

Brenda Bastiaensen

Robust Trend Regression

Master thesis, defended on June 7, 2012

Thesis advisor: P.H.C. Eilers

Specialisation: Statistical Science for the Life and Behavioural Sciences



Mathematisch Instituut, Universiteit Leiden

Robust trend regression

B.P.C. Bastiaensen ^a, P.H.C. Eilers ^{b*} and O.E. de Noord ^a

^a Shell Global Solutions International B.V., Shell Technology Centre Amsterdam, P.O. Box 38000, 1030 BN Amsterdam, The Netherlands

^b Department of Biostatistics, Erasmus University Medical Centre, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Abstract

Strong fluctuations on a short time scale in chemical process data can lead to poor fit of regression models. Robust Trend Regression (RTR) is introduced for modelling highly fluctuating variables. A robust median smoother is used to separate the data on two timescales. The median smoother avoids fitting extreme peaks such that the smooth is interpretable as the process behaviour on a slow time scale. RTR is able to detect relations which are not found when modelling the original data. A second advantage of RTR is that it outperforms models on the original data for prediction. Data from a chemical plant is used to illustrate the advantages and drawbacks of RTR.

Keywords: Smoothing, Robust Smoother, Median Smoother, Regression

1. Introduction

Chemical process data can fluctuate strongly on short time scales and show large peaks up or down. This is a nuisance when building regression models. The data can be separated into two timescales: a slow trend over time and fast changing peaks. We propose Robust Trend Regression (RTR), a method that focuses on regression modelling of trends obtained with a robust smoother. Figure 1 shows a motivating example. X1 and X2 are fractions of components in the feedstock of a chemical plant. Both the raw data and their trends are shown; the latter look quite similar. We expect that the trends show stronger relationships than the original data does.

Because of the strong peaks, standard smoothing techniques do not work well here. We modified the Whittaker smoother [1] to make it robust, essentially having the local median as its goal, instead of the mean. Robust smoothing eliminates the peaks, but it increases autocorrelation. Also it increases multicollinearity, because explanatory variables are made more similar. In the limit all trends become straight lines resulting in singular regression equations.

We study the influence of the amount of smoothing on variance inflation factors and autocorrelation and try to optimize it by cross-validation.

Time scale decomposition has been explored in econometrics. Ramsey and Lampart [2] use wavelets to study regression models on six different time scales. However, wavelet decompositions are not robust, so their approach cannot be used directly on our data.

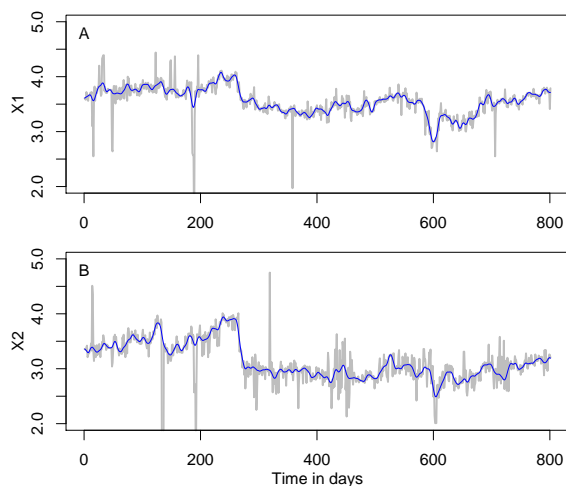


Fig. 1. (a) Time plot of X1 including robust trend, (b) Time plot of X2 including robust trend.

* To whom correspondence should be addressed. E-mail: p.eilers@erasmusmc.nl

The paper is structured as follows. Section 2 presents the statistical tools needed for RTR. Section 3 provides a case study on data from a chemical plant. Conclusions and recommendations for future research are summarized in section 4.

2. Methods

2.1. Median Smoother

Here we present a robust variant of the Whittaker smoother[1]. The Whittaker smoother (WS) minimizes an objective function that has two parts in order to balance two conflicting goals: fit the data by the sum of squared differences (L_2 norm) between the smooth and the original data and minimize roughness of the smooth by the sum of squared second order differences (L_2 norm) of the smooth. This leads to objective function Q_W :

$$Q_W = \sum_{i=1}^n w_i (y_i - z_i)^2 + \lambda \sum_{i=2}^n (\Delta^2 z_i)^2 \quad (1a)$$

In matrices:

$$Q_W = \mathbf{W} \|\mathbf{y} - \mathbf{z}\|^2 + \lambda \|\mathbf{D}_2 \mathbf{z}\|^2 \quad (1b)$$

with \mathbf{y} the original observations, \mathbf{z} the smoothed observations, \mathbf{W} a diagonal matrix with weights and λ the penalty parameter. \mathbf{D}_2 is the second order difference matrix such that $\mathbf{D}_2 \mathbf{z}$ is $\Delta^2 \mathbf{z}$. The larger the value of λ , the smoother the curve will be. \mathbf{W} can be used to give weights to the observations. As a basis, the identity matrix \mathbf{I} can be used as weight vector.

