Robust Kalman filter and smoother for errors-in-variables model with observation outliers based on Least-Trimmed-Squares

Jaafar ALMutawa
Department of Mathematics and Statistics,
King Fahd University of Petroleum and Minerals,
e-mail: jaafarm@kfupm.edu.sa

***Abstract*— In this paper, we propose a robust Kalman filter and smoother for the errors-in-variables (EIV) state space model subject to observation noise with outliers. We introduce the EIV problem with outliers and then we present the Least-Trimmed-Squares (LTS) estimator which is highly robust estimator to detect outliers. As a result, a new statistical test to check the existence of outliers which is based on the Kalman filter and smoother has been formulated. Since the LTS is combinatorial optimization problem the randomized algorithm has been proposed in order to achieve the optimal estimate. However, the uniform sampling method has a high computational cost and may lead to biased estimate, therefore we apply the subsampling method.**

Keywords: Errors-in-variables model, Least-Trimmed-Squares, Kalman filter and smoother, outliers, random search algorithm, subsampling method.

I. Introduction

A basic numerical routine for the classical EIV Kalman filter [1], [7] and smoother computes the conditional expectation which is a least squares (LS) estimate. Since the LS method is rather sensitive to outliers (non Gaussian disturbances), so is the Kalman filter and smoother. Moreover, it is well known in real applications that most practical data contain outliers with a low probability, so that a standard Gaussian assumption for observation noises might fail. Following Rousseew [6], we define the outliers to be the observations which deviate from the pattern set of the majority of the data. There are many reasons for the occurrence of outliers, e.g. misplaces decimal points, recording or transmission errors, expectational phenomena such as earthquakes or strikes, or members of different population slipping in the sample etc.

Several algorithms have been proposed to deal with outliers in the output data [2], [3], however, usually the input data are observed quantities subject to random variability. Thus, there is no reason why gross errors would only occur in the response data. In a certain sense it is more likely to have outlier in one of observed input data, because usually its dimension greater than one, and hence there are more opportunities for some thing to go wrong. As a technique for coping with this problem, Rousseeum [6] suggested the LTS estimator and [5] presented the fast LTS algorithm to compute the multivariate linear regression model. For the EIV state space model where the outliers acts in the observed input data to the best of our knowledge, there is no paper has been published in this area.

In this paper, we consider a filtering and smoothing problem in the presence of observation outliers with the aid of the LTS procedure. It is well known that the LTS is a highly robust estimator and its objective is to find h observations out of N whose square errors is minimum. However, the high computational complexity makes the LTS estimator impractical and useless. Therefore [2] proposed the random search algorithm to solve the LTS problem for the SISO linear regression model. However, applying the randomized algorithm [2] for the EIV state space model may lead to bias estimate since the structure of the data will be lost. Hence, we propose the subsampling method [8] which keeps the structural of the original data, decrease the computation time and less sensitive to outliers. Another feature of the proposed algorithm is that the algorithm can be applied to clean and dirty data as well. A minor contribution of the paper is that we derive the Kalman smoother for the EIV state space model which is required for the new statistics.

This note is organized as follows. Section 2, gives the errors-in-variables problem in the presence of outliers, and introduces the LTS estimator for the EIV state space model. In section 3, we proposed the randomized algorithm as a method to solve the LTS problem and discuss the disadvantages of the algorithm. Section 4, is dedicated to the Kalman filter and smoother with outliers and propose the subsampling method. Appendix A is devoted to Kalman filter and smoother without outliers and proof of the proposition.

II. Errors-in-variables model

As depicted in Fig. 1, consider the errors-in-variables state space model described by

$$\begin{bmatrix} x(t+1) \\ \hat{y}(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x(t) \\ \hat{u}(t) \end{bmatrix} + \begin{bmatrix} w(t) \\ 0 \end{bmatrix}, \quad (1)$$

where $x(t) \in \mathbb{R}^n$, $\hat{u}(t) \in \mathbb{R}^m$ and $\hat{y}(t) \in \mathbb{R}^p$ are unknown state, true input and output vectors respectively. Furthermore, $w(t)$ is the white Gaussian noise acting on the state whose mean is zero and has a covariance Σ_w. The measured input-output signals $u(t)$ and $y(t)$

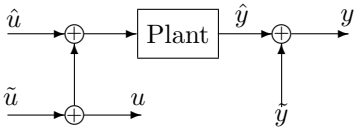

Fig. 1. Errors-in-variables model

are modelled as

$$u(t) = \hat{u}(t) + \tilde{u}(t), \tag{2}$$
$$y(t) = \hat{y}(t) + \tilde{y}(t), \tag{3}$$

where $\tilde{u}(t) \in \mathbb{R}^n$ and $\tilde{y}(t) \in \mathbb{R}^p$ are non-Gaussian white noises with zero mean and finite positive definite covariance matrices $\Sigma_{\tilde{u}}$ and $\Sigma_{\tilde{y}}$, respectively;

$$\mathbb{E}\left\{ \begin{bmatrix} \tilde{u}(t) \\ \tilde{y}(t) \end{bmatrix} \begin{bmatrix} \tilde{u}^{\mathrm{T}}(t) & \tilde{y}^{\mathrm{T}}(t) \end{bmatrix} \right\} = \begin{bmatrix} \Sigma_{\tilde{u}} & \Sigma_{\widetilde{uy}} \\ \Sigma_{\widetilde{uy}}^{\mathrm{T}} & \Sigma_{\tilde{y}} \end{bmatrix} \delta(\tau). \tag{4}$$

