

Robust Exponential Smoothing of Multivariate Time Series

Christophe Croux^{*,a}, Sarah Gelper^b, Koen Mahieu^a

^a*Faculty of Business and Economics, K.U.Leuven, Naamsestraat 69, 3000 Leuven, Belgium*

^b*Erasmus University Rotterdam, Erasmus School of Economics, Burgemeester Oudlaan 50, 3000, Rotterdam, The Netherlands*

Abstract

Multivariate time series may contain outliers of different types. In presence of such outliers, applying standard multivariate time series techniques becomes unreliable. A robust version of multivariate exponential smoothing is proposed. The method is affine equivariant, and involves the selection of a smoothing parameter matrix by minimizing a robust loss function. It is shown that the robust method results in much better forecasts than the classic approach in presence of outliers, and performs similar when the data contain no outliers. Moreover, the robust procedure yields an estimator of the smoothing parameter less subject to downward bias. As a byproduct, a cleaned version of the time series is obtained, as is illustrated by means of a real data example.

Key words: Data cleaning, Exponential smoothing, Forecasting, Multivariate time series, Robustness

*Corresponding author

Email addresses: `christophe.croux@econ.kuleuven.be` (Christophe Croux),
`gelper@ese.eur.nl` (Sarah Gelper), `koen.mahieu@econ.kuleuven.be` (Koen Mahieu)

1. Introduction

Exponential smoothing is a popular technique used to forecast time series. Thanks to its very simple recursive computing scheme, it is easy to implement. It has been shown to be competitive with respect to more complicated forecasting methods. A multivariate version of exponential smoothing was introduced by Jones (1966) and further developed by Pfefferman and Allon (1989). For a given multivariate time series $\mathbf{y}_1, \dots, \mathbf{y}_T$, the smoothed values are given by

$$\hat{\mathbf{y}}_t = \mathbf{\Lambda} \mathbf{y}_t + (\mathbf{I} - \mathbf{\Lambda}) \hat{\mathbf{y}}_{t-1}, \quad (1)$$

for $t = 2, \dots, T$, where $\mathbf{\Lambda}$ is the *smoothing matrix*. The forecast that we can make at moment T for the next value \mathbf{y}_{T+1} is then given by

$$\hat{\mathbf{y}}_{T+1|T} = \hat{\mathbf{y}}_T = \mathbf{\Lambda} \sum_{k=0}^{T-1} (\mathbf{I} - \mathbf{\Lambda})^k \mathbf{y}_{T-k}. \quad (2)$$

The forecast in (2) is a weighted linear combination of the passed values of the series. Assuming the matrix sequence $(\mathbf{I} - \mathbf{\Lambda})^k$ converges to zero, the weights decay exponentially fast and sum to the identity matrix \mathbf{I} . The forecast given in (2) is optimal when the series follows a vector IMA(1,1) model, see Reinsel (2003, page 51). The advantage of a multivariate approach is that for forecasting one component of the multivariate series, information from all components is used. Hence the covariance structure can be exploited to get more accurate forecasts. In this paper, we propose a robust version of the multivariate exponential smoothing scheme.

Classic exponential smoothing is sensitive to outliers in the data, since they affect both the update equation (1) for obtaining the smoothed values and equation (2) for computing the forecast. To alleviate this problem,

Gelper, Fried, and Croux (2009a) proposed a robust approach for *univariate* exponential smoothing. In the multivariate case the robustness problem becomes even more relevant, since an outlier in one component of the multivariate series \mathbf{y}_t will affect the smoothed values of all series. Generalizing the approach of Gelper et al. (2009a) to the multivariate case raises several new issues.

In the univariate case, the observation at time t is said to be outlying if its corresponding one-step-ahead prediction error $y_t - \hat{y}_{t|t-1}$ is large, say larger than twice the robust scale estimate of the prediction errors. A large prediction error means that the value of y_t is very different from what one expects, and hence indicates a possible outlier. In a multivariate setting the prediction errors are vectors. We declare then an observation as outlying if the robust Mahalanobis distance between the corresponding one-step-ahead prediction error and zero becomes too large. Computing this Mahalanobis distance requires a local estimate of multivariate scale.

Another issue is the selection of the smoothing matrix $\mathbf{\Lambda}$ used in equation (1). The smoothing matrix needs to be chosen such that a certain loss function computed from the one-step-ahead prediction errors is minimized. As loss function we propose the determinant of a robust estimator of the multivariate scale of the prediction errors.

In Section 2 of this paper we describe the robust multivariate exponential smoothing procedure. Its recursive scheme allows us both to detect outliers and to “clean” the time series. It then applies classic multivariate exponential smoothing to the cleaned series. The method is affine equivariant, making it different from the approach of Lanius and Gather (2009). In Section 3

we show by means of simulation experiments the improved performance of the robust version of exponential smoothing, both for forecasting and for selecting the optimal smoothing matrix. Section 4 elaborates on the use of the cleaned time series, an important byproduct of applying robust multivariate exponential smoothing. This cleaned time series can be used as an input for more complicated time series methods. We illustrate this in a real data example, where the parameters of a Vector Autoregressive (VAR) model are estimated from the cleaned time series. Finally, Section 5 contains some conclusions and ideas for further research.

