ELSEVIER
Statistics & Probability Letters 20 (1994) 143–148
STATISTICS & PROBABILITY LETTERS

Outlier detection in the state space model

Siddhartha Chib[a], Ram C. Tiwari[b,*]

[a] *John M. Olin School of Business, Washington University, St. Louis, MO 63130, USA*
[b] *Department of Mathematics, University of North Carolina, Charlotte, NC 28223, USA*

Received September 1992; revised May 1993

Abstract

Zellner (1975), Chaloner and Brant (1988), and Chaloner (1991) used the posterior distributions of the realized errors to define outliers in a linear model. The same concept is used here to define outliers in a state-space model. An effective approach to compute the posterior probabilities of observations being outliers is developed and illustrated by means of examples. The detection of two outliers is straightforward.

Key words: Kalman filter recursions; Smoothing recursions; Realized errors

1. Introduction

In a linear model with normally distributed random errors, ε_i for $i=1,\ldots,n$, with means zero and a common variance σ^2, Chaloner and Brant (1988) advocated the ith observation to be an outlier if $|\varepsilon_i|>k\sigma$ for some choice of k. They suggested to choose k so that prior probability of no outliers is large, say 0.95. This gives $k=\Phi^{-1}\{0.5+\frac{1}{2}(0.95)^{1/n}\}$. Zellner (1975), Chaloner and Brant (1988) and Chaloner (1991) thoroughly reviewed the problem of outlier detection. We use their idea to define outliers in a linear state-space model (cf. Chib and Tiwari, 1991; Harvey, 1981; Meinhold and Singpurwalla, 1983, 1989; West and Harrison, 1989) wherein the time series $\{y_t;\ t=1,2,\ \ldots\}$, given θ_t, is modelled as

$$y_t=x_t'\theta_t+\varepsilon_t, \quad \varepsilon_t\sim \mathrm{N}(0,\sigma^2), \tag{1}$$

where θ_t is the state $p\times 1$ vector at time t, x_t is a known regression $p\times 1$ vector, and ε_t is an observational error. The evolution over time of the state vector θ_t, given θ_{t-1}, is described by

$$\theta_t=G_t\theta_{t-1}+w_t, \quad w_t\sim \mathrm{N}(0,\sigma^2 W), \tag{2}$$

where G_t is a known $p\times p$ state evolution matrix and w_t the evolution error. The error sequence $\{\varepsilon_t\}$ and $\{w_t\}$ are assumed to be independent and mutually independent. The $p\times p$ matrix W is assumed known, and the

* Corresponding author.

0167-7152/94/$7.00 © 1994 – Elsevier Science B.V. All rights reserved
SSDI: 0167-7152(93)E0158-P

prior distribution of σ^2 is an inverted-gamma, $\mathrm{IG}(v_0/2,\ \delta_0/2)$, with parameters $v_0/2$ and $\delta_0/2$; i.e., $p(\sigma^2)\propto(\sigma^2)^{(v_0+1)/2}e^{-\delta_0/2\sigma^2}$.

Let D_n denote the $n\times 1$ vector $(y_1,\ldots,y_n)'$ of observations available up to and including time n. We say y_t, for $t=1,\ldots,n$, is an outlier if the event $|\varepsilon_t|>k\sigma$ occurs for some k. The choice of k is as suggested by Chaloner and Brant (1988) in the case of a linear model. To compute the posterior probability that $|\varepsilon_t|>k\sigma$ we need the posterior distribution of ε_t, given D_n, which is given by

$$\varepsilon_t|D_n,\sigma^2\sim \mathrm{N}(\hat{\varepsilon}_{t|n},\sigma^2R_{t|n}),\qquad \sigma^2|D_n\sim \mathrm{IG}(v_n/2,\ \delta_n/2),\tag{3}$$

