Efficient and robust scale estimation for trended time series

Derya Caliskan, Christophe Croux, Sarah Gelper

Efficient and robust scale estimation for trended time series

Derya Caliskan[a,b], Christophe Croux[*,a], Sarah Gelper[c]

[a] *LSTAT & Faculty of Business and Economics, K.U.Leuven, Belgium*
[b] *Department of Statistics, Hacettepe University, Turkey*
[c] *Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands*

Abstract

This paper presents a new method for robust online variability extraction in time series. The proposed estimator is simultaneously highly robust and efficient. We derive its breakdown point, influence function, and asymptotic variance and study the finite sample properties in a simulation study.

Key words: Influence function, Robustness, Scale estimation, Time Series

1. Introduction

In the recent literature, new procedures have been proposed for robust scale estimation in a time series context, see for example Nunkesser et al. (2009) and Gelper et al. (2009). These procedures are designed to monitor the variability of noisy and trended time series. A drawback of the existing methods is that they lose efficiency with respect to the standard non-robust estimator. This paper proposes a new approach which combines the quantile based estimator of Gelper et al. (2009) and the efficient τ-scale estimator of

*Christophe Croux, Faculty of Business and Economics, Naamsestraat 69, 3000 Leuven, Belgium

Yohai and Zamar (1988). As a result, we obtain a new estimator which is highly robust and at the same time attains high efficiency.

In several application fields data are automatically collected at a high frequency and need to be monitored instantaneously. Assume that we collect observations over time

$$y_t = \mu_t + \sigma_t \epsilon_t \qquad \text{for} \quad t = 1, \ldots, T \tag{1}$$

where ϵ_t is an i.i.d. sequence coming from a symmetric distribution F_0 with mean zero and variance 1. We estimate σ_t using a moving window approach. At each time t, a scale estimator $\hat{\sigma}_t$ is computed from a window of width $n < T$, containing the observations $y_{t-n+1}, ..., y_t$. The variability is allowed to change slowly over time. More precisely, the variability is only assumed to be constant in a local window which can be fairly short, for example of only 20 observations.

For the ease of notation, we drop the time index t and the observations within a fixed window are denoted by $y_1, \ldots, y_n$. We consider vertical heights of the triangles formed by triplets of successive data points. For three successive observations y_i, y_{i+1} and y_{i+2}, the height of the triangle formed by these observations is given by

$$h_i = |y_{i+1} - \frac{y_i + y_{i+2}}{2}|$$

for $i = 1, ..., n-2$. Rousseeuw and Hubert (1996) propose robust scale estimators based on these heights in the context of nonparametric regression, and Gelper et al. (2009) adapted them to the time series context. These estimators are invariant if a trend line $a + bt$ is added to the data. Moreover, they do not require to model and estimate the signal μ_t, and were

shown to be applicable for signals containing jumps, trends, trend changes and non-linearities.

The estimator Gelper et al. (2009) advocate is the $\alpha-$quantile of the heights of adjacent triangles

$$Q_{adj}^{\alpha}(y_1, ..., y_n) = c_q h_{(\lfloor \alpha(n-2) \rfloor)}, \tag{2}$$

which is the $\lfloor \alpha(n-2) \rfloor$-th value in the sequence of ordered heights, and where c_q is a consistency factor. For $\alpha = 0.25$ the highest value for the breakdown point is attained, and we denote $Q_{adj} \equiv Q_{adj}^{0.25}$. While this estimator has good properties, it has a Gaussian efficiency of only 25%. Gelper et al. (2009) proposed to increase the value of α to get a better efficiency, at the price of a lower breakdown point. The procedure proposed in this paper, the τ-adjacent estimator, maintains the high breakdown point of the Q_{adj}, but can attain an arbitrarily high efficiency.

