) = p(\phi_3)$. Therefore, $|\phi_3| < 1$ so an appropriate prior for ϕ_3 is $\phi_3 \sim U(-1, 1)$.

As mentioned above, there are several ways to do what was done above. The following is another way to derive the prior for ϕ_3 by adding (4.64) to (4.67), we can obtain

$$2\phi_1 - 2\phi_3 + 2\phi_4 < 2 \tag{4.72}$$

and by adding (4.65) to (4.66), we can obtain

$$+2\phi_1 - 2\phi_3 - 2\phi_4 < 2 \tag{4.73}$$

Therefore, with regards to the condition $\phi_3 > -1$, adding equation (4.72) to equation (4.73), we can get that $-4\phi_3 < 4 \Rightarrow \phi_3 > -1$. In order to prove that $\phi_3 < 1$, equation (4.62) can be added to equation (4.69) as follows:

$$2\phi_1 + 2\phi_3 + 2\phi_4 < 2 \tag{4.74}$$

and by adding (4.63) to (4.68) we can obtain

$$-2\phi_1 + 2\phi_3 - 2\phi_4 < 2 \tag{4.75}$$

Hence, adding (4.74) to (4.75), we can get $4\phi_3 < 4$ and thus $\phi_3 < 1$. Consequently we end up with the same result as we had from the previous method and it can now be said that ϕ_3 is independent of ϕ_4. Thus, the prior distribution of ϕ_3 is $p(\phi_3|\phi_4) \propto p(\phi_3) = \frac{1}{2}$. Thus,

$$p(\phi_3|\phi_4) \propto \frac{1}{2} \tag{4.76}$$

Moreover, having the prior distributions of $\phi_3|\phi_4$ and ϕ_4, it would be easier to derive the prior distribution of ϕ_2 by adding equations (4.62) and (4.68) together as follows:

$$2\phi_2 + 2\phi_3 < 2 \tag{4.77}$$

Next, by adding (4.62) and (4.67) we can get:

$$2\phi_2 + 2\phi_4 < 2 \tag{4.78}$$

From equations (4.77) and (4.78) we can obtain another equation which is:

$$4\phi_2 + 2\phi_3 + 2\phi_4 < 4$$

Then, by putting ϕ_2 on the left hand-side alone and taking all other terms to the other side:

$$\phi_2 < 1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4 \tag{4.79}$$

After that (4.63) and (4.66) are added to create (4.80) as follows:

$$-2\phi_2 - 2\phi_4 < 2 \tag{4.80}$$

and then (4.64) is added to (4.66) to obtain the following equation

$$-2\phi_2 - 2\phi_3 < 2 \tag{4.81}$$

Thus, another new equation is derived from combining (4.80) and (4.81), it is shown below:

$$-4\phi_2 - 2\phi_3 - 2\phi_4 < 4$$

and then

$$\phi_2 > -1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4 \tag{4.82}$$

It can be seen that all of these equations come from the nine conditions of the AR(4) model, therefore, in order to get finalize the prior distribution of ϕ_2, another equation is obtained by combining (4.79) and (4.82) as follows:

$$-1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4 < \phi_2 < 1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4 \tag{4.83}$$

Therefore, the prior distribution of ϕ_2 given ϕ_3 and ϕ_4 is as follows:

$$p(\phi_2|\phi_3, \phi_4) = \frac{1}{1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4 - (-1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4)}$$

$$p(\phi_2|\phi_3, \phi_4) = \frac{1}{2} \tag{4.84}$$

Hence the prior distribution of $p(\phi_2|\phi_3, \phi_4)$ is as follows:

$$\phi_2|\phi_3, \phi_4 \sim U(-1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4, 1 - \frac{1}{2}\phi_3 - \frac{1}{2}\phi_4) = U(-1, 1)$$

Now, we have nearly finished as there is only the prior distribution of ϕ_1 left to be derived.

From (4.62) to (4.69) we obtain inequalities for ϕ_1 as follows:

$$\phi_1 < 1 - \phi_2 - \phi_3 - \phi_4 \tag{4.85}$$

$$\phi_1 > -1 - \phi_2 + \phi_3 - \phi_4 \tag{4.86}$$

$$\phi_1 > -1 - \phi_2 - \phi_3 + \phi_4 \tag{4.87}$$

$$\phi_1 < 1 - \phi_2 + \phi_3 + \phi_4 \tag{4.88}$$

$$\phi_1 < 1 + \phi_2 + \phi_3 + \phi_4 \tag{4.89}$$

$$\phi_1 > -1 + \phi_2 - \phi_3 + \phi_4 \tag{4.90}$$

$$\phi_1 > -1 + \phi_2 + \phi_3 - \phi_4 \tag{4.91}$$

$$\phi_1 < 1 + \phi_2 - \phi_3 - \phi_4 \tag{4.92}$$

Consequently, in order to assess a range for ϕ_1 we have to look at those inequalities that are quite similar which are (4.89) and (4.92) as they both have the less than sign. So that

$$\phi_1 < 1 + \phi_2 - |\phi_3 - \phi_4| \tag{4.93}$$

If we take the absolute value of $\phi_3 + \phi_4$, we can also obtain (4.85) to (4.88). Furthermore, from equations (4.85) and (4.88), we can write them as one inequality which is,

$$\phi_1 < 1 - \phi_2 - |\phi_3 + \phi_4| \tag{4.94}$$

Again, by using the absolute value we can get the same inequality, i.e., from (4.93) and (4.94) we can obtain the following equation:

$$\phi_1 < 1 - |\phi_2| - |\phi_3 + \phi_4| \tag{4.95}$$

It can be noted that we have only worked on the inequalities that have the less than sign ($<$). In order to compute the left- hand side for ϕ_1, it is important to consider those inequalities that have the greater than sign ($>$). Taking this into the consideration, we combine (4.86) and

(4.87), and obtain:

$$\phi_1 > -1 - \phi_2 + |\phi_3 - \phi_4| \tag{4.96}$$

Likewise, we need to do the same procedure on (4.90) and (4.91) as shown below:

$$\phi_1 > -1 + \phi_2 + |\phi_3 + \phi_4| \tag{4.97}$$

Therefore, the left-hand side of ϕ_1 can be written by merging (4.96) and (4.97) as follows:

$$\phi_1 > -1 + |\phi_2| + |\phi_3 - \phi_4| \tag{4.98}$$

As a result, it can be seen that now we have both sides or one can say both limits, lower and upper, by combining (4.95) and (4.98) as written down below

$$-1 + |\phi_2| + |\phi_3 - \phi_4| < \phi_1 < 1 - |\phi_2| - |\phi_3 + \phi_4| \tag{4.99}$$

It is now known that $p(\phi_1|\phi_2, \phi_3, \phi_4)$ can be as follows:

$$p(\phi_1|\phi_2, \phi_3, \phi_4) = \frac{1}{2 - 2|\phi_2| - |\phi_3 + \phi_4| - |\phi_3 - \phi_4|} \tag{4.100}$$

Now, the prior distribution of $p(\phi_1, \phi_2, \phi_3, \phi_4)$ can be written as follows:

$$\begin{aligned} p(\phi_1, \phi_2, \phi_3, \phi_4) &= p(\phi_4) \cdot p(\phi_3|\phi_4) \cdot p(\phi_2|\phi_3, \phi_4) \cdot p(\phi_1|\phi_2, \phi_3, \phi_4) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2 - 2|\phi_2| - |\phi_3 + \phi_4| - |\phi_3 - \phi_4|} \\ &\propto \frac{1}{2 - 2|\phi_2| - |\phi_3 + \phi_4| - |\phi_3 - \phi_4|} \end{aligned} \tag{4.101}$$

It has been shown how prior distributions for the AR(4) model can be derived. These prior distributions can be used in the next section in order to find the full conditional posterior distributions for the AR(4) model.

4.17 Posterior distribution for the AR(4) model

Assume n observations are available, say $y_1, y_2, \ldots, y_n$, for the AR(4) model.

$$y_t = \phi_1 y_{t-1} - \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t,$$

where ε_t is white noise and $\varepsilon_t \sim N(0, \sigma^2)$. The unknown parameters here are ϕ_1, ϕ_2, ϕ_3, ϕ_4 and σ^2. Similar to the steps followed in the AR(2) and AR(3) models, full conditional distributions can be derived for $\phi_{(j)}|\phi_{(-j)}, y$ for $j = 1, , \ldots, p$. Thus

$$\begin{aligned}
p(\phi_1|\phi_2,\phi_3,\phi_4,y,\sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3}-\phi_4 y_{t-4})^2} p(\phi_1|\phi_2,\phi_3,\phi_4) \\
&= e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3}-\phi_4 y_{t-4})^2} . \frac{1}{2-2|\phi_2|-|\phi_3+\phi_4|-|\phi_3-\phi_4|} \\
&\propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3}-\phi_4 y_{t-4})^2} I_{[a]}.
\end{aligned}$$

where

$$I_{[a]} = \begin{cases} 1 & a \in A \\ 0 & a \notin A \end{cases}$$

for A={ $\phi_i \in$ that satisfy the conditions of the AR(4) model}. Let the variable of interest be denoted by $z_t = y_t - \phi_{(-j)} y_{t-(-j)}$ for $(j = 1, , \ldots, 4)$ in order to be able to expand the part of $\sum(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \phi_4 y_{t-4})^2$ which is inside the likelihood part of the posterior distribution. This means that when we are trying to define the posterior distribution for ϕ_1 $z_t = y_t - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \phi_4 y_{t-4}$. Therefore, the posterior distribution can be written as follows;

$$p(\phi_1|\phi_2,\phi_3,\phi_4,y,\sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum(z_t-\phi_1 y_{t-1})^2} I_{[a]}.$$

Then, by following the steps of equation (4.2) and applying it to the above equation, the posterior distribution of $\phi_1|\phi_2,\phi_3,\phi_4,y,\sigma^2$ is:

$$\phi_1|\phi_2,\phi_3,\phi_4,y,\sigma^2 \sim N_{[-a1,a1]}\left(\frac{\sum z_t y_{t-1}}{\sum y_{t-1}^2}, \frac{\sigma^2}{\sum y_{t-1}^2}\right)$$

where [-a1,a1] is the range of ϕ_1 , see equation (4.99). By adding $z_t = y_t - \phi_2 y_{t-2} - \phi_3 y_{t-3} - \phi_4 y_{t-4}$ to the above equation, the following equation can be obtained

$$\phi_1|\phi_2,\phi_3,\phi_4,y \sim N_{[-a1,a1]}\left(\frac{\sum y_t y_{t-1} - \phi_2\sum y_{t-1}y_{t-2} - \phi_3\sum y_{t-1}y_{t-3} - \phi_4\sum y_{t-1}y_{t-4}}{\sum y_{t-1}^2}, \frac{\sigma^2}{\sum y_{t-1}^2}\right) \tag{4.102}$$

In order to derive the conditional posterior distribution for $\phi_2|\phi_1,\phi_3,\phi_4,y,\sigma^2$, the procedure is:

$$\begin{aligned} p(\phi_2|\phi_1,\phi_3,\phi_4,y,\sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3}-\phi_4 y_{t-4})^2} p(\phi_2|\phi_1,\phi_3,\phi_4) \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(z_t-\phi_2 y_{t-2})^2} . \frac{1}{2-2|\phi_2|-|\phi_3+\phi_4|-|\phi_3-\phi_4|} I_{[-1,1]} \end{aligned} \tag{4.103}$$

where $z_t = y_t - \phi_1 y_{t-1} - \phi_3 y_{t-3} - \phi_4 y_{t-4}$, taking the same steps as before for $\phi_1|\phi_2,\phi_3,\phi_4,y,\sigma^2$. It is noted that the prior distribution of $\phi_2|\phi_1,\phi_3,\phi_4$ is used to derive the posterior distribution $\phi_2 \mid \phi_1,\phi_3,\phi_4,y,\sigma^2$ according to equation (4.103). However, we cannot put the posterior distribution into any form of standard distribution, so we cannot directly simulate from it. Regarding the conditional posterior distributions of $\phi_3 \mid \phi_1,\phi_2,\phi_4,y,\sigma^2$ and $\phi_4 \mid \phi_1,\phi_2,\phi_3,y,\sigma^2$, the procedure is similar, i.e.,

$$\begin{aligned} p(\phi_3|\phi_1,\phi_2,\phi_4,y,\sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3}-\phi_4 y_{t-4})^2} p(\phi_3 \mid \phi_1,\phi_2,\phi_4) \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(z_t-\phi_3 y_{t-3})^2} . \frac{1}{2-2|\phi_2|-|\phi_3+\phi_4|-|\phi_3-\phi_4|} I_{[-1,1]} \end{aligned} \tag{4.104}$$

Taking the same steps as above, where $z_t = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_4 y_{t-4}$,

$$p(\phi_4|\phi_1,\phi_2,\phi_3,y,\sigma^2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum(z_t-\phi_4 y_{t-4})^2} \cdot \frac{1}{2-2|\phi_2|-|\phi_3+\phi_4|-|\phi_3-\phi_4|} I_{[-1,1]}, \tag{4.105}$$

where $z_t = y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \phi_3 y_{t-3}$. It can be stated that the posterior distribution for $\phi_{(j)}|\phi_{(-j)}, y$ can be represented by a truncated normal distribution when $j = 1$. However, when $j > 1$, the posterior distribution is unknown. Therefore, for $j > 1$, a Metropolis step should be used where the proposed is a truncated normal distribution.

4.18 MCMC application for the AR(4) model

After realizing, as previously mentioned, that using a random walk as a proposal for the posterior distribution was unable to accurately estimate parameters of the AR(2) and AR(3) models (see Sections 4.7 and 4.11), we abandon the idea of using random walk proposal to estimate parameters of the AR(4) model. Therefore the new recommended proposals are used in order to obtain parameter estimates. Again, the idea of using the new proposal is to assume that ϕ_2 , ϕ_3 and ϕ_4 have a truncated normal distributions. It can be noted that Gibbs sampling is used to estimate ϕ_1 from the conditional distribution of equation (4.102), and the parameters ϕ_2 , ϕ_3 and ϕ_4 with posterior distributions of (4.103), (4.104) and (4.105), respectively, are estimated through Metropolis steps. Additionally, the new recommended proposals of ϕ_2 , ϕ_3 and ϕ_4 are as follows:

$$\phi_2 \mid \phi_1,\phi_3,\phi_4,y,\sigma^2 \sim N_{[-1,1]}\left(\frac{\sum y_t y_{t-2} - \phi_1\sum y_{t-1}y_{t-2} - \phi_3\sum y_{t-2}y_{t-3} - \phi_4\sum y_{t-2}y_{t-4}}{\sum y_{t-2}^2}, \frac{\sigma^2}{\sum y_{t-2}^2}\right) \tag{4.106}$$

$$\phi_3 \mid \phi_1, \phi_2, \phi_4, y, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum y_t y_{t-3} - \phi_1 \sum y_{t-1}y_{t-3} - \phi_2 \sum y_{t-2}y_{t-3} - \phi_4 \sum y_{t-3}y_{t-4}}{\sum y_{t-3}^2}, \frac{\sigma^2}{\sum y_{t-3}^2}\right) \tag{4.107}$$

$$\phi_4 \mid \phi_1, \phi_2, \phi_3, y, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum y_t y_{t-4} - \phi_1 \sum y_{t-1}y_{t-4} - \phi_2 \sum y_{t-2}y_{t-4} - \phi_3 \sum y_{t-3}y_{t-4}}{\sum y_{t-4}^2}, \frac{\sigma^2}{\sum y_{t-4}^2}\right) \tag{4.108}$$

To estimate parameters of the AR(4) model, 150 observations are simulated for each of the

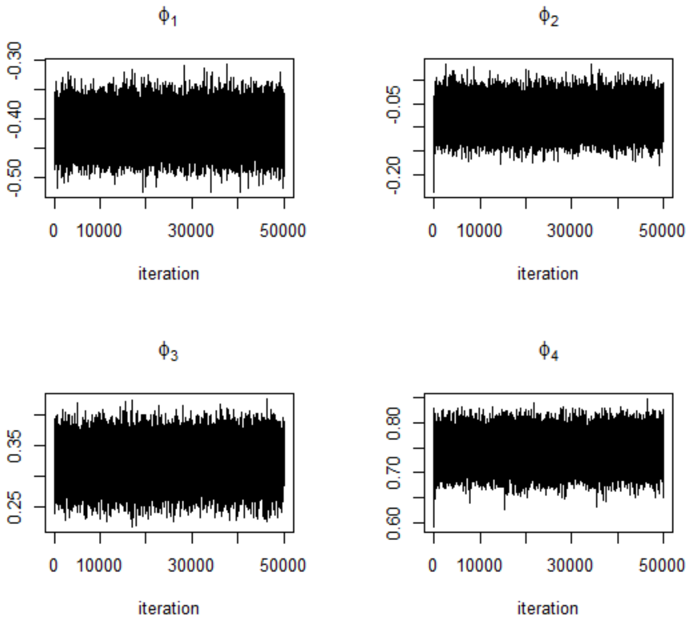


Figure 4.12: Illustration of convergence for the parameter estimates of ϕ_1, ϕ_2, ϕ_3 and ϕ_4 via MCMC of the AR(4) model with K=50000, $\phi_1 = -0.44$, $\phi_2 = -0.05$, $\phi_3 = 0.33$, $\phi_4 = 0.77$, $\alpha = 3$ and $\beta = 10$.

data sets after obtaining valid parameters of the AR(4) model. These parameters are obtained based on the corresponding relationship between partial correlations (π) and parameters (ϕ) as mentioned in Barndorff-Nielsen and Schou (1973)'s study. It is then guaranteed that the parameters have a stationary process as described in Section 4.3. We use an inverse gamma prior for σ^2 as before. Once again we choose a weakly informative prior, $\frac{1}{\sigma^2} \sim Gamma(a, b)$

assuming a=3 and b=10 (Ando, 2010).

