Afrika Statistika

Afrika Statistika
Vol. 14 (1), 2019, pages 1877 – 1889.
DOI: http://dx.doi.org/10.16929/as/2019.1877.139

ISSN 2316-090X

A Bayesian Approach for Identification of Additive Outlier in AR(p) Model

Jitendra Kumar[1,*,#] **and Saurabh Kumar**[1,*]

[1] Department of Statistics, Central University of Rajasthan, NH-8, Kishangarh, Ajmer, Rajasthan, 305817, India

Received on September 09, 2018. Accepted on February 21, 2019

Abstract. Time series is the way of data analysis and modelling in which present observation is retrieved based on past observations which is called ARIMA model in case of linear dependency. If series is contaminated by an outlier, then it affects both order and parameter(s). The present paper deals an autoregressive (AR) model with an additive outlier under Bayesian prospective. For identification of an outlier, posterior odds ratio has been derived under suitable prior assumptions. An empirical analysis and realization is carried out to get applicability of proposed testing methodology.

Résumé (French) Dans l'étude des séries temporelles, les données discrètes sont modelisées par rapport aux observations passeés, et les modèles sont appelées ARIMA dans le cas de dependances linéaires. Si la série est contaminée par un outlier, les paramètres et les valeurs sont à lois affectées. Ce papier traite du modèle autoregressif avec un outlier additionnel selon une perspective bayesienne. Pour identifier un outlier, le rapports des odds a été obtenu après voir convenablement choisi les distributions à prioiri. Une étude empirique et des études de cas sont menées pour prouver l'applicabibilité de la méthodologie utilisée.

Key words: Autoregressive model, posterior odds ratio, prior distribution.
AMS 2010 Mathematics Subject Classification : 62F03, 62F15, 62M10

* Both authors contributed equally
# **Corresponding Authour**
Email: Jitendra Kumar: vjitendrav@gmail.com, Saurabh Kumar: sonysaurabh123@gmail.com

Acknowledgement. This work was supported by Council of Scientific & Industrial Research (CSIR), India, [No. 25(0198)/11/EMR-II, 2013], and SHIATS Allahabad for the permission to carry out the project.

1. Introduction

In practice, a time series may have abnormal observations due to various reasons and not alike with the most of the observations. Such abnormal observations are called outlier which affects both order and parameter(s). In such situation, before the analysis starts it is important to handle outlier to get better understanding of data generation process and know the impacts also. One of the most important purposes of time series modelling is to forecast the future observations. The autoregressive model follows principle of dependency that the present observation linearly depends on previous observations (see, Box and Jenkins [1970]).

Outliers are classified into two categories as per the appearance in the model of the series: first model incorporates the error in the form of addition, called as additive outlier(s) and if the error is in the form of multiplication, called as multiplicative outlier(s). The dealing of the outlier(s) is (are) done by researchers in two directions first identification and second study of its affects. Outlier identification in autoregressive model was first studied by Fox [1972] using the likelihood criteria in case of known number of outlier(s). He also distinguished between additive and multiplicative outlier and showed that additive outlier should be taken more seriously than multiplicative outlier. Chang and Tiao [1983] and Chang et al. [1988] discussed the identification of both types of outliers in an autoregressive integrated moving average (ARIMA) using likelihood ratio criteria and also estimated their parameters. Barnett and Lewis [1994] proposed a test for detecting the outlier under certain conditions like type of outliers with known distribution. Battaglia and Orfei [2005] discussed a method to identify position of the outlier and then estimated magnitude of outlier in case of non-linear time series. In Bayesian approach, Abraham and Box [1979] studied the characterization of the outlier in autoregressive time series model and obtained the posterior probability. Gordon [1986] and Gordon and Smith [1988] examined the sensitivity of models for both the changes in prior and model misspecification. Gordon and Smith [1990] developed a model based extension of linear dynamic model and tested it for identification of discontinuous change in the series. Tsay [1988] and Barnett et al. [1996] discussed additive and multiplicative outliers under Bayesian framework and defined the testing procedure to identify it's region in a subset of a series (see, Siegmund et al. [2011] and Jeng et al. [2013]). Silva and Pereira [2015] proposed Bayesian approach to detect additive outliers in Poisson integer-valued AR(1) time series model. Kumar et al. [2014] proposed the identification of additive outlier for a stationary AR(1) time series with intercept term which was further extended for linear time trend (Kumar and Shukla [2015]).

The outlier is attracting the researchers because of its impact on various statistical theories as well as in applications. Usually, there are very limited and less number of observations which may act as an outlier. In independent data set, the exclusion of these observations do not affect the testing and estimation results. However, this exclusion is not possible in dependent data like time series. Since last two decades, handling of individual observations has become more popular because of data availablity and comfort with the use of high performance computation system and software. Therefore, present paper deals with Bayesian analysis of AR(p) time series model contaminated by an additive outlier. The posterior odds ratio is derived for identification of outlier by using the prior assumptions (see, Schotman and Van Dijk. [1991] and Kumar and Shukla [2015]). An empirical analysis is carried out for the series recorded on patient of Haryana state, India. In order to get more general idea, a simulation study has been considered by realizing the data, where data is generated using seed values of real data. The paper is organized in five sections, present section gives the discussion on literature, followed Section 2 which gives a brief overview of autoregressive time series model. Section 3 dwells on prior assumptions and construction of posterior odds ratio. Section 4 and Section 5 demonstrate the proposed testing methodology on real and simulated time series respectively. The last Section concludes the importance of the work and possible extension.