We will assume in the sequel, that $\tilde{u}(t)$ and $\tilde{y}(t)$ are uncorrelated with $w(t)$. Furthermore, the input and output noises $\tilde{u}(t)$ and $\tilde{y}(t)$ contain outliers with a low probability, therefore we write

$$\tilde{u}(t) = (I_m - \phi(t))\tilde{u}^n(t) + \phi(t)\tilde{u}^o(t),$$
$$\tilde{y}(t) = (I_p - \gamma(t))\tilde{y}^n(t) + \gamma(t)\tilde{y}^o(t),$$

where I_s is the $s \times s$ identity matrix for $s = \{m, p\}$, $\psi(t) = \mathrm{diag}\{\psi_{t,i}\} = \mathrm{diag}\{\psi_{t,1}, \cdots, \psi_{t,s}\}$ and $\psi_{t,i} = \{0, 1\}$ for all i and where $\psi = \{\gamma, \phi\}$. Moreover, Prob$\{\psi_{t,i} = 1\}$ is small, i.e. the majority of the observed data is clean. The noises $\{\tilde{u}^n(t), \tilde{u}^o(t), \tilde{y}^n(t), \tilde{y}^o(t)\}$ are Gaussian white noises with

$$\tilde{u}^n(t) \in \mathrm{N}(0, \Sigma_{\tilde{u}}^n), \qquad \tilde{u}^o(t) \in \mathrm{N}(0, \Sigma_{\tilde{u}}^o), \tag{5}$$
$$\tilde{y}^n(t) \in \mathrm{N}(0, \Sigma_{\tilde{u}}^n), \qquad \tilde{y}^o(t) \in \mathrm{N}(0, \Sigma_{\tilde{y}}^o), \tag{6}$$

where $\{\Sigma_{\tilde{u}}^n, \Sigma_{\tilde{u}}^o, \Sigma_{\tilde{y}}^n, \Sigma_{\tilde{u}}^o\}$ are positive definite covariance matrices. Furthermore, $\Sigma_{\tilde{u}}^o(i,i)$ and $\Sigma_{\tilde{y}}^o(i,i)$ are much larger than $\Sigma_{\tilde{u}}^n(i,i)$ and $\Sigma_{\tilde{y}}^n(i,i)$ respectively. Then, the problem of interest is to find an optimal Kalman filter and smoother estimate $\hat{u}^*(t), \hat{y}^*(t)$ and $\hat{x}(t)$ for the input-output data $\hat{u}(t), \hat{y}(t)$ and the state vector $x(t)$ given the observed input-output data. The fact that we account for the possibility that the input signal is not exactly known and it may contain outliers, makes the problem difficult, and is often referred to as an outlier-errors-in-variables (OEIV) problem.

A. Least-trimmed-squares

The LTS technique has been introduced by [6] to detect the outliers for the EIV linear regression model. Substituting (2) and (3) into (1) yields

$$\begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix}, \tag{7}$$

where $n_x(t) = -B\tilde{u}(t) + w(t)$ and $n_y(t) = -D\tilde{u}(t) + \tilde{y}(t)$. Let $\Theta = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, then the least trimmed squares estimator is defined as

$$\hat{\Theta}_{\mathrm{LTS}} = \mathrm{argmin}_{\Theta} \sum_{i=1}^{h} (r^{\mathrm{T}} \cdot r)_{[i]}(\Theta), \tag{8}$$

where $(r^{\mathrm{T}} \cdot r)_{[i]}(\Theta)$ represents the i-th order statistics among $r_1^{\mathrm{T}}(\Theta) \cdot r_1(\Theta), \cdots, r_N^{\mathrm{T}}(\Theta) \cdot r_N(\Theta)$ and where $r_i(\Theta) = \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} - \overbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}$.The so-called trimming constant h have to satisfy $\frac{N}{2} < h \leq N$. This constant determines the breakdown point of the LTS estimator since the definition (8) implies that $N - h$ observations with the largest residuals will not affect the estimator (except of the fact that the squared residuals of excluded points have to be larger than the h-th order statistics among the squared residuals).

To redefine the LTS estimator for EIV state space model, consider $\mathcal{S} = \{S \subseteq \{1, \cdots, N\} : \#S = h\}$ [1] be the collection of all subsets with cardinality h from the set $\{1, \cdots, N\}$ [2]. For any $S \in \mathcal{S}$, let $\hat{\Theta}(S)_{\mathrm{LTS}}$ be the least square estimate based on the observed data in S

$$\hat{\Theta}(S)_{\mathrm{LTS}} = \mathrm{argmin}_{\Theta} \sum_{i \in S} r_i^{\mathrm{T}}(\Theta) \cdot r_i(\Theta). \tag{9}$$

i.e. the LTS searches for a subset $S \in \mathcal{S}$ of size h that fits the observed data.

