

## *Research Article*

# **Detection of Outliers and Patches in Bilinear Time Series Models**

**Ping Chen, Ling Li, Ye Liu, and Jin-Guan Lin**

*Department of Mathematics, Southeast University, Nanjing 210096, China*

Correspondence should be addressed to Jin-Guan Lin, [jglin@seu.edu.cn](mailto:jglin@seu.edu.cn)

Received 10 January 2010; Accepted 10 February 2010

Academic Editor: Ming Li

Copyright © 2010 Ping Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a Gibbs sampling algorithm to detect additive outliers and patches of outliers in bilinear time series models based on Bayesian view. We first derive the conditional posterior distributions, and then use the results of first Gibbs run to start the second adaptive Gibbs sampling. It is shown that our procedure could reduce possible effects on masking and swamping. At last, some simulations are performed to demonstrate the efficacy of detection and estimation by Monte Carlo methods.

## **1. Introduction**

Dynamic systems or engineering time series observations [1] are often perturbed by some interrupting phenomena which generate aberrant data. They may be due to some unusual events, such as sudden disturbed factor, noise, and even recording errors. Such values are usually referred to as outliers, which may be isolated or of patch. Because outliers and patches in a time series can have adverse effects in data analysis, which maybe make the resultant inference unreliable or even invalid, it is important to detect and remove such outlier's effects. Some authors have considered the detection problem in the linear and nonlinear models. Abraham and Box [2] analyzed some outlier problems based on Bayesian methods in time series. Tsay [3] considered time series model specification in the presence of outliers. Ljung [4] presented outlier detection in time series. Justel et al. [5] studied detection of outlier patches in autoregressive time series. Battaglia and Orfei [6] considered outlier detection and estimation in nonlinear time series. Chen et al. [7] studied similar problem in ARMAX time series models. P. Chen and Y. Chen [8] gave the identification of outliers in ARMAX models via genetic algorithm.

Based on some different prior distributions, a new Gibbs sampling algorithm is proposed for identifying additive isolated outliers and patches of outliers in bilinear time series. This paper considers detection of outliers and patches in bilinear time series models, which are one of the fractal time series models [9]. Wang and Wei [10] analyzed the probabilistic structure for a rather general class of bilinear models systematically. Wang [11] considered parameter estimation and subset selection for separable lower triangular bilinear models. In particular, Chen [12] proposed a procedure for detecting additive outliers in bilinear time series, which also discusses some major problems encountered in practice, such as how one can distinguish between ARMA model with outliers and a bilinear model without outliers. Some special cases of bilinear models, such as diagonal bilinear model, vector bilinear model and so on, have been studied by many authors on not only probabilistic structure but also statistical inferences; for instance, Subba Rao [13], Subba Rao and Gabr [14], Kim et al. [15], and Sesay and Subba Rao [16], among others.

The format of the paper is as follows. Section 2 gives the outliers model of bilinear series. Section 3 presents a method to detect outliers in bilinear model via standard Gibbs sampling. In Section 4, we propose an adaptive Gibbs sampling method to detect patches of outliers in bilinear model. Section 5 gives simulated examples and some conclusions.

## 2. Outliers Models of Bilinear Time Series

The time series  $\{y_t\}$  is called a bilinear process, if it satisfies the model

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^r \sum_{j=1}^k \eta_{ij} y_{t-i} \varepsilon_{t-j}, \quad (2.1)$$

where  $\varepsilon_t \sim \text{IIDN}(0, \sigma^2)$ , and  $\phi_i, \theta_j$  and  $\eta_{ij}$  are unknown parameters. This model is called a bilinear model with order  $(p, q, r, k)$ . The bilinear time series models were introduced by Granger and Anderson [17]. It is a generalization of the linear ARMA model by incorporating the product term of time series  $y_{t-i}$  and innovation  $\varepsilon_{t-j}$ . This bilinear model has attracted much attention in the recent time series literature. In terms of potential applications, bilinear models are known to be able to mode occasional outbursts in time series, which might be useful for modeling seismological data such as records for explosions and earthquakes.

