Sociedad de Estadística e Investigación Operativa
Test (1997) *Vol.* **6**, *No.* 2, pp. 379–395

Kalman filter with outliers and missing observations

T. Cipra
Department of Statistics, Charles University of Prague
Sokolovska 83; 186 00 Prague 8, Czech Republic
R. Romera
Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Madrid 126, 28903 Getafe, Spain.

Abstract

The discrete Kalman filter which enables the treatment of incomplete data and outliers is described. The incomplete or missing observations are included in such a way as to transform the Kalman filter to the case when observations have changing dimensions. In order to treat outliers, the Kalman filter is made robust using the M-estimation principle. Some special cases are considered including a convergence result for recursive parameter estimation in an AR(1) process with innovation outliers and missing observations.

Key Words: Kalman filter, outliers, missing observations, time series analysis.

AMS subject classification: 93E20, 62F35

1 Introduction

Kalman filter and state space model formulation together provide a very powerful tool for the recursive treatment of dynamic systems (see Anderson and Moore (1979)). Many standard statistical examples arise from estimation, interpolation and prediction procedures for autoregressive time series (see, e.g., Akaike (1978), Gardner, Harvey and Phillips (1980), and Jones (1980)).

Numerous real data examples show that there are suitable modifications of the Kalman filter which allow the treatment of incomplete data and data with outliers. In the case of incomplete data for example one may observe only some components of vector data at any given time, or only an aggregate form of the components is observable, or some observations may be completely missing. When outliers appear in the observed data it would be desirable to work with a robust version of the Kalman filter

Research partially supported by CICYT grant number TIC93-0702-C02-02.
Received: November 1993; Accepted: September 1996

which is insensitive to such outliers. Early approaches to robust Kalman filtering were considered by Kleiner, Martin and Thomson (1979), Masreliez (1975), Masreliez and Martin (1977), Morris (1976) and Servi and Ho (1981). Robustness based on M-Estimation methodology (see Huber (1964, 1981)) was then developed, and a Bayesian approach to the problem is discussed in West (1981), West, Harrison and Migon (1985), Guttman and Peña (1985) and Meinhold and Singpurwalla (1989).

The above mentioned Kalman filter modifications can be used to derive recursive formulae for estimating, smoothing and predicting time series with incomplete (or missing) and contaminated observations as in Kohn and Ansley (1985, 1986), Jong (1989) and Cipra (1992). See also Cipra, Rubio and Trujillo (1991) and Cipra, Trujillo and Rubio (1995).

For applications in time series analysis, we shall consider the discrete form of the Kalman filter in the dynamic linear model given by

$$x_t = F_t x_{t-1} + w_t \tag{1.1}$$

$$y_t = H_t x_t + v_t \tag{1.2}$$

(non-linear generalizations are also possible, see, e.g., Cipra and Rubio (1991)), where

$$E(w_t) = 0,\ E(v_t) = 0,\ E(w_s, w_t') = \delta_{st} Q_t,\ E(v_s v_t') = \delta_{st} R_t,\ E(w_s v_t') = 0 \tag{1.3}$$

and certain initial conditions are satisfied. The state equation (1.1) describes the behavior of an n-dimensional state vector x_t with time while the observation equation (1.2) describes how the state is related to an m-dimensional observation vector y_t. The matrices $F_t(n \times n)$, $H_t(m \times n)$, $Q_t(n \times n)$ and $R_t(m \times m)$ are assumed known at time t.

From a practical point of view, it is important that the Kalman filter provides recursive formulae for construction of the linear minimum variance estimator $\widehat{x}_t^t$ of the state x_t and for its error covariance matrix $P_t^t = E[x_t - \widehat{x}_t^t)(x_1 - \widehat{x}_t^t)']$ at time t using all past information $\{y_0, y_1, \ldots, y_t\}$

$$\widehat{x}_t^t = \widehat{x}_t^{t-1} + P_t^{t-1} H_t' (H_t P_t^{t-1} H_t' + R_t)^{-1} (y_t - H_t \widehat{x}_t^{t-1}) \tag{1.4}$$

$$P_t^t = P_t^{t-1} - P_t^{t-1} H_t (H_t P_t^{t-1} H_t + R_t)^{-1} H_t P_t^{t-1}. \tag{1.5}$$

Here, $\widehat{x}_t^{t-1}$ and P_t^{t-1} are predictive values constructed for time t at time $t-1$:

$$\widehat{x}_t^{t-1} = F_t \widehat{x}_{t-1}^{t-1} \tag{1.6}$$

$$P_t^{t-1} = F_t P_{t-1}^{t-1} F_t' + Q_t \tag{1.7}$$

Section 2 of the paper deals with the Kalman filter for incomplete data. The main contribution of the paper appears in Section 3 where the problem of incomplete (or missing) observations and outliers in the context of the Kalman filter is studied. Some special cases are described in Section 4 including a convergence result for recursive parameter estimation in an AR(1) process with outliers and missing observations.

2 Incomplete data

The following examples of incomplete data seem to occur frequently in practice:

(i) Only those components of y_t of order $i_1(t), \ldots, i_{m_t}$ $(1 \leq i_1(t) \leq \ldots \leq i_{m_t} \leq m)$ are available at time t: then, the observed vector at time t is then the m_t-dimensional vector $y_t^* = S_t y_t$, where S_t is the $(m_t \times m)$ matrix with ones at positions $(1, i_1(t)), \ldots, (m_t, i_{m_t}(t))$ and zeroes otherwise.

