KYBERNETIKA — VOLUME 32 (1996), NUMBER 3, PAGES 289 – 306

ROBUST METHODS IN EXPONENTIAL SMOOTHING

JIŘÍ MICHÁLEK[1]

The paper deals with robust modification of exponential smoothing with additive outliers. The presented modification is based on the M-estimates approach. Simple and double exponential smoothing procedures are discussed in detail. A robust version of famous Holt's method is given, too.

1. INTRODUCTION

Exponential smoothing is a very famous and often used recursive procedure for both smoothing and prediction of time series. One of the main advantages of exponential smoothing is its numerical simplicity giving relatively good results. However, in case when the observed time series is distorted by outliers, the classical technique of exponential smoothing can fail. Such a situation can occur in practice when, e. g. an information comes out from a sensor, which can add unpleasant outliers to a basic signal caused by a wrong function of the sensor. As for the robust methods in time series the reader is referred to the survey paper by Dutter and Stockinger [3]. The paper is mainly motivated by the paper due to Cipra [2], where some recursive methods of robust exponential smoothing are derived. This paper analysed the methods described in Cipra [2] and discussed mainly the single and double robust exponential smoothing procedures. A robust version of Holt's method is given, too.

In general, exponential smoothing is a method for filtering and predicting, which is very closely connected with the method of weighted least squares, where the corresponding weights are given by a forgetting factor. The method considers a general linear regression model

$$y(t) = z^{\mathrm{T}}(t)\, a + \sigma\, e(t), \tag{1}$$

where $z(t)$ is a vector of specified basis functions, a is a vector of unknown parameters and $e(t)$ is a random noise with standard deviation 1. Here, $u^{\mathrm{T}}v$ is the usual scalar product. The parameter σ need not be known, in general. In practice, the parameters (a, σ) can usually change at time t and therefore it is reasonable to

[1]This research was supported by the Grant Agency of the Czech Republic under grant No. 201/93/0233.

generalize the model (1) into the form

$$y(t) = z^{\mathrm{T}}(t)\, a(t) + \sigma(t)\, e(t), \tag{2}$$

where unknown parameters $(a(t),\ \sigma(t))$ are estimated only on the basis of the latest observations $y(s),\ s \leq t$, which is given by a suitable choice of a forgetting parameter α. In case $\sigma(t)$ is known the unknown parameter $a(t)$ is estimated by minimizing

$$J(t) = \sum_{i=1}^{t} \left(\frac{y(i) - z^{\mathrm{T}}(i)\, a}{\sigma(t)} \right)^2 \alpha^{t-i},$$

where $0 < \alpha < 1$. When the noise $e(t)$ is distorted by outliers then Cipra in [2] suggests a modification of $J(t)$ using M-estimates, which can limit the influence of outliers. Then the "quadratic" form is

$$J(t) = \sum_{i=1}^{t} \alpha^{t-i} \rho \left(\frac{y(i) - z^{\mathrm{T}}(i)\, a}{\sigma(t)} \right),$$

where a function ρ satisfies demands that are usually required for M-estimates (e. g. see Stockinger and Dutter). The "normal equations" are of the form

$$\sum_{i=1}^{t} \alpha^{t-i} \psi \left(\frac{y(i) - z^{\mathrm{T}}(i)\, a}{\sigma(t)} \right) z(i) = 0$$

when $\psi(\cdot)$ exists. If $\sigma(t)$ is unknown, one can proceed as follows (see Stockinger and Dutter) by minimizing

$$J_1(t) = \sum_{i=1}^{t} \alpha^{t-i} \rho \left(\frac{y(i) - z(i)\, a}{\sigma} \right) \sigma + c\sigma,$$

where c is a suitable constant. After derivating we obtain the equation

$$\sum_{i=1}^{t} \alpha^{t-i} \varphi \left(\frac{y(i) - z^{\mathrm{T}}(i)\, a}{\sigma} \right) z(i) = 0$$

with $\varphi(t) = t\psi(t) - \rho(t)$. This approach leads to the following equation

$$\sum_{i=1}^{t} \psi \left(\frac{r(i)}{\hat{\sigma}} \right) z(i)\, \alpha^{t-i} = 0,$$

where $r(i) = y(i) - z^{\mathrm{T}}(i)\, \hat{a}$ are residuals and $\hat{\sigma}$ is an estimate of standard deviation. This equation can be rewritten into the form (we put $\frac{0}{0} = 1$ for simplicity)

$$\sum_{i=1}^{t} \alpha^{t-i} \frac{r(i)}{\hat{\sigma}} \frac{\psi\left(\frac{r(i)}{\hat{\sigma}}\right)}{\frac{r(i)}{\hat{\sigma}}} z(i) = 0$$

which is a form that is very similar to the method of weighted least squares with the forgetting factor α. The obtained equations are not linear and can be solved by a recurrent method (for detail see, e. g. Stockinger and Dutter). It would be very useful to have a recursive algorithm solving these equations as the time t is running $t \to t+1$. Let us introduce the denotation

$$w(i,a) = \frac{\psi\left(\frac{y(i)-z^{\mathrm{T}}(i)\,a}{\hat{\sigma}}\right)}{\frac{y(i)-z^{\mathrm{T}}(i)\,a}{\hat{\sigma}}}.$$

These quantities play the role of weights in normal equations but unfortunately they contain unknown parameters. There is an idea (e. g. Cipra [2]) the unknown parameter a is substituted at the time t by the previous estimate $\hat{a}(t-1)$ obtained at the time $t-1$. Then the variables $w(i,a)$ are not depending on the unknown parameter a and the normal equations are substituted by the following approximative normal equations

$$\sum_{i=1}^{t} \alpha^{t-i} w(i,\, \hat{a}(t-1))\,(y(i) - z^{\mathrm{T}}(i)\,\hat{a}(t-1))\,z(i) = 0$$

which are linear equations in a. As the time t is running we obtain a triangular scheme of the weights $w(i,\, \hat{a}(t-1))$

$$\begin{array}{llll}
w(1,\, \hat{a}(0)) & & & \\
w(1,\, \hat{a}(1)), & w(2,\, \hat{a}(1)) & & \\
w(1,\, \hat{a}(2)), & w(2,\, \hat{a}(2)), & w(3,\, \hat{a}(2)) & \\
\vdots & & & \\
w(1,\, \hat{a}(t-1)), & w(2,\, \hat{a}(t-1)), & \cdots, & w(t,\, \hat{a}(t-1)). \\
\vdots & \vdots & & \vdots
\end{array}$$

