A Bayesian Approach for Detecting Outliers in ARMA Time Series

GUOCHAO ZHANG
Institute of Science,
Information Engineering University
450001 Zhengzhou
CHINA
940202587@qq.com

QINGMING GUI
Institute of Science,
Information Engineering University
450001 Zhengzhou
CHINA
guiqingming@126.com

Abstract: The presence of outliers in time series can seriously affect the model specification and parameter estimation. To avoid these adverse effects, it is essential to detect these outliers and remove them from time series. By the Bayesian statistical theory, this article proposes a method for simultaneously detecting the additive outlier (AO) and innovative outlier (IO) in an autoregressive moving-average (ARMA) time series. Firstly, an approximate calculation method of the joint probability density function of the ARMA time series is given. Then, considering the situation that AO and IO may present at the same time in an ARMA time series, a model for detecting outliers with the classification variables is constructed. By this model, this article transforms the problem of detecting outliers into a multiple hypothesis testing. Thirdly, the posterior probabilities of the multiple hypotheses are calculated with a Gibbs sampling, and based on the principle of Bayesian statistical inference, the locations and types of outliers can be obtained. What's more, the abnormal magnitude of every outlier also can be calculated by the Gibbs samples. At last, the new method is tested by some experiments and compared with other methods existing. It has been proved that the new approach can simultaneously detect the AO and IO successfully and performs better in terms of detecting the outlier which is both AO and IO, and but cannot be recognized by other methods existing.

Key-words: ARMA model; Additive outlier (AO); Innovative outlier (IO); Classification variable; Bayesian hypothesis test; Gibbs sampling

1 Introduction

Time series analysis is a very important statistical method of dynamic data processing in science and engineering [1-3]. A time series often contains all kinds of outliers, such as additive outlier (AO), innovative outlier (IO), temporary change (TC), level shift (LS), etc [4, 5]. As [6] pointed, the presence of these outliers could easily mislead the conventional time series analysis procedure resulting in erroneous conclusions. So, it is important to have procedures that detect and remove such outliers effects [7].

The bayesian method for detecting outliers in a time series had been considered by [8] in the earliest time. [9] used the Gibbs sampler in the Bayesian analysis of autoregressive (AR) time series and solved the problem of detecting the AO in the AR model by the Gibbs sampling. [10] illustrated the reason of masking and swamping, and proposed a solution to the problem based on the standard Gibbs sampling. [11] developed a procedure for detecting the AOs in the ARMA model by model selection strategies and Bayesian information criterion (BIC).

However, there are some disadvantages in the existing Bayesian approaches for detecting outliers in a time series. (a) As we all know, the ARMA model is widely applied in practice than the AR model, and yet the most of existing methods aim at detection of outliers in the AR model, only a few procedures focus on the ARMA model. (b) The joint probability density function of the ARMA time series is essential to Bayesian inference. However, it is complex to be not calculated accurately when the number of observations is large because of the correlation among the observations of the ARMA time series. So, there is no way to Bayesian inference to the ARMA model. (c) It is common that all kinds of outliers in the ARMA time series may appeared at the same time. But the existing Bayesian methods cannot detect them simultaneously.

Therefore, this article proposes a method for detecting all kinds of outliers simultaneously, especially for detecting AO and IO the most common outliers, in the ARMA time series by the Bayesian statistical theory. The rest of the paper is organized as follows. In section 2, a model of detecting the AO and IO in the ARMA time series simultaneously is constructed based on the classification variables of outliers, and a rule of detecting outliers is proposed by applying the principle of Bayesian hypothesis testing. Section 3 develops a method of estimating the joint probability density function of the observations of the ARMA time series, and the conditional posterior distributions of unknown parameters are deduced. In section 4, a procedure of detecting all kinds of outliers simultaneously based on the Gibbs sampling is proposed. Section 5 shows the better performances of the approach proposed in this article comparing with other existing methods by some simulating experiments. Finally, some conclusions are given in section 6.

2 The model and rule for outlier detection with the classification variables

Assume that $\{z_t\}$ be a time series following a general ARMA (p,q) process,

$$\begin{cases} \phi(B)z_t = \theta(B)\varepsilon_t \\ \varepsilon_t \ i.i.d.\ N(0, \sigma^2) \end{cases} \tag{1}$$

where $\phi(B) = I - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\theta(B) = I - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, $p > q$, B is a backshift operator such that $B^k z_t = z_{t-k}, k = 1, 2, \cdots$, $\{\varepsilon_t\}$ is a sequence of independent random errors identically distributed $N(0, \sigma^2)$. To ensure the ARMA (p,q) model being stationary and invertible, assume that all of the zeros of $\phi(B)$ and $\theta(B)$ are on or outside the unite circle [1-3].

On the basis of the definitions of AO and IO [4-7], the observation y_t that is affected by an AO or an IO or by both of them simultaneously can be written with the classification variables [8,10,12,13] as follows:

$$y_t = z_t + \omega_t^{AO}\delta_t^{AO} + \phi^{-1}(B)\theta(B)\omega_t^{IO}\delta_t^{IO}$$

From above, a model of detecting the AO and IO in the ARMA time series $\{y_t\}$ simultaneously is constructed as follows:

$$\begin{cases} y_t = x_t + \omega_t^{AO}\delta_t^{AO} \\ x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} \\ \varepsilon_t = a_t + \omega_t^{IO}\delta_t^{IO} \\ a_t \ \ i.i.d \ \ N(0,\sigma^2) \end{cases} \quad (2)$$

Where x_t is the observation which may affected by IO. δ_t^{AO} is a classification variable of AO, and δ_t^{AO} follows a Bernoulli distribution. If $\delta_t^{AO}=1$, the observation y_t is an AO that the abnormal magnitude is ω_t^{AO}; if $\delta_t^{AO}=0$, the observation y_t is not an AO. δ_t^{IO} is a classification variable of IO, and it is also follows a Bernoulli distribution. If $\delta_t^{IO}=1$, the observation y_t is an IO that the abnormal magnitude is ω_t^{IO}; if $\delta_t^{IO}=0$, the observation y_t is not an IO.

