1987-11

Robustifying the Kalman filter

Gaver, Donald Paul; Jacobs, Patricia A.

Monterey, California. Naval Postgraduate School

NPS55-87-014

NAVAL POSTGRADUATE SCHOOL

Monterey, California

ROBUSTIFYING THE KALMAN FILTER

D. P. GAVER
P. A. JACOBS

NOVEMBER 1987

Approved for public release; distribution unlimited.

Prepared for:
Chief of Naval Research
Arlington, VA 22314

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

The work reported herein was supported in part with funds provided by the Chief of Naval Research, Arlington, VA.

This report was prepared by:

REPORT DOCUMENTATION PAGE

REPORT SECURITY CLASSIFICATION: NCLASSIFIED	1b RESTRICTIVE MARKINGS
SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
DECLASSIFICATION DOWNGRADING SCHEDULE	
PERFORMING ORGANIZATION REPORT NUMBER(S): NPS55-87-014	5 MONITORING ORGANIZATION REPORT NUMBER(S)
NAME OF PERFORMING ORGANIZATION: val Postgraduate School — 6b OFFICE SYMBOL (If applicable): Code 55	7a NAME OF MONITORING ORGANIZATION
ADDRESS (City, State, and ZIP Code): nterey, CA 93943-5000	7b ADDRESS (City, State, and ZIP Code)
NAME OF FUNDING SPONSORING ORGANIZATION: Office of the hief of Naval Research — 8b OFFICE SYMBOL (If applicable): 01122	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N0001487WR24005
ADDRESS (City, State, and ZIP Code): rlington, VA 22217-5000	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 61153N; PROJECT NO RR014-05-01; TASK NO 4114531-02; WORK UNIT ACCESSION NO

TITLE (Include Security Classification)

OBUSTIFYING THE KALMAN FILTER

PERSONAL AUTHOR(S)

aver, D. P. and Jacobs, P.A.

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year Month Day)	15 PAGE COUNT
echnical	FROM TO	Nov 87	20

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Kalman filter; student-t measurement errors; Iterative re-weighting procedure; Non-linear Filter; Biweight; Robust Estimation

ABSTRACT (Continue on reverse if necessary and identify by block number)

alman filters are tracking and prediction algorithms based on Gaussian measurement errors nd structural models. The Kalman filter performance may degrade if the measurement errors ome from a thicker-tailed-than Gaussian distribution. In this report non-linear procedures re described which are based on Kalman-type models, but work with student-t measurement rrors. This is an initial paper intended to report an approach; extensions are under evelopment. Comments are welcome.

Note: Since this report was finished comments were received by Mike West that clarify and should improve upon his approximation of our Section 3, and upon ours as well. More work is in progress.

DISTRIBUTION AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
NAME OF RESPONSIBLE INDIVIDUAL: D. P. Gaver	22b TELEPHONE (Include Area Code): (408)646-2605 — 22c OFFICE SYMBOL: Code 55Gv

ROBUSTIFYING THE KALMAN FILTER;

Protection Against Symmetrically Straggling Measurement Errors

D. P. Gaver
P. A. Jacobs

1. INTRODUCTION.

Tracking and prediction algorithms based on simple Gaussian (normal distribution) measurement errors and structural models are commonly used in practice under the name of KALMAN Filters. If (a) measurement errors are not suitably Gaussian, e.g., if occasional outliers occur or (b) true structural behavior is not simple, perhaps displaying apparently discontinuous behavior caused by unfavorable sensor-target orientation, then traditional filter performance may dramatically degrade. In this paper, we will propose and study procedures based on an elaborated model of the KALMAN-type but with the measurement errors coming from a family of possibly suitable non-Gaussian distributions (e.g., Student-t) to represent, and suitably compensate for more-thick-tailed-than-Gaussian measurement error, i.e., distributions with long straggling tails having the tendency to produce symmetric outliers.

In particular the basic stochastic model considered here is

$$\theta_n = \theta_{n-1} + \omega_n \tag{1.1}$$

$$Y_n = \theta_n + \varepsilon_n \tag{1.2}$$

where $\{\omega_n\}$ are independent normal/Gaussian random variables with mean 0 and variances $\{\tau_n\}$ and $\{\varepsilon_n\}$ are independent random variables having mean 0. The random variable θ_n is unobservable. The random variable Y_n is interpreted as the observation of θ_n made with measurement error ε_n; ε_n is not Gaussian, but controllably long-tailed. The problem is to estimate θ_n from $Y_1,..., Y_n$ in the simple recursive fashion that characterizes the classical KALMAN filter. Expression (1.1) is a simple random walk and does not represent very interesting dynamics, but does provide suggestive illustrations.

In the next, or second, section, we will describe a procedure, the ALMA (standing for KALMAN with outliers suppressed), which is based on a model in which the components of the error sequence $\{\varepsilon_n\}$ have a Student-t distribution.

