Statistics & Probability Letters 16 (1993) 259–268
North-Holland
16 March 1993

Handling spuriosity in the Kalman filter

Dennis K.J. Lin
Department of Statistics, University of Tennessee, Knoxville, TN, USA

Irwin Guttman
Department of Statistics, University of Toronto, Ont., Canada

Received February 1992

Abstract: The Kalman filter, which is in popular use in various branches of engineering, is essentially a least squares procedure. One well-recognized concern in this least squares procedure is its non-robustness to spuriously generated observations that give rise to outlying observations, rendering the Kalman filter unstable, with devastating consequences in some situations. Much evidence exists that data almost always contain a small proportion of spuriously generated observations, and indeed, one wild observation can make the Kalman filter unstable. To handle this, we introduce a new recursive estimation scheme which is found to be robust to spurious observations. Examples are given to illustrate the new scheme.

Keywords: Kullback–Leibler distances; mixture distribution; robust filter; spurious observations.

1. Introduction

A basic problem in stochastic control is that of separating the *signal* from the *noise*. There has been much work on this so-called 'filtering' problem (a *filter* is used in daily life to separate the *good* from the *bad*). One major development, called the *Kalman filter*, has been used in many areas, from on-line process control in industry to applications in economics (see e.g., Phadke, 1981).

The Kalman filter is a recursive procedure to estimate the state parameters of the system at the current time, to predict the next observation, and to update the value of the parameter state vector when the next measurement is observed. Kalman's (1960) results, popular with control engineers and other physical scientists, are reproducible using a Bayes approach with normal theory, conditional on known values of the variances and covariances involved. The structure of the standard Kalman filter is displayed in Table 1. The reader is referred to Aoki (1967) for its derivation.

Referring to Table 1, our goal is to make inference about θ_t, called the *state of nature*. The observed values of the variance of interest y_t depend on the unobservable θ_t at time t. The

Table 1
The standard Kalman filter

Model:
$$y_t = A_t\theta_t + \varepsilon_t,\quad \varepsilon_t \sim \mathrm{N}(0, C_t)$$
$$\theta_t = \Omega_t\theta_{t-1} + u_t,\quad u_t \sim \mathrm{N}(0, R_t)$$

Initial setting:
$$\theta_0 \sim \mathrm{N}(\mu_0, V_0)$$

Prediction:
$$\mu_{t|t-1} = E(\theta_t \mid y_{t-1}) = \Omega_t\mu_{t-1}$$
$$V_{t|t-1} = \mathrm{Var}(\theta_t \mid y_{t-1}) = R_t + \Omega_t V_{t-1}\Omega_t'$$
$$\hat{y}_t = A_t\mu_{t|t-1}$$
$$M_t = C_t + A_t V_{t|t-1} A_t'$$

Updating of the parameters:
$$\mu_t = \mu_{t|t-1} + V_{t|t-1}A_t'M_t^{-1}(y_t - A_t\mu_{t|t-1})$$
$$V_t = V_{t|t-1} - V_{t|t-1}A_t'M_t^{-1}A_tV_{t|t-1}$$
$$\text{Filter} = V_{t|t-1}A_t'M_t^{-1}$$

Correspondence to: Dennis K.J. Lin, Department of Statistics, University of Tennessee, Knoxville, TN 37996-0532.

0167-7152/93/$06.00 © 1993 – Elsevier Science Publishers B.V. All rights reserved

relationship between y_t and θ_t is known as the *observation equation*, whereas, the dynamic feature between θ_t and θ_{t-1} is known as the *system equation*. The matrices A_t in the observation equation, Ω_t in the system equation, as well as the covariance matrices C_t and R_t are assumed to be known. Often, the variation for the observation equation is larger than that of system equation, i.e., $C_t > R_t$ in some sense.

Under such a recursive scheme, after choices for the initial values μ_0 and V_0, an update estimation of θ_t is carried out when a new observation y_t is available, for $t = 1, 2, \ldots$. At each stage, time t say, the update estimation makes use of knowledge from both the previous estimation at time $t-1$ and the new observation y_t, as described in Table 1.

Essentially, the Kalman filter process is a least squares procedure. One well-recognized concern in the least square procedure is its non-robustness to extreme observations, with the result that the Kalman filter becomes unstable, with devastating consequences in some situations (see e.g., Kitagawa, 1987). Much evidence exists that data almost always contain a small proportion of spuriously generated observations, and indeed, one wild observation can make the Kalman filter unstable.

Several authors have suggested procedures to deal with this problem (see e.g., Harrison and Stevens, 1976; Box and Tiao, 1968; Abraham and Box, 1979; Peña and Guttman, 1989; Meinhold and Singpurwalla, 1989). In this paper, we are particularly interested in using a mixture of normals as a model for the distributions of the noise in the observation and/or the state space equations. The use of a mixture of two normals in the observation equation leads to sensible results that are easy to implement in the resulting recursive scheme, which enjoys a certain optimal property (see Peña and Guttman, 1989). We discuss the robust filter given by Peña and Guttman in Section 2, and introduce our new easily applied filter in Section 3. Two examples are given for comparison in section 4 to underline the value of the new filter.