When additive outliers (AO) are present,  $y_t$  is disturbed and unobservable. In this case, it is assumed that the observed series  $\{z_t\}$  follows

$$z_t = y_t + \beta_t \delta_t, \quad (2.2)$$

where  $\delta_t$ 's are independent and identically distributed Bernoulli random variables with  $P(\delta_t = 1) = \alpha$ , and  $\beta_t$ 's are random variables from normal distribution. That means that the observation  $z_t$  may be AO with probability  $\alpha$ ; its magnitude is  $\beta_t$  at time  $t$ .

For simplicity, let  $s = \max\{p, r\}$  and assume that  $y_1, \dots, y_s$  are fixed and  $z_t = y_t$  for  $t = 1, \dots, s$ , that is, there exist no outliers in the first  $s$  observations. The indicator vector of outliers then becomes  $\boldsymbol{\delta} = (\delta_{s+1}, \delta_{s+2}, \dots, \delta_n)'$  and the size vector is  $\boldsymbol{\beta} = (\beta_{s+1}, \beta_{s+2}, \dots, \beta_n)'$ . Assume that a vector of residuals  $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_{n-1})'$  of model (2.1) is available. And let  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_{n-1})'$ ,  $\boldsymbol{\Theta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \eta_{11}, \dots, \eta_{rk})'$ ,  $\mathbf{z} = (z_1, z_2, \dots, z_n)'$ , and

$\Phi_{t-1} = (\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}, \hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q}, \mathbf{y}_{t-1}\hat{\varepsilon}_{t-1}, \dots, \mathbf{y}_{t-r}\hat{\varepsilon}_{t-k})'$ , where  $\hat{\varepsilon}_j$  denotes the estimation of  $\varepsilon_j$  in model (2.1) without AO. Then the model (2.1) and (2.2) may be transformed as the follows

$$\begin{aligned} \mathbf{y}_t &= \Theta' \Phi_{t-1} + \varepsilon_t, \\ z_t &= \mathbf{y}_t + \beta_t \delta_t. \end{aligned} \quad (2.3)$$

### 3. Outliers Detection via Standard Gibbs Sampling

In order to detect outliers in the bilinear model, it is essential to derive the conditional posterior distributions of parameters  $\Theta, \beta, \delta, \alpha, \sigma^2, \delta_j = 1$ , and  $\beta_j$ .

For computational reason, we give the following conditions: (1) using conjugate prior distribution for parameters  $\Theta$  and  $\sigma^2$ , which distributed as multidimensional uniform distribution on  $[0, 1]$  region and inverted-Gamma distribution  $IG(\nu/2, \nu\lambda/2)$ , respectively. (2) Assume that the outlier indicator  $\delta_t$  and the outlier magnitude  $\beta_t$  are independent and distributed as *Bernoulli*( $\alpha$ ) and  $N(0, \tau^2)$ , respectively for all  $t$ . (3) The prior distribution of the contamination parameter  $\alpha$  is *Beta*( $\gamma_1, \gamma_2$ ), and  $(\beta_t)$ 's are i.i.d. for all  $t$ . (4) The hyperparameters in our model are  $\lambda, \nu, \gamma_1, \gamma_2$ , and  $\tau^2$ , all of which are assumed to be known.

Note that the prior probability of being contaminated by an outlier is the same for all observations, namely,  $P(\delta_t = 1) = \alpha$ , for  $t = s + 1, \dots, n$ . Then, under these conditions and by using the standard Bayesian method, we can obtain the following results.

**Theorem 3.1.** *The conditional posterior distribution of  $\Theta$  is  $N(\Theta^*, V_*^{-1})$ , where*

$$\Theta^* = V_*^{-1} \left( \frac{\sum_{t=s+1}^n \Phi_{t-1} \mathbf{y}_t}{\sigma^2} \right), \quad V_* = \frac{\sum_{t=s+1}^n \Phi_{t-1} \Phi_{t-1}'}{\sigma^2}. \quad (3.1)$$