2. Robust Multivariate Exponential Smoothing

At each time point t we observe a p -dimensional vector \mathbf{y}_t , for $t = 1, \dots, T$. Exponential smoothing is defined in a recursive way. Assume that we already computed the smoothed values of $\mathbf{y}_1, \dots, \mathbf{y}_{t-1}$. To obtain a robust version of the update equation (1), we simply replace \mathbf{y}_t in (1) by a “cleaned” version \mathbf{y}_t^* for any t . We now detail how this cleaned value can be computed. Denote the one-step-ahead forecast error

$$\mathbf{r}_t = \mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}, \quad (3)$$

being a vector of length p , for $t = 2, \dots, T$. The multivariate cleaned series is given by

$$\mathbf{y}_t^* = \frac{\psi_k \left(\sqrt{\mathbf{r}_t' \hat{\Sigma}_t^{-1} \mathbf{r}_t} \right)}{\sqrt{\mathbf{r}_t' \hat{\Sigma}_t^{-1} \mathbf{r}_t}} \mathbf{r}_t + \hat{\mathbf{y}}_{t|t-1} \quad (4)$$

where $\psi_k = \min(k, \max(x, -k))$ is the Huber ψ -function with boundary value k , and $\hat{\Sigma}_t$ is an estimated covariance matrix of the one-step-ahead forecast

error at time t . If k tends to infinity, $\mathbf{y}_t^* = \mathbf{y}_t$, implying that no data cleaning takes place and that the procedure reduces to classic exponential smoothing. Formula (4) is similar to the one proposed by (Masreliez, 1975) in the univariate case.

Estimation of scale: Since the covariance matrix of the \mathbf{r}_t is allowed to depend on time, it needs to be estimated locally. We propose, similar as in Cipra (1992) and Gelper et al. (2009a) for the univariate setting, the following recursive formula

$$\hat{\Sigma}_t = \lambda_\sigma \frac{\rho_{c,p} \left(\sqrt{\mathbf{r}'_t \hat{\Sigma}_{t-1}^{-1} \mathbf{r}_t} \right)}{\mathbf{r}'_t \hat{\Sigma}_{t-1}^{-1} \mathbf{r}_t} \mathbf{r}_t \mathbf{r}'_t + (1 - \lambda_\sigma) \hat{\Sigma}_{t-1} \quad (5)$$

where $0 < \lambda_\sigma < 1$ is an a priori chosen smoothing constant. For λ_σ close to zero, the importance of the incoming observation at time t is rather small, and the scale estimate will vary slowly over time, whereas for λ_σ close to one, the importance of the new observation is too large. Our simulation experiments indicated that $\lambda_\sigma = 0.2$ is a good compromise. Alternatively, one could consider a finite grid of values for λ_σ and choose the one in the grid that minimizes the determinant of a robust estimator of the covariance matrix of the forecast errors.

The real valued function $\rho_{c,p}$ is the biweight ρ -function with tuning constant c

$$\rho_{c,p}(x) = \begin{cases} \gamma_{c,p} \left(1 - (1 - (x/c)^2)^3 \right) & \text{if } |x| \leq c \\ \gamma_{c,p} & \text{otherwise,} \end{cases}$$

where the constant $\gamma_{c,p}$ is selected such that $E[\rho_{c,p}(\|X\|)] = p$, where X is a p -variate normal distribution. An extremely large value of \mathbf{r}_t will not affect the local scale estimate, since the ρ -function is bounded. The constant

k in the Huber ψ -function and c in the biweight function are taken as the square root of the 95% quantile of a chi-squared distribution with p degrees of freedom. The choice of the biweight $\rho_{c,p}$ function is common in robust scale estimation, and was also taken in Gelper et al. (2009a).

Starting values: The choice of the starting values for the recursive algorithm is crucial. For a startup period of length $m > p$, we fit the multivariate regression model $\hat{\mathbf{y}}_t = \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\beta}}t$ using the robust affine equivariant estimation method of Rousseeuw et al. (2004). We prefer a linear robust fit since exponential smoothing can also be applied on integrated time series, exhibiting local trends. Then we set $\hat{\mathbf{y}}_m = \hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\beta}}m$, and we take for $\hat{\boldsymbol{\Sigma}}_m$ a robust estimate of the covariance matrix of the residuals of this regression fit. The length of the startup period needs to be taken large enough to ensure that $\hat{\boldsymbol{\Sigma}}_m$ will have full rank. Then we start up the recursive scheme

$$\hat{\mathbf{y}}_t = \boldsymbol{\Lambda}\mathbf{y}_t^* + (\mathbf{I} - \boldsymbol{\Lambda})\hat{\mathbf{y}}_{t-1}, \quad (6)$$

where the cleaned values are computed as in (4), and the scale is updated using (5), for any $t > m$. Given that the startup values are obtained in a robust way, and that the ψ and ρ function are bounded, it is readily seen that the effect of huge outliers on the smoothed series remains limited.