2. Robust Kalman filter

Spuriously generated observations that give rise to outliers often mean that the error distributions

Table 2
The Guttman and Peña (1985) robust filter

Model:

$$y_t = A_t\theta_t + \varepsilon_t, \quad \varepsilon_t \sim \alpha_1 \mathrm{N}(0, C_{t,1}) + \alpha_2 \mathrm{N}(0, C_{t,2})$$
$$\theta_t = \Omega_t\theta_{t-1} + u_t, \quad u_t \sim \mathrm{N}(0, R_t)$$

Initial setting:

$\theta_0 \sim \mathrm{N}(\mu_0, V_0)$ and specified α_1 $(\alpha_2 = 1 - \alpha_1)$

Prediction:

$$\mu_{t|t-1} = \Omega_t\mu_{t-1}, \quad V_{t|t-1} = R_t + \Omega_t V_{t-1}\Omega_t', \quad \hat{y}_t = A_t\mu_{t|t-1}$$
$$M_{t,1} = C_{t,1} + A_t V_{t|t-1} A_t' \text{ and } M_{t,2} = C_{t,2} + A_t V_{t|t-1} A_t'$$

Compute posterior probabilities:

$$\alpha_{t,1} = [1 + \frac{\alpha_2}{\alpha_1}(\frac{\| M_{t,1} \|}{\| M_{t,2} \|})^{1/2} \exp\{\tfrac{1}{2}(y_t - A_t\mu_{t|t-1})'(M_{t,1}^{-1} - M_{t,2}^{-1})(y_t - A_t\mu_{t|t-1})\}]^{-1}$$
$$\alpha_{t,2} = 1 - \alpha_{t,1}$$
$$\alpha_{t,i} = \text{posterior probability of } y_t \sim \mathrm{N}(A_t\theta_t, C_{t,i})$$

Updating of the parameters:

$$\mu_t = \mu_{t|t-1} + V_{t|t-1}A_t'(\alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1})(y_t - A_t\mu_{t|t-1})$$
$$V_t = V_{t|t-1} - V_{t|t-1}A_t' B_t A_t V_{t|t-1}$$
$$B_t = \alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1} - H_t$$
$$H_t = \alpha_{t,1}\alpha_{t,2}(M_{t,1}^{-1} - M_{t,2}^{-1})(y_t - A_t\mu_{t|t-1})(y_t - A_t\mu_{t|t-1})'(M_{t,1}^{-1} - M_{t,2}^{-1})$$
$$\text{Filter} = V_{t|t-1}A_t'(\alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1}) = \Sigma_{i=1}^{2}\alpha_{t,i}V_{t|t-1}A_t'M_{t,i}^{-1}$$

involved have tails heavier than those of the normal distribution. Guttman and Peña (1985) thus replaced the assumption of normality by the so-called *scaled-contaminated model* (Jeffreys, 1961). Instead of a single normal distribution in the standard Kalman filter, a mixture of normals is used as a model for the observation errors ε_t's. Their approach is illustrated in Table 2.

Initially, a preliminary 'estimate' of the prior behaviour of θ is made, say $\theta_0 \sim N(\mu_0, V_0)$; as well as choices for α_i $(i=1, 2)$, $\alpha_1+\alpha_2=1$. At each stage, after observing y_t, they compute the updated estimates of the α_i's say $\alpha_{t,i}$'s, and then update the estimation of μ, using the relevant formula found in Table 2.

Based on their model, the resulting posteriors involve a mixture of normal distributions. In order to keep the recursive procedure simple, they collapse, at each stage, a certain mixture of two normals into one normal, by fitting first and second moments. This collapsing method (see Table 2) is proved to be optimum in the sense of minimizing a Kullback–Leibler distance (Peña and Guttman, 1989, Theorem 4.1).

Although precise values of $\alpha_1=\lambda$ are chosen that are specific to the user's experience, we quote Box and Tiao (1968, p. 724): "... in practical applications, α would usually be small and $\alpha=0.1$ is already too extreme a value for realistic consideration. Some insight into the reasonable range for α can be obtained by considering a 'typical' experiment involving, say, twenty runs. A fairly optimistic data analyst might perhaps expect some discrepant observations 50% of the time. A rather pessimistic analyst might expect discrepant ones 75% of the time. Using the Poisson approximation, these probabilities would correspond to values of α equal to 0.035 and 0.07 respectively. On the other hand, $\alpha=0.1$, implies that at least one of the observations will be bad (i.e., spurious) 86% of the time, a rather unacceptable situation, less often met." Because of the above, we will assume that $0 \leqslant \alpha \leqslant 0.05$.

3. A different collapsing procedure

The use of a scaled-contaminated model for the observation errors, ε_t's, leads to extremely sensible results when the experience of spurious observation is a concern. The novel idea of Guttman and Peña (1985, 1989), however, can be improved by a different collapsing scheme for the two mixture of normals.

