Kybernetika

Petr Franěk

On improving sensitivity of the Kalman filter

Kybernetika, Vol. 38 (2002), No. 4, [425]--443

Persistent URL: http://dml.cz/dmlcz/135475

KYBERNETIKA — VOLUME 38 (2002), NUMBER 4, PAGES 425 – 443

ON IMPROVING SENSITIVITY OF THE KALMAN FILTER

PETR FRANĚK

The impact of additive outliers on a performance of the Kalman filter is discussed and less outlier-sensitive modification of the Kalman filter is proposed. The improved filter is then used to obtain an improved smoothing algorithm and an improved state-space model parameters estimation.

1. INTRODUCTION

State-space models represent a powerful modeling tool which opens a unified way of dealing with a wide range of time series models. A univariate state-space model consists of two processes – an unobservable n-variate Markov chain $(x_t)_{t\in\mathbb{T}}$, the *state process*, and the *observation process* $(y_t)_{t\in\mathbb{T}}$ arising from the state process via some transformation. In a general state-space model it is assumed that (i) conditionally on $(x_t)_{t\in\mathbb{T}}$ the observations y_t are independent and (ii) y_t depends only on x_t. The prominent role among the state-space models belongs to the *linear state-space model* in which the state x_t is a linear function of the state x_{t-1} and an additive random term and the observation y_t is a linear function of the state x_t and an additive random term.

The basic task solved in the state-space model environment is an estimation of unobserved states based on observed values. In the linear state-space model the linear minimum variance estimate of x_t based on observations $y_1, \dots, y_t$ is given by the *Kalman filter* introduced in [11]. Its performance, however, may be negatively affected by *additive outliers* – outlying observations caused by additive errors entering into the linear equation that transforms a state into an observation. This type of outlying observations is quite usual in practice and it is therefore desirable to look for modifications of the Kalman filter less sensitive to such errors in data.

Robust modifications of the Kalman filter were already sought by many authors. Generally, there are two main attitudes to this problem. First group of authors applies the technique of M-estimates and the robust estimate of the state is obtained by applying L_1-norm or other Huber-like function that bounds an impact of the outlying observation. For more details about this approach refer for example to [4] or [6]. Second group of authors deals the problem from the Bayesian point of view and general heavy-tailed distributions are used for filtering. Unfortunately, the simple recursive form of the Kalman filter is lost in this situation and the filtering recursions have to be performed numerically or using Monte–Carlo simulations. This concept, quite popular recently, is described for example in [12, 14, 15, 16] or [17]. For the most recent achievements at the field of the general state-space models refer for example also to [7, 13] or [19].

Both concepts have their benefits and drawbacks. In the first approach the simple recursive and linear structure of the Kalman filter is preserved but the observations are mechanically clipped without any analysis of a nature of the outliers (the missing theory justifying this approach was lately presented in [18]). The second approach may turn into numerically demanding solutions.

A Kalman filter modification proposed in this paper combines both approaches. It preserves the simple recurrent form of the Kalman filter but the updating function determining the influence of a new observation on the filtered estimate is chosen on-line according to a nature of the outlying observation.

2. STATE–SPACE MODEL DEFINITION AND THE KALMAN FILTER

2.1. State–space model definition

A univariate linear state-space model of a time series will be used in this paper. The model is assumed to be in the form

$$\begin{aligned} y_t &= \boldsymbol{h}\, \boldsymbol{x}_t + v_t \\ \boldsymbol{x}_t &= \boldsymbol{F}\, \boldsymbol{x}_{t-1} + \boldsymbol{w}_t, \end{aligned} \tag{1}$$

where $\boldsymbol{x}_t$ is an unknown n-variate state vector and y_t is a univariate observation, $\boldsymbol{h}$ is a known $(1 \times n)$ vector, $\boldsymbol{F}$ is a known $(n \times n)$ matrix, and v_t and $\boldsymbol{w}_t$ are independent centered random residuals with $\operatorname{var} v_t = \sigma^2$ and $\operatorname{var} \boldsymbol{w}_t = \boldsymbol{R}$, σ^2 being an unknown value and $\boldsymbol{R}$ being a known $(n \times n)$ matrix. For the starting state $\boldsymbol{x}_0$ the standard assumptions $\mathrm{E}\, \boldsymbol{x}_0 = \tilde{\boldsymbol{x}}_0$ and $\operatorname{var} \boldsymbol{x}_0 = \boldsymbol{P}_0$ will be used.

If parameters $\boldsymbol{h}$, $\boldsymbol{F}$, σ^2 and $\boldsymbol{R}$ are constant in time (as is the case in this paper), the model (1) is referred to as *time-invariant*. If the errors v_t and $\boldsymbol{w}_t$ are assumed to be Gaussian, the model (1) is referred to as *Gaussian*. Unless otherwise stated the model (1) will be assumed to be Gaussian in this paper.

2.2. The Kalman filter

Denoting the history of observations up to time t as $Y_t = \{y_1, \ldots, y_t\}$ the best unbiased linear estimate $\hat{\boldsymbol{x}}_{t|t}$ of the unknown state $\boldsymbol{x}_t$ (based on Y_t) and its covariance matrix $\boldsymbol{P}_{t|t} = \mathrm{E}\,(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t})(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t})'$ are given by the Kalman filter recursions

$$\begin{aligned}
\hat{\boldsymbol{x}}_{t|t} &= \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{h}' d_t^{-2}(y_t - \boldsymbol{h}\hat{\boldsymbol{x}}_{t|t-1}) \\
\hat{\boldsymbol{x}}_{t|t-1} &= \boldsymbol{F}\hat{\boldsymbol{x}}_{t-1|t-1} \\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{h}' d_t^{-2}\boldsymbol{h}\boldsymbol{P}_{t|t-1} \\
\boldsymbol{P}_{t|t-1} &= \boldsymbol{F}\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{R} \\
d_t^2 &= \boldsymbol{h}\boldsymbol{P}_{t|t-1}\boldsymbol{h}' + \sigma^2,
\end{aligned} \tag{2}$$

where $t \in \mathbb{T} = \{1, 2, \dots\}$ and the filter is started by using $\hat{\boldsymbol{x}}_{0|0} = \tilde{\boldsymbol{x}}_0$ and $\boldsymbol{P}_{0|0} = \boldsymbol{P}_0$. In the Gaussian linear state-space model it is $\mathcal{L}(\boldsymbol{x}_t|Y_t) = N(\hat{\boldsymbol{x}}_{t|t}, \boldsymbol{P}_{t|t})$, i.e. the conditional distribution of the state given the history of observations is fully determined by the output of the Kalman filter.

Denote $\hat{y}_t = \boldsymbol{h}\,\hat{\boldsymbol{x}}_{t|t-1}$. The terms $I_t = y_t - \hat{y}_t$ are called *innovations.* According to the nature of the Kalman filter the innovations are centered, serially uncorrelated and the variables d_t^2 are their variances. In the Gaussian state-space model it is $\mathcal{L}(I_t) = N(0, d_t^2)$ and the innovations are independent (refer to [1] for further discussion).

The Kalman filter is said to converge to a *stable solution* if the covariance matrix $\boldsymbol{P}_{t|t-1}$ (and hence also the matrix $\boldsymbol{P}_{t|t}$) converges to a constant matrix. This feature is determined by the underlying state-space model. The Kalman filter run in a time invariant state-space model (1) converges to the stable solution if at least one of the following conditions applies:

- $|\lambda_i(\boldsymbol{F})| < 1$ for $i = 1, \dots, n$, $\lambda_i(\boldsymbol{F})$ being an eigenvalue of the matrix $\boldsymbol{F}$ (*asymptotical stability condition*), and $\boldsymbol{P}_0 \geq 0$;
- there is a matrix $\boldsymbol{S}$ such that $|\lambda_i(\boldsymbol{F}+\boldsymbol{G}\boldsymbol{S}')| < 1$, $i = 1, \dots, n$, $\boldsymbol{G}$ being a matrix satisfying $\boldsymbol{G}\boldsymbol{G}' = \boldsymbol{R}$ (*stabilisability condition*), there is a matrix $\boldsymbol{D}$ such that $|\lambda_i(\boldsymbol{F} + \boldsymbol{D}\boldsymbol{h})| < 1$, $i = 1, \dots, n$ (*detectability condition*), and $\boldsymbol{P}_0 \geq 0$.