Suppose that there are n observations of the time series $\{y_t\}$, say $y_1, y_2, \cdots, y_n$, and the front p observations $y_1, y_2, \cdots, y_p$, $p << n$ are not outliers[8]. If we want to judge whether y_j $(j = p+1, \cdots, n)$ is an AO or IO or not, it is necessary to test a multiple hypothesis:

$$\left.\begin{matrix} H_{1,j}: \delta_j^{AO}=0, \delta_j^{IO}=0 & H_{2,j}: \delta_j^{AO}=1, \delta_j^{IO}=0 \\ H_{3,j}: \delta_j^{AO}=1, \delta_j^{IO}=1 & H_{4,j}: \delta_j^{AO}=0, \delta_j^{IO}=1 \end{matrix}\right\} \quad (3)$$

Here, if the hypothesis $H_{1,j}$ is accepted, we can conclude that y_j is neither an AO nor an IO. If $H_{2,j}$ is accepted, y_j is an AO but not an IO. That $H_{3,j}$ is accepted means that y_j is both AO and IO. What's more, if $H_{4,j}$ is accepted, y_j is an IO but not an AO.

Based on the principle of Bayesian hypothesis testing[14], we need to choose an appropriate prior distribution for every unknown parameter and calculate the posterior probability of every hypothesis as follows:

$$P(H_{1,j}|Y), P(H_{2,j}|Y), P(H_{3,j}|Y), P(H_{4,j}|Y)$$

where $Y = (y_{p+1}, y_{p+2}, \cdots, y_n)^T$. If $P(H_{i,j}|Y) = \max \{P(H_{1,j}|Y), P(H_{2,j}|Y), P(H_{3,j}|Y), P(H_{4,j}|Y)\}$, the $H_{i,j}$ will be accepted, which means that y_j can be identified to be a normal observation or some kind of outlier.

3 Estimating the joint probability density function of observations and calculating the conditional posterior distribution of unknown parameters

3.1 Estimating the joint probability density function of observations

When the parameters $\Phi = (\phi_1, \phi_2, \cdots, \phi_p)^T$, $\Theta = (\theta_1, \theta_2, \cdots, \theta_q)^T$ and σ^2 in the ARMA (p,q) model (1) are known, assume that the observations $z_1, z_2, \cdots, z_n$ are got and the mean of every observation is zero, so $Z = (z_1, z_2, \cdots, z_n)^T$ follows an n-variate normal distribution and can be expressed as $Z \sim N_n(\vec{0},\Sigma)$ by the definition of the ARMA model. Due to the correlation among the observations of the ARMA time series, the covariance matrix Σ of Z is an n-order matrix related to the parameters $\phi_1,\phi_2,\cdots,\phi_p,\theta_1,\theta_2,\cdots,\theta_q$ and σ^2. From above, it is difficult to calculate accurately the joint probability density function of Z when the number n is very large. And further, it is more difficult to calculate accurately the joint probability density function of Y which may include all kinds of outliers even though the first p observations $y_1,y_2,\cdots,y_p$ are normal. To calculate the conditional posterior distributions of unknown parameters which will be used in the following Gibbs sampling, an approach for approximately estimating the joint probability density function of Y is considered as follows:

When $Y^*=(y_1,y_2,\cdots,y_n)^T$, $\Phi=(\phi_1,\cdots,\ \phi_p)^T$, $\Theta=(\theta_1,\cdots,\ \theta_q)^T$, $\delta^{AO}=(\delta^{AO}_{p+1},\cdots,\delta^{AO}_n)^T$, $\delta^{IO}=(\delta^{IO}_{p+1},\cdots,\delta^{IO}_n)^T$, $\omega^{AO}=(\omega^{AO}_{p+1},\cdots,\omega^{AO}_n)^T$, $\omega^{IO}=(\omega^{IO}_{p+1},\cdots,\omega^{IO}_n)^T$ and σ^2 are known, based on the model (2), we can estimate the $\varepsilon=(\varepsilon_1,\cdots,\varepsilon_n)^T$ by the equations, $\varepsilon_1=\cdots=\varepsilon_p=0$, $\varepsilon_t=y_t-\omega^{AO}_t\delta^{AO}_t-\Phi^T X_{(-t)}+\Theta^T\varepsilon_{(-t)}$, where $X_{(-t)}=(x_{t-1},\cdots,x_{t-p})^T$, $x_m=y_m-\omega^{AO}_m\delta^{AO}_m, m=t-1,\cdots,t-p$, and $\varepsilon_{(-t)}=(\varepsilon_{t-1},\cdots,\varepsilon_{t-q})^T$. By the model (2), $\varepsilon_t=y_t-\omega^{AO}_t\delta^{AO}_t-\Phi^T X_{(-t)}+\Theta^T\varepsilon_{(-t)}$, $\varepsilon_t \sim N(\omega^{IO}_t\delta^{IO}_t,\sigma^2),\ t=p+1,\cdots,n$, so the probability density function of y_t is

$$p(y_t\mid\Phi,\Theta,\sigma^2,\omega^{AO}_t,\delta^{AO}_t,\omega^{IO}_t,\delta^{IO}_t)=(2\pi)^{-\frac{1}{2}}\sigma^{-1}$$
$$\cdot\exp\{-\frac{1}{2\sigma^2}(y_t-\omega^{AO}_t\delta^{AO}_t-\omega^{IO}_t\delta^{IO}_t-\Phi^T X_{(-t)}+\Theta^T\varepsilon_{(-t)})^2\}$$