The approach to be discussed here is different from Guttman and Peña (1985, 1989). Instead of collapsing the mixture posterior at the end of each stage, here we collapse the initial mixture likelihood function. Thus, after computing the posterior probabilities that y_t comes from $N(A_t\theta_t, C_{t,1})$ or $N(A_t\theta_t, C_{t,2})$, say $\alpha_{t,1}$, or $\alpha_{t,2}=1-\alpha_{t,1}$, we approximate $\alpha_{t,1}N(0, C_{t,1}) + \alpha_{t,2}N(0, C_{t,2})$ by: $N(0, \alpha_{t,1}C_{t,1}+\alpha_{t,2}C_{t,2})$, and are then able to apply the standard Kalman filter. The new filter thus obtained combines the advantages of the filters due to Kalman (1960) and Guttman and Peña (1985). It is as efficient as the 'robust Kalman filter' developed by Guttman and Peña (1985) when spurious observations occur, and most of their good features have been retained. It is also, computationally, as simple as the standard Kalman filter (the derivation is similar to Guttman and Peña (1985), but with a different collapsing method).

We now proceed as follows, starting from time $t=1$, we first compute the posterior probabilities, $\alpha_{1,1}$ and $\alpha_{1,2}$, using the formula in Table 2 (see Guttman and Peña, 1985, for its derivation). Here, $\alpha_{1,1}$ is the posterior probability that y_1 has been generated from $N(A_1\theta_1, C_{1,1})$ and $\alpha_{1,2}$ is the posterior probability that y_1 has been generated from $N(A_1\theta_1, C_{1,2})$. Then, we collapse the estimated mixture noise distribution $\alpha_{1,1}N(0, C_{1,1}) + \alpha_{1,2}N(0, C_{1,2})$ using moments, by

$$N(0, \alpha_{1,1}C_{1,1}+\alpha_{1,2}C_{1,2}).$$

The likelihood function is then taken to be

$$y_1 \mid \theta_1 \sim N(A_1\theta_1, \alpha_{1,1}C_{1,1}+\alpha_{1,2}C_{1,2}),$$

while the prior for θ_1 is given by

$$\theta_1 \sim N(\mu_0, R_1+\Omega_1 V_0 \Omega_1').$$

It is now easy to see that the posterior of θ_1, given y_1 is

$$(\theta_1 \mid y_1) \sim N(\mu_1, V_1),$$

where

$$\mu_1 = \mu_{1|0} + V_{1|0} A_1' M_1^{-1} (y_1 - A_1 \mu_{1|0})$$

and

$$V_1 = V_{1|0} - V_{1|0} A_1' M_1^{-1} A_1 V_{1|0}$$

with

$$\mu_{1|0} = \Omega_1 \mu_0,$$

$$V_{1|0} = R_1 + \Omega_1 V_0 \Omega_1',$$

and

$$M_1 = \alpha_{1,1} C_{1,1} + \alpha_{1,2} C_{1,2} + A_1 V_{1|0} A_1'.$$

The posterior of $\theta_1 \mid y_1$ is our prior for the next stage. We continue in this way, and the resulting algorithm for proceeding in this manner at time t to $t+1$ is described in Table 3. Apart from the collapsing process at beginning, the derivation is identical to the standard Kalman filter.

4. Examples and discussion

We now reconsider two examples to illustrate the preceding mechanism and its performance.

Example 1. Consider the example given in Peña and Guttman (1989).

$$y_t = \theta_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, 4),$$

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

Here, $C_t = 4$, $R_t = 1$, $a_t = 1$, $\omega_t = 1$, for all t, and we assume that $\mu_0 = 10$, $V_0 = 10\,000$. Peña and Guttman (1989) discussed the standard Kalman filter when

$$\varepsilon_t \sim 0.95\ \mathrm{N}(0, 4) + 0.05\ \mathrm{N}(0, 100).$$

Table 4 summarizes the performance of the standard Kalman filter, the robust filter of Peña and Guttman, and the new (robust) filter of Section 3 of this paper. Note that the set of data contains 'y_{20}' that has been replaced by the value 35, and as intended, is an extreme outlier. We display the comparisons in Figures 1 to 3.

Figure 1 shows the performances of the updated mean for the three filters (μ_t's). We see that when y_t comes from a 'good run' (see Box and Tiao, 1968), all three approaches behave similarly. But if y_t is a 'bad run' (e.g., y_{20}), the new filter and the Peña and Guttman filter are stable, as Figure 1 shows. In contrast the standard Kalman filter is badly affected by y_{20}, and

Table 3
The new robust filter

Model:

$$y_t = A_t \theta_t + \varepsilon_t, \quad \varepsilon_t \sim \alpha_1 \mathrm{N}(0, C_{t,1}) + \alpha_2 \mathrm{N}(0, C_{t,2})$$
$$\theta_t = \Omega_t \theta_{t-1} + u_t, \quad u_t \sim \mathrm{N}(0, R_t)$$

Initial setting:

$\theta_0 \sim \mathrm{N}(\mu_0, V_0)$ and specified α_1 $(\alpha_2 = 1 - \alpha_1)$

