Applications of Mathematics

Christophe Croux; Sarah Gelper; Roland Fried
Computational aspects of robust Holt-Winters smoothing based on M-estimation

Applications of Mathematics, Vol. 53 (2008), No. 3, 163–176

Persistent URL: http://dml.cz/dmlcz/140313

COMPUTATIONAL ASPECTS OF ROBUST HOLT-WINTERS SMOOTHING BASED ON M-ESTIMATION*

Christophe Croux, Leuven, Sarah Gelper, Leuven,
Roland Fried, Dortmund

(Invited)

Abstract. To obtain a robust version of exponential and Holt-Winters smoothing the idea of M-estimation can be used. The difficulty is the formulation of an easy-to-use recursive formula for its computation. A first attempt was made by Cipra (Robust exponential smoothing, J. Forecast. *11* (1992), 57–69). The recursive formulation presented there, however, is unstable. In this paper, a new recursive computing scheme is proposed. A simulation study illustrates that the new recursions result in smaller forecast errors on average. The forecast performance is further improved upon by using auxiliary robust starting values and robust scale estimates.

Keywords: Holt-Winters smoothing, robust methods, time series

MSC 2010: 57R10, 62M10, 91B84

1. Introduction

Exponential smoothing is a well known method for smoothing and predicting univariate time series. It is often used because of its easy recursive computation and good forecast performance in practice. In a business environment, for instance, it can be used for predicting sales. Thanks to the decaying weights of far-away observations, the underlying process of the time series is allowed to gradually change over time. In a period of expansion, e.g. when sales go up, the exponential smoothing method can easily be adopted to incorporate trends. This method is called the Holt method. A further extension, the Holt-Winters method, also allows for seasonality. This article focusses on situations where there is only a trend, but the findings can easily be extended to account for seasonality as well. For a recent reference illustrating the practical use of the Holt-Winters method, see e.g. Kotsialos et at. [9].

* This research has been supported by the Research Fund K.U. Leuven and the Fonds voor Wetenschappelijk Onderzoek (Contract number G.0594.05).

The easy-to-use recursive scheme and the good forecast performance for many different processes, see e.g. Chatfield et al. [1], make Holt-Winters smoothing an attractive method. A major drawback of this classical method is that it can be strongly influenced by the presence of outliers in the time series. Due to the recursive formulation, one outlier results in biased smoothing values and predictions for a longer time period. This can be overcome by incorporating ideas from the robustness literature into the smoothing techniques, as discussed in Cipra [2]. He reformulates the optimization problem underlying the classical Holt-Winters method to the M-estimation approach. The difficulty then is to translate the optimization problem into easy recursive formulae, as for the classical smoothing methods. This article studies robust Holt-Winters smoothing based on M-estimation, its formulation in different recursive schemes and their computational aspects. More specifically, we first discuss why the existing method described in Cipra [2] is computationally unstable and propose a new recursive scheme. Secondly, we look at the robustness properties of the methods. Special attention is paid to the starting values and the scale estimation which are needed for the implementation of the recursive computing schemes.

The remainder of this article is organized as follows. In Section 2, we describe the classical simple exponential and Holt-Winters smoothing methods and their formulation as an optimization problem. Section 3 introduces the robust optimization problem in terms of M-estimation and its recursive computation scheme as presented in Cipra [2]. The instability of this method is explained and demonstrated in Section 4. We formulate an alternative recursive scheme for the M-estimation problem and compare both schemes in an example. Section 5 looks at further robustifications of the starting values and the scale estimation in the new recursive scheme. The simulation study in Section 6 compares the forecast performance of various Holt-Winters recursions for both contaminated and uncontaminated data.

2. Exponential and Holt-Winters smoothing

Suppose we observe a univariate time series y_t where t runs from one to T. The classical exponential smoothing method defines the value of the smoothed series at time point t, $\tilde{y}_t$, as the solution of the optimization problem

$$\tilde{y}_t = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{t} (1-\lambda)^{t-i} (y_i - \theta)^2, \tag{2.1}$$

where λ is a fixed number between zero and one, the *smoothing parameter*. At every moment t, the smoothed series $\tilde{y}_t$ equals the solution of equation (2.1). It can be shown that if $\tilde{y}_{t-1}$ is given, the solution of (2.1) is given by