Eilers and de Menezes described adapting the WS by using the sum of absolute values (L_1 norms) in both parts of the objective function for a specific type of data with sudden jumps[3]. For RTR we desire a smooth that fits a gradual changing curve, but is robustness to individual spikes. Based on the idea of using L_1 norms in the objective function the median smoother(MS) is developed, which uses the L_1 norm for the goal of fitting in order to avoid fitting extreme values and the L_2 norm for the goal of minimizing roughness.

$$Q_M = \sum_{i=1}^n w_i |y_i - z_i| + \lambda \sum_{i=2}^n (\Delta^2 z_i)^2 \quad (2a)$$

In matrices:

$$Q_M = \mathbf{1}' \mathbf{W} |\mathbf{y} - \mathbf{z}| + \lambda \|\mathbf{D}_2 \mathbf{z}\|^2 \quad (2b)$$

with $\mathbf{1}$ a vector of ones to reach summation.

The robust smoother is named the median smoother (MS), since the median of a group of values is the value that minimizes the sum of absolute values. Figure 2 shows time plots of X1 for

observations 200 till 400, with in figure 2a a smooth by the WS and figure 2b a smooth by the MS. At observation 255 we see a large down peak in the data that affects the WS while the MS ignores this.

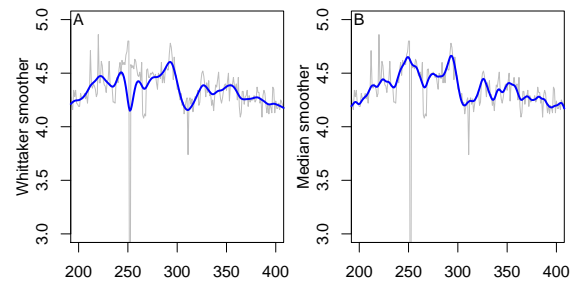


Fig. 2. (a) Timeplot of X1 including Whittaker smooth at observations 200-400 ($\lambda=100$). (b) Time plot of X1 including median smooth at observations 200-400 ($\lambda=100$).

2.1.1. Minimization objective function

Minimization of the objective function for the WS is reached by setting the partial derivatives to zero and solving the linear system.

$$(\mathbf{W} + \lambda \mathbf{D}'\mathbf{D}) \hat{\mathbf{z}} = \mathbf{W} \mathbf{y} \quad (3)$$

$\hat{\mathbf{z}}$ is the solution of the WS. Minimization of the objective function for the MS is less straightforward, since the new objective function does not have the desirable properties of the quadratic sum of squares. The MS uses the objective function of the WS for optimization, with specific weights such that the objective function is equal to Q_M .

$$|y_i - z_i| = v_i |y_i - z_i|^2 \quad \text{with} \quad v_i = \frac{1}{|y_i - z_i|} \quad (4)$$

The matrix \mathbf{V} with diagonal elements $\frac{1}{|y_i - z_i|}$ is the weight matrix for $|\mathbf{y} - \mathbf{z}|^2$. The algorithm optimizes:

$$Q_M = \mathbf{V} \|\mathbf{y} - \mathbf{z}\|^2 + \lambda \|\mathbf{D}_2 \mathbf{z}\|^2 \quad (5)$$

Optimization is done iteratively by updating the weights until $v_i = \frac{1}{|y_i - z_i|}$. The initial solution is set to the solution of the WS and the initial weights are $v_i = \frac{1}{\sqrt{(y_i - z_i)^2 + \beta^2}}$. The β is set to a fixed small value (1e-4) and is used to avoid computational problems in case \mathbf{z} is equal to \mathbf{y} . In this way v_i is an approximation of $\frac{1}{|y_i - z_i|}$. At each step in the algorithm the weighted objective function is minimized. The solutions and the weights are updated until the new solution $\hat{\mathbf{z}}$ is equal to the previous solution, where $\hat{\mathbf{z}}$ is the solution of the MS.

It is advised to normalize \mathbf{y} by mean centring and scaling by the standard deviation before using the

MS in order to make the result invariant to the scale of y . Normalizing was not needed for the WS, since scaling of y by c scales each of the two parts of the objective function by c^2 , such that the balance between the two components in the objective function is the same. This is not the case for the MS: the first part of the objective function of the MS is scaled by c while the second part is scaled by c^2 . Results of the MS can be transformed back on to the original scale.

The MS is implemented by less than 25 lines of programming code in R or Matlab (see appendix A) and needs less than a second to calculate the smooth of 1000 observations. To ensure fast calculation sparse matrices are used (R package {spam}). Solving a linear system has a computational time proportional to n^3 , while for a banded system time is proportional to n .

An advantage of smoothing the variables is the robustness with regard to missing data and outliers. As for the WS, this can be done by the choice of the weights [1]. Weights should be set to zero for the observations not to be considered and to one for the others. The smooth is based only on the observations with weight one, the rest gets automatically filled by interpolation. The MS assumes that the observations have been measured with equal distances in time. Missing values due to a low sampling frequency can be filled as missing measurements by the MS.