$t=p+1,\cdots n$. At last, the joint probability density function of Y is calculated by

$$p(Y\mid x_{(-(p+1))},\varepsilon_{(-t)},\Phi,\Theta,\sigma^2,\omega^{AO},\delta^{AO},\omega^{IO},\delta^{IO})=$$
$$\prod_{t=p+1}^{n}p(y_t\mid\Phi,\Theta,\sigma^2,\omega^{AO}_t,\delta^{AO}_t,\omega^{IO}_t,\delta^{IO}_t)=$$
$$(2\pi)^{-\frac{n-p}{2}}\sigma^{-(n-p)}\exp\{-\frac{1}{2\sigma^2}\sum_{t=p+1}^{n}(y_t-\omega^{AO}_t\delta^{AO}_t$$
$$-\omega^{IO}_t\delta^{IO}_t-\Phi^T X_{(-t)}+\Theta^T\varepsilon_{(-t)})^2\} \quad (4)$$

3.2 Calculating the conditional posterior distributions of unknown parameters

Firstly, based on the selection principles of prior distributions [15], suppose that there is a small prior probability α that every observation $y_j\ (j=p+1,\cdots,n)$ is an AO or an IO, which means that $p(\delta^{AO}_j=1)=p(\delta^{IO}_j=1)=\alpha$ [8,10]. So, the prior distributions of all unknown parameters are selected as follows:

$\Phi \sim N_p(\Phi_0,V^{-1})$, that is $p(\Phi)=(2\pi)^{-\frac{p}{2}}|V|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\Phi-\Phi_0)^T V(\Phi-\Phi_0)\}$.

$\Theta \sim N_q(\Theta_0,W^{-1})$, that is $p(\Theta)=(2\pi)^{-\frac{q}{2}}|W|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\Theta-\Theta_0)^T W(\Theta-\Theta_0)\}$.

$\omega^{AO}_j \sim N(u_1,\xi_1^2)$, $\omega^{IO}_j \sim N(u_2,\xi_2^2)$

$\delta^{AO}_j \sim b(1,\alpha)$, that is $p(\delta^{AO}_j)=\alpha^{\delta^{AO}_j}(1-\alpha)^{1-\delta^{AO}_j}$

$\delta_j^{IO} \sim b(1,\alpha)$, that is $p(\delta_j^{IO}) = \alpha^{\delta_j^{IO}} (1-\alpha)^{1-\delta_j^{IO}}$

$\sigma^2 \sim IG(\frac{v}{2},\frac{v\lambda}{2})$, where $\Phi_0, V^{-1}, \Theta_0, W^{-1}, u_1, \xi_1^2, u_2$, ξ_2^2, α, v and λ are the hyper parameters that are known.

The conditional posterior distributions of unknown parameters are acquired by the Bayesian formulas [16] as follows where $\varepsilon = (\varepsilon_1,\cdots,\varepsilon_n)^T$ and the joint posterior probability density function $p(Y \mid x_{(-(p+1))}, \varepsilon_{(-t)}, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta^{AO}, \omega^{IO}, \delta^{IO})$ of Y are estimated as described before.

(1) The conditional posterior distribution of Φ is

$$\Phi \mid Y, \Theta, \sigma^2, \omega^{AO}, \delta^{AO}, \omega^{IO}, \delta^{IO} \sim N(\hat{\Phi}_0, \hat{V}^{-1}) \quad (5)$$

where $\hat{V}^{-1} = [\frac{1}{\sigma^2}\sum_{t=p+1}^{n} X_{(-t)} X_{(-t)}^T + V]^{-1}$, $\hat{\Phi}_0 = \hat{V}^{-1}$

$[\frac{1}{\sigma^2}\sum_{t=p+1}^{n} X_{(-t)}(y_t - \omega_t^{AO}\delta_t^{AO} - \omega_t^{IO}\delta_t^{IO} + \Theta^T \varepsilon_{(-t)}) + V\Phi_0]$.

(2) The conditional posterior distribution of σ^2 is

$$\sigma^2 \mid Y, \Phi, \Theta, \omega^{AO}, \delta^{AO}, \omega^{IO}, \delta^{IO} \sim IG(a,b) \quad (6)$$

where $a = \frac{n-p+v}{2}$, $b = \frac{1}{2}[\sum_{t=p+1}^{n}(y_t - \omega_t^{AO}\delta_t^{AO} - \omega_t^{IO}\delta_t^{IO}$

$-\Phi^T X_{(-t)} + \Theta^T \varepsilon_{(-t)})^2 + v\lambda]$.

