



ELSEVIER

Statistics & Probability Letters 23 (1995) 79–92

STATISTICS &
PROBABILITY
LETTERS

Robust recursive estimation for correlated observations

Irwin Guttman^a, Dennis K.J. Lin^{b,*}

^a Department of Statistics, State University of New York, Buffalo, NY 14214-3000, USA

^b Department of Statistics, University of Tennessee, 331 Stokely Management Center, Knoxville, TN 37996-0532, USA

Received August 1992; revised March 1994

Abstract

The Kalman filter is probably the most popular recursive estimation method. It is, however, known to be non-robust to spuriously generated observations. Much attention has been focused on finding the so-called robust recursive estimation under the assumption that the observations are independent. In this paper, we show that Lin and Guttman's robust recursive estimation scheme can be easily applied to the correlated observations. Examples when the noise follows an AR(2) process with/without outliers are given for illustration.

Keywords: Box–Jenkins model; Kalman filter; Mixture distribution; Robust filter; Spurious observations

1. Introduction

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. The goal here is to make inference about θ_t , called the *state of nature*. The observed values of the variable of interest y_t depend on the unobservable θ_t at time t . The relationship between y_t and θ_t ,

$$y_t = A_t \theta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, C_t), \quad (1)$$

is known as the *observation equation*, whereas the dynamic feature between θ_t and θ_{t-1} ,

$$\theta_t = \Omega_t \theta_{t-1} + u_t, \quad u_t \sim N(0, R_t), \quad (2)$$

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 the system equation, i.e., $C_t > R_t$ in some sense.

Kalman's (1960) result, popular with control engineers and other physical scientists, is essentially a least-squares procedure. One well-recognized concern for this least-squares procedure is its non-robustness to

* Corresponding author.

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.

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). Recently, Lin and Guttman (1993) have discussed the use of 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 Guttman and Peña, 1985). When independence is assumed in the observation system (i.e., among y_i 's), the scaled-contaminated model is successfully used to handle the spuriousity in the Kalman filter. On the other hand, if the independence is not assumed in the observation system, we need to know the function of such relationships among y_i 's in order to employ any estimation scheme. In this paper, we discuss how to extend the Lin and Guttman (1993) results for the cases that y_i 's are correlated, using Box–Jenkins models as examples of the relationships among y_i 's for illustration.

2. Lin and Guttman filter

Lin and Guttman's (1993) scheme proceeds as follows. Initially, a preliminary "estimate" of the prior behavior of θ is made, say $\theta_0 \sim N(\mu_0, V_0)$. Now suppose that instead of ε_i 's of (1), we have that $\varepsilon_i \sim \alpha_1 N(0, C_{i,1}) + \alpha_2 N(0, C_{i,2})$, for α_i ($i = 1, 2$), $\alpha_1 + \alpha_2 = 1$. At each stage, after observing y_i , we compute the updated estimates of the α_i 's, say $\alpha_{i,i}$'s, and then update the estimation of μ .

Starting from time $t = 1$, we first compute the posterior probabilities, labelled $\alpha_{1,1}$ and $\alpha_{1,2}$, where $\alpha_{1,1}$ 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 $N(A_1\theta_1, C_{1,2})$. These are given by

$$\alpha_{1,1} = \left[1 + \frac{\alpha_2}{\alpha_1} \left(\frac{\|M_{1,1}\|}{\|M_{1,2}\|} \right)^{1/2} \exp \left\{ \frac{1}{2} (y_1 - A_1\mu_{1|0})' (M_{1,1}^{-1} - M_{1,2}^{-1}) (y_1 - A_1\mu_{1|0}) \right\} \right]^{-1},$$

$$\alpha_{1,2} = 1 - \alpha_{1,1}, \quad (3)$$

where

$$\mu_{1|0} = \Omega_1\mu_0, \quad V_{1|0} = R_1 + \Omega_1V_0\Omega_1' \quad (4)$$

and

$$M_{1,i} = C_{1,i} + A_1V_{1|0}A_1' \quad (i = 1, 2). \quad (5)$$

Then, a collapsing to a single normal of the estimated mixture noise distribution $\alpha_{1,1}N(0, C_{1,1}) + \alpha_{1,2}N(0, C_{1,2})$ is employed by using moments. This turns out to be, as is easily verified,

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

The likelihood function is then taken to be

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

while the prior for θ_1 is given by

$$\theta_1 \sim N(\mu_0, R_1 + \Omega_1V_0\Omega_1').$$

Table 1
The Lin and Guttman robust recursive estimation scheme

Model

$$y_t = A_t \theta_t + \varepsilon_t, \quad \varepsilon_t \sim \alpha_1 N(0, C_{t,1}) + \alpha_2 N(0, C_{t,2})$$

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

Initial setting

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

Prediction

$$\mu_{t|t-1} = \Omega_t \mu_{t-1}, \quad V_{t|t-1} = R_t + \Omega_t V_{t-1} \Omega_t', \quad y_t = A_t \mu_{t|t-1}$$

$$M_{t,1} = C_{t,1} + A_t V_{t|t-1} A_t' \quad \text{and} \quad M_{t,2} = C_{t,2} + A_t V_{t|t-1} A_t'$$

Compute posterior probabilities

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

$$\alpha_{t,2} = 1 - \alpha_{t,1}$$

$$M_t = \alpha_{t,1} M_{t,1} + \alpha_{t,2} M_{t,2} = \alpha_{t,1} C_{t,1} + \alpha_{t,2} C_{t,2} + A_t V_{t|t-1} A_t'$$

Updating of the parameters

$$\mu_t = \mu_{t|t-1} + V_{t|t-1} A_t' M_t^{-1} (y_t - A_t \mu_{t|t-1})$$

$$V_t = V_{t|t-1} - V_{t|t-1} A_t' M_t^{-1} A_t V_{t|t-1}$$

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

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

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

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 matrix $V_{1|0} A_1' M_1^{-1}$ is often referred to as the *Kalman gain filter (matrix)*.