When a variable has very sudden jumps, smoothing is not supposed to round these. Such jumps can be caused by specific events like catalyst changes or shutdowns. When the time points of the jumps are exactly known, the MS can avoid rounding such extreme jumps by adding a weight vector also in the second part of the objective function.

$$Q_M = \sum_{i=1}^n w_i |y_i - z_i| + \lambda u_i \sum_{i=2}^n (\Delta^2 z_i)^2 \quad (6a)$$

In matrices:

$$Q_M = \mathbf{1}' \mathbf{W} |\mathbf{y} - \mathbf{z}| + \lambda \mathbf{U} \|\mathbf{D}_2 \mathbf{z}\|^2 \quad (6b)$$

\mathbf{U} is a diagonal matrix with weights u_i on the diagonal, with all ones except weights $u_j = u_{j-1} = 0$ for a jump at time point j .

2.2. Robust Trend Regression

Any regression technique could be applied in Robust Trend Regression.

2.2.1. Selection of the smoothing parameter

Cross-validation is used to select the optimal smoothing parameter for which the output variable is best predicted. Bergmeir and Benitez suggest to use

blocked cross-validation in time series to overcome time dependency problems [4], this method is used with 20 blocks of 20 consecutive observations. The root mean squared prediction error (RMSE) is used as cross-validation statistic to test the predictive performance on the original data. The original data is used, since predicting the trend data for different smoothing factors is misleading; an over smoothed output variable can be perfectly predicted. The value of lambda with minimum RMSE is selected.

The main drawbacks of regression on smoothed variables comes with the risk of over-smoothing: there is stronger correlation between the input variables, variables have increased autocorrelation and there might be no sensible interpretation of the smooth.

Multicollinearity can be quantified by Variance Inflation Factors (VIF) [5]. The variance of the estimated coefficients can be written as:

$$V(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)^2} \frac{1}{1 - R_j^2} \quad (7)$$

with R_j^2 the R squared from a regression of variable j on the other variables. Marquardt describes the term $\frac{1}{1 - R_j^2}$ as the VIF [5], which reflects to what extent the variance of the estimated β_j is inflated by correlation with other regressors. No strict cut-off for the VIF can be found in literature. Marquardt suggests the maximum VIF to indicate multicollinearity should be larger than 1, but certainly not larger than 10. Snee translates this to using a cut-off of 5 or 10 as a general rule to indicate multicollinearity [6]. A maximum VIF of 5 means R_j^2 cannot be higher than 0.8, so variables cannot be explained for more than 80% by other explanatory variables. A cut-off of 10 puts a maximum of 0.9 to R_j^2 . For interpreting individual coefficients of the RTR model this research uses a VIF between 0 and 5 to indicate weak multicollinearity, between 5 and 10 to indicate moderate multicollinearity and larger than 10 to indicate strong multicollinearity. Multicollinearity only affects the interpretation of individual coefficients; the total model fit is not affected.

Most regression techniques assume independency between observations, which is not true for time series. A plot of the autocorrelation function (ACF) is used to detect non-randomness in the residuals. Levine introduced a heteroscedasticity and autocorrelation consistent (HAC) covariance matrix for models with time dependent residuals [7]. The Lumley and Heagerty estimate of the HAC is used in this research (R package {sandwich}) [8,9].

VIF's can be used to get an impression of the extent of oversmoothing. When the variables are

oversmoothed to nearly straight lines, VIF's will be extremely large. When selecting the smoothing parameters, a restriction can be put on the maximum allowed VIF: a combination of smoothing parameters for which the RTR model has a VIF over a chosen cut-off should not be considered. It is recommended to choose this cut-off larger than 10 for detecting oversmoothing, like 20, 50 or 100, depending on collinearity in the original data. The VIF's can already be quite large with no smoothing, so a VIF of 10 not necessarily indicates oversmoothing.

3. Results and Discussion

3.1. Data

This section shows a case study of Robust Trend Regression on real data over a period of 400 days. There are 22 input variables (X1-X22) and three output variables (Y1-Y3). Due to confidentiality, the nature of these variables cannot be revealed. To show the efficacy of RTR, this research uses multiple linear regression (MLR). In the remainder of the paper, RTR refers to applying MLR on smoothed data and MLR refers to applying MLR on the original data.

3.2 Robust Trend Regression

3.2.1. Selection of the smoothing parameter

For the non smoothed data the highest Variance Inflation Factor (VIF) in the model on Y1 is 5.16. A cut-off of 20 is chosen for the maximal allowed VIF to avoid oversmoothing. To avoid large computational time, the same smoothing parameter is used for each variable. The highest VIF of all variables for the Y1 RTR model is plotted in figure 3 for a range of smoothing parameters between 0 and 1000 (0, 0.001, 0.002, 0.005, 0.01, ..., 200, 500, 1000) showing a clear increase for larger lambda (note that an RTR model with a lambda of 0 is an MLR model). From lambda of 200 onwards the VIF's get larger than 20, which are excluded in the cross-validation.