(3) The conditional posterior distribution of Θ is

$$\Theta \mid Y, \Phi, \sigma^2, \omega^{AO}, \delta^{AO}, \omega^{IO}, \delta^{IO} \sim N(\hat{\Theta}_0, \hat{W}^{-1}) \quad (7)$$

where $\hat{W}^{-1} = (\frac{1}{\sigma^2}\sum_{t=p+1}^{n} \varepsilon_{(-t)} \varepsilon_{(-t)}{}^T + W)^{-1}$, $\hat{\Theta}_0 = -\hat{W}^{-1}$

$(\frac{1}{\sigma^2}\sum_{t=p+1}^{n} \varepsilon_{(-t)}(y_t - \omega_t^{AO}\delta_t^{AO} - \omega_t^{IO}\delta_t^{IO} - \Phi^T X_{(-t)}) - W\Theta_0)$.

(4) The conditional posterior distribution of $\delta_j^{AO}, \delta_j^{IO}$ can be calculated as follows:

$$P_{1j} = P(\delta_j^{AO} = 0, \delta_j^{IO} = 0 \mid Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{(-j)}^{AO}, \omega^{IO}, \delta_{(-j)}^{IO})$$
$$= \alpha^2 \exp\{-\frac{1}{2\sigma^2}\sum_{t=j}^{T}(y_t - \omega_t^{AO}\delta_t^{AO} - \omega_t^{IO}\delta_t^{IO} - \Phi^T X_{(-t)}$$
$$+\Theta^T \varepsilon_{(-t)})^2\} \quad (8)$$

$$P_{2j} = P(\delta_j^{AO} = 1, \delta_j^{IO} = 0 \mid Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{(-j)}^{AO}, \omega^{IO}, \delta_{(-j)}^{IO})$$
$$= \alpha(1-\alpha) \exp\{-\frac{1}{2\sigma^2}\sum_{t=j}^{T}(y_t - \omega_t^{AO}\delta_t^{AO} - \omega_t^{IO}\delta_t^{IO} - \Phi^T X_{(-t)}$$
$$+\Theta^T \varepsilon_{(-t)})^2\} \quad (9)$$

$$P_{3j} = P(\delta_j^{AO} = 1, \delta_j^{IO} = 1 \mid Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{(-j)}^{AO}, \omega^{IO}, \delta_{(-j)}^{IO})$$
$$= (1\text{-}\alpha)^2 \exp\{-\frac{1}{2\sigma^2}\sum_{t=j}^{T}(y_t - \omega_t^{AO}\delta_t^{AO} - \omega_t^{IO}\delta_t^{IO} - \Phi^T X_{(-t)}$$
$$+\Theta^T \varepsilon_{(-t)})^2\} \quad (10)$$

$$P_{4j} = P(\delta_j^{AO} = 0, \delta_j^{IO} = 1 \mid Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta_{(-j)}^{AO}, \omega^{IO}, \delta_{(-j)}^{IO})$$
$$= \alpha(1-\alpha) \exp\{-\frac{1}{2\sigma^2}\sum_{t=j}^{T}(y_t - \omega_t^{AO}\delta_t^{AO} - \omega_t^{IO}\delta_t^{IO} - \Phi^T X_{(-t)}$$
$$+\Theta^T \varepsilon_{(-t)})^2\} \quad (11)$$

where $T = \min(n, j+p)$.

(5) The conditional posterior distribution of ω_j^{AO} is $\omega_j^{AO} \mid Y, \Phi, \Theta, \sigma^2, \omega_{(-j)}^{AO}, \delta^{AO}, \omega^{IO}, \delta^{IO} \sim$

$$N(\hat{\omega}_j^{AO}, (\hat{\xi}_j^2)^{AO}) \quad (12)$$

where $(\hat{\xi}_j^2)^{AO} = [\frac{(\delta_j^{AO})^2}{\sigma^2}(1+\phi_1{}^2+\cdots+\phi_{N-j}^2) + \frac{1}{\xi_1{}^2}]^{-1}$,

$$\hat{\omega}_j^{AO} = (\hat{\xi}_j^2)^{AO}[\frac{\delta_j^{AO}}{\sigma^2}(y_j - \omega_j^{IO}\delta_j^{IO} - \Phi^T X_{(-j)} + \Theta^T \varepsilon_{(-j)})$$
$$-\sum_{m=j+1}^{T}\frac{\phi_{m-j}}{\sigma^2}\delta_j^{AO}(y_m - \omega_m^{AO}\delta_m^{AO} - \omega_m^{IO}\delta_m^{IO} - \phi_1 x_{m-1} - \cdots$$
$$-\phi_{m-j} y_j - \cdots - \phi_p x_{m-p} + \Theta^T \varepsilon_{(-m)}) + \frac{u_1}{\xi_1^2}]$$

(6) The conditional posterior distribution of ω_j^{IO} is $\omega_j^{IO} \mid Y, \Phi, \Theta, \sigma^2, \omega^{AO}, \delta^{AO}, \omega_{(-j)}^{IO}, \delta^{IO} \sim$

$$N(\hat{\omega}_j^{IO}, (\hat{\xi}_j^{\,2})^{IO}) \tag{13}$$

where $(\hat{\xi}_j^{\,2})^{IO} = [\frac{(\delta_j^{IO})^2}{\sigma^2} + \frac{1}{\xi_2^{\,2}}]^{-1}$, $\hat{\omega}_j^{IO} = (\hat{\xi}_j^{\,2})^{IO}[\frac{\delta_j^{IO}}{\sigma^2}(y_j - \omega_j^{AO}\delta_j^{AO} - \Phi^T X_{(-j)} + \Theta^T \varepsilon_{(-j)}) + \frac{u_2}{\xi_2^{\,2}}]$.

