https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

Bayesian Analysis Influences Autoregressive Models

Evan Abdulmajeed Hasan

Erbil, Kurdistan region of Iraq
Evan.hasah@outlook.com

***Abstract*—** *The models, principles and steps of Bayesian time series analysis and forecasting have been established extensively during the past fifty 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. The researcher, additionally 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).*

***Keywords*— *Bayesian Analysis, Autoregressive Models, Time series, Stationarity*.**

I. INTRODUCTION

The importance of Bayesian methods in econometrics has increased rapidly over the last decade. This is, no doubt, fueled by an increasing appreciation of the advantages that Bayesian inference entails. In particular, it provides us with a formal way to incorporate the prior information we often possess before seeing the data, it fits perfectly with sequential learning and decision making and it directly leads to exact small sample results. In addition, the Bayesian paradigm is particularly natural for prediction, taking into account all parameter or even model uncertainty. The predictive distribution is the sampling distribution where the parameters are integrated out with the posterior distribution and is exactly what we need for forecasting, often a key goal of time-series analysis.

The class of autoregressive models is a rather general set of models largely used to represent stationary time series. However, time series often present change points in their dynamic structure, which may have a serious impact on the analysis and lead to misleading conclusions. A change point, which is generally the effect of an external event on the phenomenon of interest, may be represented by a change in the structure of the model or simply by a change of the value of some of the parameters.

There is an extensive literature on autoregressive processes using Bayesian methods. Bayesian analysis of AR models began with the work of Zellner and Tiao (1964) who considered the AR (1) process. Valdes-Sosa, et al. (2011), Box et al. (1976), discuss the Bayesian approach to analyze the AR models. Thomas, et al. (2018) developed a numerical algorithm to produce posterior and predictive analysis for AR (1) process. Diaz and Farah (1981) devoted a Bayesian technique for computing posterior analysis of AR process with an arbitrary order. Phillips (1991) discussed the implementation of different prior distributions to develop the posterior analysis of AR models with no stationarity assumption assumed. Harrison, et al. (2003) implement a Bayesian approach to determine the posterior

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

probability density function for the mean of a p th order AR model. Zhai, et al. (2018) introduced a comparative study to some selected noninformative (objective) priors for the AR (1) model. Ibazizen and Fellag (2003), assumed a noninformative prior for the autoregressive parameter without considering the stationarity assumption for the AR (1) model. However, most literature considers a noninformative (objective) prior for the Bayesian analysis of AR (1) model without considering the stationarity assumption.

An alternative way to model unobserved volatility is to use the so-called Conditional Heteroscadatic Auto Regressive Moving Average (CHARMA) model, which allows the coefficients of an ARMA model to be random. Note that the CHARMA models have second-order properties similar to that of ARCH models. However, the presence of random coefficients makes it harder to obtain the higher order properties of CHARMA models as compared to that of ARCH models. Another class of models that are very similar in spirit to CHARMA models is known as the Random Coefficient Auto Regressive (RCAR) models. Historically an RCAR model has been used to model the conditional mean of a time series, but it can also be viewed as a volatility model. Nicholls and Quinn, henceforth, NQ, studied the RCAR models from a frequentist view point. In particular NQ obtained the stationarity condition of the RCAR model for vector valued time series. In case of univariate time series data NQ obtained the Least Square (LS) estimates of the model parameters and showed that under suitable conditions the LS estimates are strongly consistent and obey the central limit theorem. In addition, by assuming the normality of the random coefficients and the error, NQ proposed an iterative method to obtain Maximum Likelihood (ML) estimates of the parameters. They showed that the ML estimates are strongly consistent and satisfy a central limit theorem even when the error processes are not normally distributed (Gao & Bradley, 2019).

Bayesian inference is influenced by the choice of prior. Elicitation of priors is an important step of Bayesian analysis. When researchers have information on the nature of parameters of interest, they may use informative priors to reflect their beliefs. For example, Doan et al. (1984) and Litterman (1986) observed that many macroeconomic time series approximately follow random walk processes and developed an informative prior known as the “Minnesota prior” that reflects the pattern. In recent Bayesian studies, Pastor (2000) and Pastor and Stambaugh (2000) used finance theory for elicitation of informative priors (Pavanato, et al. 2018).

A simulation experiment is performed to compare frequentist properties of inferences about the covariance parameters based on maximum likelihood (ML) with those based on the proposed Jeffreys priors and a uniform prior. It is found that frequentist properties of the above Bayesian procedures are better than those of ML. In addition, frequentist properties of the above Bayesian procedures are adequate and similar to each other in most situations, except when the mean of the observations is not constant, or the spatial association is strong. In these cases, inference about the ‘spatial parameter’ based on the independence Jeffreys prior has better frequentist properties than the procedures based on the other priors. Finally, it is found that the independence Jeffreys prior is not very sensitive to some aspects of the design, such as sample size and regression design matrix, while the Jeffreys-prior displays strong sensitivity to the regression design matrix (Letham, et al. 2015).