This robust scale estimator τ_{adj} is defined as

$$\tau_{adj}(y_1, \ldots, y_n) = c_\tau \left[S_0^2 \frac{1}{n-2} \sum_{i=1}^{n-2} \rho(\frac{h_i}{S_0}) \right]^{1/2} \tag{3}$$

with initial scale estimator $S_0 = Q_{adj}(y_1, \ldots, y_n)$. The loss function ρ should be bounded, symmetric, $\rho(0) = 0$, and non-decreasing on the positive numbers. Its derivative $\psi = \rho'$ should exist and be positive at zero. In this paper we take Tukey's bisquare ρ function defined as

$$\rho(x) = \begin{cases} \frac{x^2}{2}(1 - \frac{x^2}{k^2} + \frac{x^4}{3k^4}), & \text{if } |x| \leq k \\ \frac{k^2}{6} & \text{if } |x| > k. \end{cases}$$

The value of the tuning constant k depends on the desired level of efficiency. As will be seen in Section 2, $k = 5.48$ yields a 95% asymptotic relative efficiency at the Gaussian model.

The performance of the adjacent τ-estimator is compared with the square root of the mean of squared adjacent heights,

$$MS_{adj}(y_1, \ldots, y_n) = c_s \sqrt{\frac{1}{\lfloor n-2 \rfloor} \sum_{i=1}^{\lfloor n-2 \rfloor} h_i^2} \tag{4}$$

with c_s again a consistency factor. The MS_{adj} is not robust, but it is a standard proposal for scale estimation in nonparametric regression (see Gasser et al. (1986)). It is not difficult to check that MS_{adj} equals the τ_{adj} for $k \to \infty$.

Section 2 discusses statistical properties of the τ_{adj} estimator. We show that the estimator has a breakdown point of 25%, a bounded influence function, and we compute its asymptotic variance. The good behavior of the estimator is confirmed in Section 3 by a simulation study. Section 4 summarizes the results.

2. Statistical Properties

In the section, we first derive the expression for the consistency factor c_τ for the newly proposed estimator. Expressions for the constants c_q in (2) and c_s in (4) are given in Gelper et al. (2009). We then derive the breakdown point and the influence function of the τ_{adj} estimator. The breakdown point measures the robustness under larger amounts of outliers, while the influence function measures the sensitivity of the estimator with respect to small amounts of contamination. The last subsection computes the asymptotic variance of the estimator.

2.1. Fisher Consistency

For all theoretical derivations, we assume local linearity and a constant scale within the considered time window. Hence

$$y_i = a + bi + \sigma\epsilon_i \tag{5}$$

for $i = 1, \ldots, n$ where $\epsilon_i \sim F_0$. For an appropriately chosen window width Let F be the distribution of the data and denote H_F the distribution of the corresponding triangle heights. The functional form of the $\tau_{adj}-$scale estimator (3) corresponds to

$$\tau_{adj}(F) = c_\tau \left[S_0^2(F) E_{H_F} \rho(\frac{h}{S_0(F)}) \right]^{1/2} \tag{6}$$

where $S_0(F) \equiv Q_{adj}(F) = c_q H_F^{-1}(0.25)$. The proposed scale estimator is location invariant, invariant when a trend is added to the data, and scale equivariant. Assume that the model (5) holds, and take $a = b = 0$ without loss of generality. Since the constant c_q is such that Q_{adj} is Fisher consistent, we have $S_0(F) = \sigma$. In order to achieve Fisher consistency, that is $\tau_{adj}(F) = \sigma$, we need to take

$$c_\tau = \frac{1}{\sqrt{E_{H_{F_0}} \rho(\frac{h}{S_0(F_0)})}} = \frac{1}{\sqrt{E_{H_{F_0}} \rho(h)}}. \tag{7}$$

For F_0 a standard normal $N(0,1)$ distribution, it is not difficult to verify that $E_{H_{F_0}} \rho(h) = E_{F_0} \rho(Z\sqrt{3/2})$, allowing for immediate calculation of (7). As such, we obtain for $k = 5.48$ that $c_\tau = 1.24$. To make the estimator practically unbiased at finite samples, we propose to replace c_τ by a finite sample version c_τ^n. By Monte Carlo simulation, following the approach outlined in Gelper et al. (2009), and for $k = 5.48$, we obtain

$$c_\tau^n \approx c_\tau \frac{n}{n - 1.34}\,.$$

2.2. Breakdown point

The breakdown point of a scale estimator is the minimal amount of contamination such that the estimated scale becomes either infinite (explosion) or zero (implosion). Let $\mathbf{y}_n = \{y_1, \ldots, y_n\}$ be a sample of size n. Let S be a scale estimator. Further, let $\mathbf{y}_n^m$ be a sample obtained from $\mathbf{y}_n$ but with a proportion of m/n observations altered to arbitrary values ($m \in \{1, \ldots, n\}$). The *finite sample breakdown point* of S at the sample $\mathbf{y}_n$ is defined as

$$\varepsilon^*(S, \mathbf{y}_n) = \min \frac{1}{n} \left\{ m \in \{1,\, 2,\, \ldots,\, n\} : \sup_{\mathbf{y}_n^m} |\log(\frac{S(\mathbf{y}_n^m)}{S(\mathbf{y}_n)})| = \infty \right\},$$