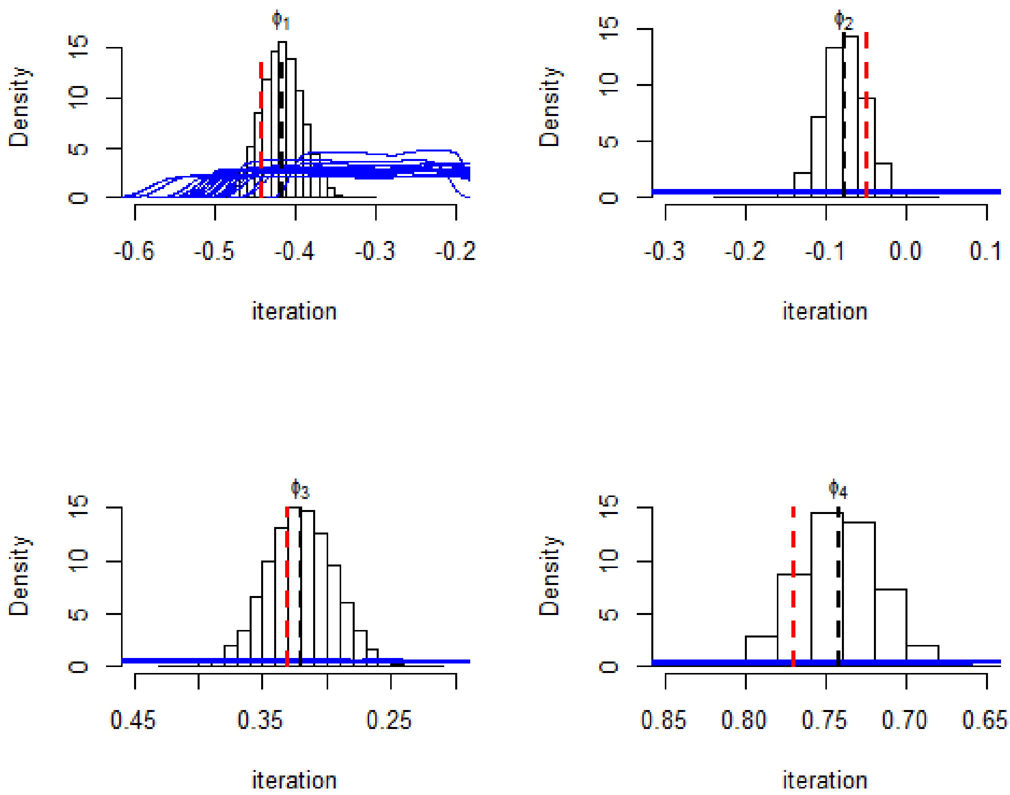


Figure 4.13: Shows simulated posterior distributions for the parameters of the AR(4) model with K=50000, $\phi_1 = -0.44$, $\phi_2 = -0.05$, $\phi_3 = 0.33$, $\phi_4 = 0.77$, $\alpha = 3$ and $\beta = 10$.

After applying MCMC, the results of the parameter estimation presented are represented in Table 4.11. To illustrate this table and to understand the MCMC results of the AR(4) model, the parameters of $\phi_1 = -0.44$, $\phi_2 = -0.05$, $\phi_3 = 0.33$ and $\phi_4 = 0.77$ are used as an example after simulating 150 observations based on the aforementioned parameters. Visual inspection of the time series plot produced by 'history' in Figure 4.12 illustrates that the MCMC has converged. It can be noted that the MCMC results of the parameters are $\phi_1 = -0.42$, $\phi_2 = -0.08$, $\phi_3 = 0.32$ and $\phi_4 = 0.74$, and errors between the true parameters and the estimated parameters are within 1 to 3%. It can be seen from Figure 4.13 that the acceptance rates are relatively high. This is

Table 4.11: MCMC results for the AR(4) model

Trial	Stuation of parameters	Iteration	ϕ_1 (Sd)	ϕ_2(Sd)	ϕ_3 (Sd)	ϕ_4 (Sd)	Acceptance rate
1	Real value	20000	-0.1	0.6	0.2	0.2	63.69%
	Estimate Value		-0.102(0.232)	0.635(0.12)	0.207(0.183)	0.236(0.124)	
2	Real value	100000	-0.1	0.6	0.2	0.2	64.03%
	Estimate Value		-0.0971(0.23)	0.638(0.12)	0.202(0.183)	0.234(0.123)	
3	Real value	40000	0.727	-0.325	-0.496	0.575	58.43%
	Estimate Value		0.707(0.374)	-0.395(0.151)	-0.49(0.229)	0.509(0.292)	
4	Real value	40000	-0.4	0.2	0.4	0.4	92.37%
	Estimate Value		-0.517(0.117)	0.2(0.127)	0.493(0.124)	0.367(0.118)	
5	Real value	50000	-0.44	-0.05	0.33	0.77	94.23%
	Estimate Value		-0.417(0.0258)	-0.0774(0.0262)	0.321(0.0263)	0.742(0.0257)	

because our proposals are close to the corresponding posterior distributions, see Hoff (2009) and Robert and Casella (2009). Additionally, general formula will be derived in the next section in order to generalize our prior and posterior distributions to AR(p) models of any order.

4.19 Generalized posterior distribution for the AR(p) model

It has been clearly mentioned that the aim of the current research is estimation of the parameters of the autoregressive model through using MCMC methods. The estimation process is conducted via AR model parameters, thus, stationary conditions, derived by parameter models in a conditional inequality way, should be taken into account (see Chapter 3). Therefore, prior distribution obtained through stationary conditions, are used to develop a Metropolis within Gibbs MCMC scheme. In order to generalize to the prior distribution for the AR(p) model, first we should consider the results of the prior distributions of AR models for orders which are lower than 5 ($p < 5$) as seen in equations (4.11), (4.45) and (4.101).

$$\text{For } p = 1, \quad p(\phi_1) = \begin{cases} \frac{1}{2} & \phi_1 \in \text{SC of AR(1)} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{For } p = 2\,, \quad p(\phi_1, \phi_2) = \begin{cases} \frac{1}{4(1-\phi_2)} & \Phi \in \text{SC of AR}(2) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{For } p = 3\,, \quad p(\phi_1, \phi_2, \phi_3) = \begin{cases} \frac{1}{4(2-|\phi_2+\phi_3|-|\phi_2-\phi_3|)} & \Phi \in \text{SC of AR}(3) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{For } p = 4, \quad p(\phi_1, \phi_2, \phi_3, \phi_4) = \begin{cases} \frac{1}{8(2-|\phi_3+\phi_4|-|\phi_3-\phi_4|-2|\phi_2|)} & \Phi \in \text{SC of AR}(4) \\ 0 & \text{otherwise} \end{cases}$$

Regarding the prior distributions of AR(5) and AR(6), please see Appendix C in order to clarify how to obtain the below priors for both of them.

$$\text{For } p = 5, \quad p(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5) = \begin{cases} \frac{1}{16(2-2|\phi_2|-2|\phi_4|-|\phi_3+\phi_5|-|\phi_3-\phi_5|)} & \Phi \in \text{SC of AR}(5) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{For } p = 6, \quad p(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6) = \begin{cases} \frac{1}{32(2-2|\phi_2|-2|\phi_4|-2|\phi_5|-|\phi_3+\phi_6|-|\phi_3-\phi_6|)} & \Phi \in \text{SC of AR}(6) \\ 0 & \text{otherwise} \end{cases}$$

Based on the above summary, the prior distribution of the AR(p) model can be generalized, for $p > 4$ as follows:

$$p(\phi_1, \phi_2, \ldots, \phi_p) = \begin{cases} \frac{1}{2^{p-1}(2-|\phi_3+\phi_p|-|\phi_3-\phi_p|-2\sum_{i=2(i\mp 3)}^{p-1}|\phi_i|)} & \Phi \in \text{SC of AR}(p) \\ 0 & \text{otherwise} \end{cases} \tag{4.109}$$

the posterior distribution for the AR(p) model can be generalized based on the posterior distributions of the AR(p) models when $p < 4$, as in equations (4.3) for AR(1), (4.16) for AR(2) and (4.49)-(4.51) for AR(3) model. For the posterior distributions of the AR(4) model see equations (4.102) to (4.105). As a result of the summarization of the aforementioned posterior distributions we can generalize the full conditional posterior distributions for the AR(p) model as follows;

$$\phi_{(j)} \mid \phi_{(-j)}, y, \sigma^2 \sim N_{[-a,a]} \left(\frac{\sum y_t y_{t-(j)} - [\phi_{(-j)}]^T \sum y_{t-(-j)} y_{t-(j)}}{\sum y_{t-(j)}^2}, \frac{\sigma^2}{\sum y_{t-(j)}^2} \right) \tag{4.110}$$

where $j \in \{1, 2, \ldots, p\}$, $(-j)$ is $\{1, 2, \ldots, p\}$ with the element j removed and p is the order of the AR model. $[-a, a]$ are the lower and upper bounds of the truncated normal distribution where $a \in$ upper boundary values that satisfy stationary conditions of $\phi(j)$ $\dot{\text{F}}$or example, in the AR(2) model when $j = 1$, $a = 1 - \phi_2$. Each of the vectors of $\phi(-j)$ is defined as follows:

$$\phi_{(-j)} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_{(j)-1} \\ \phi_{(j)+1} \\ \vdots \\ \phi_p \end{bmatrix} \quad \text{and} \quad y_{t-(-j)} = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-j+1} \\ y_{t-j-1} \\ \vdots \\ y_{t-p} \end{bmatrix}$$

Chapter 5

Prior structures and comparative results

This chapter mainly focuses on two parts in order to study and compare the current proposal with previous studies relevant to the present study. The first part compares the proposed prior distribution with the prior distributions obtained from the correspondence relationship between partial autocorrelations and parameters discussed by Barndorff-Nielsen and Schou (1973). It discusses the study by Jones (1987) in which the author generalized a Jacobian transformation based on the expressions for the parameters in terms of partial autocorrelations. One of the limitations of Jones (1987)'s study is that we cannot obtain a prior distribution for the parameters using the Jacobian transformation in the case high order of polynomial models. This is discussed in this chapter. This comparison relies on some theoretical mathematical steps and practical results when applying these prior distributions to obtain parameter estimates of the AR(p) model. We extend the work of Barnett et al. (1996) who placed uniform priors on the partial autocorrelation and proposed a Metropolis Hastings algorithm. Considering the same priors, we develop a Gibbs sampling algorithm which is easier and more routine to apply. The purpose of the second part is to apply the proposed MCMC scheme of chapter 4 to both real data and simulated data. Furthermore, this part compares the performance of the above MCMC algorithm with Box et al. (1976) as well as with the Gibbs sampling scheme of the previous section.

5.1 Prior distribution based on Barndorff-Nielsen and Schou (1973)'s study

The focus here is on the differences amongst the current study and the Barndorff-Nielsen and Schou (1973) one in terms of identifying prior distributions of autoregressive models. The recent study shows the weakness of Barndorff-Nielsen and Schou (1973)'s study in that the relations between the parameters and characteristic roots cannot be relied on. The reason is that the Barndorff-Nielsen and Schou (1973) study is unable to identify a prior distribution for the AR(p) model, especially when the order of the model is high. For this purpose, identifying prior distributions is highlighted for some primary orders $(p < 3)$. However, it can be shown that a prior distribution cannot be identified when the order of the model is higher than three $(p > 3)$. This is illustrated in detail in the next section.

5.1.1 Prior distribution of the AR(p) model when $p < 3$

As previously mentioned, defining the prior distribution of the AR(p) model is difficult in terms of mathematical procedures when the model order is $p \geq 3$. First, the prior distribution is defined when the order is $p < 3$. This is done here to make a comparison between our suggested prior distribution and the aforementioned one. In order to identify the prior distribution for the AR(1), it can be shown that $p(\phi) = \frac{1}{2} I_{[-1,1]}$ which is based on $|\phi| < 1$ and $\phi \sim U(-1, 1)$. When we have a time series of order two, we use the following relationship between the partial autocorrelations (π) and parameters (ϕ) from Barndorff-Nielsen and Schou (1973) are as follows:

$$\phi_1 = \pi_1 - \pi_1\pi_2 \tag{5.1}$$

$$\phi_2 = \pi_2 \tag{5.2}$$

We obtain $\pi_1 = \frac{\phi_1}{1-\phi_2}$ and $\pi_2 = \phi_2$. The prior distribution for the AR(2) model can be identified through the Jacobian transformation as follows:

$$p(\phi_1, \phi_2) = p(\pi_1, \pi_2).|J| \tag{5.3}$$

Therefore, the Jacobian formula is

$$|J| = \begin{vmatrix} \frac{\partial \pi_1}{\partial \phi_1} & \frac{\partial \pi_1}{\partial \phi_2} \\ \frac{\partial \pi_2}{\partial \phi_1} & \frac{\partial \pi_2}{\partial \phi_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{1-\phi_2} & \frac{\phi_1}{(1-\phi_2)^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{1-\phi_2}.$$

Since the partial autocorrelations π_1, π_2 are on (-1, 1), we can propose that π_1, π_2 are uniformly distributed on (-1, 1) and are independent. Hence, the prior distribution for the AR(2) model is

$$p(\phi_1, \phi_2) = \begin{cases} \frac{1}{4(1-\phi_2)} & \text{if } \phi_2 - 1 < \phi_1 < 1 - \phi_2 \text{ and } -1 < \phi_2 < 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

We note that our proposed prior distribution for ϕ_1, ϕ_2 (Chapter 4) coincides with (5.4). This suggests that our proposal achieves the same prior distribution as equation (4.11), but it avoids placing the priors on the partial autocorrelations. We avoid that because we can not identify the relationship between parameters and autocorrelation of more than order four. And, our aim is to place priors directly on the parameters rather than placing priors on the characteristic roots.

When identifying the prior distribution for $p = 3$ in order to establish the correspondence between parameters and partial autocorrelations described by Barndorff-Nielsen and Schou (1973), we face a somehow more difficult and complex procedure. This is discussed in the following section.