(ii) Only the aggregate value $y_{1t} + \ldots + y_{m_t}$ of all components of y_t is known at time t: the actual vector observed at time t is thus the scalar $y_t^* = S_t y_t$, where S_t is the unit m-dimensional vector.

The above two examples are special cases which occur when the dynamic linear model (1.1)-(1.3) is transformed to

$$x_t = F_t x_{t-1} + w_t \tag{2.1}$$

$$y_t^* = H_t^* x_t + v_t^* \tag{2.2}$$

where $y_t^* = S_t y_t$, $H_t^* = S_t H_t$, $v_t^* = S_t v_t$ (in particular, if $R_t^* = E(v_t^* v_t^{*\prime}) = S_t R_t S_t'$ and $S_t(m_t \times m)$ is a matrix of full row rank). In contrast to the original model (1.1)-(1.3) the dimension of the observation vector y_t^* in the new dynamic linear model (2.1)-(2.2) can change with time.

(iii) There is no new information at time t, i.e., the observation vector y_t is completely missing.

The problem of the Kalman filter in the presence of incomplete data as formulated in the examples above is solved using the following theorem which allows variable dimensions in the model (1.1)-(1.3). Although the variable dimension m_t of y_t is sufficient to deal with (2.1)-(2.2), Theorem 2.1 also allows the dimension n_t of x_t to vary.

Theorem 2.1. *Consider the model (1.1)-(1.3), where the dimensions n_t of the state vector x_t and m_t of the observation vector y_t can vary with time. Then, the formulae (1.4)-(1.7) remain valid. If the observation vector y_t is missing at time t, then (1.4) and (1.5) must be replaced by*

$$\widehat{x}_t^t = \widehat{x}^{t-1} \tag{2.3}$$

$$P_t^t = P_t^{t-1} \tag{2.4}$$

Proof. The proof is omitted here since it requires only a slight modification of the proof for the classical case with constant dimensions given, for example, in Jazwinski (1970). For further discussion of Equations (2.3) and (2.4), see also Jones (1980). It is based on the Projection Theorem in the Hilbert space of random variables. □

3 Incomplete data and outliers

In order to treat observed outliers, a relevant modified Kalman Filter has been developed. In practical applications, this can be modelled simultaneously with the case of incomplete data in such a way that one uses the model (2.1)-(2.2) with the additional assumptions

$$w_t \sim N(0, Q_t), \quad v_t \sim \varepsilon - \text{contaminated } N(0, R_t). \tag{3.1}$$

The ε-contamination in (3.1) indicates that the normal distribution, adopted as standard in the classical Kalman filter, is contaminated by a small fraction ε (e.g., $\varepsilon = 0.05$) of a symmetric distribution with heavy tails which is the source of the outliers in $\{y_t^*\}$.

It is straightforward to show that the state estimator $\widehat{x}_t^t$ in the dynamic linear model (2.1)-(2.2) corresponding to (1.4) can be obtained by minimizing

$$\widehat{x}_t^t = \operatorname{argmin}\left\{(\widehat{x}_t^{t-1} - x_t)' \left(P_t^{t-1}\right)^{-1} (\widehat{x}_t^{t-1} - x_t) + (y_t^* - H_t^* x_t)' R_t^{*-1} (y_t^* - H_t^* x_t)\right\} \tag{3.2}$$

with respect to $x_t \in R^n$ (see also Cipra and Romera (1991)), or equivalently

$$\widehat{x}_t^t = \operatorname{argmin}\left\{\sum_{i=1}^{n} (p_{it} - a_{it} x_t)^2 + \sum_{j=1}^{m_t} (s_{jt}^* - b_{jt}^* x_t)^2\right\} \tag{3.3}$$

where $p_t = (P_t^{t-1})^{-1/2}\hat{x}_t^{t-1}$, $s_t^* = (R_t^*)^{-1/2}y_t^*$, $a_t = (P_t^{t-1})^{-1/2}$ and $b_t^* = (R_t^*)^{-1/2}H_t^*$ (e.g., p_{it} is the i-th component of the vector $p_t(n \times 1)$ and $a_{it}(1 \times n)$ is the i-th row of the matrix $a_t(n \times n)$).

From among numerous methods proposed in the literature, the robust modification of the Kalman filter based on the methodology of M-estimation seems to give results which are numerically acceptable (see Cipra and Romera (1991)). An algorithm suitable for parallel computation of Kalman filtering with contaminated observations based on the square root version of the Kalman filter was introduced by Romera and Cipra (1995). In the case of M-estimation, the least squares solution (3.3) is replaced by

$$\hat{x}_t^t = \operatorname{argmin}\left\{\sum_{i=1}^{n}(p_{it} - a_{it}x_t)^2 + \sum_{j=1}^{m_t}\rho_j(s_{jt}^* - b_{jt}^*x_t)^2\right\} \tag{3.4}$$

where the ρ_j are suitable loss functions with derivatives ψ_j (the so-called psi-functions) used in robust statistics. For example, Huber's psi-function given by

$$\psi_H(z) = \begin{cases} z & \text{for } |z| \le c \\ c\ \operatorname{sgn}(z) & \text{for } |z| > c \end{cases} \tag{3.5}$$

is often used since it gives robust estimates of location which are optimal in the min-max sense, having minimum variance over the least favorable contaminating distributions. The recommended choice of c in (3.5) is $c = u_{1-\varepsilon}$, where u_α is the α-quantile of $N(0,1)$ (e.g., $c = 1.645$ for a 5% contamination of data). If the observation vector y_t is missing then (2.3) and (2.4) must be adopted.