The posterior of $\theta_1 | y_1$ 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 1. We next show how such a robust recursive estimation can be applied to the cases where the observations are correlated.

3. Correlated observations

The assumption that y_t 's are independent is much too strong in practice. The model in this paper discusses cases for which the y_t 's are not independent. Indeed, we assume that observations y_t 's are generated from an ARMA(p, q) process in the sense of Box and Jenkins (1976). Note that our approach can be applied to any process, provided the covariance matrix, Σ_t , is available. The ARMA(p, q) process is simply a special case for illustration. Also, we discuss here only the univariate case. The extension to multivariate cases is straightforward. Specifically, the model we consider can be specified as follows (where we assume that $\phi_{t,i}$'s and $\psi_{t,i}$'s are within the unit circle).

Observation equation:

$$y_t = a_t \theta_t + \xi_t, \quad (6)$$

where

$$\xi_t - \sum_{i=1}^p \phi_i \xi_{t-i} = \varepsilon_t - \sum_{i=1}^q \psi_i \varepsilon_{t-i}, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad \varepsilon_t \text{'s independent.} \quad (7)$$

System equation:

$$\theta_t = \omega_t \theta_{t-1} + u_t, \quad u_t \sim N(0, \sigma_{u_t}^2), \quad (8)$$

$$\theta_{t-1} = \mu_{t-1} + z_{t-1}, \quad z_{t-1} \sim N(0, v_{t-1}). \quad (9)$$

Eqs. (6) and (7) determine the joint likelihood function for $\underline{y}_t = (y_t, y_{t-1}, \dots, y_1)'$, which is that of $N(\theta_t^*, \Sigma_t)$ where $\theta_t^* = (a_t \theta_t, a_{t-1} \theta_{t-1}, \dots, a_1 \theta_1)'$. The variance-covariance matrix of \underline{y}_t , say $\Sigma_t = \text{Cov}(\underline{y}_t | \theta_t) = \text{Cov}(\xi)$, has a complicated form and is discussed in Box and Jenkins (1976). From (8) and (9), we may deduce that the prior for the state of nature at time t , θ_t , is $N(\mu_{t|t-1} = \omega_t \mu_{t-1}, v_{t|t-1} = \sigma_{u_t}^2 + \omega_t^2 v_{t-1})$.

We note that the y_t 's are not independent and, for further simplicity, we now consider the case $\omega_t \equiv 1$, i.e., we make the assumption that $\theta_t \equiv \theta$ so that $\theta_t = \theta + u_t$. Then the posterior for θ_t , at time t , can be shown to be normal with mean μ_t and variance v_t , with

$$\mu_t = \mu_{t-1} + \frac{a_t' \Sigma_t^{-1}}{a_t' \Sigma_t^{-1} a_t + v_{t|t-1}^{-1}} (y_t - \mu_{t-1} a_t) \quad (10)$$

and

$$v_t = [a_t' \Sigma_t^{-1} a_t + v_{t|t-1}^{-1}]^{-1} = v_{t|t-1} - v_{t|t-1} [v_{t|t-1} + (a_t' \Sigma_t^{-1} a_t)]^{-1} v_{t|t-1}, \quad (11)$$

and the Kalman gain filter here is given by

$$\frac{a_t' \Sigma_t^{-1}}{a_t' \Sigma_t^{-1} a_t + v_{t|t-1}^{-1}}, \quad (12)$$

where $\mu_{t|t-1} = \omega_t \mu_{t-1} = \mu_{t-1}$ and $v_{t|t-1} = \sigma_{u_t}^2 + v_{t-1}$. Note that Σ_t is a $t \times t$ matrix, \underline{y}_t and $a_t = (a_t, a_{t-1}, \dots, a_1)'$ are $t \times 1$ vectors, and all other quantities are scalars.

Note: Eqs. (6), (7) together with (8) imply that $y_t = (a_t \omega_t) \theta_{t-1} + a_t u_t + \xi_t$, so that we have two random components: ξ_t is an ARMA (p, q) process and u_t is normal with mean 0 and variance σ_u^2 . This is exactly the model considered by Tiao and Ali (1971), with the interest of that paper focusing on making inference about the ϕ_t 's, and ψ_t 's, if $a_t = \beta$ for all t , but when both y_t 's and θ_t 's are observed. Our concern here, however, is to make inference about the unobservable θ_t , when y_t is observed (ϕ_t 's and ξ_t 's are given).

4. The ARMA (1,1) process

As a special case of (7), take $p = 1$ and $q = 1$, namely, an ARMA (1,1) process, and also assume $\sigma_\varepsilon^2 = \sigma^2$, for all t . Then, the observation equation is

$$y_t = a_t \theta_t + \xi_t,$$

with

$$\xi_t - \phi \xi_{t-1} = \varepsilon_t - \psi \varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, \sigma^2).$$

It then turns out that (see, e.g., Box and Jenkins, 1976) $\Sigma_t = \text{Cov}(y_t | \theta) = \sigma^2 \mathbf{H}$, where \mathbf{H} has (i, j) -element, h_{ij} say,

$$h_{ij} = \begin{cases} (1 - \phi^2)^{-1}(1 + \psi^2 - 2\phi\psi) & \text{for } i = j, \\ (1 - \phi^2)^{-1}(\phi - \psi)(1 - \phi\psi)\phi^{|i-j|-1} & \text{for } i \neq j. \end{cases} \quad (13)$$