Prediction:

$$\mu_{t|t-1} = \Omega_t \mu_{t-1}, \quad V_{t|t-1} = R_t + \Omega_t V_{t-1} \Omega_t', \quad \hat{y}_t = A_t \mu_{t|t-1}$$
$$M_{t,1} = C_{t,1} + A_t V_{t|t-1} A_t' \text{ and } M_{t,2} = C_{t,2} + A_t V_{t|t-1} A_t'$$

Compute posterior probabilities:

$$\alpha_{t,1} = [1 + \frac{\alpha_2}{\alpha_1} (\frac{\| M_{t,1} \|}{\| M_{t,2} \|})^{1/2} \exp\{\tfrac{1}{2}(y_t - A_t \mu_{t|t-1})'(M_{t,1}^{-1} - M_{t,2}^{-1})(y_t - A_t \mu_{t|t-1})\}]^{-1}$$
$$\alpha_{t,2} = 1 - \alpha_{t,1}$$
$$M_t = \alpha_{t,1} M_{t,1} + \alpha_{t,2} M_{t,2} = \alpha_{t,1} C_{t,1} + \alpha_{t,2} C_{t,2} + A_t V_{t|t-1} A_t'$$

Updating of the parameters:

$$\mu_t = \mu_{t|t-1} + V_{t|t-1} A_t' M_t^{-1} (y_t - A_t \mu_{t|t-1})$$
$$V_t = V_{t|t-1} - V_{t|t-1} A_t' M_t^{-1} A_t V_{t|t-1}$$
$$\text{Filter} = V_{t|t-1} A_t' M_t^{-1} = V_{t|t-1} A_t' (\alpha_{t,1} M_{t,1} + \alpha_{t,2} M_{t,2})^{-1}$$

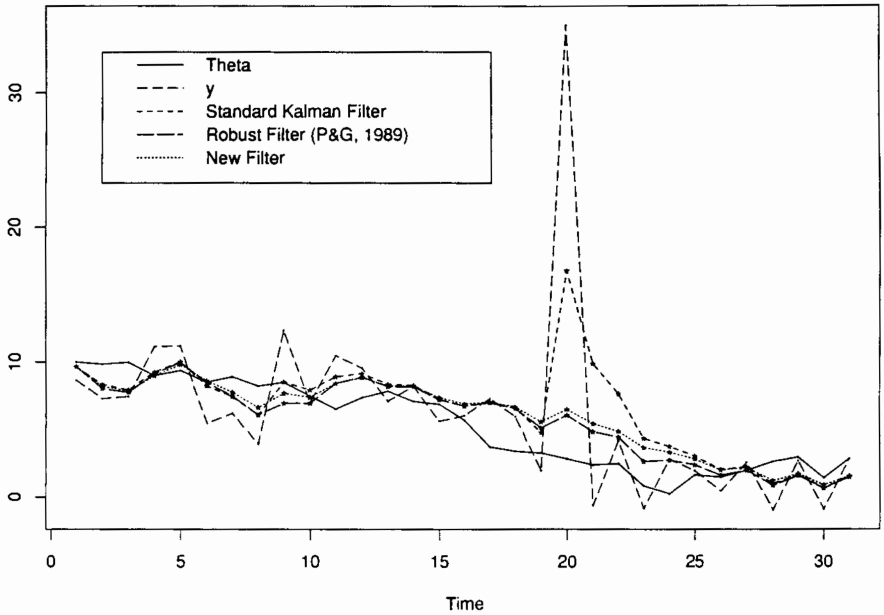

Fig. 1.

y_{20} affects μ_{21}, and μ_{22}. We note that here computationally the new filter is much simpler to handle than the Peña and Guttman filter.

Figure 2 shows the comparison of updated variances among the three filters. Frequently, the updated variance of the new filter is larger than that of Peña and Guttman filter, and of the standard Kalman filter. This will be explained later.

Figure 3 shows the comparison of $\alpha_{t,1}$ (the

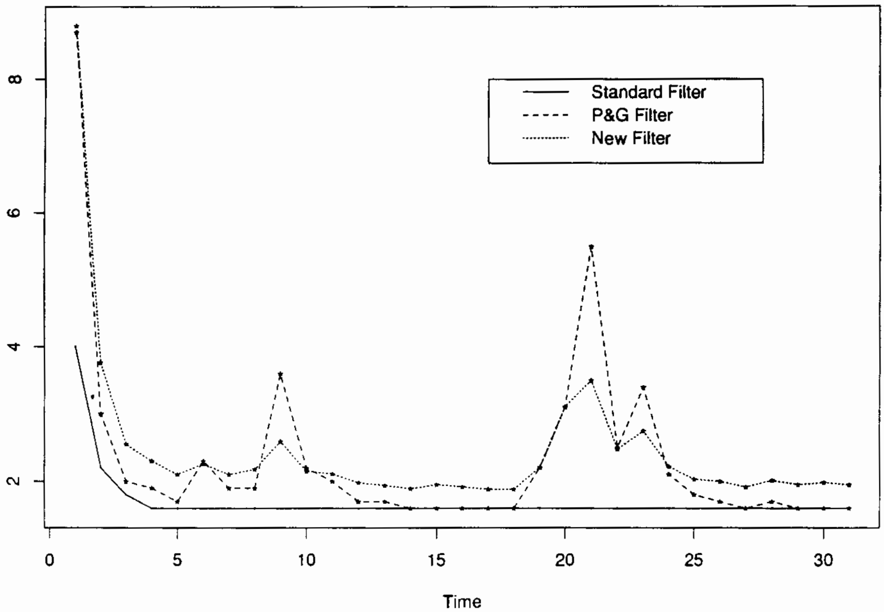

Fig. 2.