The normal equations for $\hat{x}_t^t$ corresponding to (3.4) have the form

$$\sum_{i=1}^{n} a_{it}'(p_{it} - a_{it}\hat{x}_t^t) + \sum_{j=1}^{m_t} b_{jt}^{*\prime}\psi_j(s_{jt}^* - b_{jt}^*\hat{x}_t^t) = 0 \tag{3.6}$$

and can be solved explicitly only in some special cases (see Section 4). Alternatively, one can use the following approximate normal equations

$$\sum_{i=1}^{n} a_{it}'(p_{it} - a_{it}\hat{x}_t^t) + \sum_{j=1}^{m_t} w_{jt}b_{jt}^{*\prime}(s_{jt}^* - b_{jt}^*\hat{x}_t^t) = 0 \tag{3.7}$$

where the weights w_{jt}, $j = 1, \ldots, m_t$ are defined by

$$w_{jt} = \frac{\psi_j(s^*_{jt} - b^*_{jt}\widehat{x}_t^{t-1})}{s^*_{jt} - b^*_{jt}\widehat{x}_t^{t-1}} \tag{3.8}$$

Equations (3.7) follow from (3.6) if we approximate $\widehat{x}_t^t$ by $\widehat{x}_t^{t-1}$. Using (3.7) and some simple algebra, we obtain the following recursive formulae

$$\widehat{x}_t^t = \widehat{x}_t^{t-1} + P_t^{t-1}H_t^{*\prime}[H_t^*P_t^{t-1}H_t^{*\prime} + R_t^{*1/2}W_tR_t^{*1/2}]^{-1}(y_t^* - H_t^*\widehat{x}_t^{t-1}) \tag{3.9}$$

$$P_t^t = P_t^{t-1} + P_t^{t-1}H_t^{*\prime}[H_t^*P_t^{t-1}H_t^{*\prime} + R_t^{*1/2}W_tR_t^{*1/2}]^{-1}H_t^*P_t^{t-1} \tag{3.10}$$

where $W_t = \mathrm{diag}\{w_{1t}, \ldots, w_{m_t t}\}$ and $\widehat{x}_t^{t-1}$ and P_t^{t-1} are given in (1.6) and (1.7) respectively.

These formulae (or (2.3) and (2.4) if the observation vector y_t is missing at time t) represent the approximate solution for the Kalman filter problem in the presence of incomplete data and outliers and this is suitable for application in practice.

4 Some special cases

Let us now consider some special cases of the previous results.

4.1 Kalman filter with scalar observations

Let $m = 1$ and write h_t and r_t in place of H_t and R_t in (1.2), respectively, where h_t is an n-dimensional row vector and r_t is a scalar. For the incomplete data problem,it makes sense to consider only the case of missing observations.

From (2.3), (2.4), (3.9) and (3.10), we obtain

$$\widehat{x}_t^t = \begin{cases} \widehat{x}_t^{t-1} + \frac{P_t^{t-1}h_t'}{h_tP_t^{t-1}h_t' + r_t/w_t}(y_t - h_t\widehat{x}_t^{t-1}) & \text{if } y_t \text{ is observed} \\ \widehat{x}_t^{t-1} & \text{if } y_t \text{ is missing} \end{cases} \tag{4.1}$$

$$P_t^t = \begin{cases} P_t^{t-1} - \frac{P_t^{t-1}h_t'h_tP_t^{t-1}}{h_tP_t^{t-1}h_t' + r_t/w_t} & \text{if } y_t \text{ is observed} \\ P_t^{t-1} & \text{if } y_t \text{ is missing} \end{cases} \tag{4.2}$$

where $\widehat{x}_t^{t-1}$ and P_t^{t-1} are given in (1.6) and (1.7) and

$$w_t = \frac{\psi(r_t^{-1/2}(y_t - h_t\widehat{x}_t^{t-1}))}{r_t^{-1/2}(y_t - h_t\widehat{x}_t^{t-1})} \tag{4.3}$$

In particular, if $\psi = \psi_H$, then, by solving the normal equations (3.6) exactly, we obtain

$$\widehat{x}_t^t = \begin{cases} \widehat{x}_t^{t-1} + P_t^{t-1}h_t' r_t^{-1/2}\psi_H\left(\frac{r^{-1/2}(y_t - h_t\widehat{x}_t^{t-1})}{h_t P_t^{t-1}h_t' + r_t}\right) & \text{if } y_t \text{ is observed} \\ \widehat{x}_t^{t-1} & \text{if } y_t \text{ is missing} \end{cases} \tag{4.4}$$

while the recursive formula for P_t^t may be approximated by

$$P_t^t = \begin{cases} P_t^{t-1} - \frac{P_t^{t-1}h_t' h_t P_t^{t-1}}{h_t P_t^{t-1}h_t' + r_t} & \text{if } y_t \text{ is observed} \\ P_t^{t-1} & \text{if } y_t \text{ is missing} \end{cases} \tag{4.5}$$

to avoid complex calculations.