2. Model Specification

Outlier is the observation occurring in series due to some unavoidable causes and affects the structure of the model. Let us consider that y_t follows an autoregressive model of order p (AR(p));

$$y_t = \theta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \cdots + \alpha_p y_{t-p} + u_t, \quad t = 1, 2, 3, \cdots, T. \tag{1}$$

We rewrite the Model (1) in matrix notation

$$y_t = \theta + \alpha Y_{t-1} + u_t, \tag{2}$$

where θ is the intercept term, $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_p)$ are autoregressive coefficients, u_t is the error term and $Y_{t-1} = (y_0, y_{-1}, y_{-2}, \cdots, y_{t-p})'$ are initial observations.

If time series model is contaminated by additive outlier at time point T_0, then error (u_t) is partitioned into two parts as follows: (i) error from standard error distribution and (ii) error with outlier,

$$u_t = \begin{cases} \varepsilon_t; & if\, t \neq T_0, \\ \varepsilon_t + \lambda e_t; & if\, t = T_0 \;\; and \;\; \lambda > 0. \end{cases} \tag{3}$$

Here, it is noted that ϵ_t and e_t will be in the same direction as outlier deviates more from the estimated observations. We may write Model (2) incorporating (3) as follows

$$y_t = \begin{cases} \theta + \alpha Y_{t-1} + \varepsilon_t; & if\ t \neq T_0, \\ \theta + \alpha Y_{t-1} + \lambda e_t + \varepsilon_t; & if\ t = T_0 \;\; ; \lambda > 0 \quad and \quad \varepsilon_t \sim N(0, \tau^{-1}). \end{cases} \tag{4}$$

In the present study, we are interested to know that this large deviation at time T_0 is due to outlier or not. Equivalent hypotheses are as follows:

$$\begin{array}{ll} \text{Null hypothesis} & H_0 : \lambda = 0. \\ \text{Alternative hypothesis} & H_1 : \lambda > 0. \end{array} \quad (5)$$

Here null hypothesis says that series is not contaminated by an additive outlier against the alternative that series is contaminated by an additive outlier.

3. Prior Distribution and Posterior Odds Ratio

Let us consider the following prior distribution of model parameters discussed by Broemeling [1984]; Schotman and Van Dijk. [1991] and Kumar and Shukla [2015].

$$\begin{aligned} \lambda &\sim N(\lambda_0, \tau^{-1}\omega); \; -\infty > \lambda_0 > \infty, \omega > 0, \\ \theta &\sim N(y_0, \tau^{-1}); \; -\infty > y_0 > \infty, \\ P(\tau) &= \frac{b^a}{\Gamma a}\tau^{a-1}e^{-\tau b}; \; a > 0; \; b > 0, \\ P(\alpha\,|\tau) &= \frac{\tau^{P/2}\,|P|^{1/2}}{(2\pi)^{P/2}} \exp\left\{-\frac{\tau}{2}\,(\alpha-\mu)'\,P\,(\alpha-\mu)\right\}; \; P > 0, \mu \in R^p, \end{aligned}$$

Using the above priors, we obtain the joint prior distribution $\pi(\Theta)$.

$$\begin{aligned} \pi(\Theta) &= P(\alpha/\tau)P(\lambda)P(\theta)P(\tau) \\ &= \frac{\tau^{\frac{p}{2}+a}\,b^a\,|P|^{1/2}}{\Gamma a\,(2\pi)^{\frac{p+2}{2}}\,\omega^{1/2}} \exp\left[-\frac{\tau}{2}\left\{(\alpha-\mu)'P(\alpha-\mu) + \frac{(\lambda-\lambda_0)^2}{\omega}\right.\right. \\ &\qquad \left.\left. +(\theta-y_0)^2 + 2b\right\}\right]. \quad (6) \end{aligned}$$

Main interest of present study is to derive the posterior odds ratio for the identification of additive outlier. Bayesian techniques are used with the prior belief for obtaining the posterior which is also a probability that combines samples and prior information. So, the liking of alternative and null hypothesis is core behind the testing methodology of under Bayesian setup. Therefore, posterior odds ratio (POR) is used for testing the hypothesis by comparison of probabilities. If probability of null hypothesis is more than alternative hypothesis under the assumed prior probability, i.e., POR is more than one, then null hypothesis is accepted otherwise rejected and vice versa (Bansal [2007]).

Theorem 1. *An abnormal increase/decease in magnitude of error in an AR(p) model at a time point T_0 is due to additive outlier or not. This can be identified by testing the hypothesis, i.e., $\{H_0 : \lambda = 0 \; V_s \; H_1 : \lambda > 0\}$ by using the posterior odds ratio (β_{01}) with prior odds ratio $\frac{P_0}{1-P_0}$.*

$$\beta_{01} = \frac{P_0}{1-P_0}\frac{\omega^{\frac{1}{2}}I^{\frac{1}{2}}\eta_{H_1}^{\frac{1}{2}}\xi_{H_1}^{\frac{1}{2}}\left[\Omega_{H_1}\right]^{\frac{T}{2}+\alpha}}{\eta_{H_0}^{\frac{1}{2}}\xi_{H_0}^{\frac{1}{2}}\left[\Omega_{H_0}\right]^{\frac{T}{2}+\alpha}}, \quad (7)$$

where,

$$\mathrm{I} = \left(\sum_{t=1}^{T} e_t^2 + \frac{1}{\omega}\right),$$

$$\eta_{H_1} = \left(T + 1 - \frac{1}{\mathrm{I}} \times \left(\sum_{t=1}^{T} e_t\right)^2\right),$$