where

$$\begin{aligned}&\hat{\varepsilon}_{t|n}=y_t-x_t'\hat{\theta}_{t|n},\qquad h_{t|n}=x_t'R_{t|n}x_t,\\ &v_n=v_0+n,\qquad \delta_n=\delta_0+\sum_{s=1}^{n}\hat{e}^2_{s|s-1}f^{-1}_{s|s-1},\end{aligned}\tag{4}$$

with $\hat{e}_{s|s-1}=y_s-x_s'\theta_{s|s-1}$ and $f_{s|s-1}=x_s'R_{s|s-1}x_s+1$. The smoothing recursions to compute $\hat{\theta}_{t|n}$ and $R_{t|n}$ are given by

$$\begin{aligned}&\hat{\theta}_{t|n}=\hat{\theta}_{t|t}+A_t(\theta_{t+1|n}-G_{t+1}\hat{\theta}_{t|t})\\ &R_{t|n}=R_{t|t}-A_t(R_{t+1|n}-R_{t+1|t})A_t',\end{aligned}\tag{5}$$

where $\hat{\theta}_{t|t}$ and $R_{t|t}$ are obtained through the Kalman filter recursions (see, e.g., Harvey, 1989, Section 4) and $A_t=R_{t|t}G_{t+1}'R_{t+1|t}^{-1}$.

At time $t=0$, we assume that $\theta_0|\sigma^2\sim\mathrm{N}(\hat{\theta}_{0|0},\ \sigma^2R_{0|0})$ and $\sigma^2\sim\mathrm{IG}(v_0/2,\ \delta_0/2)$ with hyperparameters $\hat{\theta}_{0|0}$, $R_{0|0}$, v_0 and δ_0 known.

2. Realized error analysis

Our approach to outlier detection builds on the framework of Bayesian error analysis, called *realized error analysis*, that is developed in (Zellner, 1975; Zellner and Moulton, 1985; Chaloner and Brant, 1988). These authors are only concerned with the linear model which is a special case of the state-space model with constant state variables. To detect which observations are outliers, define the probability $p_{t|n}$ to be $\mathrm{pr}(|\varepsilon_t|>k\sigma|D_n)$, the posterior probability that the tth observation is what we have defined to be an outlier. Let $\Phi(z)$ denote the standard normal cumulative distribution function. Further, let

$$u_{t|n}=(k\sigma-\hat{\varepsilon}_{t|n})/\sigma\sqrt{h_{t|n}},\qquad v_{t|n}=(-k\sigma-\hat{\varepsilon}_{t|n})/\sigma\sqrt{h_{t|n}},\tag{6}$$

then we have

$$p_{t|n}=\mathrm{pr}(|\varepsilon_t|>k\sigma|D_n)=\int_0^\infty\{1-\Phi(u_{t|n})+\Phi(v_{t|n})\}\,p(\sigma^2|D_n)\,\mathrm{d}\sigma^2.\tag{7}$$

The $p_{t|n}$'s can be compared with the prior probability $2\Phi(-k)$. It is easy to see that $p_{t|n}$ is an increasing function of $h_{t|n}$, often referred to as leverage.

As in (Chaloner and Brant, 1988), the posterior probability that ε_s and ε_t are both outliers can be computed. For this, we use the posterior distribution of $(\varepsilon_s,\varepsilon_t)$, $1\leqslant s\leqslant t\leqslant n$, given by

$$(\varepsilon_s,\varepsilon_t)'|D_n,\sigma^2\sim\mathrm{N}((\hat{\varepsilon}_{s|n},\hat{\varepsilon}_{t|n})',\ \sigma^2H_{s,t|n}),\tag{8}$$