5.1.2 Prior distribution of the AR(p) model when $p \geq 3$

This section shows that the procedure of defining a prior distribution, when using the relationship between partial autocorrelations and parameters in Barndorff-Nielsen and Schou (1973)'s study, faces some difficulties. This is because partial derivatives cannot be found easily when $p \geq 3$. In order to find the prior distribution for the AR(3) model, the mapping of partial autocorrelations π into parameters ϕ can be used from Section 3.6 as has been shown from equations of (3.70)-(3.72). It can be noticed that the range of π_i are between (-1, 1) then we have used the distribution of π_i is uniformly distributed, $\pi_i \sim U(-1,1)$, for $i = 1,2,3$ and $p(\pi_i) = \frac{1}{2}I_{[-1,1]}$. In fact, there are other alternative prior distributions that we could have used such as normal prior distribution. Therefore, the joint prior for the partial autocorrelations is as follows:

$$p(\pi_1, \pi_2, \pi_3) = p(\pi_1).p(\pi_2).p(\pi_3) = \left(\frac{1}{2}\right)^3 I_{[-1,1]} \tag{5.5}$$

To find priors for the AR(3) model, the concept of a derivative of a coordinate transformation can be explored which is known as the Jacobian transformation, therefore, the equation of $\{\pi(\phi_i)\}^{-1}$ can be transformed into an equation of $\phi(\pi_i)$ and finding the Jacobian as follows:

$$p(\phi_1, \phi_2, \phi_3) = p(\pi_1, \pi_2, \pi_3).|J| \tag{5.6}$$

Thus, the determinate of the Jacobian matrix is needed in order to find the prior for the AR(3) model. Therefore, the Jacobian of π_i with respect to ϕ_i is as follows:

$$|J| = \begin{vmatrix} \frac{\partial \pi_1}{\partial \phi_1} & \frac{\partial \pi_1}{\partial \phi_2} & \frac{\partial \pi_1}{\partial \phi_3} \\ \frac{\partial \pi_2}{\partial \phi_1} & \frac{\partial \pi_2}{\partial \phi_2} & \frac{\partial \pi_2}{\partial \phi_3} \\ \frac{\partial \pi_3}{\partial \phi_1} & \frac{\partial \pi_3}{\partial \phi_2} & \frac{\partial \pi_3}{\partial \phi_3} \end{vmatrix}$$

From the fact that $\pi_3 = \phi_3$, and $\frac{\partial \pi_3}{\partial \phi_1} = \frac{\partial \pi_3}{\partial \phi_2} = 0$ and $\frac{\partial \pi_3}{\partial \phi_3} = 1$, the above equation simplifies to $|J| = \begin{vmatrix} \frac{\partial \pi_1}{\partial \phi_1} & \frac{\partial \pi_1}{\partial \phi_2} \\ \frac{\partial \pi_2}{\partial \phi_1} & \frac{\partial \pi_2}{\partial \phi_2} \end{vmatrix}$.

The partial derivative $\frac{\partial \pi_1}{\partial \phi_1}$ can be calculated as follows:

$$\frac{\partial \pi_1}{\partial \phi_1} = \frac{1 - \phi_2 + \phi_2\phi_3^2 - \phi_3^2}{(1 - \phi_2 - \phi_1\phi_3 - \phi_3^2)^2} = \frac{(1 - \phi_2)(1 - \phi_3^2)}{(1 - \phi_2 - \phi_1\phi_3 - \phi_3^2)^2} \tag{5.7}$$

Additionally, the partial derivative $\frac{\partial \pi_1}{\partial \phi_2}$ can be calculated as follows:

$$\begin{aligned}\frac{\partial \pi_1}{\partial \phi_2} = \frac{-\phi_1(-1)}{(1 - \phi_2 - \phi_1\phi_3 - \phi_3^2)^2} + \frac{\phi_3(1 - \phi_2 - \phi_1\phi_3 - \phi_3^2) - \phi_2\phi_3(-1)}{(1 - \phi_2 - \phi_1\phi_3 - \phi_3^2)^2} \\ = \frac{(\phi_1 + \phi_3)(1 - \phi_3^2)}{(1 - \phi_2 - \phi_1\phi_3 - \phi_3^2)^2}\end{aligned} \tag{5.8}$$

the partial derivative $\frac{\partial \pi_2}{\partial \phi_1}$ is:

$$\frac{\partial \pi_2}{\partial \phi_1} = \frac{\phi_3}{1 - \phi_3^2} \tag{5.9}$$

and the parial derivative $\frac{\partial \pi_2}{\partial \phi_2}$ can be calculated as follows:

$$\frac{\partial \pi_2}{\partial \phi_2} = \frac{1}{1 - \phi_3^2} \tag{5.10}$$

Hence, from equations (5.7), (5.8), (5.9) and (5.10), the Jacobian can be obtained as follow;

$$|J| = \begin{vmatrix} \frac{(1-\phi_2)(1-\phi_3^2)}{(1-\phi_2-\phi_1\phi_3-\phi_3^2)^2} & \frac{(\phi_1+\phi_3)(1-\phi_3^2)^2}{(1-\phi_2-\phi_1\phi_3-\phi_3^2)} \\ \frac{\phi_3}{1-\phi_3^2} & \frac{1}{1-\phi_3^2} \end{vmatrix}$$

Thus, the determinate of the Jacobian matrix is as follows:

$$|J| = \frac{1}{1 - \phi_2 - \phi_1\phi_3 - \phi_3^2} \tag{5.11}$$

After the Jacobian has been found from the above steps, now, from equations (5.5) and (5.11) the prior distribution for the AR(3) model can be written as follows:

$$p(\phi_1, \phi_2, \phi_3) = \begin{cases} \frac{1}{2^3(1-\phi_2-\phi_1\phi_3-\phi_3^2)} & \text{if } |\phi_2 - \phi_3| - 1 < \phi_1 < 1 - |\phi_2 + \phi_3|,\ |\phi_2| < 1 \text{ and } |\phi_3| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.12}$$

It can be seen from the above steps that the prior distribution for the AR(3) model is obtained using the Jacobian transformation. But if, in the same way, the prior distribution is to be identified for the AR(p) model when $p = 4$, it cannot be obtained the same way as has been done for orders $p = 2$ and $p = 3$. This is because when $p > 3$, partial derivatives between characteristic roots and parameters cannot be found or obtained. This is due to the complicated relationship that can be seen from the mapping of characteristic roots to parameters as obtained in equations (5.7) and (5.8). In addition, procedures for calculating the Jacobian determinant for matrices of size $r \times r$ when $r \geq 3$, are mathematically complicated. Using Maple 18 software, we attempted to obtain the determinant of the 4×4 Jacobian matrix, but because of the complexity of the process, the result was not achieved. Thus, it can be said that the information described by Barndorff-Nielsen and Schou (1973) on the AR model cannot be relied upon when we estimate parameters of the AR(p) model and apply MCMC. This is when parameters are directly estimated via parameters rather than characteristic roots. The proposed prior distribution of Chapter 4 does not face these difficulties.

5.2 Prior distribution for the AR(p) model based on Jones (1987)

In the previous section, we noted the difficulties of using the Jacobian transformation in obtaining a prior distribution for the AR(p) model. But Jones (1987) has derived a Jacobian transformation based on the correspondence between parameters and partial autocorrelations obtained from Barndorff-Nielsen and Schou (1973)'s study.

$$|J| = \prod_{k=2}^{p} (1-\pi_k)^{[\frac{k}{2}]}(1+\pi_k)^{[\frac{k-1}{2}]} \tag{5.13}$$

Thus, when a prior distribution is defined for the AR(p) model through $J(\pi \to \phi)$ then

$$p(\phi_1, \ldots, \phi_p) = p(\pi_1, \ldots, \pi_p)\, |J|^{-1}$$

As we know $\pi_i \sim U(-1, 1)$ when $i = 1, 2, \ldots, p$. Therefore, $p(\pi_1 \ldots, \pi_p) = p(\pi_1) \ldots p(\pi_p)$ based on the π 's being independent. Thus, when a prior distribution is defined for the AR(2) model, we are able to use the Jacobian transformation proposed by Jones (1987) as follows

$$\begin{aligned} p(\phi_1, \phi_2) &= p(\pi_1, \pi_2)\, |J|^{-1} \\ \text{where} \quad |J| &= (1-\pi_2)^{[1]}(1+\pi_2)^{[\frac{1}{2}]=1-\pi_2} \\ &= 1 - \phi_2 \end{aligned}$$

Then the prior distribution of the AR(2) parameters is

$$p(\phi_1, \phi_2) = \begin{cases} \frac{1}{4(1-\phi_2)} & -1 < \phi_1 < 1 \quad -1 < \phi_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

When a prior distribution is defined for the AR(3) model, it can be derived in the same way as for the AR(2).

$$p(\phi_1, \phi_2, \phi_3) = p(\pi_1, \pi_2, \pi_3)\, |J|^{-1} = \frac{1}{2^3} . \, |J|^{-1}$$

$|J|$ can be obtained through equation (5.13), as follows

$$\begin{aligned} |J| &= (1-\pi_2)(1+\pi_2)^{[\frac{1}{2}]}.(1-\pi_3)^{[\frac{3}{2}]}(1+\pi_3) \\ &= (1-\pi_2).(1-\pi_3).(1+\pi_3) \\ &= (1-\pi_2).(1-\pi_3^2) \end{aligned}$$

By substituting into the above equation the expressions for the partial autocorrelations π_2 and π_3 in terms of the parameters ϕ_1, ϕ_2 and ϕ_3 (as obtained in equation (3.70) and (3.71)), we obtain

$$|J| = \left(1 - \frac{\phi_2 + \phi_1\phi_3}{(1-\phi_3^2)}\right).(1-\phi_3^2) = 1 - \phi_2 - \phi_1\phi_3 - \phi_3^2$$

The prior distribution for the AR(3) model is this

$$p(\phi_1, \phi_2, \phi_3) = \begin{cases} \frac{1}{8(1-\phi_2-\phi_1\phi_3-\phi_3^2)} & \Phi \in \text{ SC of AR(3)} \\ 0 & \textit{otherwise} \end{cases}$$

Regarding the AR(4) model, the prior distribution cannot be obtained in the same way as it was obtained for the AR(2) and AR(3). In general, there is no explicit relationship of $|J|$ in terms of the ϕ's. For $p = 4$, the Jacobian is

$$\begin{aligned} |J| &= (1-\pi_2)(1+\pi_2)^{[\frac{1}{2}]}.(1-\pi_3)^{[\frac{3}{2}]}(1+\pi_3).(1-\pi_4)^2(1+\pi_4)^{[\frac{3}{2}]} \\ &= (1-\pi_2)(1-\pi_3^2).(1-\pi_4)^2(1+\pi_4) \end{aligned}$$

and by using equations (3.88) to (3.90), we can obtain an expression of $|J|$ with respect to the ϕ's. However, due to (3.88) this will be too complicated. Thus, as mentioned previously, we cannot obtain prior distributions based on the mapping between partial autocorrelations and parameters. Likewise, we cannot rely on the Jones (1987) method for defining prior distribution for the AR(p) model. Our proposed a prior structure offers a general framework, which is considerably simpler.

5.3 Gibbs Sampler for the AR(2) model using partial autocorrelations

The aim of this section is to derive full conditional posterior distributions for the AR(2) model based on the corresponding relationship between the ϕ 's and the π 's. The parameters of the AR(2) model ϕ_1 and ϕ_2 are converted to partial autocorrelations, π_1 and π_2 based on the equations provided by Barndorff-Nielsen and Schou (1973). Here, we attempt to apply MCMC through the Gibbs sampler in order to estimate parameters of the AR model. We extend the work of Barnett et al. (1996) who placed uniform priors on the partial autocorrelations and proposed a Metropolis-Hastings algorithm. Considering the same priors, we develop a Gibbs sampling algorithm which is easier and more routine to apply. Assume n observations are available, say $y_1, y_2, \ldots, y_n$. The aim is to estimate ϕ_1, ϕ_2 and σ^2 . We use the AR(2) model $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$ where ε_t is white noise and $\varepsilon_t \sim N(0, \sigma^2)$. The unknown parameters here are ϕ_1, ϕ_2 and σ^2 . The posterior distribution for the AR(2) model using partial autocorrelations is given by:

$$p(\phi_1, \phi_2 \mid y, \sigma^2) \quad \propto \quad p(y \mid \phi_1, \phi_2, \sigma^2).p(\phi_1, \phi_2)$$

It has been known from the corresponding relationship between the ϕ 's and the π 's based on equations provided by Barndorff-Nielsen and Schou (1973) that:

$$\phi_1 = \pi_1(1 - \pi_2) \tag{5.14}$$

$$\phi_2 = \pi_2 \tag{5.15}$$

When both π_1 and π_2 are uniformly distributed between -1 and 1 (i.e., $\pi_1, \pi_2 \sim U[-1,1]$), then

$$p(\phi_1, \phi_2 \mid y, \sigma^2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2})^2}.I_{[-1,1]},$$

where $I_{[-1,1]}$ indicates the indicator function on [-1,1]. For ϕ_2 , the conditional posterior distribution can be written as follows:

$$p(\pi_2 \mid y, \sigma^2, \pi_1) \quad \propto \quad p(y \mid \pi_1, \pi_2, \sigma^2).p(\pi_2 \mid \pi_1)$$

Because of the fact that π_1 and π_2 are independent, therefore $p(\pi_1, \pi_2) = p(\pi_1)$ and we have

$$p(\pi_2 \mid y, \sigma^2, \pi_1) \quad \propto \quad p(y \mid \pi_1, \pi_2).p(\pi_1)$$

Hence,

$$\begin{aligned} p(\pi_2 \mid y, \sigma^2, \pi_1) \quad &\propto \quad e^{-\frac{1}{2\sigma^2}\sum(y_t-\pi_1(1-\pi_2)y_{t-1}-\pi_2 y_{t-2})^2}.I_{[-1,1]} \\ &\propto \quad e^{-\frac{1}{2\sigma^2}\sum(y_t-\pi_1 y_{t-1}-\pi_2(y_{t-2}-\pi_1 y_{t-1}))^2}.I_{[-1,1]} \end{aligned}$$

Let us denote $z_{s1} = y_t - \pi_1 y_{t-1}$ and $x_{s1} = y_{t-2} - \pi_1 y_{t-1}$. Thus, the conditional posterior distributions using partial autocorrelation for the AR(2) model is given by:

$$p(\pi_2 \mid y, \sigma^2, \pi_1) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum(z_{s1}-\pi_2 x_{s1})^2}.I_{[-1,1]} \tag{5.16}$$

We need to find an expression for the mean of the Normal distribution p($\pi_2 \mid z, \sigma^2$) such that

$$p(\pi_2 \mid z, \sigma^2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum(\pi_1-z))^2}$$

Then, $\pi_2 \mid z, \sigma^2$ is normally distributed with mean equal to z and variance equal to σ^2 Therefore, we take the sum in the right-hand side of equation (5.16) and changed as the follows:

$$\begin{aligned}\sum(z_{s1} - \pi_2 x_{s1})^2 &= \sum(z_{s1}^2 + \pi_2^2 x_{s1}^2 - 2\pi_2 z_{s1} x_{s1}) \\ &= \sum z_{s1}^2 + \pi_2^2 \sum x_{s1}^2 - 2\pi_2 \sum z_{s1} x_{s1} + \left(\frac{\sum z_{s1}.x_{s1}}{\sum x_{s1}^2}\right)^2 - \left(\frac{\sum z_{s1}.x_{s1}}{\sum x_{s1}^2}\right)^2\end{aligned}$$

Then,

$$\sum(z_{s1} - \pi_2 x_{s1})^2 \propto \sum x_{s1}^2 \left(\pi_2 - \frac{\sum z_s.x_{s1}}{\sum x_{s1}^2}\right)^2, \tag{5.17}$$

and by substituting (5.17) into (5.16) we get

$$p(\pi_2 \mid y, \pi_1, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum x_{s1}^2 \left(\pi_2 - \frac{\sum z_s.x_{s1}}{\sum x_{s1}^2}\right)^2} .I_{[-1,1]},$$

If we compare the above equation with the normal distribution, we obtain the posterior distribution for π_2, which is truncated normally distributed, as follows:

$$\pi_2 \mid y, \pi_1, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum z_{s1} x_{s1}}{\sum x_{s1}^2}, \frac{\sigma^2}{\sum x_{s1}^2}\right) \tag{5.18}$$

where $z_{s1} = y_t - \pi_1 y_{t-1}$, and $x_{s1} = y_{t-2} - \pi_1 y_{t-1}$. Therefore, the posterior distribution for π_2 is given by:

$$\pi_2 \mid y, \pi_1, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum(y_t - \pi_1 y_{t-1})(y_{t-2} - \pi_1 y_{t-1})}{\sum y_{t-2} - \pi_1 y_{t-1}^2}, \frac{\sigma^2}{\sum(y_{t-2} - \pi_1 y_{t-1})^2}\right) \tag{5.19}$$

Likewise,the conditional posterior distribution for π_1 can be derived as follows:

$$p(\pi_1 \mid y, \pi_2, \sigma^2) \propto p(y \mid \pi_1, \pi_2, \sigma^2).p(\pi_1 \mid \pi_2)$$

Because of the fact that π_1 and π_2 are independent, therefore $p(\pi_1|\pi_2) = p(\pi_1) = \frac{1}{2}I_{[-1,1]}$.

$$
\begin{aligned}
p(\pi_1 \mid y, \pi_2, \sigma^2) &\propto p(y \mid \pi_1, \pi_2, \sigma^2).p(\pi_2) \\
&\propto e^{-\frac{1}{2\sigma^2}\sum(z_{s2}-\pi_1 x_{s2})^2}.I_{[-1,1]}
\end{aligned} \tag{5.20}
$$

where $z_{s2} = y_t$ - $\pi_2 y_{t-2}$ and $x_{s2} = (1-\pi_2)y_{t-1}$. Then, we re-express the sum in the right-hand side of equation (5.20). It is noted that the same mathematical steps are used to obtain equation (5.17) with the adaption of changing z_{s1} to z_{s2}, as we end up with

$$
\sum(z_{s2} - \pi_1 x_{s2})^2 \propto \sum x_{s2}^2 \left(\pi_1 - \frac{\sum z_{s2}.x_{s2}}{\sum x_{s2}^2}\right)^2 \tag{5.21}
$$