The weights are random variables, which are dependent stochastically, in general, because

$$w(i,\, \hat{a}(t-1)) = f(i,\, e(i) - z^{\mathrm{T}}(i)\,\hat{\theta}(t-1))$$

where $\hat{\theta}(t-1) = \hat{a}(t-1) - a$. In case we will consider the Huber function $\psi_{\mathrm{H}}(x) = x$ for $|x| \le c$ and $\psi_{\mathrm{H}}(x) = c\,\mathrm{sign}(x)$ otherwise, we can see that

$$0 < w(i,\, \hat{a}(t-1)) \le 1.$$

Similar results can be obtained in the case of Welsh's function, which is defined as

$$\psi_{\mathrm{W}}(x) = x \cdot e^{-cx^2},$$

where c is another suitable constant, too. Further we will assume that $0 < w(i,a) \le 1$ in every case. With respect to the fact that in each step $t \to t+1$ all the weights $w(i,\, \hat{a}(t))$, $i = 1, 2, \ldots, t+1$ are changing one cannot derive the recursive algorithm in such a way as described in Cipra [2], where the weights lying on the main diagonal

are used only. The role of weights $w(i, \hat{a}(t-1))$ can be explained as follows: the greater error of prediction $y(t) - z^{\mathrm{T}}(t)\,\hat{a}(t-1)$, the smaller value of the weight.

Then, the system of "approximative" normal equations, which are linear, has a solution given by

$$\hat{a}(t) = \left[\sum_{i=1}^{t} \alpha^{t-i} w(i, \hat{a}(t-1))\, z(i)\, z^{\mathrm{T}}(i)\right]^{-1} \sum_{i=1}^{t} \alpha^{t-i} w(i, \hat{a}(t-1))\, y(i)\, z(i).$$

Let us try to substitute this calculation by a recursive algorithm. First, we find the inversion of the matrix

$$[P(t)]^{-1} = \left[\sum_{i=1}^{t} \alpha^{t-i} w(i, \hat{a}(t-1))\, z(i)\, z^{\mathrm{T}}(i)\right]^{-1}.$$

The existence of an inverse matrix immediately follows from the form of the matrix $z(i)\, z^{\mathrm{T}}(i)$ and from the fact that α and $w(i, a)$ are positive. It is evident that

$$\begin{aligned} P(t) &= \sum_{i=1}^{t} \alpha^{t-i} w(i, \hat{a}(t-1))\, z(i)\, z^{\mathrm{T}}(i) = \alpha \sum_{1}^{t-1} \alpha^{t-1-i} w(i, \hat{a}(t-1))\, z(i)\, z^{\mathrm{T}}(i) \\ &\qquad + w(t, \hat{a}(t-1))\, z(t)\, z^{\mathrm{T}}(t) \\ &= \alpha\, P(t-1) + \alpha \sum_{i=1}^{t} \alpha^{t-1-i} \Delta\, w(i, \hat{a}(t-2))\, z(i)\, z^{\mathrm{T}}(i), \end{aligned}$$

where

$$\Delta\, w(i, \hat{a}(t-2)) = w(i, \hat{a}(t-1) - w(i, \hat{a}(t-2)) \quad \text{for } i = 1, 2, \ldots, t-1$$

and

$$\Delta\, w(t, \hat{a}(t-1)) = w(t, \hat{a}(t-1)).$$

In this way we have reached the relation

$$P(t) = \alpha\, P(t-1) + \alpha \sum_{i=1}^{t} \alpha^{t-1-i} \Delta\, w(i, \hat{a}(t-2))\, z(i)\, z^{\mathrm{T}}(i)$$

which is different from the recursive relation given in Cipra [2], p. 67 because he puts

$$\Delta\, w(i, \hat{a}(t-2)) = 0 \quad \text{for } i = 1, 2, \ldots, t-1.$$

For simplicity let us denote $\Delta\, w(i, \hat{a}(t-2))\, z(i)\, z^{\mathrm{T}}(i) = R_i$ then

$$P(t)^{-1} = \frac{1}{\alpha}\left[P(t-1) + \sum_{i=1}^{t} \alpha^{t-1-i} R_i\right]^{-1}.$$

At this moment we use a well-known lemma about inverse matrices, namely

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B(DA^{-1}B + C^{-1})^{-1} DA^{-1}$$

in such a way that this inversion will be used t-times. In the first step

$$P(t)^{-1} = \frac{1}{\alpha}\left[P(t-1) + \sum_{i=1}^{t-1} \alpha^{t-1-i} R_i + \frac{R_t}{\alpha}\right]^{-1}.$$

Denote, for a better orientation, $A_t(t-1) = P(t-1) + \sum_{i=1}^{t-1} \alpha^{t-1-i} R_i$, then

$$P(t)^{-1} = \frac{1}{\alpha}\left(A_t^{-1}(t-1) - A_t^{-1}(t-1)\, z(t) \times \left[z^{\mathrm{T}}(t)\, A_t^{-1}(t-1)\, z(t) + \frac{\alpha}{w(t, \hat{a}(t-1))}\right]^{-1} z^{\mathrm{T}}(t)\, A_t^{-1}(t-1)\right).$$

Then, in the second step, let us express

$$A_t(t-1) = A_t(t-2) + R_{t-1} = A_t(t-2) + \Delta\, w(t-1,\, \hat{a}(t-1))\, z(t-1)\, z^{\mathrm{T}}(t-1),$$

and hence

$$\begin{aligned} A_t^{-1}(t-1) &= \left[A_t(t-2) + \Delta\, w(t-1,\, \hat{a}(t-1))\, z(t-1)\, z^{\mathrm{T}}(t-1)\right]^{-1} \\ &= A_t^{-1}(t-2) - A_t^{-1}(t-2)\, z(t-1)\left[z^{\mathrm{T}}(t-1)\, A_t^{-1}(t-2)\, z(t-1) + \frac{1}{\Delta\, w(t-1,\, \hat{a}(t-1))}\right]^{-1} \\ &\quad \times z^{\mathrm{T}}(t-1)\, A_t^{-1}(t-2) \end{aligned}$$

and this relation will be substituted into the previous formula instead of $A_t^{-1}(t-1)$. In this way we will proceed to the last relation

$$A_t(2) = P(t-1) + \alpha^{t-2}\, \Delta\, w(1,\, \hat{a}(t-1))\, z(1)\, z^{\mathrm{T}}(1)$$

and after t steps we will finally obtain the formula for $P(t)^{-1}$ via $P^{-1}(t-1)$, which gives a recursive relation. However, this procedure is very complicated and numerically exacting and its complexity is growing up as t grows. On the other hand, one can expect that for relatively small t the value α^t will be almost negligible and hence the expressions of the form

$$\left[z^{\mathrm{T}}(i)\, P^{-1}(t-1)\, z(i) + \frac{1}{\alpha^t \Delta w(i, \hat{a}(t-1))}\right]^{-1}$$

will be very small as $|\Delta\, w(i,\, \hat{a}(t-1))| \leq 2$.