Table 1
Pena and Guttman simulated data

t	y_t	t	y_t	t	y_t	t	y_t
1	12.18	9	8.86	17	4.88	25	29.00
2	9.32	10	1.00	18	3.34	26	0.35
3	11.20	11	7.79	19	2.08	27	3.42
4	9.59	12	7.79	20	3.53	28	1.64
5	7.41	13	7.621	21	1.25	29	2.17
6	7.69	14	7.19	22	2.70	30	2.64
7	9.06	15	4.71	23	0.48	31	3.87
8	8.17	16	6.28	24	0.19		

where

$$H_{s,t|n}=\begin{bmatrix} h_{s|n} & x_s'C_{s,t|n}x_t \\ - & h_{t|n} \end{bmatrix},$$

with $C_{s,t|n}=A_sC_{s+1,t|n}$, where $C_{t,t|n}=R_{t|n}$ is used to initialize the recursion, and A_s is defined in (5). This elegant expression for $C_{s,t|n}$ is due to Jong and Mackinnon (1988). Let $r_{s,t|n}=x_s'C_{s,t|n}x_t/\sqrt{h_{s|n}h_{t|n}}$ be the correlation between ε_s and ε_t given D_n. Further let $B(a, b, \rho)$ be the standard bivariate normal cumulative distribution function with correlation ρ. Then the posteriori probability that ε_s and ε_t are both outliers is

$$\begin{aligned} p_{s,t|n}&=\text{pr}(|\varepsilon_s|>k\sigma,\ |\varepsilon_t|>k\sigma\,|\,D_n) \\ &=\int_0^\infty \{\bar{B}(u_{s|n},\ u_{t|n},\ v_{s|n},\ v_{t|n},\ r_{s,t|n})\}\,p(\sigma^2|D_n)\,\mathrm{d}\sigma^2, \end{aligned} \tag{9}$$

where $\bar{B}(a, b, c, d, \rho)\equiv B(-a, -b, \rho)+B(-a, d, -\rho)+B(c, -b, -\rho)+B(c, d, \rho)$. The $p_{s,t|n}$'s can be compared to the prior probability $\{2\Phi(-k)\}^2$.

In many applications, for example, in the structural time series models considered by Harvey (1989), the elements of matrix W depend on a set of unknown parameters λ. Assuming $p(\lambda)$ to be the prior distribution of λ, the posterior probabilities outliers can be computed by integrating $p_{t|n}$ and $p_{s,t|n}$ with respect to $p(\lambda|D_n)$, see (Chib, et al., 1990).

The approach developed does not rely on an outlier generating mechanism and is extremely easy to implement. Also, our approach can be adapted to deal with problems where the underlying time series experiences an abrupt break, as, for example, in the multi-process state-space model of (cf. West and Harrison 1989).

3. Examples

The first example is the Pena and Guttman simulated data (Pena and Guttman, 1988, pp. 244, 245) obtained from the state-space model (see Table 1).

$$y_t=\theta_t+\varepsilon_t,\quad \varepsilon_t\sim \mathrm{N}(0, 1),$$

$$\theta_t=\theta_{t-1}+w_t,\quad w_t\sim \mathrm{N}(0, 1),\qquad t=1,\ldots,31.$$

At time $t=1$, $\theta_1=10$. The observations y_{10} and y_{25} are outliers.

Table 2
Table of smoothed residuals $\hat{\varepsilon}_{t|n}$, leverages $h_{t|n}$ and posterior probabilities $\mathrm{pr}(|\varepsilon_t|>3.11\sigma|D_n)=p_{t|n}$ (3.11) for Pena and Guttman simulated data

t	$\hat{\varepsilon}_{t\|n}$	$h_{t\|n}$	$p_{t\|n}(3.11)$
1	0.9492	0.6180	0.0002
2	−0.9613	0.4721	
3	0.9069	0.4508	
4	0.1921	0.4477	
5	−0.9007	0.4473	
6	−0.4343	0.4472	
7	0.6879	0.4472	
8	0.2381	0.4472	
9	1.6063	0.4472	
10	−3.9691	0.4472	0.0063
11	1.1365	0.4472	
12	0.5885	0.4472	
13	0.4589	0.4472	
14	0.5283	0.4472	
15	−0.9241	0.4472	
16	0.7494	0.4472	
17	0.2024	0.4472	
18	−0.2824	0.4472	
19	−0.7694	0.4472	
20	0.6840	0.4472	
21	−0.9084	0.4472	
22	0.3207	0.4472	
23	−1.7996	0.4472	0.0001
24	−3.7895	0.4472	0.0047
25	10.5312	0.4472	0.7504
26	−4.0769	0.4472	0.0074
27	−0.0420	0.4473	
28	−0.8990	0.4477	
29	−0.3450	0.4508	
30	−0.1960	0.4721	
31	0.5170	0.6180	0.0001