In most cases, it is not feasible to generate all possible subsets provided that N is large due to computational cost. In the next section, we will generate finite number of subsets which will lead to a feasible solution that will converge with probability one to the true solution by using the randomized algorithm.

III. The random search algorithm

It is obvious that the objective function, $\hat{\Theta}(S)_{\mathrm{LTS}}$ in (9) can be found by searching for the best subset $S \in \mathcal{S}$ that minimizes the squares of the errors. In fact there are S_i subsets in $\mathcal{S}$ for $i = 1, \cdots, \binom{N}{M}$, so that finding the best subset that minimizes the value of the objective function is a very difficult combinatorial problem. However, we can easily calculate the value of the objective function (9), for each subset $S \in \mathcal{S}$ and then sort them in increasing order, i.e.

$$\hat{\Theta}(S)_{[1]} = \mathrm{argmin}_{\Theta} \sum_{i \in S} r_i^{\mathrm{T}}(\Theta) \cdot r_i(\Theta) \leq \cdots$$
$$\leq \mathrm{argmin}_{\Theta} \sum_{i \in S} r_i^{\mathrm{T}}(\Theta) \cdot r_i(\Theta) = \max_{S \in \mathcal{S}} \det \mathrm{cov}(S).$$

Now we think of $S_i \in \mathcal{S}$ as a random variable that is uniformly distributed, and hence $\hat{\Theta}(S_i)$ is also a random variable depending on S_i. Let $F(\hat{\Theta}(S_i))$ denote the unknown probability distribution function of $\hat{\Theta}(S_i)$ for

[1] $\# :=$ cardinality of the subset S.
[2] $[\cdot]$ is the greatest integer number.

$i = 1, \cdots, L$ be L, independently generated samples of $S_i \in \mathcal{S}$. Furthermore, let $\bar{S} \in \{S\}_{r=1}^{L}$ be such that $\hat{\Theta}(\bar{S}) = \min_{1 \le r \le L} \hat{\Theta}(S_r)$. We can derive the following theorem by using the result of Bai [2]. The theorem finds the sample size L so that $\hat{\Theta}(\bar{S})$ converges to the true solution with probability close to one.

Theorem 1: For the EIV model (1), we can show that the following (i) $\sim$ (ii) hold:

(i) For all $0 < F\left(\min_{S \in \mathcal{S}} \hat{\Theta}(S)\right) < \epsilon < 1$ and for all $0 < \delta < 1$, if $L \ge \frac{\ln(1/\delta)}{\ln(1/(1-\epsilon))}$, then

$$\text{Prob}\left\{ F\left(\min_{1 \le r \le L} \hat{\Theta}(S_r) \right) \le \epsilon \right\} \ge 1 - \delta.$$

(ii) Let S_r for $r = 1, \cdots, k$ be k-th disjoint subsets such that $\cup_{r=1}^{k} S_r = \{1, \cdots, N\}$ and run the randomized algorithm in each subset. Then the overall probability that the confidence statement are simultaneously true is $1 - \sum_{i=1}^{k} \alpha_i$.

Proof: (i) Let $\hat{\Theta}(S)_{[k]}$ denote the maximum $\hat{\Theta}(S)$ that satisfies $F(\hat{\Theta}(S)) \le \epsilon$, i.e

$$F\left(\hat{\Theta}(S)_{[\binom{N}{M}]}\right) \ge \cdots \ge F\left(\hat{\Theta}(S)_{[k+1]}\right) > \epsilon.$$

It is easy to see that $F\left(\min_{1 \le i \le L} \hat{\Theta}(S)\right) \le \epsilon$ if and only if $\min_{1 \le i \le L} \hat{\Theta}(S_i) \le \hat{\Theta}_{[k]}$, implying that

$$\begin{aligned}
&\text{Prob}\left\{ F\left(\min_{1 \le i \le L} \hat{\Theta}(S_i) \right) \le \epsilon \right\} \\
&\quad = \text{Prob}\left\{ \min_{1 \le i \le L} \hat{\Theta}(S_i) \le \hat{\Theta}(S)_{[k]} \right\} \\
&\quad = 1 - \text{Prob}\left\{ \min_{1 \le i \le L} \hat{\Theta}(S_i) > \hat{\Theta}_{[k]} \right\} \\
&\quad = 1 - \text{Prob}\left\{ \hat{\Theta}(S_1) \ge \hat{\Theta}(S)_{[k+1]} \right\} \\
&\quad \times \cdots \times \text{Prob}\left\{ \hat{\Theta}(S_L)) \ge \hat{\Theta}(S)_{[k+1]} \right\} \\
&\quad \ge 1 - (1-\epsilon)^L.
\end{aligned}$$