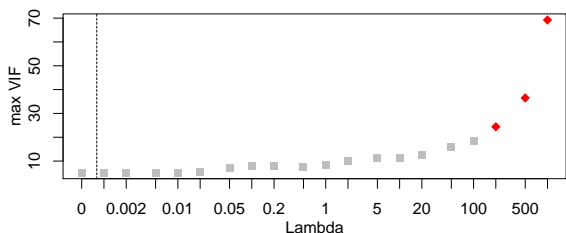


Fig. 3. Highest VIF of RTR Y1 model for several lambdas.

For each output variable cross-validation is applied on the range of lambdas up to and including 100. For each lambda the RMSE is plotted for Y1, Y2 and Y3 in figure 4. For Y1 and Y3 a lambda of 5 is

selected and for Y2 a lambda of 20. For each output variable RTR with optimal chosen smoothing parameter has a better predictive performance than MLR. Notice that for some values of lambda the predictive performance of RTR may be worse. The output variables including their smooth are plotted in figure 5.

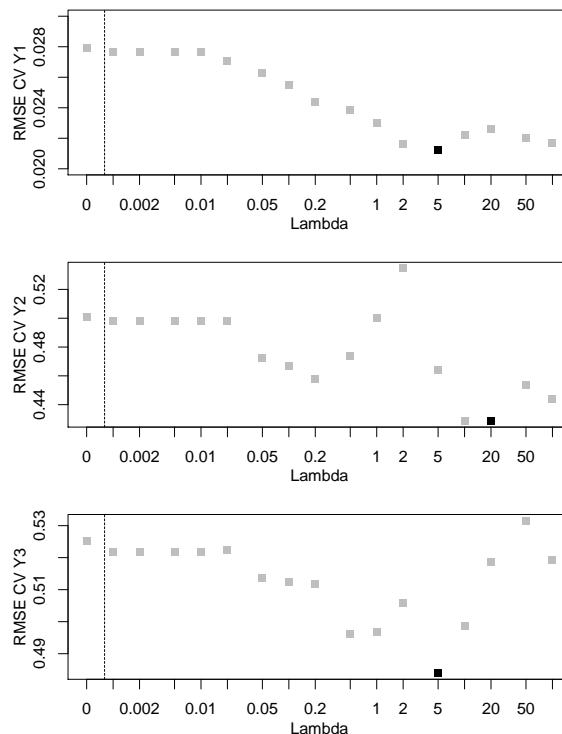


Fig. 4. Cross-validation RMSE for lambda between 0 and 100 for each output variable. Black indicates the minimal RMSE (a) Y1, (b) Y2, (c) Y3.

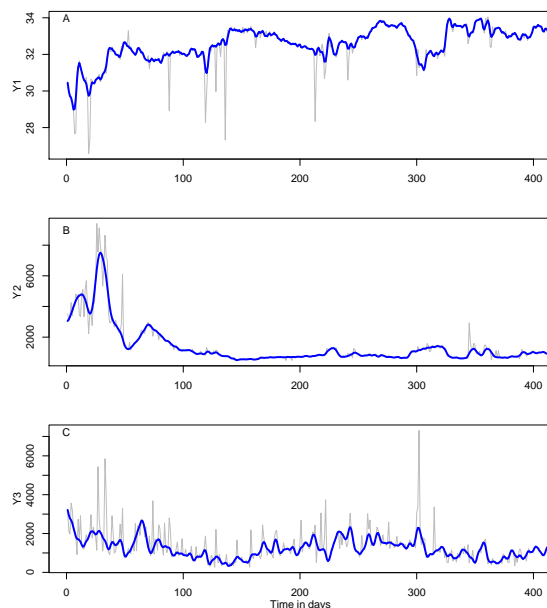


Fig. 5. Time plots of output variables including smoothed trend: (a) Y1 ($\lambda=5$), (b) Y2($\lambda=20$), (c) Y3 ($\lambda=5$).

3.2.1. Model Diagnostics

To see the effect of smoothing on the models the following checks are performed: 1. Residuals against fitted values, to check linearity and heteroscedasticity: For all three output variables there is no indication of nonlinearity. For Y1 there is an indication for heteroscedasticity in the MLR model, which is solved in the RTR model. 2. QQ plot of residuals to check residual normality: For all output variables the QQ plot of the MLR model shows a shaped curve. The RTR models have no indication for non-normal residuals. 3. ACF plot of residuals to check autocorrelation: For the MLR models the ACF plots show strong time dependency for Y1 and Y2 and moderate for Y3. Time dependency is very strong for the residuals of the RTR models for all output variables. We can conclude that for each output variable, smoothing the data improved heteroscedasticity and normality of the residuals, but smoothing brought more time dependency in the errors. Figure 6 shows the diagnostic plots for the MLR and RTR model on Y3 (plots for Y1, Y2 are not presented since they are very similar). Figure 7 shows time plots of the residuals of the RTR and MLR model on Y3 to provide extra insight in the time dependency. For both models the errors are not randomly fluctuating around the zero line.