Note: Posterior probabilities less than 10^{-4} have been omitted.

At $t=0$, we assume that $\hat{\theta}_{0|0}=10$, $R_{0|0}=10^3$, $v_0=0.1$ and $\delta_0=2$. Setting the prior probability of no outliers to be 0.95 gives $k=3.11$ to define an outlier. Table 2 gives smoothed residual, $\hat{\varepsilon}_{t|n}$, leverage, $h_{t|n}$, and $p_{t|n}$, for $t=1,\ldots,31$. Comparing probabilities $p_{t|n}$ with the prior probability $2\Phi(k)=0.0019$ shows that observations 10 and 25 are outliers. In addition, observations 23 and 26 also appear as outliers.

For selected values of s and t (and for $k=2$), Table 3 gives the posterior probabilities $p_{s,t|n}$ and the correlation coefficient $r_{s,t|n}$. Once again the observations 10 and 25 are outlier.

The second example uses the sales data consisting of 72 observations given in (West and Harrison, 1989, p. 334). The variables of interest is an index of sales on a standardized deflated scale of a well-established food product in the UK market. Here $x_t=(1, x_{2t}, e_2'; e_2'; e_2')$ and $G_t=G=\text{block diag}[1; 1; F_1; F_3; F_4]$, where x_{2t} is a covariate measuring price and costs, and

$$e_2=(1,0)'; \quad \mathrm{F}_r=\begin{bmatrix} \cos(\pi r/6) & \sin(\pi r/6) \\ -\sin(\pi r/6) & \cos(\pi r/6) \end{bmatrix}, \quad r=1,3,4.$$

Table 3
Table of correlation coefficients $r_{s,t|n}$ and posterior probabilities $\mathrm{pr}(|\varepsilon_s|>2\sigma, |\varepsilon_t|>2\sigma)=p_{s,t|n}(2)$ for selected values of s and $t(s<t)$ for Pena and Guttman data

s,t	$r_{s,t\|n}$	$p_{s,t\|n}(2)$
10,11	0.5279	0.0002
10,24		0.0339
10,25		0.1816
24,25	0.5279	0.1527
25,26	0.5279	0.1905

Note: Correlation coefficients less than 10^{-4} have been omitted.

Table 4
Table of correlation coefficients $r_{s,t|n}$ and posterior probabilities $\mathrm{pr}(|\varepsilon_s|>2\sigma, |\varepsilon_t|>2\sigma|D_n)=p_{s,t|n}(2)$ $(s<t)$ for West and Harrison sales data

s, t	$r_{s,t\|n}$	$P_{s,t\|n}(2)$
2,31	−0.2501	
2,32	−0.1329	
7,32	0.1180	0.1144
31,32	0.7197	0.3169
31,33	0.2749	0.0002
31,34	0.7422	0.3941
47,49	0.4698	0.0193
49,55	0.0994	0.2026
55,56	0.6334	0.0002
70,71	0.5217	0.0003

Note: Posterior probabilities loess than 10^{-4} have been omitted.