A traditional approach to lattice data uses continuously indexed geostatistical spatial process models after spatially aggregating over the areal units that define the lattice grid cells (or points). These approaches do not, however, scale well with lattice dimension. For example, deformation approaches, spatial moving-average models and other non-stationary covariance models incorporate spatial dependencies through the covariance structure of a Gaussian process. Bayesian posterior sampling algorithms for these models are prohibitively slow for large lattices as they require full (non-sparse) matrix inversions/Cholesky decompositions in every iteration. The computational cost is O (n 3) floating point operations (FLOPs) where n is the number of point realizations from the continuous spatial process. Typically, n is much larger than the lattice size since the regional aggregations use Monte Carlo integration with many points from each areal unit. Dimension reduction approximations are rarely applicable here, due of the need to integrate the resulting continuous process over the region (Jóhannesson, et al. 2016).

In the absence of pre-sample information, Bayesian VAR inference can be thought of as adopting ‘non-informative’ (or ‘diffuse’ or ‘flat’) priors, that express complete ignorance about the model parameters, in the light of the sample evidence summarized by the likelihood function (i.e. the probability density function of the data as a function of the parameters). Often, in such a case, Bayesian probability statements about the unknown parameters

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

(conditional on the data) are very similar to classical confidence statements about the probability of random intervals around the true parameters value. For example, for a VAR with Gaussian errors and a flat prior on the model coefficients, the posterior distribution is centered at the maximum likelihood estimator (MLE), with variance given by the variance-covariance matrix of the residuals (Sheriff, et al. 2018).

One of the key challenges of high-dimensional models is the complex interactions among variables and the inferential difficulty associated with handling large datasets. For instance, in large VAR models, econometricians encounter the curse of dimensionality problem due to high number of variables relative to the number of data points. The standard Bayesian VAR approach to this problem is to apply Minnesota prior by Doan et al. (1984), as a solution to overfitting. This approach is however inefficient to deal with the problem of indeterminacy, i.e. when the number of parameters in a system of equations exceeds the number of observations. Two common approaches to the indeterminacy issue discussed in the literature are based alternatively on dimension reduction or variable selection methodologies. For dimension reduction, dynamic factor models, factor augmented VAR and Bayesian model averaging have been extensively discussed and widely considered to extract useful information from a large number of predictors. For variable selection, standard techniques have been applied to reduce the number of predictors, e.g., the Least Absolute Shrinkage and Selection Operator (LASSO) of Tibshirani (1996), and its variants. The method considered in this paper is related to the latter, thus to variable selection. Variable selection is a fundamental problem in high-dimensional models, and this 2 is closely related to the possibility to describe the model with sparsity (Benavoli, et al. 2017). The idea of sparsity is associated with the notion that a large variation in the dependent variables is explained by a small proportion of predictors. Modeling sparsity has received attention in recent years in many fields, including econometrics (Moore, et al. 2016).

II. LITERATURE REVIEW

Time series

Many statistical methods relate to data which are independent, or at least uncorrelated. There are many practical situations where data might be correlated. This is particularly so where repeated observations on a given system are made sequentially in time. A time series is a sequential set of data points, measured typically over successive times. It is mathematically defined as a set of vectors x(t),t = 0,1,2,... where t represents the time elapsed. The variable) x(t is treated as a random variable. The measurements taken during an event in a time series are arranged in a proper chronological order. A time series containing records of a single variable is termed as univariate. But if records of more than one variable are considered, it is termed as multivariate. A time series can be continuous or discrete. In a continuous time, series observations are measured at every instance of time, whereas a discrete time series contains observations measured at discrete points of time. For example, temperature readings, flow of a river, concentration of a chemical process etc. can be recorded as a continuous time series. On the other hand, population of a particular city, production of a company, exchange rates between two different currencies may represent discrete time series. Usually in a discrete time series the consecutive observations are recorded at equally spaced time intervals such as hourly, daily, weekly, monthly or yearly time separations. As mentioned in, the variable being observed in a discrete time series is assumed to be measured as a continuous variable using the real number scale. Furthermore, a continuous time series can be easily transformed to a discrete one by merging data together over a specified time interval (Goligher, et al. 2018).

The methods of time series analysis pre-date those for general stochastic processes and Markov Chains. The aims of time series analysis are to describe and summaries time series data, fit low-dimensional models, and make forecasts. We write our real-valued series of observations as . . ., X−2, X−1, X0, X1, X2, . . ., a doubly infinite sequence of real-valued random variables indexed by Z. The impact of time series analysis on scientific applications can be partially documented by producing an abbreviated listing of the diverse fields in which important time series problems may arise. For example, many familiar time series occur in the field of economics, where we are continually exposed to daily stock market quotations or monthly unemployment figures. Social scientists follow population series, such as birthrates or school enrollments. An epidemiologist might be interested in the number of influenza cases observed over some time period. In medicine, blood pressure measurements traced over time could be useful for evaluating drugs used in treating hypertension. Functional magnetic resonance imaging of brain-wave time series patterns might be used to study how the brain reacts to

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

certain stimuli under various experimental conditions (Bergstroem, et al. 2015).