*Proof.* Under the conditions above, we have that

$$\begin{aligned} p(\Theta | \mathbf{z}, \delta, \beta, \alpha, \sigma^2) &\propto L(\Theta, \sigma^2, \delta, \beta, \alpha | \mathbf{z}) p(\Theta) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=s+1}^n (\mathbf{y}_t - \Theta' \Phi_{t-1})^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} \left[ \Theta' \sum_{t=s+1}^n \Phi_{t-1} \Phi_{t-1}' \Theta - 2\Theta' \sum_{t=s+1}^n \Phi_{t-1} \mathbf{y}_t + \sum_{t=s+1}^n \mathbf{y}_t^2 \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \Theta' \sum_{t=s+1}^n \frac{\Phi_{t-1} \Phi_{t-1}'}{\sigma^2} \Theta - 2\Theta' \frac{\sum_{t=s+1}^n \Phi_{t-1} \mathbf{y}_t}{\sigma^2} + \Theta^* \frac{\Phi_{t-1} \Phi_{t-1}'}{\sigma^2} \Theta^* \right] \right\} \\ &\vdots \\ &\propto \exp \left\{ -\frac{1}{2} (\Theta - \Theta^*)' V_* (\Theta - \Theta^*) \right\} \\ &\sim N(\Theta^*, V_*^{-1}). \end{aligned} \quad (3.2)$$

This completes the proof of the theorem.  $\square$

**Theorem 3.2.** (1) The conditional posterior distribution of  $\sigma^2$  is  $\text{IG}((\nu + (n - s))/2, (\nu\lambda + \mathbb{S}^2)/2)$ , where  $\mathbb{S}^2 = \sum_{t=s+1}^n (\mathbf{y}_t - \Theta'\Phi_{t-1})^2$ .

(2) The conditional distribution of  $\alpha$  given the sample and the other parameters is  $\text{Beta}(\gamma_1 + k, \gamma_2 + n - s - k)$ , where  $k$  denotes the number of  $\delta_j = 1$ ,  $j = 1, \dots, n$ .

(3) When  $\delta_j = 0$ , there is no new information about the posterior distribution of  $\beta_j$ , namely,  $\beta_j$  distributed as  $N(0, \tau^2)$ . When  $\delta_j = 1$ , since  $\mathbf{z}_t$  contains information of  $\beta_j$ , we have that

$$p(\beta_j | \mathbf{z}, \boldsymbol{\delta}, \beta_{(-j)}, \Theta, \sigma^2, \alpha) \sim N(\beta_j^*, \tau_j^*), \quad (3.3)$$

where  $\beta_{(-j)}$  is obtained  $\boldsymbol{\beta}$  by eliminating the element  $\beta_j$ , and  $\beta_j^* = \mathbb{A}\tau^2 / (\mathbb{B}\tau^2 + \sigma^2)$ ,  $\tau_j^* = \sigma^2\tau^2 / (\mathbb{B}\tau^2 + \sigma^2)$ , where  $\mathbb{A} = -\sum_{t=j}^{T_j} a_t(1)\varphi_{t-j}$  and  $\mathbb{B} = \sum_{t=j}^{T_j} \varphi_{t-j}^2$ . The proof of this Theorem is very tedious and similar to that in MuCulloch and Tsay [18]; we omit it here.

**Theorem 3.3.** For the conditional posterior distribution of  $\delta_j = 1$ , we have

$$p(\delta_j = 1 | \mathbf{z}, \boldsymbol{\delta}_{(-j)}, \boldsymbol{\beta}, \Theta, \sigma^2, \alpha) = \left[ 1 + \frac{1 - \alpha}{\alpha} B_{10}(j) \right]^{-1}, \quad (3.4)$$

where  $\boldsymbol{\delta}_{(-j)}$  is obtained  $\boldsymbol{\delta}$  by eliminating the element  $\delta_j$ ,  $T_j = \min\{n, j + s\}$ , and

$$B_{10}(j) = \exp \left\{ \frac{1}{2\sigma^2} \left[ \sum_{t=j}^{T_j} a_t^2(1) - \sum_{t=j}^{T_j} a_t^2(0) \right] \right\}; \quad (3.5)$$

here  $a_t(1) = (\mathbf{y}_t - \Theta'\Phi_{t-1})_{\delta_j=1}$ ,  $a_t(0) = (\mathbf{y}_t - \Theta'\Phi_{t-1})_{\delta_j=0}$ , and  $a_t(0) = a_t(1) + \varphi_{t-j}\beta_j$ . When  $s = 0$ , then  $\varphi_i = 0$  for all  $i$ ; when  $s > 0$ , we have that

$$\varphi_i = \begin{cases} -1, & i = 0, \\ \phi_i, & i = 1, \dots, s, \\ 0, & i \geq s + 1. \end{cases} \quad (3.6)$$