In a similar way we can also calculate the estimates $\hat{a}(t)$ of unknown parameters a. Because

$$\begin{aligned} \hat{a}(t) &= P^{-1}(t) \sum_{1}^{t} \alpha^{t-i} w(i,\, \hat{a}(t-1))\, z(i)\, y(i) \\ &= P^{-1}(t)\left[\alpha \sum_{1}^{t-1} \alpha^{t-1-i} w(i,\, \hat{a}(t-2))\, z(i)\, y(i) \right. \\ &\quad \left. + \alpha \sum_{1}^{t} \alpha^{t-1-i} \Delta\, w(i,\, \hat{a}(t-2))\, z(i)\, y(i)\right] \end{aligned}$$

we get the following recursive relation

$$\hat{a}(t) = \alpha\, P^{-1}(t)\, P(t-1)\, \hat{a}(t-1) + P^{-1}(t)\alpha \sum_{i=1}^{t} \alpha^{t-1-i} \Delta\, w(i,\, \hat{a}(t-2))\, z(i)\, y(i).$$

The above described algorithm is entitled in such a case only when the unknown parameter a is almost not changing in time. Under more rapid changes it is reasonable to substitute the series

$$\sum_{i=1}^{t} \alpha^{t-1-i} \Delta\, w(i,\, \hat{a}(t-2))$$

by its last member $w(t, \hat{a}(t-2))$ only in the recursive relation calculating $P^{-1}(t)$ and in that calculating $\hat{a}(t)$, too. Then the obtained recursive algorithm would be the same as described in Cipra [2], p. 62. In the next text we will deal with the approach presented in Cipra [2] because it is simpler with respect to calculation than that considered here earlier. We will concentrate mainly to simple and double exponential smoothing procedures, which are commonly used in practice.

2. ROBUST SIMPLE AND DOUBLE EXPONENTIAL SMOOTHING

First, we will derive a general robust exponential smoothing when we follow the approach commonly used in a nonrobust form. We will start from the following model that is quite usual in exponential smoothing, e. g. see Abraham & Ledolter [1], namely

$$y(t+j) = \sum_{k=1}^{p} \beta_k(t)\, z_k(j) + e(t+j), \quad j = 0, \pm 1, \pm 2, \ldots$$

where

$$\beta_t^{\mathrm{T}} = (\beta_1(t), \beta_2(t), \ldots, \beta_p(t))$$

is a vector of unknown parameters, $z^{\mathrm{T}}(j) = (z_1(j), z_2(j), \ldots, z_p(j))$ is a vector of specified basis functions and $\{e(\cdot)\}$ is noise. The parameter t presents a running time, the parameter j expresses a local time in the model considered at the time t. At this time we will assume the knowledge of standard deviation σ (for simplicity we will put $\sigma = 1$). The case of unknown σ will be solved later. This model is usually written in a matrix form

$$y(t+j) = z^{\mathrm{T}}(j)\, \beta(t) + e(t+j).$$

The main idea of exponential smoothing is based on the use of the latest observations for estimating unknown parameters $\beta_k(t)$, which can change in time. In this way we get to the model

$$y(t-j) = z^{\mathrm{T}}(-j)\, \beta(t) + e(t-j), \quad j = 0, 1, 2, \ldots$$

where the influence of older observations is suppressed using a forgetting factor α. The classic method of exponential smoothing uses the estimates obtained from the minimization of a quadratic form

$$\sum_{j=0}^{\infty} \alpha^j (y(t-j) - z^{\mathrm{T}}(-j)\,\beta)^2.$$

In the case of a robust version let's start with the quadratic form

$$\sum_{j=0}^{\infty} \alpha^j \rho(y(t-j) - z^{\mathrm{T}}(-j)\,\beta),$$

where ρ is a suitable function satisfying usual demands on M-estimates. At this moment we leave the question of the existence of a derivative apart and let's perform a formal differentiation of this form. In this way we obtain normal equations

$$\sum_{j=0}^{\infty} \alpha^j \psi(y(t-j) - z^{\mathrm{T}}(-j)\,\beta)\, z(-j) = 0.$$

These normal equations are very difficult to solve and therefore we use the idea of substitution by approximative linear equations again. Let's define adaptive weights

$$w(t-j,\,\beta) = \frac{\psi(y(t-j) - z^{\mathrm{T}}(-j)\,\beta)}{y(t-j) - z^{\mathrm{T}}(-j)\,\beta}$$

where the unknown parameter β will be substituted by a suitable estimate. For complexity, let's put $\frac{\psi(0)}{0} = 1$. As the parameter β can change in course of time it is reasonable to substitute β in the weight $w(t-j,\,\beta)$ by the estimate $\hat{\beta}(t-j-1)$ obtained at the time $t-j-1$. In this way the original weight $w(t-j,\,\beta)$ will be replaced by the weight $w(t-j,\,\hat{\beta}(t-j-1))$. At this moment we get to the system of approximative linear equations

$$\sum_{j=0}^{\infty} \alpha^j\, w(t-j,\,\hat{\beta}(t-j-1))\, z(-j)\,(y(t-j) - z^{\mathrm{T}}(-j)\,\beta) = 0.$$

Now, we must ensure the existence of these infinite series. First, we will discuss the series

$$\sum_{j=0}^{\infty} \alpha^j\, w(t-j,\,\hat{\beta}(t-j-1))\, z(-j)\, y(t-j). \qquad (*)$$

As $y(t-j) = z^{\mathrm{T}}(-j)\,\beta + e(t-j)$, this infinite series is the sum of two series, namely

$$\sum_{j=0}^{\infty} \alpha^j\, w(t-j,\,\hat{\beta}(t-j-1))\, z(-j)\, z^{\mathrm{T}}(-j)\,\beta$$

and

$$\sum_{j=0}^{\infty} \alpha^j\, w(t-j,\,\hat{\beta}(t-j-1))\, z(-j)\, e(t-j).$$

The conditions for existence in the sense a.s. are given in the following

Lemma. Let the noise $\{e(\cdot)\}$ be centered having mutually independent components with finite bounded dispersions. Further, let the following series be convergent

$$\sum_{j=0}^{\infty} \alpha^j \|z(-j)\|,$$

then the series $(*)$ is convergent a.s.

Remark. $\|\cdot\|$ is a usual Euclidean vector norm.