We represent time series measurements with Y1, . . ., YT where T is the total number of measurements. In order to analyze a time series, it is useful to set down a statistical model in the form of a stochastic process. A stochastic process can be described as a statistical phenomenon that evolves in time. While most statistical problems are concerned with estimating properties of a population from a sample, in time series analysis there is a different situation. Although it might be possible to vary the length of the observed sample, it is usually impossible to make multiple observations at any single time (for example, one can't observe today's mortality count more than once). This makes the conventional statistical procedures, based on large sample estimates, inappropriate. Stationarity is a convenient assumption that permits us to describe the statistical properties of a time series (Andrade, et al. 2018).

A time series model for the observed data {xt} is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables {Xt} of which {xt} is postulated to be a realization. Water is a very valuable commodity in any society, irrespective of the society's level of development. This resource is a key component in the provision of food and sanitary conditions, dong with economic development in areas such as industry and agriculture. With the approach of the twenty-first century, the world is faced with challenges, such as rapid population growth, increasing water demands, decreasing water quality, pollution and the associated health impacts, ground water depletion, conflict over shared water resources, and uncertainty over climate change. All of these challenges will ultimately affect water management. Therefore, as we move into the twenty-first century, efficient water resources management is of paramount importance in the quest for economic and social sustainability (Reefhuis, et al. 2015).

Reservoir management is important because it focuses on balancing present consumption with conservation for future consumption and on a smaller scale, it deals with controlling the supply and storage of water of a reservoir system. Storage in a reservoir is a function of the net inflow into the reservoir. Reservoir is generally used for temporary storage but one disadvantage to this use is the increased volume of water lost through seepage and evaporation, but the benefits gained from using a reservoir significantly outweighs the losses. The main advantage is the conservation of water because storage controls the consumption when it is needed, rather than when nature allows (Kruschke& Liddell, 2018).

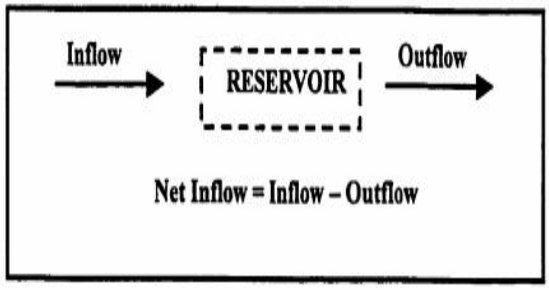

Fig.1: Flow chart for net inflow of a reservoir

Some problems incurred in reservoir management include financial, technological and human resources issues, such as water shortage, water pollution, floods, exceedance, lack of capital, fiscal failure, change of population and culture change. From an engineering point of view, one of the main problems experienced when attempting to efficiently manage the system is the lack of reliable information on the fixture water demand and the net inflow; since inflow is a natural phenomenon that cannot be predicted with certainty due to randomness (Dembo, et al. 2015).

In practice, the net Mow is determined using the water balance equation:

$$\Delta S = (I + P) - (O + O_g + E)$$

Where AS is the net storage, Le. the net inflow, 1 is the inflow, P is the precipitation, O is the surface outflow, 0, is the subsurface seepage and E is the evaporation.

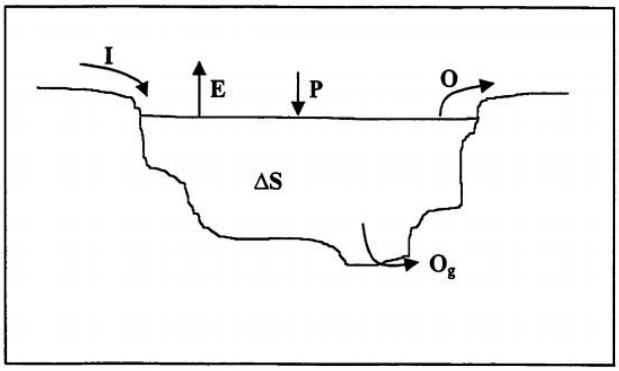

Fig.2: Reservoir Water Balance

This method of net inflow estimation has its limitations, which are its inability to obtain direct and reliable measurements of the parameters such as the subsurface flow, subsurface seepage and evaporation. As a result, net inflow is generality estimated by determining the mergence in water level in the reservoir. This is a very inaccurate method because it is not possible to accurately measure the

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

water level of a large water reservoir. Therefore, large errors are incurred, which results in inefficient planning and excessive cost. Hence improvements in forecasting techniques are needed in order to assist in reducing or alleviating these costs (Bierkens, et al. 2019).