*Proof.* The conditional posterior distribution of  $\delta_j$  ( $j = s + 1, \dots, n$ ) is *Bernoulli* with probability

$$\begin{aligned}
& p(\delta_j = 1 \mid \mathbf{z}, \delta_{(-j)}, \boldsymbol{\beta}, \Theta, \sigma^2, \alpha) \\
&= \frac{\alpha L(\Theta, \sigma^2, \delta_j = 1, \boldsymbol{\beta}, \alpha \mid \mathbf{z})}{\alpha L(\Theta, \sigma^2, \delta_j = 1, \boldsymbol{\beta}, \alpha \mid \mathbf{z}) + (1 - \alpha) L(\Theta, \sigma^2, \delta_j = 0, \boldsymbol{\beta}, \alpha \mid \mathbf{z})} \\
&= \frac{\alpha \exp \left\{ -(1/2\sigma^2) \sum_{t=j}^{T_j} (y_t - \Theta' \Phi_{t-1})^2 \right\}_{\delta_j=1}}{\alpha \exp \left\{ -(1/2\sigma^2) \sum_{t=j}^{T_j} (y_t - \Theta' \Phi_{t-1})^2 \right\}_{\delta_j=1} + (1 - \alpha) \exp \left\{ -(1/2\sigma^2) \sum_{t=j}^{T_j} (y_t - \Theta' \Phi_{t-1})^2 \right\}_{\delta_j=0}} \\
&= \frac{\alpha}{\alpha + (1 - \alpha) \exp \left\{ (1/2\sigma^2) \left[ \sum_{t=j}^{T_j} [a_t(1)]^2 - \sum_{t=j}^{T_j} [a_t(0)]^2 \right] \right\}} \\
&= \left[ 1 + \frac{1 - \alpha}{\alpha} B_{10}(j) \right]^{-1}.
\end{aligned} \tag{3.7}$$

So the conclusion is obtained.  $\square$

#### 4. Detection of Outlier Patches via Adaptive Gibbs Sampling

Similar to Justel et al. [5], 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  $\hat{\Theta}^{(m)}$ ,  $\hat{\sigma}^{(m)}$ ,  $\hat{\boldsymbol{\beta}}^{(m)}$ , and  $\hat{\alpha}^{(m)}$  be the posterior means of  $\Theta$ ,  $\sigma^2$ ,  $\boldsymbol{\beta}$ , and  $\alpha$  respectively based on the  $m$  iterations of the first Gibbs run. First, we select an appropriate critical value  $c_1$  to identify potential outliers. An observation  $z_j$  is identified as an outlier if the posterior probability  $\hat{p}_j^{(m)} > c_1$ . Let  $\{t_1, \dots, t_g\}$  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  $2h$  around the identified outlier to search for the boundary points of a possible outlier patch by a forward-backward method. Examplng an identified outlier  $z_{t_i}$ , for the  $h$  observations before  $z_{t_i}$  if their posterior probabilities  $\hat{p}_j^{(m)} > c_2$ , 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 beginning point of the outlier patch. Denote the point by  $z_{t_i-k_i}$ . Then we do the same for the  $h$  observations after  $z_{t_i}$  and select the farthest point from  $z_{t_i}$  with  $\hat{p}_j^{(m)} > c_2$  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 is denoted by  $(z_{t_i-k_i}, \dots, z_{t_i+v_i})$ . Consecutive or overlapping patches should be merged to form a larger patch. Lastly, draw Gibbs samples jointly within a patch. Suppose that a patch of  $d$  outliers starting at time index  $j$  is specified. Denote the vectors of outlier indicators and magnitudes by  $\boldsymbol{\delta}_{j,d} = (\delta_j, \dots, \delta_{j+d-1})'$  and  $\boldsymbol{\beta}_{j,d} = (\beta_j, \dots, \beta_{j+d-1})'$ ,

respectively, associated with the patch. We give the conditional posterior distributions of  $\delta_{j,d}$  and  $\beta_{j,d}$  in the next proposition; its derivation is similar to Theorem 1 of Justel et al. [5].