4.2 Holt method

The Holt method is a smoothing and predicting method recommended for univariate time series $\{y_t\}$ wich have a locally linear trend. The corresponding linear model has the form

$$L_t = L_{t-1} + T_{t-1} + \delta L_t, \quad \delta L_t \sim \text{ iid } N(0, \sigma_1^2) \tag{4.6}$$

$$T_t = T_{t-1} + \delta T_t, \quad \delta T_t \sim \text{ iid } N(0, \sigma_2^2) \tag{4.7}$$

$$y_t = L_t + \varepsilon_t, \quad \varepsilon_t \sim \text{ iid } N(0, \sigma^2) \tag{4.8}$$

where L_t and T_t denote the level and trend at time t, respectively. The residuals $\{\delta L_t\}$, $\{\delta T_t\}$ and $\{\varepsilon_t\}$ are assumed to be mutually independent.

Although exact recursive formulae can be derived using the Kalman filter methodology, the Holt method is usually represented for practical purposes in the following simplified form

$$\widehat{L}_t = \widehat{L}_{t-1} + \widehat{T}_{t-1} + \alpha e_t^{t-1} \tag{4.9}$$

$$\widehat{T}_t = \widehat{T}_{t-1} + \alpha\gamma e_t^{t-1} \tag{4.10}$$

$$\widehat{y}_{t+k}^t = \widehat{L}_t + k\widehat{T}_t, \quad k \geq 0 \tag{4.11}$$

where $\widehat{y}_{t+k}^{t}$ is the prediction of y_{t+k} at time t (when $k = 0$, it reduces to the smoothed value of y_t), $e_t^{t-1} = y_t - \widehat{y}_t^{t-1}$ is the one-step-ahead prediction error and α, γ $(0 < \alpha, \gamma < 1)$ are smoothing constants which approximate the exact coefficients (variable with time) provided by the Kalman filter (e.g., α, $\gamma \sim 0.1$).

In the case of a time series $\{y_t\}$ with missing observations and outliers, we can modify (4.9) and (4.10) in accordance with Section 3 to the form

$$\widehat{L}_t = \begin{cases} \widehat{L}_{t-1} + \widehat{T}_{t-1} + \alpha s_t \psi_H(e_t^{t-1}/s_t) & \text{if } y_t \text{ is observed} \\ \widehat{L}_{t-1} & \text{if } y_t \text{ is missing} \end{cases} \tag{4.12}$$

$$\widehat{T}_t = \begin{cases} \widehat{T}_{t-1} + \alpha\gamma s_t \psi_H(e_t^{t-1}/s_t) & \text{if } y_t \text{ is observed} \\ \widehat{T}_{t-1} & \text{if } y_t \text{ is missing} \end{cases} \tag{4.13}$$

(Equation (4.11) remains unchanged). Here, s_t is a suitable estimate of the standard deviation $\sigma(e_t^{t-1})$. For example,

$$s_t = \begin{cases} 1.25\nu|e_t^{t-1}| + (1-\nu)s_{t-1} & \text{if } y_t \text{ is observed} \\ s_{t-1} & \text{if } y_t \text{ is missing} \end{cases} \tag{4.14}$$

where ν $(0 < \nu < 1)$ is a constant chosen to be close to zero and 1.25 approximates $(\pi/2)^{1/2}$ (see, e.g., Montgomery and Johnson (1976)).

Remark 4.1. A more sophisticated procedure of robustification, which follows more closely the Kalman filter results of Section 4.1, replaces the expression $\alpha s_t \psi_H(e_t^{t-1}/s_t)$ in (4.12) by $\alpha(1-\alpha)^{-1/2} s_t \psi_H((1-\alpha)^{1/2} e_t^{t-1}/s_t)$ and the expression $\alpha\gamma s_t \psi_H(e_t^{t-1}/s_t)$ in (4.13) by $\alpha\gamma(1-\alpha)^{-1/2} s_t \psi_H((1-\alpha)^{1/2} e_t^{t-1}/s_t)$ (for details see Cipra, Rubio and Canal (1992)).

4.3 Parameter estimation for the AR(1) process

Consider an AR(1) process

$$y_t = \varphi y_{t-1} + \varepsilon_t \tag{4.15}$$

where $\{\varepsilon_t\}$ is white noise with $\operatorname{var}\varepsilon_t = \sigma^2$. If we want to estimate the parameter φ recursively in the case of missing observations and innovation outliers (e.g., $\varepsilon_t \sim \varepsilon$-contaminated $N(0, \sigma^2)$) then, using the dynamic linear model

$$x_t = x_{t-1} \tag{4.16}$$

$$y_t = y_{t-1} x_t + \varepsilon_t \tag{4.17}$$

(i.e. $F_t = 1, H_t = h_t = y_{t-1}$) and the formulae (1.6), (1.7), (4.4) and (4.5) we obtain

$$\widehat{x}_t^t = \begin{cases} \widehat{x}_{t-1}^{t-1} + P_t^t y_{t-1}\sigma^{-1}\psi_H\left(\frac{\sigma(y_t - y_{t-1}\widehat{x}_{t-1}^{t-1})}{P_t^t y_{t-1}^2 + \sigma^2}\right) & \text{if } y_{t-1},\ y_t \text{ are observed} \\ \widehat{x}_{t-1}^{t-1} & \text{otherwise} \end{cases} \tag{4.18}$$

$$P_t^t = \begin{cases} \frac{P_{t-1}^{t-1}\sigma^2}{P_{t-1}^{t-1}y_{t-1}^2 + \sigma^2} & \text{if } y_{t-1},\ y_t \text{ are observed} \\ P_{t-1}^{t-1} & \text{otherwise} \end{cases} \tag{4.19}$$

(with respect to Theorem 4.1, we have used P_t^t instead of $P_t^{t-1} = P_{t-1}^{t-1}$ in (4.18)).