In the third section, the traditional KALMAN procedure will be described. It is based on the assumption that components of $\{\varepsilon_n\}$ have iid normal distributions. Finally, a robust procedure due to West [1981] will be described.

In section 4 results of an extensive simulation experiment will be presented and discussed. The simulation experiment compares the various procedures. The results indicate that the ALMA procedure is significantly better than the KALMAN when the true measurement error distribution is Student-t. Further, there is not much lost in using the ALMA procedure instead of the KALMAN when the true measurement error distribution is normal.

2. THE ALMA FILTER AND RELATED PROCEDURES.

While many measurement errors of physical quantities are approximately normal, especially "in the middle" of their distribution, there can well be thicker-than-normal/Gauss tails and also occasional extreme outliers; that these can have seriously degrading effects in regression-like problems has been the subject of considerable research; we cite books by Mosteller and Tukey (1977), Huber (1981), Hampel (1986); in the time-series context the article by Martin and Yohai (1986), which contains many references; also lately the articles by West and his associates (1981,1985); it is to West's approach that our methodology should best be compared.

One way to model these features is to extend the tails of the normal by continuous scale mixing. Such an approach can lead to the Student-t form, and to many other useful forms as well. We will assume here that $\{\varepsilon_n\}$ are independent random variables, now having in the Student-t distribution with mean 0, scale σ_n (*not* the standard deviation) and d degrees of freedom; that is,

$$p_{\varepsilon_n}(u) = c(d)\,\frac{1}{\sigma_n}\left[1 + \left(\frac{u}{\sigma_n}\right)^2 \frac{1}{d}\right]^{-\frac{d+1}{2}}. \tag{2.1}$$

Let y_i denote the i^{th} measurement and $y^n = (y_1, ..., y_n)$. Assume that $\theta_{n-1}| y^{n-1}$ has a normal distribution with mean m_{n-1} and variance C_{n-1}. Since ω_n is assumed to have a normal distribution with variance τ_n, $\theta_n| y^{n-1}$ has a normal distribution with mean m_{n-1} and variance $C_n^{\#}=C_{n-1}+\tau_n$. Thus, from (1.1), (1.2), and (2.1)

$$P\{\theta_n \in d\theta, Y_n \in dy \mid Y_1=y_1,\ldots,Y_{n-1}=y_{n-1}\}$$

$$= K \exp\left\{-\frac{1}{2}\frac{(\theta-m_{n-1})^2}{C_n^{\#}} - \frac{1}{2}(d+1)\ln\left[1+\left(\frac{\theta-y}{\sigma_n}\right)^2\frac{1}{d}\right]\right\}d\theta dy \qquad (2.2)$$

$$\approx K \exp\left[-\frac{1}{2}\frac{(\theta-\mu(y))^2}{C(y)} + \frac{1}{2}Q(y)\right]d\theta dy$$

where the approximation replaces the expression in the exponent by an approximating quadratic in θ.

2.1 The ALMA Procedure.

The ALMA procedure provides a Gaussian approximation to the distribution of $\theta_n \mid y^n$, but one that emphatically differs from the classical linear-in-observations form. Following an argument in Gaver et al. [1986], differentiate both sides of (2.2) with respect to θ to obtain

$$\frac{\theta-\mu(y)}{C(y)} = \frac{\theta-m_{n-1}}{C_n^{\#}} + \frac{d+1}{d}\,\frac{\theta-y}{\sigma_n^2}\,\frac{1}{1+\left(\frac{\theta-y}{\sigma_n}\right)^2\frac{1}{d}}. \qquad (2.3)$$

Equating the terms involving θ results in the following equation:

$$\theta: \frac{1}{C(y)} = \frac{1}{C_n^{\#}} + w(y)\frac{1}{\sigma_n^2} \qquad (2.4)$$

where the *weight*

$$w(y) = \frac{d+1}{d} \frac{1}{1 + \left(\frac{\theta - y}{\sigma_n}\right)^2 \frac{1}{d}} . \tag{2.5}$$

Furthermore, equating the constant terms results in

$$\frac{\mu(y)}{C(y)} = \frac{m_{n-1}}{C_n^{\#}} + w(y)\frac{y}{\sigma_n^2} . \tag{2.6}$$

The ALMA procedure approximates $\theta_n \mid y^n$ by the normal distribution having mean

$$m_n = C_n \left[m_{n-1} \frac{1}{C_n^{\#}} + w(y_n) y_n \frac{1}{\sigma_n^2} \right] \tag{2.7}$$

and variance

$$C_n = \left[\frac{1}{C_n^{\#}} + w(y_n) \frac{1}{\sigma_n^2} \right]^{-1} \tag{2.8}$$

where

$$w(y_n) = \frac{d+1}{d} \frac{1}{1 + \left(\frac{\theta - y_n}{\sigma_n}\right)^2 \frac{1}{d}} . \tag{2.9}$$

Note that the weight $w(y_n)$ involves the unknown θ. One implementation uses approximate weights of the form

$$w_k(y_n) = \frac{d+1}{d} \frac{1}{1 + \left(\frac{y_n - m_{n-1}}{\sigma_n}\right)^2 \frac{k}{d}} . \tag{2.10}$$

When k=l, m_{n-1} is used in place of θ in (2.9).