$$\tilde{y}_t = \lambda y_t + (1-\lambda)\tilde{y}_{t-1}. \tag{2.2}$$

As a result, solving problem (2.1) at every time point boils down to a recursive computing scheme. The update equation requires a starting value, which is usually obtained from a startup period consisting of the first m observations, e.g. the mean of the first m observations

$$\tilde{y}_m = \frac{1}{m}\sum_{i=1}^{m} y_i.$$

The exponential smoothing method easily allows to make predictions. At every moment t, the h-step-ahead prediction based on all previous observations $1, \ldots, t$ is denoted by $\hat{y}_{t+h|t}$ and simply equals the most recent smoothed value

$$\hat{y}_{t+h|t} = \tilde{y}_t. \tag{2.3}$$

If the time series y_t contains a trend, however, the exponential smoothing method can be improved upon. The Holt-Winters method explicitly allows for local linear trends in the data. The method was proposed in Holt [8] and Winters [13]. At every time point t, we have a local level a_t and a local linear trend F_t. The local level and trend are the solution of the optimization problem

$$(a_t, F_t) = \operatorname*{argmin}_{a,F} \sum_{i=1}^{t} (1-\lambda)^{t-i}(y_i - (a + Fi))^2. \tag{2.4}$$

Again, λ is the smoothing parameter taking values between zero and one. The smoothed value at time t, $\tilde{y}_t$, then equals the local level a_t:

$$\tilde{y}_t = a_t.$$

Equivalently as in the simple exponential smoothing case, the solution of problem (2.4) can easily be obtained recursively. There is an update equation both for the level and the trend component:

$$\begin{aligned} \tilde{y}_t &= \lambda_1 y_t + (1-\lambda_1)(\tilde{y}_{t-1} + F_{t-1}), \\ F_t &= \lambda_2(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\lambda_2)F_{t-1}. \end{aligned} \tag{2.5}$$

Corresponding to the optimization in equation (2.4), we have $\lambda_1 = \lambda_2 = \lambda$. In practice, however, it is common to allow for different smoothing parameters in the level and the trend equation. To be able to start the recursive calculations, we need starting values. Using the first m observations as the startup period, one usually applies a linear regression fit

$$\hat{y}_t = \hat{\alpha}_0 + \hat{\beta}_0 t, \quad \text{for } t = 1, \ldots, m. \tag{2.6}$$

The parameter estimates for the intercept α_0 and slope β_0 can e.g. be obtained by ordinary least squares. The smoothed series at time m is then given by the fitted value at m, and the trend by the fitted slope parameter:

$$\tilde{y}_m = \hat{\alpha}_0 + \hat{\beta}_0 m \quad \text{and} \quad F_m = \hat{\beta}_0.$$

Forecasts for y_{t+h} can be obtained by linear extrapolation:

$$\hat{y}_{t+h|t} = \tilde{y}_t + F_t h$$

for $t = m, \ldots, T$.

3. A smoothing algorithm based on M-estimation

Following the general idea of M-estimation, Cipra [2] proposed to make both equations (2.1) and (2.4) robust by replacing the square by a suitable loss-function, denoted by ϱ. A suitable loss-function $\varrho(x)$ is assumed to be non-decreasing in $|x|$, to be bounded and to satisfy $\varrho(0) = 0$. For the exponential smoothing problem, we thus get the optimization problem

$$\tilde{y}_t = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{t} (1-\lambda)^{t-i} \varrho\Big(\frac{y_i - \theta}{\sigma_t}\Big), \tag{3.1}$$

and equivalently for the Holt-Winters method

$$(a_t, F_t) = \underset{a,F}{\operatorname{argmin}} \sum_{i=1}^{t} (1-\lambda)^{t-i} \varrho\Big(\frac{y_i - (a + iF)}{\sigma_t}\Big). \tag{3.2}$$

Here, σ_t is the scale of $y_t - \tilde{y}_t$. The role of the auxiliary scale σ_t is important, and more details follow below.