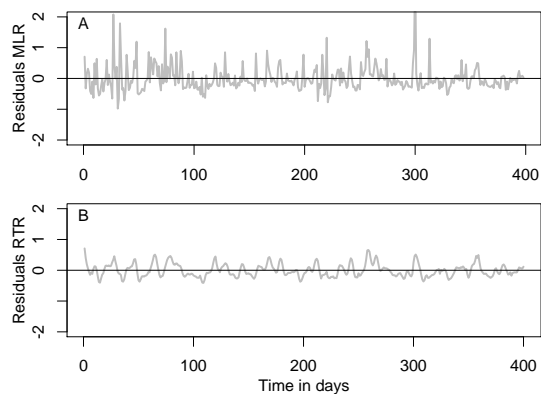


Fig. 7. Time plot residuals Y3 models (a) MLR (b) RTR.

To get insight in multicollinearity VIF's are given for Y3 models in table 1 (VIF's are nearly equal for Y1 and Y2). All VIF's are larger for smoothed data. In the original data one VIF is larger than 5 and no VIF's are larger than 10, indicating moderate multicollinearity for X21. In the RTR model there is one VIF larger than 10 and seven VIF's between 5 and 10, showing moderate multicollinearity for X2, X9, X10, X14, X16, X19 and X21 and strong multicollinearity for X4.

Table 1. Variance Inflation Factors for Y3 RTR and MLR models.

| Variable | VIF Original Data | VIF Smoothed Data |
|----------|-------------------|-------------------|
| X1 | 1.67 | 2.16 |
| X2 | 4.87 | 9.86 |
| X3 | 1.18 | 2.58 |
| X4 | 4.44 | 11.19 |
| X5 | 1.35 | 3.69 |
| X6 | 1.29 | 3.11 |
| X7 | 2.71 | 3.73 |
| X8 | 1.66 | 3.29 |
| X9 | 3.53 | 7.43 |
| X10 | 3.15 | 8.38 |
| X11 | 1.41 | 2.28 |
| X12 | 1.60 | 2.24 |
| X13 | 2.05 | 2.89 |
| X14 | 3.14 | 5.20 |
| X15 | 2.46 | 3.54 |
| X16 | 4.13 | 6.08 |
| X17 | 2.56 | 3.86 |
| X18 | 1.69 | 3.24 |
| X19 | 4.84 | 6.74 |
| X20 | 2.77 | 4.21 |
| X21 | 5.16 | 7.22 |
| X22 | 2.88 | 3.50 |

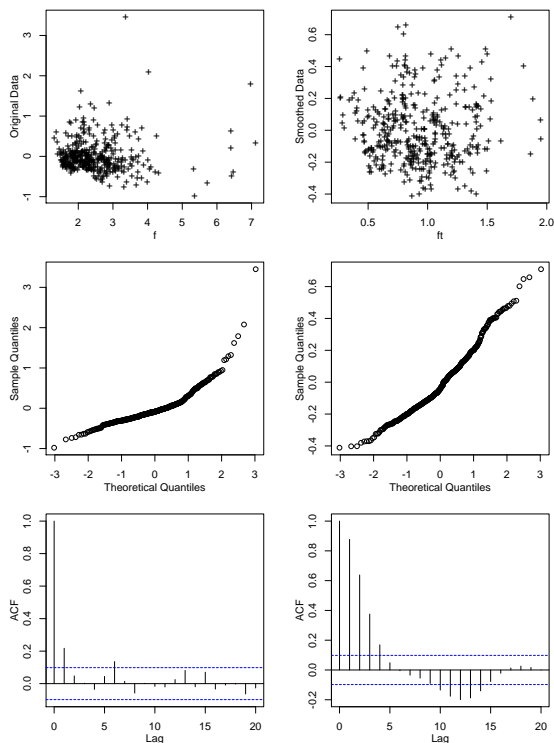


Fig. 6 Diagnostic plots of residuals Y3 MLR (left) and RTR(right) models: Row 1. Residuals against fitted values, Row 2. QQ plot residuals, Row 3. ACF plot residuals.

3.2.2. Model fit

To overcome nonnormality of the residuals the log of the output variables is used for all models. For original data extreme outliers are removed, the smoothed data contains no outliers. The data is scaled

by its standard deviation to compare coefficients between variables.

Table 2 shows the fit of the MLR and RTR model together with the fit of an MLR model on the peak data (the difference between the trend and the original data). MLR models on the peak data are not able to explain a reasonable part of the variance in the peaks for Y2 and Y3. Since the peak variation is also captured in the original data, a poor fit on the peaks explains why we perform better in modelling the trend than in modelling the original data.