For example, when $\phi = 0$ (i.e., dealing with an MA(1) process), we would then have

$$\Sigma_t = \sigma^2 \begin{bmatrix} 1 + \psi^2 & -\psi & 0 & \dots & 0 & 0 \\ -\psi & 1 + \psi^2 & -\psi & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \dots & 1 + \psi^2 & -\psi \\ 0 & 0 & \cdot & \dots & -\psi & 1 + \psi^2 \end{bmatrix}. \quad (14)$$

When $\psi = 0$ (i.e., dealing with an autoregressive AR(1) process), we have

$$\Sigma_t = \frac{\sigma^2}{1 - \phi^2} \begin{bmatrix} 1 & \phi & \dots & \phi^{t-1} \\ \phi & 1 & \dots & \phi^{t-2} \\ \cdot & \cdot & \cdot & \cdot \\ \phi^{t-1} & \phi^{t-2} & \dots & 1 \end{bmatrix}. \quad (15)$$

From (15), we see that

$$\Sigma_t^{-1} = \sigma^{-2} \begin{bmatrix} 1 & -\phi & 0 & \dots & \cdot & 0 \\ -\phi & 1 + \phi^2 & -\phi & \dots & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \dots & 1 + \phi^2 & -\phi \\ 0 & 0 & \cdot & \dots & -\phi & 1 \end{bmatrix}. \quad (16)$$

Thus,

$$\underline{a}'_t \Sigma_t^{-1} \underline{a}_t = \sigma^{-2} (a_t - \phi a_{t-1}, -\phi a_t + (1 + \phi^2)a_{t-1} - \phi a_{t-2}, \dots, -\phi a_3 + (1 + \phi^2)a_2 - \phi a_1, -\phi a_2 + a_1),$$

and

$$\underline{a}'_t \Sigma_t^{-1} \underline{a}_t = \sigma^{-2} [a_t(a_t - \phi a_{t-1}) + \sum_{j=2}^{t-1} a_j [-\phi a_{j+1} + (1 + \phi^2)a_j^2 - \phi a_{j-1}] + a_1(-\phi a_2 + a_1)].$$

The above can be substituted in (10–12) to obtain updated parameters.

In particular, when $a_t \equiv 1$, for all t , we find

$$\mu_t = \mu_{t-1} + \left[\frac{(1 - \phi) [t(1 - \phi) + 2\phi]}{(1 - \phi)[t(1 - \phi) + 2\phi] + \sigma^2 v_{t|t-1}^{-1}} \right] \left[\frac{y_t + (1 - \phi) \sum_{i=2}^{t-1} y_i + y_1}{1 + (t - 2)(1 - \phi) + 1} - \mu_{t-1} \right], \quad (17)$$

$$v_t = \frac{\sigma^2 v_{t|t-1}}{(1 - \phi) [t(1 - \phi) + 2\phi] v_{t|t-1} + \sigma^2}, \quad (18)$$

and have that the Kalman gain filter is

$$\frac{v_{t|t-1}(1-\phi)[1, 1-\phi, \dots, 1-\phi, 1]}{(1-\phi)[t(1-\phi) + 2\phi]v_{t|t-1} + \sigma^2}. \quad (19)$$

As we can see, because σ^2 is given and $v_{t|t-1} = \sigma_{ut}^2 + v_{t-1}$ is bounded, as t becomes large, the updated variance (18) tends to zero, and the updated mean (17) goes to $\bar{y} = \sum_{i=1}^t y_i/t$. Note that the standard Kalman filter setting is intimately related to the AR (1) process considered above. Some simulation results show that similar estimation will be obtained via the standard Kalman filter approach and via (17), (18).

5. The AR (2) process

We now consider another special case of (2), when $p = 2$ and $q = 0$, i.e., an AR (2) process with $\sigma_t^2 \equiv \sigma^2$, for all t ; namely,

$$\xi_t = \phi_1 \xi_{t-1} + \phi_2 \xi_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2).$$

It follows that (see, for example, Box and Jenkins, 1976) $\Sigma_t = \text{Cov}(y_t | \theta) = \sigma^2 K$, where K has its (i, j) th element, k_{ij} , say,

$$k_{ij} = \left[\frac{1 - \phi_2}{1 + \phi_2} \right] \frac{\rho(|i - j|)}{(1 - \phi_2)^2 - \phi_1^2}, \quad (20)$$

and $\rho(\cdot)$ follows the Yule–Walker equation of order 2, namely,

$$\rho(t) = \phi_1 \rho(t-1) + \phi_2 \rho(t-2) \quad \text{for all } t \geq 2,$$

with $\rho(0) = 1$ and $\rho(1) = \phi_1/(1 - \phi_2)$. It turns out that, for $t \geq 5$,

$$\Sigma_t^{-1} = \sigma^{-2} \begin{bmatrix} 1 & -\phi_1 & -\phi_2 & 0 & 0 & \dots & 0 & 0 & 0 \\ -\phi_1 & 1 + \phi_1^2 & -\phi_1 + \phi_1 \phi_2 & -\phi_2 & 0 & \dots & 0 & 0 & 0 \\ -\phi_2 & -\phi_1 + \phi_1 \phi_2 & 1 + \phi_1^2 + \phi_2^2 & -\phi_1 + \phi_1 \phi_2 & -\phi_2 & 0 & 0 & 0 & 0 \\ 0 & -\phi_2 & -\phi_1 + \phi_1 \phi_2 & 1 + \phi_1^2 + \phi_2^2 & -\phi_1 + \phi_1 \phi_2 & -\phi_2 & \dots & \cdot & \cdot \\ 0 & 0 & -\phi_2 & -\phi_1 + \phi_1 \phi_2 & \cdot & \cdot & & & \\ 0 & 0 & 0 & -\phi_2 & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \dots & -\phi_2 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 1 + \phi_1^2 + \phi_2^2 & -\phi_1 + \phi_1 \phi_2 & -\phi_2 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & -\phi_1 + \phi_1 \phi_2 & 1 + \phi_1^2 & -\phi_1 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & -\phi_2 & -\phi_1 & 1 \end{bmatrix}$$