Now $L \ge \frac{\ln(\frac{1}{\delta})}{\ln \frac{1}{(1-\epsilon)}} \Rightarrow (1-\epsilon)^L \le \delta$. Consequently,

$$\text{Prob}\left\{ F\left(\min_{1 \le i \le L} \hat{\Theta}(S_i) \right) \le \epsilon \right\} \ge 1 - (1-\epsilon)^L \ge 1 - \delta.$$

(ii) Let $\mathcal{E}_i$, $(i = 1, \cdots, k)$ be the ith statement corresponds to the subset S_i, and assume that the ith statement $\mathcal{E}_i$, $(i = 1, \cdots, k)$ is correct, i.e.

$$\text{Prob}[\mathcal{E}_i] = 1 - \alpha_i,$$

and let $\bar{\mathcal{E}}_i$ be the complementary event of $\mathcal{E}_i$, then

$$\begin{aligned}
\text{Prob}[\cap \mathcal{E}_i] &= 1 - \text{Prob}[\overline{\cap_i \mathcal{E}_i}] = 1 - \text{Prob}[\cup_i \bar{\mathcal{E}}_i] \\
&\ge 1 - \sum \text{Prob}[\bar{\mathcal{E}}_i] = 1 - \sum \alpha_i,
\end{aligned}$$

if $\alpha_i = \alpha$ for $i = 1. \cdots, k$. Then

$$\text{Prob}[\cap \mathcal{E}_i] \ge 1 - k\alpha.$$

Theorem 1 means that, whenever we generate L independent random subsets $S_L = \{S_i\}_{i=1}^{L}$ and compute the value of the objective function (9) for each subset $S_i \in S_L$, a subset $\bar{S} \in S_L$ with minimum value of the objective function (9) will improve our estimate. However, it may be noted that in the worst case this improvement is not considerable comparing to the LS estimate by using all observed data. In fact, if the number of the observed data is very large, then the probability of finding a subset $S \in \mathcal{S}$ with cardinality equal to h that does not contain any outlier approaches zero, i.e.

$$P_{\mathcal{I}} = \frac{\binom{\mathcal{I}}{h}}{\binom{N}{h}} = \frac{\mathcal{I}!(N-h)!}{(\mathcal{I}-h)!N!} = \prod_{j=0}^{h-1} \frac{\mathcal{I}-j}{N-j}, \qquad (10)$$

where $\mathcal{I}$ stands for the number of clean data. According to (10) the random search algorithm can be improved by taking S with small cardinality and by finding the smallest h relative Mahalanobis distances d_i. This will increase the probability of finding a subset S_i from $\mathcal{S}$ that does not contain any outliers.

At this stage, we will derive the optimal estimate for the true input-output and the associated error covariances for the OEIV model using the Kalman filter and smoother.

IV. Kalman filter for the errors-in-variables model with outliers

Let $z(t) = y(t) - Du(t)$, then (7) can be written as

$$\begin{bmatrix} x(t+1) \\ z(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix}. \quad (11)$$

In addition, let $\mathcal{Z}(t) = \{z(0), \cdots, z(t)\}$, $\Phi(t) = \{\phi(0), \cdots, \phi(t)\}$ and $\Gamma(t) = \{\gamma(0), \cdots, \gamma(t)\}$ and define [3]

$$x(t \mid t) \equiv \mathbb{E}[x(t) \mid \mathcal{Z}(t), \Phi(t), \Gamma(t)], \quad (12)$$

$$x(t+1 \mid t) \equiv \mathbb{E}[x(t+1) \mid \mathcal{Z}(t), \Phi(t), \Gamma(t)], \quad (13)$$

$$y(t+1 \mid t) \equiv \mathbb{E}[y(t+1) \mid \mathcal{Z}(t), \Phi(t), \Gamma(t)], \quad (14)$$

$$P(t \mid t) \equiv \mathbb{E}[(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^{\mathrm{T}} \mid \mathcal{Z}(t), \Phi(t), \Gamma(t)], \quad (15)$$

$$P(t+1 \mid t) \equiv \mathbb{E}[(x(t+1) - \hat{x}(t+1))(x(t+1) - \hat{x}(t+1))^{\mathrm{T}} \mid \mathcal{Z}(t), \Phi(t), \Gamma(t)], \quad (16)$$

then the Kalman filter is given by

$$z(t+1 \mid t) = Cx(t+1 \mid t), \quad (17)$$

$$x(t+1 \mid t) = Ax(t \mid t) + Bu(t), \quad (18)$$

and we could compute the covariance of the errors as

$$\begin{aligned}
&\mathbb{E}\{(z(t+1) - z(t+1 \mid t))(x(t+1) - x(t+1 \mid t))^{\mathrm{T}}\} \\
&\quad = CP(t+1 \mid t), \\
&\mathbb{E}\{(z(t+1) - z(t+1 \mid t))(z(t+1) - z(t+1 \mid t))^{\mathrm{T}}\} \\
&\quad = CP(t+1 \mid t)C^{\mathrm{T}} + \gamma(t)\Sigma_{\tilde{y}}^{n} + (I_p - \gamma(t))\Sigma_{\tilde{y}}^{o} \\
&\qquad + D[\phi(t)\Sigma_{\tilde{u}}^{n} + (I_m - \phi(t))\Sigma_{\tilde{u}}^{o}]D^{\mathrm{T}}, \qquad (19)
\end{aligned}$$