Suppose that $\mathbf{y}_n$ is in general position, meaning that no three observations (i, y_i) lie on the same line for $1 \leq i \leq n$. Gelper et al. (2009) have shown that, for the initial estimator $S_0 = Q_{adj}$,

$$\varepsilon^*(S_0, \mathbf{y}_n) = \frac{1}{n} \min \left\{ \left\lceil \frac{n - 1 - \lfloor \alpha(n-2) \rfloor}{3} \right\rceil, \lfloor \alpha(n-2) \rfloor \right\}. \tag{8}$$

The highest possible value for the breakdown point is attained for

$$\alpha = \frac{n+1}{4(n-2)} \approx 0.25. \tag{9}$$

The finite sample breakdown point tends to the *asymptotic breakdown point* of 25%. In the following proposition we state that the breakdown point of the τ_{adj} estimator is the same as that of the initial estimator S_0.

Proposition 1. *Let $\mathbf{y}_n$ be a sample in general position. For the τ_{adj} estimator with a bounded loss function ρ we have that*

$$\varepsilon^*(\tau_{adj}, \mathbf{y}_n) = \varepsilon^*(S_0, \mathbf{y}_n).$$

Proof: From definition (3) of the τ_{adj} estimator it is readily seen that the estimator tends to infinity if and only if the initial scale estimator S_0 does, since ρ is bounded. Implosion of the τ_{adj} estimator occurs if either (i) S_0 implodes to zero (ii) all heights h_i are equal to zero (iii) S_0 is so large that $\rho(h_i/S_0)$ is arbitrarily small for all $1 \leq i \leq n-2$. Since we assumed that the sample is in general position, (ii) cannot occur. Furthermore, by definition of S_0, about 75% of the heights is larger than S_0, and $\rho(1) > 0$, hence also (iii) is excluded. We conclude that the τ_{adj} implodes if and only of the initial scale estimator implodes.

2.3. Influence function

The influence function of the functional S at the distribution F measures the effect on S of adding a small mass at the point w, standardized by the mass of the contamination. If we denote the point mass at w by Δ_w and consider the contaminated distribution $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_w$ then the influence function is given by

$$IF(w; S, F) = \lim_{\varepsilon \to 0} \left[\frac{S(F_\varepsilon) - S(F)}{\varepsilon} \right]. \tag{10}$$

In the Appendix we derive an explicit expression for the influence function of τ_{adj}, assuming model (5) holds with $F_0 = N(0,1)$. Figure 1 pictures this influence function at this normal model with $\sigma = 1$. We see that the τ_{adj} has the desirable property of a bounded IF, hence it is B-robust. Its influence function is smooth and has a quadratic shape close to the center of the distribution.

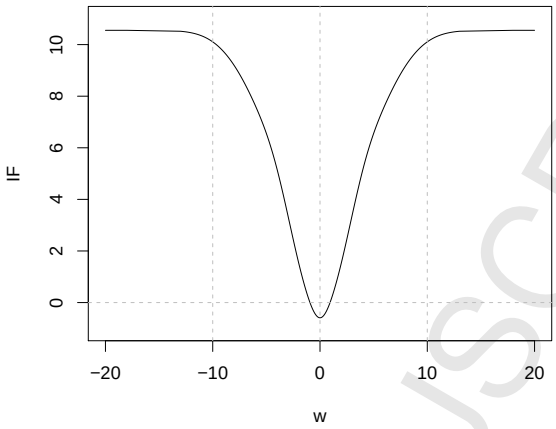

Figure 1: Influence function of the τ_{adj} estimator at the normal model with $\sigma = 1$.