If the Kalman filter converges to the stable solution, the impact of the assumption made about the unknown starting state (i.e. $\tilde{\boldsymbol{x}}_0$ and $\boldsymbol{P}_0$) is forgotten in time. It may be seen from the stabilisability and detectability conditions that this diminishing effect depends on the positive definiteness of the matrix $\boldsymbol{R}$. Proofs of these features may be found in [1] or in [14].

3. KALMAN FILTER PERFORMANCE ON DATA WITH ADDITIVE OUTLIERS

3.1. Model of additive outliers

Additive outliers are usually modeled by replacing the observation error term v_t with the following error term:

$$(1 - Z_t)v_t + Z_t q_t. \tag{3}$$

Here $v_t \sim N(0, \sigma^2)$ corresponds to errors from the model (1), $q_t \sim H_t$ (H_t being a centered symmetric distribution with a variance ψ_t^2, $\psi_t^2 > \sigma^2$) are errors producing the outliers independent of $(v_t)_{t\in\mathbb{T}}$ and Z_t are deterministic or random indicators (independent of v_t and q_t) of outliers (e.g. $Z_t \sim Alt(\gamma)$, $\gamma \in (0; 1)$).

3.2. Impact of one outlying observation

Assume that y_t is the outlying observation and there are no other outliers, i. e. $Z_j = 1$ for $j = t$ and $Z_j = 0$ otherwise. Then $\operatorname{var} v_t = \psi_t^2 > \sigma^2$ is the right variance at time t. Denoting $(d_t^R)^2$ the right value of the innovation variance at this time and observing

$$d_t^2 = \boldsymbol{h}\boldsymbol{P}_{t|t-1}\boldsymbol{h}' + \sigma_t^2 = (d_t^R)^2 - (\psi_t^2 - \sigma_t^2)$$
$$d_t^{-2} = (d_t^R)^{-2} + \frac{\psi_t^2 - \sigma_t^2}{(d_t^R)^2((d_t^R)^2 - (\psi_t^2 - \sigma_t^2))} = (d_t^R)^{-2} + \frac{\psi_t^2 - \sigma_t^2}{(d_t^R)^2 d_t^2}$$

we get

$$\begin{aligned} \hat{\boldsymbol{x}}_{t|t} &= \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{h}' d_t^{-2}(y_t - \boldsymbol{h}\hat{\boldsymbol{x}}_{t|t-1}) \\ &= \hat{\boldsymbol{x}}_{t|t}^R + \boldsymbol{P}_{t|t-1}\boldsymbol{h}' \frac{\psi_t^2 - \sigma_t^2}{(d_t^R)^2 d_t^2}(y_t - \boldsymbol{h}\hat{\boldsymbol{x}}_{t|t-1}) \\ &= \hat{\boldsymbol{x}}_{t|t}^R + \Delta\hat{\boldsymbol{x}}_{t|t}. \end{aligned} \tag{4}$$

Here $\hat{\boldsymbol{x}}_{t|t}^R$ is the right filtered value (filtered value obtained by using the true observation error variance ψ_t^2). According to this formula the estimate of the state is moved away from the right value and the magnitude of this move depends on the value of the innovation and on the difference between the real variance and the variance used for filtering.

We may study how the error propagates to following estimates. We have

$$\hat{\boldsymbol{x}}_{t+1|t} = \boldsymbol{F}\hat{\boldsymbol{x}}_{t|t} = \boldsymbol{F}\hat{\boldsymbol{x}}_{t|t}^R + \boldsymbol{F}\Delta\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t+1|t}^R + \Delta\hat{\boldsymbol{x}}_{t+1|t} \tag{5}$$

and then

$$\begin{aligned} \hat{\boldsymbol{x}}_{t+1|t+1} &= \hat{\boldsymbol{x}}_{t+1|t} + \boldsymbol{P}_{t+1|t}\boldsymbol{h}' d_{t+1}^{-2}(y_{t+1} - \boldsymbol{h}\hat{\boldsymbol{x}}_{t+1|t}) \\ &= \hat{\boldsymbol{x}}_{t+1|t}^R + \Delta\hat{\boldsymbol{x}}_{t+1|t} + \boldsymbol{P}_{t+1|t}\boldsymbol{h}' d_{t+1}^{-2}(y_{t+1} - \boldsymbol{h}\hat{\boldsymbol{x}}_{t+1|t}^R - \boldsymbol{h}\Delta\hat{\boldsymbol{x}}_{t+1|t}) \\ &= \hat{\boldsymbol{x}}_{t+1|t+1}^R + (\boldsymbol{I} - d_{t+1}^{-2}\boldsymbol{P}_{t+1|t}\boldsymbol{h}'\boldsymbol{h})\boldsymbol{F}\Delta\hat{\boldsymbol{x}}_{t|t} \\ &= \hat{\boldsymbol{x}}_{t+1|t+1}^R + \Delta\hat{\boldsymbol{x}}_{t+1|t+1}. \end{aligned} \tag{6}$$

Following the same steps we finally get for $n \in \mathrm{N}$

$$\begin{aligned} \hat{\boldsymbol{x}}_{t+n|t+n} &= \hat{\boldsymbol{x}}_{t+n|t+n}^R + \left(\prod_{i=1}^{n}(\boldsymbol{I} - d_{t+i}^{-2}\boldsymbol{P}_{t+i|t+i-1}\boldsymbol{h}'\boldsymbol{h})\boldsymbol{F}\right)\Delta\hat{\boldsymbol{x}}_{t|t} \\ &= \hat{\boldsymbol{x}}_{t+n|t+n}^R + \Delta\hat{\boldsymbol{x}}_{t+n|t+n}. \end{aligned} \tag{7}$$

Denoting

$$\begin{aligned} \boldsymbol{A}_0^t &= \boldsymbol{I} \\ \boldsymbol{A}_n^t &= \prod_{i=1}^{n}(\boldsymbol{I} - d_{t+i}^{-2}\boldsymbol{P}_{t+i|t+i-1}\boldsymbol{h}'\boldsymbol{h})\boldsymbol{F} \end{aligned} \tag{8}$$

we may conclude that

$$\hat{x}_{t+j|t+j} = \hat{x}^R_{t+j|t+j} + A^t_j \Delta \hat{x}_{t|t}, \quad j = 0, 1, \dots \tag{9}$$

Thus the impact of the outlying observation diminishes if A^t_n tends to zero. For example, this is the case if the filter converges to a stable solution. As the recursive formula for A^t_n does not depend on observations, it may be calculated before the filtering starts and may be used as an indicator of a sensitivity of the system to outlying observations.

Similar result may be observed in the case of $P_{t|t}$. Now we get

$$\begin{aligned} P_{t|t} &= P_{t|t-1} - P_{t|t-1} h' d_t^{-2} h P'_{t|t-1} \\ &= P^R_{t|t} - P_{t|t-1} h' \frac{\psi_t^2 - \sigma_t^2}{(d_t^R)^2 d_t^2} h P'_{t|t-1} \\ &= P^R_{t|t} - C(\psi_t^2 - \sigma_t^2). \end{aligned} \tag{10}$$

If compared with the right value, $P_{t|t}$ is also affected and in the case y_t is an outlier it underestimates the true variance of $\hat{x}_{t|t}$. It is, however, questionable if the impact of the outlying observation should be reflected in this matrix. Believing the model (1) is correct, the matrix $P_{t|t}$ should not depend on observations. According to this argument, the impact of outliers will be reflected only in the filtered estimates of the state further in this paper.

4. IDENTIFICATION AND PROCESSING OF OUTLIERS

4.1. Outliers detection in a completely identified model

If all parameters of the state space model (1) are known, the identification of the outlying observations is easy. Since $I_t = y_t - \hat{y}_t \sim N(0, d_t^2)$, the observation y_t may be identified as an outlier (on the probability level α) if

$$\frac{|y_t - \hat{y}_t|}{d_t} \geq u_{1-\alpha/2}, \tag{11}$$

$u_{1-\alpha/2}$ being an appropriate quantile of the standardized normal distribution.