Both optimization problems (3.1) and (3.2) resemble the definition of a weighted M-estimator in linear regression. A well known computational method for solving M-estimation in a regression problem is the iteratively reweighted least squares. Starting from initial parameter estimates, the observations are reweighted and the parameters re-estimated in every iteration step until some convergence criterion is met. For problem (3.2), we obtain the first order condition as a weighted least squares problem:

$$\sum_{i=1}^{t}(1-\lambda)^{t-i}w_{i,t}(a,F)\Big(\frac{y_i-a-iF}{\sigma_t}\Big)z_i=0 \tag{3.3}$$

with

$$z_i=\begin{bmatrix}1\\ i\end{bmatrix}$$

and with $w_{i,t}(a,F)$ the weight given to observation i at time point t,

$$w_{i,t}(a,F)=\psi\Big(\frac{y_i-a-iF}{\sigma_t}\Big)\Big/\Big(\frac{y_i-a-iF}{\sigma_t}\Big). \tag{3.4}$$

Here, $\psi(x)$ is the first order derivative of $\varrho(x)$. Note that $w_{i,t}(a,F)$ in equation (3.4) still depends on t. For a new observation y_t for which we want to obtain a smoothed value, all the weights $w_{i,t}$ need to be re-computed. Recomputing the weights can be time consuming and computationally intensive when t gets large. This can be avoided, though, by making the following approximation of the weights:

$$w_{i,t}(a,F)\approx w_i:=\psi\Big(\frac{y_i-\hat{y}_{i|i-1}}{\hat{\sigma}_{i-1}}\Big)\Big/\Big(\frac{y_i-\hat{y}_{i|i-1}}{\hat{\sigma}_{i-1}}\Big). \tag{3.5}$$

This approximation avoids reweighting in every step and allows the WLS problem to be written recursively. Let z_t denote the vector $(1,t)'$ and $s_t=(a_t,F_t)'$. Using this notation, the following recursive scheme is derived in Cipra [2]:

$$\begin{aligned}
&\text{(i)} && r_t=y_t-z_t'\hat{a}_{t-1},\\
&\text{(ii)} && w_{t-1}=\frac{\hat{\sigma}_{t-1}}{r_t}\psi\Big(\frac{r_t}{\hat{\sigma}_{t-1}}\Big),\\
&\text{(iii)} && M_t=\frac{1}{1-\lambda}\Big(M_{t-1}-\frac{M_{t-1}z_t'z_tM_{t-1}'}{(1-\lambda)/w_{t-1}+z_t'M_{t-1}z_t}\Big),\\
&\text{(iv)} && s_t=s_{t-1}+\frac{M_{t-1}z_t}{(1-\lambda)/w_{t-1}+z_t'M_{t-1}z_t}r_t,\\
&\text{(v)} && \tilde{y}_t=z_t's_t,
\end{aligned} \tag{3.6}$$

for $t=1,\dots,T$. The first equation defines the one-step-ahead forecast error r_t. The second equation calculates the weight w_{t-1} of observation $t-1$. Since we made approximation (3.5), the weights of all previous observations remain unchanged and the weights do not need to be recalculated in every step. The third equation defines an auxiliary quantity M_t which is a (2×2) matrix. Equation (iv) defines s_t, a vector of length two needed to obtain the smoothed value in the last equation (v).

The set of update equations in (3.6) makes use of a scale estimate $\hat{\sigma}_t$ that is obtained by

$$\hat{\sigma}_t = \gamma|r_t| + (1-\gamma)\hat{\sigma}_{t-1} \tag{3.7}$$

for $t > m$. This equation estimates the scale $\hat{\sigma}_t$ using the one-step-ahead forecast errors. The scale estimate is recursively obtained and the smoothing parameter γ should be taken close to zero. The starting value for the scale estimator is given by

$$\hat{\sigma}_m^2 = \frac{1}{m-2}\sum_{i=1}^{m}(y_i - \hat{\alpha}_0 - \hat{\beta}_0 i)^2, \tag{3.8}$$

where $\hat{\alpha}_0$ and $\hat{\beta}_0$ are respectively the intercept and the slope of a linear OLS regression in the startup period, as in equation (2.6). That is, the first scale estimate is the standard deviation of the OLS residuals in the startup period.

To be able to start the recursions in (3.6), Cipra [2] defined the following starting values based on a startup period with m observations:

$$M_m = \left(\sum_{t=1}^{m} z_t z_t'\right)^{-1} \quad \text{and} \quad s_m = M_m \sum_{t=1}^{m} z_t y_t.$$

Equation (3.2) defines the robust smoothing based on M-estimation, and makes use of a ϱ-function. However, in the recursive solution (3.6), only a ψ-function appears, with $\varrho' = \psi$. In the remainder of this paper, we work with the Huber ψ-function

$$\psi(x) = \begin{cases} x & \text{if } |x| < k, \\ \operatorname{sign}(x)k & \text{otherwise,} \end{cases}$$

where it is, based on the standard normal distribution, common to choose $k = 2$. The Huber ψ-function is also used in Cipra [2].