Table 2. Model fit of the models for the three output variables for Original, Smoothed and Peak Data.

| Data | R^2 | R^2 | R^2 |
|----------------|-------|-------|-------|
| | Y1 | Y2 | Y3 |
| Original (MLR) | 0.67 | 0.79 | 0.43 |
| Smoothed (RTR) | 0.91 | 0.91 | 0.69 |
| Peak | 0.68 | 0.11 | 0.11 |

Figure 8 shows the data including the fitted values by the RTR and MLR models. The fitted curve by RTR is way smoother than the curve fitted by MLR. MLR tries to model the peaks in which it only succeeds at some timepoints. Figure 9 shows observed against fitted plots for each model. From the scatterplots we see two effects of RTR: the outlying observations in the MLR models are not seen in the scatterplots of the RTR models and the main cloud of observations is smaller for RTR than for MLR, indicating a stronger linear relation in smoothed data.

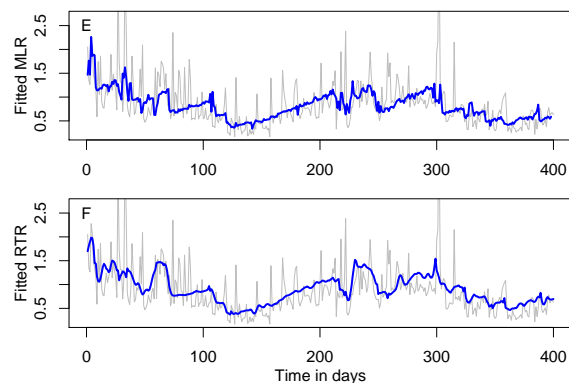
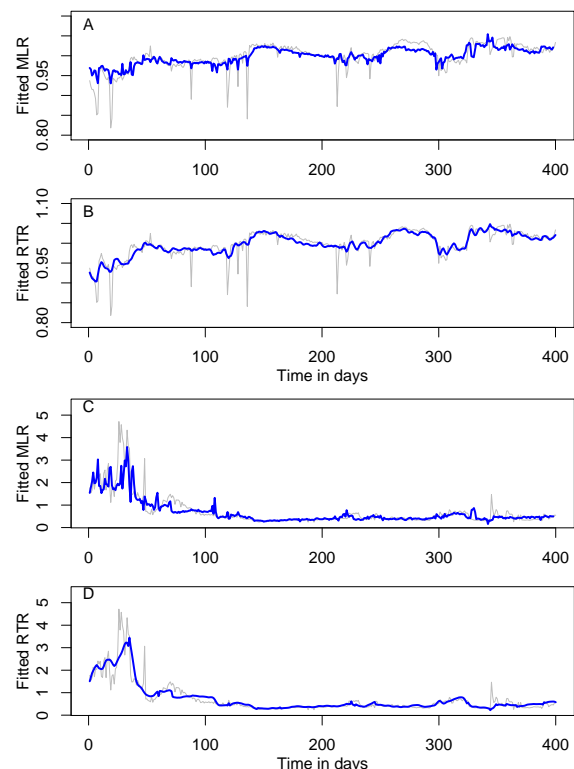


Fig. 8. Time plots of original data including fitted Trend by: (a) MLR Y1 (b) RTR Y1 (c) MLR Y2 (d) RTR Y2 (e) MLR Y3 (f) RTR Y3.

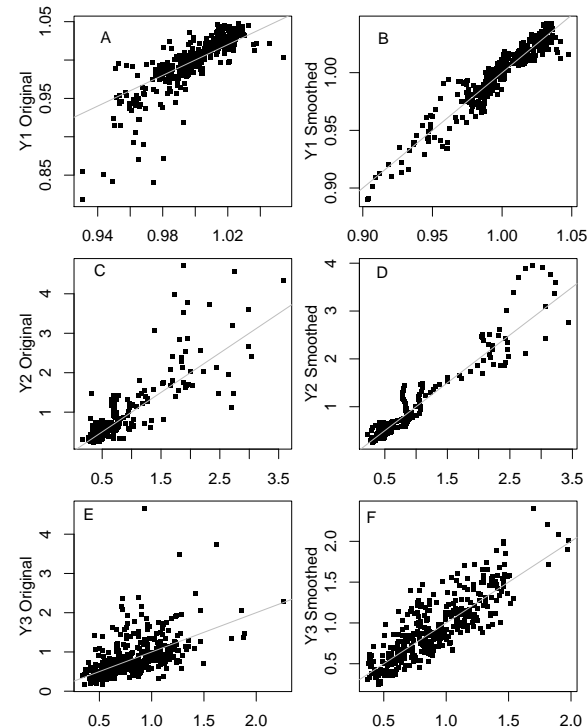


Fig. 9. Observed against fitted plots. Left MLR, right RTR. Top row Y1, middle row Y2, bottom row Y3.

3.2.3 Coefficients

To get insight in the effect of smoothing on the coefficient estimates figure 10 shows the estimated coefficients of the input variables for each value of lambda (black coefficients are significantly different from zero; only variables significantly different from zero for at least one value of lambda are plotted). There are variables of which the coefficients and significance drastically change when lambda gets larger, confirming RTR can indicate different relations between the variables than MLR.