4.2. Outliers detection in a model with unknown σ^2

In practical applications, however, the observation error variance σ^2 is usually unknown. It has to be estimated so that the identification of outlying observations might be performed. To keep the on-line property of the Kalman filter this estimate should also be estimated on-line using only the history of observations.

4.2.1. Adaptive on-line estimation of the observation variance σ^2

Suppose all parameters except σ^2 are known. Assuming the Kalman filter gets into the stable state after some time $t_0(\sigma_S^2)$ (note that this time depends on the value σ_S^2

used to start the filter) we know that innovations are uncorrelated $N(0, d^2)$ random variables after this time. This feature may be used to construct the estimate $\hat{\sigma}^2$.

Taking additional $m > 2$ observations after time $t_0 = t_0(\sigma_S^2)$ to start the process the following recursive estimate of the innovation variance may be used:

$$\begin{aligned}\hat{d}^2_{t_0+m+s} &= \frac{1}{m+s} \sum_{j=t_0}^{t_0+m+s} I_j^2 \\ &= \frac{1}{m+s} \sum_{j=t_0}^{t_0+m+s-1} I_j^2 + \frac{1}{m+s} I^2_{t_0+m+s} \\ &= \frac{m+s-1}{m+s} \hat{d}^2_{t_0+m+s-1} + \frac{1}{m+s} I^2_{t_0+m+s}, \qquad s = 1, 2, \ldots \end{aligned} \tag{12}$$

The Kalman filter with d_t^2 replaced with this estimate converges to the right value of $\boldsymbol{P}_{t|t}$ but this convergence is rather slow. The reason is obvious. The time $t_0(\sigma_S^2)$ as well as the innovations obtained after this time depend on the value σ_S^2 used to start the filter. The impact of the wrong observation variance σ_S^2 is present in all historical innovations (and thus in the estimate $\hat{d}^2_{t_0+m+s}$ as well) and only slowly forgotten by the system. To improve the speed of the convergence it would be helpful to start the filter with an estimate of σ^2 based on the observation history (i. e. to set $\sigma_S^2 = \hat{\sigma}^2_{t_0+m+s-1}$) and to recalculate all innovations.

The estimate of $\hat{\sigma}^2_{t_0+m+s-1}$ may be obtained by using several techniques with different additional computational burden.

EM estimate. The EM algorithm applied in the state-space model environment is probably the most universal solution (the EM-algorithm is briefly introduced in Section 5.4). Good features of the resulting estimate $\hat{\sigma}^2_{t_0+m+s-1}$ may be expected since σ^2 is assumed to be the only unknown parameter of the state-space model. However, the computation of the estimate for each observation may result in too lengthy processing of the data, even if the number of the iterations is limited. This computational complexity may be lowered by processing the data from a window covering only $M > t_0(\hat{\sigma}^2_{t_0+m+s-1}) + m$ most recent observations.

Bayesian on-line estimate. Several authors proposed an on-line estimate of the unknown parameters of the linear state-space model (refer for example to [3] or [13]). Among these approaches a bank of several Kalman filters run with different values of σ^2 is a popular solution. The estimation of σ^2 is then performed on-line using the Bayesian approach – refer to [1] for more details. The processing would be faster than in the previous case because no repeated processing of the data is involved. However, at least some prior information about σ^2 is required when defining the bank of the Kalman filters.

Simple recursive estimate. The following simple recursive estimate gives good practical results for univariate time series and is not too computationally demanding.

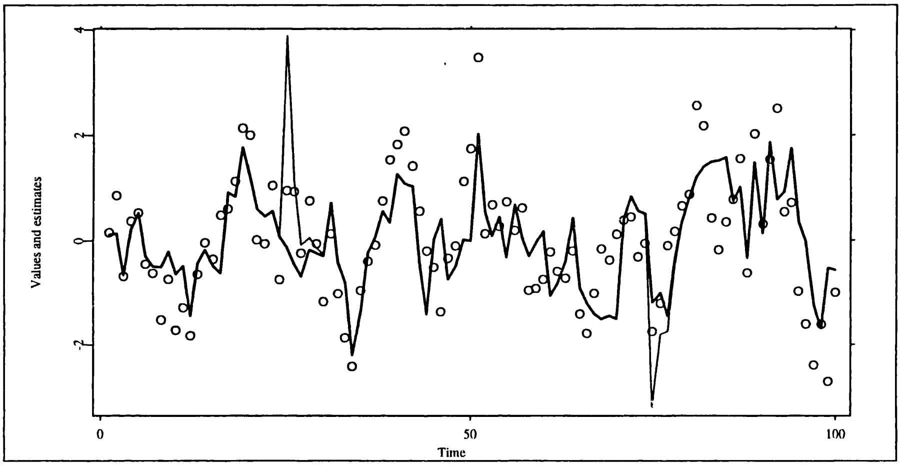

Fig. 1. True (circles) and filtered state obtained from the Kalman filter run on clean (thick) and contaminated (thin) data.

Denote $\lambda_s = 1/(m+s-1)$. After time $t_0(\sigma_S^2)$ it is $d_t^2 = d^2$ and $\boldsymbol{P}_{t|t-1} = \boldsymbol{P}$ and $d^2 = \boldsymbol{hPh}' + \sigma^2$. We then may get the recursive estimate of the observation variance as follows

$$\begin{aligned}\hat{\sigma}^2_{t_0+m+s} &= \hat{d}^2_{t_0+m+s} - \boldsymbol{hPh}' \\ &= (1-\lambda_s)\hat{\sigma}^2_{t_0+m+s-1} + \lambda_s(I^2_{t_0+m+s} - \boldsymbol{hPh}'). \end{aligned} \tag{13}$$

For practical purposes this estimate is not very satisfactory from the following two reasons: (i) for small m the estimate is negative with high probability, (ii) for small s there is too much weight given to a newly observed innovation and the resulting estimate is not smooth. It is therefore better to replace it with some modification, for example

$$\hat{\sigma}^2_{t_0+m+s} = (1-\lambda_s)\hat{\sigma}^2_{t_0+m+s-1} + \lambda_s(I^2_{t_0+m+s} - \boldsymbol{hPh}')^+ \tag{14}$$

with λ_s set as above or as an arbitrary constant (close to zero). As may be easily seen, this second estimate is not asymptotically unbiased but gives good practical results.

More computing effort is again the cost of improved speed of the convergence. Increased computational intensity may be bound if the recalculation of the innovations does not start at the beginning of the time series but only M observations back in history, $M > t_0(\hat{\sigma}^2_{t_0+m+s-1}) + m$, or by recalculating the estimate after obtaining more than one observation. The value m should be selected to obtain an applicable properties of the estimate $\hat{d}^2_{t_0+m+1}$, $m = 10$ proved to be sufficient.

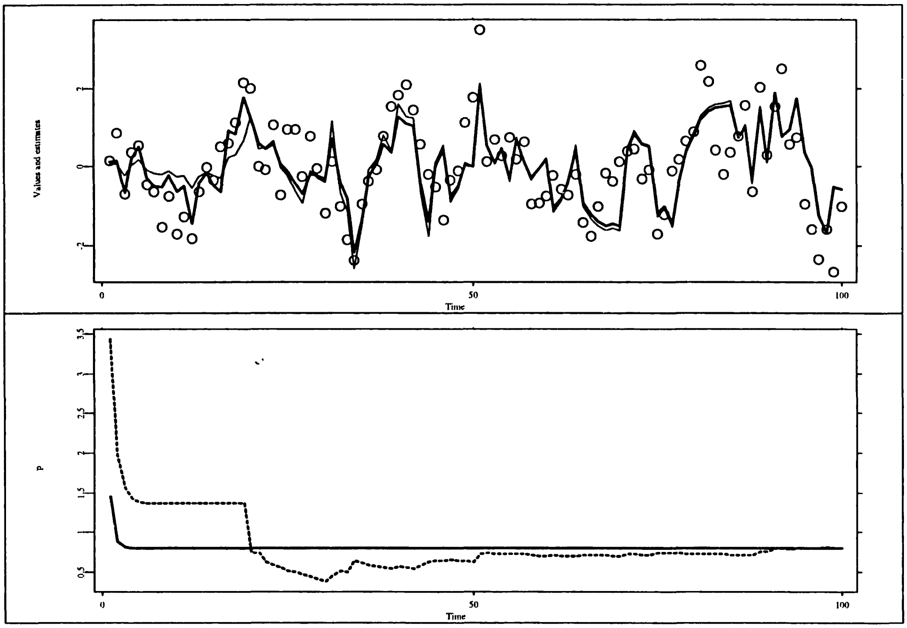

Fig. 2. True (circles) and filtered state obtained from the Kalman filter (thick) and the modified filter (thin) run on the data $\boldsymbol{Y}$ – upper panel; variance $\boldsymbol{P}_{t|t}$ obtained from the Kalman filter (solid) and the modified filter (dashed) – lower panel.