The following convergence result holds:

Theorem 4.1. *In the AR(1) process (4.15), let an estimate $\widehat{x}_t^t$ of the parameter φ be given by means of the recursive formulae (4.18) and (4.19) with initial (random) values $\widehat{x}_0^0$ and P_0^0. Furthermore, let the following conditions be satisfied*

(i) $|\varphi| < 1$;

(ii) $\varepsilon_t \sim$ *iid*, $E\varepsilon_t = 0$, $\operatorname{var}\varepsilon_t = \sigma^2$, $(0 < \sigma^2 < \infty)$;

(iii) the distribution of ε_t *is symmetric, such that* $F_\varepsilon(-\delta) < F_\varepsilon(\delta)$ *for each* $\delta > 0$;

(iv) $E\widehat{x}_0^{02} < \infty$, $P_0^0 > 0$ *a.s., and* $\widehat{x}_0^0$, P_0^0, ε_t *are independent;*

(v)

$$\sum_{t=1}^{\infty} S_t / S(t) = \infty, \tag{4.20}$$

where

$$S_t = \begin{cases} 1 & \text{if } y_{t-1},\ y_t \text{ are observed and } S(t) = \sum_{i=1}^{t} S_i \\ 0 & \text{otherwise} \end{cases} \tag{4.21}$$

Then,

$$\widehat{x}_t^t \to \varphi \quad a.s. \tag{4.22}$$

Proof. See Section 5. □

Remark 4.2. Assumption (v) of Theorem 4.1 ensures an admissible form for the scheme of missing observations. It is satisfied, for example, in the situation when $\{y_t\}$ can be divided into segments of the same length k, where each segment contains at least one observable value y_t. Then, it obviously follows that

$$\sum_{t=1}^{\infty} S_t/S(t) \geq \sum_{i=1}^{\infty} 1/(ik) = \infty$$

The case of recursive parameter estimation in autoregressive processes with additive outliers is studied in Cipra, Rubio and Canal (1993).

Remark 4.3. Theorem 4.1 can be proved for other types of limited psi-functions ψ besides ψ_H in (4.18). It is sufficient to replace assumption (iii) by

(iii') the distribution of ε_t is such that, for an arbitrary $b \neq 0$ and $u > 0$

$$b \int_{-\infty}^{\infty} \Psi(u(b+x)) dF_\varepsilon(x) > 0$$

(see (5.10) and (5.11) in the proof of Lemma 5.1).

Remark 4.4. In practice, σ will be replaced by an estimate s_{t-1} of this parameter in (4.18) and (4.19) (e.g., an estimate of the type (4.14)). Moreover, Theorem 4.1 holds if we assume, in addition to (i)-(v), the following:

(vi) $s_t \to \sigma$ a.s.

(vii) $\left(\frac{1}{S(t)}\right) \sum_{i=1}^{t} \frac{S_i y_{i-1}^2}{s_{i-1}^2} \to \frac{\sigma_y^2}{\sigma^2}$ a.s.

(see also Example 4.1).

Example 4.1. The AR(1) process $y_t = 0.5y_{t-1} + \varepsilon_t$ has been generated for the heavy-tailed distribution $\varepsilon_t \sim 0.95N(0,1) + 0.05R(-15,15)$, where $R(a,b)$ denotes the uniform distribution on the interval (a,b). The robust recursive estimates $\widehat{x}_t^t$ and P_t^t obtained from (4.18) and (4.19) with $\sigma = 1$, $\widehat{x}_0^0 = 0$, $P_0^0 = 1$, $c = 1.645$ and given in Table 1 seem satisfactory. Moreover,

if we replace σ by its estimate s_{t-1} as given by (4.14) with $\nu = 0.05$ and with s_0 estimated from the initial segment of 20 observations y_t by using classical method for an AR(1) process then, the results are very similar to those given in Table 1.

	with $\sigma = 1$		with σ estimated by s_{t-1}	
t	$\widehat{x}_t^t$	P_t^t	$\widehat{x}_t^t$	P_t^t
10	-0.2778	0.2182	-0.3106	0.1405
20	0.3963	0.0741	0.2871	0.0490
30	0.4813	0.0425	0.3918	0.0307
40	0.5519	0.0234	0.4908	0.0191
50	0.5690	0.0172	0.5165	0.0155
60	0.4619	0.0090	0.4227	0.0110
70	0.4666	0.0088	0.4292	0.0107
80	0.4480	0.0078	0.4151	0.0097
90	0.4417	0.0062	0.4135	0.0084
100	0.4644	0.0024	0.4501	0.0055

Table 1: Recursive estimation in the simulated AR(1) model $y_t = 0.5y_{t-1} + \varepsilon_t$ with innovation outliers.