$$\Phi = \sum_{t=1}^{T} Y_{t-1} - \frac{1}{I}\sum_{t=1}^{T} e_t Y_{t-1} \sum_{t=1}^{T} e_t,$$

$$\xi_{H_1} = \sum_{t=1}^{T} Y_{t-1}^{'} Y_{t-1} + P - \frac{1}{\mathrm{I}}\left(\sum_{t=1}^{T} e_t Y_{t-1}\right)^{'}\left(\sum_{t=1}^{T} e_t Y_{t-1}\right) - \frac{1}{\eta_{H_1}}\Phi'\Phi,$$

$$\Lambda = \sum_{t=1}^{T} y_t + y_0 - \frac{1}{I}\frac{\lambda_0}{\omega}\sum_{t=1}^{T} e_t - \frac{1}{I}\sum_{t=1}^{T} e_t Y_{t-1}\sum_{t=1}^{T} e_t,$$

$$\hat{\alpha}_{H_1} = \frac{1}{\xi_{H_1}}\left[\sum_{t=1}^{T} y_t Y_{t-1} + P\mu - \frac{1}{\eta_{H_1}}\Phi'\Lambda - \frac{1}{\mathrm{I}}\left(\sum_{t=1}^{T} e_t Y_{t-1}\right)\left(\frac{\lambda_0}{\omega} + \sum_{t=1}^{T} e_t y_t\right)\right],$$

$$\Omega_{H_1} = \sum_{t=1}^{T} y_t^2 - \frac{1}{\eta_{H_1}}\Lambda'\Lambda - \hat{\alpha}'_{H_1}\xi_{H_1}\hat{\alpha}_{H_1} + \mu' P\mu + y_0^2 - \frac{1}{\mathrm{I}}\left(\frac{\lambda_0}{\omega} + \sum_{t=1}^{T} e_t y_t\right)^2 + \frac{\lambda_0^2}{\omega} + 2b,$$

$$\eta_{H_0} = (T+1),$$

$$\xi_{H_0} = \sum_{t=1}^{T} Y_{t-1}^{'} Y_{t-1} + P - \frac{1}{\eta_{H_0}}\left(\sum_{t=1}^{T} Y_{t-1}\right)^{'}\left(\sum_{t=1}^{T} Y_{t-1}\right),$$

$$\hat{\alpha}_{H_0} = \frac{1}{\xi_{H_0}}\left(\sum_{t=1}^{T} y_t Y_{t-1} + P\mu - \frac{1}{\eta_{H_0}}\left(y_0 + \sum_{t=1}^{T} y_t\right)\left(\sum_{t=1}^{T} Y_{t-1}\right)\right),$$

and

$$\Omega_{H_0} = \sum_{t=1}^{T} y_t^2 + \mu' P\mu + 2b + y_0^2 - \frac{1}{\eta_{H_0}}\left(y_0 + \sum_{t=1}^{T} y_t\right)^2 - \hat{\alpha}_{H_0}^{'}\xi_{H_0}\hat{\alpha}_{H_0}. \tag{8}$$

Proof. The Likelihood function under H_1 is given by

$$\begin{aligned} L(\theta, \alpha, \tau, \lambda\,|y) &= L(y_1, y_2, \cdots, y_T\,|\theta, \lambda,\ \alpha, \tau) \\ &= \frac{\tau^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}}\exp\left\{-\frac{\tau}{2}\sum_{t=1}^{T}(y_t - \theta - \alpha Y_{t-1} - \lambda e_t)^2\right\}, \end{aligned} \tag{9}$$

Similarly, likelihood function under H_0 is given by

$$\begin{aligned} L(\theta, \alpha, \tau\,|y) &= L(y_1, y_2, \cdots, y_T\,|\theta,\ \alpha, \tau) \\ &= \frac{\tau^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}}\exp\left\{-\frac{\tau}{2}\sum_{t=1}^{T}(y_t - \theta - \alpha Y_{t-1})^2\right\}, \end{aligned} \tag{10}$$

Case 1: Under the alternative hypothesis, model is contaminated by an additive outlier, i.e, $\lambda > 0$. We get the posterior probability by combining the likelihood function (9) and joint prior distribution (6).

$$\begin{aligned} P(y|H_1) &= \int_0^\infty \int_{R^p} \int_{-\infty}^\infty \int_{-\infty}^\infty L(y\,|\theta,\ \alpha,\tau,\lambda)\,P(\lambda)\,P(\theta|\tau)\,P(\alpha|\tau)\,P(\tau)\,d\lambda\,d\theta\,d\alpha\,d\tau \\ &= \int_0^\infty \int_{R^p} \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{\tau^{\frac{T+p}{2}+a}\,b^a\,|P|^{\frac{1}{2}}}{\Gamma a\,(2\pi)^{\frac{T+p+2}{2}}\,\omega^{\frac{1}{2}}} \exp\left[-\frac{\tau}{2}\left\{\sum_{t=1}^T (y_t-\theta-\alpha Y_{t-1}-\lambda e_t)^2\right.\right. \\ &\qquad \left.\left. +(\alpha-\mu)'P(\alpha-\mu)+\frac{(\lambda-\lambda_0)^2}{\omega}+(\theta-y_0)^2+2b\right\}\right]\,d\lambda\,d\theta\,d\alpha\,d\tau \end{aligned} \tag{11}$$