posterior probability that y_t was generated as a 'good run') between the Peña and Guttman filter and the new filters. The Peña and Guttman filter always has smaller $\alpha_{t,1}$ than that of the new filter, and the $\alpha_{t,1}$ values of Peña and Guttman filter for moderately behaved observations may seriously underestimate the probability that such an observation is truly generated from the source N(0, 4) — see, for example, Figure 3 for $\alpha_{t,1}$'s. We shall provide a more detailed discussion after the next example.

Example 2. Consider the example given in Meinhold and Singpurwalla (1983).

$$y_t = a_t\theta_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, 2),$$
$$\theta_t = \omega_t\theta_{t-1} + u_t, \quad u_t \sim \mathrm{N}(0, 1),$$

where $C_t = 2$, $R_t = 1$, $\omega_t = \frac{1}{2}(-1)^t$. The coefficients a_t's are also given by Meinhold and Singpurwalla (1983, Table 1). The initial values were taken as $\mu_0 = 4.183$, $V_0 = 1$. We assume that

$$\varepsilon_t \sim 0.95\ \mathrm{N}(0, 2) + 0.05\ \mathrm{N}(0, 30).$$

Note that it is assumed that there is *no* outlier in the original example, and we replaced their y_{17} by 10, which is indicated as an outlier.

Table 5 summarizes the performance of the updated mean for the standard Kalman filter, the robust filter in Peña and Guttman, and the new filter. Note that the posterior probabilities are identical, up to the second decimal, for the Peña and Guttman filter and the new filter of this paper, for the data of this example. Figures 4 and

Table 4
Comparison among three filters (Example 1)

t	y_t	θ_t	Mean (μ_t)			Variance (V_t)			$\alpha_{t,1}$	
			μ_{std}	$\mu_{P\&G}$	μ_{new}	V_{std}	$V_{P\&G}$	V_{new}	$\alpha_{P\&G}$	α_{new}
1	8.65	10.00	9.66	9.66	9.66	4.0	8.7	8.8	0.95	0.95
2	7.28	9.83	8.34	8.01	8.19	2.2	3.0	3.8	0.97	0.99
3	7.44	9.98	7.94	7.73	7.84	1.8	2.0	2.5	0.99	0.99
4	11.13	8.99	9.25	9.12	8.99	1.6	1.9	2.3	0.94	0.97
5	11.18	9.36	10.02	9.97	9.79	1.6	1.7	2.1	0.98	0.98
6	5.45	8.50	8.22	8.46	8.61	1.6	2.3	2.3	0.81	0.95
7	6.17	8.90	7.42	7.45	7.75	1.6	1.9	2.1	0.97	0.98
8	3.92	8.20	6.05	6.07	6.61	1.6	1.9	2.2	0.93	0.96
9	12.32	8.47	8.50	6.91	7.67	1.6	3.6	2.6	0.27	0.90
10	6.95	7.46	7.90	6.93	7.38	1.6	2.2	2.1	0.99	0.99
11	10.46	6.49	8.90	8.40	8.40	1.6	2.0	2.1	0.93	0.97
12	9.54	7.34	9.15	8.88	8.82	1.6	1.7	2.0	0.98	0.99
13	7.07	7.82	8.33	8.16	8.21	1.6	1.7	1.9	0.98	0.98
14	8.17	7.06	8.27	8.16	8.19	1.6	1.6	1.9	0.99	0.99
15	5.59	6.85	7.22	7.18	7.35	1.6	1.6	1.9	0.97	0.98
16	5.99	5.67	6.74	6.71	6.87	1.6	1.6	1.9	0.98	0.99
17	7.29	3.69	6.95	6.94	7.02	1.6	1.6	1.9	0.99	0.99
18	5.94	3.37	6.56	6.55	6.64	1.6	1.6	1.9	0.98	0.99
19	1.96	3.25	4.76	5.10	5.55	1.6	2.2	2.2	0.79	0.94
20	35.00	2.81	16.76	6.04	6.47	1.6	3.1	3.1	0.00	0.00
21	−0.62	2.36	9.86	4.81	5.41	1.6	5.5	3.5	0.31	0.80
22	4.13	2.46	7.62	4.40	4.84	1.6	2.5	2.5	0.98	0.98
23	−0.84	0.82	4.32	2.61	3.64	1.6	3.4	2.7	0.71	0.90
24	2.78	0.24	3.72	2.70	3.29	1.6	2.1	2.2	0.99	0.99
25	1.93	1.62	3.02	2.37	2.79	1.6	1.8	2.0	0.99	0.99
26	0.45	1.46	2.02	1.60	1.99	1.6	1.7	2.0	0.98	0.98
27	2.54	1.96	2.22	1.97	2.19	1.6	1.6	1.9	0.99	0.99
28	−0.95	2.62	0.98	0.86	1.21	1.6	1.7	2.0	0.96	0.97
29	2.69	2.95	1.65	1.58	1.74	1.6	1.6	1.9	0.98	0.98
30	−0.89	1.40	0.66	0.63	0.88	1.6	1.6	2.0	0.97	0.98
31	2.83	2.84	1.51	1.48	1.55	1.6	1.6	1.9	0.97	0.98