**Proposition 4.1.** *Let  $z = (z_1, z_2, \dots, z_n)$  be a vector of observations according to (2.3), with no outliers in the first  $s$  data points. Assume  $\delta_j \sim \text{Bernoulli}(\alpha)$ ,  $j = s + 1, s + 2, \dots, n$ , and*

$$p(\Theta, \sigma^2, \alpha, \beta) \propto \left(\frac{1}{\sigma^2}\right)^{v/2+1} \alpha^{\gamma_1-1} (1-\alpha)^{\gamma_2-1} \exp\left\{-\frac{1}{2\tau^2} \sum_{t=s+1}^n (\beta_t)^2 - \frac{v\lambda}{2} \frac{1}{\sigma^2}\right\}, \quad (4.1)$$

where the parameters  $\gamma_1, \gamma_2, v, \lambda$ , and  $\tau^2$  are known. Let  $e_t(0) = y_t - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j \hat{\varepsilon}_{t-j} - \sum_{i=1}^r \sum_{j=1}^k \eta_{ij} y_{t-i} \hat{\varepsilon}_{t-j}$  be the residual at time  $t$  when the series is adjusted for all identified outliers not in the interval  $[j, j + d - 1]$  with the outliers identified in  $\delta_{j,d}$ ,  $T_{j,d} = \min\{n, j + d + s - 1\}$ , and

$$e_t(\delta_{j,d}) = \begin{cases} e_t(0) + \sum_{i=0}^{t-j} \pi_i \delta_{t-i} \beta_{t-i}, & t = j, \dots, j + d - 1, \\ e_t(0) + \sum_{i=t-j-d+1}^{t-j} \pi_i \delta_{t-i} \beta_{t-i}, & t > j + d - 1, \end{cases} \quad (4.2)$$

where  $\pi_0 = -1$ ,  $\pi_i = \phi_i$  if  $i = 1, \dots, s$ , and  $\pi_i = 0$  if  $i \geq s + 1$ . Then we have the following.

- (a) The conditional posterior probability of a block configuration  $\delta_{j,d}$ , given the sample and the other parameters, is

$$p(\delta_{j,d} | z, \theta_{\delta_{j,d}}) = C \alpha^{s_{j,d}} (1-\alpha)^{d-s_{j,d}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=j}^{T_{j,d}} e_t(\delta_{j,d})^2\right\}, \quad (4.3)$$

where  $s_{j,d} = \sum_{t=j}^{j+d-1} \delta_t$ ,  $\theta_{\delta_{j,d}}$  denotes all the other parameters except  $\delta_{j,d}$ , and  $C$  is a normalization constant so that the total probability of the  $2^d$  possible configurations of  $\delta_{j,d}$  is one.

- (b) The conditional posterior distribution of  $\beta_{j,d}$  given the sample and other parameters is  $N_d(\beta_{j,d}^*, \Omega_{j,d}^{-1})$ , where

$$\begin{aligned} \beta_{j,d}^* &= -\frac{1}{\sigma^2} \Omega_{j,d}^{-1} \sum_{t=j}^{T_{j,d}} e_t(0) D_{j,d} \Pi_{t-j}, \\ \Omega_{j,d} &= \frac{1}{\sigma^2} D_{j,d} \left( \sum_{t=j}^{T_{j,d}} \Pi_{t-j} \Pi_{t-j}' \right) D_{j,d} + \frac{1}{\tau^2} I, \end{aligned} \quad (4.4)$$

$D_{j,d} = \text{diag}\{\delta_j, \dots, \delta_{j+d-1}\}$ , and  $\Pi_t = (\pi_t, \dots, \pi_{t-d+1})'$  with  $\pi_0 = -1$ ,  $\pi_i = \phi_i$  for  $i = 1, \dots, s$ , and  $\pi_i = 0$  for  $i < 0$  or  $i > s$ .

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 results of Proposition 4.1 to draw  $\delta_{j,d}$  and  $\beta_{j,d}$  in the second Gibbs sampling, which is also run for  $m$  iterations.

The starting values of  $\delta_t$  are as follows:  $\delta_t^{(0)} = 1$  if  $\hat{p}_t^{(m)} > c_2$ , otherwise,  $\delta_t^{(0)} = 0$ . Then the prior distributions of  $\beta_t$  are as follows.