Proof. Let $\{\hat{\beta}(t-j-1)\}$ be quite arbitrary estimates, then thanks to the construction all the weights $w(t-j, \hat{\beta}(t-j-1))$ are positive a.s. and bounded between 0 and 1. Denote

$$S_N^i = \sum_{j=1}^{N} \alpha^j |z_i(-j)| \cdot |e(t-j)|, \quad i = 1, 2, \ldots, p.$$

There is no problem to show that $\{S_N^i\}$ is nonnegative supermartingal for each $i = 1, 2, \ldots, p$ satisfying

$$\mathsf{E}\{S_{N+1}^i / S_N^i\} = S_N^i + \alpha^{N+1} |z_i(-N-1)| \, \mathsf{E}\{|e(t-N-1)|\}.$$

If $\mathsf{E}\{|e(t-j)|\}$ are bounded from above, then thanks to the above inequality one can easily prove the existence of the limit

$$\lim_{N \to \infty} S_N^i$$

in the sense a.s. As $0 < w(t-j, \hat{\beta}(t-j-1)) \leq 1$ then the series

$$\sum_{j=0}^{\infty} \alpha^j w(t-j, \hat{\beta}(t-j-1)) \, z_i(-j) \, e(t-j)$$

is also a.s. convergent.

The convergence of the series

$$\sum_{0}^{\infty} \alpha^j w(t-j, \hat{\beta}(t-j-1)) \, z_i(-j) \, z_k(-j), \quad i, k = 1, 2, \ldots, p$$

is an easy consequence of the convergence of $\sum_{j=0}^{\infty} \alpha^j \|z(t-j)\|^2$, because

$$|z_i(-j) \, z_k(-j)| \leq \frac{1}{2} |z_i(-j)|^2 + \frac{1}{2} |z_k(-j)|^2 \leq \frac{1}{2} \|z(-j)\|^2.$$

□

Then, the unknown parameter β at the time t is estimated as

$$\hat{\beta}_t = \left[\sum_{j=0}^{\infty} \alpha^j \, w(t-j)\, z(-j)\, z^{\mathrm{T}}(-j)\right]^{-1} \sum_{j=0}^{\infty} \alpha^j \, w(t-j)\, y(t-j)\, z(-j)$$

where for the sake of simplicity $w(t-j,\, \hat{\beta}(t-j-1)) = w(t-j)$. Let's try to find a recursive relation between $\hat{\beta}_t$ and $\hat{\beta}_{t+1}$. First, we need a recursive formula between the inversion matrices P_{t+1}^{-1}, P_t^{-1} where

$$P_{t+1} = \sum_{0}^{\infty} \alpha^j \, w(t+1-j)\, z(-j)\, z^{\mathrm{T}}(-j) \quad \text{and} \quad P_t = \sum_{0}^{\infty} \alpha^j \, w(t-j)\, z(-j)\, z^{\mathrm{T}}(-j).$$

It is evident that

$$P(t+1) = w(t+1)\, z(0)\, z^{\mathrm{T}}(0) + \alpha \sum_{j=0}^{\infty} \alpha^j \, w(t-j)\, z(-j-1)\, z^{\mathrm{T}}(-j-1).$$

In order to obtain a recursive formula we are obliged to assume a relation between $z(-k-1)$ and $z(-k)$. This relation is expressed via a matrix operator L in the classical version of exponential smoothing and we will use the same approach, namely

$$z(-k-1) = L^{-1}\, z(-k)$$

then

$$P(t+1) = w(t+1)\, z(0)\, z^{\mathrm{T}}(0) + \alpha\, L^{-1}\, P(t)\, (L^{-1})^{\mathrm{T}}.$$

At this moment we are ready to use the well known lemma on matrix inversion and thus we can write

$$P^{-1}(t+1) = \frac{1}{\alpha}\, L^{\mathrm{T}}\, P^{-1}(t)\, L - \frac{L^{\mathrm{T}} P^{-1} L z(0) z^{\mathrm{T}}(0) L^{\mathrm{T}}(t) L w(t+1)}{\alpha(\alpha + w(t+1) z^{\mathrm{T}}(0) L^{\mathrm{T}} P^{-1}(t) L z(0))}. \qquad (**)$$

The regularity of the matrix $P(t+1)$ can be proved by mathematical induction if we start with any regular symmetric matrix $P(0)$. We also need the regularity of the matrix L. After this we can write

$$\hat{\beta}_{t+1} = P^{-1}(t+1)\, (w(t+1)\, y(t+1)\, z(0) + \alpha\, L^{-1}\, p(t)),$$

where the vector $p(t) \doteq \sum_{k=0}^{\infty} \alpha^k w(t-k)\, y(t-k)\, z(-k)$. Using the recursive relation $(**)$ and after not complicated calculation we can write

$$\hat{\beta}(t+1) = L^{\mathrm{T}}\, \hat{\beta}(t) + \frac{G(t)\, z(0)\, w(t+1)}{\alpha + w(t+1)\, z^{\mathrm{T}}(0)\, G(t)\, z(0)} [y(t+1) - (L\, z(0))^{\mathrm{T}} \hat{\beta}(t)]$$

where we have used the denotation $G(t) = L^{\mathrm{T}}\, P^{-1}(t)\, L$. At the first sight we see that this robust version of exponential smoothing is very similar to the Kalman filter

because $y(t+1) - z^{\mathrm{T}}(1)\,\hat{\beta}(t)$ plays the role of an innovation sequence and the vector function

$$\frac{w(t+1)\,G(t)\,z(0)}{\alpha + z^{\mathrm{T}}(0)\,G(t)\,z(0)\,w(t+1)}$$

can be understood as a gain function. Let's multiply both the sides of the last equality by $z^{\mathrm{T}}(0)$. In this way we obtain the recursive relation for robust exponential smoothing

$$z^{\mathrm{T}}(0)\,\hat{\beta}_{t+1} = (L\,z(0))^{\mathrm{T}}\,\hat{\beta}_t + \frac{z^{\mathrm{T}}(0)\,G(t)\,z(0)\,w(t+1)}{\alpha + z^{\mathrm{T}}(0)\,G(t)\,z(0)\,w(t+1)}[y(t+1) - (L\,z(0))^{\mathrm{T}}\,\hat{\beta}_t].$$

As $L\,z(0) = z(1)$ this last equality gives a relation between robust filtering and robust smoothing because $y(t+1) - z^{\mathrm{T}}(1)\,\hat{\beta}_t$ is an error of one-step-ahead prediction, $z^{\mathrm{T}}(0)\,\hat{\beta}_{t+1}$ is de facto $\hat{y}(t+1|t+1)$, i.e. a filtered value of $y(t+1)$ and $z^{\mathrm{T}}(1)\,\hat{\beta}_t$ is a predicted value $\hat{y}(t+1|t)$ based on information up to time t. Using this facts we can rewrite the last equality into the following form

$$\hat{y}(t+1|t+1) = \hat{y}(t+1|t) + \frac{U(t)\,w(t+1)}{\alpha_1 + U(t)\,w(t+1)}[y(t+1) - \hat{y}(t+1|t)]$$

if $U(t) = z^{\mathrm{T}}(0)\,G(t)\,z(0)$. The relation between $U(t)$ and $U(t+1)$ is not so simple although $U(t) = z^{\mathrm{T}}(1)\,P^{-1}(t)\,z(1)$ as one can simply show. Using the relation between $P^{-1}(t+1)$ and $P^{-1}(t)$ we can easily prove that