When $k=\frac{1}{4}$, $0.5(m_{n-1} + y_n)$ is used in place of θ.

The basic ALMA procedure is to evaluate $w_k(y_n)$ and then use it to find

$$C_n = \left[\frac{1}{C_n^{\#}} + w_k(y_n)\frac{1}{\sigma_n^2}\right]^{-1} \tag{2.11}$$

and

$$m_n = C_n\left(\frac{m_{n-1}}{C_n^{\#}} + \frac{w_k(y_n)y_n}{\sigma_n^2}\right). \tag{2.12}$$

The point estimate of θ_n given y^n is $\theta_n = m_n$ and an estimate of the variance of θ_n is C_n. Thus the procedure provides a particular *Gaussian* posterior approximation. In other similar contexts, non-linear filters for example, it has been suggested that the procedure (2.10) - (2.12) be iterated with the newly-computed m_n, replacing m_{n-1} in (2.10) - (2.12) in each iteration. In the simulations 0, 1 and 2 iterations were implemented, and the results compared.

2.2 The Biweight.

The ALMA procedure is an iterative reweighting procedure. In the ordinary regression context another weight has been suggested: the so-called (Tukey) biweight, cf. Mosteller and Tukey (1977). In our context, the *biweight* procedure can replace the weight $w_k(y)$ in the ALMA procedure with the biweight

$$w_B(y)=\begin{cases}\left[1 - k\left((y-m_{n-1})\left[a\sigma_n\sqrt{\frac{d}{d-2}}\right]^{-1}\right)^2\right]^2 & \text{if } k\left((y-m_{n-1})\left[a\sigma_n\sqrt{\frac{d}{d-2}}\right]^{-1}\right)^2<1 \\ 0 & \text{otherwise.}\end{cases} \tag{2.13}$$

The variance of a Student-t distribution with d degrees of freedom and scale σ is $\sigma^2\frac{d}{d-2}$ if $d>3$, otherwise being infinite. Hence the (bi)weight $w_B(y)$ uses the measurement y if $|y|$ is within a standard deviations of m_{n-1}, the estimate of θ_{n-1}. The weight is zero if the deviation is greater.

As was done in the basic ALMA procedure, 0, 1, and 2 iterations of (2.10)–(2.12) were tried, with $w_B(y_n)$ replacing $w_k(y_n)$, for values of a=5, 7, 9 and k=1, 0.25.

2.3 Aspects of the Likelihood Procedure.

It is possible for the likelihood function (2.2) to exhibit two local θ-maxima. In such a case, the likelihood procedure approximates the local maxima and chooses the one which globally maximizes the likelihood.

To examine the details let

$$f(\theta) = \frac{d}{d\theta} \ln P\{\theta_n \in d\theta, Y_n \in dy\}$$

$$= \left(\frac{-(\theta - m_{n-1})}{C_n^{\#}} + \frac{d+1}{d} \left[1 + \left(\frac{y-\theta}{\sigma_n} \right)^2 \frac{1}{d} \right]^{-1} \frac{y-\theta}{\sigma_n^2} \right) d\theta dy. \tag{2.14}$$

Now it is clearly possible for f(θ)=0 to have multiple roots. To be specific, f(θ)=0 for those θ satisfying

$$0 = \theta^3 + \theta^2(-2y - m_{n-1}) \tag{2.15}$$

$$+ \theta \left[\sigma_n^2 d + y^2 + (d+1) C_n^{\#} + 2y m_{n-1} \right]$$

$$+ \left[-m_{n-1} \sigma_n^2 d - m_{n-1} y^2 - (d+1) y C_n^{\#} \right].$$

The properties of this cubic-in-θ equation can be deduced from classical results.

Let

$$D = \left(\frac{y-m_{n-1}}{\sigma_n}\right)^4 \tag{2.16}$$

$$+ \left(\frac{y-m_{n-1}}{\sigma_n}\right)^2 \left\{ 2d^2 - 5d(d+1)\frac{C_n^{\#}}{\sigma_n^2} - \frac{1}{4}(d+1)^2\left[\frac{C_n^{\#}}{\sigma_n^2}\right]^2 \right\}$$

$$+ \left[d + (d+1)\frac{C_n^{\#}}{\sigma_n^2} \right]^3 ;$$

then if

D>0 (2.15) has 1 real root and two conjugate imaginary roots;

D=0 (2.15) has 3 real roots, at least two of which are equal;

D<0 (2.15) possesses 3 real and unequal roots.