Hence

$$\begin{aligned} \underline{a}_t^{\prime} \Sigma_t^{-1} &= \sigma^{-2} (a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2}, -\phi_1 a_t + (1 + \phi_1^2) a_{t-1} + (-\phi_1 + \phi_1 \phi_2) a_{t-2} - \phi_2 a_{t-3}, \dots, \\ &\quad -\phi_2 a_{j+2} + (-\phi_1 + \phi_1 \phi_2) a_{j+1} + (1 + \phi_1^2 + \phi_2^2) a_j + (-\phi_1 + \phi_1 \phi_2) a_{j-1} - \phi_2 a_{j-2}, \dots, \\ &\quad -\phi_2 a_4 + (-\phi_1 + \phi_1 \phi_2) a_3 + (1 + \phi_1^2) a_2 - \phi_1 a_1, -\phi_2 a_3 - \phi_1 a_2 + a_1), \end{aligned}$$

from which we may easily find the explicit form of $\underline{a}'_t \Sigma_t^{-1} \underline{a}_t$. In particular, when $a_t \equiv 1$, for all t ,

$$\underline{1}' \Sigma_t^{-1} \underline{1} = \sigma^{-2}(1 - \phi_1 - \phi_2)(1, (1 - \phi_1), (1 - \phi_1 - \phi_2), \dots, (1 - \phi_1 - \phi_2), (1 - \phi_1 - \phi_2), 1 - \phi_1, 1),$$

and

$$\underline{1}' \Sigma_t^{-1} \underline{1} = \sigma^{-2}(1 - \phi_1 - \phi_2)[(t - 4)(1 - \phi_1 - \phi_2) + 2(1 - \phi_1) + 2].$$

Therefore, we obtain the updated estimates as

$$\begin{aligned} \mu_t = \mu_{t-1} + & \left[\frac{2 + 2(1 - \phi_1) + (t - 4)(1 - \phi_1 - \phi_2)}{2 + 2(1 - \phi_1) + (t - 4)(1 - \phi_1 - \phi_2) + \sigma^2 v_{t|t-1}^{-1} (1 - \phi_1 - \phi_2)^{-1}} \right] \\ & \times \left[\frac{y_t + (1 - \phi_1)y_{t-1} + (1 - \phi_1 - \phi_2) \sum_{j=3}^{t-2} y_j + (1 - \phi_1)y_2 + y_1}{1 + (1 - \phi_1) + (t - 4)(1 - \phi_1 - \phi_2) + (1 - \phi_1) + 1} - \mu_{t-1} \right], \end{aligned} \quad (21)$$

$$v_t = [(t - 4)(1 - \phi_1 - \phi_2) + 2(1 - \phi_1) + 2](1 - \phi_1 - \phi_2)\sigma^{-2} + v_{t|t-1}^{-1}]^{-1}. \quad (22)$$

By taking $\phi_2 = 0$ and $\phi_1 = \phi$ (an AR(1) process), of course, (21) becomes (17), and (22) becomes (18). In *non-stationary* cases (i.e., the roots of $m^2 - \phi_1 m - \phi_2 = 0$ are not inside the unit circle), the values of y_t 's are unstable, rendering wild behavior for the Kalman filter estimate as time increases. But some examples carried out by the authors based on simulated data show that this approach yields stable estimation for the unobservable θ .

6. Illustrative examples

In this section, we provide two simulated examples with an AR(2) process. The data are simulated so that an outlier can be easily inserted.

Example 1. Consider the following stationary AR(2) process with $\sigma^2 = 3^2$, $\phi_1 = \frac{1}{4}$, $\phi_2 = -\frac{3}{4}$.

Observation equation:

$$\begin{aligned} y_t &= \theta_t + \xi_t, \\ \xi_t &= \frac{1}{4}\xi_{t-1} - \frac{3}{4}\xi_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, 3^2). \end{aligned}$$

System equation:

$$\theta_t = \theta + u_t, \quad u_t \sim N(0, 1).$$

The unobservable in this simulation is assigned as $\theta = 10$. The initial (prior) setting is taken to be $\theta_0 \sim N(12, 12)$. Table 2 summarizes the performance of the updated mean and variance for the standard Kalman filter and the new approach discussed above. Figs. 1 and 2 provide a visual comparison of the updated mean and variance, respectively.

In Fig. 1, where the y_t values and the updated means are plotted, a typical AR(2) pattern in the y_t 's is shown as expected. The standard Kalman filter, however, is influenced by these observations and has a similar behavior to the y_t 's, but closer to the goal $\theta = 10$. In contrast, by knowing the AR(2) structure beforehand, we see that the new filter for the AR(2) process is consistently forecasting in a better fashion.