By substituting (5.21) into (5.20) we can obtain the conditional posterior distribution as follows:

$$
p(\pi_1 \mid y, \pi_2, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum x_{s2}^2\left(\pi_1 - \frac{\sum z_{s2}.x_{s2}}{\sum x_{s2}^2}\right)^2}.I_{[-1,1]}
$$

Thus, the posterior distribution of the partial autocorrelation π_1 is a truncated normal distribution:

$$
\pi_1 \mid y, \pi_2, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum(y_t - \pi_2 y_{t-2})(1-\pi_2)y_{t-1}}{\sum((1-\pi_2)y_{t-1})^2}, \frac{\sigma^2}{\sum((1-\pi_2)y_{t-1})^2}\right) \tag{5.22}
$$

5.4 Gibbs sampler for the AR(3) model using partial autocorelations

The objective of this section is to derive the conditional posterior distribution for the AR(3) model based on the corresponding relationship between ϕ_s and π_s . The parameters of the AR(3) model ϕ_1 , ϕ_2 and ϕ_3 are converted using the same steps as in Section (5.3). It is well known from corresponding relationship between ϕ_s and π_s , based on equations provided by

Barndorff-Nielsen and Schou (1973), that:

$$\phi_1 = \pi_1 - \pi_1\pi_2 - \pi_2\pi_3 \tag{5.23}$$

$$\phi_2 = \pi_2 - \pi_1\pi_3 - \pi_1\pi_2\pi_3 \tag{5.24}$$

$$\phi_3 = \pi_3 \tag{5.25}$$

where π_1 , π_2 and π_3 are uniformly distributed between -1 and 1 i.e., ($\pi_1, \pi_2, \pi_3 \sim U[-1, 1]$). As we know from Bayes' theorem,

$$p(\phi_1 \mid y) \propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3})^2}.p(\phi_1) \tag{5.26}$$

$$p(\phi_2 \mid y) \propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3})^2}.p(\phi_2) \tag{5.27}$$

$$p(\phi_3 \mid y) \propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\phi_1 y_{t-1}-\phi_2 y_{t-2}-\phi_3 y_{t-3})^2}.p(\phi_3) \tag{5.28}$$

The posterior conditional distribution for π_3 of the AR(3) model using partial autocorrelations is given by:

$$\begin{aligned} p(\pi_3 \mid y, \pi_1, \pi_2, \sigma^2) &\propto p(y \mid \pi_1, \pi_2, \pi_3, \sigma^2).p(\pi_3 \mid \pi_1, \pi_2) \\ &\propto p(y \mid \pi_1, \pi_2, \pi_3, \sigma^2).\frac{1}{2}I_{[-1,1]} \end{aligned}$$

Note: $p(\pi_i|\pi_{-i}) = \frac{1}{2}.I_{[-1,1]}$, because $\pi_1, \ldots, \pi_p$ are independent and $\pi_i \sim U(-1, 1)$, where $i = 1, \ldots, p$. By substituting the corresponding relationship into equation (5.28), we can get

$$\begin{aligned} p(\pi_3 \mid y, \pi_1, \pi_2, \sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t-(\pi_1-\pi_1\pi_2-\pi_2\pi_3)y_{t-1}-(\pi_2-\pi_1\pi_3-\pi_1\pi_2\pi_3)y_{t-2}-\pi_3 y_{t-3})^2}.I_{[-1,1]} \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t-\pi_1 y_{t-1}+\pi_1\pi_2 y_{t-1}-\pi_2 y_{t-2}-\pi_3(\pi_1\pi_2 y_{t-2}-\pi_2 y_{t-1}-\pi_1 y_{t-2}+y_{t-3}))^2}.I_{[-1,1]} \end{aligned}$$

Denote $z_{s3} = y_t - \pi_1 y_{t-1} + \pi_1\pi_2 y_{t-1} - \pi_2 y_{t-2}$, and $x_{s3} = \pi_1\pi_2 y_{t-2} - \pi_2 y_{t-1} - \pi_1 y_{t-2} + y_{t-3}$. Then,

$$p(\pi_3 \mid y, \pi_1, \pi_2, \sigma^2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum (z_{s3}-\pi_3 x_{s3})^2} I_{[-1,1]} \tag{5.29}$$

As constant terms can be added, we can obtain

$$p(\pi_3 \mid y, \pi_1, \pi_2, \sigma^2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\left(\sum z_{s3}^2 + \pi_3^2 \sum x_{s3}^2 - 2\pi_3 \sum z_{s3} x_{s3} + (\frac{\sum z_{s3}.x_{s3}}{\sum x_{s3}^2})^2\right)} I_{[-1,1]}$$

Then, the posterior conditional distribution for π_3 is given by

$$p(\pi_3 \mid y, \pi_1, \pi_2, \sigma^2) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum x_{s3}^2 \left(\pi_3 - \frac{(\sum z_{s3}.x_{s3})}{\sum x_{s3}^2}\right)^2} .I_{[-1,1]}, \tag{5.30}$$

so that given below the posterior distribution of π_3 is the truncated normally distribution with posterior mean and variance

$$\pi_3 \mid y, \pi_1, \pi_2, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum z_{s3}.x_{s3}}{\sum x_{s3}^2}, \frac{\sigma^2}{\sum x_{s3}^2}\right)$$

As $z_{s3} = y_t - \pi_1 y_{t-1} + \pi_1\pi_2 y_{t-1} - \pi_2 y_{t-2}$ and $x_{s3} = \pi_1\pi_2 y_{t-2} - \pi_2 y_{t-1} - \pi_1 y_{t-2} + y_{t-3}$, the posterior conditional distribution of π_3 is as follows

$$\begin{aligned}\pi_3 \mid y, \pi_1, \pi_2, \sigma^2 \quad \sim \quad & N_{[-1,1]}\left(\frac{\sum(y_t - \pi_1 y_{t-1} + \pi_1\pi_2 y_{t-1} - \pi_2 y_{t-2})(\pi_1\pi_2 y_{t-2} - \pi_2 y_{t-1} - \pi_1 y_{t-2} + y_{t-3})}{\sum(\pi_1\pi_2 y_{t-2} - \pi_2 y_{t-1} - \pi_1 y_{t-2} + y_{t-3})^2},\right. \\ & \left.\frac{\sigma^2}{\sum(\pi_1\pi_2 y_{t-2} - \pi_2 y_{t-1} - \pi_1 y_{t-2} + y_{t-3})^2}\right)\end{aligned} \tag{5.31}$$

Similarly, the conditional posterior distribution for π_2 can be derived as follows

$$p(\pi_2 \mid y, \pi_1, \pi_3, \sigma^2) \quad \propto \quad p(y \mid \pi_1, \pi_2, \pi_3, \sigma^2).p(\pi_2 \mid \pi_1 \pi_3)$$

Again, by substituting the corresponding relationship into equation (5.27), we can get

$$\begin{aligned} p(\pi_2 \mid y, \pi_1, \pi_3, \sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\sum(y_t-(\pi_1-\pi_1\pi_2-\pi_2\pi_3)y_{t-1}-(\pi_2-\pi_1\pi_3-\pi_1\pi_2\pi_3)y_{t-2}-\pi_3 y_{t-3})^2}.I_{[-1,1]} \\ &\propto e^{-\frac{1}{2\sigma^2}\sum(z_{s4}-\pi_2 x_{s4})^2} I_{[-1,1]}, \end{aligned}$$

where $z_{s4} = y_t - \pi_1 y_{t-1} + \pi_1\pi_3 y_{t-2} - \pi_3 y_{t-3}$ and $x_{s4} = y_{t-2} - \pi_1\pi_3 y_{t-2} - \pi_1 y_{t-1} - \pi_3 y_{t-1}$. Then,

$$\begin{aligned} p(\pi_2 \mid y, \pi_1, \pi_3, \sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\left(\sum z_{s4}^2+\pi_2^2\sum x_{s4}^2-2\pi_2\sum z_{s4}x_{s4}+(\frac{\sum z_{s4}x_{s4}}{\sum x_{s4}^2})^2\right)} \\ &\propto e^{-\frac{1}{2\sigma^2}\sum x_{s4}^2\left(\pi_2-\frac{\sum z_{s4}.x_{s4}}{\sum x_{s4}^2}\right)^2}.I_{[-1,1]} \end{aligned}$$

Thus, the posterior distribution of π_2 is given by

$$\pi_2 \mid y, \pi_1, \pi_3, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum z_{s4}.x_{s4}}{\sum x_{s4}^2}, \frac{\sigma^2}{\sum x_{s4}^2}\right),$$

and we can conclude that

$$\begin{aligned} \pi_2 \mid y, \pi_1, \pi_3, \sigma^2 \quad \sim \quad & N_{[-1,1]}\Bigg(\frac{\sum(y_t - \pi_1 y_{t-1} + \pi_1\pi_3 y_{t-2} - \pi_3 y_{t-3})(y_{t-2} - \pi_1\pi_3 y_{t-2} - \pi_1 y_{t-1} - \pi_3 y_{t-1})}{\sum(y_{t-2} - \pi_1\pi_3 y_{t-2} - \pi_1 y_{t-1} - \pi_3 y_{t-1})^2}, \\ & \frac{\sigma^2}{\sum(y_{t-2} - \pi_1\pi_3 y_{t-2} - \pi_1 y_{t-1} - \pi_3 y_{t-1})^2}\Bigg) \end{aligned} \tag{5.32}$$

Now, the posterior conditional distribution for π_1 has to be derived. From Bayes' theorem, we know that

$$\begin{aligned} p(\pi_1 \mid y, \pi_2, \pi_3, \sigma^2) &\propto p(y \mid \pi_1, \pi_2, \pi_3, \sigma^2).p(\pi_1 \mid \pi_2, \pi_3) \\ &\propto p(y \mid \pi_1, \pi_2, \pi_3, \sigma^2).I_{[-1,1]} \end{aligned}$$

By using the corresponding relationship between parameters and partial autocorrelations that was proposed by Barndorff-Nielsen and Schou (1973), we can amend the equation as follows

$$\begin{aligned} p(\pi_1 \mid y, \pi_2, \pi_3, \sigma^2) &\propto e^{-\frac{1}{2\sigma^2}\sum (y_t - (\pi_1 - \pi_1\pi_2 - \pi_2\pi_3) y_{t-1} - (\pi_2 - \pi_1\pi_3 - \pi_1\pi_2\pi_3) y_{t-2} - \pi_3 y_{t-3})^2} . I_{[-1,1]} \\ &\propto e^{-\frac{1}{2\sigma^2}\sum (z_{s5} - \pi_1 x_{s5})^2} . I_{[-1,1]}, \end{aligned}$$

where $z_{s5} = y_t + \pi_2\pi_3 y_{t-1} - \pi_2 y_{t-2} - \pi_3 y_{t-3}$ and $x_{s5} = \pi_2 y_{t-1} + y_{t-1} - \pi_3 y_{t-2} - \pi_2\pi_3 y_{t-2}$. Then, the posterior distribution for π_1 is given by

$$p(\pi_1 \mid y, \pi_2, \pi_3, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}\sum x_{s5}^2 \left(\pi_1 - \frac{\sum z_{s5} . x_{s5}}{\sum x_{s5}^2}\right)^2} . I_{[-1,1]},$$

and thus

$$\pi_1 \mid y, \pi_2, \pi_3, \sigma^2 \sim N_{[-1,1]}\left(\frac{\sum z_{s5} . x_{s5}}{\sum x_{s5}^2}, \frac{\sigma^2}{\sum x_{s5}^2}\right).$$

After substituting the equation for z_{s5} and x_{s5}, the posterior distribution of π_1 can be written as follows:

$$\begin{aligned} \pi_1 \mid y, \pi_2, \pi_3, \sigma^2 \sim\ & N_{[-1,1]}\left(\frac{\sum (y_t + \pi_2\pi_3 y_{t-1} - \pi_2 y_{t-2} - \pi_3 y_{t-3})(\pi_2 y_{t-1} + y_{t-1} - \pi_3 y_{t-3} - \pi_2\pi_3 y_{t-2})}{\sum (\pi_2 y_{t-1} + y_{t-1} - \pi_3 y_{t-2} - \pi_2\pi_3 y_{t-2})^2},\right. \\ & \left.\frac{\sigma^2}{\sum (\pi_2 y_{t-1} + y_{t-1} - \pi_3 y_{t-2} - \pi_2\pi_3 y_{t-2})^2}\right) \end{aligned} \quad (5.33)$$

5.5 MCMC results for the AR(2) and AR(3) models using partial autocorrelation priors

In order to estimate parameters of the AR(2) and AR(3) models, a suitable MCMC scheme is used whereby priors are placed on the partial autocorrelations. This results in a posterior inference for the partial autocorrelation; then posterior estimates of the AR parameters are implied

Table 5.1: Pseudo-code of the MCMC procedure for the AR(2) model based on the partial autocorrelation.

Update 1: $\pi_1,\ \phi_1$ **initiate values**: $\pi_1^{(0)}, \pi_2^{(0)}$ randomly chosen from U(-1, 1) $\pi_{\mathbf{1}}^{\mathbf{(new)}}$: sample from the truncated normal distribution of equation (5.22) $\pi_{\mathbf{1}}^{\mathbf{(new)}}$: calculated based on $\pi_1^{(old)}$ and $\pi_2^{(old)}$ $\phi_{\mathbf{1}}^{\mathbf{(new)}} = \pi_1^{(new)} - \pi_1^{(new)}\pi_2^{(old)}$
Update 2: $\pi_2,\ \phi_2$ $\pi_{\mathbf{2}}^{\mathbf{(new)}}$: sample from the truncated normal distribution of equation (5.19) $\pi_{\mathbf{2}}^{\mathbf{(new)}}$: calculate based on $\pi_1^{(new)}$ and $\pi_2^{(old)}$ $\phi_{\mathbf{2}}^{\mathbf{(new)}} = \pi_2^{(new)}$

Table 5.2: Pseudo-code of the MCMC procedure for the AR(3) model based on the partial autocorrelation.

Update 1: $\pi_1,\ \phi_1$ **initiate values**: $\pi_1^{(0)}, \pi_2^{(0)}$ and $\pi_3^{(0)}$ randomly chosen from U(-1, 1) $\pi_{\mathbf{1}}^{\mathbf{(new)}}$: sample from the truncated normal distribution of equation (5.33) $\pi_{\mathbf{1}}^{\mathbf{(new)}}$: calculate based on $\pi_1^{(old)}$, $\pi_2^{(old)}$ and $\pi_3^{(old)}$ $\phi_{\mathbf{1}}^{\mathbf{(new)}} = \pi_1^{(new)} - \pi_1^{(new)}\pi_2^{(old)} - \pi_2^{(old)}\pi_3^{(old)}$
Update 2: $\pi_2,\ \phi_2$ $\pi_{\mathbf{2}}^{\mathbf{(new)}}$: sample from the truncated normal distribution of equation (5.32) $\pi_{\mathbf{2}}^{\mathbf{(new)}}$: calculate based on $\pi_1^{(new)}$, $\pi_2^{(old)}$ and $\pi_3^{(old)}$ $\phi_{\mathbf{2}}^{\mathbf{(new)}} = \pi_2^{(new)} - \pi_1^{(new)}\pi_3^{(old)} - \pi_1^{(new)}\pi_2^{(new)}\pi_3^{(old)}$
Update 3: $\pi_3,\ \phi_3$ $\pi_{\mathbf{3}}^{\mathbf{(new)}}$: sample from the truncated normal distribution of equation (5.31) $\pi_{\mathbf{3}}^{\mathbf{(new)}}$: calculate based on $\pi_1^{(new)}, \pi_2^{(new)}$ and $\pi_3^{(old)}$ $\phi_{\mathbf{3}}^{\mathbf{(new)}} = \pi_3^{(new)}$

as in Barndorff-Nielsen and Schou (1973). Using the Gibbs sampler for the AR(2) and AR(3) models we estimate partial autocorrelations. For the AR(2) model, partial autocorrelations of $\pi_2\,,\ \pi_1$ from equations (5.19) and (5.22), respectively, are simulated. For the AR(3) model, partial autocorrelations of $\pi_3\,, \pi_2\,, \pi_1$ from equations (5.31), (5.32) and (5.33) , respectively, are simulated. Then, we calculate the parameters of the AR(2) model using equations (5.14) and (5.15), and equations (5.23), (5.24) and (5.25) for the AR(3) model. One of the limitations of estimating partial autocorrelations as described above is that the results sometimes are undefined. This is because of the fact that, from equations of (5.31), (5.32) and (5.33) in section 5.4, it is possible that the denominator part of means and variances of the posterior conditional distributions of these partial autocorrelations can be zero. Thus, we have used rejection sampling in order to reject the obtained undefined results. The results of parameter estimations obtained from partial autocorrelations are not accurate in comparison with our new recommended proposal. We have applied MCMC to a simulated set of observations from the AR(2) model using equations (5.19) and (5.22). Then we calculate parameters through equations (5.14) and (5.15). We have used the true values of $\phi_1 = 0.3$ and $\phi_2 = 0.2$. The parameter estimates using the partial autocorrelations based approach of Section 5.3 are $\phi_1 = 0.14$ and $\phi_2 = 0.16$. The obtained results using our proposed approach are $\phi_1 = 0.306$ and $\phi_2 = 0.203$. The percentage of the differences between the results obtained from partial autocorrelation and the true values are 53% and 20% for ϕ_1 and ϕ_2, respectively, and the percentage of the differences between the results obtained from the proposed approach and the true values are 2% and 1.5% for ϕ_1 and ϕ_2, respectively.