4. Computational stability and the new proposal

The recursive scheme of Cipra [2], presented in equations (3.6), is obtained by making use of the matrix inversion lemma. This allows to write the matrix M_t as in (3.6) (iii):

$$M_t = \left(\sum_{i=1}^{t} \beta^{t-i} w_i z_i z_i'\right)^{-1} = \frac{1}{1-\lambda}\left(M_{t-1} - \frac{M_{t-1} z_t' z_t M_{t-1}'}{(1-\lambda)/w_{t-1} + z_t' M_{t-1} z_t}\right).$$

The denominator of the latter expression can become very small. So due to the use of the matrix inversion lemma, small errors might occur. In a recursive computation scheme, these small errors are accumulated and may result in serious biases. This is illustrated in Fig. 1. In the left panel, we see an artificial time series and its smoothed values as computed by (3.6). A large peak in the smoothed series all of a sudden occurs. This is the result of accumulating rounding errors. For the robust *exponential* smoothing based on M-estimation and using the recursive scheme in (3.6), this problem does not occur. The reason is that for exponential smoothing, M_t reduces to a scalar.

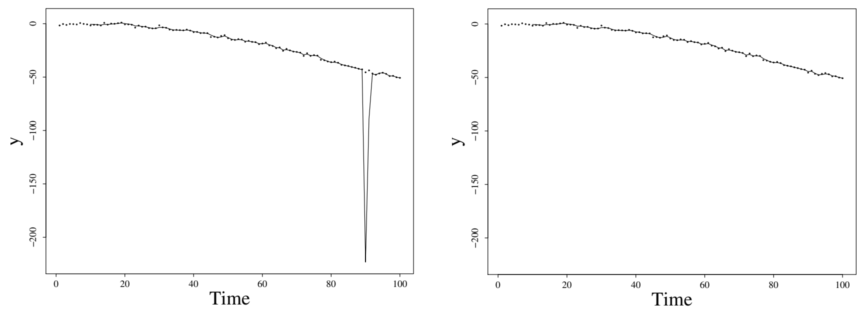

Figure 1. Illustration of the instability of the recursive method proposed by Cipra [2], using artificial data. The dots represent the data and the line the smoothed series according to the update equations (3.6) in the left panel, and according to (4.1) in the right panel. A startup period of $m = 10$ is taken.

To overcome the computational instability of the recursive equations in (3.6) for robust Holt-Winters smoothing, we present another way to compute a solution of the first order condition (3.3). We use weights $\tilde{w}_{i,t}$ defined as

$$\tilde{w}_{i,t} = (1-\lambda)^i \, \frac{\psi\big(y_i - a - iF/\hat{\sigma}_{i-1}\big)}{\big(y_i - a - iF/\hat{\sigma}_{i-1}\big)}$$

to rewrite the first order condition as

$$\sum_{i=1}^{t} \tilde{w}_{i,t} \Big(\frac{y_i - a - iF}{\hat{\sigma}_{i-1}}\Big) z_i = 0.$$

Straightforward calculations lead to the solution

$$a_t = \frac{N_t^y}{N_t^c} - F_t \frac{N_t^x}{N_t^c}, \quad F_t = \frac{N_t^c N_t^{xy} - N_t^x N_t^y}{N_t^c N_t^{xx} - (N_t^x)^2} \quad \text{and} \quad \tilde{y}_t = a_t. \tag{4.1}$$

The following quantities are needed:

$$N_t^c = \sum_{i=1}^{t} \lambda^{t-i} w_i, \quad N_t^y = \sum_{i=1}^{t} \lambda^{t-i} w_i y_i, \quad N_t^x = \sum_{i=1}^{t} \lambda^{t-i} w_i i,$$
$$N_t^{xx} = \sum_{i=1}^{t} \lambda^{t-i} w_i \, i^2, \quad \text{and} \quad N_t^{xy} = \sum_{i=1}^{t} \lambda^{t-i} w_i (i \, y_i).$$

As discussed before, one of the nice properties of the classical Holt-Winters is the recursive representation. The smoothed level a_t and trend F_t as in the robust Holt-Winters method in equation (4.1) can be obtained by computing the above quantities recursively as

$$\begin{aligned}
&\text{(i)} && N_t^c = \lambda N_{t-1}^c + w_t, \\
&\text{(ii)} && N_t^y = \lambda N_{t-1}^y + w_t y_t, \\
&\text{(iii)} && N_t^x = \lambda N_{t-1}^x + w_t t, \\
&\text{(iv)} && N_t^{xx} = \lambda N_{t-1}^{xx} + w_t t^2, \\
&\text{(v)} && N_t^{xy} = \lambda N_{t-1}^{xy} + w_t t \, y_t.
\end{aligned} \tag{4.2}$$

Equations (4.1) and (4.2) define a new recursive scheme for Holt-Winters smoothing based on M-estimation. The new update equations do not involve any matrix inversion and are thus expected to be more stable.