In Fig. 2, where the updated variances are plotted, we see that the updated variance for the standard Kalman filter soon converges to the value 2.54 which is the root of $x = (1 + x) - (1 + x)(1 + 9 + x)^{-1}(1 + x)$, as discussed in Section 4. The updated variance for the new filter is much smaller than that of the standard Kalman filter except the very beginning one which depends heavily on the initial setting;

Table 2
Comparison of two filters (Example 1)

t	y_t	Mean (μ_t)		Variance (V_t)	
		μ_{std}	μ_{new}	V_{std}	V_{new}
1	8.74	10.07	10.75	5.32	8.03
2	6.11	8.44	9.33	3.71	0.08
3	10.04	8.99	9.34	3.09	0.26
4	11.52	9.78	9.30	2.81	0.40
5	14.07	11.06	9.80	2.68	0.53
6	15.12	12.24	10.53	2.61	0.63
7	6.35	10.55	10.55	2.58	0.73
8	4.66	8.88	10.45	2.56	0.83
9	15.88	10.86	10.69	2.55	0.92
10	20.01	13.45	10.96	2.55	1.00
11	7.07	11.65	10.87	2.54	1.08
12	- 2.69	7.59	10.54	2.54	1.15
13	11.26	8.63	10.56	2.54	1.22
14	20.66	12.03	10.57	2.54	1.29
15	6.46	10.46	10.30	2.54	1.36
16	1.12	7.82	10.25	2.54	1.42
17	12.02	9.01	10.29	2.54	1.48
18	24.72	13.44	10.56	2.54	1.55
19	10.41	12.59	10.49	2.54	1.60
20	- 5.28	7.54	10.32	2.54	1.66
21	- 1.59	4.96	10.05	2.54	1.72
22	17.83	8.60	10.01	2.54	1.77
23	23.56	12.82	10.09	2.54	1.82
24	4.68	10.52	10.01	2.54	1.88
25	- 1.50	7.13	10.00	2.54	1.93
26	11.29	8.30	10.01	2.54	1.98
27	17.24	10.82	9.97	2.54	2.03
28	6.10	9.49	9.85	2.54	2.07
29	6.42	8.62	9.92	2.54	2.12
30	18.76	11.49	10.07	2.54	2.17

The unobservable $\theta = 10$.

however, it is increasing as the time t increases. This is simply because the Σ_t matrix is expanding at each stage resulting in a larger updated variance at each stage. If we constrain the Σ_t matrix to be the same size for all t , say 3×3 for the AR (2) case (see the next example), the updated variances turn out to be consistently small. This procedure gives rise to a filter that is robust to outliers in general, as is to be discussed below.

If we are concerned with the possibility of spuriousness that may cause outliers, we replace the assumption $\varepsilon_t \sim N(0, \sigma_t^2)$ in (7) by

$$\varepsilon_t \sim \alpha_1 N(0, \sigma_{t,1}^2) + \alpha_2 N(0, \sigma_{t,2}^2).$$

Then, after evaluating the posterior probabilities $\alpha_{t,1}$ and $\alpha_{t,2} = 1 - \alpha_{t,1}$, we collapse