Figure 5.1 illustrates that the results obtained via our proposed prior distributions are more sufficient and are almost the same as the true values. On the other hand, the zoomed histogram of Figure 5.1 indicates that parameters estimated via partial autocorrelations are far from their true values, and trace plots show that the MCMC chains reached convergence. It can be noted that the obtained parameter estimates via partial autocorrelations are not accurate compared with the results obtained from the new proposal. Therefore, in this thesis (Chapter 4) we have proposed a new prior distribution placed directly on the parameters of the AR(p) model.

5.6 Comparison between our proposed approach and the method of Box et al. (1976)

Our aim in this section is to compare our results with those from the method of Box et al. (1976) by using both real and simulated data. In order to focus on the estimation of the AR parameters an affected by the estimation of σ^2, we fix σ^2 and assume it is known or prespecified.

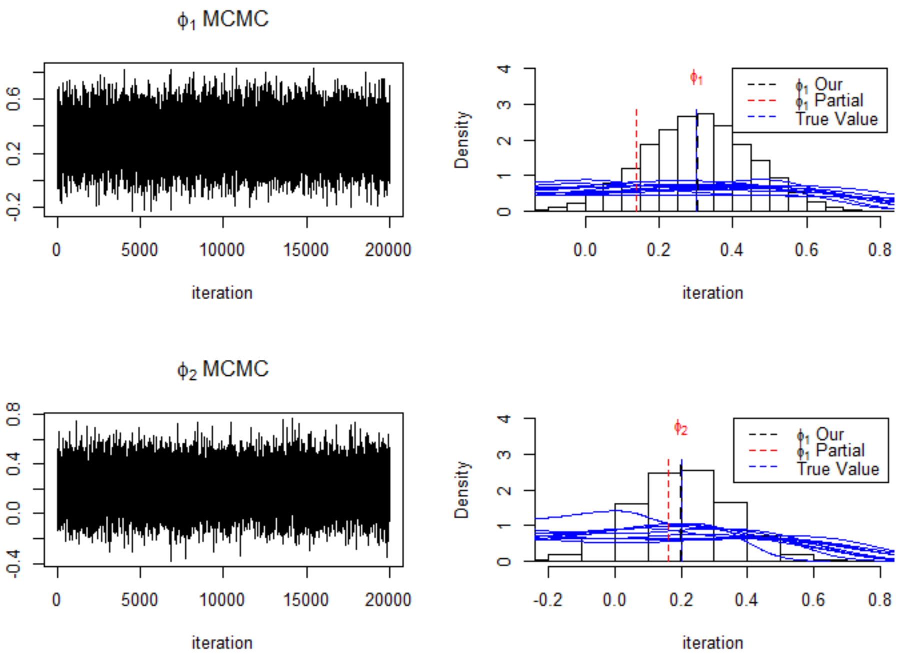


Figure 5.1: Illustrates the results of a comparison between our proposed approach and priors obtained from the correspondence relationship between partial autocorrelations and parameters in Barndorff-Nielsen and Schou (1973) of the AR(2) model. The blue lines of the two histograms indicate that different priors is used for ϕ_1 and ϕ_2.

Based on Box et al. (1976)'s analysis we chose a relatively large value for σ^2 between 2 and 7. Therefore, we are able to isolate the estimation of the AR coefficients alone. Later on, following a fully Bayesian framework, we adopt a weakly informative gamma prior for the precision $\frac{1}{\sigma^2}$ as outlined in Section 4.2. We illustrate the above using both AR(2) and AR(3) models.

5.6.1 Simulation study for the AR(2) model

We have simulated data from the AR(2) model in order to use both the proposed and Box-Jenkins estimation methods. We used the true values $\phi_1 = 0.8$ and $\phi_2 = -0.8$. The obtained results for the AR(2) model using the new proposed approach was $\phi_1 = 0.7976$ and $\phi_2 = -0.7895$ with standard deviations of 0.0512 and 0.0514, respectively. The results obtained using Box et al.

(1976)'s approach are $\phi_1 = 0.7929$ and $\phi_2 = -0.7803$. It can be noted that the obtained results for the AR(2) model using both approaches are approximately the same with almost 1% errors. Trace plots and histograms were obtained for both parameters. The samples seemed to stabilize

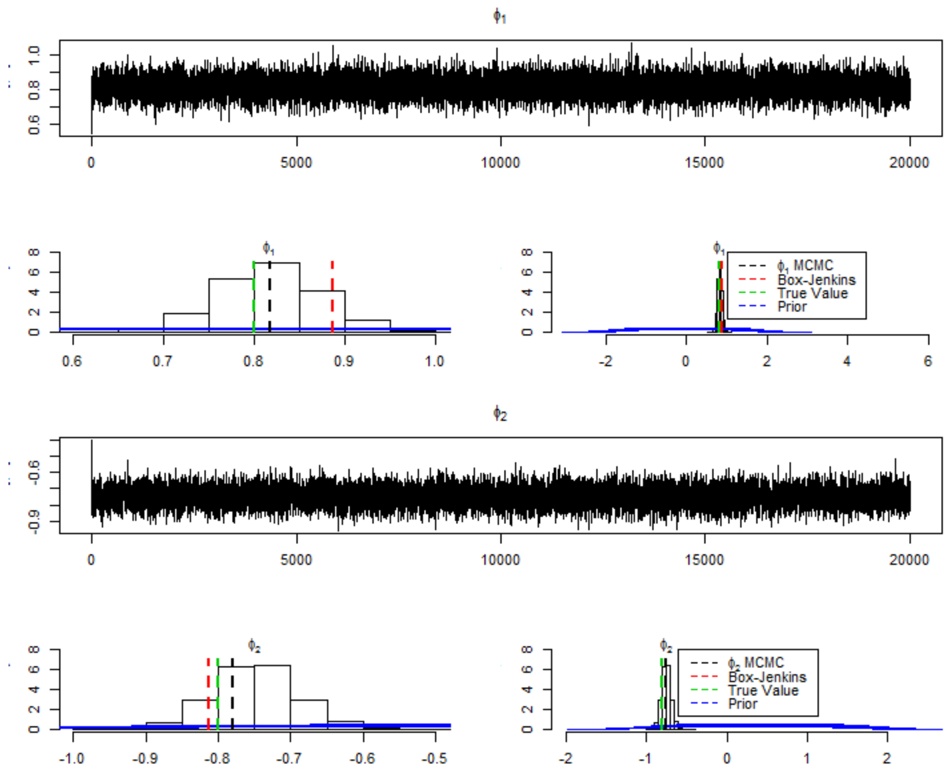


Figure 5.2: Illustrates results of a comparison between the proposed approach and Box et al. (1976)'s method for the AR(2) model. The blue lines indicate that different priors is used for ϕ_1 and ϕ_2, and histograms of the right-hand side are zoomed from histograms histograms of the left-hand side.

reasonably early, and the number of iterations was 20,000 iterations and the length of the burn-in period was taken to be 10% of the iterations. The acceptance rate is high because our proposals are close to the posterior distributions. The left-hand histogram of Figure 5.2 is zoomed in order to clearly present differences between the parameter estimates obtained from Box-Jenkins and our proposed approach, and how accurate these results are accurate in comparison to the true values. Figure 5.2 indicates that there are no appreciable differences in the parameter estimates of both approaches. This means that the obtained results for the AR(2) model using our proposed approach is satisfactory and sufficient. However, our proposed method benefits from being able to easily quantify parameter uncertainty as in a Bayesian setting we are able to provide credible intervals and to assess the quality of the estimates based on a sample from the posterior distribution.

5.6.2 Simulation study for the AR(3) model

A set of observations are simulated in order to obtain parameter estimates of the AR(3) model using our new MCMC proposal presented in Section 4.12. The aim is to compare our approach with Box et al. (1976)'s method. The simulated data from Section 4.12 with true AR values of $\phi_1 = -0.4\,,$ $\phi_2 = -0.8$ and $\phi_3 = -0.6$ were used for this comparison. The obtained parameter estimates for the AR(3) model using our proposal are $\phi_1 = -0.414\,,$ $\phi_2 = -0.792$ and $\phi_3 = -0.599$ with errors of 1.4%, 0.8% and 0.1%, respectively. The results obtained using Box et al. (1976) are $\phi_1 = -0.367\,,$ $\phi_2 = -0.784$ and $\phi_3 = -0.557$ with errors of 4.7%, 0.8% and 4.2%, respectively . It can be noted that the MCMC estimates are closer to the true values when compared to maximum likelihood estimates (Box et al., 1976); leading to smaller residuals (errors).

Trace plots of Figures 4.8 and 4.9 indicate that the chain converged quickly, it was run for 20,000 iterations with a burn-in period of 10% of the iterations. In conclusion, the obtained results using our recommended proposal are more accurate than the parameter estimates obtained using Box-Jenkins. Similar comments to the AR(2) study apply about the access to uncertainty of the parameter estimates.

5.6.3 Illustration for AR(2): monthly Sheffield temperatures

In this section we use real data in order to compare the performance of our proposed prior distribution and that using Box et al. (1976). Data of monthly temperature for a Sheffield

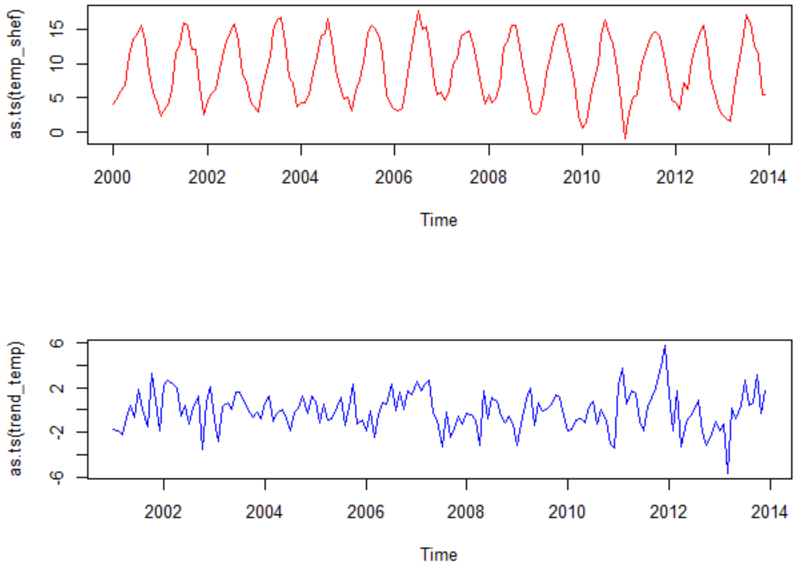


Figure 5.3: Shows the seasonality and the result of removing the seasonality pattern of the monthly temperature data in Sheffield from January 2000 to December 2013.

site were taken from the Climate Research Unit, a subset of the CRU-TS3.22 dataset (http://badc.nerc.ac.uk/data/cru/) (these data are presented in Appendix D). Figure 5.3 indicates that seasonal variation exists and it can be noted that the seasonality is removed. The AIC statistics is used in order to select an AR model because the PACF indicates that an AR(p) is suitable. After using MCMC in order to estimate parameters of the AR(2) model using both maximum likelihood and the proposed approach, the results can be compared.

Several models were fitted in order to select the final model based on Akaike information criterion (AIC) results. The fitted models and their AIC results are presented in Table 5.3. We have selected the AR(2) model because it has the lowest value of AIC (AIC=591.8 and $\sigma^2 = 2.46$). The obtained results for the AR(2) model using the new proposed approach for this data set are $\phi_1 = 0.29$ and $\phi_2 = 0.049$ with standard deviations of 0.159 and 0.160, respectively. The results obtained using Box et al. (1976) are $\phi_1 = 0.289$ and $\phi_2 = 0.051$, with standard deviations 0.080 and 0.080, respectively. It is noted that the MCMC estimates for this model using the proposed

Table 5.3: Shows different fitted time series models their their σ^2 and AIC

Model	Coefficients (Sd)	σ^2	AIC
AR(1)	0.305 (0.076)	2.56	595.8
AR(2)	0.289 0.051 (0.080) (0.080)	2.46	591.8
AR(3)	0.286 0.201 0.109 (0.076) (0.083) (0.080)	2.53	595.7
MA(1)	0.287 (0.076)	2.59	595.5
MA(2)	0.285 0.0536 (0.0813) (0.077)	2.58	597.1
MA(3)	0.289 0.137 0.267 (0.077) (0.078) (0.096)	2.56	595.2

method are almost the same in comparison to the maximum likelihood ones. The errors between the two approaches are 0.001 (for ϕ_1) and 0.002 (for ϕ_2). Trace plots in Figure 5.4 illustrate that the chain converged reasonably early and the number of iterations was 20,000 with a burn-in period of 10% of the iterations. The left-hand zoomed histograms of ϕ_1 and ϕ_2 in Figure 5.4 are the sampled posterior distributions. The black and red dashed lines show that the estimates (posterior mode and MLE) using both approaches are almost the same. The blue dashed lines in each graph illustrates that our prior distribution is a flat prior. In conclusion, the obtained results from the monthly temperature data using our proposed approach were approximately the same as parameter estimates obtained using Box et al. (1976). When fixing $\sigma^2 = 2.46$ both the MCMC and maximum likelihood estimates are very similar (as we have seen). However, in practice we will not know σ^2 and we shall not rely upon maximum likelihood estimation for the variance. Following standard Bayesian procedures we place a gamma prior for the precision

$$\frac{1}{\sigma^2} \sim G(a, b),$$

where a and b are to be specified. Then the Metropolis within Gibbs MCMC scheme (see the pseudo-code of the MCMC procedure in Table 4.9) is used. Table 5.6 shows posterior modes for ϕ_1, ϕ_2 and σ^2 for a range of values a and b. We have chosen $\alpha = 3$ and $\beta = 10$ corresponding to low precision with a mean of 0.3 and a variance of 0.03. This suggests a relatively weakly

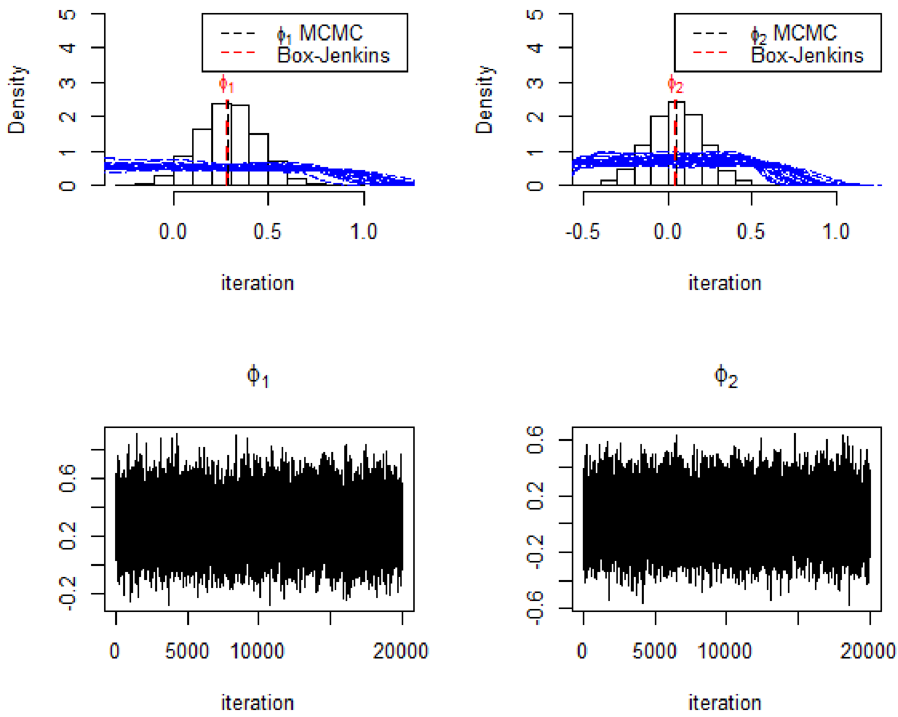


Figure 5.4: Illustrates results of a comparison between the proposed approach with Box et al. (1976)'s method for the AR(2) model. The blue lines indicate that different priors is used for ϕ_1 and ϕ_2.

informative prior for σ^2 ($\sigma^2 \sim IG(3, 10)$) with mean equal to 5, mode equal to 3.333 and variance equal to 25. Figure 5.5 shows the trace plot and the histogram of σ^2 with the posterior mode being equal to 0.1787. In comparison to the MLEs the posterior modes are close to them (as well as close to MCMC using $\sigma^2 = 2.46$), but the significantly lower $\sigma^2 = 0.1787$ results in much tighter credible intervals and hence more accurate estimates. Similar comments apply about the access to uncertainty of parameter estimation.