As before, we use a startup period of length m to begin the recursive scheme. The starting values are given by

$$N_m^c = m, \quad N_m^y = \sum_{i=1}^{m} y_i, \quad N_m^x = \sum_{i=1}^{m} i,$$
$$N_m^{xx} = \sum_{i=1}^{m} i^2 \quad \text{and} \quad N_m^{xy} = \sum_{i=1}^{m} i \, y_i.$$

The weights used in the recursive equations in (4.2) depend on the scale estimator. The same scale update recursion as in (3.7) is used here, with the same starting value given by equation (3.8). Note that this update equation is not really robust in the sense that one extreme value of r_t can still make the estimated scale arbitrarily large. The same holds for the starting values. This will be further discussed in Section 5.

In theory, both the solutions for the robust smoothing based on M-estimation, equation (3.6) on the one hand and equations (4.1) and (4.2) on the other hand, should give exactly the same result for the smoothed values $\tilde{y}_t$. This is not the case in practice, however, as can be seen from Fig. 1. The right panel shows the smoothed series as obtained by the new smoothing equations in (4.1) and (4.2). For the new update equations, no sudden bump occurs as in the right panel where the smoothed series according to scheme (3.6) is shown. To illustrate the difference between both methods, we look at a time series where seemingly there is no numerical problem with the Cipra-approach. The left panel of Fig. 2 shows such a time series, together with the smoothed series according to the two different schemes. The right panel plots the difference between these smoothed values, where we see that the difference gradually increases. Although this increase is only visible starting from observation 80, it already accumulates from the beginning. To study the forecasting performance of the two methods in detail, a simulation study is carried out in Section 6.

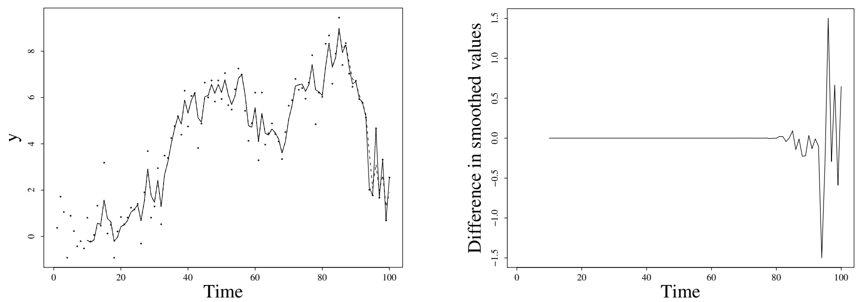

Figure 2. Left panel: a simulated time series (dots) with smoothed values according to recursions (3.6) (full line) and (4.1)–(4.2) (dashed line). Right panel: the difference between the smoothed values according to recursions (3.6) and (4.1)–(4.2).

5. Robustification of starting values and scale

The previous sections describe two recursive formulations for robust smoothing based on M-estimation. Auxiliary equations are needed for the starting values and scale estimates. So far, these are not robust. In this section we describe how to obtain a Holt-Winters smoothing method which is stable and robust for both the update equations and the starting values. As an alternative to the non-robust scale update equation in (3.7), we propose to estimate the scale recursively based on a τ^2-scale estimator as proposed in Yohai and Zamar [14]. We then obtain the recursion

$$\hat{\sigma}_t^2 = \gamma\, \varrho\Big(\frac{r_t}{\hat{\sigma}_{t-1}}\Big)\hat{\sigma}_{t-1}^2 + (1-\gamma)\hat{\sigma}_{t-1}^2, \tag{5.1}$$

where γ is a smoothing parameter and ϱ is a bounded loss function. We take the biweight ϱ-function given by

$$\varrho(x) = \begin{cases} c_k\big(1-(1-(x/k)^2)^3\big) & \text{if } |x| \leqslant k, \\ c_k & \text{otherwise,} \end{cases} \tag{5.2}$$

