COMPSTAT 2010——The 19th International Conference on Computational Statistics ***Paris-France, August 22-27***

Estimation and detection of outliers and patches in nonlinear time series models

CHEN Ping

Department of Mathematics, Southeast University,
Nanjing, 210096, China
Email: cp18@263.net.cn

1. Introduction

2. Outliers models and identification of ARMAX models

3. Outliers detection for ARMAX model via standard Gibbs sampling

4. Detection of outlier patches via adaptive Gibbs sampling

5. Simulation studies and conclusions

6. References

1. Introduction

Time series observations are often perturbed by some unusual events, such as sudden political factor, economic crises, and even typing or recording errors. Such values are usually referred to as outliers. There may be isolated outliers or patches of outliers in a time series. Outliers may have a significant impact on model identification and parameter estimation for time series. A special case of multiple outliers is a patch of additive outliers, Justel et al.(2001) proposed a procedure to detect outlier patches in an autoregressive process. Chen(1997) did a lot in the detection of additive outliers in bilinear time series. We know that the ARMAX model is more complex than ARMA model. It is widely used in engineering, finance and signal management.

In this paper, based on some different prior distributions, an adaptive Gibbs sampling algorithm is proposed for identifying additive isolated outliers and patches of outliers in nonlinear time series. First, we introduce Outliers models and identification of ARMAX series. Second, we propose Gibbs sampling methods to mine outliers and patches in the view of Bayesian. At last, some case studies show that the algorithm is effective in detecting the locations of outliers and patches and in estimating their size for the ARMAX models and bilinear models.

2. Outliers models and identification of ARMAX series

An ARMA model with input process is called ARMAX model, which is defined as

$$Z_t = \sum_{i=1}^{d} v_i(B) X_{i,t} + n_t, \tag{2.1}$$

where $v_i(B) = (\omega_i(B)/\delta_i(B))B^{k_i}$ is the transfer function of ith input process, $n_t = (\theta(B)/\phi(B))\varepsilon_t$ is noise process, $\{Z_t\}$ is called response process. And $X_{i,t}$ denotes the ith input process or the difference of ith input process at time t, k_i presents the influence's time delay of ith input process, $\{\varepsilon_t\}$ is normal white noise process, $\phi(B)$ and $\theta(B)$ is defined as: $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$, and the B is the backshift operator. When $v_i(B) = 0, i = 1, ..., d$, (2.1) is ARMA model, when some $v_i(B)$ is nonzero constant, $i = 1, ..., d$, then (2.1) is regression model with ARMA error.

The additive outliers(AO) model is as follows:

Suppose that only the jth point z_j is an AO, with influence magnitude β_{tj}, then we have

$$Z_t = \sum_{i=1}^{d} v_i(B) X_{i,t} + \beta_{tj} \delta_{t,tj} + n_t, \tag{2.2}$$

where $\delta_{t,tj}$ is Kronecker symbol: If $t = tj$, then $\delta_{t,tj} = 1$, else $\delta_{t,tj} = 0$.

Considerable simplification in the identification process would occur if the input to the system were white noise. Similar to (2.1), suppose that the ARMAX model of only one input process is as follows:

$$Z_t = \delta^{-1}(B)\omega(B)X_{t-m} + n_t = v(B)X_t + n_t, \qquad (2.3)$$

after 'prewhitening' the input, the cross correlation function between the prewhitened input and correspondingly transformed output is directly proportional to the response function. In practice, we do not know the theoretical function $\rho_{\xi\eta}(k)$, so we must substitute estimates in v_k to give $\hat{v}_k$. The preliminary estimates $\hat{v}_k$ can provide a rough basis for selecting suitable transfer function model. First, we may use the estimates $\hat{v}_k$ so obtained to make guesses of the order $r1$ and $r2$ of $\delta(B)$ and $\omega(B)$, and of the delay parameter m. Second, we do not consider the noise n_t now, substituting $Z_t = \hat{v}(B)X_t$ in the equation $\delta(B)Z_t = \omega(B)B^m X_t$, based on equating coefficients of B , to obtain initial estimates of the parameters $\delta(B)$ and $\omega(B)$ in (2.3).

3. Outliers detection for ARMAX model via standard Gibbs sampling

We detect AO type outliers in ARMAX model by Gibbs sampling based on Bayesian method. The idea is as follows: Suppose the probability that observation is outlier has some prior information. Based on the method of conjugate priors, we proved some theorems which gives the expressions of some posterior distributions, then we compute the posterior probabilities for each data point to be an AO type outlier using techniques of Gibbs sampling. If the posterior probability is larger than some prescribed value, then we consider it as an AO type outlier.

Since output z_t may be an outlier at each time point, we let $\delta_t = 1$ if the observation at this time point is an additive outlier. Let $\delta_t = 0$ if it is an outlier-free time point, and denote $P(\delta_t = 1) = \alpha$. Then the general ARMA model with additive outlier is as follows

$$\begin{cases} y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\ z_t = y_t + \delta_t \beta_t, \end{cases} \qquad \varepsilon_t \sim N(0, \sigma^2). \qquad (3.1)$$

It means that the observation z_t may be AO with probability α, its magnitude is β_t at time t . For simplicity, assume that $y_1, ..., y_p$ are fixed and $z_t = y_t$ for $t = 1, ..., p$, i.e. there exist no outliers in the first p observations. The indicator vector of outliers then becomes $\delta = (\delta_{p+1}, \delta_{p+2}, ..., \delta_n)'$ and the size vector is $\beta = (\beta_{p+1}, \beta_{p+2}, ..., \beta_n)'$. Let $\hat{\varepsilon}_t$ denote the residual estimation of model (3.1) without AO. And let $\hat{\varepsilon} = (\hat{\varepsilon}_1, ..., \hat{\varepsilon}_{n-1})'$, $\varepsilon = (\varepsilon_1, ..., \varepsilon_{n-1})'$, $\Theta = (\phi_1, \phi_2, ..., \phi_p, \theta_1, \theta_2, ..., \theta_q)'$, $z = (z_1, z_2, ..., z_n)'$ and $\Phi_{t-1} = (y_{t-1}, y_{t-2}, ..., y_{t-p}, \hat{\varepsilon}_{t-1}, ..., \hat{\varepsilon}_{t-q})'$. By $y_t = \Theta'\Phi_{t-1} + \varepsilon_t$, we can obtain the likelihood function

$$L(\Theta, \sigma^2, \delta, \beta, \alpha \mid z, \hat{\varepsilon}) \propto \sigma^{n-p} exp\{-\frac{1}{2\sigma^2} \sum_{t=p+1}^{n} (y_t - \Theta'\Phi_{t-1})^2\}, \quad (3.2)$$

where $y_t = z_t - \delta_t \beta_t$.