2.4. Asymptotic variance

The estimators are based on heights of triangles. While the observations themselves are assumed to be independent, the heights will be autocorrelated up to order two. As in Genton (1998), the asymptotic variance of an estimator S based on the heights h_i is given by

$$\begin{aligned} ASV(S,F) &= E(IF^2(h_i;S,H_F)) + 2E(IF(h_i;S,H_F)IF(h_{i+1};S,H_F)) \\ &+ 2E(IF(h_i;S,H_F)IF(h_{i+2};S,H_F)). \end{aligned}$$

where $IF(h;S,H_F)$ is the influence function of the estimator S at the distribution H_F. At model (5) with $F_0 = N(0,1)$, the influence function of the heights can be obtained by straightforward calculus. Without loss of generality we may further assume that $\sigma = 1$. Then,

$$IF(h;S,H_N) = IF(h;S_0,H_N)\left(c_\tau\sqrt{d} - \frac{c_\tau}{2\sqrt{d}}E[\psi(Z)Z]\right) + \frac{\sigma_0}{2}\left(\frac{c_\tau}{\sqrt{d}}\rho(\frac{h}{\sigma_0}) - \sqrt{d}c_\tau\right)$$

where $\sigma_0 = \sqrt{\frac{3}{2}}$, $d = E[\rho(Z)]$, $Z \sim N(0,1)$, $\psi = \rho'$, and N is an index referring to the assumption of normality. The influence function of the initial scale estimator Q_{adj} is given by

$$IF(h; S_0, H_N) = c_q[\frac{0.25 - I(h < Q_N^{\alpha})}{2\sqrt{2/3}\varphi(\sqrt{2/3}Q_N^{0.25})}]$$

where $Q_N^{0.25} = H_N^{-1}(0.25)$ is the first quartile of the distribution of the heights at the standard normal distribution, and I is the indicator function, see Gelper et al. (2009).

The exact value of the ASV for the non-trimmed mean-squared-heights estimator MS_{adj} is 35/36, at the normal model with $\sigma = 1$. For the $\tau-$scale estimator, the ASV is obtained by numerical integration. The asymptotic relative efficiency of the τ_{adj} estimator w.r.t. MS_{adj} is then defined as

$$\text{Eff}(S, F) = \frac{ASV(MS_{adj}, F)}{ASV(S, F)}.$$

Figure 2 present the asymptotic efficiency of the τ_{adj} as a function of k, at the normal model. An efficiency of 0.95 is attained at $k = 5.48$. We also simulated, over $M = 10000$ simulation runs, finite-sample efficiencies for a window width of $n = 20$. One sees that the asymptotic results provide a good approximation for the finite sample setting. Furthermore, for k converging to zero, the efficiency of the Q_{adj} and τ_{adj} estimators coincide.

3. Simulation

In this section, we simulate a time series generated from model (1) of length $T = 1000$, with constant location and scale $\sigma_t = 1$. We consider three types of outliers: (a) isolated additive outliers, (b) patches of outliers,

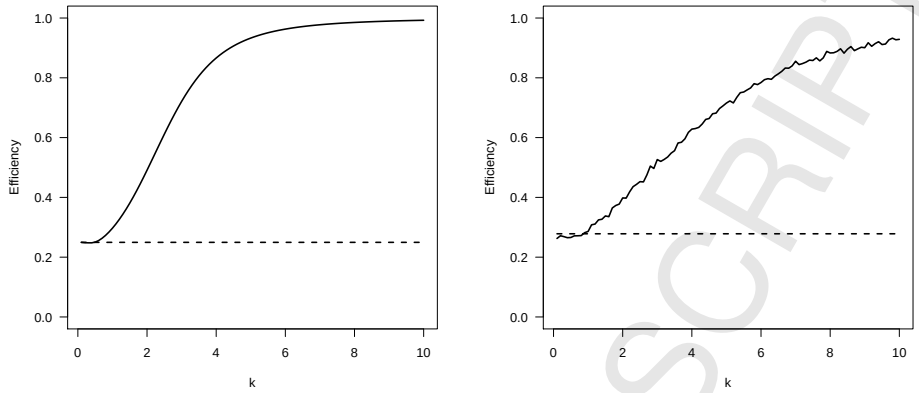

Figure 2: Asymptotic efficiencies (left) and finite sample efficiencies for $n = 20$ (right) of the τ_{adj} estimator (solid line) and the Q_{adj} estimator (dashed line) as a function of the tuning constant k at the normal model.