Let us consider $\hat{\lambda}_{H_1} = \frac{1}{I}\left(\frac{\lambda_0}{\omega} + \sum_{t=0}^T e_t\,(y_t-\theta-\alpha Y_{t-1})\right)$ and integrate Equation (11) with respect to λ,

$$\begin{aligned} P(y|H_1) &= \int_0^\infty \int_{R^p} \int_{-\infty}^\infty \frac{\tau^{\frac{T+p-1}{2}+a}\,b^a\,|P|^{\frac{1}{2}}}{\Gamma a\,(2\pi)^{\frac{T+p+1}{2}}\,\omega^{\frac{1}{2}}\,I^{\frac{1}{2}}} \exp\left[-\frac{\tau}{2}\left\{\sum_{t=1}^T (y_t-\theta-\alpha Y_{t-1})^2\right.\right. \\ &\quad \left.\left. +(\alpha-\mu)'P(\alpha-\mu)+\frac{\lambda_0^2}{\omega}+(\theta-y_0)^2-\hat{\lambda}^2 I+2b\right\}\right]\,d\theta\,d\alpha\,d\tau \end{aligned}$$

Considering $\hat{\theta}_{H_1} = \frac{1}{\eta_{H_1}}\,(\Lambda-\alpha\Phi)$ and integrating the above equation with respect to θ, α, and τ, we get

$$\begin{aligned} P(y|H_1) &= \int_0^\infty \int_{R^p} \frac{\tau^{\frac{T+p}{2}+a-1}\,b^a\,|P|^{\frac{1}{2}}}{\Gamma a\,(2\pi)^{\frac{T+p}{2}}\,\omega^{\frac{1}{2}}\,I^{\frac{1}{2}}\,\eta_{H_1}^{\frac{1}{2}}} \exp\left[-\frac{\tau}{2}\left\{\sum_{t=1}^T (y_t-\alpha Y_{t-1})^2-\hat{\theta}_{H_1}^2\eta_{H_1}\right.\right. \\ &\quad \left.\left. +(\alpha-\mu)'P(\alpha-\mu)+\frac{\lambda_0^2}{\omega}+y_0^2+2b-\frac{1}{I}\left(\frac{\lambda_0}{\omega}+\sum_{t=1}^T e_t(y_t-\alpha Y_{t-1})\right)^2\right\}\right]\,d\alpha\,d\tau \\ &= \frac{b^a\,|P|^{\frac{1}{2}}}{\Gamma a\,(2\pi)^{\frac{T}{2}}\,\omega^{\frac{1}{2}}\,I^{\frac{1}{2}}\,\eta_{H_1}^{\frac{1}{2}}\,\xi_{H_1}^{\frac{1}{2}}} \int_0^\infty \tau^{\frac{T}{2}+a-1}\exp\left[-\frac{\tau}{2}\Omega_{H_1}\right]d\tau \\ &= \frac{b^a\,|P|^{\frac{1}{2}}}{\Gamma a\,(2\pi)^{\frac{T}{2}}\,\omega^{\frac{1}{2}}\,I^{\frac{1}{2}}\,\eta_{H_1}^{\frac{1}{2}}\,\xi_{H_1}^{\frac{1}{2}}}\;\frac{\Gamma\left(\frac{T}{2}+a\right)}{\left[\Omega_{H_1}\right]^{\frac{T}{2}+a}} \end{aligned} \tag{12}$$

Case 2: Under the null hypothesis, when model is not contaminated by an additive outlier, i.e, $\lambda = 0$. The posterior probability under H_0 is

$$\begin{aligned} P(y|H_0) &= \int_0^\infty \int_{R^p} \int_{-\infty}^\infty \frac{\tau^{\frac{T+p+1}{2}+a}\,b^a\,|P|^{\frac{1}{2}}}{\Gamma a\,(2\pi)^{\frac{T+p+1}{2}}} \exp\left[-\frac{\tau}{2}\left\{\sum_{t=1}^T (y_t-\alpha Y_{t-1})^2-2\theta\sum_{t=1}^T (y_t-\alpha Y_{t-1})\right.\right. \\ &\qquad \left.\left. +(\alpha-\mu)'P(\alpha-\mu)-2\theta y_0+(T+1)\theta^2+y_0^2+2b\right\}\right]\,d\theta\,d\alpha\,d\tau. \end{aligned}$$

Let us define $\hat{\theta}_{H_0} = \frac{1}{\eta_{H_0}}\left(\sum_{t=1}^T (y_t-\alpha Y_{t-1})+y_0\right)$ and $\hat{\alpha}_{H_0}$ given by Equation (8) and integrating with respect to θ, α and τ, we get

$$P(y|H_0) = \int_{R^p}\int_{-\infty}^{\infty}\int_0^{\infty} \frac{\tau^{\frac{T+p-1}{2}+a}\, b^a\, |P|^{\frac{1}{2}}}{\Gamma a\, (2\pi)^{\frac{T+p+1}{2}}} \exp\left[-\frac{\tau}{2}\left\{\sum_{t=1}^{T}(y_t-\alpha Y_{t-1})^2 -2\theta\hat{\theta}_{H_0} + (\alpha-\mu)'P(\alpha-\mu) + \eta_{H_0}\theta^2 + y_0^2 + 2b\right\}\right] d\theta\, d\alpha\, d\tau$$