Affine equivariance: An important property of the proposed procedure is affine equivariance. If we consider the time series $\mathbf{z}_t = B\mathbf{y}_t$, with B a non-singular $p \times p$ matrix, then the cleaned and smoothed series are given by $\mathbf{z}_t^* = B\mathbf{y}_t^*$ and $\hat{\mathbf{z}}_t = B\hat{\mathbf{y}}_t$. Applying univariate robust exponential smoothing on each component separately will not have this affine equivariance property.

Selection of the smoothing parameter matrix: Both the robust and classic

multivariate exponential smoothing and forecasting method depend on a smoothing matrix $\mathbf{\Lambda}$. We propose to select $\mathbf{\Lambda}$ using a data-driven approach, on the basis of the observed time series during a certain training period. After this training period, the matrix $\mathbf{\Lambda}$ remains fixed. More precisely, $\mathbf{\Lambda}$ is selected by minimizing the determinant of the estimated covariance matrix of the one-step-ahead forecast errors. As a further simplification, we assume that the smoothing matrix is symmetric. While in the univariate case $\mathbf{\Lambda}$ is simply a scalar in the closed interval $[0, 1]$, in the multivariate case we require that $\mathbf{\Lambda}$ is a matrix with all eigenvalues in $[0, 1]$, similar as in Pfefferman and Allon (1989). Let $R := \{\mathbf{r}_{m+1}, \dots, \mathbf{r}_T\}$ be the set of the one-step-ahead forecast errors, then

$$\mathbf{\Lambda}_{\text{opt}} := \underset{\mathbf{\Lambda} \in S_1(p)}{\operatorname{argmin}} \det \widehat{\operatorname{Cov}}(R), \quad (7)$$

where $S_1(p)$ is the set of all $p \times p$ symmetric matrices with all eigenvalues in the interval $[0, 1]$.

For classic multivariate exponential smoothing, the estimator of the covariance matrix of the one-step-ahead forecast errors is just taken equal to the sample covariance matrix with mean fixed at zero:

$$\widehat{\operatorname{Cov}}(R) := \hat{\Sigma}(R) = \frac{1}{T - m} \sum_{t=m+1}^T \mathbf{r}_t \mathbf{r}_t'. \quad (8)$$

The one-step-ahead forecast errors \mathbf{r}_t will contain outliers at the places where the observed series has outliers. Therefore we use a robust estimation of the covariance matrix called *Minimum Covariance Determinant* (MCD) estimator (Rousseeuw and Van Driessen, 1999). For any integer h such that $1 \leq h \leq T - m$ define

$$L^h = \{A \subset R \mid \#A = h\} \subset 2^R$$

of all subsamples of size h of the one-step-ahead forecast errors. This set is finite for $T \in \mathbb{N}$, hence there exists a set $L_{\text{opt}} \in L^h$ such that

$$L_{\text{opt}} = \underset{A \in L^h}{\operatorname{argmin}} \det \hat{\Sigma}(A),$$

where $\hat{\Sigma}(A)$ is the sample covariance matrix (with mean equal to zero) of the subsample $A \subset R$, as in (8). We define the MCD estimator of scale as

$$\hat{\Sigma}_{\text{MCD}}^{(h)}(R) := \hat{\Sigma}(L_{\text{opt}}).$$

A common choice in the literature is $h = \lfloor \frac{T-m+p+1}{2} \rfloor$, which yields the highest breakdown point, but low efficiency. We take $h = \lfloor 0.75(T-m) \rfloor$ which is still resistant to outliers (25% breakdown point), but has a higher efficiency (Croux and Haesbroeck, 1999).

3. Simulation Study

In this section we study the effect of *additive outliers* and *correlation outliers* on both the classic and the robust multivariate exponential smoothing method. We compare the one-step-ahead forecast accuracy, and the selection of the smoothing parameter matrix by both methods. Forecast accuracy is measured by the determinant of the MCD estimator on the scatter of the one-step-ahead forecast errors. We prefer to use a robust measure of forecast accuracy, since we want to avoid that the forecasts made for unpredictable outliers dominate the analysis.

We generate time series $\mathbf{y}_1, \dots, \mathbf{y}_T$ from a multivariate random walk plus noise model:

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \tag{9}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}_t,$$

for $t = 1, 2, \dots$, with $\boldsymbol{\mu}_0 = 0$, and where $\{\boldsymbol{\varepsilon}_t\}$ and $\{\boldsymbol{\eta}_t\}$ are two independent serially uncorrelated zero mean bivariate normal processes with constant covariance matrices $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_\eta$ respectively. In Harvey (1986) it is shown that, if there exists a $q \in \mathbb{R}$ (the so-called signal-to-noise ratio) such that $\boldsymbol{\Sigma}_\eta = q\boldsymbol{\Sigma}_\varepsilon$, the theoretical optimal smoothing matrix for the classic method is given by

$$\boldsymbol{\Lambda}_{\text{opt}} = \frac{-q + \sqrt{q^2 + 4q}}{2} I_p, \quad (10)$$

where I_p is the $p \times p$ identity matrix.