Stationarity

Broadly speaking, a time series is said to be stationary if there is no systematic trend, no systematic change in variance, and if strictly periodic variations or seasonality do not exist. Most processes in nature appear to be non-stationary. Yet much of the theory in time-series literature is only applicable to stationary processes. Time series analysis is about the study of data collected through time. The field of time series is a vast one that pervades many areas of science and engineering particularly statistics and signal processing: this short article can only be an advertisement. Hence, the first thing to say is that there are several excellent texts on time series analysis. Most statistical books concentrate on stationary time series and some texts have good coverage of "globally non-stationary" series such as those often used in financial time series. For a general, elementary introduction to time series analysis the author highly recommends. The core of Chatfield's book is a highly readable account of various topics in time series including time series models, forecasting, time series in the frequency domain and spectrum estimation and also linear systems. More recent editions contain useful, well-presented and well-referenced information on important new research areas (Smitherman, et al. 2018).

It is a common assumption in many time series techniques. Time series observed in the practice are sometimes non-stationary. In this case, they should be transformed to some stationary time series, if possible, and then be analyzed.

Two types of stationarity exist: strong (or strict) and weak stationarity. Weak stationarity is sufficient for our purposes. A weak stationary process has the property that the mean, variance and autocovariance structure do not change over time. In mathematical terms:

$$E(X_t) = \mu \quad \text{for all } t$$

$$E(X_t^2) = \sigma^2 \quad \text{for all } t$$

$$\text{cov}(X_t, X_k) = \text{cov}(X_{t+s}, X_{k+s}) \quad \text{for all} \quad t, k, s$$

In other words, we mean flat looking series, without trend, with constant variance over time and with no periodic fluctuations (seasonality) or autocorrelation. There are several formal tests of stationarity; quite popular is Augmented Dickey-Fuller test.

A key idea in time series is that of stationarity. Roughly speaking, a time series is stationary if its behavior does not change over time. This means, for example, that the values always tend to vary about the same level and that their variability is constant over time. Stationary series have a rich theory and their behavior is well understood. This means that they play a fundamental role in the study of time series (Smitherman, et al. 2018).

A process { } x(t),t = 0,1, 2,... is Strongly Stationary or Strictly Stationary if the joint probability distribution function of {xt−s , xt−s+1 ,..., xt ,...xt+s−1 , xt+s} is independent of t for all s. Thus, for a strong stationary process the joint distribution of any possible set of random variables from the process is independent of time. However, for practical applications, the assumption of strong stationarity is not always needed and so a somewhat weaker form is considered. A stochastic process is said to be Weakly Stationary of order k if the statistical moments of the process up to that order depend only on time differences and not upon the time of occurrences of the data being used to estimate the moments. For example a stochastic process {x(t),t = 0,1, 2,...} is second order stationary if it has time independent mean and variance and the covariance values (,) t t s Cov x x − depend only on s.

It is important to note that neither strong nor weak stationarity implies the other. However, a weakly stationary process following normal distribution is also strongly stationary. Some mathematical tests like the one given by Dickey and Fuller are generally used to detect stationarity in a time series data. As mentioned in, the concept of stationarity is a mathematical idea constructed to simplify the theoretical and practical development of stochastic processes. To design a proper model, adequate for future forecasting, the underlying time series is expected to be stationary. Unfortunately, it is not always the case. As stated by Hipel and McLeod, the greater the time span of historical observations, the greater is the chance that the time series will exhibit non-stationary characteristics. However, for relatively short time span, one can reasonably model the series using a stationary stochastic process. Usually time series, showing trend or seasonal patterns are non-stationary in nature. In such cases, differencing and power transformations are often used to remove the trend and to make the series stationary (Aquino-López, et al. 2018).

Increased need to water resources with proper quality and quantity besides possessing temporal and spatial distribution adapted with operation needs required

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

engineers and investigators of water sources to make more efficient managerial systems for hydro-systems. It is needless telling that accuracy in predicting the streams of forthcoming periods has valuable effect on the efficiency of decision support systems for operating the reservoir. Prediction includes approximating the future situation of a parameter with four dimensions: quality, quantity, space and time [3]. Regarding to the statistics in Iran, it seems that time series models are acceptable variants for developing the flow prediction model. The basic theory for developing mentioned models is that the future is a reflection of past and any statistical relation that could be found in the historical statistics can be generalized to the future (Kruschke& Liddell, 2018).