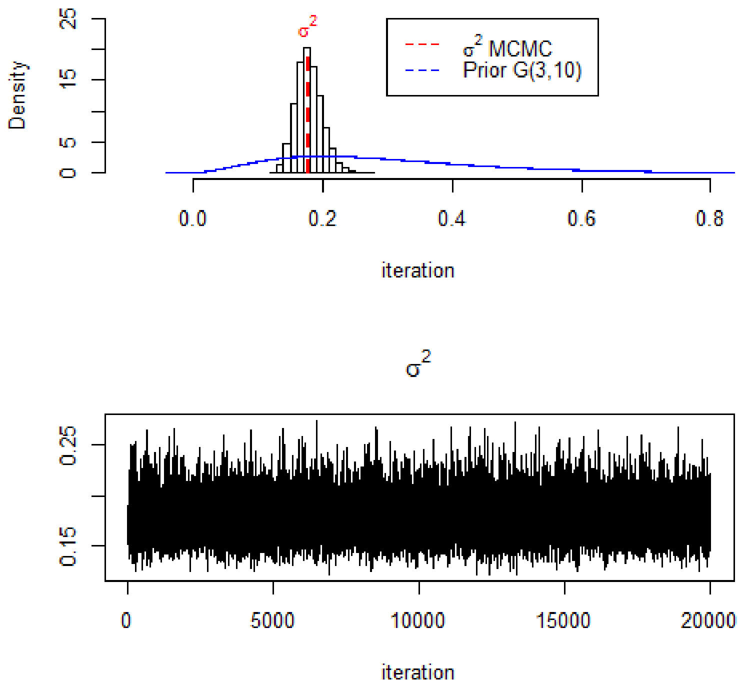


Figure 5.5: Illustrates estimated σ^2 via MCMC for the AR(2) model fitted to the Sheffield temperature data. The blue line indicates the prior distribution for σ^2.

5.6.4 Illustration for AR(3): daily Sheffield temperatures

After using a simulated dataset to estimate parameters of the AR(3) model, now we use real data in order to obtain parameter estimates using the current proposal and to compare with the method of Box et al. (1976). In real data, results of a time series are not based on the particular choice of parameter configurations. In this instance, Sheffield minimum average temperature data taken using a Campbell Stokes recorder from January 2000 to December 2015 was analysed. The data is presented on the `http://www.metoffice.gov.uk/pub/data/weather/uk/climate/stationdata/sheffielddata.txt` website. It can be noted from Figure 5.6 that there exist seasonal patterns. The seasonality is removed (see the lower panel of Figure 5.6) and as above we use the AIC to identify the AR model. Table 5.5 illustrates that the AR(3) model

Table 5.4: Illustrates results of the AR(2) parameters and σ^2 using our proposed approach with some different α and β

α	β	$\phi_1(Sd)$	$\phi_2(Sd)$	$\sigma^2(Sd)$
3	10	0.2898(0.0093)	0.0502(0.0091)	0.1787(0.0202)
3	5	0.2899(0.0052)	0.0501(0.0057)	0.1163(0.0133)
3	1	0.2899(0.0033)	0.0501(0.0033)	0.0661(0.0075)
4	10	0.2900(0.0089)	0.0501(0.0088)	0.1765(0.0200)
4	5	0.2900(0.0057)	0.0501(0.0057)	0.1146(0.0129)
4	1	0.2899(0.0033)	0.0501(0.0033)	0.0655(0.0073)
5	10	0.2900(0.0088)	0.0500(0.0088)	0.1745(0.0197)
5	5	0.2900(0.0057)	0.0501(0.0056)	0.1133(0.0127)
5	1	0.2900(0.0032)	0.0501(0.0032)	0.0646(0.0071)

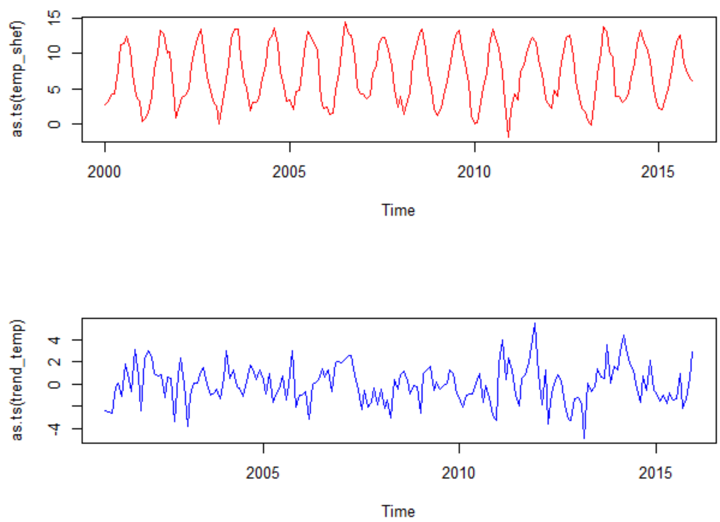


Figure 5.6: Illustrates the minimum average temperature data of Sheffield from January 2000 to December 2015.

is selected based on the lowest value of AIC (AIC = 520 with $\sigma^2 = 1.01$) in order to obtain parameters estimates.

The Metropolis within Gibbs MCMC scheme (see the pseudo-code of the MCMC procedure in Table 4.10) is used. The obtained parameter estimates for the AR(3) model using our current proposal are $\phi_1 = 0.3268$, $\phi_2 = 0.0071$ and $\phi_3 = 0.0977$ with standard deviations of 0.0739, 0.0779 and 0.0741, respectively. The results obtained using maximum likelihood estimates are

$\phi_1 = 0.327$, $\phi_2 = 0.007$ and $\phi_3 = 0.098$. The percentage of difference between results obtained from the current proposal and Box et al. (1976) are for ϕ_1 0.13%, ϕ_2 0.09% and ϕ_3 0.05%. These are approximately the same results obtained using our recommended proposal in comparison to the MLEs. The left-hand zoomed histograms of ϕ_1 , ϕ_2 and ϕ_3 of Figures 5.7, 5.8 and 5.9 show the sampled posterior distributions. The black and red dashed lines show that the estimates (posterior mode and MLE) using both approaches are almost the same. The blue line in each graph illustrates that our prior distribution is a flat prior. Table 5.6 shows posterior modes for ϕ_1 , ϕ_2 , ϕ_3 and σ^2 for a range of values a and b. We have chosen $\alpha = 3$ and $\beta = 10$ corresponding to low precision with a mean of 0.3 and a variance of 0.03. This suggests a relatively weekly informative prior for σ^2 ($\sigma^2 \sim IG(3, 10)$) with mean equal to 5, mode equal to 3.333 and variance equal to 25. Figure 5.10 shows the trace plot and the histogram of σ^2 with the posterior mode being equal to 0.103. In comparison to the MLE s of ϕ_1 , ϕ_2 and ϕ_3 the posterior modes of ϕ_1 , ϕ_2 and ϕ_3 are close to them and are also close to the posterior modes when using fixed $\sigma^2 = 1.01$. However, the significantly lower $\sigma^2 = 0.103$ results in much tighter credible intervals and hence more accurate estimates.

Trace plots in Figures 5.7, 5.8, 5.9 and 5.10 of ϕ_1 , ϕ_2 , ϕ_3 and σ^2 , respectively, illustrate that the chain converged reasonably early and the number of iterations was 20,000 with a burn-in period of 10% of the iterations. In conclusion, the obtained results from daily minimum Sheffield temperatures using our proposed approach were almost the same as the parameter estimates obtained using Maximum likelihood estimation.

Table 5.5: Shows different fitted time series models and their AIC.

Model	**Coefficients (Sd)**	σ^2	**AIC**
AR(1)	0.3438 (0.0029)	1.018	520.18
AR(2)	0.3305 0.0387 (0.0742) (0.0745)	1.017	521.91
AR(3)	0.3268 0.0071 0.0977 (0.0739) (0.0779) (0.0741)	1.01	520.00
MA(1)	0.3214 (0.0678)	1.032	522.27
ARIMA(1,1)	0.5006 -0.1799 (0.2039) (0.2348)	1.015	521.7
ARIMA(2,1)	0.9105 -0.1505 -0.5862 (0.6050) (0.2418) (0.5940)	1.014	523.49
MA(2)	0.3284 0.0647 (0.0739) (0.0674)	1.027	523.66
ARIMA(3,1)	0.0531 0.0952 0.1158 0.2767 (0.4144) (0.1574) (0.0741) (0.4133)	1.004	523.74
ARIMA(1,2)	0.693 -0.3583 -0.1097 (0.281) (0.2905) (0.1464)	1.013	523.21
ARIMA(1,3)	.-0.1464 -0.869 0.4513 1.0074 (0.0486) (0.051) (0.0732) (0.0557)	0.9424	523.05
ARIMA(3,3)	0.1924 -0.8426 0.3234 0.1422 0.9858 0.0794 (0.1994) (0.0466) (0.1805)(0.2126) (0.0453) (0.1192)	0.9301	523.37

Table 5.6: Illustrates results of the AR(3) parameters and σ^2 using our proposed approach with different α and β.

α	β	$\phi_1(Sd)$	$\phi_2(Sd)$	$\phi_3(Sd)$	$\sigma^2(Sd)$
3	10	0.329(0.012)	0.006(0.013)	0.098(0.012)	0.103(0.017)
3	5	0.329(0.004)	0.006(0.004)	0.097(0.004)	0.052(0.005)
3	1	0.329(0.001)	0.006(0.001)	0.097(0.001)	0.011(0.001)
4	10	0.329(0.012)	0.006(0.008)	0.098(0.008)	0.108(0.017)
4	5	0.329(0.004)	0.006(0.004)	0.097(0.004)	0.054(0.005)
4	1	0.329(0.001)	0.006(0.001)	0.097(0.001)	0.011(0.001)
5	10	0.329(0.001)	0.006(0.008)	0.098(0.008)	0.107(0.017)
5	5	0.329(0.004)	0.006(0.004)	0.097(0.004)	0.053(0.005)
5	1	0.329(0.001)	0.006(0.001)	0.097(0.001)	0.011(0.001)

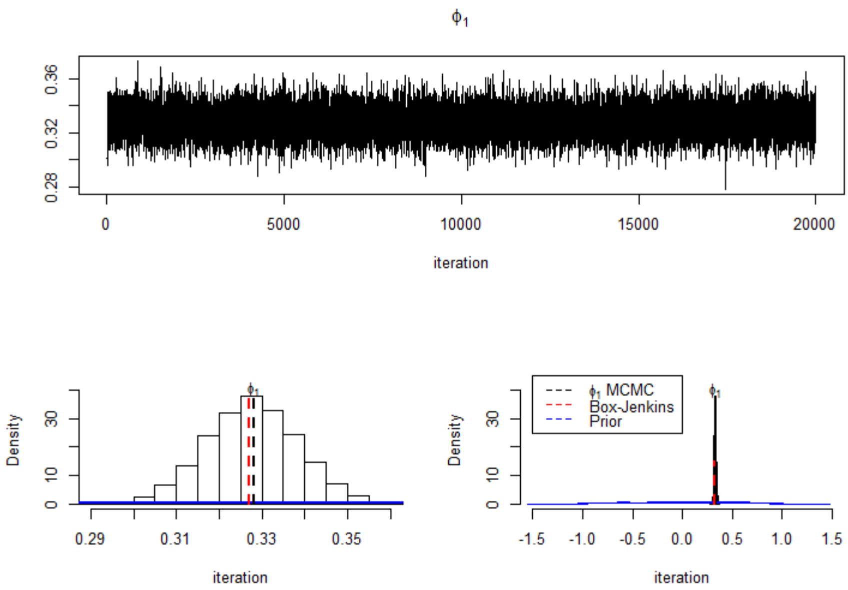


Figure 5.7: Trace plot of ϕ_1 obtained from the MCMC of the minimum temperature data. The blue lines indicate that different priors is used for ϕ_1, and histogram of the right-hand side is zoomed from histograms histogram of the left-hand side.

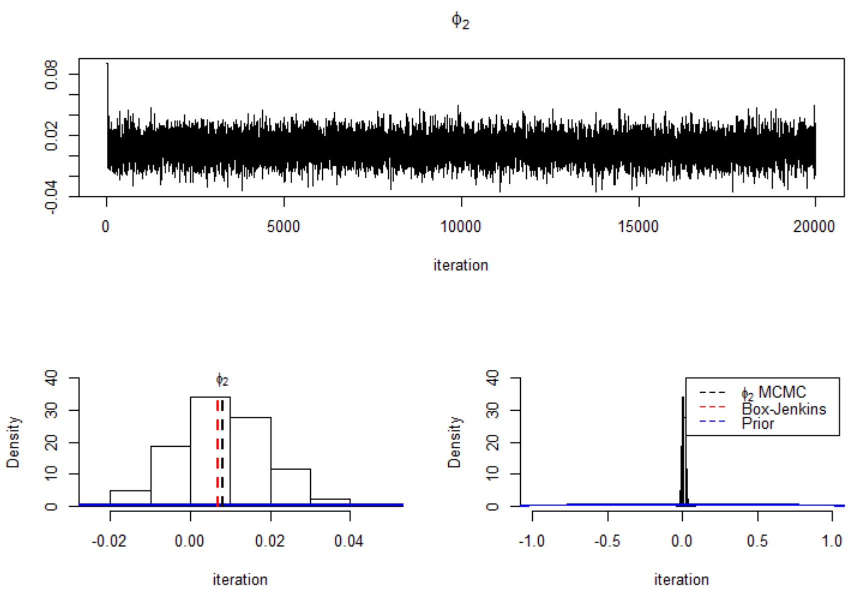


Figure 5.8: Trace plot of ϕ_2 obtained from the MCMC of the minimum temperature data. The blue lines indicate that different priors is used for ϕ_2, and histogram of the right-hand side is zoomed from histograms histogram of the left-hand side.

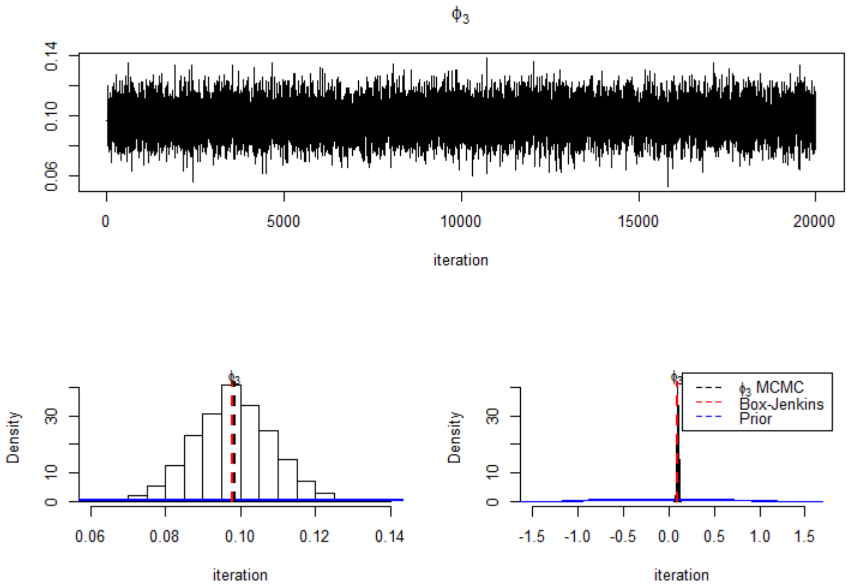


Figure 5.9: Trace plot of ϕ_3 obtained from the MCMC of the minimum temperature data. The blue lines indicate that different priors is used for ϕ_3, and histogram of the right-hand side is zoomed from histograms histogram of the left-hand side.