For computational reason, we use conjugate prior distribution for parameters Θ and σ^2, which distributed as multidimensional uniform distribution on [0,1] region and inverted-Gamma distribution $IG(\frac{\nu}{2}, \frac{\nu\lambda}{2})$ respectively. Assume that the outlier indicator δ_t and the outlier magnitude β_t are independent and distributed as $Bernoulli(\alpha)$ and $N(0, \tau^2)$ respectively for all t. Then, the prior probability of being contaminated by an outlier is the same for all observations, namely $P(\delta_t = 1) = \alpha$, for $t = p + 1, ..., n$. The prior distribution of the contamination parameter α is $Beta(\gamma_1, \gamma_2)$, and $(\beta_t)'s$ are $i.i.d.$ for all t. The hyperparameters in our model are $\lambda, \nu, \gamma_1, \gamma_2$ and τ^2, all of which are assumed known.

It is obvious that we must obtain the conditional posterior distributions of parameters $(\Theta, \beta, \delta, \alpha, \sigma^2)$ for detecting outliers in the model. The important thing is to conduct the conditional posterior distributions of $\delta_j = 1$ and β_j. Let $z = (z_1, z_2, ..., z_n)$ denote the observed vector of model (3.1). Using the standard Bayesian method, under the above conditions, we have the following results:

Theorem (1) For the conditional posterior distribution of $\delta_j = 1$, we have

$$p(\delta_j = 1 | z, \delta_{(-j)}, \beta, \Theta, \sigma^2, \alpha) = [1 + \frac{1-\alpha}{\alpha} B_{10}(j)]^{-1}$$

where $\delta_{(-j)}$ is obtained from δ by eliminating the element δ_j, $T_j = min\{n, j+p\}$, and

$$B_{10}(j) = exp\{\frac{1}{2\sigma^2}[\sum_{t=j}^{T_j} a_t^2(1) - \sum_{t=j}^{T_j} a_t^2(0)]\}$$

where $a_t(1) = (y_t - \Theta'\Phi_{t-1})_{\delta_j=1}$, $a_t(0) = (y_t - \Theta'\Phi_{t-1})_{\delta_j=0}$ and $a_t(0) = a_t(1) + \varphi_{t-j}\beta_j$. When $p = 0$, then $\varphi_i = 0$ for all i ; when $p > 0$, we have that $\varphi_i = -1$ if $i = 0$, $\varphi_i = \phi_i$ if $i = 1, ..., p$, and $\varphi_i = 0$ if $i \geq p + 1$.

(2) When $\delta_j = 0$, there is not any new information about the posterior distribution of β_j, namely, β_j distributed as $N(0, \tau^2)$. When $\delta_j = 1$, since z_t contains information of β_j, so we have that

$$p(\beta_j | z, \delta, \beta_{(-j)}, \Theta, \sigma^2, \alpha) \sim N(\beta_j^*, \tau_j^*),$$

where $\beta_{(-j)}$ is obtained from β by eliminating the element β_j, and $\beta_j^* = \frac{\mathbb{A}\tau^2}{\mathbb{B}\tau^2+\sigma^2}$, $\tau_j^* = \frac{\sigma^2\tau^2}{\mathbb{B}\tau^2+\sigma^2}$. where $\mathbb{A} = -\sum_{t=j}^{T_j} a_t(1)\varphi_{t-j}$, *and* $\mathbb{B} = \sum_{t=j}^{T_j} \varphi_{t-j}^2$.

We give the Gibbs method for detecting AO in ARMAX model as follows: **a.** Given the starting point $(\lambda, \nu, \gamma_1, \gamma_2, \tau^2, \alpha)$, where-after this algorithm iterates the following loop: **b.** sample $\Theta^{(t)}$ from $p(\Theta|z, \delta^{(t-1)}, \beta^{(t-1)}, \alpha^{(t-1)}, \sigma^{2(t-1)})$, **c.** sample $\sigma^{2(t)}$ from $p(\sigma^2|z, \Theta^{(t)}, \delta^{(t-1)}, \beta^{(t-1)}, \alpha^{(t-1)})$, **d.** sample $\delta_j^{(t)}$ from $p(\delta_j|z, \Theta^{(t)}, \sigma^{2(t)}, \beta^{(t-1)}, \alpha^{(t-1)})$, **e.** sample $\beta_j^{(t)}$ from $p(\beta_j|z, \Theta^{(t)}, \sigma^{2(t)}, \delta^{(t)}, \alpha^{(t-1)})$, **f.** sample $\alpha^{(t)}$ from $Beta(\gamma_1 + k, \gamma_2 + n - p - k)$, repeat the above steps till it is convergence. From the Bayesian principle, we suppose that if the outlying posterior probability is larger then c_1, then believe it is an outlier.