where c_k is a constant, to achieve consistency of the scale parameter under a normal error distribution. For the common choice of $k = 2$ we have $c_k = 2.52$. As a starting value for the recursion in (5.1) we take the robust τ-scale estimation computed from the residuals of a robust regression in the startup period. We could choose any robust regression method but the repeated median, introduced by Siegel [11], shows good performance in smoothing applications, for references see e.g. Davies et al. [4], Fried [5], and Gather et al. [6]. For the linear regression fit in equation (2.6) the repeated median parameter estimates for the intercept $\hat{\alpha}_0$ and slope $\hat{\beta}_0$ are defined as

$$\hat{\beta}_0 = \operatorname*{med}_i \left(\operatorname*{med}_{j \neq i} \frac{y_i - y_j}{i - j} \right) \quad \text{and} \quad \hat{\alpha}_0 = \operatorname*{med}_i (y_i - \hat{\beta}_0\, i).$$

This regression can also be used to obtain robust starting values for the recursive scheme in (4.1)–(4.2). Therefore, we replace the values of y_t in the startup period by their fitted values $\hat{y}_t$ as obtained from the repeated median regression.

6. Simulation study

To compare the forecast performance of various Holt-Winters smoothing methods, we carry out a simulation study. We compare the performance of four Holt-Winters methods described in the previous sections. The first method is the classical Holt-Winters method, as described by the system of equations (2.5), which will be referred to as HW. The other three methods are based on M-estimation, but they use different recursive schemes and starting values. We compare the weighted regression (WR) method as presented in Cipra [2], using the recursive equations in (3.6), with the new recursive formulae in (4.1)–(4.2). The latter method will be referred to as the new weighted regression (NWR). The fourth method uses the same recursions as the NWR method, but uses robust starting values and robust scale estimation as described in Section 5. We use the abbreviation RNWR for it. For all four methods we use the same fixed smoothing parameter $\lambda = 0.3$. This is an arbitrary choice but other simulations, not reported here, show that the simulation results following below are similar to those for other values of λ. For the scale update equations in (3.7) and (5.1) we choose $\gamma = 0.1$.

The data we use for comparing the forecast performance of the four methods are simulated according to the local linear trend model. In this model, the observed time series y_t is composed of a local level α_t and a local linear trend β_t. More specifically,

$$y_t = \alpha_t + e_t, \quad e_t \sim N(0, \sigma^2), \tag{6.1}$$

where α_t and β_t are given by

$$\begin{aligned} \alpha_t &= \alpha_{t-1} + \beta_{t-1} + \eta_t, && \eta_t \sim N(0, \sigma_\eta^2) \\ \beta_t &= \beta_{t-1} + \nu_t, && \nu_t \sim N(0, \sigma_\nu^2). \end{aligned} \tag{6.2}$$

The noise components e_t, η_t and ν_t have zero mean, are independent of one another and serially uncorrelated. The components η_t and ν_t in equation (6.2) both follow an $N(0, 0.1)$ distribution. For the noise component e_t in equation (6.1) we consider four different scenario's. In the first setting, we look at data without contamination, clean data (CD). The noise component e_t is $N(0, 1)$ distributed and no outliers are included. In the second setting, we include symmetric outliers (SO): every observation is with 95% probability from an $N(0, 1)$ distribution and with 5% probability from an $N(0, 20)$. The outliers have the same mean as the other observations but a larger variance. In the third setting, the outliers have the same variance but a different mean from the bulk of the data, generating asymmetric outliers (AO). In both the SO and AO setting, we do not include outliers in the forecast period, i.e. from observation 201 to 205, as the outlier generating process is unpredictable by definition. The fourth setting simulates the e_t from a fat-tailed (FT) distribution, a Student's-t_3. A summary of the four simulation schemes is presented in Tab. 1

Setting		Description
CD	Clean data	$e_t \overset{iid}{\sim} N(0, 1)$
SO	Symmetric outliers	$e_t \overset{iid}{\sim} (1 - \varepsilon)N(0, 1) + \varepsilon N(0, 20)$, with $\varepsilon = 0.05$
AO	Asymmetric outliers	$e_t \overset{iid}{\sim} (1 - \varepsilon)N(0, 1) + \varepsilon N(20, 1)$, with $\varepsilon = 0.05$
FT	Fat tailed data	$e_t \overset{iid}{\sim} t_3$

Table 1. Simulation schemes.