$$\alpha_{t,1} N(0, \sigma_{t,1}^2) + \alpha_{t,2} N(0, \sigma_{t,2}^2)$$

to

$$N(0, \alpha_{t,1} \sigma_{t,1}^2 + \alpha_{t,2} \sigma_{t,2}^2),$$

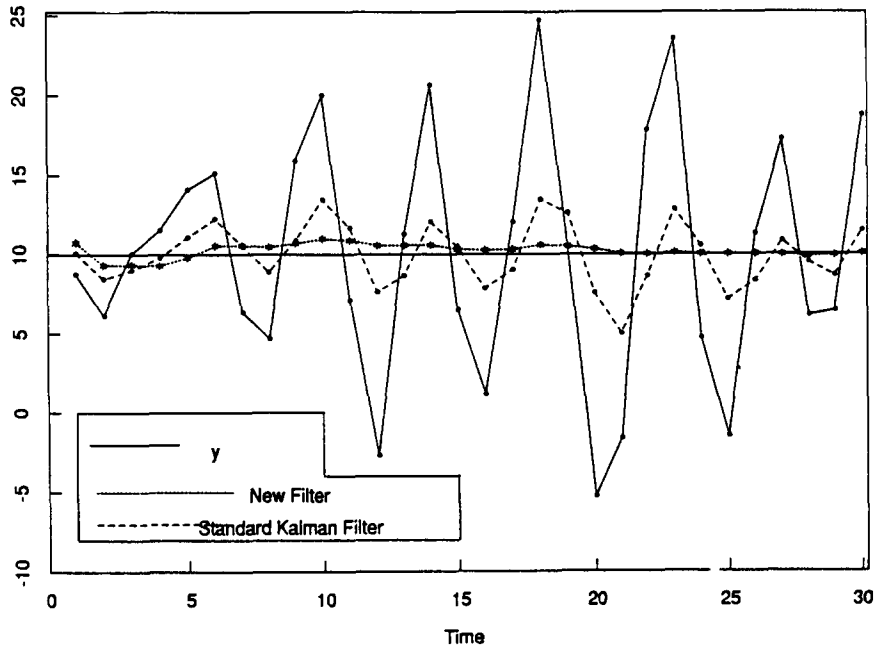


Fig. 1. Comparison plot for Example 1

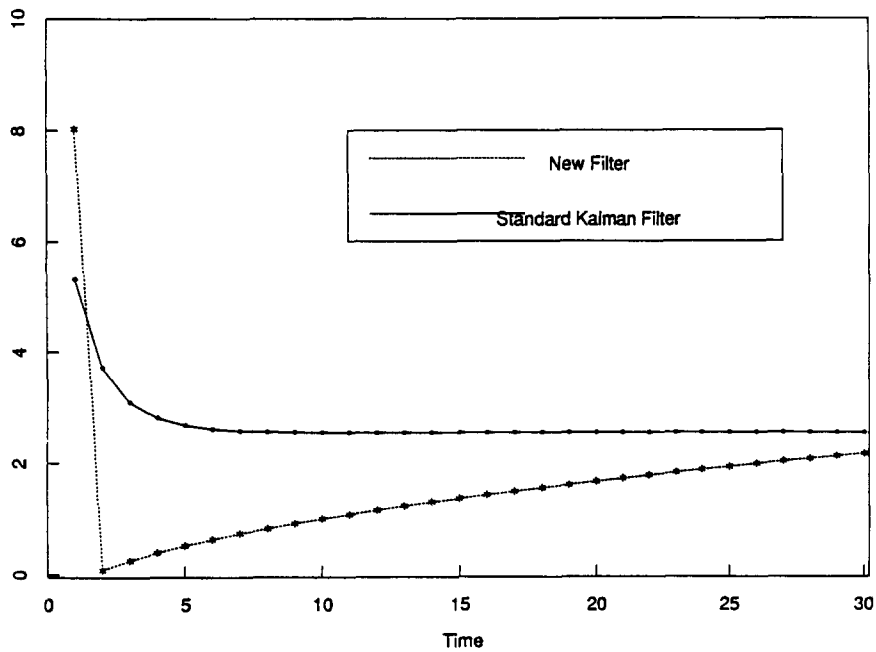


Fig. 2. Comparison of variance in Example 1

and then set $\alpha_{t,1}\sigma_{t,1}^2 + \alpha_{t,2}\sigma_{t,2}^2 = \sigma_t^2$, and apply the methodology of the previous section to the processes (10)–(11). For the AR(2) process, this becomes

$$\begin{aligned}
 M_{t,1} &= \left[\frac{1 - \phi_2}{1 + \phi_2} \right] \frac{\sigma_{t,1}^2}{(1 - \phi_2)^2 - \phi_1^2} + V_{t|t-1}, \\
 M_{t,2} &= \left[\frac{1 - \phi_2}{1 + \phi_2} \right] \frac{\sigma_{t,2}^2}{(1 - \phi_2)^2 - \phi_1^2} + V_{t|t-1}.
 \end{aligned}
 \tag{23}$$

The updated scheme in (10)–(12) involves all observations up to time t . Thus, when an outlying observation arises, this updated method can be unstable as in the standard Kalman filter. To handle this, we reconstruct (10)–(12), using only the last $(p + 1)$ observations, rather than all t , for we are involved with an AR(p) process, so that y_t is only associated with y_{t-1}, \dots, y_{t-p} . As an example, we do the above for an AR(2) process. The Σ_t^* matrix is 3×3 and is equal to $A^* \text{Cov}(\xi_t) A^{*-1}$ where A^* now is the 3×3 left-upper submatrix of the A matrix used previously. Also, $y_t^* = (y_{t-2}, y_{t-1}, y_t)'$, and $a_t^* = (a_{t-2}, a_{t-1}, a_t)'$. The formulae (10)–(12) are then applied. When such a modified scheme was used in Example 1, we obtain a similar result for the updated mean but with consistently smaller updated variance.

Example 2. We now reconsider Example 1 (where y_t 's are not independent and are generated via an AR(2) process) by replacing y_{11} by 65, clearly an outlier. Table 3 summarizes the performance of the updated mean and variance for the standard Kalman filter and the new (modified) approach. Figs. 3 and 4 provide a visual comparison of the updated mean and variance, respectively.

The y_t values and the updated means are plotted in Fig. 3. We see that the standard Kalman filter shows that for this AR(2) pattern, the outlier y_{11} influences this standard filter unduly. The new filter, however, smooths out both effects, and consistently close to the target $\theta = 10$. The value μ_{13} from the new approach shows that the outlier y_{11} also influences the new approach but with much smaller effect and not immediately.

The updated variances, plotted in Fig. 4, confirm our discussion that the updated variance of the new approach is not increasing but consistently small, except, of course, at the very beginning, $t = 1$, which depends heavily upon the initial setting V_0 . The updated variance for the standard Kalman filter is, of course, identical to Fig. 2 (Example 1) because it is *independent* of the y_t 's.

The last column of Table 3 shows the posterior probability, $\alpha_{t,1}$, for y_t being generated from the “good run”. Note that apart from $\alpha_{11,1} = 0$ (as expected), there are some other observations that are classified as outlying observations (e.g., y_{20}), using $\alpha_{t,2} = 1 - \alpha_{t,1}$.

Note: In general, the likelihood function for the new observation y_t is determined as follows. We have that $y_t = a_t\theta_t + \xi_t$, and $\xi_t - \sum_{i=1}^p \phi_i \xi_{t-i} = \varepsilon_t - \sum_{i=1}^q \psi_i \varepsilon_{t-i}$ where $\varepsilon_t \sim \alpha_{t,1} \mathbf{N}(0, \sigma_{t,1}^2) + \alpha_{t,2} \mathbf{N}(0, \sigma_{t,2}^2)$. Then the likelihood based on y_t is

$$\begin{cases} \mathbf{N}(\theta_t, \gamma_0^{(1)}) & \text{with probability } \alpha_{t,1}, \\ \mathbf{N}(\theta_t, \gamma_0^{(2)}) & \text{with probability } \alpha_{t,2} (= 1 - \alpha_{t,1}), \end{cases}$$

where $\gamma_0^{(i)} = \text{Cov}(Y_t, Y_{t-0}) = \text{Var}(Y_t)$ evaluated via the assumption of stationarity. Thus, we have

$$M_{t,i} = \gamma_0^{(i)} + v_{t|t-1}.
 \tag{24}$$

Table 3
Comparison of two filters (Example 2)