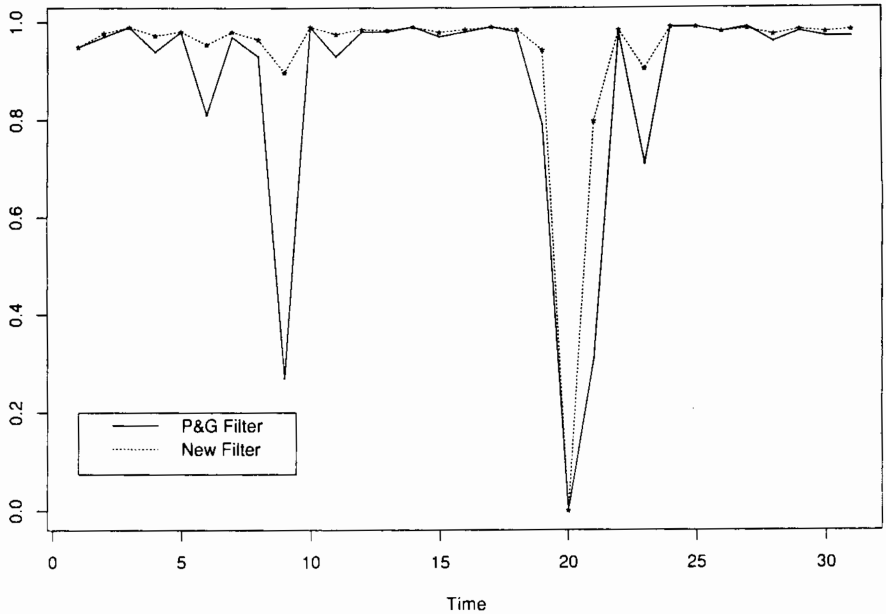

Fig. 3.

Table 5
Comparison among three filters (Example 2)

t	y_t	θ_t	Mean (μ_t)			Variance (V_t)			$\alpha_{t,1}$	
			μ_{std}	$\mu_{P\&G}$	μ_{new}	V_{std}	$V_{P\&G}$	V_{new}	$\alpha_{1,P\&G}$	$\alpha_{1,new}$
1	1.01	1.06	−0.62	−0.72	−1.12	0.61	0.77	0.83	0.924	0.924
2	−0.37	−0.49	−0.35	−0.39	−0.54	0.84	0.87	0.92	0.984	0.984
3	−1.76	−0.96	−0.53	−0.50	−0.31	0.81	0.84	0.91	0.972	0.971
4	1.28	−0.33	0.34	0.34	0.32	0.70	0.72	0.78	0.977	0.978
5	−0.90	−0.16	−0.43	−0.43	−0.40	0.64	0.65	0.71	0.981	0.981
6	0.11	0.37	−0.10	−0.10	−0.10	0.73	0.74	0.80	0.983	0.983
7	−1.52	0.22	−0.55	−0.54	−0.46	0.69	0.71	0.78	0.977	0.977
8	−2.41	−0.96	−1.05	−1.02	−0.84	0.79	0.82	0.90	0.968	0.967
9	1.04	0.67	0.73	0.72	0.63	0.81	0.82	0.87	0.983	0.983
10	0.37	0.07	0.37	0.36	0.33	0.75	0.76	0.82	0.983	0.983
11	−0.30	0.37	−0.21	−0.21	−0.20	0.64	0.65	0.71	0.982	0.982
12	−1.66	−0.60	−0.64	−0.63	−0.54	0.85	0.86	0.92	0.977	0.977
13	2.04	0.69	0.97	0.95	0.83	0.70	0.72	0.79	0.975	0.975
14	−1.30	−0.41	−0.04	−0.04	0.00	0.91	0.93	0.98	0.976	0.977
15	−0.92	0.09	−0.32	−0.32	−0.29	0.82	0.83	0.89	0.982	0.982
16	1.43	0.57	0.44	0.42	0.35	0.75	0.77	0.84	0.977	0.977
17	10.00	−0.06	3.74	0.29	0.33	0.59	1.12	1.13	0.001	0.001
18	−0.35	−0.39	0.98	−0.05	−0.02	0.68	0.73	0.79	0.982	0.982
19	1.64	−0.71	0.36	0.63	0.54	0.63	0.66	0.73	0.976	0.976
20	0.37	−1.24	0.25	0.34	0.31	0.79	0.80	0.85	0.984	0.984
21	−1.23	0.64	−0.50	−0.53	−0.46	0.93	0.93	0.98	0.981	0.981
22	1.64	0.95	0.29	0.27	0.19	1.01	1.02	1.07	0.973	0.974
23	−1.55	−0.79	−0.69	−0.68	−0.58	0.71	0.73	0.80	0.978	0.978
24	−1.19	−0.61	−0.66	−0.65	−0.58	0.74	0.75	0.81	0.982	0.981
25	0.12	2.23	0.26	0.26	0.24	0.80	0.81	0.86	0.984	0.984