5 Proof of Theorem 4.1

Lemma 5.1. *Let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}$ be a sequence of σ-algebras in a probability space $(\Omega, \mathcal{F}, P)$. Let z_t, β_t, ξ_t and η_t $(t = 0, 1, \ldots)$ be non-negative $\mathcal{F}$-measurable random variables such that*

$$E(z_t/\mathcal{F}_{t-1}) \leq (1 + \beta_{t-1})z_{t-1} + \zeta_{t-1} - \eta_{t-1} \quad t = 1, 2, \ldots \tag{5.1}$$

$$\sum_{t=0}^{\infty} \beta_t < \infty \quad a.s., \quad \sum_{t=0}^{\infty} \zeta_t < \infty \quad a.s. \tag{5.2}$$

Then, the sequence z_t converges a.s. and

$$\sum_{t=0}^{\infty} \eta_t < \infty. \tag{5.3}$$

Proof. See Robbins and Siegmund (1971). □

Lemma 5.2. *In the AR(1) process (4.15), let $\widehat{x}_t^t$ denote an estimate of the parameter φ given by means of the recursive formula*

$$\widehat{x}_t^t = \begin{cases} \widehat{x}_{t-1}^{t-1} + a_{t-1}y_{t-1}\psi_H(u_{t-1}(y_t - y_{t-1}\widehat{x}_{t-1}^{t-1})) & \text{if } y_{t-1},\ y_t \text{ are observed} \\ \widehat{x}_{t-1}^{t-1} & \text{otherwise} \end{cases} \tag{5.4}$$

with an initial (random) value $\widehat{x}_0^0$. Here, a_{t-1} and u_{t-1} $(t = 1, 2 \ldots)$ are $\mathcal{F}_{t-1}$-measurable random variables for $\mathcal{F}_{t-1} = \sigma\{\widehat{x}_0^0, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ fulfilling

$$0 < a_{t-1}^{(1)} \le a_{t-1} \le a_{t-1}^{(2)} \ \text{a.s.}, \ \sum_{t=1}^{\infty} S_t a_{t-1}^{(1)} = \infty, \ \sum_{t=1}^{\infty} S_t (a_{t=1}^{(2)})^2 = \infty, \tag{5.5}$$

$$0 < k \le u_{t-1} \le K < \infty \quad \text{a.s.} \tag{5.6}$$

for deterministic sequences $a_{t-1}^{(1)}$ and $a_{t-1}^{(2)}$ (S_t is defined in (4.21)) and constants k and K. Let the assumptions (i)-(v) of Theorem 4.1 be satisfied. Then, the limit relation (4.21) holds.

Proof. We omit the superscripts for simplicity (e.g. we write $\widehat{x}_t$ instead of $\widehat{x}_t^t$). If we put

$$\bar{x}_t = \widehat{x}_t - \varphi \tag{5.7}$$

then, we can rewrite (5.4) in the form

$$\bar{x}_t = \bar{x}_{t-1} - S_t a_{t-1} y_{t-1} \varphi_H(u_{t-1}(y_{t-1}\bar{x}_{t-1} - \varepsilon_t))$$

so that

$$\bar{x}_t^2 \le \bar{x}_{t-1}^2 - 2S_t a_{t-1} y_{t-1} \bar{x}_{t-1} \psi_H(u_{t-1}(y_{t-1}\bar{x}_{t-1} - \varepsilon_t)) + S_t (a_{t-1}^{(2)} c y_{t-1})^2 \tag{5.8}$$

and conditional expectations are given by

$$\begin{aligned} E(\bar{x}_t^2/\mathcal{F}_{t-1}) \ &\le \ \bar{x}_{t-1}^2 + S_t (a_{t-1}^{(2)} c y_{t-1})^2 - 2S_t a_{t-1} y_{t-1} \bar{x}_{t-1} \\ &\cdot \ E\{\psi_H(u_{t-1}(y_{t-1}\bar{x}_{t-1} - \varepsilon_t))/\mathcal{F}_{t-1}\} \end{aligned} \tag{5.9}$$

We apply Lemma 5.1 to the inequality (5.9) to get

$$z_t = \bar{x}_t^2, \quad \beta_{t-1} = 0, \quad \zeta_{t-1} = S_t(a_{t-1}^{(2)} c y_{t-1})^2,$$

$$\eta_{t-1} = 2S_t a_{t-1} y_{t-1} \bar{x}_{t-1} E\{\psi_H(u_{t-1}(y_{t-1}\bar{x}_{t-1} - \varepsilon_t))/\mathcal{F}_{t-1}\}$$

It now remains to verify that $\eta_t \geq 0$ a.s. Let us denote

$$g(b,u) = E_\varepsilon\{\psi_H(u(b+\varepsilon))\} = \int_{-\infty}^{\infty} \psi_H(u(b+x))dF_\varepsilon(x) \tag{5.10}$$

for $-\infty < b < \infty$ and $k \leq u \leq K$. Then, by assumptions (ii) and (iii), we have that

$$bg(b,u) > 0, \quad b \neq 0, \quad k \leq u \leq K \tag{5.11}$$

so that η_t is a special case. By Lemma 5.1, there exists a (finite) random variable $\bar{x}$ such that