3.1. Forecast accuracy

We generate $M = 1000$ time series from model (9) with

$$\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \text{ and } q = \frac{1}{4}.$$

We consider four different sampling schemes. In the first scheme, the data are clean or uncontaminated. The second and third sampling scheme consider additive outliers. In the second scheme, 10% contamination is added to the first component of the multivariate time series. More specifically, we include additive outliers with a size of $K = 12$ times the standard deviation of the error term. The third scheme is similar to the second scheme, but here both components contain 5% contamination, yielding 10% overall contamination. The outliers are added such that they do not occur at the same time points in both time series. In the description of the results, we refer to the second and third simulation schemes as ‘Additive1’ and ‘Additive2’ respectively. In the last sampling scheme, we include 10% correlation outliers by reversing the sign of the off-diagonal elements in the correlation matrix $\boldsymbol{\Sigma}_\varepsilon$.

To compare the performance of the classic and the robust exponential smoothing schemes, we focus on the one-step-ahead forecast errors. Since these are multivariate, they are summarized by the value of the determinant of their covariance matrix, as estimated by the MCD, averaged over all M simulation runs. Outliers are expected to affect the multivariate smoothing procedure in two ways. There is a direct effect on the forecasted value and an indirect effect via the selection of the smoothing matrix $\mathbf{\Lambda}$. To be able to distinguish between these two effects, we first study the forecast performance using the known value of the optimal smoothing matrix $\mathbf{\Lambda}_{\text{opt}}$ as given in equation (10). In a second experiment, $\mathbf{\Lambda}$ is chosen in a data driven manner as explained in Section 2.

In the first experiment, where we use the optimal $\mathbf{\Lambda}$ according to equation (10), we consider time series of lengths $T = 20, 40, 60$ and 100 . A startup period of $m = 10$ is used and the one-step-ahead forecast errors \mathbf{r}_t are evaluated over the period $t = m + 1, \dots, T$. Table 1 reports the average determinant of the MCD estimator of the forecast error covariance matrix over 1000 simulation runs. When the difference between the classic and the robust procedure is significant at the 5 % level, as tested for by a paired t -test, the smallest value is reported in bold.

Table 1 shows that for uncontaminated data, the classic approach is slightly better than the robust approach, but the difference is very small for longer time series. When additive outliers are included, however, the robust procedure clearly outperforms the classical one. There is no clear difference in forecast accuracy between the second and the third simulation setting from which we conclude that the proposed procedure can easily deal

Table 1: Average value, over 1000 simulation runs, of the determinant of the MCD estimator of the one-step-ahead forecast errors, for a test period of length $n = 20, 40, 60, 100$, and for 4 different sampling scheme. If the difference between the classic non-robust (C) and the robust (R) method is significant at 5 %, the smallest value is reported in bold. The smoothing matrix is set at its theoretical optimal value.

n	Clean		Additive1		Additive2		Correlation	
	C	R	C	R	C	R	C	R
20	2.58	3.08	12.02	5.64	11.46	5.62	2.84	3.32
40	2.23	2.34	10.65	4.22	11.43	4.14	2.31	2.41
60	2.13	2.20	10.58	4.00	11.34	3.91	2.28	2.34
100	2.07	2.11	10.46	3.90	11.11	3.73	2.17	2.21

Table 2: As Table 1, but now with the smoothing matrix estimated from the data.

n	Clean		Additive1		Additive2		Correlation	
	C	R	C	R	C	R	C	R
20	3.14	3.58	17.99	6.58	17.15	6.45	3.32	3.49
40	3.22	3.41	22.00	7.40	22.88	5.82	3.58	3.33
60	3.42	3.77	25.70	8.51	27.87	6.85	3.60	3.50
100	3.66	4.27	32.05	9.62	41.61	9.29	4.30	4.49

with additive outliers in all components of a multivariate series. Finally, we compare the performance of both methods for uncontaminated data and data including correlation outliers. From Table 1 it is clear that the forecast performance of either methods is hardly affected by the correlation outliers. The difference between the classic and the robust approach remains small.

The difference between the robust and classic approach is most visible for additive outliers with size $K = 12$ standard deviations of the error term. One might wonder how the results depend on the value of K . In Figure 1 we plot the magnitude of the forecast errors, as measured by the value of the determinant of the MCD estimator of the one-step-ahead forecast errors averaged over 1000 simulations, and with $n = 100$, for $K = 0, 1, \dots, 12$. We see that up to $K = 3$, the performance is very similar. Hence for small additive outliers, there is not much difference between the two methods. However, for moderate to extreme outliers, the advantage of using the robust method is again clear. Note that while the magnitude of the forecast errors continues to increase with K for the classical method, this is not the case for the robust method. The effect of placing additive outliers at $K = 6$ or at $K = 12$ on the robust procedure is about the same.