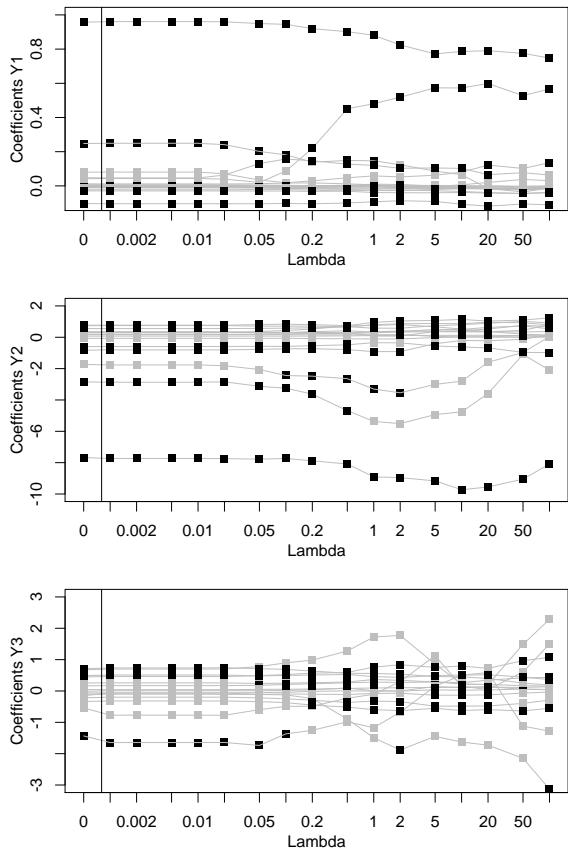


Fig. 10. Estimated coefficients against lambda (a) Y1, (b) Y2, (c) Y3. Black indicates significantly different from zero.

In figure 11 standardized coefficients of MLR and RTR with optimal lambda are provided, including error bars of two times the HAC estimated standard deviation, showing larger standard errors in RTR models due to increased autocorrelation. In the RTR models there was an indication for multicollinearity, but also for variables not suffering from multicollinearity (X3, X5, X6, X12 and X22) RTR has different coefficients and significance.

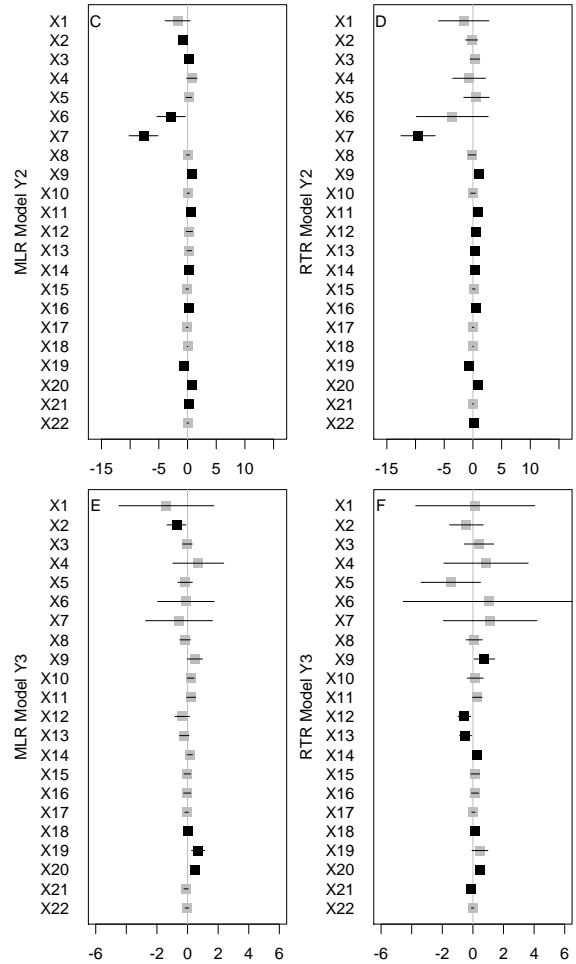


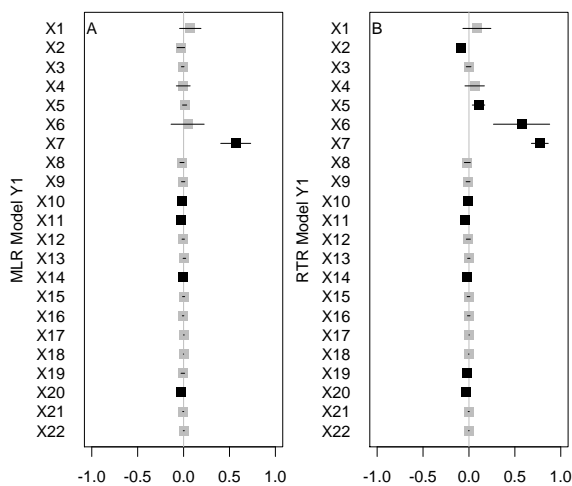
Fig. 11. Plots of the estimated coefficient including error bars of two times the HAC estimated standard error. Coefficient with significant p-values are black. Left MLR, right RTR. Top row Y1, middle row Y2, bottom row Y3.