For each of the four simulation settings in Tab. 1, we simulate $N = 1000$ time series of length 205. For each of the N simulated time series, we apply four smoothing methods up to observation 200, and forecast observations 201 and 205. The forecasts are then compared with the corresponding realized values. As such we obtain a series of N forecast errors, both one- and five-steps-ahead, for each of the four methods considered and for each of the four settings presented in Tab. 1. We compare the forecast errors according to two criteria. The first criterion is the Mean Squared Forecast Error (MSFE) which is defined as

$$\mathrm{MSFE}(r_1, \ldots, r_N) = \sum_{i=1}^{N} r_i^2,$$

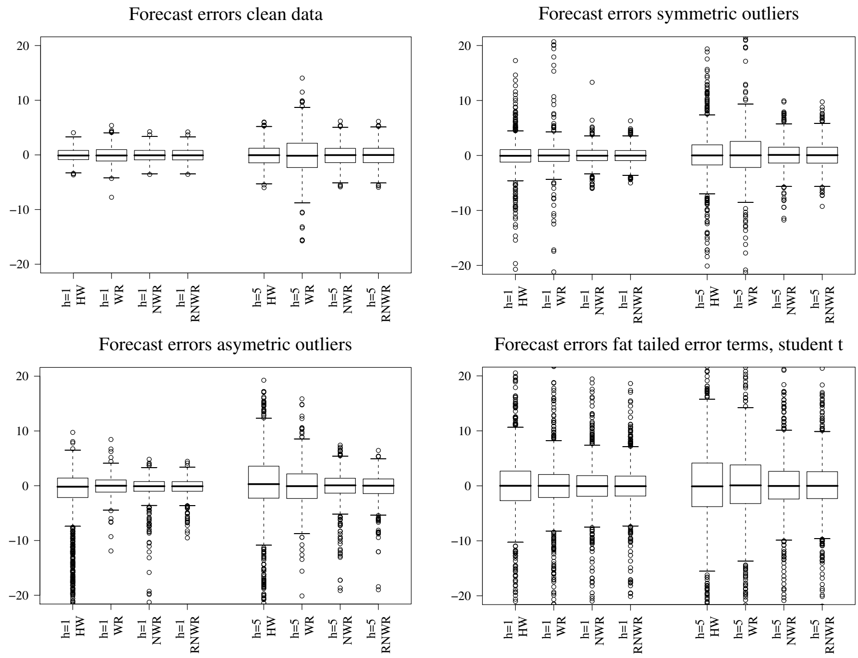

Figure 3. Boxplots of one-step-ahead ($h = 1$) and five-step-ahead ($h = 5$) forecast errors for 1000 replications of a time series of length 200, simulated according to simulation schemes CD (clean data, top left), SO (symmetric outliers, top right), AO (asymmetric outliers, bottom left) and FT (Student's-t error terms, bottom right). We consider standard Holt-Winters (HW), Weighted Regression (WR), Weighted Regression according to the new recursive scheme (NWR) and with robust starting values and scale estimation (RNRW).

where $r_1, \ldots r_N$ denote a series of forecast errors. Large prediction errors have a big influence on the MSFE, as all errors enter the sum as a square. To have a better idea about the spread of the bulk of the forecast errors, we also look at the τ^2-scale measure

$$\tau^2(r_1, \ldots, r_N) = s_N^2 \frac{1}{N} \sum_{i=1}^{N} \varrho\Big(\frac{r_i}{s_N}\Big),$$

where $s_N = \text{Med}_i\, |r_i|$ and we use the biweight ϱ-function as defined in equation (5.2). In particular, for fat-tailed error distributions, like in simulation scheme FT, the τ^2-scale measure is appropriate to use.

The simulation results are graphically presented in Fig. 3. The upper-left panel shows one- and five-step-ahead forecast errors for the clean data setting. The RW method performs bad as compared to the other three methods. This can also be observed from Tab. 2 which presents the MSFE and τ^2 measure of the one-step-ahead forecast errors. Both the MSFE and τ^2 measure are high for the WR method, while the other methods yield comparable results. The performance of the NWR and RNWR methods is almost as good as that of the classical HW method when no outliers are present.