$$= \int_0^{\infty} \frac{\tau^{\frac{T}{2}+a-1}\, b^a\, |P|^{\frac{1}{2}}}{\Gamma a\, (2\pi)^{\frac{T}{2}}\, \eta_{H_0}^{\frac{1}{2}}\, \xi_{H_0}^{\frac{1}{2}}} \exp\left[-\frac{\tau}{2}\left\{\sum_{t=1}^{T} y_t^2 + \mu' P\mu + 2b + y_0^2 - \frac{1}{\eta_{H_0}}\left(y_0+\sum_{t=1}^{T} y_t\right)^2 - \hat{\alpha}'_{H_0}\xi_{H_0}\hat{\alpha}_{H_0}\right\}\right] d\tau$$

$$= \frac{b^a\, |P|^{\frac{1}{2}}}{\Gamma a\, (2\pi)^{\frac{T}{2}}\, \eta_{H_0}^{\frac{1}{2}}\, \xi_{H_0}^{\frac{1}{2}}} \frac{\Gamma(\frac{T}{2}+a)}{[\Omega_{H_0}]^{\frac{T}{2}+a}} \tag{13}$$

By utilizing the posterior probability under H_1 (Equation (12)) and H_0 (Equation (13)) with prior odds ratio $\frac{P_0}{1-P_0}$, we get the posterior odds ratio β_{01} (7). □

4. Empirical Analysis

If time series data is contaminated by an additive outlier then there will be a sudden jump from a particular point. For this reason, present study is exploring the testing procedure under Bayesian framework to identify such type of observations and examine whether this jump is due to outlier or not. For this, we have analyzed monthly time series of recorded patient of communicable disease of Haryana State, India. There are 20 communicable diseases recorded by Central Bureau of Health Intelligence (CBHI), Ministry of Health, India during the period January 2012 to December 2014, where we get a series namely Enteric fever series (as shown in Fig. 1) having a suspected major deviation in September 2012 which may be an outlier. Enteric fever, commonly known as Typhoid fever is caused by Salmonella Typhoid bacteria. Typhoid fever is rare in industrialised countries. Some of these occur throughout the year and some especially in rainy and post-rainy season (see, Tyagi et al. [2011]). During the study, it is noted that outlier has occurred in the month of September 2012 which is the last month of the rainy seasons in Haryana state. Model has been fitted for all possible AR processes using R-software and got AR(1) as the best fitted model and recorded maximum ups at time point ($T_0 = 9$), which is shown in residual plot given in Fig. 1.

In practice, an observation which is lying beyond the 3σ limit may be considered as an outlier. In such case, error is partitioned into two parts, such as error from assumed distribution and magnitude of an outlier. The residual at extreme point is 4341.15 at $T_0 = 9$ and it can be partitioned into two parts such as 3σ and λe_t values which is given in Table 1 and value of coefficients for model (4) are also recorded in Table 1. Here, we have combined error, i.e., $\varepsilon_t + \lambda e_t$ at T_0 and considered maximum permissible errors 3σ which may be extended for the values from error distribution and then difference may be taken equal to λe_t.

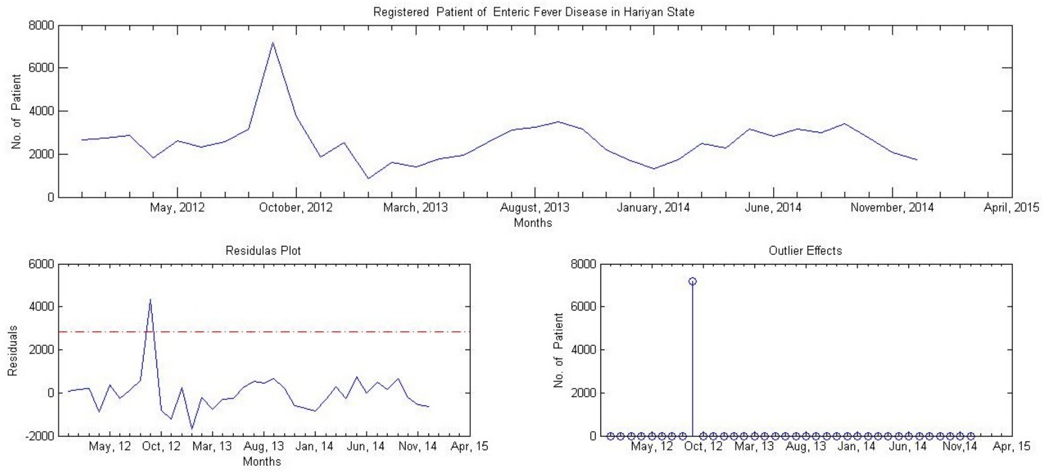

Fig. 1. Realization of the Enteric fever series

Table 1. Real data model fitting

Model		Coefficient	AIC	$\pm 3\sigma$ limit	Position of Outlier	λe_t
ARIMA(1,0,0)	θ	2571.3501 (-264.881)	600.3401	± 2823.5032	9	1517.6493
	α_1	0.4275 (-0.1503)				

Present study is aims to test the presence of an outlier for the best-fitted model, i.e., AR(1). In empirical study, it is noticed that at time point T_0, an additional variability came in the form of product of two values λ and e_t. Therefore, we have taken various combinations of λ and e_t in such a way that product of these two values must be equal to 1517.6493 and we have calculated β_{01} for different value of $\lambda = \{20, 40, 60, 80\}$ and corresponding value of e_t. β_{01} is shown in Fig. 2. Here, it is observed that β_{01} is increased as value of e_t increases for fixed λ. We also observed that β_{01} is almost constant with the increase of value of λ for fixed e_t as reported in Fig. 2.