Note that if $d=\infty$ so that ε_n has a normal distribution, then certainly D>0 and (2.2) has a unique maximum. If $d<\infty$ and $C_n^{\#}\sigma_n^{-2}$ is small enough, then D>0 and once again (2.2) will have a unique maximum. If $d<\infty$ and $C_n^{\#}\sigma_n^{-2}$ is large enough (actually, larger than $\frac{(6\sqrt{3}-10)d}{d+1} \approx \frac{4d}{d+1}$), then D<0 for an interval of values of $(y-m_{n-1})^2$ and (2.15) will have 3 real unequal roots; in this case (2.2) will have two local maxima.

The likelihood procedure computes D. If $D \geq 0$ it uses the ALMA procedure with weight

$$w_k(y) = \frac{d+1}{d}\left[1+\left(\frac{y-m_{n-1}}{\sigma_n}\right)^2 \frac{k}{d}\right]^{-1} \tag{2.17}$$

to compute θ_n. If $D<0$, then two candidate estimates θ_1, and θ_2 of θ are computed. Both estimates are obtained via the ALMA procedure (2.7)-(2.9). One approximates weight (2.9) by setting $\theta = m_{n-1}$ as in (2.10); think of the result as *prior-dominated*. The other approximates weight (2.9) by setting $\theta = y$, so that $w(y) = \frac{d+1}{d}$; the result is *data-determinated*. The likelihood function is then evaluated at each value of $\theta: \theta_1$ and θ_2. The quoted estimate of θ_n is set equal to the θ_i that comes closest to maximizing the global likelihood; the estimate of the variance is set equal to the corresponding C_n.

3. THE KALMAN AND WEST PROCEDURES.

In this subsection, the traditional KALMAN procedure will be described for the model (1.1)-(1.2). A procedure proposed by West (1981) will also be discussed.

3.1 The KALMAN Procedure.

The KALMAN filter finds the estimate $\hat{\theta}_n$ of θ_n which minimizes the conditional mean square error of $(\hat{\theta}_n - \theta_n)$ given y^n. If $\{\varepsilon_n\}$ are independently normally distributed with mean 0 and variances $\{\gamma_n\}$, then the KALMAN filter can be viewed as a Bayesian updating procedure; see Meinhold and Singpurwalla (1983).

The Bayesian KALMAN procedure assumes $\theta_{n-1}|y^{n-1}$ is normal with mean m_{n-1} and variance C_{n-1}. Thus, from (1.1) $\theta_n|y^{n-1}$ is normal with mean m_{n-1} and variance $C^{\#}_n = C_{n-1} + \tau_n$. From (1.2)

$$P(\theta_n \in d\theta, Y_n \in dy \mid y^{n-1}) = K \exp\left[-\frac{1}{2}\frac{(\theta_n - m_{n-1})^2}{C^{\#}_n} - \frac{1}{2}\frac{(y-\theta_n)^2}{\gamma_n}\right] d\theta dy \tag{3.1}$$

$$= \tilde{K} \exp\left\{ -\frac{1}{2}\left[\frac{1}{C_n^{\#}} + \frac{1}{\gamma_n}\right]\left[\theta - \left(\frac{m_{n-1}}{C_n^{\#}} + \frac{y}{\gamma_n}\right)\left(\frac{1}{C_n^{\#}} + \frac{1}{\gamma_n}\right)^{-1}\right]^2 \right\} d\theta dy. \tag{3.2}$$

Thus $\theta_n | y^n$ has a normal distribution with mean

$$m_n = C_n\left[\frac{m_{n-1}}{C_n^{\#}} + \frac{y_n}{\gamma_n}\right] \tag{3.3}$$

and variance

$$C_n = \left[\frac{1}{C_n^{\#}} + \frac{1}{\gamma_n}\right]^{-1}. \tag{3.4}$$

The estimate of θ_n given y^n is then

$$\hat{\theta}_n = m_n \tag{3.5}$$

and an estimate of the variance of θ_n is C_n.

Comparing (3.3)-(3.4) with (2.10)-(2.12) indicates that, if y_n is close to m_{n-1}, then the ALMA procedure will closely resemble the KALMAN. In particular, if $\gamma_n = \sigma_n^2$ and $d \to \infty$, the 2 estimators are identical. However, if y_n is far from m_{n-1}, then the ALMA procedure will tend to discount that observation, relying on its estimate of θ_{n-1} to strongly influence its estimate of θ_n. This behavior implies that the ALMA procedure will be less quickly responsive to changes in the values of θ_n than will be the KALMAN. This is the price paid for robustness to outlying measurement errors: KALMAN treats *all* changes in observations as representative of structural (θ_n) changes; ALMA is more tentative. Of course ALMA may be tuned towards KALMAN by increasing the d-value.