The ARMA model

ARMA models have at least three generals, possibly overlapping uses in ecology. First, they can be used when a researcher has one or a few time series in hand and wants to investigate potential processes underlying their dynamics. When performing detailed analyses on the dynamics of a particular system, we generally advocate for a research approach involving mechanistic models tailored specifically for the system. Nonetheless, fitting a simple ARMA model might be useful as a first step, for example, identifying the lagged structure of the data (i.e., values of p and q). Because it is linear, the ARMA model is the simplest model that includes lagged effects in both densities and environmental (random) fluctuations. Although ecological time series are unlikely to be linear, by the Zhang, et al. (2018), representation theorem any stochastic process can be represented by a MA process that has identical statistical moments, and under mild restrictions a pure MA process can be written as an ARMA process. Therefore, although equation 1 is linear, it can nonetheless be used to approximate any nonlinear stochastic process. In practice, this might not be a useful result, because the MA process representation may be infinite (q ¼ ‘) and the number of lags required to well-approximate the dynamics with an ARMA process may be too large to allow practical model fitting. Even in situations of strongly nonlinear dynamics, however, fitting a linear ARMA model to the data may be valuable. For example, in a study investigating nonlinear dynamical phenomena such as chaos or alternative states, a best-fitting linear model can serve as a null hypothesis against which to compare the fits of nonlinear models. This provides a test for the existence of complex dynamics that cannot be well explained by linear processes. A second use for ARMA models is to give a quantitative estimate of some qualitative descriptor of dynamics. For example, an ARMA model could be used to estimate the mean and variance of the stationary distribution of a stochastic process, or some measure of the ‘‘stability’’ of the process. In this case, the quantity in question is a function of the ARMA coefficients bi and/or aj, and we are more interested in this function than the actual coefficients. We then judge our ability to fit the ARMA model to data based on the bias and precision of the estimates of this function rather than the bias and precision of the estimates of the specific coefficients. An informative summary measure of the dynamics of a system is its characteristic return time, or more precisely, the rate at which the stochastic process approaches its stationary distribution (i.e., the distribution that a process settles to after sufficient time). The characteristic return time gives a measure of the stability of the stochastic process, with greater stability corresponding to more rapid return to stationarity. A third use for ARMA models is to conduct broad surveys of multiple time-series data sets. When analyzing large numbers of time series from different sources and possibly heterogeneous systems (e.g., taxonomically diverse species), it is not practical to construct separate mechanistic, nonlinear models appropriate for each system. Instead, ARMA models can be fit to all time series, and the resulting fitted models used to compare them. In broad surveys, we are likely to be interested in functions of ARMA coefficients, like the characteristic return time discussed above, rather than the coefficients themselves. When comparing multiple data sets, the sample size is the number of data sets (rather than the number of points in any one data set). Therefore, we are more concerned about bias than precision. While we would like high precision in the individual estimates for each time series, any imprecision will merely make it harder to statistically infer patterns. Bias, on the other hand, could give us false results. The importance of bias over precision separates the use of ARMA models for broad surveys from the other two uses of ARMA models that we described above for which both bias and precision are important (Martín-Sánchez, et al. 2018).

The MA(q) average has the feature that after q lags there isn’t any correlation between two random variables. On the other hand, there are correlations at all lags for an AR(p) model. In addition, as we shall see later on, it is much easier to estimate the parameters of an AR model than an MA. Therefore, there are several advantages in fitting an AR model to the data (note that when the roots are of the characteristic polynomial lie inside the unit circle, then the

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

AR can also be written as an MA (∞), since it is causal). However, if we do fit an AR model to the data, what order of model should we use? Usually one uses the AIC (BIC or similar criterion) to determine the order. But for many data sets, the selected order tends to be relatively large, for example order 14. The large order is usually chosen when correlations tend to decay slowly and/or the autocorrelations structure 96 is quite complex (not just monotonically decaying). However, a model involving 10-15 unknown parameters is not particularly parsimonious and more parsimonious models which can model the same behavior would be useful. A very useful generalization which can be more flexible (and parsimonious) is the ARMA (p, q) model, in this case Xt satisfies

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

Bayesian Analysis

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. The prior distribution gives a summary of everything that is obtained about parameter, except the data; moreover, in the Bayesian approach, conclusions are normally reached using probability statements (Aktekin, et al. 2018).

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. In Bayesian inference, the model for the observed quantity y = (y1,y2,...,yn)T is defined via a vector of unknown parameters φ using a probability distribution p(y|φ) where it is assumed that φ is a random quantity having a prior distribution p(φ). Thus, inference about φ is based on its posterior density p(φ|y), given by:

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

where p(y) =Pp(φ)p(y|φ) if φ is a discrete random variable and p(y) =Rp(φ)p(y|φ)dφ in the continuous case. Equation may then be stated in a proportional form:

$$p(\phi|y) \propto p(\phi)p(y|\phi)$$

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

A Bayesian analysis starts by choosing some values for the prior probabilities. We have our two competing hypotheses BB and BW, and we need to choose some probability values to describe how sure we are that each of these is true. Since we are talking about two hypotheses, there will be two prior probabilities, one for BB and one for BW. For simplicity, we will assume that we don't have much of an idea which is true, and so we will use the following prior probabilities (Heavens &Sellentin, 2018):

$$P(\mathrm{BB}) = 0.5 \quad (3.1)$$
$$P(\mathrm{BW}) = 0.5. \quad (3.2)$$

Pay attention to the notation. The upper-case P stands for probability, and if we just write P(whatever), that means we are talking about the prior probability of whatever. We will see the notation for the posterior probability shortly. Note also that since the two hypotheses are mutually exclusive (they can't both be true) and exhaustive (one of these is true, it can't be some undefined third option). We will almost always consider mutually exclusive and exhaustive hypotheses.