Example 4.1. One hundred observations were generated from the following model:

$$\begin{aligned} y_t &= x_t + v_t \\ x_t &= 0.65\, x_{t-1} + w_t, \end{aligned} \tag{15}$$

where $v_t \sim N(0,2)$ and $w_t \sim N(0,1)$ (data $\boldsymbol{Y}$). Then observations number 25 and 75, respectively, were shifted by 10 upwards and 5 downwards, respectively, (data $\boldsymbol{Y_{AO}}$).

Figure 1 shows the filtered state obtained by the Kalman filter run on data sets $\boldsymbol{Y}$ and $\boldsymbol{Y_{AO}}$. As may be seen, in the case of the contaminated data $\boldsymbol{Y_{AO}}$ the state estimates are spoiled after the outlier occurred. The impact of the outliers was fully absorbed – up to a third decimal place – after 9 observations in both cases.

Figure 2 shows the filtered state obtained by the Kalman filter and the modified filter with the on-line observation variance estimation run on the data set $\boldsymbol{Y}$. The Kalman filter was run using the right values of the parameters, the modified filter was run using $\sigma^2 = 10$ as the starting value and $m = 10$. The filter returned the final estimate $\hat{\sigma}^2_{100} = 3.033$, but the standard deviation of the filtered state converged to the right value faster (after 50 observations the absolute difference between standard deviations returned by the two respective filters was less than 0.04).

4.2.2. Outliers detection

Having the on-line estimate $\hat{d}_t^2$ of the innovation variance, $t > t_0(\hat{\sigma}_{t-1}^2) + m$, the observation y_t may be detected as an outlier if

$$\frac{|y_t - \hat{y}_t|}{\hat{d}_t} \geq u_{1-\alpha/2}, \quad t > t_0(\hat{\sigma}_{t-1}^2) + m, \tag{16}$$

$u_{1-\alpha/2}$ being an appropriate quantile of the standardized normal distribution (but now α can not be interpreted as an exact confidence level as the estimate of innovation variance is used – instead it reflects a level of insurance against outliers). Some ideas how to set α may be found in [18] but in this paper all parameters of the state-space model are assumed to be known.

4.3. On-line processing of outlying observations

Outlying observations are usually treated by replacing the innovation term $d_t^{-2}(y_t - \hat{y}_t)$ in the Kalman filter with some general function $g(I_t)$. The Huber function

$$g_{\mathrm{Hub}}(I_t) = \begin{cases} I_t/\hat{d}_t^2 & \text{if } I_t \in \langle 0; K\hat{d}_t\rangle \\ K/\hat{d}_t & \text{if } I_t > K\hat{d}_t \\ -g_{\mathrm{Hub}}(-I_t) & \text{if } I_t < 0, \end{cases} \tag{17}$$

where K is usually some quantile of the standardized normal distribution (e.g. $u_{1-\alpha/2}$) is quite popular. However, this choice is made *ad hoc*. Two other updating functions based on the model of additive outliers are proposed in this paper.

If the observation y_t is identified as an outlier using the rule (11) or (16), the estimate $\hat{x}_{t|t}$ should be constructed with respect to this fact. The impact of the suspected observation on the filtered value should be reduced but simply omitting it would be too strict – some measure of a distance of the particular observation from what was expected according to the history of observations should be taken into account. For this purpose a system of distributions is proposed from which the appropriate distribution H_t is selected to model the outlier.

4.3.1. System of normal distributions

The simplest applicable system is in the form

$$\mathcal{L}_1 = \left\{ H(q); \frac{\partial H}{\partial q} = h(q) = \frac{1}{\sqrt{2\pi}\rho} \exp\left[-\frac{1}{2}\frac{q^2}{\rho^2}\right], \rho \in (0, \infty) \right\}. \tag{18}$$

It is a system of normal distributions with the standard deviation ρ used as the factor that drives the observation error variance. This parameter should correspond to a distance of the observed value y_t from the expected value $\hat{y}_t$. For this purpose the attained level p_t of the test (11) or (16) may be used, for example $p_t = 2\,(1 - \Phi(|I_t/\hat{d}_t|))$, where Φ is a cumulative distribution function of the standardized normal distribution. Then $\rho_t = \rho(p_t) = \rho(y_t - \hat{y}_t)$ may be any function satisfying $\rho_t \to \hat{\sigma}_t$ for $p_t \to \alpha-$, $\rho_t \to \infty$ for $p_t \to 0+$ and $\rho_t = \hat{\sigma}_t$ for $p_t \geq \alpha$.

Denoting $s_t^2 = \boldsymbol{h}\boldsymbol{P}_{t|t-1}\boldsymbol{h}'$ the updating function in this system is in the form

$$g_{L1}(I_t) = I_t/(s_t^2 + \rho^2(I_t)). \tag{19}$$

4.3.2. System of generalized error distributions

The previous system has good practical results but modeling outliers using normal distribution may be considered as improper. To avoid this criticism the following system proposed in [2] (and used to model errors in data also by other authors) may be used:

$$\mathcal{L}_2 = \left\{ H(q); \frac{\partial H}{\partial q} = h(q) = \frac{k}{\varphi} \exp\left[-\frac{1}{2} \left| \frac{q}{\varphi} \right|^{\frac{2}{1+\rho}} \right], \varphi > 0, \rho \in [0,1] \right\}, \tag{20}$$

where $1/k = \Gamma((3+\rho)/2)\, 2^{(3+\rho)/2}$. The variance of a random variable X with the density $h(q)$ taken from this system is

$$\operatorname{var} X = 2^{1+\rho} \varphi^2 \frac{\Gamma\left(\frac{3}{2}(1+\rho)\right)}{\Gamma\left(\frac{1}{2}(1+\rho)\right)}.$$

As may be seen, the value $\rho = 0$ corresponds to the normal distribution, the value $\rho = 1$ corresponds to the double-exponential distribution which is the heaviest one from this system (this system may be enlarged by taking $\rho \in [0, b], b > 1$ with a reasonable bound $b < 10$). The parameter φ^2 should be related to the estimate $\hat{\sigma}_t^2$ so that for $\rho = 0$ the variance of the selected distribution H_t was equal to φ^2, i.e. $\varphi^2 = \hat{\sigma}_t^2$.

The parameter ρ_t should then again correspond to a distance of the observed value y_t from the expected value $\hat{y}_t$ and should satisfy that $\rho_t \to 0$ for $p_t \to \alpha-$, $\rho_t \to 1$ for $p_t \to 0+$ and $\rho_t = 0$ for $p_t \geq \alpha$.

To obtain the updating function for this system the following theorem may be used (its proof may be found in [15]).