4. Detection of outlier patches via adaptive Gibbs sampling

Similar to Justel et al.(2001), our procedure also consists of two Gibbs runs. In the first run, the standard Gibbs sampling based on the results of Section 3 is carried out. The results of this Gibbs run are then used to implement a second Gibbs sampling that is adaptive in treating identified outliers and in using block interpolation to reduce possible masking and swamping effects. Let $\hat{\Theta}^{(s)}$, $\hat{\sigma}^{(s)}$, $\hat{\beta}^{(s)}$ and $\hat{\alpha}^{(s)}$ be the posterior means of Θ, σ^2, β and α respectively based on the s iterations of the first Gibbs run. First, we select a appropriate critical value c_1 to identify potential outliers. An observation z_j is identified as an outlier if the posterior probability $\hat{p}_j^{(s)} > c_1$. Let

$\{t_1, ..., t_m\}$ be the collection of time indexes of outliers identified by the first Gibbs run. Second, let $c_2, c_2 \leq c_1$ be another appropriate critical value to specify the beginning and end points of a potential outlier patch. We select a window of length $2p$ around the identified outlier to search for the boundary points of a possible outlier patch by a forward-backward method. For example, consider an identified outlier z_{t_i} . For the p observations before z_{t_i} , if their posterior probabilities $\hat{p}_j^{(s)} > c_2$, then these points are regarded as possible outlier patch associated with z_{t_i}. We then select the farthest point from z_{t_i} as the begining point of the outlier patch. Denote the point by $z_{t_i - k_i}$. Then we do the same for the p observations after z_{t_i} and select the farthest point from z_{t_i} with $\hat{p}_j^{(s)} > c_2$ as the end point of the outlier patch. Denote the end point by $z_{t_i+v_i}$. Combine the two blocks to form a possible outlier patch associated with z_{t_i}, which denoted by $(z_{t_i-k_i}, \cdots, z_{t_i+v_i})$. Consecutive or overlapping patchs should be merged to form a larger patch. Lastly, draw Gibbs samples jointly within a patch. Suppose that a patch of k outliers starting at time index j is specified. Denote the vectors of outlier indicators and magnitudes by $\delta_{j,k} = (\delta_j, \cdots, \delta_{j+k-1})'$ and $\beta_{j,k} = (\beta_j, \cdots, \beta_{j+k-1})'$, respectively, associated with the patch. Similar to the Theorem 1 of Justel et al.(2001), we may obtain the conditional posterior distributions of $\delta_{j,k}$ and $\beta_{j,k}$.

For the second adaptive Gibbs sampling, we use the results of the first Gibbs run to start the second Gibbs sampling and to specify prior distributions of the parameters. For each outlier patch, we use the conditional posterior distributions to draw $\delta_{j,k}$ and $\beta_{j,k}$ in the second Gibbs sampling, which is also run for s iterations. The starting values of δ_t are as follows: $\delta_t^{(0)} = 1$ if $\hat{p}_t^{(s)} > 0.3$, otherwise, $\delta_t^{(0)} = 0$. Then the prior distributions of β_t are as follows.

(a) If z_t is identified as an isolated outlier, then the prior distribution of β_t is $N(\hat{\beta}_t^{(s)}, \tau^2)$, where $\hat{\beta}_t^{(s)}$ is the Gibbs estimate of β_t from the first Gibbs run.

(b) If z_t belongs to an outlier patch, then the prior distribution of β_t is $N(\tilde{\beta}_t^{(s)}, \tau^2)$, where $\tilde{\beta}_t^{(s)}$ is the conditional posterior mean as follows:

$$\tilde{\beta}_{j,k} = (D_{j,k} \sum_{t=j}^{T_{j,k}} \Pi_{t-j} \Pi'_{t-j} D_{j,k})^{-1} (- \sum_{t=j}^{T_{j,k}} e_t(0) D_{j,k} \Pi_{t-j})$$

(c) If z_t does not belong to any outlier patch, and is not an isolated outlier, then the prior distribution of β_t is $N(0, \tau^2)$.

5. Simulation studies and conclusions

Example A. In the simulations, we consider the ARMAX(1,1,2) model:

$$\begin{cases} (1 - 0.78B + 0.3B^2)x_t = e_t \\ (1 - 0.7B)y_t = (1 - 0.7B)(1 - 0.48B)x_t + (1 - 0.27B + 0.96B^2)\varepsilon_t \\ z_t = y_t - 11\delta_{t,31} + 10\delta_{t,32} - 9\delta_{t,33} + 10\delta_{t,34} - 9\delta_{t,35} + 10\delta_{t,50}, \end{cases} \tag{5.1}$$

where $\{e_t\}$ and $\{\varepsilon_t\}$ are all normal white noise, their means are zero and variance $\sigma^2 = 1$.

We create 101 observations $x_0, x_1, \cdots, x_{100}$ of x_t and 100 observations $z_1, z_2, \cdots, z_{100}$ of z_t by simulation. It is obvious that the input process is an AR(2) series, the transfer function of the ARMAX model is $(1 - 0.48B)$, a patch of five consecutive additive outliers have been introduced from $t = 31$ to $t = 35$, a single AO has been add at $t = 50$, and the outlier magnitudes are $\beta_{31} = -11, \beta_{32} = 10, \beta_{33} = -9, \beta_{34} = 10, \beta_{35} = -9$ and $\beta_{50} = 10$ respectively. Applying our method to the above simulate series $\{z_t\}$

and prewhitening the input series. Making $\{x_t\}$ follows an ARMA model: $(1 - 0.41117B)x_t = \varepsilon_t$. Then we take the same manipulation of prewhitening the $\{z_t\}$. By analyzing filtered cross correlation coefficient of $\{z_t\}$ and $\{x_t\}$, we obtain the transfer function $1.00597 - 0.1568B$ for $\{x_t\}$. Note that the transfer function was influenced by outliers. In order to delete the influence of input process $\{x_t\}$ in response process $\{z_t\}$, we let $z_t^* = z_t - (1.00597 - 0.1568B)x_t$. Then we can apply the method described above to detect the outliers in $\{z_t^*\}$, which are just the outliers of $\{z_t\}$.