4 The implementation for outlier detection

When the orders p and q of the ARMA model (1) are known, but the parameters $\Phi = (\phi_1, \phi_2, \cdots, \phi_p)^T$, $\Theta = (\theta_1, \theta_2, \cdots, \theta_q)^T$ and σ^2 are unknown, the posterior probability $P(H_{i,j} \mid Y)$ cannot be calculated directly by the Bayesian formula. Thus, we can use the Gibbs sampling based on the conditional posterior distributions of unknown parameters to estimate $P(H_{i,j} \mid Y)$.

Above all, a procedure for detecting the AO and IO in an ARMA time series simultaneously is proposed based on the Gibbs sampling [15, 17] as follows:

Step 1: Choose the hyper parameters $\Phi_0, V^{-1}, \Theta_0, W^{-1}, u_1, \xi_1^2, u_2, \xi_2^2, \alpha, v$ and λ for the prior distributions of all unknown parameters in the ARMA model.

Step 2: Choose the initial values $\Phi^{(0)}, \Theta^{(0)}, (\sigma^2)^{(0)}, (\omega^{AO})^{(0)}, (\delta^{AO})^{(0)}, (\omega^{IO})^{(0)}$ and $(\delta^{IO})^{(0)}$ for the Gibbs sampling.

Step 3: Implement the Gibbs sampling as follows and get the Gibbs samples. Suppose that the (k-1)-th sample $(\Phi^{(k-1)}, \Theta^{(k-1)}, (\sigma^2)^{(k-1)}, (\omega^{AO})^{(k-1)}, (\delta^{AO})^{(k-1)}, (\omega^{IO})^{(k-1)}, (\delta^{IO})^{(k-1)})$ has been acquired. Then, the k-th sample can be obtained by the following procedure:

(a) Estimate the $\varepsilon^{(k)}$ by the equations, $\varepsilon_1 = \cdots = \varepsilon_p = 0$ an $\varepsilon_t = y_t - (\omega_t^{AO})^{(k-1)}(\delta_t^{AO})^{(k-1)} - (\Phi^{(k-1)})^T X_{(-t)} + (\Theta^{(k-1)})^T \varepsilon_{(-t)}$, $t = p+1, \cdots, n$.

(b) Obtain $\Phi^{(k)}$ from $\Phi \mid Y, \Theta^{(k-1)}, (\sigma^2)^{(k-1)}, (\omega^{AO})^{(k-1)}, (\delta^{AO})^{(k-1)}, (\omega^{IO})^{(k-1)}, (\delta^{IO})^{(k-1)}, \varepsilon^{(k)}$.

(c) Obtain $(\sigma^2)^{(k)}$ from $\sigma^2 \mid Y, \Phi^{(k)}, \Theta^{(k-1)}, (\omega^{AO})^{(k-1)}, (\delta^{AO})^{(k-1)}, (\omega^{IO})^{(k-1)}, (\delta^{IO})^{(k-1)}, \varepsilon^{(k)}$.

(d) Obtain $\Theta^{(k)}$ from $\Theta \mid Y, \Phi^{(k)}, (\sigma^2)^{(k)}, (\omega^{AO})^{(k-1)}, (\delta^{AO})^{(k-1)}, (\omega^{IO})^{(k-1)}, (\delta^{IO})^{(k-1)}, \varepsilon^{(k)}$.

(e) Calculate $P_{1j}^{(k)}, P_{2j}^{(k)}, P_{3j}^{(k)}$ and $P_{4j}^{(k)}$, and obtain $((\delta_j^{AO})^{(k)}, (\delta_j^{IO})^{(k)})$ from $\delta_j^{AO}, \delta_j^{IO} \mid Y, \Phi^{(k)}, \Theta^{(k)}, (\sigma^2)^{(k)}, (\omega^{AO})^{(k-1)}, (\delta_{(-j)}^{AO})^{(k-1,k)}, (\omega^{IO})^{(k-1)}, (\delta_{(-j)}^{IO})^{(k-1,k)}, \varepsilon^{(k)}$, where $(\delta_{(-j)}^{AO})^{(k-1,k)} = ((\delta_{p+1}^{AO})^{(k)}, \cdots, (\delta_{j-1}^{AO})^{(k)}, (\delta_{j+1}^{AO})^{(k-1)}, \cdots, (\delta_n^{AO})^{(k-1)})$ and $(\delta_{(-j)}^{IO})^{(k-1,k)} = ((\delta_{p+1}^{IO})^{(k)}, \cdots, (\delta_{j-1}^{IO})^{(k)}, (\delta_{j+1}^{IO})^{(k-1)}, \cdots, (\delta_n^{IO})^{(k-1)})$.

(f) Obtain $(\omega_j^{AO})^{(k)}$ from $\omega_j^{AO} \mid Y, \Phi^{(k)}, \Theta^{(k)}, (\sigma^2)^{(k)}, (\omega_{(-j)}^{AO})^{(k,k-1)}, (\delta^{AO})^{(k)}, (\omega^{IO})^{(k-1)}, (\delta^{IO})^{(k-1)}$

$,\varepsilon^{(k)}$, where $(\omega_{(-j)}^{AO})^{(k-1,k)} = ((\omega_{p+1}^{AO})^{(k)}, \cdots, (\omega_{j-1}^{AO})^{(k)}, (\omega_{j+1}^{AO})^{(k-1)}, \cdots, (\omega_{n}^{AO})^{(k-1)})$.