$$U(t+1) = \frac{1}{\alpha}\,z^{\mathrm{T}}(2)\,P^{-1}(t)\,z(2) - \frac{1}{\alpha}\,\frac{z^{\mathrm{T}}(2)\,P^{-1}(t)\,z(1)\,z^{\mathrm{T}}(1)\,P^{-1}(t)\,z(2)}{\alpha + w(t+1)\,U(t)}\,w(t+1)$$

which gives a more formal relation, namely,

$$\|z(1)\|^2_{t+1} = \frac{1}{\alpha}\left(\frac{\alpha\|z(2)\|^2_t + w(t+1)\,\left[\|z(2)\|^2_t\,\|z(1)\|^2_t - \langle z(1), z(2)\rangle^2_t\right]}{\alpha + w(t+1)\,\|z(1)\|^2_t}\right)$$

if $\|z(i)\|^2_k = z^{\mathrm{T}}(i)\,P^{-1}(k)\,z(i)$, because $P^{-1}(\cdot)$ is positive definite with probability 1, as follows from the recursive relation.

Next, we will concentrate ourselves on the two most important cases of exponential smoothing used in practice. First, we will deal with simple exponential smoothing, which is described by a model with a locally constant trend

$$y(t+j) = \beta_t + e(t+j). \tag{3}$$

Here, the parameter β is a scalar, $z(j) = 1$ for every j and $L = 1$. Then the recursive algorithm for estimating β is given by

$$\hat{\beta}_{t+1} = \hat{\beta}_t + \frac{w(t+1)\,G(t)}{\alpha + w(t+1)\,G(t)}\,[y(t+1) - \hat{\beta}_t].$$

As easy to see, here $P^{-1}(t) = G(t)$ and hence

$$P^{-1}(t+1) = \frac{P^{-1}(t)}{\alpha + w(t+1)\,P^{-1}(t)}.$$

If, at this moment we compare this result with that given in Cipra [2] for the case of simple exponential smoothing, we see that both the algorithms are identical. This algorithm is relatively very simple and is very similar to the classical simple exponential smoothing because

$$\hat{\beta}_{t+1} = \frac{\alpha}{\alpha + w(t+1)\,P^{-1}(t)}\,\hat{\beta}_t + \frac{w(t+1)\,P^{-1}(t)}{\alpha + w(t+1)\,P^{-1}(t)}\,y(t+1).$$

There is no problem to show that $P(t+1) = \alpha\,P_t + w(t+1)$, which is a very simple relation. Using this, the robust simple exponential smoothing can be expressed as

$$\hat{\beta}_{t+1} = \hat{\beta}_t + \frac{w(t+1)}{\alpha\,P_t + w(t+1)}\,(y(t+1) - \hat{\beta}_t)$$

where

$$P_t = \alpha^t\,P_0 + \sum_{j=1}^{t} \alpha^{t-j}\,w(j).$$

Let's denote $\theta_t = \hat{\beta}_t - \beta_t$, then the above relation can be rewritten as

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \frac{w(t+1)}{\alpha\,P_t + w(t+1)}\,(e(t+1) - \hat{\theta}_t)$$

which gives

$$\hat{\theta}_{t+1} = \frac{\alpha^{t+1}\,\hat{\theta}_0}{\sum_{j=1}^{t+1-j} w(t+1-j)} + \frac{\sum_{j=0}^{t+1} w(t+1-j)\,e(t+1-j)}{\sum_{j=0}^{t+1} \alpha^{t+1-j} w(t+1-j)},$$

if for simplicity we put $P_0 = 1$ and $w(0) = 1$. Thus, we obtained the expression for $\hat{\theta}_{t+1}$ using $\hat{\theta}_0$, $\{e(\cdot)\}$ and the weights $\{w(\cdot)\}$ only.

Remark 1. In case $\alpha < 1$ there is no possibility to investigate an asymptotic behaviour of $\hat{\theta}(t)$ for lack of information. Although the case with $\alpha = 1$ does not belong to exponential smoothing procedures, despite of this fact we can consider the above given algorithm. It would be very nice to prove the consistence of $\hat{\theta}(t)$ in the stable case $\beta_t \to \beta_\infty$ if $t \to \infty$. Such a result would prove the consistence of the proposed robust procedure. But, the author of the paper has not so far met with success in proving such a result when the weights are derived from the classical Huber function. In the following theorem the consistence of $\hat{\theta}(t)$ is proved under the use of a modified Huber function. On the basis of this result we cannot, of course speak about the robustness in the classical sense due to Huber and Hampel, but we can speak about the "practical" robustness because Theorem 1 ensures the stability of this algorithm for $\alpha = 1$ in practice.

The following theorem describes the asymptotic behaviour of $\hat{\theta}_t$ if $\alpha = 1$ and t runs to infinity.

Theorem 1. Let in the model (3) with locally constant trend $\lim_{t\to\infty}\beta_t = \beta$ exist finite and $\{e(\cdot)\}$ be iid random variables with a symmetric distribution function with vanishing mean and finite dispersion. Let adaptive weights $w(\cdot)$ be derived from a modified Huber function $\psi_{\mathrm{HMOD}}(\cdot)$, which is defined as

$$\begin{aligned} &\psi_{\mathrm{HMOD}}(x) = x && \text{for } 0 \le x \le c \\ &\psi_{\mathrm{HMOD}}(x) = c && \text{for } c < x \le C_\infty \\ &\psi_{\mathrm{HMOD}}(x) = \varepsilon(x - C_\infty) + c && \text{for } x \ge C_\infty \\ \text{and}\quad &\psi_{\mathrm{HMOD}}(x) = -\psi_{\mathrm{HMOD}}(-x) && \text{for } x < 0, \end{aligned}$$

where ε is a small positive real number, $c < C_\infty$, $c > C_\infty\varepsilon$. The adaptive weights are then defined as

$$w(x) = \frac{\psi_{\mathrm{HMOD}}(x)}{x}.$$

Then $\hat\theta_t \to 0$ a.s. and in the quadratic mean sense, too.