<i>t</i>	<i>y_t</i>	Mean (<i>μ_t</i>)		Variance (<i>V_t</i>)		<i>α_{t,1}</i>
		<i>μ_{std}</i>	<i>μ_{new}</i>	<i>V_{std}</i>	<i>V_{new}</i>	
1	8.74	10.07	10.32	5.32	6.31	0.985
2	6.11	8.44	9.43	3.71	0.06	0.984
3	10.04	8.99	9.43	3.09	0.21	0.989
4	11.52	9.78	9.29	2.81	0.21	0.988
5	14.07	11.06	9.91	2.68	0.20	0.983
6	15.12	12.24	10.65	2.61	0.20	0.981
7	6.35	10.55	10.54	2.58	0.20	0.984
8	4.66	8.88	10.40	2.56	0.19	0.978
9	15.88	10.86	10.55	2.55	0.19	0.980
10	20.01	13.45	10.77	2.55	0.13	0.931
11	65.00	28.02	11.48	2.54	0.05	0.000
12	− 2.69	19.34	11.43	2.54	0.04	0.542
13	11.26	17.06	16.40	2.54	0.17	0.989
14	20.66	18.08	15.26	2.54	0.18	0.984
15	6.46	14.80	14.36	2.54	0.15	0.946
16	1.12	10.93	14.29	2.54	0.07	0.677
17	12.02	11.24	13.55	2.54	0.18	0.988
18	24.72	15.05	13.58	2.54	0.08	0.862
19	10.41	13.74	12.85	2.54	0.19	0.987
20	− 5.28	8.37	12.78	2.54	0.05	0.071
21	− 1.59	5.56	12.51	2.54	0.05	0.511
22	17.83	9.02	11.99	2.54	0.15	0.980
23	23.56	13.13	12.17	2.54	0.07	0.837
24	4.68	10.74	11.92	2.54	0.15	0.965
25	− 1.50	7.28	11.77	2.54	0.05	0.654
26	11.29	8.42	11.29	2.54	0.19	0.989
27	17.24	10.91	10.90	2.54	0.17	0.977
28	6.10	9.55	10.42	2.54	0.20	0.983
29	6.42	8.67	10.72	2.54	0.20	0.985
30	18.76	11.52	10.92	2.54	0.16	0.958

The unobservable $\theta = 10$.

In the AR(1) process, for example, this is

$$M_{t,1} = \frac{\sigma_{t,1}^2}{1 - \phi} + V_{t|t-1} \quad \text{and} \quad M_{t,2} = \frac{\sigma_{t,2}^2}{1 - \phi} + V_{t|t-1}. \tag{25}$$

Once we collapse the mixture likelihood function, the assumption of homogeneity of variance is lost, causing non-stationarity. The exact form of Σ_t is complicated. However, computationally, for any ARMA(*p, q*) process, write

$$A\xi_t = B\varepsilon_t$$

with appropriate values in the square matrices **A** and **B**. Then $\xi_t = A^{-1}B\varepsilon_t$ leads to

$$\Sigma_t = \text{Cov}(y_t | \theta_t) = \text{Cov}(\xi_t) = B^{-1}A \text{Cov}(\varepsilon_t) (B^{-1}A)^{-1}.$$