where

$$\begin{aligned}
P(t+1 \mid t) &= \mathbb{E}[(x(t+1) - x(t+1 \mid t))(x(t+1) - x(t+1 \mid t))^{\mathrm{T}}] \\
&= AP_{t|t}A^{\mathrm{T}} + \Sigma_w + B\phi(t)\Sigma_{\tilde{u}}^{n}B^{\mathrm{T}} + B(I_m - \phi(t))\Sigma_{\tilde{u}}^{o}B^{\mathrm{T}}. \qquad (20)
\end{aligned}$$

[3]The Kalman filter and smoother without outliers is given in Appendix .

The optimal Kalman filter estimate for the state $x(t)$ is

$$x(t+1 \mid t+1) = x(t+1 \mid t) + P(t+1 \mid t)C^{\mathrm{T}}\Sigma_{\epsilon}(t)^{-1}\epsilon(t), \tag{21}$$

while $\epsilon(t)$ and $\Sigma_{\epsilon}(t)$ denote the innovation of $z(t)$ and its covariance matrix given by

$$\begin{aligned}\epsilon(t) &= z(t) - Cx(t \mid t) \\ &= Cx(t) + n_y(t) - Cx(t \mid t) \qquad (22)\\ \Sigma_{\epsilon}(t) &= \mathbb{E}[\epsilon(t)\epsilon(t)^{\mathrm{T}}] \\ &= CP(t \mid t)C^{\mathrm{T}} + \gamma(t)\Sigma_{\tilde{y}}^{n} + (I_p - \gamma(t))\Sigma_{\tilde{y}}^{o} \\ &\quad + D[\phi(t)\Sigma]_{\tilde{u}}^{n} + (I_m - \phi(t))\Sigma_{\tilde{u}}^{o}]D^{\mathrm{T}}. \qquad (23)\end{aligned}$$

The optimal smooth estimates $\hat{u}(t \mid N)$, $\hat{y}(t \mid N)$ of $\hat{u}(t)$, $\hat{y}(t)$ that can be obtained from $\{u(0), y(0), \cdots, u(N), y(N)\}$, under constraints (1)-(3) are given by

$$\hat{u}(t \mid N) = u(t) - \tilde{u}(t \mid N) = u(t) - \mathbb{E}\{\tilde{u}(t) \mid z(0), \cdots, z(N)\}, \tag{24}$$

$$\hat{y}(t \mid N) = y(t) - \tilde{y}(t \mid N) = y(t) - \mathbb{E}\{\tilde{y}(t) \mid z(0), \cdots, z(N)\}, \tag{25}$$

where $\tilde{u}(t \mid N) = \mathbb{E}\{\tilde{u}(t) \mid z(0), \cdots, z(N)\}$ and $\tilde{y}(t \mid N) = \mathbb{E}\{\tilde{y}(t) \mid z(0), \cdots, z(N)\}$ are the optimal estimate for $\tilde{u}(t)$ and $\tilde{y}(t)$ respectively. To compute $\tilde{u}(t \mid N)$ and $\tilde{y}(t \mid N)$ we replace $z(t)$ by its innovation

$$\begin{aligned}\tilde{u}(t \mid N) &= \mathbb{E}[\tilde{u}(t) \mid z(0), \cdots, z(t), \epsilon(t+1), \cdots, \epsilon(N)] \\ &= \mathbb{E}[\tilde{u}(t) \mid z(0), \cdots, z(t)] + \mathbb{E}[\tilde{u}(t) \mid \epsilon(t+1), \cdots, \epsilon(N)] \\ &= \tilde{u}(t \mid t) + \sum_{s=t+1}^{N} \mathrm{cov}\{\tilde{u}(t), \epsilon(s)\}\Sigma_{\epsilon}(s)^{-1}\epsilon(s) \qquad (26)\\ \tilde{y}(t \mid N) &= \mathbb{E}[\tilde{y}(t) \mid z(0), \cdots, z(t), \epsilon(t+1), \cdots, \epsilon(N)] \\ &= \mathbb{E}[\tilde{y}(t) \mid z(0), \cdots, z(t)] + \mathbb{E}[\tilde{y}(t) \mid \epsilon(t+1), \cdots, \epsilon(N)] \\ &= \tilde{y}(t \mid t) + \sum_{s=t+1}^{N} \mathrm{cov}\{\tilde{y}(t), \epsilon(s)\}\Sigma_{\epsilon}(s)^{-1}\epsilon(s), \qquad (27)\end{aligned}$$

where $\tilde{u}(t \mid t)$ and $\tilde{y}(t \mid t)$ are given in Appendix . Now the covariances can be found as follows