Proof. Using the relation $\psi_{\mathrm{HMOD}}(x) = w(x)\,x$ the robust algorithm for simple exponential smoothing can be expressed as

$$\hat\theta_{t+1} = \hat\theta_t + \frac{\psi_{\mathrm{HMOD}}(e(t+1) - \hat\theta_t)}{S_{t+1}}$$

where $S_{t+1} = \sum_{j=0}^{t+1} w(j)$, $w(0) = 1$. One can easily show that $\varepsilon < w(x) \le 1$ and hence

$$\varepsilon(t+1) < S_{t+1} \le t+1 \quad \text{a.s.}.$$

Then

$$\hat\theta_{t+1}^2 = \hat\theta_t^2 - 2\hat\theta_t \frac{\psi_{\mathrm{HMOD}}(\hat\theta_t - e(t+1))}{S_{t+1}} + \frac{w^2(t+1)\,(e(t+1) - \hat\theta_t)^2}{S_{t+1}^2}.$$

Let $F_t = \sigma(e(t), e(t-1), \ldots)$. As there exists $\mathsf{E}\,\hat\theta_{t+1}^2$ (it follows from the recursive relation and properties of $\{e(\cdot)\}$), we can consider the conditional expected value

$$\begin{aligned} \mathsf{E}(\hat\theta_{t+1}^2 | F_t) \;=\; & \hat\theta_t^2 - 2\hat\theta_t \mathsf{E}\left\{ \frac{\psi_{\mathrm{HMOD}}(\hat\theta_t - e(t+1))}{S_{t+1}} \middle| F_t \right\} \\ & + \mathsf{E}\left(\frac{w^2(t+1)\,(\theta_t - e(t+1))^2}{S_{t+1}^2} \middle| F_t \right). \end{aligned}$$

The definition of $\psi_{\mathrm{HMOD}}(\cdot)$ gives

$$\begin{aligned} \mathsf{E}\left\{ \frac{w^2(t+1)\,(e(t+1) - \hat\theta_t)^2}{S_{t+1}^2} \middle| F_t \right\} \;&\le\; \mathsf{E}\left\{ \frac{(e(t+1) - \hat\theta_t)^2}{\varepsilon^2(t+1)^2} \middle| F_t \right\} \\ &= \frac{1}{\varepsilon^2}\frac{\sigma^2}{(t+1)^2} + \frac{1}{\varepsilon^2}\frac{1}{(t+1)^2}\hat\theta_t^2 \quad \text{a.s.} \end{aligned}$$

because $\hat{\theta}_t$ is F_t-measurable and $\mathsf{E}\{e^2(t+1)\} = \sigma^2$. In the next step we will show that the quantity

$$\hat{\theta}_t \, \mathsf{E}\left\{ \frac{\psi_{\mathrm{HMOD}}(\hat{\theta}_t - e(t+1))}{S_{t+1}} \middle| F_t \right\} \geq 0 \quad \text{a.s.}.$$

Thanks to independence $e(t+1)$ and F_t we can write

$$\mathsf{E}\left\{ \frac{\psi_{\mathrm{HMOD}}(\hat{\theta}_t - e(t+1))}{S_{t+1}} \middle| F_t \right\} = \int_{-\infty}^{\infty} \frac{\psi_{\mathrm{HMOD}}(\hat{\theta}_t - x)}{S_t + w(|\hat{\theta}_t - x|)} \, \mathrm{d}F(x) \quad \text{a.s.}$$

where $F(\cdot)$ is a symmetric distribution function of $e(t+1)$. The random variable S_t is F_t-measurable. A very important fact following from the definition of $w(\cdot)$ is that the function

$$\frac{\psi_{\mathrm{HMOD}}(x)}{S_t + w(|x|)}$$

is a nondecreasing odd function in $x \in R$, which is negative for $x < 0$, vanishing at 0 and positive for $x > 0$. Then we have

$$\int_{-\infty}^{\infty} \frac{\psi_{\mathrm{HMOD}}(x)}{S_t + w(|x|)} \, \mathrm{d}F(x) = 0.$$

Then, for an arbitrary $\hat{\theta}_t < 0$

$$\frac{\psi_{\mathrm{HMOD}}(x - \hat{\theta}_t)}{S_t + w(|x - \hat{\theta}_t|)} \geq \frac{\psi_{\mathrm{HMOD}}(x)}{S_t + w(|x|)}$$

and hence

$$\int_{-\infty}^{\infty} \frac{\psi_{\mathrm{HMOD}}(x - \hat{\theta}_t)}{S_t + w(|x - \hat{\theta}_t|)} \geq 0,$$

which gives

$$\hat{\theta}_t \, \mathsf{E}\left\{ \left(\frac{\psi_{\mathrm{HMOD}}(e(t+1) - \hat{\theta}_t)}{S_{t+1}} \middle| F_t \right) \right\} \leq 0.$$

A similar approach can be used for the case $\hat{\theta}_t > 0$. Let's denote

$$\eta_t = 2\hat{\theta}_t \, \mathsf{E}\left\{ \left(\frac{\psi_{\mathrm{HMOD}}(\hat{\theta}_t - e(t+1))}{S_{t+1}} \middle| F_t \right) \right\}.$$

We proved that $\eta_t \geq 0$ a.s. for each $t \in \mathcal{N}$. Then

$$\hat{\theta}^2(t+1) \leq \left(1 + \frac{1}{\varepsilon^2 (t+1)^2} \right) \hat{\theta}_t^2 + \frac{1}{\varepsilon^2} \frac{\sigma^2}{(t+1)^2} - \eta_t.$$

As the series $\sum_{t=0}^{\infty} \frac{\sigma^2}{\varepsilon^2 (t+1)} < \infty$, it is possible to use the result on supermartingales, see Robbins and Siegmund [4], which proves the convergence $\{\hat{\theta}_t^2\}$ in the sense a.s.

This result also proves the convergence of the series $\sum_{t=0}^{\infty} \eta_t$, which will be used now to prove $\hat{\theta}_t \to 0$ a.s.

As

$$\begin{aligned} \eta_t &= 2\hat{\theta}_t \, \mathrm{E}\left\{ \frac{\psi_{\mathrm{HMOD}}(\hat{\theta}_t - e(t+1))}{S_t + w(|\hat{\theta}_t - e(t+1)|)} \middle| F_t \right\} \\ &= 2\int_{-\infty}^{\infty} \frac{\hat{\theta}_t \, \psi_{\mathrm{HMOD}}(\hat{\theta}_t - x)}{S_t + w(|\hat{\theta}_t - x|)} \, \mathrm{d}F(x) \end{aligned}$$

and $\sum_{t=1}^{\infty} S_t^{-1} = +\infty$ a.s. although $S_t^{-1} \underset{t\to\infty}{\longrightarrow} 0$, we can η_t expressed as

$$\eta_t = \frac{2}{S_t} \int_{-\infty}^{\infty} \frac{\hat{\theta}_t \, \psi_{\mathrm{HMOD}}(\hat{\theta}_t - x)}{1 + \frac{w(|\hat{\theta}_t - x|)}{S_t}} \, \mathrm{d}F(x)$$

where the last integral must be nonnegative as $\eta_t \geq 0$ a.s. and $S_t > 0$ a.s., too. Thanks to the construction of $\psi_{\mathrm{HMOD}}(\cdot)$

$$\eta_t \geq \frac{2}{t} \int_{-\infty}^{\infty} \frac{\hat{\theta}_t \, \psi_{\mathrm{HMOD}}(\hat{\theta}_t - x)}{1 + \frac{w(\hat{\theta}_t - x)}{S_t}} \, \mathrm{d}F(x) \geq 0 \quad \text{a.s.}$$

and hence

$$\sum_{t=1}^{\infty} \frac{1}{t} \int_{-\infty}^{\infty} \frac{\hat{\theta}_t \, \psi_{\mathrm{HMOD}}(\hat{\theta}_t - x)}{1 + \frac{w(\hat{\theta}_t - x)}{S_t}} < \infty \quad \text{a.s.}$$