There are three essential ingredients underlying Bayesian statistics first described by T. Bayes in 1774 (da Costa, et al. 2018). Briefly, these ingredients can be described as follows (these will be explained in more detail in the following sections).

- The first ingredient is the background knowledge on the parameters of the model being tested. his first ingredient refers to all knowledge available before seeing the data and is captured in the so-called prior distribution, for example, a normal distribution. The variance of this prior distribution reflects our level of uncertainty about the population value of the parameter of interest: The larger the variance, the more uncertain we are. The prior variance is expressed as precision, which is simply the inverse of the variance. The smaller the prior variance, the higher the precision, and the more confident one is that the prior mean reflects the population mean. In this study we will vary the specification of the prior distribution to evaluate its influence on the final results.

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

- The second ingredient is the information in the data themselves. It is the observed evidence expressed in terms of the likelihood function of the data given the parameters. In other words, the likelihood function asks: Given a set of parameters, such as the mean and/or the variance, what is the likelihood or probability of the data in hand?
- The third ingredient is based on combining the first two ingredients, which is called posterior inference. Both (1) and (2) are combined via Bayes' theorem (and are summarized by the so-called posterior distribution, which is a compromise of the prior knowledge and the observed evidence. The posterior distribution reflects one's updated knowledge, balancing prior knowledge with observed data.

Bayesian statistics is motivated by Bayes' Theorem, named after Thomas Bayes (1701-1761). Bayesian statistics takes prior knowledge about the parameter and uses newly collected data or information to update our prior beliefs. Furthermore, parameters are treated as unknown random variables that have density of mass functions. The process starts with a 'prior distribution' that reflects previous knowledge or beliefs. Then, similar to frequentist, we gather data to create a 'likelihood' model. Lastly, we combine the two using Bayes' Rule to achieve a 'posterior distribution'. If new data is gathered, we can then use our posterior distribution as a new prior distribution in a new model; combined with new data, we can create a new posterior distribution (Etherton, et al. 2018).

In recent years, Bayesian approach has been widely applied to clinical trials, research in education and psychology, and decision analyses. However, some statisticians still consider it as an interesting alternative to the classical theory based on relative frequency. These frequentists argue that the introduction of prior distributions violates the objective view point of conventional statistics. Interestingly, the feature of applying priors is also the reason why Bayesian approach is superior to frequentist. The following table briefly summarizes the differences between frequentist and Bayesian approaches. Then, I simply list the cons and pros of Bayesian statistics and suggest situations to which Bayesian statistics is more applicable (Thapa, et al. 2018).

	Frequentist	Bayesian
parameter of model	• fixed, unknown constants • can NOT make probabilistic statements about the parameters	• random variables (parameters can't be determined exactly, uncertainty is expressed in probability statements or distributions) • can make probability statements about the parameters
probability	objective, relative frequency	subjective, degree of belief
main outcomes	point estimates with standard error	posterior distribution
estimate/inference	use data to best estimate unknown parameters	• pinpoint a value of parameter space as well as possible by using data to update belief • all inference follow posterior • use simulation method: generate samples from the posterior and use them to estimate the quantities of interest
interval estimate	Confidence Interval: a claim that the region covers the true parameter, reflecting uncertainty in sampling procedure. e.g: 95%CI=(a, b) implies the interval (a, b) covers the true parameter among 95% of the experiments	Credible Interval: a claim that the true parameter is inside the region with measurable probability. One can make a direct probability statement about parameters. e.g: 95%CI=(a, b) implies the chance that the true parameter falls in (a, b) is 95%.

III. METHODS AND FINDINGS

Applications

Stationary AR processes

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

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

where φi are the parameters of the AR(p) model and εt is white noise (εt is iid with 0mean and some variance). Then, yt is a stationary process if and only if the roots of φ(z) = 1−φ1z−···−φpzp lie outside the unit circle. This proof is different from the previous proofs that have been done

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

through the polynomials of φ(z) for ARMA(p,q) as can be seen in (Brockwell and Davis, 2001). Let Xt be a vector of time series following the vector AR model of first order;

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

MCMC methods for autoregressive Models

As stated earlier a major focus of our project is on the estimation of autoregressiveparameters through MCMC methods. As is established in the literature (BarndorffNielsen and Schou (1973) and Huerta and West (1999)) a prior distribution on parametersor transformations of them, must respect the requirement of stationarity which imposesconditions on those parameters. Thus, in this section, we define a new prior distributionplaced directly on the AR parameters. We go on to propose suitable MCMC schemes forestimation. This is achieved through information obtained on the stationary conditions for the AR(p) model, forrelatively 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 takeadvantage 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. MCMChas been applied on a set of simulated data; the data have been simulated on the basisof an AR model.