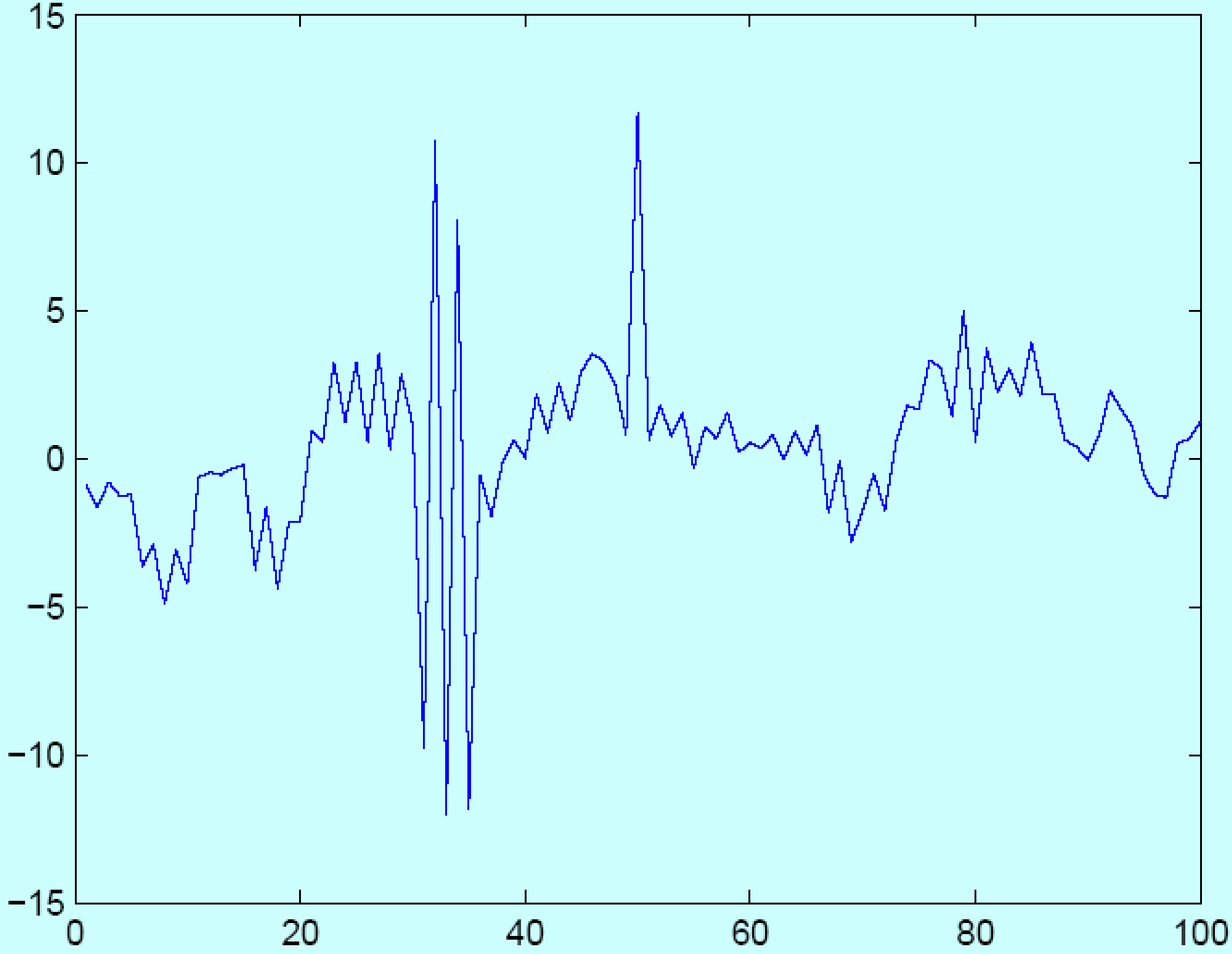

Fig.1 The curve of observations $\{z_t\}$ in example A.

Let $\gamma_1 = 5$, $\gamma_2 = 95$, $\nu = 3$, $\lambda = \widetilde{\sigma}^2/3$, $\alpha = 0.5$, $c_1 = 0.5$, $c_2 = 0.3$ and $\tau = 3\widetilde{\sigma}$, where $\widetilde{\sigma}^2$ is mean square error of $\{z_t\}$. Here $\gamma_1 = 5$ means that we believe the prior probability of each point is an outlier approximate to 0.05. First, we detect the outliers in $\{z_t\}$ using the standard Gibbs sampling. Limit to the computational ability, we take 100 iterations by usual Gibbs algorithm. We obtain the posterior probabilities that each observation is outlier, the Figure 2 shows that the posterior probabilities of being an outlier for data points at $t = 31, 33, 34, 35, 36, 50$ are large. Meanwhile, the outlying posterior probabilities of other observations are low. We see that the algorithm fails to detect the inner point at $t = 32$ of the patch, resulting in the masking effect. On the other hand, the algorithm misspecifies the

’good’ data point at $t = 36$ as outlier because the outlying posterior probability of this point is larger, so the ’good’ data point at $t = 36$ is swamped by the patch of outliers. Second, in order to avoid the presence of masking and swamping problem, we use the method given by section 4, and take 900 iterations by the adaptive Gibbs algorithm. The Figure 3 shows the posterior probability of outlier for each data point, which clearly shows that the patch of outliers and the isolated outlier in $\{z_t\}$ process are detected triumphantly, and there is not any misjudgement. The posterior means of the sizes of these outliers are $\hat{\beta}_{31} = -5.9849$, $\hat{\beta}_{32} = 14.4163$, $\hat{\beta}_{33} = -16.5910$, $\hat{\beta}_{34} = 14.4383$, $\hat{\beta}_{35} = -14.6142$ and $\hat{\beta}_{50} = 3.8853$, respectively.