3.2 The West Procedure.

West proposes an estimation procedure for θ_n given y^n in the case in which the density p_{ε_n} is symmetric about 0. In the special case in which p_{ε_n} is normal, West's procedure reduces to the KALMAN filter.

Once again, assume $\theta_{n-1} \mid y^{n-1}$ is normal with mean m_{n-1} and variance C_{n-1} so that $\theta_n | y^{n-1}$ is normal with mean m_{n-1} and variance $C_n^{\#} = C_{n-1} + \tau_n$.

$$P\{\theta_n \in d\theta,\ Y_n \in dy \mid Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}\}$$

$$= K \exp\left[-\frac{1}{2}(\theta - m_{n-1})^2 \frac{1}{C_n^{\#}} + \ln p_{\varepsilon_n}(y - \theta) \right] d\theta dy \qquad (3.6)$$

$$\approx K \exp\left[-\frac{1}{2}(\theta - m_{n-1})^2 \frac{1}{C_n^{\#}} + \left(\ln p_{\varepsilon_n}(y - m_{n-1}) + g(y - m_{n-1})(\theta - m_{n-1}) - G(y - m_{n-1}) \frac{(\theta - m_{n-1})^2}{2} \right) \right] d\theta dy \qquad (3.7)$$

where a Taylor expansion provides

$$g(u) = \frac{-d}{du} p_{\varepsilon_n}(u) \qquad (3.8)$$

$$G(u) = \frac{-d^2}{du^2} p_{\varepsilon_n}(u). \qquad (3.9)$$

Completing the square in (3.7) results in

$$P\,[\theta_n \in d\theta, Y_n \in dy \mid Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}]$$

$$\approx \bar{K} \exp\left\{ -\frac{1}{2}\left[\frac{1}{C_n^{\#}} + G(y - m_{n-1}) \right] \left((\theta - m_{n-1}) - g(y - m_{n-1}) \left[\frac{1}{C_n^{\#}} + G(y - m_{n-1}) \right]^{-1} \right)^2 \right\}. \qquad (3.10)$$

Hence, $P\{\theta_n \in d\theta \mid Y_1 = y_1, \ldots, Y_n = y_n\}$ is approximated by a normal distribution having mean

$$m_n = m_{n-1} + C_n g(y_n - m_{n-1}) \qquad (3.11)$$

and variance

$$C_n = \left[\frac{1}{C_n^{\#}} + G(y_n - m_{n-1})\right]^{-1} . \tag{3.12}$$

In the special case in which ε_n has a Student-t distribution with d degrees of freedom and scale parameter σ_n,

$$p_{\varepsilon_n}(u) = c(d)\frac{1}{\sigma_n}\left[1 + \left(\frac{u}{\sigma_n}\right)^2 \frac{1}{d}\right]^{-\frac{(d+1)}{2}} \tag{3.13}$$

$$g(u) = \frac{d+1}{d}\left[1 + \left(\frac{u}{\sigma_n}\right)^2 \frac{1}{d}\right] \frac{u}{\sigma_n^2} \tag{3.14}$$

and

$$G(u) = \frac{d+1}{d}\frac{1}{\sigma^2}\left[1 + \left(\frac{u}{\sigma}\right)^2 \frac{1}{d}\right]^{-2}\left[1 - \left(\frac{u}{\sigma}\right)^2 \frac{1}{d}\right] . \tag{3.15}$$

Since $G(y_n - m_{n-1})$ is playing the role of a variance in (3.10), but may become embarrassingly negative for large u, West suggests that it be replaced by $\max(0, G(y_n - m_{n-1}))$; this step has been taken in the simulations that illustrate the various procedures proposed here. West suggests another possibility in West et al. (1985).

4. A SIMULATION EXPERIMENT.

All simulations were carried out on an IBM 3033AP computer at the Naval Postgraduate School. Random numbers were generated using the LLRANDOMII random number package; cf. Lewis and Uribe (1981).

For each replication of the simulation the model of (1.1)-(1.2) is generated for n=0,1,...,100. In the simulations reported below $\{\omega_n\}$ are iid normal with mean zero and variance one. For each replication, estimates $\hat{\theta}_n$ of θ_n given y^n are computed using each of the procedures described above. The data collected are the estimation error $\hat{\theta}_n - \theta_n$ for n=25, 50, 75, 100 and the estimate of variance C_n, n=25, 50, 75, 100. The number of independent replications is 1000.

Tables 1 and 2 report results of the KALMAN and ALMA procedures for simulations in which $\{\varepsilon_n\}$ are iid normal with mean zero and variance one. The ALMA procedure actually uses the incorrect measurement error model that $\{\varepsilon_n\}$ are iid Student-t with d=3 degrees of freedom and variance equal to one. Results for the ALMA procedure are shown for weights as in (2.10), for k=1.0 and k=0.25. The procedure was iterated 0, 1, and 2 times.