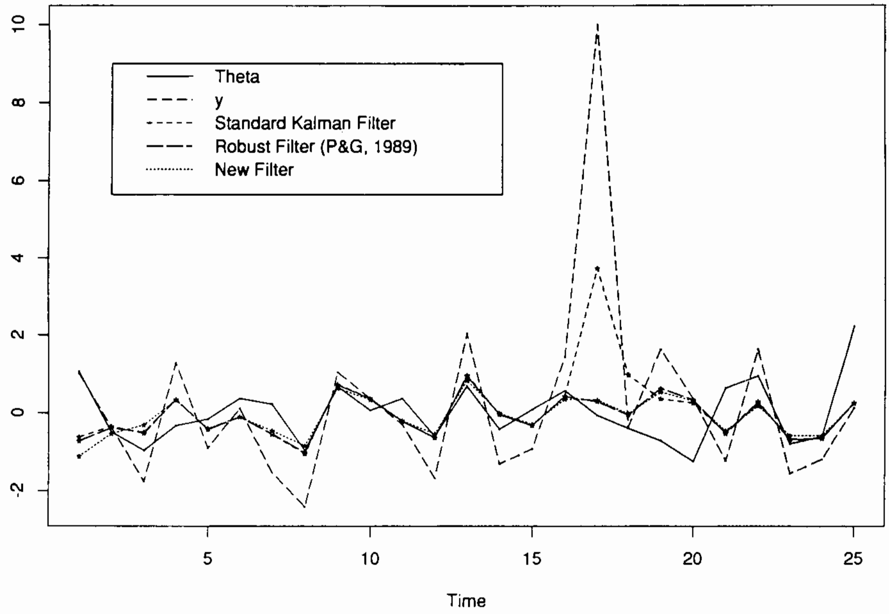

Fig. 4.

5 provide a visual comparison, of the updated mean and variance respectively.

Figure 4 shows the near-identical performances of the two robust filters, and both behave closely to the standard Kalman filter except for y_{17} where the standard Kalman filter, as expected, becomes unstable (this also affects μ_{18}). Figure 5 shows the comparison of updated variances among the three filters. It is of interest to note that the updated variance of the new filter is

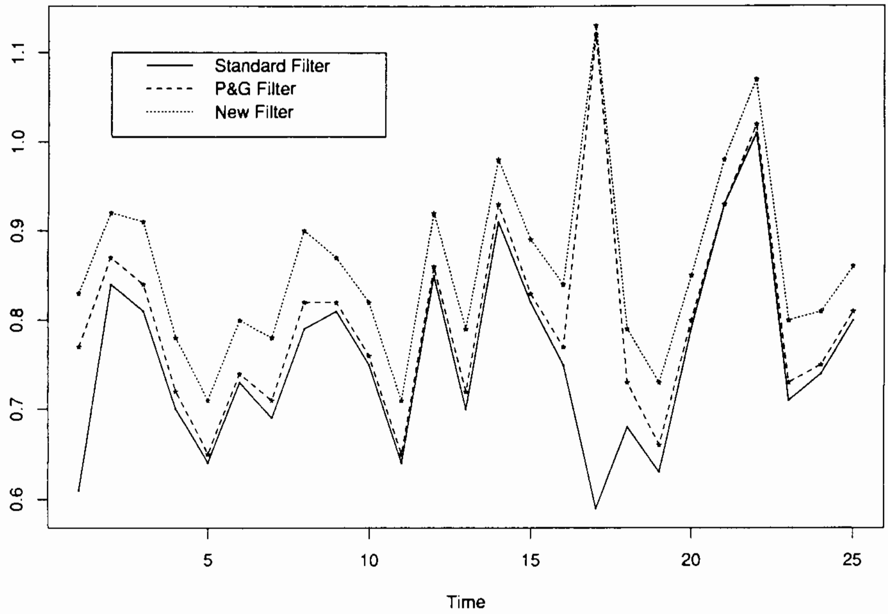

Fig. 5.

consistently larger than that of Peña and Guttman filter, and that of the standard filter. The differences are not substantial, however.

Remarks. To compare the different filtering processes under discussion, we note that the gain matrices of the three approaches have similar structure, but are indeed different as follows.