But $\sum_1^{\infty} \frac{1}{t} = +\infty$ and hence a subsequence $\{t_j\}_{j=1}^{\infty}$ must exist such that

$$\int_{-\infty}^{\infty} \frac{\hat{\theta}_{t_j} \, \psi_{\mathrm{HMOD}}(\hat{\theta}_{t_j} - x)}{1 + \frac{w(|\hat{\theta}_{t_j} - x|)}{S_{t_j}}} \, \mathrm{d}F(x) \underset{j\to\infty}{\longrightarrow} 0.$$

We know that $\{\hat{\theta}_{t_j}^2\}$ is a.s. convergent. Thus there exists a subsequence $\{\hat{\theta}_{t_{j_k}}\}$ such that

$$\lim_{k\to\infty} \hat{\theta}_{t_{j_k}} = \hat{\theta}_{\infty} \quad \text{a.s.}$$

In this way, we have proved that

$$\theta_{\infty} \int_{-\infty}^{\infty} \psi_{\mathrm{HMOD}}(\theta_{\infty} - x) \, \mathrm{d}F(x) = 0$$

because $|\psi_{\mathrm{HMOD}}(z)| \leq |z|$ and we assume the existence of $\int_{-\infty}^{\infty} z^2 \, \mathrm{d}F(x) = \sigma^2$. At this moment we have two possibilities. Either $\theta_{\infty} = 0$ or

$$\int_{-\infty}^{\infty} \psi_{\mathrm{HMOD}}(\theta_{\infty} - x) \, \mathrm{d}F(x) = 0,$$

which implies thanks to the symmetry of $F(\cdot)$, the property $\psi_{\mathrm{HMOD}}(-z) = -\psi_{\mathrm{HMOD}}(z)$ and its strict monotony in a neighbourhood of 0 that in this case $\theta_{\infty} = 0$, too. This completes the proof. □

Remark 2. The modification of the Huber function $\psi_H(\cdot)$ is necessary owing to the convergence of the series

$$\sum_{t=0}^{\infty} S_{t+1}^{-2}$$

as we need the boundedness of adaptive weights from below $w(\cdot) \geq \varepsilon$. But such an assumption is only a technical matter because C_∞ can be arbitrarily large and ε, in the opposite, can be almost zero but positive. In practice, these assumptions are automatically satisfied by using a computer.

Secondly, we will study double exponential smoothing in detail. Here, we start with the model

$$y(t+j) = \beta^0(t) + \beta^1(t)\, j + e(t+1), \quad j = 0, \pm 1, \pm 2, \ldots$$

and we are looking for estimates $\hat{\beta}^0(t)$, $\hat{\beta}^1(t)$ satisfying approximative normal equations

$$\sum_{j=0}^{\infty} \alpha^j\, w(t-j)\,(y(t-j) - \beta_0 + \beta_1 j) = 0$$

$$\sum_{j=0}^{\infty} \alpha^j\, j\, w(t-j)\,(y(t-j) - \beta_0 + \beta_1 j) = 0,$$

where the adaptive weights $w(t-j)$ are functions of one-step-ahead prediction

$$y(t-j) - \hat{\beta}_0(t-j-1) - \hat{\beta}_1(t-j-1).$$

Here, in the matrix denotation we have

$$\boldsymbol{\beta}_t = \begin{bmatrix} \beta_0(t) \\ \beta_1(t) \end{bmatrix}, \quad \boldsymbol{z}(j) = \begin{bmatrix} 1 \\ -j \end{bmatrix} \quad \text{and} \quad \boldsymbol{L} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Using these relations after a short calculation we obtain

$$\hat{\beta}_0(t+1) = \hat{\beta}_0(t) + \hat{\beta}_1(t) + \frac{w(t+1)\, A_2(t+1)\,[y(t+1) - \hat{\beta}_0(t) - \hat{\beta}_1(t)]}{\alpha^2\, R(t) + w(t+1)\, A_2(t+1)}$$

$$\hat{\beta}_1(t+1) = \hat{\beta}_1(t) + \frac{w(t+1)\, A_1(t+1)\,[y(t+1) - \hat{\beta}_0(t) - \hat{\beta}_1(t)]}{\alpha^2\, R(t) + w(t+1)\, A_2(t+1)}$$

where $A_1(\cdot)$, $A_2(\cdot)$ and $R(\cdot)$ are calculated recursively by

$$\begin{aligned} R(t+1) &= \alpha^2\, R(t) + A_2(t+1)\, w(t+1) \\ A_2(t+1) &= \alpha(A_1(t+1) + A_2(t) + A_1(t)) \\ A_1(t+1) &= \alpha(A_0(t) + A_1(t)) \\ A_0(t+1) &= w(t+1) + \alpha\, A_0(t). \end{aligned}$$

At the first sight we see that this robust version of double exponential smoothing is very similar to the classical version.

3. A ROBUST HOLT'S METHOD

The classical Holt's method, which is frequently used in practice is based on the choice of two independent forgetting factors α_1, α_2 as seen from the following relations

$$\begin{aligned}\hat{\beta}_0(t+1) &= \alpha_1\, y(t+1) + (1-\alpha_1)\,(\hat{\beta}_0(t) + \hat{\beta}_1(t)) \\ &= \hat{\beta}(t) + \hat{\beta}_1(t) + \alpha(y(t+1) - \hat{\beta}_0(t) - \hat{\beta}_1(t)) \\ \hat{\beta}_1(t+1) &= \alpha_2(\hat{\beta}_0(t+1) - \hat{\beta}_0(t)) + (1-\alpha_2)\,\hat{\beta}_1(t) \\ &= \hat{\beta}_1(t) + \alpha_2\alpha_1(y(t+1) - \hat{\beta}_0(t) - \hat{\beta}_1(t))\end{aligned}$$

where $0 < \alpha_i \leq 1,\ i = 1, 2$.