Table 1 shows statistics of $\hat{\theta}_n - \theta_n$ for n=25, 50, 75, 100. As anticipated, the KALMAN procedure which uses the correct (normal) model exhibits the smallest variance of $\hat{\theta}_n - \theta_n$. The ALMA procedure with k=0.25 and 0 iterations and the ALMA procedure with k=l and 1 iteration have the smallest variances for the ALMA procedures.

Table 2 exhibits the estimates of the variance of θ_n, namely C_n, for the ALMA procedure for n=25, 50, 75, 100. The KALMAN estimate of the variance is the constant 0.618 for all of these n. This constant is the limiting solution to equation (3.4) with $\tau_n = \gamma_n = 1$; that is, with $C = \lim_{n \to \infty} C_n$

$$C = \frac{1}{\frac{1}{C+1} + 1}$$

a simple quadratic with appropriate solution

$$C = \frac{1+\sqrt{5}}{2} = 0.618.$$

The variance of $\hat{\theta}_n - \theta_n$ for the KALMAN procedure in Table 1 is close to the calculated 0.618.

The mean values of C_n for the ALMA procedure with k=0.25 and 0 iterations and k=1 with 1 iteration are about half that of the corresponding variances of $\hat{\theta}_n - \theta_n$ in Table 1.

Tables 3-4 report results for a simulation in which $\{\varepsilon_n\}$ are iid Student–t with 3 degrees of freedom and variance equal to 1. Table 3 reports statistics of the estimation error, $\hat{\theta}_n$-θ_n, for the KALMAN, ALMA, Biweight, Likelihood, and West procedures. As usual, the KALMAN procedure assumes $\{\varepsilon_n\}$ are iid normal with mean 0 and variance 1. The other procedures assume $\{\varepsilon_n\}$ are iid Student–t with 3 degrees of freedom and variance equal to 1. The ALMA procedure with k=.25 and no iterations exhibits the smallest variance of $\hat{\theta}_n - \theta_n$. The more complicated Likelihood procedure with k=0.25 and no iterations exhibits the next-smallest variance. The ALMA with k=1 and 1 iteration exhibits the third smallest variance.

The Biweight procedure was implemented with the constants in the weight (2.13) a=5,7,9 and k=.25 and 1; the procedure was iterated 0,1, and 2 times. The results for a=5 were much worse than those for a=7 and 9 indicating that a=5 is not large enough to suppress outlying values; they are not reported. Iterating the biweight procedure 1 and 2 times did not improve the results for any values of a. The results of Table 3 indicate that the biweight procedure with the smallest variance uses k=1.0 and a=7 with no iterations.

The West procedure described in West (1981) as currently implemented does not do as well as the KALMAN. The statistics of C_n in Table 4 seem to indicate that the difficulty is with the estimate of variance, C_n; the fix for negative $G(y-m_{n-1})$ makes it possible for C_n to increase by one in successive times over long periods of time.

Table 4 exhibits the statistics of C_n. The KALMAN procedure, the ALMA procedure with k=0.25 and 0 iterations, the ALMA procedure with k=1 and 1 iteration, the Likelihood procedure with k=0.25 and 0 iterations and the Biweight with k=1, a=7 all have mean C_n approximately half the variance of $\hat{\theta}_n$–θ_n.

5. CONCLUSIONS.

The simulation results obtained to date indicate that a satisfactory robust KALMAN=ALMA procedure utilizes the k=0.25 weight-starting option and requires no iteration. While the above filter is about 7% *less* efficient than the KALMAN when measurement errors are ideally Gaussian, it is about 6% *more* efficient when errors are long-tailed non-Gaussian; efficiency is in terms of ratios of (estimated) variances and is not the only meaningful criterion.

Examination of Table 3 reveals through values of skewness, and kurtosis , that as anticipated, the robust ALMA estimation errors are substantially more closely Gaussian than are the corresponding KALMAN products when measurement errors are Student-t.

REFERENCES

R. G. Brown. *Introduction to Random Signal Analysis and KALMAN Filtering.* John Wiley and Sons New York, 1983.

D. P. Gaver, P. A. Jacobs, I. O'Muircheartaigh, A. Meldrum. *Problems of Identification..* Naval Postgraduate School Technical Report. NPS55-86-021, Monterey, CA.

R. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York, 1986.

P. J. Huber. *Robust Statistics.* Wiley, New York, 1981.

P. A. W. Lewis and L. Uribe. *The New Naval Postgraduate School random number package - LLRANDOMII.* Naval Postgraduate School Technical Report NPS55-81-005. Monterey, CA

R. D. Martin and V. J. Yohai. Influence Functionals for Time Series (with discussion). Ann. Statist. 14 (1986).

R. J. Meinhold and N. D. Singpurwalla. Understanding the KALMAN Filter. *The American Statistician 37. (1983)* pp. 123-127.