Theorem. (Masreliez, 1975) Suppose $\mathcal{L}(\boldsymbol{x}_t|Y_{t-1}) = N(\hat{\boldsymbol{x}}_{t|t-1}, \boldsymbol{P}_{t|t-1})$ and that the conditional density $p(y_t|Y_{t-1})$ is twice differentiable. Then

$$\begin{aligned} \hat{\boldsymbol{x}}_{t|t} &= \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{h}' g_t(y_t) \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{h}' G_t(y_t) \boldsymbol{h} \boldsymbol{P}_{t|t-1} \\ \boldsymbol{P}_{t+1|t} &= \boldsymbol{F}\boldsymbol{P}_{t|t}\boldsymbol{F}' + \boldsymbol{R}, \end{aligned} \tag{21}$$

where

$$g_t(y_t) = -\frac{1}{p(y_t|Y_{t-1})} \frac{\partial p(y_t|Y_{t-1})}{\partial y_t} \qquad \text{and} \qquad G_t(y_t) = \frac{\partial g_t(y_t)}{\partial y_t}.$$

This theorem requires the density $p(y_t|Y_{t-1})$ and its first and second derivatives in the point y_t. In a fact it is necessary to find the convolution $N(\boldsymbol{h}\hat{\boldsymbol{x}}_{t|t-1}, \boldsymbol{h}\boldsymbol{P}_{t|t-1}\boldsymbol{h}') * H_t$, $H_t \in \mathcal{L}_2$, i.e. to find the integral

$$p(y_t|Y_{t-1}) = \frac{1}{\sqrt{2\pi}\hat{\tau}}\frac{k}{\varphi}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\frac{(x-\hat{y}_t)^2}{s_t^2}\right)\exp\left(-\frac{1}{2}\left|\frac{x-y_t}{\varphi}\right|^{\frac{2}{1+\rho}}\right)\mathrm{d}x. \tag{22}$$

A closed form solution does not exist. But using the Taylor expansion and the fact that the absolute value of a normally distributed random value has the folded-normal distribution we may use the following asymptotically equivalent approximation

$$\begin{aligned} p(y_t|Y_{t-1}) &= \frac{1}{\sqrt{2\pi}\hat{\tau}}\frac{k}{\varphi}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\frac{(z+y_t-\hat{y}_t)^2}{s_t^2}\right)\exp\left(-\frac{1}{2}\left|\frac{z}{\varphi}\right|^{\frac{2}{1+\rho}}\right)\mathrm{d}x \\ &= \frac{k}{\varphi}\mathrm{E}_{N(-I_t,s_t^2)}\exp\left(-\frac{1}{2}\left|\frac{Z}{\varphi}\right|^{\frac{2}{1+\rho}}\right) \\ &\approx \frac{k}{\varphi}\exp\left(-\frac{1}{2}\left(\frac{\mathrm{E}\,|Z|}{\varphi}\right)^{\frac{2}{1+\rho}}\right), \end{aligned}$$

where

$$\mathrm{E}\,|Z| = \sqrt{\frac{2}{\pi}}\,\hat{\tau}\exp\left(-\frac{1}{2}\left(\frac{y_t-\hat{y}_t}{s_t}\right)^2\right) - (y_t-\hat{y}_t)\left(1-2\Phi\left(\frac{y_t-\hat{y}_t}{s_t}\right)\right).$$

Further it is

$$\begin{aligned} \frac{\partial p(y_t|Y_{t-1})}{\partial y_t} &= p(y_t|Y_{t-1})\frac{1}{\varphi^{\frac{2}{1+\rho}}}\frac{1}{1+\rho}\times \\ &\times\left[\sqrt{\frac{2}{\pi}}\hat{\tau}\exp\left(-\frac{1}{2}\left(\frac{y_t-\hat{y}_t}{s_t}\right)^2\right) - (y_t-\hat{y}_t)\left(1-2\Phi\left(\frac{y_t-\hat{y}_t}{s_t}\right)\right)\right]^{\frac{1-\rho}{1+\rho}}\times \\ &\times\left[-\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{2}\left(\frac{y_t-\hat{y}_t}{s_t}\right)^2\right)\frac{y_t-\hat{y}_t}{s_t} - \right. \\ &\left. -\left(1-2\Phi\left(\frac{y_t-\hat{y}_t}{s_t}\right)\right) + 2\Phi'\left(\frac{y_t-\hat{y}_t}{s_t}\right)(y_t-\hat{y}_t)\right] \end{aligned}$$

and therefore

$$\frac{\partial p(y_t|Y_{t-1})}{\partial y_t} = p(y_t|Y_{t-1})\frac{1}{\varphi^{\frac{2}{1+\rho}}}\frac{1}{1+\rho}\left(2\Phi\left(\frac{y_t-\hat{y}_t}{s_t}\right)-1\right)(\mathrm{E}\,|Z|)^{\frac{1-\rho}{1+\rho}}.$$

Finally

$$g_A(y_t) = -\frac{\partial p(y_t|Y_{t-1})/\partial y_t}{p(y_t\mid Y_{t-1})} = \frac{1}{\varphi^{\frac{2}{1+\rho}}}\frac{1}{1+\rho}\left(2\Phi\left(\frac{y_t-\hat{y}_t}{s_t}\right)-1\right)(\mathrm{E}\,|Z|)^{\frac{1-\rho}{1+\rho}}$$

is an approximation of the function $g(y_t)$. Now this approximation has to be adjusted so that for non-outlying observations the modified filter corresponds to the Kalman filter. Define

$$g_A^*(y_t) = g_A(y_t) - \left(g_A(\operatorname{sgn}(y_t - \hat{y}_t)\, K\, \hat{d}_t) - \frac{\operatorname{sgn}(y_t - \hat{y}_t)\, K}{\hat{d}_t} \right) e^{-(|y_t - \hat{y}_t| - K\hat{d}_t)}$$

or

$$g_A^*(y_t) = g_A(y_t) \left(\frac{\operatorname{sgn}(y_t - \hat{y}_t)\, K}{\hat{d}_t} \frac{1}{g_A(\operatorname{sgn}(y_t - \hat{y}_t)\, K\, \hat{d}_t)} - 1 \right) \left(1 + e^{-(|y_t - \hat{y}_t| - K\hat{d}_t)} \right),$$

where K is the appropriate quantile (see formula (11) or (16)). The updating function may now be defined as

$$g_{L2}(I_t) = \begin{cases} I/\hat{d}_t^2 & \text{if } I_t \in \langle 0; K\hat{d}_t \rangle \\ \min\left\{ g_A^*(y_t), I_t/\hat{d}_t^2 \right\} & \text{if } I_t > K\hat{d}_t \\ -g_{L2}(-I_t) & \text{if } I_t < 0. \end{cases} \tag{23}$$

Example 4.2. The data $\boldsymbol{Y_{AO}}$ were processed using the Kalman filter and the modified filter with outlier detection and processing using the system $\mathcal{L}_2$ ($\alpha = 0.005$). Figure 3 shows the result. The modified filter detected outliers at times 25, 51 and 75. The impact of outliers in the case of the modified filter was less destructive than in the case of the Kalman filter and was fully absorbed after two time periods. The filter returned a final estimate $\hat{\sigma}_{100}^2 = 2.964$, again the standard deviation of the filtered state converged to the right value quite fast.

The three respective updating functions are displayed in Figure 4. The impact of outliers is bound for all three updating functions. New updating functions, however, are not so strictly rejecting moderately outlying observations. This corresponds to uncertainty about observations that are close to a non-outliers region. On the other hand, observations that are very far from the expected value are treated more strictly by the new updating functions than by the Huber function.