In practice, the optimal smoothing matrix is unknown. We therefore consider a second experiment where the selection of the smoothing matrix is data-driven based on a training period of length k , as described in detail in Section 2. We generate time series of lengths $T = k + 20, k + 40, k + 60$ and $k + 100$ and use a training period of $k = 50$ observations including a startup period of length $m = 10$. Similar as in the previous experiment, the forecast accuracy is evaluated by the average determinant of the MCD estimator for

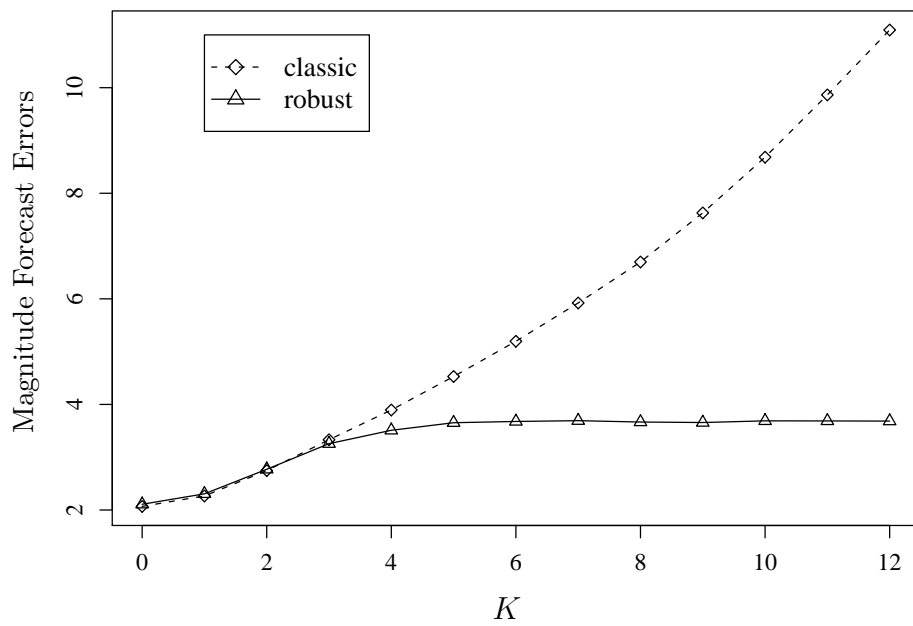


Figure 1: Average value, over 1000 simulation runs, of the determinant of the MCD estimator of the one-step-ahead forecast errors, for a test period of length $n = 100$, and for the Additive2 simulation scheme, as a function of the size K of the outliers

the covariance matrix of \mathbf{r}_t , where $t = k + 1, \dots, T$.

The results of this second, more realistic, experiment are reported in Table 2. First of all, notice that there is a loss in statistical efficiency due to fact that the smoothing matrix needs to be selected. For uncontaminated data, both methods perform comparably. Including additive outliers strongly affects the forecast accuracy of the classic method, and to a far lesser extend of the robust method. In presence of correlation outliers, the forecast accuracy of the two methods is again comparable. A comparison of Table 1 and 2 suggests that outliers have a severe effect on the forecasts, both directly and indirectly via the selection of the smoothing matrix. To study the last phenomenon in more depth, the next subsection presents a numerical experiment on the data driven selection of the smoothing matrix.

3.2. Selection of the smoothing parameter matrix

The smoothing matrix is selected to minimize the determinant of the sample covariance matrix (in the classic case) or the MCD-estimator (in the robust case) of the one-step-ahead forecast errors in the training period. To visualize the target function in both the classic and the robust case, with and without outliers, we fix the non-diagonal elements of the smoothing matrix to zero and generate 100 time series of length 60 from the same data generating process as before. We apply the classic and the robust multivariate exponential smoothing method, with smoothing matrix

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix},$$

where λ takes values on a grid of the interval $[0, 1]$, and using a startup period of length $m = 10$. For each value of λ , the average of the observed values of

the target functions is plotted in Figure 2.

The vertical dashed line indicates the optimal value of λ according to expression (10). The solid curves are the averaged values of the target function with 95% pointwise confidence bounds (dotted). Both methods have similar target functions. To illustrate the effect of outliers, we add one large additive outlier to the first component of the bivariate time series at time point 35. The resulting target functions of both methods are plotted in Figure 3. The selection of λ in the classic case is clearly biased towards zero, due to only one outlier, whereas the robust parameter selection remains nearly unchanged. This can be explained using equation (10) and the condition $q \Sigma_\varepsilon = \Sigma_\eta$. When outliers are present in the data, the method considers them as extra noise. Hence the signal-to-noise ratio q will decrease. By (10), the diagonal elements of the smoothing matrix will decrease as well, and thus λ will decrease. The proposed robust method does not suffer from this problem.

4. Real Data Example

The robust multivariate exponential smoothing scheme provides a cleaned version \mathbf{y}_t^* of the time series. As such, an affine equivariant data cleaning method for multivariate time series is obtained. In this example, we illustrate how a cleaned series can be used as input for further time series analysis.