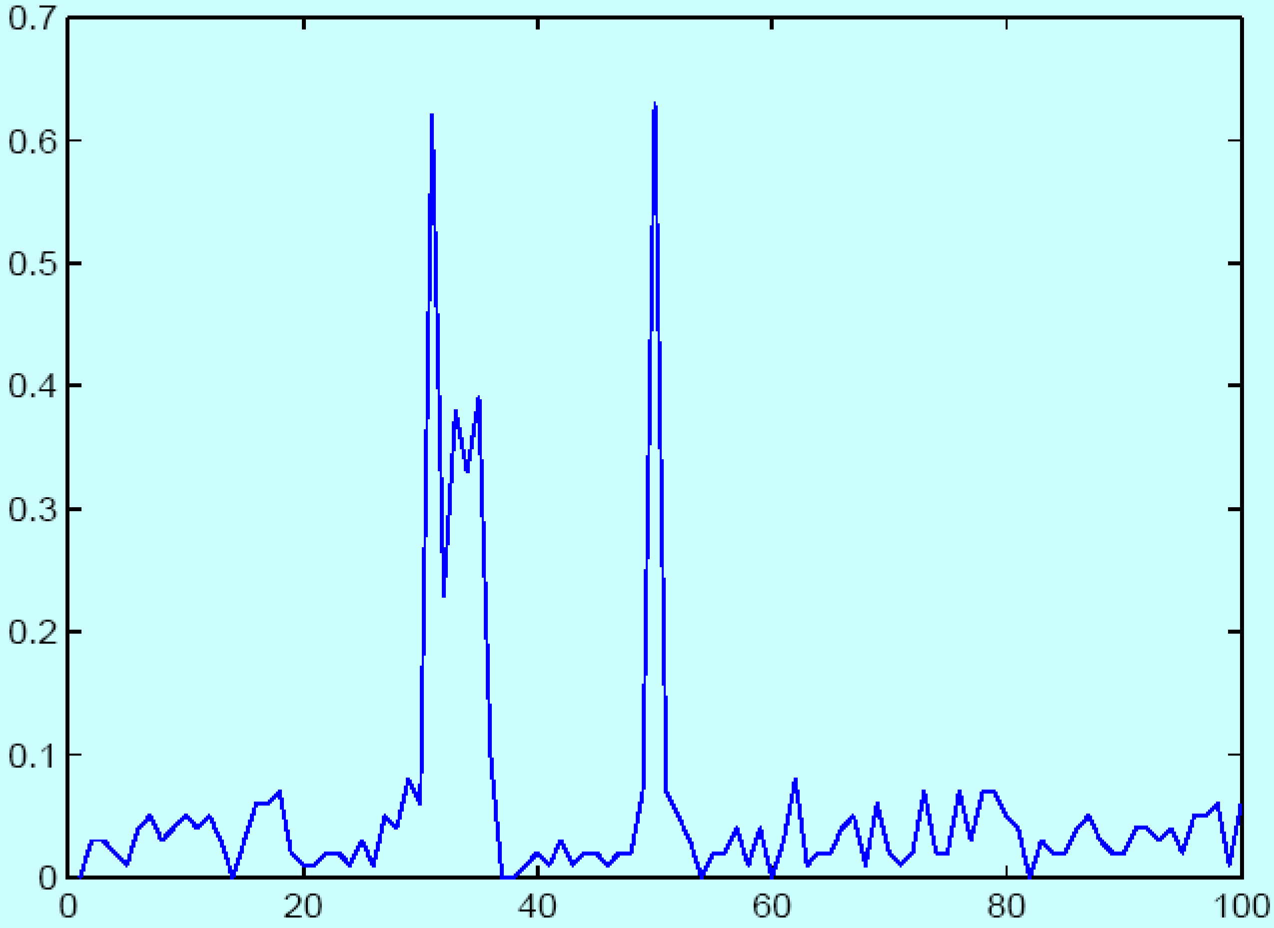

Fig.2 The posterior probability that each point is an outlier via standard Gibbs sampling in example A.