(g) Obtain $(\omega_j^{IO})^{(k)}$ from $\omega_j^{IO} \mid Y, \Phi^{(k)}, \Theta^{(k)}, (\sigma^2)^{(k)}, (\omega^{AO})^{(k)}, (\delta^{AO})^{(k)}, (\omega_{(-j)}^{IO})^{(k-1,k)}, (\delta^{IO})^{(k)}, \varepsilon^{(k)}$, where $(\omega_{(-j)}^{IO})^{(k-1,k)} = ((\omega_{p+1}^{IO})^{(k)}, \cdots, (\omega_{j-1}^{IO})^{(k)}, (\omega_{j+1}^{IO})^{(k-1)}, \cdots, (\omega_{n}^{IO})^{(k-1)})$.

Implement and end the iterative procedure after the Gibbs sampling is convergent.

Step 4: Make the Bayesian inference. Supposing that N samples are acquired and the Gibbs sampling is convergent after acquiring the M-th sample, we use the last N-M samples to make the following Bayesian inference.

(a) Get the locations and types of outliers.

By the above Gibbs sampling, the posterior probabilities of four hypotheses can be calculated approximately by

$$P(H_{1,j} \mid Y) = P(\delta_j^{AO} = 0, \delta_j^{IO} = 0 \mid Y) \approx \frac{1}{N-M} \sum_{k=M+1}^{N} \frac{P_{1j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \tag{14}$$

$$P(H_{2,j} \mid Y) = P(\delta_j^{AO} = 1, \delta_j^{IO} = 0 \mid Y) \approx \frac{1}{N-M} \sum_{k=M+1}^{N} \frac{P_{2j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \tag{15}$$

$$P(H_{3,j} \mid Y) = P(\delta_j^{AO} = 1, \delta_j^{IO} = 1 \mid Y) \approx \frac{1}{N-M} \sum_{k=M+1}^{N} \frac{P_{3j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \tag{16}$$

$$P(H_{4,j} \mid Y) = P(\delta_j^{AO} = 0, \delta_j^{IO} = 1 \mid Y) \approx \frac{1}{N-M} \sum_{k=M+1}^{N} \frac{P_{4j}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} \tag{17}$$

where $\frac{P_{ij}^{(k)}}{P_{1j}^{(k)} + P_{2j}^{(k)} + P_{3j}^{(k)} + P_{4j}^{(k)}} (i = 1,2,3,4)$ is the conditional posterior probability of $H_{i,j} (i = 1,2,3,4)$ at the k-th Gibbs sampling.

If $P(H_{i,j} \mid Y) = \max\{P(H_{1,j} \mid Y), P(H_{2,j} \mid Y), P(H_{3,j} \mid Y), P(H_{4,j} \mid Y)\}$, the hypothesis $H_{i,j}$ is accepted, and y_j can be identified to be a normal observation or some kind of outlier.

(b) Estimate the abnormal magnitudes of outliers.

If y_j $(j = p+1, \cdots, n)$ is recognized as an AO, its abnormal magnitude may estimated by

$$\hat{\omega}_j^{AO} = \frac{\sum_{k=M+1}^{N} (\omega_j^{AO})^{(k)}}{N-M} \tag{18}$$

If y_j $(j = p+1, \cdots, n)$ is recognized as an IO, its abnormal magnitude may estimated by

$$\hat{\omega}_j^{IO} = \frac{\sum_{k=M+1}^{N} (\omega_j^{IO})^{(k)}}{N-M} \tag{19}$$

5 Examples and analysis

In order to illustrate the performance of the approach for detecting outliers in the ARMA model proposed by this article, three examples are designed based on 100 observations from an ARMA model of orders (2,2),

$$\begin{cases} x_t = 0.5x_{t-1} + 0.3x_{t-2} + \varepsilon_t - 0.15\varepsilon_{t-1} - 0.1\varepsilon_{t-2} \\ \varepsilon_t \quad i.i.d \quad N(0,1) \end{cases},$$

Example 1: Add an IO of the abnormal magnitude equals to 5 at t=30; add an AO of the abnormal magnitude equals to 10 at t=80. Calculate the posterior probabilities of four hypotheses, and they are shown in Fig.1.