F. Mosteller and J. W. Tukey. *Data Analysis and Regression.* Addison - Wesley, 1977.

M. West. Robust sequential approximate Bayesian estimation. *J. R. Statist. Soc.* B 43 (1981) pp. 157-166.

M. West, P. J. Harrison, and H. S. Migon. Dynamic generalized linear models and Bayesian forecasting. *JASA 80* (1985) 73-83.

Table 1

Statistics of $\hat{\theta}_n - \theta_n$

Normal Measurement Errors with Variance 1

Time n:			25				50				75				100			
Proc	Nbr Iter	k	M	V	S	K	M	V	S	K	M	V	S	K	M	V	S	K
K	-	-	0.00	0.61	-0.04	0.01	0.02	0.62	-0.01	-0.15	-0.03	0.63	0.02	-0.29	0.00	0.60	-0.05	-0.19
A	0	1.0	-0.02	0.91	-0.07	0.60	0.03	0.77	0.21	1.40	-0.01	0.78	0.01	0.18	0.01	0.84	-0.19	0.78
		0.25	0.00	0.65	-0.02	-0.07	0.03	0.65	0.04	0.03	-0.02	0.64	0.01	-0.14	0.00	0.67	-0.05	-0.17
A	1	1.0	0.01	0.70	-0.05	-0.02	0.04	0.69	0.15	0.62	-0.01	0.68	-0.00	0.02	0.03	0.73	-0.06	-0.01
		0.25	0.02	0.70	0.00	-0.07	0.03	0.76	0.02	-0.09	-0.03	0.74	0.04	-0.10	0.00	0.75	0.02	-0.30
A	2	1.0	0.01	0.71	0.02	-0.09	0.04	0.75	0.06	0.06	-0.02	0.72	0.01	-0.04	0.00	0.76	-0.01	-0.25
		0.25	0.02	0.77	-0.01	-0.11	0.02	0.83	0.02	-0.10	-0.03	0.82	0.04	-0.09	0.00	0.81	0.04	-0.31

Procedure (Proc.)

K = KALMAN

A = ALMA

Statistics

M = Mean

V = Variance

S = Skewness

K = Kurtosis

Table 2
Statistics of C_n
Normal Measurement Errors with Variance 1

Time n:			25		50		75		100	
Proc	Nbr Iter	k	M	V	M	V	M	V	M	V
A	0	1.0	.50	.08	.49	.07	.48	.06	.48	.07
		0.25	.31	.01	.30	.01	.30	.01	.30	.01
	1	1.0	.23	.02	.22	.02	.22	.02	.22	.02
		0.25	.04	.00	.14	.00	.14	.00	.13	.00
	2	1.0	.14	.01	.13	.01	.13	.01	.13	.01
		0.25	.09	.00	.09	.00	.09	.00	.09	.00

Procedure (Proc.)
A = ALMA

Statistics
M = Mean
V = Variance

Table 3

Statistics of $\hat{\theta}_n - \theta_n$

Student-t Measurement Errors with 3 degrees of freedom and Variance 1.