5. IMPROVED SMOOTHING AND PARAMETER ESTIMATION

5.1. Kalman smoother

Having the observation history Y_T we may be interested in looking for the estimates of unknown states at time t, $t = 0, 1, \ldots, T$. This problem is called *(fixed-interval) smoothing*. In the environment of the state-space models the smoothed estimates may be obtained by using the following backward recursions:

$$\begin{aligned} \hat{\boldsymbol{x}}_{t|T} &= \hat{\boldsymbol{x}}_{t|t} + \boldsymbol{P}_t^* (\hat{\boldsymbol{x}}_{t+1|T} - \boldsymbol{F}\hat{\boldsymbol{x}}_{t|t}), \\ \boldsymbol{P}_{t|T} &= \boldsymbol{P}_{t|t} - \boldsymbol{P}_t^* (\boldsymbol{P}_{t+1|T} - \boldsymbol{P}_{t+1|t}) \boldsymbol{P}_t^{*\prime}, \\ \boldsymbol{P}_t^* &= \boldsymbol{P}_{t|t} \boldsymbol{F}' \boldsymbol{P}_{t+1|t}^{-1} \end{aligned} \tag{24}$$

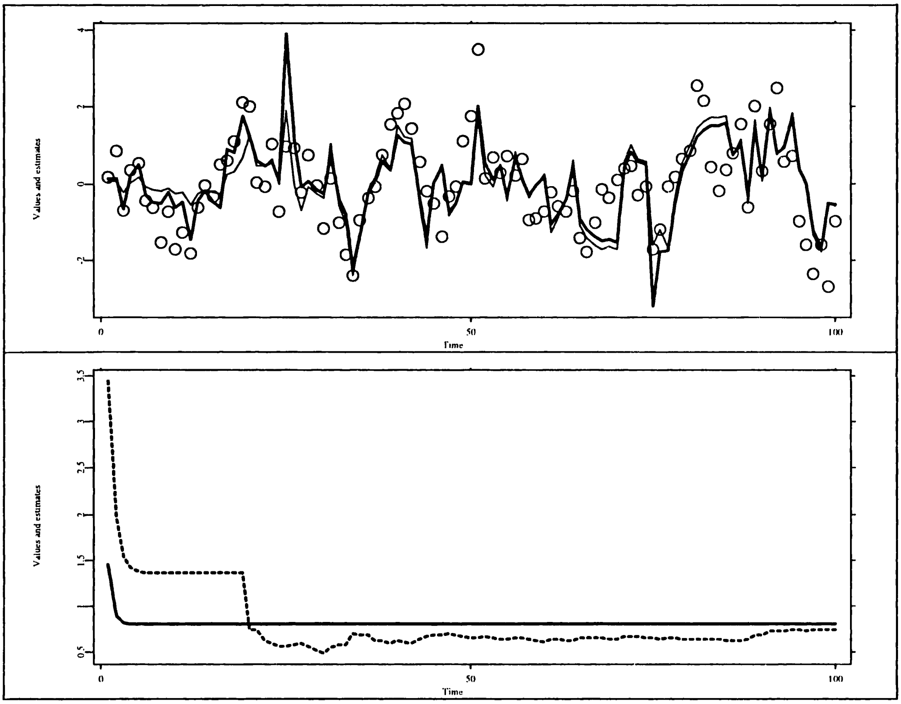

Fig. 3. True (circles) and filtered state obtained from the Kalman filter (thick) and the modified filter (thin) using the data set Y_{AO}.

defined for $t = T-1, T-2, \dots, 0$. Hence, the filtered estimates are the basis of the smoothing algorithm. As may be expected, the smoothed estimates are also affected by outlying observations. We may study an impact of a single outlier y_t (continuing from Section 3.2) as follows.

5.2. Kalman smoothing performance in a presence of outlying observations

Suppose now that y_t is the only outlier and suppose further that $\boldsymbol{A}_n^t \to 0$ for $n \to \infty$. If the convergence does not hold, the impact of a single outlier at time t affects the performance of the filter and smoothing in this case is not reasonable – in such situation remodeling the state-space equations or transforming the observations is advisable. Denote $s = t+u$ the last time when the impact matrix $\boldsymbol{A}_u^t$ is treated as a non-zero matrix, i. e. $u = \min\left\{k; ||\boldsymbol{A}_k^t||^2 < \varepsilon\right\}$. Till this time the filtered estimates

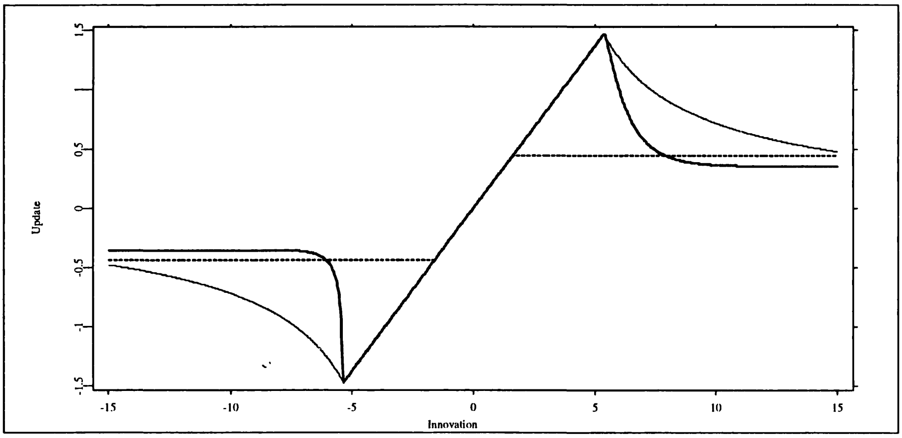

Fig. 4. Huber (dotted), $g_{L1}(I_t)$ (thin) and $g_{L2}(I_t)$ (thick) at time 25.

contain a measurable error (in the sense of the modeled time series). Then we get

$$\begin{aligned}\hat{x}_{s|T} &= \hat{x}_{s|s} + P_s^* \left(\hat{x}_{s+1|T} - F\,\hat{x}_{s|s}\right)\\ &= \hat{x}_{s|s}^R + \Delta\hat{x}_{s|s} + P_s^* \left(\hat{x}_{s+1|T} - F\,\hat{x}_{s|s}^R - F\,\Delta\hat{x}_{s|s}\right)\\ &= \hat{x}_{s|T}^R + (I - P_s^*\,F)\,\Delta\hat{x}_{s|s}\\ &= \hat{x}_{s|T}^R + (I - P_s^*\,F)\,A_u^t\,\Delta\hat{x}_{t|t}\\ &= \hat{x}_{s|T}^R + \Delta\hat{x}_{s|T}. \end{aligned}\tag{25}$$

Denote

$$B_u^t = (I - P_{t+u}^*\,F)\,A_u^t, \qquad i = 1,\dots,u-1. \tag{26}$$

Next

$$\begin{aligned}\hat{x}_{s-1|T} &= \hat{x}_{s-1|s-1} + P_{s-1}^*(\hat{x}_{s|T} - F\,\hat{x}_{s-1|s-1})\\ &= \hat{x}_{s-1|s-1}^R + \Delta\hat{x}_{s-1|s-1} + P_{s-1}^*(\hat{x}_{s|T}^R + \Delta\hat{x}_{s|T} - F\,\hat{x}_{s-1|s-1}^R - F\,\Delta\hat{x}_{s-1|s-1})\\ &= \hat{x}_{s-1|T}^R + \Delta\hat{x}_{s-1|s-1} + P_{s-1}^*(\Delta\hat{x}_{s|T} - F\,\Delta\hat{x}_{s-1|s-1})\\ &= \hat{x}_{s-1|T}^R + A_{u-1}^t\,\Delta\hat{x}_{t|t} + P_{s-1}^*\,(I - P_s^*\,F)\;A_u^t\Delta\hat{x}_{t|t} - P_{s-1}^*\,F\,A_{u-1}^t\,\Delta\hat{x}_{t|t}\\ &= \hat{x}_{s-1|T}^R + \left(A_{u-1}^t + P_{s-1}^*\,A_u^t + P_{s-1}^*\,P_s^*\,F\,A_u^t - P_{s-1}^*\,F\,A_{u-1}^t\right)\Delta\hat{x}_{t|t}\\ &= \hat{x}_{s-1|T}^R + \Delta\hat{x}_{s-1|T}. \end{aligned}\tag{27}$$