In Bayesian framework, decision of acceptance and rejection of hypothesis is taken based on posterior probability through POR. Here, considering equal prior probability, i.e., $P_0 = 0.5$, if β_{01} is more than one, it means that there are more chances to accept the null hypothesis. In all cases β_{01} is less than one, which means there is no evidence to accept the null hypothesis and we conclude that

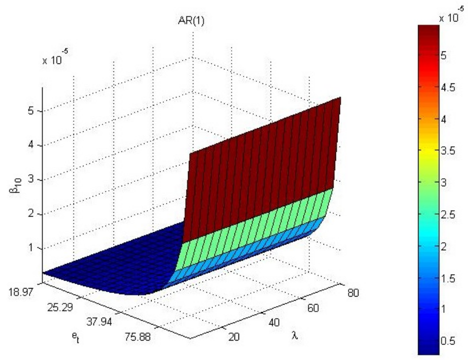

Fig. 2. Posterior odds ratio of the Enteric fever series

series is contaminated by the outlier at $T_0 = 9$.

It is also noticed that value of β_{01} is less than one for all value of λ and e_t because the real series is having one outlier. The empirical study shows the significance of Theorem 1, but value of β_{01} may be affected due to size of the series as well as variation within series. So in this connection, we have explored a simulation study in the next section.

5. Realization

Simulation is the realization of real world to explain the process for various studies. We have taken coefficient values from real data and considered the models AR(p), where p = 1, 2 and 3. The coefficient of pre-assumed model is calculated after fitting the model on Enteric fever series and recorded in Table 1. We have generated a time series of size $T = \{40, 60, 80 \; and \; 120\}$ observations from AR(p) model. For simulation purpose, value of error variance τ^{-1} is 1.25, $\omega = \{0.001, 0.003, 0.005, 0.007, 0.009\}$; $\lambda_0 = \{1, 5, 10, 15, 20\}$ and $\{\mu_1, \mu_2, \mu_3, \cdots, \mu_p\}$. These are taken as same values as recorded in empirical analysis. For AR(1), initial value $y_0 = 100$, for AR(2) $y_0 = 100$, $y_1 = 200$ and for the AR(3) $y_0 = 100, y_1 = 200, y_2 = 300$. The position of the outlier at $T_0 = 9$ is fixed for all the simulated series, i.e., additive outlier will occur at known positions in all the simulated series.
Here, we have injected different values of an outlier at T_0 that appear in the form of λe_t. For the numerical purpose, we have taken fixed $e_t = 10$ and generated λ considering various values of λ_0 ranging from 1 to 20. The value of POR is calculated with varying λ_0 and T for different five values of ω and obtained β_{01} for AR(1) model. For this, Jeffrey's hypothesis testing criterion is used to calculate POR from the following steps:

1. Simulate time series of size T for AR(1) and taking the seed values from the estimated values of coefficient(s) of best model.
2. Inject an outlier (e_t) at T_0.

Table 2. Realization of Enteric fever series for Simulation

Model	Coefficient				AIC
	θ	α_1	α_2	α_3	
AR(1)	2571.3531 (264.8805)	0.4321 (0.1502)	-	-	600.3401
AR(2)	2576.0901 (240.8312)	0.4702 (0.1602)	-0.1101 (0.1605)	-	601.9302
AR(3)	2578.2312 (215.2401)	0.4602 (0.1620)	-0.0448 (0.1812)	-0.1240 (0.1602)	603.3431

3. Write the likelihood function under a particular hypothesis with appropriate prior distribution.
4. Compute posterior probability under null and alternative hypothesis.
5. Reject H_0 if POR is less than one, otherwise accept.

This procedure was also repeated for AR(2) and AR(3) models. A single series could not interpret the model appropriately. Therefore, process is replicated 5000 times and recorded average value of β_{01} for all three models with respective set of given parameters. The β_{01} are represented for model AR(1), AR(2) and AR(3) in Fig. 3 and value of β_{01} as also recorded in Appendix (Table A1, Table A2 and Table A3).

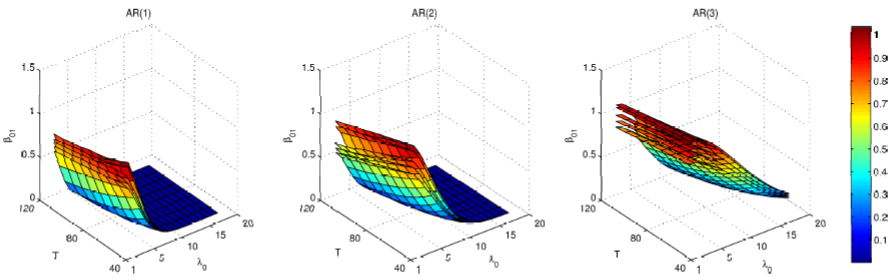

Fig. 3. Variation in POR with Size of the series (T), λ and ω for AR(1), AR(2) and AR(3)

The main motive of present section is to understand the behaviour of β_{01} in respect to size of series and magnitude of outlier (e_t). If series is contaminated by an outlier and the best model is correctly identified, then Theorem 1 detects outlier properly. On the basis of Enteric fever series considered in the empirical analysis shows that the best fitted model was AR(1). It is possible that an AR(1) process may be miss-identified due to outlier. Therefore, we have investigated it through our simulation for an autoregressive process of higher order, i.e., p=2 and p=3. We found a similar pattern of β_{01} for model AR(2) and AR(3). The slope of β_{01} for AR(2) process is more rather than with AR(1), but in case of AR(3) model, the slope of β_{01} is rapidly increasing and tends to one. This indicates that β_{01} values increase with the increase in size of the series (T). It is also observed that if value of ω increases then impact of λ_0 on the β_{01} is less. In these situations, no effect is coming from the outlier either with the increase of size of series or with the variation within the series. In the all cases, value of β_{01} tends to one or more than one. In case of AR for large time series and less value of λ, null hypothesis is accepted which means series is not contaminated by an outlier. This may be because of two reasons, one large series may manage the impact of outlier and record due to small λ.