- (a) If  $z_t$  is identified as an isolated outlier, then the prior distribution of  $\beta_t$  is  $N(\hat{\beta}_t^{(m)}, \tau^2)$ , where  $\hat{\beta}_t^{(m)}$  is the Gibbs estimate of  $\beta_t$  from the first Gibbs run.
- (b) If  $z_t$  belongs to an outlier patch, then the prior distribution of  $\beta_t$  is  $N(\tilde{\beta}_t^{(m)}, \tau^2)$ , where  $\tilde{\beta}_t^{(m)}$  is the conditional posterior mean as follows:

$$\tilde{\beta}_{j,d} = \left( D_{j,d} \sum_{t=j}^{T_{j,d}} \Pi_{t-j} \Pi'_{t-j} D_{j,d} \right)^{-1} \left( - \sum_{t=j}^{T_{j,d}} e_t(0) D_{j,d} \Pi_{t-j} \right). \quad (4.5)$$

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

## 5. Simulated Example and Conclusions

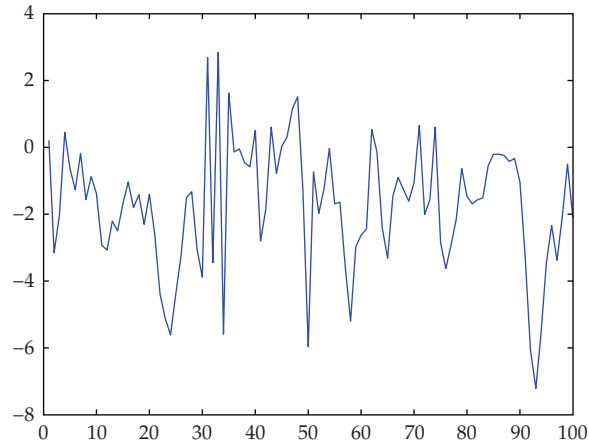
*Example 5.1.* In the simulations, the designed bilinear model is BL(2,2,1,1) model:

$$\begin{aligned} y_t &= 0.5y_{t-1} + 0.3y_{t-2} + \varepsilon_t - 0.34\varepsilon_{t-1} - 0.3\varepsilon_{t-2} + 0.4y_{t-1}\varepsilon_{t-1}, \\ z_t &= y_t + 5\delta_{t,31} - 4\delta_{t,32} + 5\delta_{t,33} - 4\delta_{t,34} + 3\delta_{t,35} - 6\delta_{t,50}, \end{aligned} \quad (5.1)$$

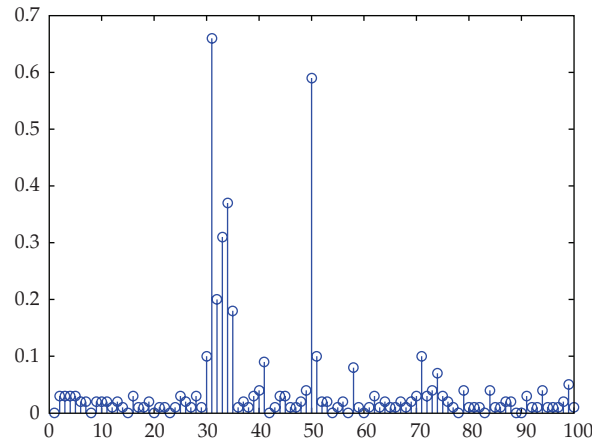
where  $\delta_{t,k}$  is Kronecker symbol: if  $t = k$ , then  $\delta_{t,k} = 1$ , else  $\delta_{t,k} = 0$ , and  $\{\varepsilon_t\}$  is standard normal white noise.

We produce a set of observations of the bilinear model (5.1) by simulation. In order to obtain stationary data, we produce 1000 observations of  $y_t$  firstly; then we use the last 100 observations as our simulated sample, in which a patch of five consecutive additive outliers has been introduced from  $t = 31$  to  $t = 35$ , a single AO has been added at  $t = 50$ , and the outlier magnitudes are  $\beta_{31} = 5$ ,  $\beta_{32} = -4$ ,  $\beta_{33} = 5$ ,  $\beta_{34} = -4$ ,  $\beta_{35} = 3$ , and  $\beta_{50} = -6$ , respectively. Figure 1 shows the trend of simulated series  $\{z_t\}$ . It is obvious that the curve of observations had large volatility; it would be very difficult to distinguish between "outliers" and normal points of nonlinear model.