where a patch is a groups of 3 consecutive outliers having the same value, and (c) innovation outliers. For schemes (a) and (b) we consider independent error terms $\varepsilon_t \sim N(0, 1)$, and induce outliers by replacing a proportion ε of the observations, for $\varepsilon = 0, 0.01, 0.05$, and 0.10, by values coming from a $N(0, 5)$. In simulation scheme (c) we consider a first order autoregressive model $\varepsilon_t = \theta\varepsilon_{t-1} + v_t$, with $v_t \sim N(0, 1)$, and $\theta = 0.5$. Outliers are then induced by replacing a proportion ε of the v_t by values coming from a $N(0, 5)$. For every simulated series, we compute the root mean squared error (RMSE)

$$RMSE = \left(\frac{1}{T - n + 1} \sum_{t=n}^{T} \frac{(\hat{\sigma}_t - \sigma_t)^2}{\sigma_t^2} \right)^{1/2}.$$

Here $n = 20$ is the window width, T is the length of the time series, and $\hat{\sigma}_t$ is the estimated scale.

In Table 1 the average RMSE over the $M = 10000$ simulated time series is reported. In absence of outliers, the robust τ_{adj} is almost as efficient as

Table 1: Simulated RMSE for clean data, 1%, 5% and 10% outliers, for window width $n = 20$, $T = 1000$, and averaged over $M = 10000$ simulation runs. Three types of outlier configurations are considered.

	Additive outliers			Patches of outliers			Innovation outliers		
	Q_{adj}	τ_{adj}	MS_{adj}	Q_{adj}	τ_{adj}	MS_{adj}	Q_{adj}	τ_{adj}	MS_{adj}
$\varepsilon = 0.00$	0.44	0.24	0.22	0.44	0.24	0.22	0.41	0.30	0.29
$\varepsilon = 0.01$	0.45	0.29	0.36	0.45	0.28	0.29	0.40	0.30	0.32
$\varepsilon = 0.05$	0.51	0.46	0.70	0.49	0.38	0.50	0.40	0.30	0.44
$\varepsilon = 0.10$	0.61	0.67	1.01	0.55	0.48	0.70	0.40	0.37	0.62

the MS_{adj} estimator. This is as expected, since the tuning constant k was selected to achieve a 95% relative efficiency. This in contrast with the Q_{adj} estimator, having a RMSE which is almost twice as large. In presence of even only 1% of contamination, the MS_{adj} is no longer the most precise, and the τ-estimator behaves best. For larger amounts of contamination, i.e. 10%, the Q_{adj} estimator is slightly better than the τ_{adj}, at least for additive outliers. For patches of outliers and innovation outliers, the τ_{adj} remains to have the smallest RMSE. We conclude from the simulation study that in presence of amounts of contamination up to 10%, the τ_{adj} estimator is to be preferred. If we have a larger (but less than 25%) proportion of outliers, then the bias of the τ_{adj} remains bounded, given its high breakdown point, but the Q_{adj} estimator tends to perform better. Since we expect that most univariate time series contain few outliers and that it is in practice quite rare to have large amounts of outliers, we recommend the τ_{adj} estimator.

4. Conclusion

Robust scale estimators based on the heights of triangles formed by triplets of successive observation are used for monitoring the scale of nonlinear noisy time series, as is documented in Gelper et al (2009). In this paper we propose to use $\tau-$estimators. These estimators keep the high breakdown point of the initial estimator, while they may have an arbitrarily high efficiency. The efficiency of the τ_{adj} estimator depends on a tuning constant k. We computed the value of k yielding a 95% relative efficiency with respect to the standard estimator. Monte Carlo simulations illustrate the good performance of the proposed procedure.

A major question we did not addressed is the choice of the window width n. It needs to be small enough for (5) to hold, but large enough to still provide accurate estimates. Another topic for future research is to investigate the properties of the τ_{adj} estimator for dependent data. The scale estimator can then still be applied, and will maintain the high breakdown point. The asymptotic variance, however, will depend on the dependency structure in the data.