$$\begin{aligned}\mathrm{cov}\{\tilde{u}(t), \epsilon(s)\} &= \mathrm{cov}\{\tilde{u}(t) - \tilde{u}(t \mid t) + \tilde{u}(t \mid t), \epsilon(s)\} \\ &= \mathrm{cov}\{\tilde{u}(t \mid t), \epsilon(s)\} = [\Sigma_{\widetilde{u}\widetilde{y}} - \Sigma_{\tilde{u}}D^{\mathrm{T}}]\Sigma_{\epsilon}(t)^{-1}\Sigma_{\epsilon}(t,s) \\ &= [\Sigma_{\widetilde{u}\widetilde{y}} - \Sigma_{\tilde{u}}D^{\mathrm{T}}]\Sigma_{\epsilon}(t)^{-1}CP(t \mid t-1)L(s-1,t)^{\mathrm{T}}C^{\mathrm{T}}, \qquad (28)\\ \mathrm{cov}\{\tilde{y}(t), \epsilon(s)\} &= \mathrm{cov}\{\tilde{y}(t) - \tilde{y}(t \mid t) + \tilde{y}(t \mid t), \epsilon(s)\} \\ &= \mathrm{cov}\{\tilde{y}(t \mid t), \epsilon(s)\} = [\Sigma_{\tilde{y}} - \Sigma_{\widetilde{u}\widetilde{y}}^{\mathrm{T}}D^{\mathrm{T}}]\Sigma_{\epsilon}(t)^{-1}\Sigma_{\epsilon}(t,s) \\ &= [\Sigma_{\tilde{y}} - \Sigma_{\widetilde{u}\widetilde{y}}^{\mathrm{T}}D^{\mathrm{T}}]\Sigma_{\epsilon}(t)^{-1}CP(t \mid t-1)L(s-1,t)^{\mathrm{T}}C^{\mathrm{T}}, \qquad (29)\end{aligned}$$

where $\Sigma_{\tilde{u}} = (I_m - \phi(t))\Sigma_{\tilde{u}}^{n} + \phi(t)\Sigma_{\tilde{u}}^{o}$ and $\Sigma_{\tilde{y}} = (I_p - \gamma(t))\Sigma_{\tilde{y}}^{n} + \gamma(t)\Sigma_{\tilde{y}}^{o}$ and $\Sigma_{\widetilde{u}\widetilde{y}} = (I_m - \phi(t))\Sigma_{\widetilde{u}\widetilde{y}}^{n}(I_p - \gamma(t)) + \phi(t)\Sigma_{\widetilde{u}\widetilde{y}}^{o}\gamma(t)$. The $L(s-1,t)$ and $\Sigma_{\epsilon}(t,s)$ are defined and calculated in Proposition 2(given in Appendix).

Proposition 1: Let π_t be a random integer number from 1 to N, and formulate the set $S = \{\pi_t : t = 1, \cdots, h\} \in \mathcal{S}$. Furthermore, let $u(\pi_t \mid S)$ and $y(\pi_t \mid S)$ be the Kalman smoother as in (24) and (25). Then the LTS cost function can be written as

$$\begin{aligned}\hat{\Theta}(S)_{\mathrm{LTS}} = \ & \mathrm{argmin}_{\Theta} \sum_{i \in S} \left(\begin{bmatrix} u(i) \\ y(i) \end{bmatrix} - \begin{bmatrix} u(i \mid S) \\ y(i \mid S) \end{bmatrix} \right)^{\mathrm{T}} \\ & \times \left(\begin{bmatrix} u(i) \\ y(i) \end{bmatrix} - \begin{bmatrix} u(i \mid S) \\ y(i \mid S) \end{bmatrix} \right). \qquad (30)\end{aligned}$$

It should be noted that if i is included in the subset S, then $\phi(i)$ and $\gamma(i)$ will be the identity matrices, otherwise they are the zero matrices. In Proposition 1, if we apply the uniform sampling method then we will lose the structure of the original data and consequently the estimate will be biased. Therefore, we apply another sampling method which is called subsampling method [8].

A. Subsampling method

Instead of generating a random subsets from the observed input-output data we generate blocks of contiguous observations of fixed dimension b. That is, we divide the last $(N-n)$ observations into k subsets, where each subset contains the first initial data $(\omega(1), \cdots, \omega(n))$ and a set of $[(N-n)/k]$ contiguous observations. In other words, the subsets can be described as $S_r^{(b+n)} = \{\omega(1), \cdots, \omega(n), \omega(n+1+(r-1)b), \cdots, \omega(n+br)\}$, where $r = 1, \cdots, k$. Then we perform an exhaustive search of all possible blocks and choose the one which gives the minimum value for the objective function. It should be noted that, if $(N-n)/k$ is an integer then we have exactly k subsets. In general there are $k+1$ subsets, where the first k of size $n + [(N-n)/k]$, and the last of size $N - [(N-n)/k]k$. For the seek of simplicity and without loss of generality we assume that b is an integer where $b = (N-n)/k$.