Let  $\gamma_1 = 5$ ,  $\gamma_2 = 95$ ,  $\nu = 3$ ,  $\lambda = \sigma^2/2$ ,  $\alpha = 0.5$ ,  $\tau = 2\sqrt{5}\sigma$ , and  $c_1 = 0.5$ . 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. We can obtain the posterior probabilities that each observation is an outlier by using the given methods in Section 3, which are shown in Figure 2. It displayed that the posterior probabilities of being outliers only at  $t = 31$  and  $t = 50$  were larger than 0.5. Meanwhile, the outlying posterior probabilities of other observations were lower; the posterior probability of being an outlier at



**Figure 1:** The curve of the simulated observations  $\{z_t\}$

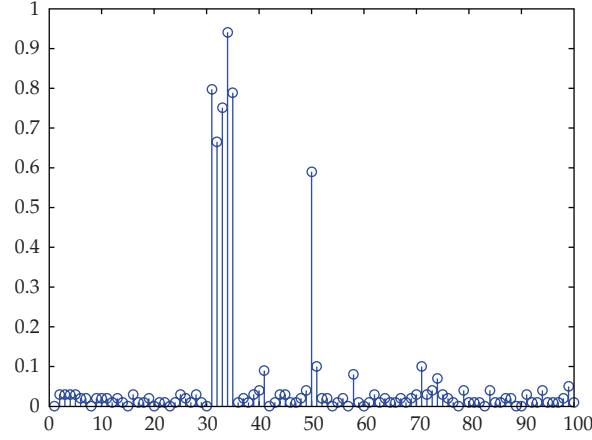


**Figure 2:** The posterior probabilities that each point is an outlier via standard Gibbs sampling.

$t = 35$  was even smaller than 0.2. We see that the standard Gibbs sampling failed to detect the inner and border points at  $t = 32, 33, 35$ , that resulted in the masking effect.

Second, in order to avoid the presence of masking and swamping problem, we use the method given in Section 4, and take 900 iterations by the adaptive Gibbs algorithm. Figure 3 gives the plot of the estimated posterior probability that each point is seen as an outlier via adaptive Gibbs sampling in simulation, where the window width of search was 6. It is obvious that the posterior probabilities of being an outlier obtained by adaptive Gibbs algorithm for data points at  $t = 31, \dots, 35$  and  $t = 50$  are larger than 0.5, which meant that these points were possible outliers. Meanwhile, the outlying posterior probabilities of other observations are very small. Actually, if we select the critical values  $c_1 = 0.5$  and  $c_2 = 0.15$ , then we could identify the patch. On the other hand, many normal points like outliers in Figure 1 were not misspecified as outliers because the outlying posterior probabilities of these points were smaller than 0.2, which shows that the adaptive Gibbs sampling was more effective than the standard Gibbs sampling in mining the additive outlier patch for bilinear time series.





**Figure 3:** The posterior probability that each point is an outlier via adaptive Gibbs algorithm.

**Table 1:** The simulated results of the model BL(2, 2, 1, 1) via adaptive Gibbs algorithm.

	$\beta_{31}$	$\beta_{32}$	$\beta_{33}$	$\beta_{34}$	$\beta_{35}$	$\beta_{50}$
True value	5	-4	5	-4	3	-6
Posterior mean	4.7139	-3.6529	4.1193	-6.9487	4.7747	-2.6025
Standard deviation	0.6211	0.6555	0.6044	0.6374	0.6711	0.5813

Third, the posterior means of the sizes of these outliers are  $\hat{\beta}_{31} = 4.6714$ ,  $\hat{\beta}_{32} = -3.4534$ ,  $\hat{\beta}_{33} = 4.3176$ ,  $\hat{\beta}_{34} = -6.9497$ ,  $\hat{\beta}_{35} = 5.1142$ , and  $\hat{\beta}_{50} = -2.2523$ , respectively.

Furthermore, we did 200 replications for the above simulation. Table 1 gives the average values of the posterior means of  $\beta_{31}, \dots, \beta_{35}$  and  $\beta_{50}$  and their standard deviations. In these replications, the isolated outliers were identified by standard and adaptive Gibbs sampling procedures. Through computing, we find that the standard Gibbs sampling failed to detect all the outlier patches in 200 replications, but in 179 replications, the adaptive Gibbs sampling successfully could specify all outliers and patches. On the other hand, we note that the last three parameters are actually not estimated well by the adaptive Gibbs algorithm in many replications. Investigating its reason, we think it may be the influence of nonlinearity or the frontal outliers of patch. Such situation deserves careful investigation in the future.