Similarly we get

$$\hat{x}_{s-j|T} = \hat{x}_{s-j|T}^R + \Delta\hat{x}_{s-j|T}, \qquad j = 1,\dots,u, \tag{28}$$

where

$$\Delta\hat{\boldsymbol{x}}_{s-j|T} = \boldsymbol{B}^t_{u-j}\Delta\hat{\boldsymbol{x}}_{t|t}$$

and the impact matrix $\boldsymbol{B}^t_{u-j}$ is given by the following backward recursions:

$$\begin{aligned}
\boldsymbol{B}^t_u &= (\boldsymbol{I} - \boldsymbol{P}^*_{t+u}\boldsymbol{F})\boldsymbol{A}^t_u \\
\boldsymbol{B}^t_{u-1} &= (\boldsymbol{I} - \boldsymbol{P}^*_{t+u-1}\boldsymbol{F})\boldsymbol{A}^t_{u-1} + \boldsymbol{P}^*_{t+u-1}\boldsymbol{B}^t_u \\
\boldsymbol{B}^t_{u-2} &= (\boldsymbol{I} - \boldsymbol{P}^*_{t+u-2}\boldsymbol{F})\boldsymbol{A}^t_{u-2} + \boldsymbol{P}^*_{t+u-1}\boldsymbol{B}^t_{u-1} \\
&\vdots \\
\boldsymbol{B}^t_1 &= (\boldsymbol{I} - \boldsymbol{P}^*_{t+1}\boldsymbol{F})\boldsymbol{A}^t_1 + \boldsymbol{P}^*_{t+1}\boldsymbol{B}^t_2 \\
\boldsymbol{B}^t_0 &= \boldsymbol{A}^t_0 + \boldsymbol{P}^*_t\boldsymbol{B}^t_1,
\end{aligned}$$

i. e. the impact matrix $\boldsymbol{B}^t_{u-j}$ evolves as

$$\boldsymbol{B}^t_u = (\boldsymbol{I} - \boldsymbol{P}^*_{t+u-j}\boldsymbol{F})\boldsymbol{A}^t_{u-j} + \boldsymbol{P}^*_{t+u-j}\boldsymbol{B}^t_{u-j+1}, \qquad j = 1, \ldots, u-1. \tag{29}$$

Prior time t the impact of the outlier y_t on the smoothed estimate may be expressed as follows:

$$\begin{aligned}
\hat{\boldsymbol{x}}_{t-1|T} &= \hat{\boldsymbol{x}}_{t-1|t-1} + \boldsymbol{P}^*_{t-1}(\hat{\boldsymbol{x}}^R_{t|T} + \Delta\hat{\boldsymbol{x}}_{t|T} - \boldsymbol{F}\hat{\boldsymbol{x}}^R_{t-1|t-1}) \\
&= \hat{\boldsymbol{x}}^R_{t-1|T} + \boldsymbol{P}^*_{t-1}\boldsymbol{B}^t_0\Delta\hat{\boldsymbol{x}}_{t|t}
\end{aligned} \tag{30}$$

and hence the impact matrix prior time t evolves as follows:

$$\begin{aligned}
\boldsymbol{B}^t_{-1} &= \boldsymbol{P}^*_{t-1}\boldsymbol{B}^t_0 \\
\boldsymbol{B}^t_{-2} &= \boldsymbol{P}^*_{t-2}\boldsymbol{B}^t_{-1} = \boldsymbol{P}^*_{t-2}\boldsymbol{P}^*_{t-1}\boldsymbol{B}^t_0 \\
&\vdots
\end{aligned} \tag{31}$$

5.3. Improved smoothing

Hence, the outlying observation affects smoothed estimates prior the time of its occurrence. Again, a magnitude of this impact depends on the filtering error $\Delta\hat{\boldsymbol{x}}_{t|t}$ and the impact matrices $\boldsymbol{B}^t_j$. These may be calculated for a given state-space model before any data are observed. Robust modification of the smoothing recursions may be obtained when using an output of the modified filter instead of the output of the Kalman filter.

5.4. EM-algorithm

So far all parameters of the state-space model except σ^2 were assumed to be known. In practical applications, however, the starting state mean $\tilde{x}_0$ is not known and the matrices $\boldsymbol{F}$ and $\boldsymbol{R}$ may contain unknown elements that have to be estimated. The EM-algorithm proved to be an efficient tool for completing this task.

Assume $(\hat{\boldsymbol{x}}_{0,r}, \boldsymbol{F}_r, \sigma_r^2, \boldsymbol{R}_r)$ is a set of estimates obtained in the rth step of the algorithm. Assuming $\mathcal{L}(y_j|\boldsymbol{x}_j) \sim N(\boldsymbol{h}\boldsymbol{x}_j, \sigma^2)$ for $j = 1, \ldots, T$ we may write the log-likelihood of the history Y_T as (up to a constant)

$$\ln L(Y_T) = -\frac{T}{2} \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum_{j=1}^{T} (y_j - \boldsymbol{h}\boldsymbol{x}_j)^2. \tag{32}$$

In the E-step of the algorithm the conditional expectation

$$G(\hat{\boldsymbol{x}}_{0,r}, \boldsymbol{F}_r, \sigma_r^2, \boldsymbol{R}_r) = \mathrm{E}\,(\ln L(Y_T)|Y_T)$$

is found. In our case it is easy to get

$$G(\hat{\boldsymbol{x}}_{0,r}, \boldsymbol{F}_r, \sigma_r^2, \boldsymbol{R}_r) = -\frac{T}{2} \ln \sigma_r^2 - \frac{1}{2\,\sigma_r^2} \sum_{j=1}^{T} \left[(y_j - \boldsymbol{h}\hat{\boldsymbol{x}}_{j|T})^2 + \boldsymbol{h}\boldsymbol{P}_{j|T}\boldsymbol{h}' \right]. \tag{33}$$

In the M-step of the algorithm the expression (33) is maximized according to the unknown parameters. Using some matrix calculus it is not too complicated to solve the likelihood equation and to express the updated estimates as

$$\begin{aligned} \boldsymbol{F}_{r+1} &= \boldsymbol{B}\boldsymbol{A}^{-1}, \\ \boldsymbol{R}_{r+1} &= T^{-1}(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}'), \\ \hat{\sigma}_{r+1}^2 &= T^{-1} \sum_{t=1}^{T} [(y_t - \boldsymbol{h}\hat{\boldsymbol{x}}_{t|T})^2 + \boldsymbol{h}\boldsymbol{P}_{t|T}\boldsymbol{h}'], \\ \hat{\boldsymbol{x}}_{0,r+1} &= \hat{\boldsymbol{x}}_{0|T}, \end{aligned} \tag{34}$$

where

$$\begin{aligned} \boldsymbol{A} &= \sum_{t=1}^{T} (\boldsymbol{P}_{t-1|T} + \hat{\boldsymbol{x}}_{t-1|T}\hat{\boldsymbol{x}}'_{t-1|T}), \quad \boldsymbol{B} = \sum_{t=1}^{T} (\boldsymbol{P}_{t,t-1|T} + \hat{\boldsymbol{x}}_{t|T}\hat{\boldsymbol{x}}'_{t-1|T}), \\ \boldsymbol{C} &= \sum_{t=1}^{T} (\boldsymbol{P}_{t|T} + \hat{\boldsymbol{x}}_{t|T}\hat{\boldsymbol{x}}'_{t|T}) \end{aligned}$$

and the matrix $\boldsymbol{P}_{t,t-1|T}$ may be obtained recursively as

$$\begin{aligned} \boldsymbol{P}_{T,T-1|T} &= (\boldsymbol{I} - \boldsymbol{P}_{T|T-1}\boldsymbol{h}d_T^{-2})\boldsymbol{F}\boldsymbol{P}_{T|T}, \\ \boldsymbol{P}_{t-1,t-2|T} &= \boldsymbol{P}_{t|t}\boldsymbol{P}^{*\prime}_{t-2} + \\ & \quad \boldsymbol{P}^{*}_{t-1}(\boldsymbol{P}_{t,t-1|T} - \boldsymbol{F}\boldsymbol{P}_{t-1|t-1})\boldsymbol{P}^{*\prime}_{t-2}, \quad t = T, T-1, \ldots, 2. \end{aligned}$$

Now these estimates may be used as an initial value for the new cycle of the EM-algorithm until the algorithm converges to some stable solution. The EM estimates may be modified to incorporate some structural assumptions about elements of $\boldsymbol{F}$ and $\boldsymbol{R}$ (refer to [9] for more details).

Table 1. Results of the EM-algorithm and its less sensitive modification.