Furthermore, if the number of the subsets k is large, then the probability of having at least a clean subset of data which does not contain any outlier will increase. However, if k is large, then the cardinality of each subset will be small, and consequently the estimate of the parameters can be unstable. P. Heagerty and T. Lumley [8] suggest that $b \approx \sqrt{N}$ to ensure a balance between the statistical properties of the estimated parameters and the robustness of the method.

Theorem 2: Let $\mid S_1^{(b+n)} \mid = h$ and put

$$\begin{aligned}J_1 := \sum_{i \in S_1^{(b+n)}} & \left(\begin{bmatrix} u(i) \\ y(i) \end{bmatrix} - \begin{bmatrix} u(i \mid S_1^{(b+n)}) \\ y(i \mid S_1^{(b+n)}) \end{bmatrix} \right)^{\mathrm{T}} \\ & \times \left(\begin{bmatrix} u(i) \\ y(i) \end{bmatrix} - \begin{bmatrix} u(i \mid S_1^{(b+n)}) \\ y(i \mid S_1^{(b+n)}) \end{bmatrix} \right).\end{aligned}$$

Now let $r_1(i) \in S_1^{(b+n)}$ and take $S_2^{(b+n)}$ such that $\{\mid r_1(i) \mid; i \in S_2^{(b+n)}\} = \{(r_1)_{[1]}, \cdots, (r_1)_{[h]}\}$, where $(r_1)_{[1]} \leq \cdots \leq (r_1)_{[h]}$. This yields $r_2(i)$ for all $i = 1, \cdots, N$ and $J_2 := \sum_{i \in S_2^{(b+n)}} r_i^{\mathrm{T}}(\Theta) \cdot r_i(\Theta)$. Then

$$J_2 \leq J_1$$

The proof of Theorem 2 is a direct application of Theorem 1 from [4]. It should noted that constructing a new subset $S_2^{(b+n)}$ from $S_1^{(b+n)}$ is called C-step where following Rousseeuw and etc [4], C stands for

"concentration" because the new subset $S_2^{(b+n)}$ gives a lower value for the objective than $S_1^{(b+n)}$ does.

Random search algorithm:

Let $\cup_{i=1}^{k} S_i^{(b+n)} = \{1, 2, \cdots, N\}$,

- **Step 1:** Generate all subsamples of $S_i^{(b+n)}$, and for each sub-sample $S_i^{(b+n)}$, calculate $\hat{\Theta}(S_i^{(b+n)})_{\mathrm{LTS}}$ and consequently find $\hat{\Theta}(\bar{S})_{\mathrm{LTS}} = \min_{S_i^{(b+n)} \in \mathcal{S}} \hat{\Theta}_{\mathrm{LTS}}(S_i^{(b+n)})$.
- **Step 2:** Using Chi-square distribution detect the outliers and put $S_1 = \{\pi_i : i = 1, \cdots, h\}$.
- **Step 3:** Repeat step 1 to step 2, until convergent.

V. Conclusion

In this paper, we have studied the Kalman filter and smoother for the Error-In-Variables state space models with outliers. The outliers have been detected using highly robust estimator called minimum covariance determinant which requires the Kalman filter and smoother to be computed. In order to achieve the optimal solution of the LTS problem, the random search algorithm has been proposed. However, applying the uniform sampling method to the randomized algorithm leads to complex calculation and biased estimate. Thus, we applied the subsampling method in order to keep the same dependence structure as the original data. The subsampling method leads to unbiased estimate and decrease the complexity issue of calculations. The proposed algorithm is highly robust to the effect of outliers.

Acknowledgment This work was supported by King Fahd University of Petroleum and Minerals

References

[1] R. Diversi, R. Guidorzi and U. Soverini: Kalman filtering in extended noise environments; *IEEE Trans. Automatic Control*, Vol. 50, No. 9, pp. 1396-1402 (2005)

[2] E. Bai: A random least-trimmed-squares identification algorithm; *Automatica*, Vol. 39, No. 9, pp. 1651-1659 (2003)

[3] T. Proietti: Leave-k-out diagnostics in state-space models, *J. of Time Series Analysis*, Vol. 24, No. 2, pp. 221-236 (2003).

[4] P. J. Rousseeum, S. Van Aelst, K. Van Driessen and J. Agulló: Robust multivariate regression; *Technometrics*, Vol. 46, No. 3, pp. 293–305 (2004)

[5] P. J. Rousseeum and K. Van Driessen: Computing LTS regression for large data sets; *Data Mining and Knowledge Discovery*, Vol. 12, pp. 29–45 (2006)

[6] P. Rousseeum: Least median of squares regression; *J. American Statistical Assoc.*, Vol. 79, pp. 871-880 (1984)

[7] Ivan Markovsky and Bart De Moor: Linear dynamic filtering with noisy input and output; *Automatica*, Vol. 41, No. 1, pp. 167-171 (2005)

[8] P. Heagerty and T. Lumley: Window subsampling of estimating functions with application to regression models; *J. American Statistical Assoc.*, Vol. 95, pp. 197-211 (2000)