$$\bar{x}_t \to \bar{x} \tag{5.12}$$

From (5.8), it follows that, for an arbitrary n

$$\begin{aligned} \bar{x}_n^2 &\leq \bar{x}_0^2 - 2\sum_{t=1}^{n} S_t a_{t-1} y_{t-1} \bar{x}_{t-1} \psi_H(u_{t-1}(y_{t-1}\bar{x}_{t-1} - \varepsilon_t)) \\ &+ c^2 \sum_{t=1}^{n} S_t (a_{t-1}^{(2)})^2 y_{t-1}^2 \end{aligned}$$

and hence

$$2\sum_{t=1}^{\infty} S_t E\{a_{t-1} y_{t-1} \bar{x}_{t-1} \psi_H(u_{t-1}(y_{t-1}\bar{x}_{t-1} - \varepsilon_t))\} \leq E\bar{x}_0^2 + c^2\sigma_y^2 \sum_{t=1}^{\infty} S_t (a_{t-1}^{(2)})^2$$

where $\sigma_y^2 = \operatorname{var} y_t = E y_t^2$. Therefore, by (5.5), we have that

$$\sum_{t=1}^{\infty} S_t a_{t-1}^{(1)} E\{y_{t-1}\bar{x}_{t-1}\psi_H(u_{t-1}(y_{t-1}\bar{x}_{t-1} - \varepsilon_t))\} < \infty.$$

Since $\sum_{t=1}^{\infty} S_t a_{t-1}^{(1)} < \infty$, a subsequence must exist such that

$$\sum_{j=1}^{\infty} E\{y_{t_j-1}\bar{x}_{t_j-1}\psi_H(u_{t_j-1}(y_{t_j-1}\bar{x}_{t_j-1} - \varepsilon_{t_j}))\} < \infty.$$

Hence,

$$y_{t_j-1}\bar{x}_{t_j-1}E\{\psi_H(u_{t_j-1}(y_{t_j-1}\bar{x}_{t_j-1}-\varepsilon_{t_j}))/\mathcal{F}_{t_j-1}\}\to 0 \quad \text{a.s.}$$

or, equivalently,

$$y_{t_j-1}\bar{x}_{t_j-1}g(y_{t_j-1}\bar{x}_{t_j-1},u_{t_j-1})\to 0 \quad \text{a.s.}$$

From (5.11),

$$\bar{y}_{t_j-1}\bar{x}_{t_j-1}\to 0 \quad \text{a.s.} \tag{5.13}$$

Furthermore, we can write

$$\varepsilon_{t_j}\bar{x}_{t_j-1}=y_{t_j}(\bar{x}_{t_j-1}-\bar{x}_{t_j})+y_{t_j}\bar{x}_{t_j}-\varphi y_{t_j-1}\bar{x}_{t_j-1} \tag{5.14}$$

Since the y_{t_j} are identically distributed and the limit expressions (5.12) and (5.13) are satisfied, all three summands on the right-hand side of (5.14) converge in probability to zero, i.e.

$$\varepsilon_{t_j}\bar{x}_{t_j-1}\to 0 \quad \text{in probability} \tag{5.15}$$

Finally, due to the independence of $\bar{x}_{t_j-1}$ and ε_{t_j}, where the ε_{t_j} are identically distributed, and by (5.12), we have that

$$\bar{x}_t\to 0 \quad \text{a.s.}$$

□

Proof of Theorem 4.1.

We have that

$$P_t=\left(P_{t-1}^{-1}+S_ty_{t-1}^2/\sigma^2\right)^{-1}=\left[P_0^{-1}+(S_1y_0^2+\ldots+S_ty_{t-1}^2)/\sigma^2\right]^{-1}$$

If we denote

$$u_{t-1}=\sigma/(P_ty_{t-1}^2+\sigma^2)$$

then, for times t such that $S_t=1$, it is obvious that

$$(2\sigma)^{-1}\le u_{t-1}\le\sigma^{-1}$$

Furthermore, due to properties of the process y_t (see, e.g., Hannan (1970)), we have

$$S(t)P_t \to \sigma^2/\sigma_y^2 \quad \text{a.s.} \tag{5.16}$$

Choose an arbitrary $\varepsilon > 0$ and $0 < \delta < \sigma^2/\sigma_y^2$. From (5.16), there exists t_0 such that

$$P(\sup_{t \geq t_0} |S(t)P_t - \sigma^2/\sigma_y^2| < \delta) > 1 - \varepsilon$$

and hence

$$P(\cap_{t \geq t_0}[|S(t)P_t - \sigma^2/\sigma_y^2| < \delta]) > 1 - \varepsilon.$$

Put

$$\bar{x}_t = \begin{cases} \hat{x}_t & \text{for } t = 0, 1, \ldots, t_0 - 1 \\ \bar{x}_{t-1} + S_t P_t y_{t-1} \sigma^{-1} \Psi_H(u_{t-1}(y_t - y_{t-1}\bar{x}_{t-1})) & \text{for } t \geq t_0,\ |S(t)P_t - \sigma^2/\sigma_y^2| < \delta \\ \bar{x}_{t-1} + S_t S(t)^{-1}(\sigma^2/\sigma_y^2) y_{t-1} \Psi_H(u_{t-1}(y_t - y_{t-1}\bar{x}_{t-1})) & \text{for } t \geq t_0,\ |S(t)P_t - \sigma^2/\sigma_y^2| \geq \delta \end{cases}$$