	$\hat{x}_0$	$\hat{F}$	$\hat{R}$	$\hat{\sigma}^2$
Ordinary EM on data $\boldsymbol{Y}$	0.1311	0.5490	1.0573	1.9231
Ordinary EM on data $\boldsymbol{Y_{AO}}$	0.1267	0.4978	1.1790	3.3953
Improved EM on data $\boldsymbol{Y}$	0.2068	0.5042	1.2868	1.5752
Improved EM on data $\boldsymbol{Y_{AO}}$	0.3259	0.4064	1.7067	1.7223

5.5. Improved EM algorithm

If there are outlying observations in the data, the EM estimates of the state-space model parameters tend to overestimate σ^2 and underestimate the matrix $\boldsymbol{F}$. Having less outlier-sensitive smoothing algorithm, its output may be used to get improved EM estimates. However, minor changes are necessary in this case. In the case of the contaminated data the modified filter bounds the impact of the outlying observations on $\hat{\boldsymbol{x}}_{t|t}$ and $\hat{\boldsymbol{x}}_{t|T}$. Thus the terms $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are correct (i. e. close to the reality) but the smoothing errors $y_t - \boldsymbol{h}\hat{\boldsymbol{x}}_{t|T}$ will be too large if y_t was detected as an additive outlier. In the modified EM-algorithm it is thus necessary to replace the third term of (34) by the term

$$\hat{\sigma}^2_{r+1} = T^{-1}\sum_{t=1}^{T}[\varphi(y_t - \boldsymbol{h}\hat{\boldsymbol{x}}_{t|T}) + \boldsymbol{h}\boldsymbol{P}_{t|T}\boldsymbol{h}'],$$

where $\varphi(\cdot)$ is some bounded function.

Example 5.1. The ordinary and improved EM-algorithm were run on data sets $\boldsymbol{Y}$ and $\boldsymbol{Y_{AO}}$ with the following starting values: $\boldsymbol{x}_0 = 0$, $\boldsymbol{F} = -0.1$, $\boldsymbol{R} = 10$ and $\sigma^2 = 10$. Results of 400 iterations of each algorithm are summarized in Table 1.

Note that the ordinary EM algorithm estimated much higher observation variance in the case of the data set $\boldsymbol{Y_{AO}}$. Results of the improved EM algorithm run on the data set $\boldsymbol{Y_{AO}}$ are comparable to the results of the ordinary EM algorithm run on the data set $\boldsymbol{Y}$. The improved EM algorithm returned a better estimate of the observation variance even in the case of contaminated data. Please note that the EM-algorithm may be easily generalized for multivariate state $\boldsymbol{x}$ and $\boldsymbol{F}$ with linear constrains.

6. EXAMPLE OF ANALYSIS OF REAL-LIFE DATA

Average monthly prices (multiples of 10000 pesetas/100 kg) of a lamb meat in Spain observed in a time period 1985–1988 were processed using the Kalman smoother and the modified smoothing algorithm. The data were taken from the article [6]. A state-space representation of a seasonal time series with 12 seasons was used with

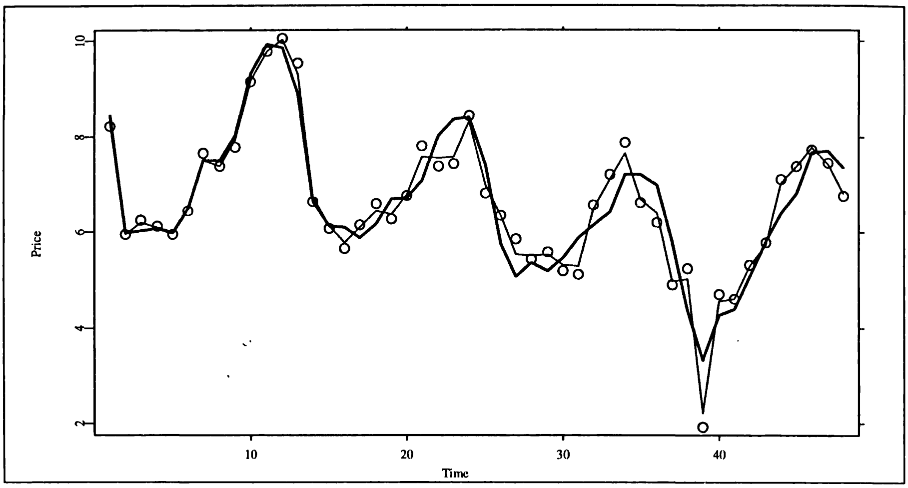

Fig. 5. Smoothed monthly prices of a lamb meat in Spain (1985–1988). Estimates obtained by the Kalman smoother are displayed by using the thin line, estimates obtained by the modified smoother are displayed by using the thick line.

initial values of the state-space model parameters proposed in the mentioned article. The result is displayed in Figure 5. The impact of the outlying observation number 39 is less serious in the case of the modified smoother.

Note 6.1. The algorithms described in this paper were implemented in the statistical environment XploRe.

ACKNOWLEDGEMENT

The work on this paper was supported by the Grant Agency of the Czech Republic under grant No. 201/00/0770.

(Received August 31, 2001.)

REFERENCES

[1] B. D. O. Anderson and J. B. Moore: Optimal Filtering. Prentice–Hall, Englewood Cliffs, N. J. 1979.
[2] G. E. P. Box and G. C. Tiao: Bayesian Inference in Statistical Analysis. Addison–Wesley, London 1973.
[3] N. Cantarelis and F. R. Johnston: On-line variance estimation for the steady state Bayesian forecasting model. J. Time Ser. Analysis *3* (1982), 225–234.
[4] T. Cipra: Some modifications of recursive time series methods. In: ROBUST '96 – Collection of Discussion Papers, Union of the Czech Mathematicians and Physicists, Praha 1997.

[5] T. Cipra and A. Rubio: Kalman filter with a non-linear non-gaussian observation relation. Trabajos de Estadistica *6* (1991), 111–119.
[6] T. Cipra, A. Rubio, and J. L. Canal: Robust smoothing and forecasting procedures. Central European J. Oper. Research *1* (1992), 41–56.
[7] J. Durbin and S. J. Koopman: Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. J. Roy. Statist. Soc. *62* (2000), 3–56.
[8] P. Franěk: Kalman Filter. KPMS Seminair Works Series, MFF UK, Praha 1999.
[9] J. R. M. Hosking, J. S. Pai, and L. S. Y. Wu: An algorithm for estimating parameters of state-space models. Statist. Probab. Letters *28* (1996), 99–106.
[10] A. H. Jazwinski: Stochastic Processes and Filtering Theory. Academic Press, New York 1972.
[11] R. E. Kalman: A new approach to linear filtering and prediction problems. Trans. Amer. Soc. Mech. Eng. – J. Basic Eng. *82* (1960), 35–45.
[12] G. Kitagawa: Non-Gaussian state-space modelling of nonstationary time series. J. Amer. Statist. Assoc. *82* (1987), 1032–1050.
[13] G. Kitagawa: Self-organizing state-space model. J. Amer. Statist. Assoc. *93* (1998), 1203-1215.
[14] H. R. Künsch: State space and hidden Markov models. In: Complex Stochastic Systems (O. E. Barndorf-Nielsen, D. R. Cox, and C. Klüppelberg, eds.), Chapman & Hall / CRC, Boca Raton 2001, pp. 109–173.
[15] C. J. Masreliez: Approximate non-Gaussian filtering with linear state and observation relations. IEEE Trans. Automat. Control *AC-20* (1975), 107–110.
[16] C. J. Masreliez and R. D. Martin: Robust Bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Automat. Control *AC-22* (1975), 361–371.
[17] R. J. Meinhold and N. Z. Singpurwalla: Robustification of the Kalman filter. J. Amer. Statist. Assoc. *84* (1989), 479–486.
[18] P. Ruckdeschel: Ansätze zur Robustifizierung des Kalman-filters. Doctoral Thesis, Bayreuth University, Bayreuth 2001.
[19] H. Tanizaki: Nonlinear and non-Gaussian state estimation: A quasi-optimal estimator. Comm. Statist. – Theory Methods *29* (2000), 2805–2834.

Mgr. Petr Franěk, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics – Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: petr.franek@cnb.cz