Appendix

The Kalman filter is given by

$$z(t+1 \mid t) = Cx(t+1 \mid t) \tag{31}$$

$$x(t+1 \mid t) = Ax(t \mid t-1) + Bu(t) + K(t)\epsilon(t) \tag{32}$$

$$K(t) = [AP(t \mid t-1)C^{\mathrm{T}} + S(t)]\Sigma_\epsilon(t)^{-1} \tag{33}$$

$$P(t+1 \mid t) = AP(t \mid t-1)A^{\mathrm{T}} + Q(t) - [AP(t \mid t-1)C^{\mathrm{T}} + S(t)] \times \Sigma_\epsilon(t)^{-1}[AP(t \mid t-1)C^{\mathrm{T}} + S(t)]^{\mathrm{T}} \tag{34}$$

and the Kalman smoother for $t = N, N-1, \cdots, 1$ is given by

$$x(t-1 \mid N) = x(t-1 \mid t-1) + J(t-1)[x(t \mid N) - x(t \mid t-1)] \tag{35}$$

$$P(t-1 \mid N) = P(t-1 \mid t-1) + J(t-1)[P(t \mid N) - P(t \mid t-1)]J(t-1)^{\mathrm{T}} \tag{36}$$

$$J(t-1) = P(t-1 \mid t-1)AP(t \mid t-1)^{-1}$$

$$\tilde{u}(t \mid t) = [\Sigma_{\widetilde{u}\widetilde{y}}(t) - \Sigma_{\tilde{u}}D^{\mathrm{T}}]\Sigma_\epsilon(t)^{-1}\epsilon(t) \tag{37}$$

$$\tilde{y}(t \mid t) = [\Sigma_{\tilde{y}} - \Sigma_{\widetilde{u}\widetilde{y}}^{\mathrm{T}}D^{\mathrm{T}}]\Sigma_\epsilon(t)^{-1}\epsilon(t) \tag{38}$$

By using (37) and (38), the minimal variance estimates of $\hat{y}(t)$ and $\hat{u}(t)$ can be written in the form

$$\hat{u}(t \mid t) = u(t) - [\Sigma_{\widetilde{u}\widetilde{y}} - \Sigma_{\tilde{u}}D^{\mathrm{T}}]\Sigma_\epsilon(t)^{-1}\epsilon(t) \tag{39}$$

$$\hat{y}(t \mid t) = y(t) - [\Sigma_{\tilde{y}} - \Sigma_{\widetilde{u}\widetilde{y}}^{\mathrm{T}}D^{\mathrm{T}}]\Sigma_\epsilon(t)^{-1}\epsilon(t) \tag{40}$$

Proposition 2: For $1 \le t \le s$, the followings hold

(i)

$$P(t,s) = \mathbb{E}\{(x(t) - x(t \mid t-1))(x(s) - x(s \mid s-1))^{\mathrm{T}}\} = P(t \mid t-1)L(s-1,t)^{\mathrm{T}}. \tag{41}$$

where $L(s,t) = L(s)\cdots L(t)$ and $L(s) = A - K(s)C$.

(ii)

$$\Sigma_\epsilon(t,s) = \mathbb{E}\{\epsilon(t)\epsilon(s)^{\mathrm{T}}\} = CP(t \mid t-1)L(s-1,t)^{\mathrm{T}}C^{\mathrm{T}} \tag{42}$$

Proof: (i)

$$\begin{aligned} &x(s+1) - x(s+1 \mid s) \\ &= A\left(x(s) - x(s \mid s-1)\right) + n_x(s) - K(s)\epsilon(s) \\ &= \left(A - K(s)C\right)\left(x(s) - x(s \mid s-1)\right) + n_x(s) - K(s)n_y(s) \\ &= G(s)\left(x(s) - x(s \mid s-1)\right) + n_x(s) - K(s)n_y(s), \end{aligned}$$

where $G(s) = (A - K(s)C)$, hence

$$\begin{aligned} P(t,s) &= \mathbb{E}\{(x(t) - x(t \mid t-1))(x(s+1) - x(s+1 \mid s))^{\mathrm{T}}\} \\ &= \mathbb{E}\{(x(t) - x(t \mid t-1))(T(s)\left(x(s) - x(s \mid s-1)\right) \\ &\quad + n_x(s) - K(s)n_y(s))^{\mathrm{T}}\} \\ &= \mathbb{E}\{(x(t) - x(t \mid t-1))(x(s) - x(s \mid s-1))\}T(s)^{\mathrm{T}} \\ &= P(t \mid t-1)L(s-1,t)^{\mathrm{T}}. \end{aligned} \tag{43}$$

(ii)

$$\begin{aligned} \mathbb{E}\{\epsilon(t)\epsilon(s)^{\mathrm{T}}\} &= \mathbb{E}\{C(x(t) - x(t \mid t-1))(x(s) - x(s \mid s-1))^{\mathrm{T}}C^{\mathrm{T}}\} \\ &= CP(t \mid t-1)L(s-1,t)^{\mathrm{T}}C^{\mathrm{T}} \end{aligned} \tag{44}$$