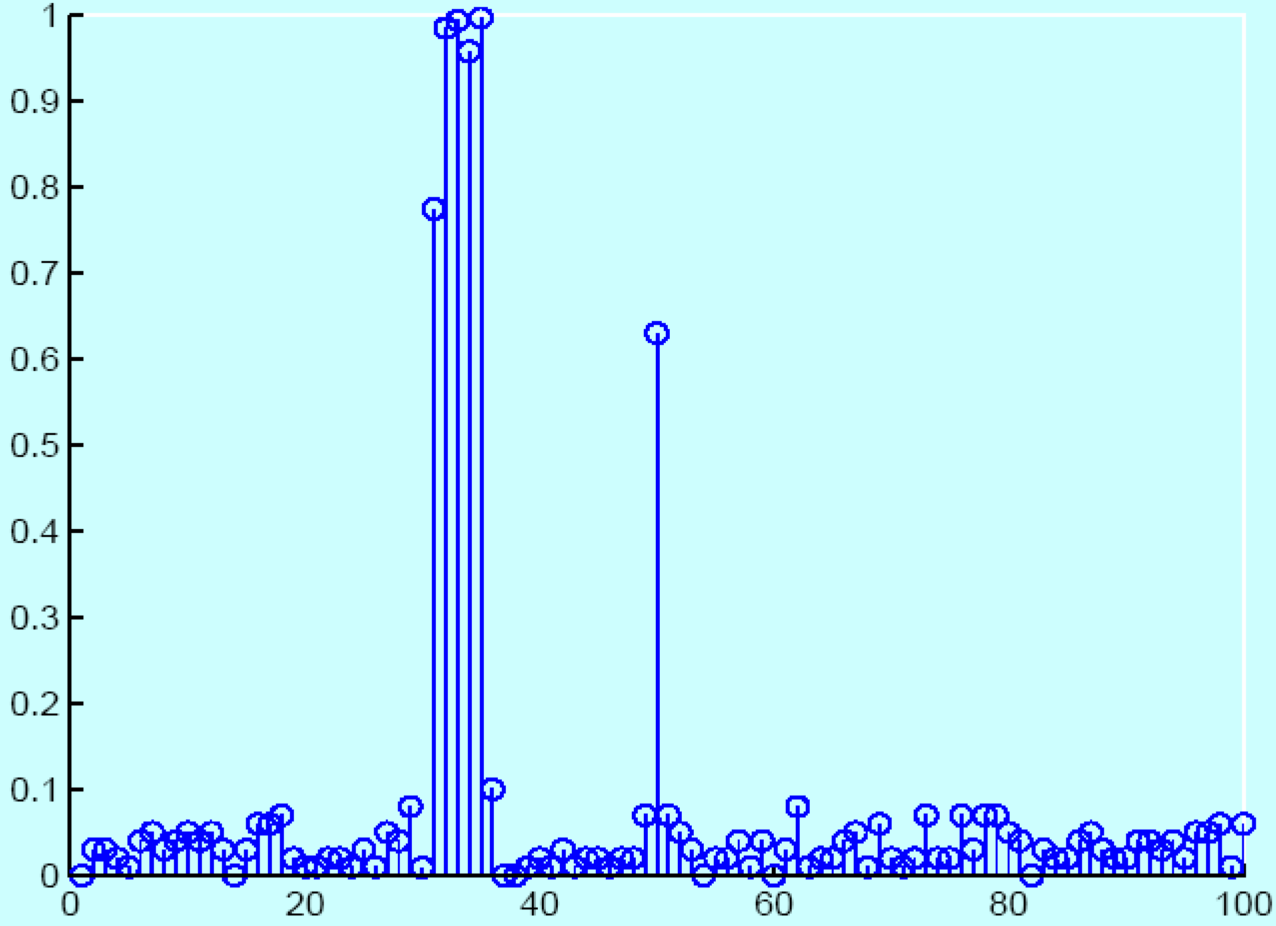

Fig.3 The posterior probability that each point is an outlier via adaptive Gibbs algorithm in example A.

Similar to the method above, we also could detect the outliers and patches in bilinear time series models by adaptive Gibbs sampling algorithm. We omit the theoretics and give an example as follows

Example B. In the following example, we consider the BL(1,1,1,1) model:

$$\begin{cases} y_t = 0.7y_{t-1} + \varepsilon_t - 0.3\varepsilon_{t-1} + 0.31y_{t-1}\varepsilon_{t-1} \\ z_t = y_t + 4\delta_{t,40} - 3\delta_{t,41} + 3\delta_{t,42}, \end{cases} \tag{5.2}$$

where $\{\varepsilon_t\}$ is standard normal white noise.

We create a set of observations of the bilinear model (5.2) by simulation, where a patch of three consecutive additive outliers have been introduced from $t = 40$ to $t = 42$, and the outlier magnitudes are $\beta_{40} = 4, \beta_{41} = -3$ and $\beta_{42} = 3$ respectively. The figure shows that the curve of observations has large volatility, it would be very difficult to distinguish between 'outliers' and normal points of nonlinear model.

First, we detect the outliers in $\{z_t\}$ using the standard Gibbs sampling. We obtain the posterior probabilities that each observation is outlier, the Figure 5 obviously shows that the posterior probabilities of being outlier only at $t = 40$ is large than 0.5. Meanwhile,

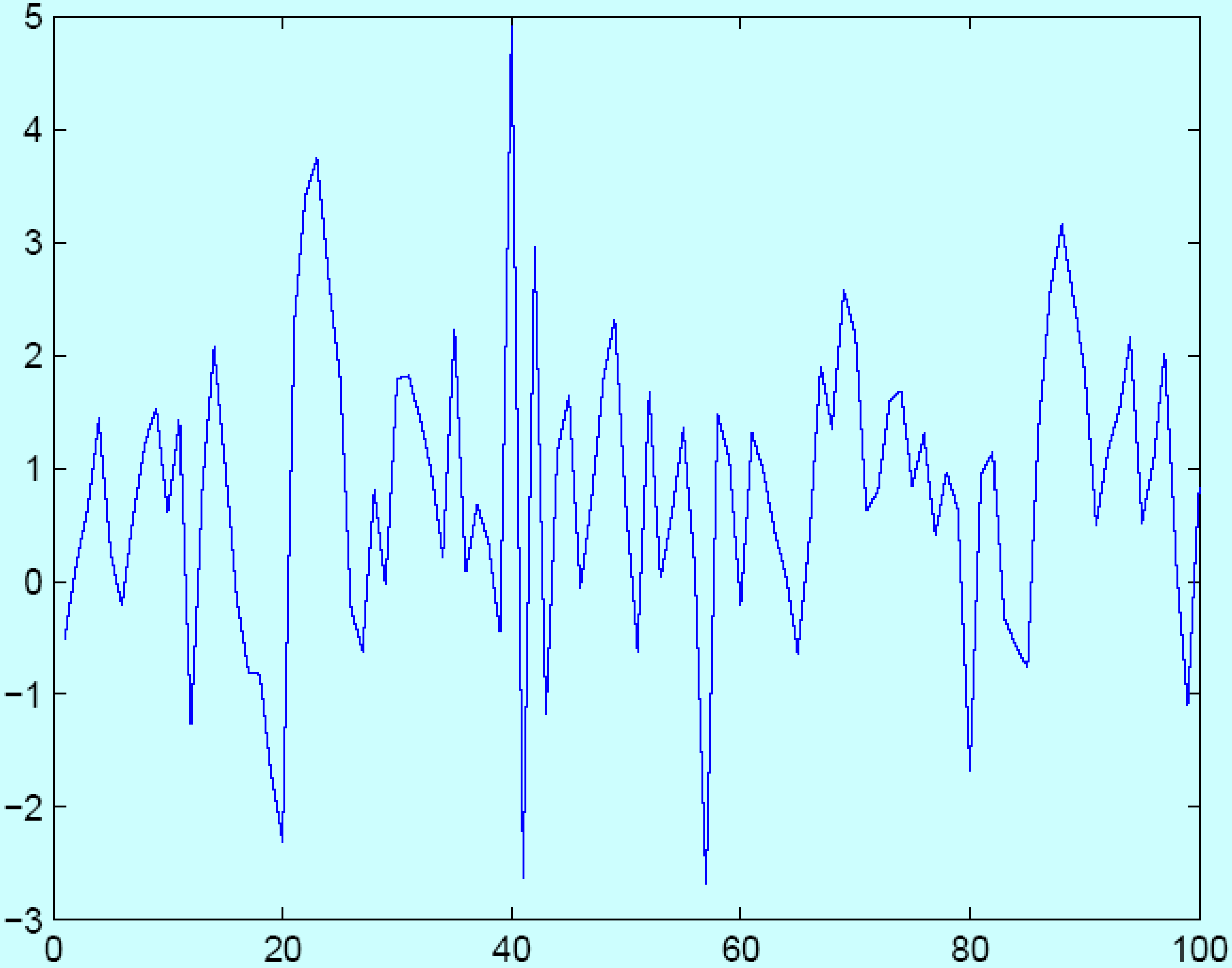

Fig.4 The curve of observations $\{z_t\}$ in example B.