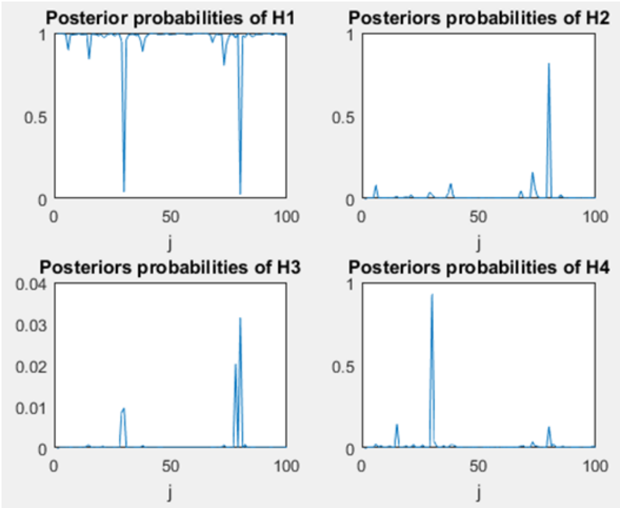

Fig.1 The posterior probabilities of example 1

The Fig.1 shows that $P(H_{4,30}|Y)=\max\{P(H_{1,30}|Y),P(H_{2,30}|Y),P(H_{3,30}|Y),P(H_{4,30}|Y)\}$, $P(H_{2,80}|y)=\max\{P(H_{1,80}|Y),P(H_{2,80}|Y),P(H_{3,80}|Y),P(H_{4,80}|Y)\}$,and $P(H_{1,j}|Y)=\max\{P(H_{1,j}|Y),P(H_{2,j}|Y),P(H_{3,j}|Y),P(H_{4,j}|Y)\}$, $j\neq 30,80$. What's more, the abnormal magnitudes of outliers can be calculated by formula (18) and (19). At last, the results of detecting outliers are that the 30^{th} observation is an IO and the 80^{th} observation is an AO, and the estimates of their abnormal magnitudes are $\hat{\omega}_{30}^{IO}=4.6494$ and $\hat{\omega}_{80}^{AO}=8.9637$, respectively.

We also use the iterative procedure proposed by [18] to outlier detection and the results are that both 30^{th} and 80^{th} observations are AOs and their abnormal magnitudes are $\hat{\omega}_{30}^{AO}=3.7522$ and $\hat{\omega}_{80}^{AO}=8.2626$, respectively, the 35^{th} observation is an IO and its abnormal magnitude is $\hat{\omega}_{35}^{IO}=-3.7044$.

Compared two results above, we can see clearly that the approach proposed in this article can locate and recognize precisely the outliers in an ARMA time series, and the abnormal magnitudes of these outliers can be estimated accurately. However, the erroneous identification for the locations and types of outliers is presented in the detection results of the iterate procedure proposed by [18].

Example 2: Add an AO of the abnormal magnitude equals to 6, 10 at t=30, 80, respectively. The posterior probabilities of four hypotheses are shown in Fig.2.

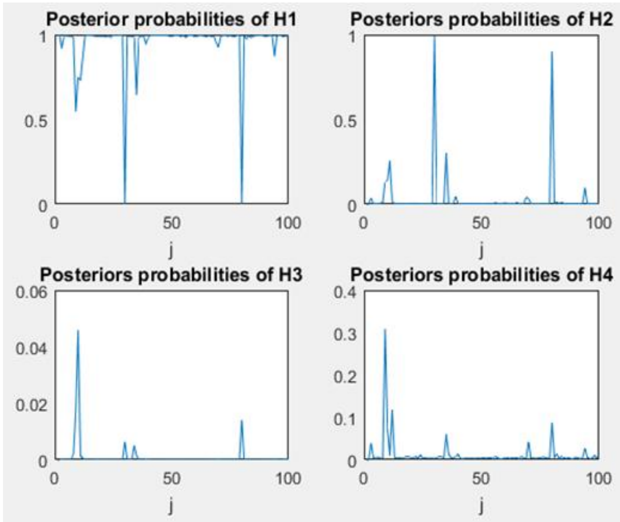

Fig.2 The posterior probabilities of example 2

The Fig.2 shows that $P(H_{2,30}|Y)=\max\{P(H_{1,30}|Y),P(H_{2,30}|Y),P(H_{3,30}|Y),P(H_{4,30}|Y)\}$, $P(H_{2,80}|Y)=\max\{P(H_{1,80}|Y),P(H_{2,80}|Y),P(H_{3,80}|Y),P(H_{4,80}|Y)\}$ and $P(H_{1,j}|Y)=\max\{P(H_{1,j}|Y),P(H_{2,j}|Y),P(H_{3,j}|Y),P(H_{4,j}|Y)\}$, $j\neq 30,80$. The abnormal magnitudes of outliers also can be calculated by formula (18). At last, the results of detecting outliers are that both 30th and 80th observations are AOs and the estimates of their abnormal magnitudes are $\hat{\omega}_{30}^{AO}=5.2475$ and $\hat{\omega}_{80}^{AO}=8.2772$, respectively.

Since the method proposed by [11] only can detect the AO in the ARMA time series, we use it to detect the outliers and get the results that the 30^{th}, 53th and 80th all are AOs and their abnormal magnitudes are $\hat{\omega}_{30}^{AO} = 2.62$, $\hat{\omega}_{53}^{AO} = 3.6966$ and $\hat{\omega}_{80}^{AO} = 5.9822$, respectively.

Compared the results above, it is obvious that the method proposed in this article can detect precisely the multiple outliers in an ARMA time series. But, the method proposed by [11] cannot locate the outliers accurately and may cause misjudgment of AO. What's more, the abnormal magnitudes of AOs still include a seriously bias.

Example 3: Add an AO of the abnormal magnitude equals to 10 and an IO of the abnormal magnitude equals to 8 at t=50 simultaneously; add an AO of the abnormal magnitude equals to 10 at t=80.The posterior probabilities of four hypotheses are shown in Fig.3.