The matrix W is specified as

$$W=\text{block diag}[0.009;\ 0.0002;\ 0.0003 I_6],$$

where I_6 is the 6×6 identity matrix. The initial prior information is specified directly for θ_1 and is given by

$$\hat{\theta}_{1|0}=[9.5;\ -0.7;\ 0.691;\ 1.159;\ 0.283;\ -0.050;\ -0.217;\ 0.144]',$$

and

$$R_{1|0}=20\times\text{block diag}[0.09;\ 0.01;\ 0.0067 I_6],$$

with $v_0=6$ and $\delta_0=0.0990$.

To quote West and Harrison (1989, p. 354), "a typical outlier shows up in the January, 1980 error ... the error in this month (is) way outside the 90% interval, and is in fact outside the 99% limit ... Another apparent outlier occurs is February, 1976, but this one is explained as due to the inappropriateness of the initial prior estimate of the seasonal factor for that month."

The posterior probabilities $p_{t|n}$ with $k=2$ given in Fig. 1 confirm that January 1980 is an outlier. Interestingly, February 1976 does not show up as an outlier, and other observations do. To determine if February 1976 is masked as an outlier, the probability of two outliers is computed and reported in Table 4. It is clear that masking is not a problem as the probability of February 1976 and June 1978 (and February 1976 and July 1979) being outliers is negligible.

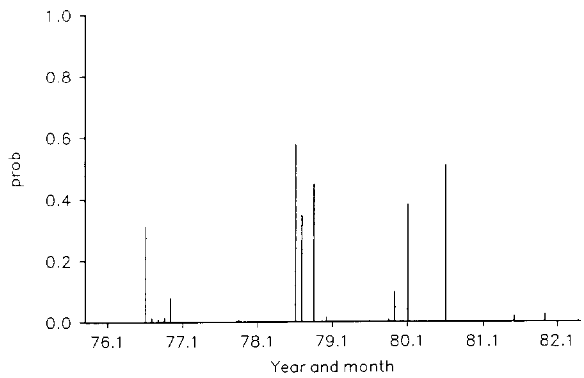

Figure 1. Plot of Posterior probabilities $pr(|\varepsilon_t| > 2\sigma | D_n)$: West & Harrison sales data.

References

Chaloner, K. (1991), Bayesian residual analysis in the presence of censoring, *Biometrika* **78**, 637–644.

Chaloner, K. and R. Brant (1988), A Bayesian approach to outlier detection and residual analysis, *Biometrika* **75**, 651–659.

Chib, S., S.R. Jammalamadaka and R.C. Tiwari (1990), On Bayes smoothing in a time varying regression model, in: S. Geisser, J.S. Hodges, S.J. Press and A. Zellner, eds., *Bayesian and Likelihood Methods in Statistics and Econometrics* (North-Holland, Amsterdam).

Chib, S. and R.C. Tiwari (1991), Extreme bounds analysis in the Kalman filter, *Amer. Statist.* **45**, 113–114.

Harvey, A.C. (1981), *Time Series Models* (Wiley, New York).

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge Univ. Press, New York).

Jong, P.D. and M.J. MacKinnon (1988), Covariances for smoothed estimates in state space models, *Biometrika* **75**, 601–602.

Meinhold, R.J. and N.D. Singpurwalla (1983), Understanding the Kalman filter, *Amer. Statist.* **37**, 123–127.

Meinhold, R.J. and N.D. Singpurwalla, (1989), Robustification of Kalman filter models, *J. Amer. Statist. Assoc.* **87**, 479–486.

Pena, D. and I. Guttman (1988), Bayesian approach to robustifying the Kalman filter, In: J. Spall, ed. *Bayesian Analysis of Time Series and Dynamic Models* (Marcel Dekker, New York) pp. 227–253.

West, M. and P.J. Harrison (1989), *Bayesian Forecasting and Dynamic Models* (Springer, New York).

Zellner, A. (1975), Bayesian analysis of regression error terms, *J. Amer. Statist. Assoc.* **70**, 138–144.

Zellner, A. and B.R. Moulton (1985), Bayesian regression diagnostics with applications to international consumption and income data, *J. Econometrics* **29**, 187–211.