Since the bilinear series will by itself produce what appears to be outliers, the detection of the locations of outliers and patches is more difficult, and the estimation of the magnitude of outliers and patches is more complex than autoregressive series. By a number of simulations of detecting the outlier patches in bilinear model by adaptive Gibbs sampling, we discovered that the critical value  $c_2$  should be selected smaller than that in ARMA model. It may be that the variation of bilinear series is larger than that of ARMA series. On the other hand, in the process of running, the Gibbs sampler was repeated several times with different hyper-parameters 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. Some other case studies also show that our procedure could reduce possible masking and swamping effects, which is an improvement and extension on bilinear time series model over the existing outliers detection methods.

## Acknowledgments

The authors thank the editor and three referees for their careful and constructive comments that led to substantial improvements of the paper. The project is supported by NSFCJS BK2008284.

## References

- [1] M. Li and W. Zhao, "Representation of a stochastic traffic bound," *IEEE Transactions on Parallel and Distributed Systems*, preprint.
- [2] B. Abraham and G. E. P. Box, "Bayesian analysis of some outlier problems in time series," *Biometrika*, vol. 66, no. 2, pp. 229–236, 1979.
- [3] R. S. Tsay, "Time series model specification in the presence of outliers," *Journal of the American Statistical Association*, vol. 81, pp. 132–141, 1986.
- [4] G. M. Ljung, "On outlier detection in time series," *Journal of the Royal Statistical Society. Series B*, vol. 55, no. 2, pp. 559–567, 1993.
- [5] A. Justel, D. Peña, and R. S. Tsay, "Detection of outlier patches in autoregressive time series," *Statistica Sinica*, vol. 11, no. 3, pp. 651–673, 2001.
- [6] F. Battaglia and L. Orfei, "Outlier detection and estimation in nonlinear time series," *Journal of Time Series Analysis*, vol. 26, no. 1, pp. 107–121, 2005.
- [7] P. Chen, F. R. Yan, Y. Y. Wu, and Y. Chen, "Detection of outliers in ARMAX time series models," *Advances in Systems Science and Applications*, vol. 9, pp. 97–103, 2009.
- [8] P. Chen and Y. Chen, "The identification of outliers in ARMAX models via genetic algorithm," to appear in *Kybernetes*.
- [9] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [10] H.-B. Wang and B.-C. Wei, "Separable lower triangular bilinear model," *Journal of Applied Probability*, vol. 41, no. 1, pp. 221–235, 2004.
- [11] H.-B. Wang, "Parameter estimation and subset selection for separable lower triangular bilinear models," *Journal of Time Series Analysis*, vol. 26, no. 5, pp. 743–757, 2005.
- [12] C. W. S. Chen, "Detection of additive outliers in bilinear time series," *Computational Statistics & Data Analysis*, vol. 24, no. 3, pp. 283–294, 1997.
- [13] T. Subba Rao, "On the theory of bilinear time series models," *Journal of the Royal Statistical Society B*, vol. 43, no. 2, pp. 244–255, 1981.
- [14] T. Subba Rao and M. M. Gabr, *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, vol. 24 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1984.
- [15] W. K. Kim, L. Billard, and I. V. Basawa, "Estimation for the first-order diagonal bilinear time series model," *Journal of Time Series Analysis*, vol. 11, no. 3, pp. 215–229, 1990.
- [16] S. A. O. Sesay and T. Subba Rao, "Difference equations for higher-order moments and cumulants for the bilinear time series model  $BL(p, 0, p, 1)$ ," *Journal of Time Series Analysis*, vol. 12, no. 2, pp. 159–177, 1991.
- [17] C. W. J. Granger and A. P. Andersen, *An Introduction to Bilinear Time Series Models*, Vandenhoeck & Ruprecht, Göttingen, Germany, 1978.
- [18] R. E. McCulloch and R. S. Tsay, "Bayesian analysis of autoregressive time series via the Gibbs sampler," *Journal of Time Series Analysis*, vol. 15, no. 2, pp. 235–250, 1994.