Consider the housing data from the book of Diebold (2001) and used in Croux and Joossens (2008). It concerns a bivariate time series of monthly data. The first component contains housing starts and the second component contains housing completions. The data are from January 1968 until June 1996. A plot of the data can be found in Figure 4, indicated by asterisks (*).

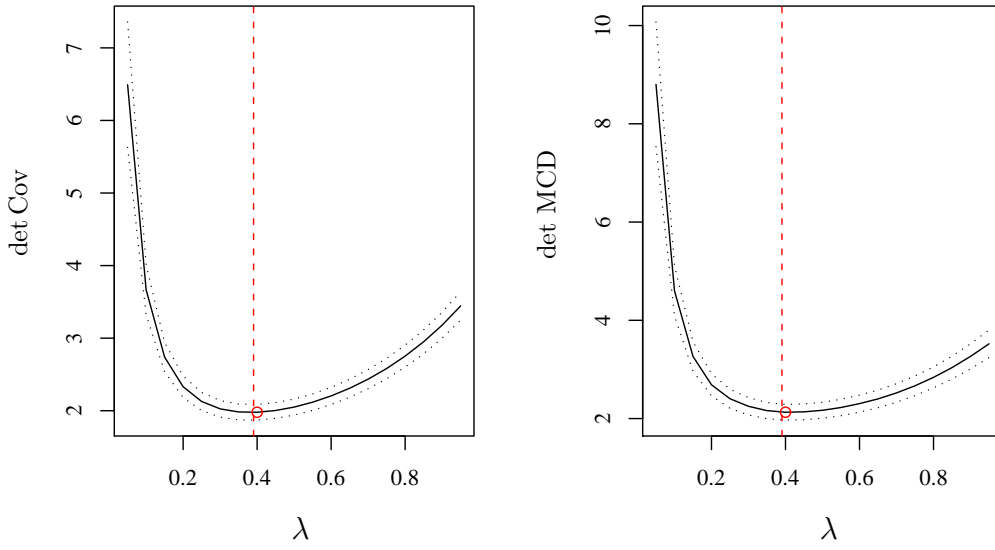


Figure 2: Simulated target function for the classic (left) and the robust method (right), with clean time series. The minimum value is indicated with a circle, the dashed line corresponds to the optimal value of λ .

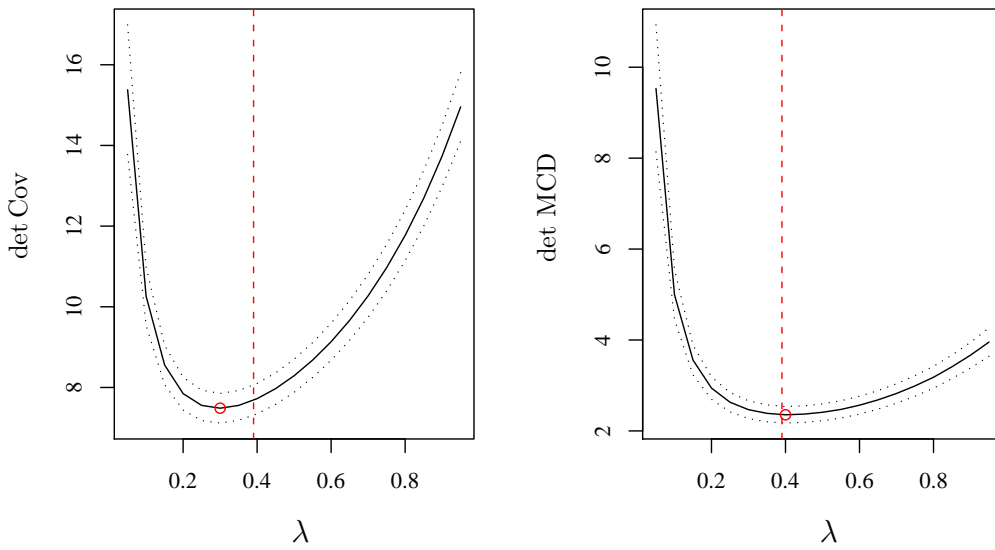


Figure 3: Simulated target function for the classic (left) and the robust method (right), with one large additive outlier. The minimum value is indicated with a circle, the dashed line corresponds to the optimal value of λ .

We immediately notice two large outliers, one near 1971 and another near 1977, both in the first component (housing starts). Moreover, the time series contains correlation outliers, but these are hard to detect in the time series plot. By applying robust exponential smoothing, we know that the results will be stable in presence of such correlation outliers.

We use a startup period of $m = 10$ and the complete series is used as training sample for selecting the smoothing matrix. We get

$$\mathbf{\Lambda} = \begin{pmatrix} 0.68 & 0.04 \\ 0.04 & 0.62 \end{pmatrix}.$$

Figure 4 shows the original series, together with the cleaned version. The cleaning procedure clearly eliminates the large outliers from the original series. Moreover, other smaller outliers, which we could not immediately detect from the plot, are flattened out.