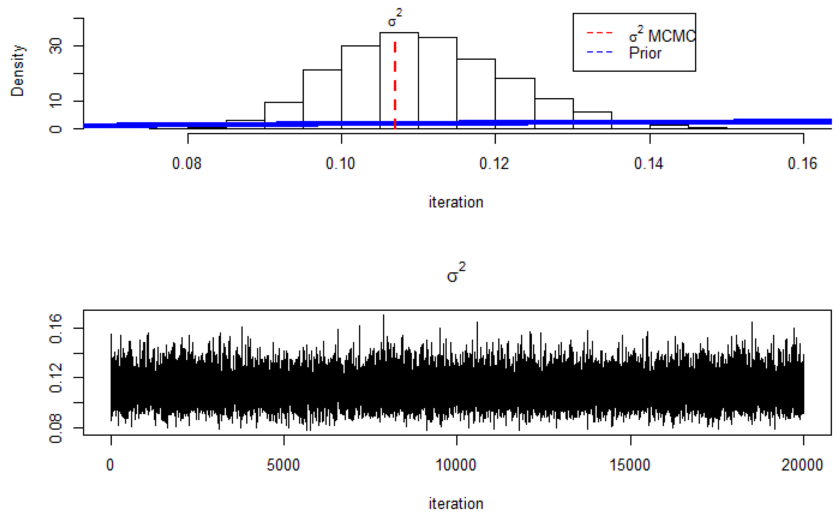


Figure 5.10: Trace plot of σ^2 obtained from the MCMC of the minimum temperature data. based on choosing a $G(3, 10)$ prior for the precision $\frac{1}{\sigma^2}$. The blue lines indicate that different priors is used for σ^2.

Chapter 6

Conclusions and Discussion

6.1 Conclusions

The objective of the current work was to estimate parameters of the AR(p) model using a MCMC procedure. The estimations were obtained using both Gibbs sampling and Metropolis steps. We propose a new flexible prior distribution placed directly on the AR parameters of the AR(p) model. This was motivated by priors proposed for the AR(1), AR(2), ... , AR(6) model, which take advantage of the range of the AR parameters. We then developed a Metropolis step within a Gibbs sampler for estimation of parameters. This scheme was illustrated using simulated data, for AR(2), AR(3) and AR(4) models, and we extended it to models with higher lag order.

MCMC has been applied on a set of simulated data; the data have been simulated on the basis of an AR on model. We have applied MCMC on the application of real data in order to estimate parameters of the AR(2) and AR(3) models using the proposed approach and Box et al. (1976). Our proposed approach gave approximately the same results as Box et al. (1976), but our method benefits from being able to quantify parameter uncertainty, as in a Bayesian setting we are able to provide credible intervals and to assess the quality of the estimates based on a sample from the posterior distribution.

We advocate the use of prior distributions placed directly on the parameters. Thus, the stationarity conditions were revisited because this restricts the space of the parameters. Furthermore, one of the advantages of this study is that we developed and derived stationarity conditions for the AR(p) model by determining the region of the stationarity conditions for the model. The prior distribution for the AR(2) model placed directly on the parameters of the model provided the same prior as that implied by placing uniform priors on the partial autocorrelations.

We determined the restriction of the stationarity conditions for the AR(3) model using a three dimensional graph. This was done by simulating parameters of the AR(3) model using rejection sampling. We have found that our new flexible prior distribution is more suitable than the prior distributions obtained from the correspondence relationship between partial autocorrelations and parameters discussed by Barndorff-Nielsen and Schou (1973) and Jones (1987) when applying MCMC to estimate parameters of the AR(p) model, especially when $p \geq 3$.

We concluded a study on simulated data to evaluate the performance of our new proposed prior distribution for the AR(3) model. We have used Bayes factors in order to distinguish between models.

There are a number of limitations that could be addressed in a future study. First, there is not much information available on the stationarity conditions. A general formula for stationarity conditions does not exist for the higher order polynomial model. Additionally, we cannot control all parameters simultaneously in order to estimate parameters of the AR model using a Metropolis approach.

6.2 Extensions and future work

The models and methods developed in this thesis can be extended in various ways. Here we outline some possibilities.

In Chapter 4 we developed MCMC schemes for inference of autoregressive models (AR). At the core of our work is the specification of priors that respect stationarity. The proposed methodology can be extended in order to get prior for moving average (MA) time series models. Indeed by taking into account the duality between stationarity in autoregressive models and invertibility in moving average models we can propose suitable MCMC inference for moving average models. The stationarity conditions proposed in Chapter 3 for autoregressive models are directly translated to invertibility conditions for a suitable moving average model. Then Chapter 4 can propose priors and MCMC inference for moving average models too. It is believed that with some extensions the methodology in Chapter 4 can be extended to mixed-models, i.e., autoregressive moving average (ARMA) models. Indeed, the priors for stationarity should be as in Chapter 4 and the priors for invertibility should be as discussed above. A suitable Metropolis within Gibbs sampling scheme should be relatively simple. The proposal would involve separate Metropolis steps for the AR and MA parts.

The starting point of this research was to propose priors placed on the AR coefficients rather than placed indirectly on functions of the coefficients, such as the roots of the AR characteristic polynomial or the partial autocorrelations. The proposed priors use uniform distributions and respect the proposed inequalities which are imposed by the stationarity assumption. However, the uniform distribution choice for the priors may be dropped. One could replace the uniform with a truncated normal distribution in all our studies. This is somewhat an advantage of our proposal, in that many distributions may be used to build the priors (as long as the inequalities are respected). A future line of research could be directed in comparing the effect of different priors (in particular building priors using the uniform against the truncated normal distribution).

In Chapter 3 we have proposed a set of new stationarity conditions for the AR model. These conditions offer greater simplicity in comparison to the existing conditions, because they consist of linear inequalities. These inequalities can be exploited in order to propose a test for stationarity, for example to be applied to co-integration. It is believed that there is a pattern over which stationarity conditions (similar to those proposed in this thesis) of order more than four can be built. It is hoped that an inductive algorithm can build groups of stationarity conditions from one lag order to the next and a computer program could implement tests or checks of stationarity very efficiently.

In Chapter 3 we have proposed a set of sufficient stationarity conditions for autoregressive models of any order. However, the quest for necessary and sufficient conditions is still open and one future direction of research would be to discover necessary and sufficient conditions for any order, or at least for up to some lag order. In our study, we have found that the sufficient conditions work well as we place priors based on them, but for testing or for other estimation problems, as well as for theoretical purposes, it is desirable to have necessary and sufficient conditions.

The methodology in Chapter 4 may be extended to slowly-varying time-varying AR models. There are a few opportunities here, such as the threshold autoregressive models, which are suitable for locally-stationary data, or other AR models with time-varying AR coefficients. Such models have been proposed, especially in the econometrics, finance, literature, and an extension of our work could offer future lines of research.

Bibliography

Albert, J. H., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business & Economic Statistics 11 (1), 1–15.

Anderson, T. W., 1971. The statistical analysis of time series. Vol. 19. John Wiley & Sons.

Ando, T., 2010. Bayesian model selection and statistical modeling. Taylor & Francis US.

Barndorff-Nielsen, O., Schou, G., 1973. On the parametrization of autoregressive models by partial autocorrelations. Journal of Multivariate Analysis 3 (4), 408–419.

Barnett, G., Kohn, R., Sheather, S., 1996. Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. Journal of Econometrics 74 (2), 237–254.

Barnett, G., Kohn, R., Sheather, S., 1997. Robust Bayesian estimation of autoregressive–moving-average models. Journal of Time Series Analysis 18 (1), 11–28.

Box, G. E., Jenkins, G. M., Reinsel, G. C., 1976. Time series analysis: forecasting and control. Wiley.

Box, G. E., Jenkins, G. M., Reinsel, G. C., 2008. Time series analysis: forecasting and control. Wiley. com.

Brass, W., 1974. Perspectives in population prediction: Illustrated by the statistics of England and Wales. Journal of the Royal Statistical Society. Series A (General), 532–583.

Brockwell, P. J., Davis, R. A., 2001. Time series: theory and methods. Springer.

Brockwell, P. J., Davis, R. A., 2002. Introduction to time series and forecasting. Taylor & Francis US.

Brockwell, P. J., Davis, R. A., 2006. Introduction to time series and forecasting. Springer Science & Business Media.

Casella, G., George, E. I., 1992. Explaining the Gibbs sampler. The American Statistician 46 (3), 167–174.

Chatfield, C., 2003. The analysis of time series: an introduction. CRC press.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The american statistician 49 (4), 327–335.

Chipman, J. S., 1950. The multi-sector multiplier. Econometrica: Journal of the Econometric Society, 355–374.

Cohn, A., 1922. Über die anzahl der wurzeln einer algebraischen gleichung in einem kreise. Mathematische Zeitschrift 14 (1), 110–148.

Cowpertwait, P. S., Metcalfe, A. V., 2009. Introductory time series with R. Springer Science & Business Media.

Dunson, D. B., 2001. Commentary: practical advantages of Bayesian analysis of epidemiologic data. American Journal of Epidemiology 153 (12), 1222–1226.

Enders, W., 2008. Applied econometric time series. John Wiley & Sons.

Farebrother, R., 1973. Simplified Samuelson conditions for cubic and quartic equations. The Manchester School 41 (4), 396–400.

Farebrother, R., 1992. A note on the Schur and Samuelson conditions. The Manchester [S]chool 60, 79–81.

Gargantini, I., 1971. The numerical stability of the Schur-Cohn criterion. SIAM Journal on Numerical Analysis 8 (1), 24–29.

Gelfand, A. E., Smith, A. F., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85 (410), 398–409.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6), 721–741.

Gilgen, H., 2006. Univariate time series in geosciences. Springer.

Gill, J., 2014. Bayesian methods: A social and behavioral sciences approach. Vol. 20. CRC press.

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97–109.

Hipel, K. W., McLeod, A. I., 1994. Time series modelling of water resources and environmental systems. Elsevier.

Hoff, P. D., 2009. A first course in Bayesian statistical methods. Springer Science & Business Media.

Hoffman, K., Kunze, R., 1971. Álgebra lineal. Fondo Educativo Interamericano. M México.

Huerta, G., West, M., 1999. Priors and component structures in autoregressive time series models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (4), 881–899.

Jeffreys, H., 1998. The theory of probability. OUP Oxford.

Jones, M., 1987. Randomly choosing parameters from the stationarity and invertibility region of autoregressive-moving average models. Applied Statistics, 134–138.

Kass, R. E., Raftery, A. E., 1995. Bayes factors. Journal of the American Statistical Association 90 (430), 773–795.

Kulahci, M., Bisgaard, S., 2011. Time series analysis and forecasting by example. Wiley.

Marmol, F., 1995. The stationarity conditions for an AR (2) process and Schur's theorem. Econometric Theory 11 (05), 1180–1182.

Marriott, J., Ravishanker, N., Gelfand, A., Pai, J., 1995. Bayesian analysis of ARMA processes: Complete sampling based inference under full likelihoods. Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner, ed. DA Berry, KM Chaloner and JK Geweke. New York: Wiley.

McCulloch, R. E., Tsay, R. S., 1993. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. Journal of the American Statistical Association 88 (423), 968–978.

McLeod, A. I., Zhang, Y., 2008. Improved subset autoregression: With R package. Journal of Statistical Software 28 (2), 1–28.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21, 1087.

Monahan, J. F., 1983. Fully Bayesian analysis of ARMA time series models. Journal of Econometrics 21 (3), 307–331.

Monahan, J. F., 1984. A note on enforcing stationarity in autoregressive-moving average models. Biometrika 71 (2), 403–404.

Najim, M., 2010. Modeling, estimation and optimal filtration in signal processing. Vol. 25. John Wiley & Sons.

Okuguchi, K., Irie, K., 1990. The Schur and Samuelson conditions for a cubic equation. The Manchester School 58 (4), 414–418.

Peña, D., Tiao, G. C., Tsay, R. S., 2011. A course in time series analysis. Vol. 322. John Wiley & Sons.

Petris, G., Petrone, S., Campagnoli, P., 2009. Dynamic linear models with R. Springer Science & Business Media.

Priestley, M. B., 1981. Spectral analysis and time series.

Rao, S. S., 2008. A course in time series analysis. Tech. rep., Technical report, Texas A&M University.

Robert, C., Casella, G., 2009. Introducing Monte Carlo methods with R. Springer Science & Business Media.

Robert, C., Casella, G., 2010. Introducing Monte Carlo methods with R. Springer.

Samuelson, P. A., 1941. Conditions that the roots of a polynomial be less than unity in absolute value. The Annals of Mathematical Statistics 12 (3), 360–364.

Schur, J., 1918. ber potenzreihen, die im innern des einheitskreises beschrnkt sind. Journal fr die reine und angewandte Mathematik 148, 122–145.
URL http://eudml.org/doc/149476

Shumway, R. H., Stoffer, D. S., 2011. Time series analysis and its applications: with R examples. Springer.

Sorensen, D., Gianola, D., 2002. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer Science & Business Media.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (4), 583–639.

Steel, M., 2008. Bayesian time series analysis. The New Palgrave Dictionary of Economics, 2nd ed. Palgrave Macmillan, London.

Triantafyllopoulos, K., Bersimis, S., 2016. Phase II control charts for autocorrelated processes. Quality Technology & Quantitative Management 13 (1), 88–108.

Wei, W. W.-S., 1994. Time series analysis. Addison-Wesley Redwood City, California.

West, M., Harrison, J., 1999. Bayesian Forecasting & Dynamic Models. Springer, 2nd ed.

Wise, J., 1956. Stationarity conditions for stochastic processes of the autoregressive and moving-average type. Biometrika 43 (1/2), 215–219.

Wold, H., 1939. A study in the analysis of stationary time series. Ph.D. thesis, Almquist & Wilksell.

Appendixes

A MCMC application for AR(1)

In this appendix, MCMC is applied for the AR(1) model in a way that iterations are calculated throughout mean and median. Several trials are conducted for each application using different parametrization and different sample size. Additionally, Monte Carlo experiment has also applied for each application.

A.1 MCMC for AR(1) through mean calculation

The result of Monte Carlo experiment (N=100) via Gibbs sampler for AR(1) through mean.

$n = 500$	$\phi(SD)$	$\sigma^2(SD)$
ϕ	$\sigma^2 = 9$	
0.3	**0.3035695(0.04241648)**	**11.45019(58.3374072289382)**
0.5	**0.4993794(0.03542788)**	**9.58903(49.1396869310546)**
0.8	**0.7942672(0.02782708)**	**12.17568(35.3621794418434)**
ϕ	$\sigma^2 = 16$	
0.3	**0.2975381(0.04088179)**	**19.93656(75.8073674148096)**
0.5	**0.4982909(0.03408459)**	**16.04293(64.3705684886756)**
0.8	**0.7987673(0.02294168)**	**14.68617(46.1567575800935)**
ϕ	$\sigma^2 = 100$	
0.3	**0.3020121(0.03750834)**	**125.3021(184.697515338205)**

Continued on next page

Table 1 – ***Continued from previous page***

0.5	**0.5024354(0.03325)**	**106.5376(195.88715332871)**
0.8	**0.7991675(0.02439708)**	**122.9728(119.118920229871)**
ϕ	$\sigma^2 = 10000$	
0.3	**0.2988242(0.03364664)**	**10365.51(1590.02482028744)**
0.5	**0.4932745(0.02981814)**	**9706.166(826.199786675215)**
0.8	**0.7997934(0.02329609)**	**11151.09(1745.25217235901)**
$n = 1000$	$\phi(SD)$	$\sigma^2(SD)$
ϕ	$\sigma^2 = 9$	
0.3	**0.297507(0.02870523)**	**10.35882(81.0632252625435)**
0.5	**0.5034632(0.02648413)**	**10.49218(75.1360526072586)**
0.8	**0.8000454(0.01704109)**	**9.160845(52.9613609798699)**
ϕ	$\sigma^2 = 16$	
0.3	**0.2979963(0.02833023)**	**19.29028(103.98212006649)**
0.5	**0.5041558(0.0244048)**	**19.01725(91.0089525506624)**
0.8	**0.7980846(0.01839579)**	**18.91693(75.088604553373)**
ϕ	$\sigma^2 = 100$	
0.3	**0.2989155(0.02195994)**	**84.14596(220.080835690949)**
0.5	**0.4956312(0.02334992)**	**107.6356(245.838325900002)**
0.8	**0.8020782(0.01686264)**	**125.0334(178.781528793033)**
ϕ	$\sigma^2 = 10000$	
0.3	**0.3013527(0.0241207)**	**9965.448(4438.13440357465)**
0.5	**0.5036863(0.02218887)**	**9745.297(16640.5880916956)**
0.8	**0.7934447(0.01574845)**	**10821.34(1718.80962120874)**

B MCMC for AR(1) through median calculation

The result of Monte Carlo experiment (N=100) via Gibbs sampler for AR(1) through median.