the outlying posterior probabilities of other observations are low, the posterior probability of being an outlier at $t = 41$ is even small than 0.2. However, the posterior probability of being an outlier at $t = 57$ is larger than at $t = 41, 42$. We see that the standard Gibbs sampling fails to detect the inner and border points at $t = 41, 42$. resulting in the masking effect. On the other hand, the algorithm is apt to misspecify the 'good' data point at $t = 57$ as outlier because the outlying posterior probability of this point is larger than every points but the point at $t = 40$, so that the 'good' data point at $t = 57$ may be swamped by the patch of outliers.

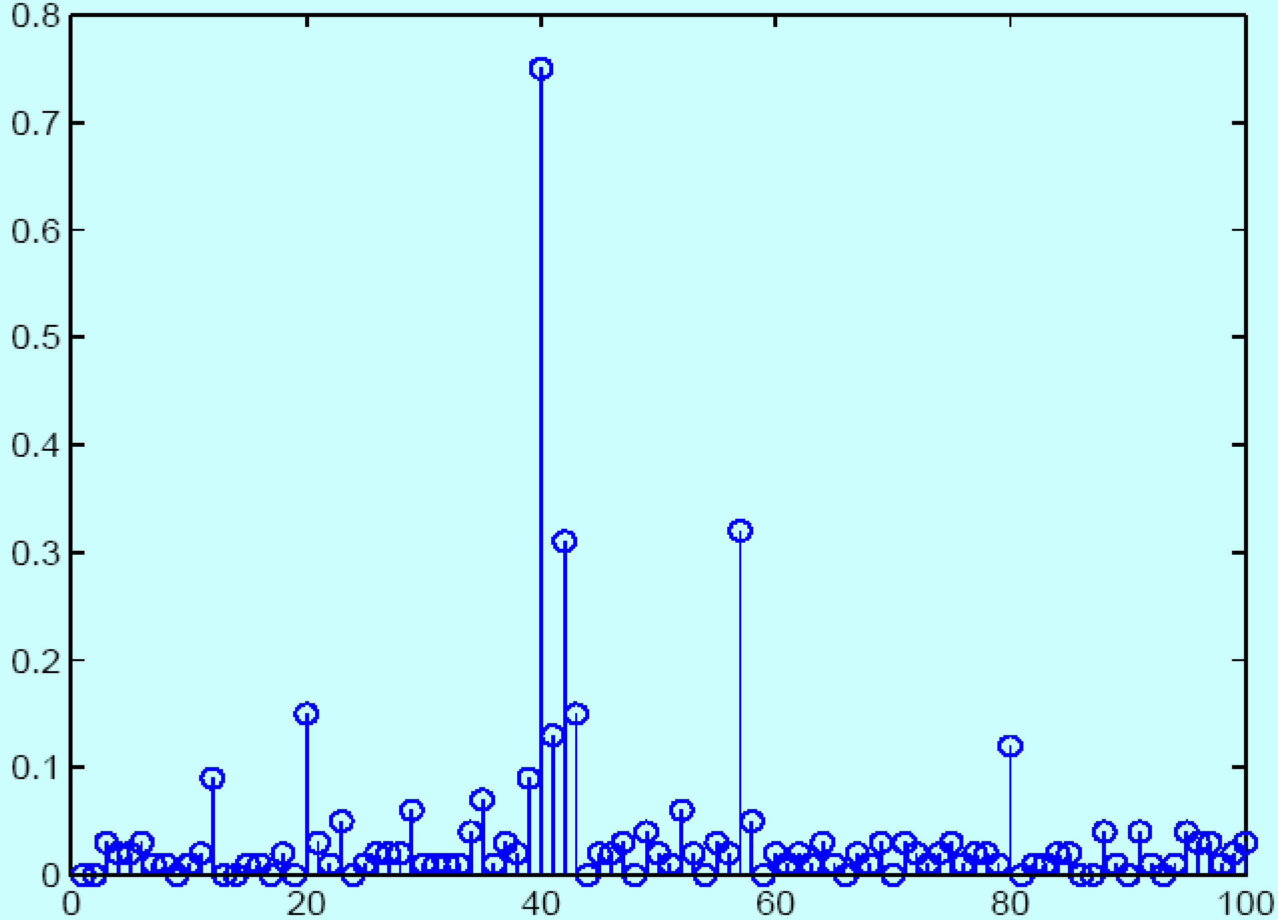

Fig.5 The posterior probability that each point is an outlier via standard Gibbs sampling in Example B.

Second, in order to avoid the presence of masking and swamping problem, we apply the similar method given by section 4, and take 900 iterations by the adaptive Gibbs algorithm. In the process of running, we let the initial distribution of Θ be $N(\mathbf{0}, 0.3\mathbf{I})$. The Gibbs sampler was repeated several times with different hyperparameters and different numbers of iterations to reanalyze the data. The results show that the locations of possible outliers and patch are stable, even though the estimated outlying probabilities may vary slightly between the Gibbs samples. The figure 6 gives the time plot of the estimated posterior probability that each point is an outlier via adaptive Gibbs sampling, the window width of search was 4.