Since this cannot be due to multicollinearity, it suggests different drivers for the slow and fast variation.

4. Conclusion

Extreme spikes and fast fluctuations in time series data lead to poor fit in regression models, which does not necessarily mean that no relations exist. Separating the data on two time scales has shown that each of these time scales can have different drivers of the variance. The robust median smoother is the ideal smoother to determine the process behaviour on a slow time scale, since it is not influenced by individual peaks in the data.

It forms the basis of Robust Trend Regression, which applies regression on the smoothed input and output variables. It outperforms regression techniques on original data in identifying drivers for slow process variations, which can be obscured when



original data is modelled. Prediction accuracy of RTR models was also found to be better than for MLR models based on original data.

Modelling smoothed data has both some positive and negative effects on model diagnostics. The smooth is less influenced by outliers and by missing data. Heteroscedasticity and normality of the residuals are improved. The main drawbacks of RTR are related to the risk of oversmoothing. Firstly, a sensible interpretation of the curves might be difficult. Secondly, when applying smoothing autocorrelation increases. It is recommended to use autocorrelation consistent covariance matrix estimates in a RTR model to ensure reliable test statistics on individual coefficients. Thirdly, also multicollinearity between variables gets stronger. Variance Inflation Factors (VIF's) can be used to test for multicollinearity before interpreting individual results. VIF's can also give insight in the degree of oversmoothing, since VIF's will be very large for a high degree of smoothing. Future research could investigate the efficacy of RTR when using regression techniques that are more robust to multicollinearity.

Blocked cross-validation was used to find the optimal degree of smoothing, where the same smoothing parameter for each variable was used for computational reasons. Future research could focus on finding an efficient way to select the optimal degree of smoothing for each variable separately.

By focusing on modelling the Trend, no conclusions can be drawn on the peaks in the data. This needs to be studied separately.

References

- [1] P.H.C. Eilers, A Perfect Smoother, *Anal. Chem.* 75 (2003) 3299-3304
- [2] J.B. Ramsey, C. Lampart, Decomposition of economic relationships by timescale using wavelets, *Macroecon. Dyn.* 2 (1998) 49-71
- [3] P.H.C. Eilers, R.X. de Menezes, Quantile smoothing of array CGH data, *Bioinformat.* 21(7) (2005) 1146-1153
- [4] C. Bergmeir, J. Benitez, On the use of cross-validation for time series predictor evaluation, *Inform. Sc.* 191 (2012) 192-213
- [5] D.W. Marquardt, Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, *Technom.* 12 (3) (1970) 591-612
- [6] R.D. Snee, Validation of Regression Models: Methods and Examples, *Technom.* 19 (4) (1977) 415-428
- [7] D. Levine, A Remark on Serial Correlation in Maximum Likelihood, *J. Economet.* 23 (1983) 337-342

[8] T. Lumley, P. Heagerty, Weighted Empirical Adaptive Variance Estimators for Correlated Data Regression, *J. Economet.* 29 (1999) 305-325

[9] A. Zeileis, Econometric Computing with HC and HAC Covariance Matrix Estimators, *J. Stat. Soft* 11 (10) (2004)

Appendix A

R code median smoother:

```
library(spam)
m=length(y)
s=as.double(scale(y))
E = diag.spam(m)
D = diff(E, diff = 2)
P = lambda * t(D) %*% D
      z1 = solve(E + P, s)
# z1 is the result of the Whittaker Smoother
z2 = z1
beta = 1e-4
for (it in 1:100) {
  r = s - z2
  v = 1 / sqrt(r ^ 2 + beta ^ 2)
  V = diag.spam(v)
  znew = solve(V + P, v * s)
  dz = max(abs(znew - z2))
  cat(it, dz, '\n')
  if (dz < 1e-4) break
  z2 = as.vector(znew)
}
z<-(z2*sd(y))+mean(y)
# z is the result of the Median Smoother
```

Matlab code median smoother:

```
m = length(y);
sdy = std(y);
s = y / sdy;
E = speye(m);
D = diff(E, 2);
lambda = 100;
P = lambda * D' * D;
z1 = (E + P) \ s;
# z1 is the result of the Whittaker smoother
z2 = z1;
beta = 1e-4;
for it = 1:100
  r = s - z2;
  v = 1 ./ sqrt(r.^2 + beta ^ 2);
  V = spdiags(v, 0, m, m);
  znew = (V + P) \ (v .* s);
  dz = max(abs(znew - z2));
  disp(dz)
  z2 = znew;
  if dz < 1e-4
    break
  end
end
z = (z2*std(y))+mean(y)
# z is the result of the Median smoother
```