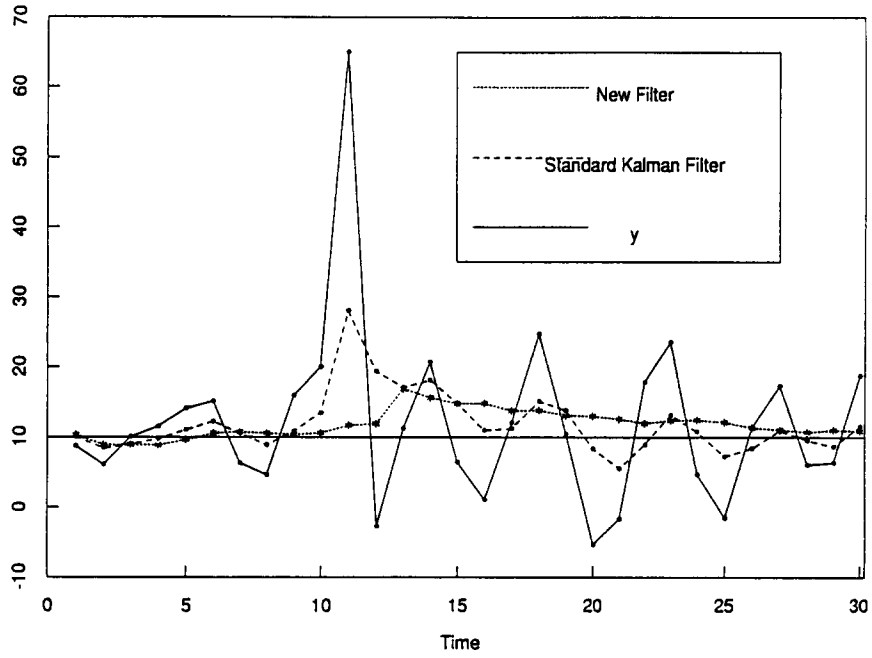


Fig. 3. Comparison plot for Example 2

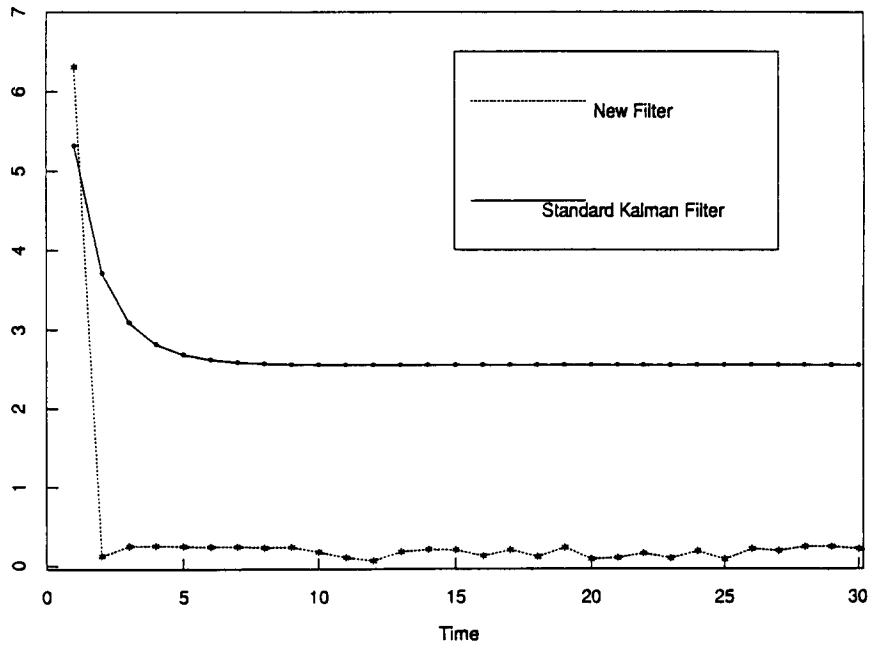


Fig. 4. Comparison of variance in Example 2

For the AR(2) case, we have $\mathbf{B} = \mathbf{I}$, and

$$\mathbf{A} = \begin{bmatrix} c_1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ c_2 & c_3 & 0 & 0 & \dots & 0 & 0 & 0 \\ -\phi_2 & -\phi_1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\phi_2 & -\phi_1 & 1 & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & & -\phi_1 & 1 & 0 \\ 0 & 0 & 0 & 0 & & -\phi_2 & -\phi_1 & 1 \end{bmatrix},$$

where the choice of c_i 's depends upon the initial assumption. One popular approach is to take $c_1 = c_3 = 1$, and $c_2 = -\phi_1$; namely, assuming $\xi_{-1} = \xi_0 = 0$. Here, however, to be consistent with the previous example, we take

$$c_1 = \left[\frac{1 + \phi_2}{1 - \phi_2} [(1 - \phi_2)^2 - \phi_1^2] \right]^{1/2},$$

$$c_2 = -\phi_1 \left[\frac{1 + \phi_2}{1 - \phi_2} \right]^{1/2},$$

$$c_3 = [1 - \phi_2^2]^{1/2}$$

as used in Example 1. This is derived via solving the inverse matrix of Σ_t in (20), assuming stationarity. Also, note that $\text{Cov}(\varepsilon_t)$ is a diagonal matrix with its diagonal elements σ_t^2 's (these σ_t^2 's are not necessarily the same).

7. Remarks

In this paper, we apply the Lin and Guttman robust filter to the correlated observations when the correlation structure is known. We use the popular ARMA(p, q) structure to illustrate our core idea. By taking the advantage of the given correlation structure, it is shown that such a scheme is robust to outliers. Recently, there have been several articles on Kalman filtering with correlated data, most of these *Signal Processing*, where the phenomenon is referred to as "colored noise". See, for example, Balakrishnan (1984), Anderson and Moore (1979), and Gibson et al. (1991). These methods mainly deal with the correlation structure, but not with outliers. Thus, as expected, they do not perform well when the outliers are present.

When the correlation structure is not precisely known, the outlier problem is very difficult. The performance of the Lin and Guttman filter will depend on the "distance" between the correlation matrix used and the "true" correlation matrix. The former is generally calculated from some reference signal, while the later is unknown.

Acknowledgements

We thank one referee for helpful comments on an earlier version of this article. Irwin Guttman was partially supported by Grant A8743 from NSERC (Canada), and Dennis K.J. Lin was partially supported by the Professional Development Award, the University of Tennessee and by the National Science Foundation under Grant DMS-9204007.

References

- Abraham, B. and G.E.P. Box (1979), Bayesian analysis of some outlier problems in time series, *Biometrika* **66**, 229–236.
- Anderson, B.D.O. and J.B. Moore (1979), *Optimal Filtering* (Prentice-Hall, Englewood Cliffs, NJ).
- Balakrishnan, A.V. (1984), *Kalman Filtering Theory* (Optimization Software, Inc., New York).
- Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis Forecasting and Control* (Holden-Day, San Francisco, 2nd ed.).
- Box, G.E.P. and G.C. Tiao (1968), A Bayesian approach to some outlier problems, *Biometrika* **55**, 119–129.
- Gibson, J.D., B. Koo and S.D. Gray (1991), Filtering of colored noise for speech enhancement and coding, *IEEE Trans. Signal Processing* **39**, 1732–1741.
- Guttman, I. and D. Peña (1985), Robust Kalman filtering and its applications, Tech. Report No. 85-1, Department of Statistics, University of Toronto.
- Harrison, P.J. and C.F. Stevens (1976), Bayesian Forecasting (with discussion), *J. Roy. Statist. Soc. Ser. B* **38**, 205–247.
- Kalman, R.E. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME* **82D**, 35–47.
- Lin, D.K.J. and I. Guttman (1993), Handling spuriousity in the Kalman filter, *Statist. Probab. Lett.* **16**, 259–268.
- 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.
- Tiao, G.C. and M.M. Ali (1971), Analysis of correlated random effects: linear model with two random components, *Biometrika* **58**, 37–51.