Time n:				25				50				75				100			
Proc	Nbr Iter	k	a	M	V	S	K	M	V	S	K	M	V	S	K	M	V	S	K
K	-	-	-	0.01	0.57	-0.48	2.7	0.02	0.53	-0.02	2.2	0.02	0.67	0.78	8.7	0.04	0.54	-0.17	1.7
A	0	1.0	-	0.03	0.67	-0.07	1.0	0.02	0.58	-0.08	1.1	0.02	0.71	-0.02	1.4	0.01	0.65	-0.10	1.7
		0.25	-	0.01	0.53	-0.16	1.6	0.01	0.48	-0.08	1.5	0.02	0.57	-0.01	2.1	0.02	0.50	-0.26	2.3
A	1	1.0	-	0.01	0.55	-0.09	1.5	0.01	0.49	-0.09	1.2	0.02	0.61	-0.01	1.8	0.01	0.52	-0.23	2.4
		0.25	-	-0.01	0.63	-0.46	4.1	0.01	0.58	0.03	3.1	0.03	0.69	0.06	3.5	0.03	0.58	0.20	2.7
A	2	1.0	-	-0.01	0.58	-0.16	2.4	-0.01	0.54	-0.05	1.6	0.03	0.64	0.02	1.9	0.02	0.54	-0.26	2.6
		0.25	-	-0.02	0.71	-0.66	5.5	0.01	0.66	0.09	4.3	0.03	0.79	0.18	4.9	0.04	0.64	-0.05	2.9
B	0	1.0	7	0.01	0.56	-0.17	2.5	0.01	0.51	-0.07	2.2	0.02	0.59	-0.09	2.6	0.02	0.52	-0.31	2.4
		0.25	7	0.01	0.61	-0.68	5.2	0.01	0.57	0.04	3.9	0.03	0.69	0.35	6.0	0.04	0.56	-0.13	2.7
B	0	1.0	9	-0.01	0.57	-0.42	3.5	0.01	0.54	-0.00	3.0	0.02	0.62	-0.07	3.3	0.03	0.53	-0.22	2.5
		0.25	9	-0.01	0.63	-0.76	5.7	0.01	0.57	0.05	4.1	0.03	0.72	0.75	9.9	0.04	0.57	-0.11	2.8
L	0	1.0	-	0.03	0.68	-0.12	1.5	0.03	0.56	-0.09	1.0	0.03	0.69	-0.02	1.5	0.01	0.65	-0.13	1.8
		0.25	-	0.01	0.54	-0.23	1.7	0.01	0.48	-0.11	1.5	0.03	0.60	-0.04	2.1	0.02	0.52	0.29	2.5
L	1	1.0	-	0.01	0.55	-0.13	1.7	0.01	0.49	-0.09	1.2	-0.02	0.61	0.01	1.8	0.01	0.52	0.23	2.4
		0.25	-	-0.01	0.63	-0.53	4.1	0.00	0.57	-0.07	2.8	0.03	0.69	0.07	3.4	0.03	0.58	0.20	2.7
L	2	1.0	-	0.00	0.59	-0.22	2.5	0.01	0.53	-0.08	1.6	0.03	0.53	0.02	1.9	0.02	0.54	-0.26	2.6
		0.25	-	-0.02	0.72	-0.70	5.5	0.00	0.65	-0.01	3.9	0.03	0.79	0.18	4.9	0.04	0.64	-0.05	2.9
W	-	-	-	-0.06	112	-0.26	3.2	-0.51	384	0.14	4.3	0.08	774	0.03	5.1	-0.79	1249	-0.25	5.2

Procedure (Proc.)

K = KALMAN
A = ALMA
B = Biweight
L = Likelihood
W = West

Statistics

M = Mean
V = Variance
S = Skewness
K = Kurtosis

Table 4
Statistics of C_n
Student-t Measurement Errors with 3 degrees of freedom and Variance 1.

Time n:				25		50		75		100	
Proc	Nbr Iter	k	a	M	V	M	V	M	V	M	V
A	0	1.0	-	.46	.07	.44	.06	.48	.07	.45	.06
		0.25	-	.29	.01	.28	.01	.29	.01	.29	.01
A	1	1.0	-	.21	.02	.20	.02	.21	.02	.21	.02
		0.25	-	.13	.00	.13	.00	.14	.00	.14	.00
A	2	1.0	-	.13	.01	.12	.01	.13	.01	.13	.01
		0.25	-	.09	.00	.09	.00	.09	.00	.09	.00
B	0	1.0	7	.29	.01	.29	.00	.29	.01	.29	.00
		0.25	7	.27	.00	.27	.00	.27	.00	.27	.00
B	0	1.0	9	.28	.00	.28	.00	.28	.00	.28	.00
		0.25	9	.27	.00	.27	.00	.27	.00	.27	.00
L	0	1.0	-	.44	.06	.44	.06	.46	.06	.46	.06
		0.25	-	.29	.01	.28	.01	.29	.01	.29	.01
L	1	1.0	-	.21	.02	.20	.02	.21	.02	.21	.02
		0.25	-	.14	.00	.14	.00	.14	.00	.14	.00
L	2	1.0	-	.13	.01	.12	.01	.13	.01	.13	.01
		0.25	-	.09	.00	.09	.00	.09	.00	.09	.00
W	-	-	-	8.8	64	16	240	23	543	29	946

Procedures (Proc.)
A = ALMA
B = Biweight
L = Likelihood
W = West

Statistics
M = Mean
V = Variance

DISTRIBUTION LIST

	NO. OF COPIES
Library (Code 0142) Naval Postgraduate School Monterey, CA 93943-5000	2
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Office of Research Administration (Code 012) Naval Postgraduate School Monterey, CA 93943-5000	1
Center for Naval Analyses 4401 Ford Ave. Alexandria, VA 22302-0268	1
Library (Code 55) Naval Postgraduate School Monterey, CA 93943-5000	1
Operations Research Center, Rm E40-164 Massachusetts Institute of Technology Attn: R. C. Larson and J. F. Shapiro Cambridge, MA 02139	1
Koh Peng Kong OA Branch, DSO Ministry of Defense Blk 29 Middlesex Road SINGAPORE 1024	1
Arthur P. Hurter, Jr. Professor and Chairman Dept of Industrial Engineering and Management Sciences Northwestern University Evanston, IL 60201-9990	1
Institute for Defense Analysis 1800 North Beauregard Alexandria, VA 22311	1