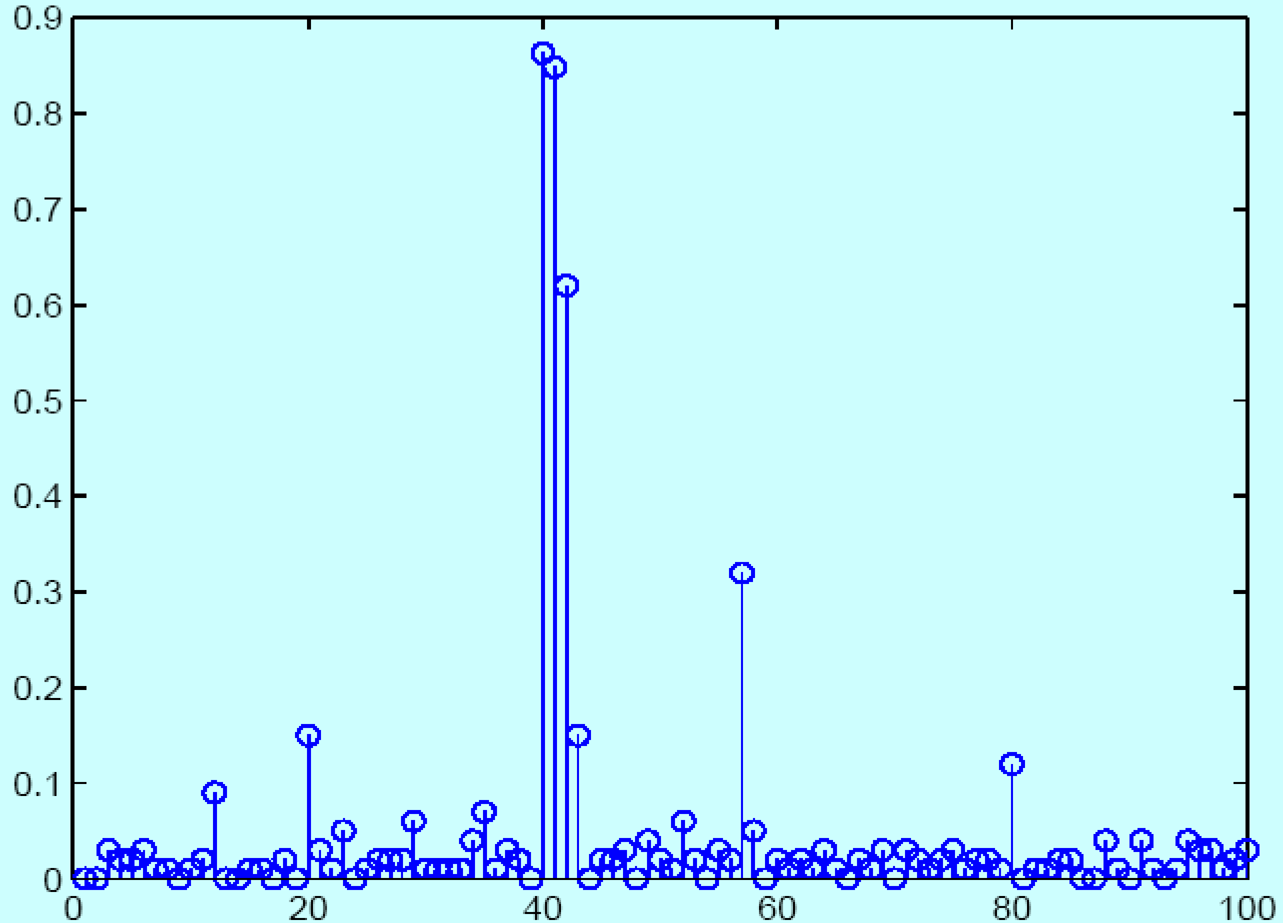

Fig.6 The posterior probability that each point is an outlier via adaptive Gibbs algorithm in Example B.

It obviously shows that the posterior probability of being an outlier obtained by adaptive Gibbs algorithm for data points at $t = 40, 41, 42$ are large. Meanwhile, the outlying posterior probabilities of other observations are low. Actually, if select the critical value $c_2 = 0.3$, then we could identify the patch. On the other hand, many normal points which were similar to outliers do not be misspecified as outliers because the outlying posterior probabilities of these points are smaller than 0.3, which show that the adaptive Gibbs sampling is effective in mining the additive outlier patch of bilinear time series.

Furthermore, the posterior means of the sizes of these outliers are $\hat{\beta}_{40} = 4.1652$, $\hat{\beta}_{41} = -5.7693$ and $\hat{\beta}_{42} = 3.4172$, respectively. By a number of simulations which detect the patches in bilinear model by adaptive Gibbs sampling, we discovered that the critical value c_2 should be selected smaller than ARMA model. Investigate its reason, it may be the volatility of bilinear series is larger than ARMA series, and itself could often produce some normal points which appear to be outliers.

References

BOX, G.E.P., JENKINS, G.M. and REINSEL, G.C.(1994): *Time Series Analysis: Forecasting and Control, Third edition*. Prentice-Hall & Englewood Cliffs, NJ.

CHEN, C. W. S.(1997):Detection of additive outliers in bilinear time series. *Comput. Statist. Data Anal. 24, 283-294*.

CHEN, P., LI, L., LIU, Y. and LIN, J.G.(2010). Detection of outliers and patches in bilinear time series models. *Mathematical Problems in Engineering, vol.2010, Article ID 580583, 10 pages*.

CHEN, P., YAN, F.R., WU, Y.Y. and CHEN, Y. (2009): Detection of outliers in ARMAX time series models. *Advances in Systems Science and Applications 9, 97-103*.

JUSTEL, A., PEÑA, D. and TSAY, R.S. (2001): Detection of outlier patches in autoregressive time series. *Statistica Sinica 11, 651-673*.

MCCULLOCH, R.E., TSAY, R.S.(1994): Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis 15, 235-250.*

Thank You!