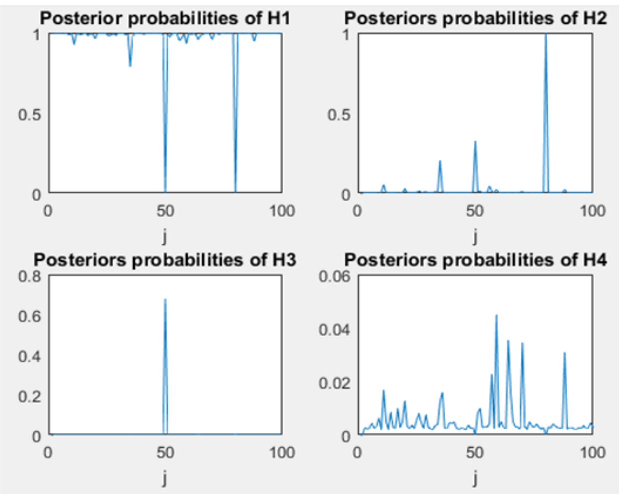

Fig.3 The posterior probabilities of example 3

Similarly, the 50th observation is recognized as both an AO of the abnormal magnitude equals to 9.7421and an IO of the abnormal magnitude equals to 7.3540, and the 80th observation is recognized as an AO of the abnormal magnitude equals to 9.0487.

From the three examples above, we can see that the Bayesian approach proposed in this article can get the locations, types, abnormal magnitudes of AO and IO in an ARMA time series precisely. Especially, the approach has a good performance of detecting the observation which is both AO and IO.

In addition, by the comparison of example 1 and 2, the phenomenon can be found that the Bayesian approach proposed in this article can more accurate detect the AO and IO in an ARMA time series than the previous methods in the literatures.

6 Conclusions

In view of some difficulties existed in the detection of outliers in the ARMA time series, this paper suggests many solutions.

Firstly, if the number of the observations from the ARMA time series is very large, calculating the joint probability density function of the observations is very difficult. To solve this problem, a method of estimating the function is considered. And this method also be embed into the Gibbs sampling latter in order to realize the outlier detection for the ARMA time series.

Secondly, in order to simultaneously detect the AO and IO in an ARMA time series, this paper constructs a model which can be used to reflect the observation affected by AO and IO at the same time and conclude the problem of simultaneously detecting AO and IO to a multiple hypothesis testing.

Nextly, the Gibbs sampling is suggested to calculate the posterior probability of every hypothesis, and then the multiple hypothesis is tested based on the principle of Bayesian hypothesis testing, so that the kinds of outliers can be detected.

In addition, this article shows a completely procedure from Bayesian approach for detecting outliers in an ARMA time series after solving the problems mentioned above.

Finally, in order to show the effect of approach proposed in this article, three simulation experiments are designed. The results of this procedure are compared with other existing methods, which shows that the approach for simultaneously detecting AO and IO in an ARMA time series can get the locations with types of outliers accurately and has a better detection results than other methods. Especially, the observation that is both AO and IO can be recognized easily by this procedure proposed in this article.

Acknowledgements: This research was supported by National Science Foundation of China (No.41474009).

References:

[1] R. H. Shumway, D. S. Stoffer, *Time Series Analysis and Its Applications with R Examples,* Springer, 2006.

[2] R. S. Tsay, *Analysis of Financial Time Series,* Wiley, 2002.

[3] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, *Time Series Analysis, Forecasting, and Control,* Prentice Hall, 1994.

[4] Y. Z. Cai, N. Davies, A Simple Diagnostic Method of Outlier Detection for Stationary Gaussian Time Series, *Journal of Applied Statistics*, Vol.30, No.2, 2003, pp. 205-223.

[5] C. Chen, L. M. Liu, Joint Estimation of Model Parameters and Outlier Effects in Time Series, *Journal of the American Statistical Association*, Vol.88, No.421, 1993, pp. 284-297.

[6] R. S. Tsay, Outliers, Level Shifts, and Variance Changes in Time Series, *Journal of Forecasting*, Vol.7, No.1, 1988, pp. 1–20.

[7] K. Choy, Outlier Detection for Stationary Time Series, *Journal of Statistical Planning and Inference*, Vol.99, No.2, 2001, pp. 111–127.

[8] B. Abraham, G. E. P. Box, Bayesian Analysis of Some Outlier Problems in Time Series, *Biometrika*, Vol.66, No.2, 1979, pp. 229-236.

[9] R. E. Mcculloch, R. S. Tsay, Bayesian Analysis of Autoregressive Time Series via the Gibbs Sampler, *Journal of Time Series Analysis*, Vol.15, No.2, 1994, pp. 235-250.

[10] A. Justel, D. Peña, R. S. Tsay, Detection of Outlier Patches in Autoregressive Time Series, *Statistica Sinica*, Vol.11, No.3, 1998, pp. 651-673.

[11] D. Peña, P. Galeano, *Economic Time Series: Modeling and Seasonality,* Chapman & Hall, 2012.

[12] J. Lu, L. Shi, F. Chen, Outlier Detection in Time Series Models Using Local Influence Method, *Communication in Statistics- Theory and Methods*, Vol.41, No.12, 2012, pp. 2202-2220.

[13] H. Louni, Outlier Detection in ARMA Models, *Journal of time series analysis,* Vol.29, No.6, 2008, pp. 1057-1065.

[14] G. E. P. Box, G. C. Tiao, *Bayesian Inference in Statistical Analysis Models*, Addison-Wesley, 1973.

[15] J. Lund, M. Rudemo, Generalised Gibbs Sampler and Multigrid Monte Carlo for Bayesian Computation, *Biometrika*, Vol.87, No.2, 2000, pp. 353-369.

[16] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis,* Wiley, 1985.

[17] A. E. Gelfand, Gibbs Sampling, *Journal of the American Statistical Association,* Vol.95, No.452, 2000, pp. 1300-1304.

[18] R. S. Tsay, Time Series Model Specification in the Presence of Outlier, *Journal of the American Statistical Association*, Vol.81, No.393, 1986, pp. 132-141.