Assume n observations are available, say y1,y2,...,yn . The aim is to estimate the unknown parameters of φ and σ2 . We use the AR(1) model yt = φyt−1 + εt where εt is white noise and εt∼ N(0,σ2). 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 1 σ2 (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,σ2).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

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

IV. RESULTS

This section mainly focuses on two parts in order to study and compare the current proposalwith previous studies relevant to the present study. The first part compares the proposed priordistribution with the prior distributions obtained from the correspondence relationship betweenpartial autocorrelations and parameters discussed by Barndorff-Nielsen and Schou (1973). Itdiscusses the study by Jones (1987) in which the author generalized a Jacobian transformationbased 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 parametersusing the Jacobian transformation in the case high order of polynomial models. This is discussedin this section. This comparison relies on some theoretical mathematical steps and practicalresults 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 partialautocorrelation and proposed a Metropolis Hastings algorithm. Considering the same priors, wedevelop a Gibbs sampling algorithm which is easier and more routine to apply. The purpose ofthe second part is to apply the proposed MCMC scheme of section to both real data and sim ulated data. Furthermore, this part compares the performance of the above MCMC algorithmwith Box et al. (1976) as well as with the Gibbs sampling scheme of the previous section.

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(φ) = 1 2I[−1,1] which is based on |φ| < 1 and φ ∼

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

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:

$$\phi_1 = \pi_1 - \pi_1\pi_2$$

$$\phi_2 = \pi_2$$

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 π1,π2 are on (-1, 1), we can propose that π1,π2 are uniformly
distributed on (-1, 1) and are independent. Hence, the prior distribution for the AR(2) model is

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

Prior distribution of the AR(p) model when p ≥ 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 ≥ 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 πi are between (-1, 1) then we have used the distribution of πi is uniformly distributed, πi~U(−1,1), for i = 1,2,3 and p(πi) = 1 2I[−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:

$$p(\pi_1,\pi_2,\pi_3) = p(\pi_1).p(\pi_2).p(\pi_3) = \left(\frac{1}{2}\right)^3 I_{[-1,1]}$$

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 {π(φi)}−1 can be transformed into an equation of φ(πi) and finding the Jacobian as follows:

$$p(\phi_1,\phi_2,\phi_3) = p(\pi_1,\pi_2,\pi_3).|J|$$

A set of observations are simulated in order to obtain parameter estimates of the AR(3) modelusing our new MCMC proposal presented earlier.
with Box et al. (1976)'s method. The simulated data from Section 4.12 with true AR values of φ1 = −0.4, φ2 = −0.8 and φ3 = −0.6 were used for this comparison. The obtained parameter estimates for the AR (3) model using our proposal are φ1 = −0.414, φ2 = −0.792 and φ3 = −0.599 with errors of 1.4%, 0.8% and 0.1%, respectively. The results obtained using Box et al. (1976) are φ1 = −0.367, φ2 = −0.784 and φ3 = −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 valueswhen compared to maximum likelihood estimates (Box et al., 1976); leading to smaller residuals(errors).

V. 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 station arity 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

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

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 station arity 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.

REFERENCES

[1] Aktekin, T., Dutta, D. K., &Sohl, J. E. (2018). Entrepreneurial firms and financial attractiveness for securing debt capital: a Bayesian analysis. *Venture Capital*, *20*(1), 27-50.

[2] Andrade, U., Bengaly, C. A. P., Alcaniz, J. S., & Santos, B. (2018). Isotropy of low redshift type Ia supernovae: A Bayesian analysis. *Physical Review D*, *97*(8), 083518.

[3] Aquino-López, M. A., Blaauw, M., Christen, J. A., & Sanderson, N. K. (2018). Bayesian Analysis of $$^{} $$ Pb Dating. *Journal of Agricultural, Biological and Environmental Statistics*, *23*(3), 317-333.

[4] Benavoli, A., Corani, G., Demšar, J., &Zaffalon, M. (2017). Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis. *The Journal of Machine Learning Research*, *18*(1), 2653-2688.

[5] Bergstroem, J., Gonzalez-Garcia, M. C., Maltoni, M., &Schwetz, T. (2015). Bayesian global analysis of neutrino oscillation data. *Journal of High Energy Physics*, *2015*(9), 200.

[6] Bierkens, J., Fearnhead, P., & Roberts, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *The Annals of Statistics*, *47*(3), 1288-1320.