Then, by Lemma 5.2 with

$$a_{t-1}^{(1)} = S(t)^{-1}(\sigma^2/\sigma_y^2 - \delta)\sigma^{-1}, \quad a_{t-1}^{(2)} = S(t)^{-1}(\sigma^2/\sigma_y^2 + \delta)\sigma^{-1},$$

$$k = (2\sigma)^{-1} \quad \text{and} \quad K = \sigma^{-1},$$

we obtain

$$\bar{x}_t \to \delta \quad \text{a.s.}$$

Finally, we can write

$$\begin{aligned} P(\hat{x}_t \to \varphi) &\geq P(\cap_{t \geq t_0}[\bar{x}_t = \hat{x}_t] \cap [\bar{x} \to \varphi]) = P(\cap_{t \geq t_0}[\bar{x} = \hat{x}_t]) \\ &\geq P(\cap_{t \geq t_0}[|S(t)P_t - \sigma^2/\sigma_y^2| < \delta]) > 1 - \varepsilon \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary we must have $\hat{x} \to \varphi$ a.s.

Acknowledgements

We thank the Associate Editor and referees for their careful reading of the manuscript and valuable comments.

References

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering.* New Jersey: Prentice-Hall, Englewood Cliffs.

Akaike, H. (1978). Covariance matrix Computation of the State Variable of a Stationary Gaussian Process. *Annals of the Institute of Statistical Mathematics*, **30**, Part B, 499-504.

Cipra, T. (1992). Robust exponential smoothing. *Journal of Forecasting*, **11**, 57-69.

Cipra, T. and R. Romera (1991). Robust Kalman filter and its application in time series analysis. *Kybernetika*, **27**, 481-494.

Cipra, T. and A. Rubio (1991). Kalman filter with a non-linear non-Gaussian observation relation. *Trabajos de Estadística*, **6**, 111-119.

Cipra, T., A. Rubio and J. L. Canal (1993). Recursive estimation in autoregressive models with additive outliers. *Kybernetika*, **29**, 62-72.

Cipra, T., A. Rubio and J. L. Trujillo (1991). Time series analysis: recursive methods and their modification for time series with outliers and missing observations. *Extracta Mathematicae*, **6**, 64-95.

Cipra, T., J.L. Trujillo and A. Rubio (1995). Holt-Winters method with missing observation. *Management Science* **41**, 174-178.

Gardner, G., A.C. Harvey and G.D.A. Phillips (1980). An Algorithm for Exact Maximum Likelihood Estimation by Means of Kalman Filtering. *Applied Statistics*, **29**, 311-322.

Guttman, I. and D. Peña (1985). Robust Filtering. Comment on Dynamic Generalized Linear Models and Bayesian Forecasting by West, Harrison and Migon, *Journal of the American Statistical Association*, **80**, 91-92.

Hannan, E. J. (1970). *Multiple Time Series*, J. Wiley, New York.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of the Institute of Statistical Mathematics*, **35**, 73-101.

Huber, P. J. (1981). Robust Statistics. J. Wiley, New York.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory.* New York: Academic Press.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, **22**, 389-395.

Jong, P. de (1989). Smoothing and Interpolation With the State Space Model. *Journal of the American Statistical Association*, **84**, 408-809.

Kleiner, B., R.D. Martin and D.J. Thompson (1979). Robust estimation of power spectra. *Journal of the Royal Statistical Association B*, **41**(3), 313-351.

Kohn, R. and C. F. Ansley (1985). Efficient Estimation and Prediction along time Series Regression Models. *Biometrika*, **72**, 694-497.

Kohn, R. and C.F. Ansley (1986). Estimation, Prediction and Interpolation for ARIMA Models With Missing Data. *Journal of the American Statistical Association*, **81**, 751-761.

Masreliez, C. J. (1975). Approximate non-Gaussian filtering with linear state and observation relations. *IEEE Transactions on Automation and Control*, **AC-20**, 107-110.

Masreliez, C. J. and R.D. Martin (1977). Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Automation and Control*, **AC-22**, 361-371.

Meinhold, R. J. and N.D. Singpurwalla (1989). Robustification of Kalman filter models. *Journal of the American Statistical Association*, **84**, 479-486.

Montgomery, D. C. and L.A. Johnson (1976). *Forecasting and Time Series Analysis*. New York: McGraw-Hill.

Morris, J. M. (1976). The Kalman Filter: A Robust Estimator for Some Classes of Linear Quadratic Problems. *IEEE Transactions on Information Theory*, **IT-22**, 526-534.

Robbins, H. and D. Siegmund (1971). A convergence theorem for non negative almost supermartingales and some applications. *Optimizing Methods in Statistics* (J. S. Rustagi, de.). New York: Academic Press, 233-257.

Romera, R. and T. Cipra (1993). On practical implementation of Robust Kalman Filtering. *Communications in Statistics: Simulation and Computation*, **24**, 461-488.

Servi, L. D. and Y.C. Ho (1981). Recursive estimation in the presence of uniformly distributed measurement noise. *IEEE Transactions on Automation and Control*, **AC-26**, 563-565.

West, M. (1981). Robust Sequential Approximate Bayesian Estimation. *Journal of the Royal Statistical Association B*, **43**, 157-166.

West, M., P.J. Harrison and H.S. Migon (1985). Dynamic Generalized Linear Models and Bayesian Forecasting. *Journal of the American Statistical Association*, **80**, 73-97.