$n = 500$	$\phi(SD)$	$\sigma^2(SD)$
ϕ	$\sigma^2 = 9$	
0.3	**0.3036262(0.04241648)**	**11.27793(58.3374072289382)**
0.5	**0.4993833(0.03542788)**	**9.419552(49.1396869310546)**
0.8	**0.7942666(0.02782708)**	**11.60681(35.3621794418434)**
ϕ	$\sigma^2 = 16$	
0.3	**0.2975074(0.04088179)**	**19.65716(75.8073674148096)**
0.5	**0.4983044(0.03408459)**	**15.76756(64.3705684886756)**

Continued on next page

Table 2 – ***Continued from previous page***

0.8	**0.7987807(0.02294168)**	**14.14226(46.1567575800935)**
ϕ	$\sigma^2 = 100$	
0.3	**0.3020622(0.03750834)**	**122.9993(184.697515338205)**
0.5	**0.5024318(0.03325)**	**104.4498(195.88715332871)**
0.8	**0.799232(0.02439708)**	**115.3406(119.118920229871)**
ϕ	$\sigma^2 = 10000$	
0.3	**0.2988705(0.03364664)**	**10209.84(1590.02482028744)**
0.5	**0.4932792(0.02981814)**	**9515.633(826.199786675215)**
0.8	**0.7998268(0.02329609)**	**10516.75(1745.25217235901)**
$n = 1000$	$\phi(SD)$	$\sigma^2(SD)$
ϕ	$\sigma^2 = 9$	
0.3	**0.2975043(0.02870523)**	**10.29701(81.0632252625435)**
0.5	**0.5034656(0.02648413)**	**10.41061(75.1360526072586)**
0.8	**0.800052(0.01704109)**	**8.976618(52.9613609798699)**
ϕ	$\sigma^2 = 16$	
0.3	**0.2979671(0.02833023)**	**19.16991(103.98212006649)**
0.5	**0.5041792(0.0244048)**	**18.76385(91.0089525506624)**
0.8	**0.7980757(0.01839579)**	**18.47265(75.088604553373)**
ϕ	$\sigma^2 = 100$	
0.3	**0.2989267(0.02195994)**	**83.62193(220.080835690949)**
0.5	**0.4956375(0.02334992)**	**106.5391(245.838325900002)**
0.8	**0.8020535(0.01686264)**	**120.3887(178.781528793033)**
ϕ	$\sigma^2 = 10000$	
0.3	**0.3014118(0.0241207)**	**9894.474(4438.13440357465)**
0.5	**0.5036773(0.02218887)**	**9645.914(16640.5880916956)**
0.8	**0.7934537(0.01574845)**	**10511.05(1718.80962120874)**

C Stationary conditions and Prior distribution for AR(5) and AR(6) model

In order to understand more about the stationary conditions and prior distribution of AR(p) , determining stationary conditions and prior distribution have been done for AR(5) and AR(6).

C.1 Stationary conditions and Prior distribution for AR(5) model

Determining stationary condition for the AR(5) model can be done through equation (3.70) which we know as follow;

$$A_5 = \begin{bmatrix} A_4 & I_2^3 \\ -A_4 & I_2^*3 \end{bmatrix}$$

$$A_5\Phi \prec \mathbf{1}, \tag{1}$$

$$\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 < 1 \tag{2}$$

$$\phi_1 - \phi_2 - \phi_3 - \phi_4 - \phi_5 < 1 \tag{3}$$

$$-\phi_1 + \phi_2 - \phi_3 + \phi_4 + \phi_5 < 1 \tag{4}$$

$$-\phi_1 - \phi_2 + \phi_3 - \phi_4 - \phi_5 < 1 \tag{5}$$

$$-\phi_1 - \phi_2 - \phi_3 + \phi_4 + \phi_5 < 1 \tag{6}$$

$$-\phi_1 + \phi_2 + \phi_3 - \phi_4 - \phi_5 < 1 \tag{7}$$

$$\phi_1 - \phi_2 + \phi_3 + \phi_4 + \phi_5 < 1 \tag{8}$$

$$\phi_1 + \phi_2 - \phi_3 - \phi_4 - \phi_5 < 1 \tag{9}$$

$$-\phi_1 - \phi_2 - \phi_3 + \phi_4 + \phi_5 < 1 \tag{10}$$

$$-\phi_1 + \phi_2 + \phi_3 - \phi_4 - \phi_5 < 1 \tag{11}$$

$$\phi_1 - \phi_2 + \phi_3 + \phi_4 + \phi_5 < 1 \tag{12}$$

$$\phi_1 + \phi_2 - \phi_3 - \phi_4 - \phi_5 < 1 \tag{13}$$

$$\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 < 1 \tag{14}$$

$$\phi_1 - \phi_2 - \phi_3 - \phi_4 - \phi_5 < 1 \tag{15}$$

$$-\phi_1 + \phi_2 - \phi_3 + \phi_4 + \phi_5 < 1 \tag{16}$$

$$-\phi_1 - \phi_2 + \phi_3 - \phi_4 - \phi_5 < 1 \tag{17}$$

From equations of (20) to (27), we can obtain right sides for ϕ_1 as follow;

$$\phi_1 < 1 - |\phi_2| - |\phi_3| - |\phi_3 + \phi_5|$$

However, from equations of (28) to (41), we can determine left side for ϕ_1 which is;

$$\phi_1 > -1 + |\phi_2| + |\phi_4| - |\phi_3 - \phi_5|$$

Hence, $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$ **is uniformly distributed with upper and lower ranges which is** $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5) \sim U(|\phi_2| + |\phi_4| - |\phi_3 - \phi_5| - 1, 1 - |\phi_2| - |\phi_3| - |\phi_3 + \phi_5|)$ **where** $(i = 2, 3, 4, 5)$ **.Therefore the prior distribution for AR(5) is;**

$$p(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5) = \frac{1}{16(2 - 2|\phi_2| - 2|\phi_4| - |\phi_3 + \phi_5| - |\phi_3 - \phi_5|)} \tag{18}$$

C.2 Stationary conditions and Prior distribution for AR(6) model

Determining stationary condition for the AR(6) model can be done through equation (3.70) which we know as follow;

$$A_6 = \begin{bmatrix} A_5 & I_2^4 \\ -A_5 & I_2^* 4 \end{bmatrix}$$

$$A_6 \Phi \prec \mathbf{1}, \tag{19}$$

It can be seen that the above are one set of equations of stationary conditions,but these can be divided into tow columns in order to simplify equations. As it has been known that there are 32 stationary conditions in AR(6) model. Out of 32 equations, 32 of them are those $\phi_1 s$ **in the column B which their signs are +. The other remaining 16 stationary conditions are those that have been shown in the column D as the**

follows;

Set A conditions	**Set B conditions**	
$\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 < 1 - \phi_2 - \phi_3 - \phi_4 - \phi_5 - \phi_6$	(20)
$\phi_1 - \phi_2 - \phi_3 - \phi_4 - \phi_5 - \phi_6 < 1$	$\phi_1 < 1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 + \phi_6$	(21)
$\phi_1 - \phi_2 + \phi_3 + \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 < 1 + \phi_2 - \phi_3 - \phi_4 - \phi_5 - \phi_6$	(22)
$\phi_1 + \phi_2 - \phi_3 - \phi_4 - \phi_5 - \phi_6 < 1$	$\phi_1 < 1 - \phi_2 + \phi_3 + \phi_4 + \phi_5 + \phi_6$	(23)
$\phi_1 - \phi_2 + \phi_3 - \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 < 1 + \phi_2 - \phi_3 + \phi_4 - \phi_5 - \phi_6$	(24)
$\phi_1 + \phi_2 - \phi_3 + \phi_4 - \phi_5 - \phi_6 < 1$	$\phi_1 < 1 - \phi_2 + \phi_3 - \phi_4 + \phi_5 + \phi_6$	(25)
$\phi_1 + \phi_2 + \phi_3 - \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 < 1 - \phi_2 - \phi_3 + \phi_4 - \phi_5 - \phi_6$	(26)
$\phi_1 - \phi_2 - \phi_3 + \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 < 1 + \phi_2 + \phi_3 - \phi_4 + \phi_5 - \phi_6$	(27)
$\phi_1 - \phi_2 + \phi_3 - \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 < 1 + \phi_2 - \phi_3 + \phi_4 + \phi_5 - \phi_6$	(28)
$\phi_1 + \phi_2 - \phi_3 + \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 < 1 - \phi_2 + \phi_3 - \phi_4 - \phi_5 + \phi_6$	(29)
$\phi_1 + \phi_2 + \phi_3 - \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 < 1 - \phi_2 - \phi_3 + \phi_4 + \phi_5 - \phi_6$	(30)
$\phi_1 - \phi_2 - \phi_3 + \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 < 1 + \phi_2 + \phi_3 - \phi_4 - \phi_5 + \phi_6$	(31)
$\phi_1 + \phi_2 + \phi_3 + \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 < 1 - \phi_2 - \phi_3 - \phi_4 + \phi_5 - \phi_6$	(32)
$\phi_1 - \phi_2 - \phi_3 - \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 < 1 + \phi_2 + \phi_3 + \phi_4 - \phi_5 + \phi_6$	(33)
$\phi_1 - \phi_2 + \phi_3 + \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 < 1 + \phi_2 - \phi_3 - \phi_4 + \phi_5 - \phi_6$	(34)
$\phi_1 + \phi_2 - \phi_3 - \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 < 1 - \phi_2 + \phi_3 + \phi_4 - \phi_5 + \phi_6$	(35)

Set C conditions	Set D conditions	
$-\phi_1 + \phi_2 - \phi_3 + \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 > -1 + \phi_2 - \phi_3 + \phi_4 + \phi_5 + \phi_6$	(36)
$-\phi_1 - \phi_2 + \phi_3 - \phi_4 - \phi_5 - \phi_6 < 1$	$\phi_1 > -1 - \phi_2 + \phi_3 - \phi_4 - \phi_5 - \phi_6$	(37)
$-\phi_1 - \phi_2 - \phi_3 + \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 > -1 - \phi_2 - \phi_3 + \phi_4 + \phi_5 + \phi_6$	(38)
$-\phi_1 + \phi_2 + \phi_3 - \phi_4 - \phi_5 - \phi_6 < 1$	$\phi_1 > -1 + \phi_2 + \phi_3 - \phi_4 - \phi_5 - \phi_6$	(39)
$-\phi_1 - \phi_2 - \phi_3 - \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 > -1 - \phi_2 - \phi_3 - \phi_4 + \phi_5 + \phi_6$	(40)
$-\phi_1 + \phi_2 + \phi_3 + \phi_4 - \phi_5 - \phi_6 < 1$	$\phi_1 > -1 + \phi_2 + \phi_3 + \phi_4 - \phi_5 - \phi_6$	(41)
$-\phi_1 + \phi_2 - \phi_3 - \phi_4 + \phi_5 + \phi_6 < 1$	$\phi_1 > -1 + \phi_2 - \phi_3 - \phi_4 + \phi_5 + \phi_6$	(42)
$-\phi_1 - \phi_2 + \phi_3 + \phi_4 - \phi_5 - \phi_6 < 1$	$\phi_1 > -1 - \phi_2 + \phi_3 + \phi_4 - \phi_5 - \phi_6$	(43)
$-\phi_1 - \phi_2 - \phi_3 - \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 > -1 - \phi_2 - \phi_3 - \phi_4 - \phi_5 + \phi_6$	(44)
$-\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 > -1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 - \phi_6$	(45)
$-\phi_1 + \phi_2 - \phi_3 - \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 > -1 + \phi_2 - \phi_3 - \phi_4 - \phi_5 + \phi_6$	(46)
$-\phi_1 - \phi_2 + \phi_3 + \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 > -1 - \phi_2 + \phi_3 + \phi_4 + \phi_5 - \phi_6$	(47)
$-\phi_1 + \phi_2 - \phi_3 + \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 > -1 + \phi_2 - \phi_3 + \phi_4 - \phi_5 + \phi_6$	(48)
$-\phi_1 - \phi_2 + \phi_3 - \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 > -1 - \phi_2 + \phi_3 - \phi_4 + \phi_5 - \phi_6$	(49)
$-\phi_1 - \phi_2 - \phi_3 + \phi_4 - \phi_5 + \phi_6 < 1$	$\phi_1 > -1 - \phi_2 - \phi_3 + \phi_4 - \phi_5 + \phi_6$	(50)
$-\phi_1 + \phi_2 + \phi_3 - \phi_4 + \phi_5 - \phi_6 < 1$	$\phi_1 > -1 + \phi_2 + \phi_3 - \phi_4 + \phi_5 - \phi_6$	(51)

From equations of (20) to (41), we can obtain right sides for ϕ_1 as follow;

$$\phi_1 < 1 - |\phi_2| - |\phi_3| - |\phi_3 + \phi_5| - |\phi_3 + \phi_6|$$

However, from equations of (28) to (41), we can determine left side for ϕ_1 which is;

$$\phi_1 > -1 + |\phi_2| + |\phi_4| + |\phi_5| + |\phi_3 + \phi_6|$$

therefore; Hence, $(\phi_1|\phi_2,\phi_3,\phi_4,\phi_5)$ **is uniformly distributed with upper and lower ranges which is** $\phi_1|\phi_2,\phi_3,\phi_4,\phi_5,\phi_6 \sim U(|\phi_2|+|\phi_4|+|\phi_5|+|\phi_3+\phi_6|-1, 1-|\phi_2|-|\phi_4|-|\phi_5|-|\phi_3+\phi_6|$ **where** $(i=2,3,4,5,6)$ **.Therefore the prior distribution for AR(6) is;**

$$\textbf{p =6}, \quad p(\phi_1,\phi_2,\phi_3,\phi_4,\phi_5,\phi_6) = \begin{cases} \frac{1}{32(2-2|\phi_2|-2|\phi_4|-2|\phi_5|-|\phi_3+\phi_6|-|\phi_3-\phi_6|)} & \Phi \in SCofAR(6) \\ 0 & otherwise \end{cases}$$

D Monthly temperature for Sheffield

Data of monthly temperature for Sheffield site were taken from Climate Research Unit, CRU-TS3.22 dataset (http://badc.nerc.ac.uk/data/cru/) **normalsize**

Y/M	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000	4	4.8	6.1	6.8	11.1	13.6	14.2	15.5	13.5	8.8	5.8	4.2
2001	2.3	3	3.9	6.1	11.5	12.9	16	15.6	12	12.1	6.6	2.4
2002	4.5	5.7	6.2	8.1	10.9	13.3	14.7	15.8	13.2	8.6	7.3	4.5
2003	3.6	2.9	6.5	8.8	11	14.9	16.3	16.7	13.2	7.9	7.1	3.7
2004	4.3	4.2	5.5	8.5	11.1	14.1	14.4	16.5	13.4	9.2	6.8	4.9
2005	5.1	3.1	6	7.6	10.3	14.4	15.5	15.1	13.7	11.5	5.5	4
2006	3.3	3	3.5	7.2	11	14.9	17.8	15	15.3	11.6	7.2	5.4
2007	5.8	4.7	5.8	9.9	10.8	14	14.5	14.8	12.9	10	6.7	4.1
2008	5.5	4.3	4.9	6.7	12.5	13.2	15.6	15.6	12.4	8.9	6.2	2.8
2009	2.4	3	5.9	8.6	11.1	13.9	15.5	15.7	12.9	10.3	7.3	2.3
2010	0.6	1.3	5	7.8	10	14.2	16.3	14.4	12.9	9.4	4.3	-1.1
2011	2.8	5.1	5.5	9.5	11.5	13.2	14.5	14.7	14	11.3	8.2	4.7
2012	4.3	3.2	7.2	6.2	10.5	12.6	14.5	15.6	12	8.2	5.8	3.7
2013	2.4	2	1.5	6.4	9.7	13	17.2	16	12.7	11.3	5.5	5.4