A further analysis of the cleaned series leads to the specification of a Vector Autoregressive (VAR) model for the cleaned series in differences. The lag length selected by the Bayesian Information Criterion equals one. The model is estimated equation by equation by non robust ordinary least squares, since we know that the cleaned series do not contain outliers anymore. We get

$$\Delta \mathbf{y}_t^* = \begin{pmatrix} -3.6 \cdot 10^{-4} \\ 5.8 \cdot 10^{-5} \end{pmatrix} + \begin{pmatrix} -0.28 \cdot 10^{-4} & 0.119 \\ 0.005 & -0.411 \end{pmatrix} \Delta \mathbf{y}_{t-1}^* + \hat{\boldsymbol{\varepsilon}}_t. \quad (11)$$

5. Conclusion

For univariate time series analysis, robust estimation procedures are well developed, see Maronna et al. (2006, Chapter 8) for an overview. To avoid

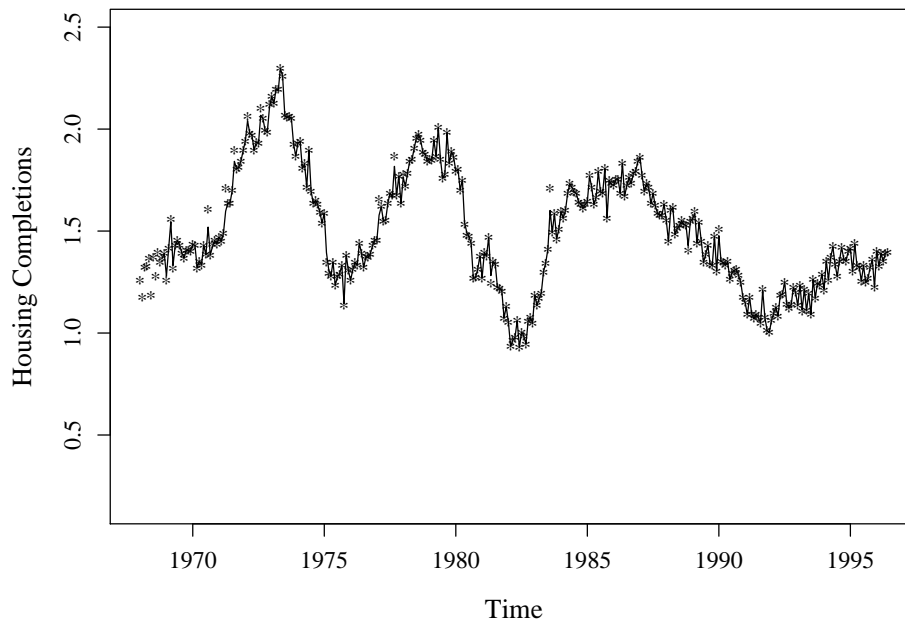
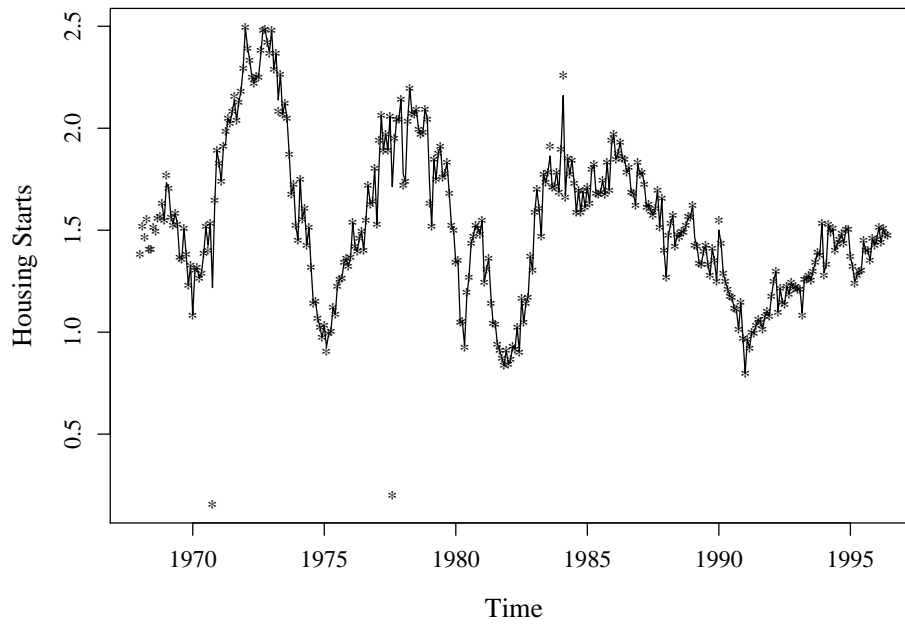


Figure 4: Top: the housing starts (*) with the cleaned series (solid). Bottom: housing completions (*) with cleaned series (solid) (in thousands).

the propagation effect of outliers, a cleaning step is advised, that goes along with the robust estimation procedure (e.g. Muler, Peña, and Yohai (2008)). For resistant analysis of multivariate time series much less work has been done. Estimation of robust VAR models is proposed in Ben et al. (1999) and Croux and Joossens (2008), and a projection-pursuit based outlier detection method by Galeano et al. (2006).