		HW	WR	NWR	RNWR
CD	MSFE	**1.62**	2.47	1.65	1.64
	τ^2	**1.02**	1.55	**1.02**	**1.02**
SO	MSFE	8.65	24.42	2.38	**2.08**
	τ^2	1.84	1.65	1.20	**1.17**
AO	MSFE	43.78	106.63	7.95	**3.03**
	τ^2	3.86	1.64	1.12	**1.08**
FT	MSFE	3227.58	4543.68	2574.95	**2546.67**
	τ^2	10.08	6.09	4.96	**4.61**

Table 2. The MSFE and τ^2-scale for the one-step-ahead forecast errors in the four simulation settings. The smallest value for each row is in bold.

Not only in the CD setting, but also in the contaminated settings SO and AO as well as in the fat-tailed setting, the WR method shows large forecast errors. In every setting, the MSFE of the WR method is larger than that of the HW method. This shows the need for new robust Holt-Winters methods. The NWR method performs much better, it has always smaller MSFE and τ^2 than the WR method. However, the boxplots in Fig. 3 indicate that, especially in presence of asymmetric outliers, the NWR method still shows serious prediction bias. Overall, the RNWR method has smaller MSFE and τ^2 measures than the other methods. This shows the usefulness of robust starting values and scale estimation in combination with the new update formulae.

7. Conclusion

This article studies robust Holt-Winters smoothing based on M-estimation. The existing recursive scheme for Holt-Winters M-estimation, as formulated in [2], inherits instabilities due to the use of the matrix inversion lemma. Small errors are accumulated and result in large biases. We present an alternative recursive scheme for the same optimization problem. As the Holt-Winters smoothing algorithm is often used to obtain forecasts, the simulation study compares different recursive schemes according to forecast-accuracy criteria. It is illustrated that the new recursive scheme performs better, both in presence and absence of outliers. Moreover, further robustification of the starting values and the scale estimates allow to obtain even more accurate predictions, in particular in presence of severe outliers. Note that using the M-estimation approach is not the only possibility for robustifying the Holt-Winters method. Other robust Holt-Winters methods have been discussed e.g. in [3], [10], [12] and [7].

References

[1] *C. Chatfield, A. Koehler, J. Ord, R. Snyder*: A new look at models for exponential smoothing. The Statistician *50* (2001), 147–159.

[2] *T. Cipra*: Robust exponential smoothing. J. Forecast. *11* (1992), 57–69.

[3] *T. Cipra, R. Romera*: Kalman filter with outliers and missing observations. Test *6* (1997), 379–395.

[4] *P. Davies, R. Fried, U. Gather*: Robust signal extraction for on-line monitoring data. J. Stat. Plann. Inference *122* (2004), 65–78.

[5] *R. Fried*: Robust filtering of time series with trends. J. Nonparametric Stat. *16* (2004), 313–328.

[6] *U. Gather, K. Schettlinger, R. Fried*: Online signal extraction by robust linear regression. Comput. Stat. *21* (2006), 33–51.

[7] *S. Gelper, R. Fried, C. Croux*: Robust forecasting with exponential and Holt-Winters smoothing. Preprint. 2007.

[8] *C. Holt*: Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum 52. 1959.

[9] *A. Kotsialos, M. Papageorgiou, A. Poulimenos*: Long-term sales forecasting using Holt-Winters and neural network methods. J. Forecast. *24* (2005), 353–368.

[10] *R. Romera, T. Cipra*: On practical implementation of robust Kalman filtering. Comm. Stat., Simulation Comput. *24* (1995), 461–488.

[11] *A. Siegel*: Robust regression using repeated medians. Biometrika *69* (1982), 242–244.

[12] *J. Taylor*: Forecasting daily supermarket sales using exponentially weighted quantile regression. Eur. J. Oper. Res. *178* (2007), 154–167.

[13] *P. Winters*: Forecasting sales by exponentially weighted moving averages. Manage. Sci. *6* (1960), 324–342.

[14] *V. Yohai, R. Zamar*: High breakdown-point estimates of regression by means of the minimization of an efficient scale. J. Am. Stat. Assoc. *83* (1988), 406–413.

Authors' addresses: *C. Croux*, *S. Gelper*, Fac. of Economics and Business, K. U. Leuven, Naamsestraat 69, 3000 Leuven, Belgium, e-mail: `christophe.croux@econ.kuleuven.be`, `sarah.gelper@econ.kuleuven.be`; *R. Fried*, Department of Statistics, University of Dortmund, 44221 Dortmund, Germany, e-mail: `fried@statistik.uni-dortmund.de`.