We see that Holt's method is a combination of two double exponential smoothing procedures with forgetting parameters α and β, where $\alpha \geq \beta$. On the basis of this fact we propose the following robust version of Holt's method, namely

$$\begin{aligned}\hat{\beta}_0(t+1) &= \hat{\beta}_0(t) + \hat{\beta}_1(t) + \frac{w(t+1)\,A_2(\alpha, t+1)}{R(\alpha, t+1)}\,[y(t+1) - \hat{\beta}_0(t) - \hat{\beta}_1(t)] \\ \hat{\beta}_1(t+1) &= \hat{\beta}_1(t) + \frac{w(t+1)\,A_1(\beta, t+1)}{R(\beta, t+1)}\,[y(t+1) - \hat{\beta}_0(t) - \hat{\beta}_1(t)]\end{aligned}$$

where $A_2(\alpha, j)$, $R(\alpha, j)$ are derived using a forgetting factor α, $A_1(\beta, j)$ and $R(\beta, j)$ are calculated using a forgetting factor β, where of course $\alpha \geq \beta$ as it is described in the previous robust version of double exponential smoothing.

4. ROBUST VERSION WITH UNKNOWN σ

As written in Introduction the construction of an M-estimate is based on minimization of the "quadratic" form

$$\sum_{j=1}^{t} \rho\left(\frac{y(j) - z^{\mathrm{T}}(j)\,a}{\sigma}\right)\sigma + c\,\sigma$$

with respect to both a and σ (for details see Dutter & Stockinger). Here σ is a standard deviation of noise. In practice we don't know usually the parameter σ and hence the proposed robust version of simple and double exponential smoothing must be modified in this direction. We can follow the approach given in Cipra [2]. First, we must modify adaptive weights $w(t-j,\, \hat{a}(t-j))$, $j = 0, 1, 2, \ldots$. The weight $w(\cdot)$ is defined then as the ratio

$$w(t-j, a) = \frac{\psi\left(\frac{y(t-j) - z^{\mathrm{T}}(t-j)\,a}{\sigma}\right)}{\frac{y(t-j) - z^{\mathrm{T}}(t-j)\,a}{\sigma}}$$

in the case of uknown σ. As the unknown parameter a is substituted by the latest estimate $\hat{a}(t-j-1)$ it is reasonable to estimate σ similarly, i.e. by an estimate

$\hat{\sigma}(t-j-1)$ obtained at the time $t-j-1$. Then the random variable

$$\frac{y(t-j)-z^{\mathrm{T}}(t-j)\,\hat{a}(t-j-1)}{\hat{\sigma}(t-j-1)}$$

is the standardized one-step-ahead prediction error. The calculation of $\hat{\sigma}(t-j)$ can be given by a very simple recursive relation proposed in Cipra [2], namely

$$\hat{\sigma}(t-j)=\gamma|y(t-j)-z^{\mathrm{T}}(t-j)\,\hat{a}(t-j-1)|+(1-\gamma)\,\hat{\sigma}(t-j-1)$$

where γ is chosen close to zero.

Despite of this complication we can state the following very interesting

Theorem 2. Let all the assumptions from Theorem 1 be satisfied and further, let the unknown parameter σ be estimated by a recursive procedure

$$\hat{\sigma}(t+1)=f_{t+1}(\hat{\sigma}(t),e(t+1)),$$

where $\{f_{t+1}(\cdot,\cdot)\}$ are chosen in such a way that $\hat{\sigma}(t)>0$ a.s. for each $t\in\mathcal{N}$. Then $\{\hat{\theta}^2(t)\}$ is a.s. convergent when t runs to infinity.

Proof. The robust algorithm with unknown σ can be rewritten as follows

$$\hat{\theta}_{t+1}=\hat{\theta}_t+\frac{\hat{\sigma}(t)\cdot\psi_{\mathrm{HMOD}}\left(\frac{e(t+1)-\hat{\theta}_t}{\hat{\sigma}(t)}\right)}{S(t)+w(t+1)},$$

with $S(t)=\sum_0^t w(t)$, $w(0)=1$. After making powers on both the sides we obtain

$$\hat{\theta}_{t+1}^2=\hat{\theta}_t^2-2\hat{\theta}_t\frac{\hat{\sigma}(t)\cdot\psi_{\mathrm{HMOD}}\left(\frac{\hat{\theta}(t)-e(t+1)}{\hat{\sigma}(t)}\right)}{S(t)+w(t+1)}+\frac{w^2(t+1)\,(e(t+1)-\hat{\theta}_t)^2}{S^2(t+1)}.$$

Then, thanks to the definitions of $\psi_{\mathrm{HMOD}}(\cdot)$ and $w(\cdot)$ and positivity of $\hat{\sigma}(t)$ we can assert that the function

$$\frac{\psi_{\mathrm{HMOD}}\left(\frac{x}{\hat{\sigma}_t}\right)}{S_t+w\left(\left|\frac{x}{\hat{\sigma}_t}\right|\right)}$$

is nondecreasing and odd. In case $\hat{\theta}_t<0$ we obtain similarly as in the proof of Theorem 1 that

$$\hat{\theta}_t\,\hat{\sigma}_t\int_{-\infty}^{\infty}\frac{\psi_{\mathrm{HMOD}}\left(\frac{x-\hat{\theta}_t}{\hat{\sigma}_t}\right)}{S(t)+w\left(\left|\frac{x-\hat{\theta}_t}{\hat{\sigma}_t}\right|\right)}\,\mathrm{d}F(x)\le 0$$

because

$$\int_{-\infty}^{\infty}\frac{\psi_{\mathrm{HMOD}}\left(\frac{x}{\hat{\sigma}_t}\right)}{S(t)+w\left(\left|\frac{x}{\hat{\sigma}_t}\right|\right)}\,\mathrm{d}F(x)=0.$$

The same approach can be applied in case $\hat{\theta}_t > 0$. There is no problem to show that $\varepsilon < w\left(\frac{x}{\sigma}\right) \leq 1$ for each $x \in R_1$ and each $\sigma > 0$. Using the inequality for supermartingales mentioned in the proof of Theorem 1, we can state that the sequence $\{\{\hat{\theta}_t\}^2\}$ is convergent. But, in order to prove the convergence to zero we need the convergence of the estimate $\{\hat{\sigma}(t)\}$ to any positive finite number. This demand is more complicated owing to a mutual relation binding $\hat{\theta}_t$ and $\hat{\sigma}_t$ together in recursive algorithms. □

ACKNOWLEDGEMENT

The author thanks the referees for their comments and suggestions.

(Received February 14, 1995.)

REFERENCES

[1] B. Abraham and J. Ledolter: Statistical Methods for Forecasting. Wiley, New York 1983.
[2] T. Cipra: Robust exponential smoothing. J. Forecasting *11* (1992), 57–69.
[3] R. Dutter and N. Stockinger: Robust time series analysis: a survey. Supplement to Kybernetika *23* (1987).
[4] H. Robbins and D. Siegmund: A convergent theorem for nonnegative almost supermartingales and some applications. In: Optimizing Methods in Statistics (J. S. Rustagi, ed.), Academic Press, New York 1971, pp. 233–257.

RNDr. Jiří Michálek, CSc., Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 4, 18208 Praha 8. Czech Republic.