[7] da Costa, S. S., Benetti, M., &Alcaniz, J. (2018). A Bayesian analysis of inflationary primordial spectrum models using Planck data. *Journal of Cosmology and Astroparticle Physics*, *2018*(03), 004.

[8] Dembo, M., Matzke, N. J., Mooers, A. Ø., & Collard, M. (2015). Bayesian analysis of a morphological supermatrix sheds light on controversial fossil hominin relationships. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20150943.

[9] Etherton, J. L., Osborne, R., Stephenson, K., Grace, M., Jones, C., & De Nadai, A. S. (2018). Bayesian analysis of multimethod ego-depletion studies favours the null hypothesis. *British Journal of Social Psychology*, *57*(2), 367-385.

[10] Gao, H., & Bradley, J. R. (2019). Bayesian analysis of areal data with unknown adjacencies using the stochastic edge mixed effects model. *Spatial Statistics*, 100357.

[11] Goligher, E. C., Tomlinson, G., Hajage, D., Wijeysundera, D. N., Fan, E., Jüni, P., ... & Combes, A. (2018). Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome and posterior probability of mortality benefit in a post hoc Bayesian analysis of a randomized clinical trial. *Jama*, *320*(21), 2251-2259.

[12] Harrison, L., Penny, W. D., &Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage*, *19*(4), 1477-1491. Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., &Friston, K. (2011). Effective connectivity: influence, causality and biophysical modeling. *Neuroimage*, *58*(2), 339-361.

[13] Heavens, A. F., &Sellentin, E. (2018). Objective Bayesian analysis of neutrino masses and hierarchy. *Journal of Cosmology and Astroparticle Physics*, *2018*(04), 047.

[14] Jóhannesson, G., de Austri, R. R., Vincent, A. C., Moskalenko, I. V., Orlando, E., Porter, T. A., ... & Hobson, M. P. (2016). Bayesian analysis of cosmic ray propagation: evidence against homogeneous diffusion. *The Astrophysical Journal*, *824*(1), 16.

[15] Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic bulletin & review*, *25*(1), 155-177.

https://dx.doi.org/10.22161/ijebm.3.3.2 *ISSN: 2456-7817*

[16] Letham, B., Rudin, C., McCormick, T. H., & Madigan, D. (2015). Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics*, *9*(3), 1350-1371.

[17] Martín-Sánchez, J. C., Lunet, N., González-Marrón, A., Lidón-Moyano, C., Matilla-Santander, N., Clèries, R., ... & Ferro, A. (2018). Projections in breast and lung cancer mortality among women: A Bayesian analysis of 52 countries worldwide. *Cancer research*, *78*(15), 4436-4442.

[18] Moore, B. R., Höhna, S., May, M. R., Rannala, B., &Huelsenbeck, J. P. (2016). Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. *Proceedings of the National Academy of Sciences*, *113*(34), 9569-9574.

[19] Pavanato, H. J., Mayer, F. P., Wedekin, L. L., Engel, M. H., & Kinas, P. G. (2018). Prediction of humpback whale group densities along the Brazilian coast using spatial autoregressive models. *Marine Mammal Science*, *34*(3), 734-754.

[20] Reefhuis, J., Devine, O., Friedman, J. M., Louik, C., &Honein, M. A. (2015). Specific SSRIs and birth defects: bayesian analysis to interpret new data in the context of previous reports. *bmj*, *351*, h3190.

[21] Sheriff, S. L., Sun, D., He, Z., Vangilder, L. D., & Isabelle, J. L. (2018). Model selection for wild turkey hunter success rates using small-area estimation methods. *Wildlife Society Bulletin*, *42*(4), 622-631.

[22] Smitherman, T. A., Kuka, A. J., Calhoun, A. H., Walters, A. B. P., Davis-Martin, R. E., Ambrose, C. E., ... & Houle, T. T. (2018). Cognitive-behavioral therapy for insomnia to reduce chronic migraine: A sequential Bayesian analysis. *Headache: The Journal of Head and Face Pain*, *58*(7), 1052-1059.

[23] Thapa, S., Lomholt, M. A., Krog, J., Cherstvy, A. G., & Metzler, R. (2018). Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data. *Physical Chemistry Chemical Physics*, *20*(46), 29018-29037.

[24] Thomas, Z. M., MacEachern, S. N., &Peruggia, M. (2018). Reconciling Curvature and Importance Sampling Based Procedures for Summarizing Case Influence in Bayesian Models. *Journal of the American Statistical Association*, *113*(524), 1669-1683.

[25] Zhai, X., Huang, H., Xu, P., & Sze, N. N. (2018). The influence of zonal configurations on macro-level crash modeling. *Transportmetrica A: Transport Science*, 1-18.

[26] Zhang, Y., Tian, L., Sleiman, P., Ghosh, S., &Hakonarson, H. (2018). Bayesian analysis of genome-wide inflammatory bowel disease data sets reveals new risk loci. *European Journal of Human Genetics*, *26*(2), 265.