6. Conclusion

In this paper, we consider a Bayesian mechanism for an autoregressive time series model contaminated with additive outlier (AO-AR(p)). A posterior odds ratio is derived in the presence of additive outlier with the known prior assumptions. An empirical analysis on monthly patient of Enteric fever series is considered to observe the impact of the outlier and conclude that series have a suspected major deviation due to outlier. With the help of simulation study, different magnitudes and different size of the series are considered to know the significance of the studied model and it is suitable to justify the mechanism. The work may also be extended for panel AR(p) and vector autoregressive model.

References

Abraham, B. and Box, G.E. (1979). Bayesian analysis of some outlier problems in time series. *Biometrika*, 229–236.

Bansal, A.K. (2007). *Bayesian parametric inference*. Alpha Science International Limited.

Barnett, G., Kohn, R. and Sheather, S. (1996). Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics*, 74(2): 237–254.

Barnett, V. and Lewis, T. (1994). Outliers in statistical data, ($3rd$ ed.). John Wiley & Sons, Chichester.

Battaglia, F. and Orfei, L. (2005). Outlier detection and estimation in nonlinear time series. *Journal of Time Series Analysis*, 26(1): 107–121.

Box and Jenkins (1970). Engineering and statistics handbook-ARMA Models. INSEAD, 64–65.

Broemeling, L.D. (1984). Bayesian Analysis of Linear Models, (1^{st} ed.). CRC Press.

Chang, I. and Tiao, G.C. (1983). Estimation of time series parameters in the presence of outliers. *Technical Report 8, Statistics Research Centre*, University of Chicago.

Chang, I., Tiao, G.C. and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 3: 193–204.

Fox, A.J. (1972). Outliers in time series. *Journal of the Royal Statistical Society*, Series B (Methodological), 350–363.

Gordon, K. (1986). Modelling and monitoring of medical time series, (Ph.D. diss.). University of Nottingham.

Gordon, K. and Smith, A.F.M. (1988). Modeling and Monitoring Biomedical Time Series in Bayesian Analysis of Time Series and Dynamic Models. *ed. J. Spall*, New York: Marcel Dekker, 359–391.

Gordon, K. and Smith, A.F.M. (1990). Modeling and monitoring biomedical time series. *Journal of the American Statistical Association*, 85(410): 328–337.

Jeng, X.J., Cai, T.T. and Li, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika*, 100(1): 157–157.

Kumar, J. and Shukla, A. (2015). Identification of Additive Outlier in Stationary AR (1): A Bayesian Approach. *International Journal of Intelligent Technologies and Applied Statistics*, 8(3): 275–290.

Kumar, J., Shukla, A. and Tiwari, N. (2014). Bayesian Analysis of a Stationary AR (1) model and outlier. *Electronic Journal of Applied Statistical Analysis*, 7(1): 81–93.

Siegmund, D., Yakir, B. and Zhang, N.R. (2011). Detecting simultaneous variant intervals in aligned sequences. *The Annals of Applied Statistics*, 645–668.

Schotman, P.C., and Van Dijk, H.K. (1991). On Bayesian routes to unit roots. *Journal of Applied Econometrics*, 6(4): 387–401.

Silva, M.E. and Pereira, I. (2015). Detection of additive outliers in Poisson INAR (1) time series. *In Mathematics of Energy and Climate Change*, 377–388. Springer International Publishing.

Tsay, R.S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of forecasting*, 7(1): 1–20.

Tyagi, P.K., Tyagi, S., Kumar, R. and Panday, C.S. (2011). Bacteriological analysis of air of kitchens in rural and urban areas of Panipat district in Haryana (India). *Int. J. Pharm. Biol. Sci.*, 2(1): 247–256.

Appendex

Table A1. Average posterior odds ratio for different combination of λ_0,T and ω AR(1) model