Standard Kalman filter:

$$V_{t|t-1}A_t'M_t^{-1}, \qquad M_t = C_t + A_t V_{t|t-1} A_t'.$$

Peña and Guttman filter:

$$V_{t|t-1}A_t'\left(\alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1}\right),$$
$$M_{t,i} = C_{t,i} + A_t V_{t|t-1} A_t'.$$

The new filter:

$$V_{t|t-1}A_t'M_t^{-1} = V_{t|t-1}A_t'(\alpha_{t,1}M_{t,1} + \alpha_{t,2}M_{t,2})^{-1}.$$

We will use the notation $C_1 \geqslant C_2$, for two matrices C_1 and C_2, when $C_1 - C_2$ is positive semidefinite. Now it can be shown that

$$\left[\alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1}\right] \geqslant \left[\alpha_{t,1}M_{t,1} + \alpha_{t,2}M_{t,2}\right]^{-1}, \tag{4.1}$$

and in fact,

$$\begin{aligned}&\left[\alpha_{t,1}M_{t,1}^{-1} + \alpha_{t,2}M_{t,2}^{-1}\right] - \left[\alpha_{t,1}M_{t,1} + \alpha_{t,2}M_{t,2}\right]^{-1}\\ &\quad = \alpha_{t,1}\alpha_{t,2}\left(M_{t,1}^{-1} - M_{t,2}^{-1}\right)\left(M_{t,2} - M_{t,1}\right)\\ &\qquad \times\left[\alpha_{t,1}M_{t,1} + \alpha_{t,2}M_{t,2}\right]^{-1}\\ &\quad \geqslant 0 \quad (\text{positive semidefinite}).\end{aligned}$$

Because of (4.1), if the initial conditions are the same for all three filters, then the updated variance of the new filter, which uses an approximate likelihood will certainly be larger than the updated variance of the Peña and Guttman filter for similar α_t's. This explains the behavior of Figure 5, for the two robust filters happen to have near-identical α_t's.

The equation (4.1) also implies that the updated mean of the new filter is much (robustly) closer to the mean of the likelihood function than that of the Peña and Guttman filter, a desirable property. (See Lin and Guttman (1991), for additional explanations.) This explains the behavior of Figure 3 in which the posterior probability $\alpha_{t,1}$ in the new filter is more 'stable' than that of Peña and Guttman filter.

We also note that the updated variance in the standard Kalman filter,

$$v_t = v_{t|t-1} - v_{t|t-1}A_t'M_t^{-1}A_t v_{t|t-1}$$

is independent of the observed y_t's. In particular, when $\Omega_t = A_t = I$, it will soon converge to the positive root of equation

$$V = (R+V) - (R+V)M^{-1}(R+V),$$
$$M = R + C + V.$$

This is illustrated by Example 1 where the value of 1.6, the updated variance for $t \geqslant 4$, is the positive root of

$$x = (1+x) - (1+x)(1+4+x)^{-1}(1+x).$$

The examples given above, take $C_{t,1}$ in the robust filters equal to C_t of the standard Kalman filter and also we take $C_{t,2} \gg C_{t,1}$, leading to an error with larger variance in the observation equation. Therefore, the updated variances for the standard Kalman filter are always less than that of the two robust filters. Note that the new filter is much easier to evaluate (cf. Table 2 and Table 3).

Acknowledgement

Irwin Guttman was partially supported by Grant A8743 from NSERC (Canada), and Dennis K.J. Lin was partially supported by a Faculty Research Fellowship through the College of Business Administration, University of Tennessee.

References

Abraham, B. and G.E.P. Box (1979), Bayesian analysis of some outlier problems in time series, *Biometrika* **66**, 229–236.

Aoki, M. (1967), *Optimization of Stochastic Systems* (Academic Press, New York).

Box, G.E.P. and G.C. Tiao (1968). A Bayesian approach to some outlier problems, *Biometrika* **55**, 119–129.

Guttman, I. and D. Peña (1985), Robust Kalman filtering and its applications, Tech. Rept. No. 85-1, Dept. of Statist., Univ. of Toronto (Toronto, Ont.).

Harrison, P.J. and C.F. Stevens (1976), Bayesian forecasting (with discussion), *J. Roy. Statist. Soc. Ser. B* **38**, 205–247.

Jefferys, H. (1961), *Theory of Probability* (Clarendon Press, Oxford, 3rd ed.).

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME* **82D**, 35–47.

Kitagawa, G. (1987), Non-Gaussian state-space modeling of nonstationary time series (with discussion), *J. Amer. Statist. Assoc.* **82**, 1032–1063.

Lin, D.K.J. and I. Guttman (1991), Handling spuriosity in the Kalman filter, Tech. Rept. No. 91-16, Dept. of Statist., Univ. of Toronto (Toronto, Ont.).

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.* **84**, 479–486.

Peña, D. and I. Guttman (1989), Optimal collapsing of mixture distributions in robust recursive estimation, *Comm. Statist.* **18**, 817–833.

Phadke, M.S. (1981), Quality audit using adaptive Kalman filtering, *ASQC Quality Congress Transaction* – San Francisco, CA, pp. 1045–1052.