In this paper we propose an affine equivariant robust exponential smoothing approach for multivariate time series. Thanks to its recursive definition, it is applicable for online monitoring. An important byproduct of the method is that a cleaned version of the time series is obtained. Cleaning of time series is of major importance in applications, and several simple cleaning methods were proposed for univariate time series (e.g. Pearson (2005)). Our paper contains one of the first proposals for cleaning of multivariate time series.

For any given value of the smoothing parameter matrix, the procedure is fast to compute and affine equivariant. Finding the optimal $\mathbf{\Lambda}$ in a robust way is computationally more demanding. In this paper a grid-search was applied, working well for bivariate data, but not being applicable in higher dimension. The construction of feasible algorithms for the optimal selection of the smoothing parameter matrix, or proposals for easy-to-use rules of thumb for suboptimal selection of $\mathbf{\Lambda}$ are topics for future research. As we have shown in Section 3, a crucial aspect is that the selection of the smoothing parameters needs to be done in a robust way, see (Boente and Rodriguez, 2008) for a related problem.

Other areas for further research are the robust online monitoring of multivariate scale. In the univariate setting, this problem was already studied

by (Nunkesser et al., 2009) and (Gelper et al., 2009b). The sequence of local scale estimates $\hat{\Sigma}_t$, as defined in (5), could serve as a first proposal in this direction. Finally, extensions of the robust exponential smoothing algorithm to spatial or spatio-temporal processes (LeSage et al., 2009) are also of interest.

References

- Ben, M., Martinez, E., Yohai, V., 1999. Robust estimation in vector arma models. *Journal of Time Series Analysis* 20, 381–399.
- Boente, G., Rodriguez, D., 2008. Robust bandwidth selection in semiparametric partly linear regression models: Monte carlo study and influential analysis. *Computational Statistics and Data Analysis* 52 (5), 2808–2828.
- Cipra, T., 1992. Robust exponential smoothing. *Journal of Forecasting* 11 (1), 57–69.
- Croux, C., Haesbroeck, G., 1999. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis* 71, 161–190.
- Croux, C., Joossens, K., 2008. *Compstat2008*. Physica-Verlag HD, Ch. Robust Estimation of the Vector Autoregressive Model by a Least Trimmed Squares Procedure, pp. 489–501.
- Diebold, F. X., 2001. *Elements of forecasting*, 2nd Edition. South-Western.
- Galeano, P., Peña, D., Tsay, R., 2006. Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association* 101, 654–669.

- Gelper, S., Fried, R., Croux, C., 2009a. Robust forecasting with exponential and Holt-Winters smoothing, to appear in “Journal of forecasting”.
- Gelper, S., Schettlinger, K., Croux, C., Gather, U., 2009b. Robust online scale estimation in time series: A model-free approach. *Journal of Statistical Planning & Inference* 139, 335–349.
- Harvey, A. C., 1986. Analysis and generalisation of a multivariate exponential smoothing model. *Management Science* 32 (3), 374–380.
- Jones, R. H., 1966. Exponential smoothing for multivariate time series. *Journal of the Royal Statistical Society. Series B (Methodological)* 28 (1), 241–251.
- Lanius, V., Gather, U., 2009. Robust online signal extraction from multivariate time series, to appear in “Computational Statistics and Data Analysis”.
- LeSage, J., Banerjee, S., Fischer, M. M., Congdon, P., 2009. Spatial statistics: Methods, models and computation. *Computational Statistics and Data Analysis* 53 (8), 2781 – 2785.
- Maronna, R. A., Martin, R. D., Yohai, V. J., 2006. *Robust statistics: Theory and Methods*. Wiley.
- Masreliez, C. J., Feb 1975. Approximate non-gaussian filtering with linear state and observation relations. *IEEE Transactions on Automatic Control* 20 (1), 107–110.
- Muler, N., Peña, D., Yohai, V. J., 2008. Robust estimation for arma models. *The Annals of Statistics*, to appear.

- Nunkesser, R., Fried, R., Schettlinger, K., Gather, U., 2009. Online analysis of time series by the qn estimator. *Computational Statistics and Data Analysis* 53 (6), 2354–2362.
- Pearson, R. K., 2005. *Mining Imperfect Data*. Siam.
- Pfefferman, D., Allon, J., 1989. Multivariate exponential smoothing: Method and practice. *International Journal of Forecasting* 5 (1), 83–98.
- Reinsel, G. C., November 2003. *Elements of Multivariate Time Series Analysis*, 2nd Edition. Springer Verlag.
- Rousseeuw, P. J., Van Aelst, S., Van Driessen, K., Agullo, J. A., 2004. Robust multivariate regression. *Technometrics* 46 (3), 293–305.
- Rousseeuw, P. J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41 (3), 212–223.