ω	T	$\lambda_o = 1$	$\lambda_o = 5$	$\lambda_o = 10$	$\lambda_o = 15$	$\lambda_o = 18$
0.001	40	7.87E-01	2.76E-01	1.05E-02	1.10E-04	5.31E-06
	60	7.35E-01	1.48E-01	1.14E-03	1.28E-06	1.45E-08
	80	6.94E-01	7.95E-02	1.22E-04	1.52E-08	3.99E-11
	120	6.12E-01	2.25E-02	1.41E-06	2.13E-12	3.07E-16
0.003	40	8.50E-01	2.99E-01	1.15E-02	1.19E-04	5.79E-06
	60	8.03E-01	1.62E-01	1.23E-03	1.40E-06	1.59E-08
	80	7.54E-01	8.72E-02	1.32E-04	1.67E-08	4.35E-11
	120	6.67E-01	2.45E-02	1.52E-06	2.32E-12	3.35E-16
0.005	40	9.09E-01	3.21E-01	1.23E-02	1.28E-04	6.21E-06
	60	8.62E-01	1.74E-01	1.33E-03	1.51E-06	1.69E-08
	80	8.10E-01	9.28E-02	1.42E-04	1.78E-08	4.69E-11
	120	7.18E-01	2.65E-02	1.66E-06	2.50E-12	3.56E-16
0.007	40	9.82E-01	3.61E-01	1.30E-02	1.32E-04	6.31E-06
	60	9.06E-01	1.96E-01	1.41E-03	1.54E-06	1.70E-08
	80	8.48E-01	1.07E-01	1.53E-04	1.80E-08	4.55E-11
	120	7.50E-01	3.19E-02	1.80E-06	2.50E-12	3.34E-16
0.009	40	1.03E+00	3.80E-01	1.38E-02	1.39E-04	6.66E-06
	60	9.51E-01	2.08E-01	1.48E-03	1.62E-06	1.80E-08
	80	9.06E-01	1.13E-01	1.61E-04	1.90E-08	4.80E-11
	120	8.11E-01	3.37E-02	1.88E-06	2.62E-12	3.52E-16

Table A2. Average posterior odds ratio for different combination of λ_0,T and ω AR(2) model

ω	T	$\lambda_o = 1$	$\lambda_o = 5$	$\lambda_o = 10$	$\lambda_o = 15$	$\lambda_o = 18$
	40	9.14E-01	5.43E-01	1.04E-01	8.67E-03	1.44E-03
	60	8.33E-01	3.87E-01	3.38E-02	8.53E-04	6.08E-05
0.001	80	7.47E-01	2.75E-01	1.09E-02	8.42E-05	2.59E-06
	120	6.00E-01	1.37E-01	1.12E-03	8.32E-07	4.65E-09
	40	9.95E-01	5.90E-01	1.13E-01	9.27E-03	1.56E-03
	60	8.99E-01	4.19E-01	3.68E-02	9.20E-04	6.60E-05
0.003	80	8.14E-01	2.97E-01	1.19E-02	9.22E-05	2.81E-06
	120	6.49E-01	1.45E-01	1.25E-03	9.14E-07	5.11E-09
	40	1.06E+00	6.32E-01	1.22E-01	9.97E-03	1.68E-03
	60	9.70E-01	4.51E-01	3.93E-02	9.90E-04	7.06E-05
0.005	80	8.80E-01	3.18E-01	1.28E-02	9.88E-05	3.01E-06
	120	7.05E-01	1.58E-01	1.33E-03	9.82E-07	5.52E-09
	40	1.18E+00	7.24E-01	1.48E-01	1.32E-02	2.31E-03
	60	1.12E+00	5.43E-01	5.11E-02	1.40E-03	1.07E-04
0.007	80	1.05E+00	4.05E-01	1.77E-02	1.50E-04	4.89E-06
	120	9.03E-01	2.23E-01	2.10E-03	1.70E-06	1.04E-08
	40	1.25E+00	7.65E-01	1.56E-01	1.39E-02	2.45E-03
	60	1.18E+00	5.74E-01	5.40E-02	1.48E-03	1.13E-04
0.009	80	1.11E+00	4.27E-01	1.86E-02	1.58E-04	5.18E-06
	120	9.64E-01	2.36E-01	2.22E-03	1.80E-06	1.10E-08

Table A3. Average posterior odds ratio for different combination of λ_0,T and ω AR(3) model

ω	T	$\lambda_o = 1$	$\lambda_o = 5$	$\lambda_o = 10$	$\lambda_o = 15$	$\lambda_o = 18$
	40	1.01E+00	8.92E-01	5.84E-01	2.94E-01	1.73E-01
	60	9.82E-01	8.22E-01	4.39E-01	1.58E-01	7.22E-02
0.001	80	9.51E-01	7.53E-01	3.29E-01	8.50E-02	3.00E-02
	120	8.84E-01	6.30E-01	1.84E-01	2.46E-02	5.19E-03
	40	1.09E+00	9.68E-01	6.34E-01	3.19E-01	1.88E-01
	60	1.07E+00	8.93E-01	4.77E-01	1.72E-01	7.83E-02
0.003	80	1.03E+00	8.18E-01	3.57E-01	9.25E-02	3.26E-02
	120	9.57E-01	6.85E-01	2.00E-01	2.66E-02	5.62E-03
	40	1.17E+00	1.04E+00	6.80E-01	3.42E-01	2.02E-01
	60	1.14E+00	9.58E-01	5.12E-01	1.84E-01	8.40E-02
0.005	80	1.10E+00	8.79E-01	3.84E-01	9.91E-02	3.50E-02
	120	1.03E+00	7.32E-01	2.15E-01	2.86E-02	6.06E-03
	40	1.24E+00	1.11E+00	7.35E-01	3.73E-01	2.21E-01
	60	1.21E+00	1.03E+00	5.58E-01	2.03E-01	9.34E-02
0.007	80	1.18E+00	9.58E-01	4.24E-01	1.11E-01	3.94E-02
	120	1.12E+00	8.18E-01	2.43E-01	3.29E-02	7.02E-03
	40	1.28E+00	1.15E+00	7.57E-01	3.84E-01	2.27E-01
	60	1.25E+00	1.06E+00	5.74E-01	2.09E-01	9.60E-02
0.009	80	1.22E+00	9.85E-01	4.36E-01	1.14E-01	4.06E-02
	120	1.15E+00	8.42E-01	2.50E-01	3.38E-02	7.22E-03