Appendix

Derivation of the influence function of the τ_{adj} estimator at the normal model: Assume that model (5) holds, with $F_0 = N(0,1)$. Without loss of generality, assume $a = b = 0$, and $\sigma = 1$, such that $F = N(0,1)$. From (6) it follows

that

$$\begin{aligned} IF(w;\tau_{adj},F) &= \frac{\partial}{\partial\varepsilon}\tau_{adj}(F_\varepsilon)|_{\varepsilon=0} \\ &= c_\tau \frac{\partial}{\partial\varepsilon}[S_0^2(F_\varepsilon)E_{H_{F_\varepsilon}}\rho(\frac{h}{S_0(F_\varepsilon)})]^{1/2})|_{\varepsilon=0} \\ &= \frac{c_\tau}{2}E_{H_F}\rho(h)^{-1/2}\frac{\partial}{\partial\varepsilon}[S_0^2(F_\varepsilon)E_{H_{F_\varepsilon}}\rho(\frac{h}{S_0(F_\varepsilon)})]|_{\varepsilon=0} \\ &= \frac{c_\tau}{2}E_{H_F}\rho(h)^{-1/2}\left(2IF(w;S_0,F)E_{H_F}\rho(h) + \underbrace{\frac{\partial}{\partial\varepsilon}E_{H_{F_\varepsilon}}\rho(\frac{h}{S_0(F_\varepsilon)})|_{\varepsilon=0}}_{A}\right) \end{aligned}$$

where we used that $S(F_0)=1$. In the above expression, with $\psi=\rho'$,

$$\begin{aligned} A &= \frac{\partial}{\partial\varepsilon}\int_0^\infty \rho(\frac{h}{S_0(F_\varepsilon)})dH_{F_\epsilon}(h)|_{\varepsilon=0} \\ &= -IF(w;S_0,F)\int_0^\infty \psi(h)hdH_F(h) + \int_0^\infty \rho(h)d\underbrace{\frac{\partial}{\partial\varepsilon}H_{F_\epsilon}(h)}_{B}|_{\varepsilon=0}. \end{aligned}$$

One has $H_F(u)=\Phi(\sqrt{2/3}u)-\Phi(-\sqrt{2/3}u)$, giving us for B:

$$\begin{aligned} \frac{\partial H_{F_\epsilon}(u)}{\partial\varepsilon}|_{\varepsilon=0} &= -3(2\Phi(\sqrt{2/3}h)-1)+\Phi(\sqrt{2}(h+w))-\Phi(\sqrt{2}(w-h)) \\ &+ 2(\Phi\sqrt{4/5}(w/2+h))-\Phi\sqrt{4/5}(w/2-h) \\ &:= G(h,w). \end{aligned}$$

So we can write A

$$A=-IF(w;S_0,F)\int_0^\infty \psi(h)hdH_F(h)+\int_0^\infty \rho(h)dG(h,w).$$

We conclude that, with $c_\tau^2=1/E_{H_{F_0}}\rho(h)$,

$$IF(w;\tau_{adj},F)=IF(w;S_0,F)\left[1-c_\tau^2\int_0^\infty \psi(h)hdH_F(h)\right]+\frac{c_\tau^2}{2}\int_0^\infty \rho(h)dG(h,w). \tag{11}$$

The above integrals can be computed either analytically or numerically. An expression for $IF(w;Q_{adj},F)$ is given in Gelper et al. (2009).

References

Gasser, T., Sroka, L., Jennen-Steinmetz, C., 1986. Residual variance and residual pattern in nonlinear regression. Biometrika 73 (3), 625–633.

Gelper, S., Schettlinger, K., Croux, C., Gather, U., 2009. Robust online scale estimation in time series: A regression-free approach. Journal of Statistical Planning and Inference 139, 335–339.

Genton, M. G., 1998. Asymptotic variance of M-estimators for dependent Gaussian random variables. Statist. Probab. Lett. 38 (3), 255–261.

Nunkesser, R., Fried, R., Schettlinger, K., Gather, U., 2009. Online analysis of time series by the q_n estimator. Computational Statistics and Data Analysis 53, 2354–2362.

Rousseeuw, P., Hubert, M., 1996. Regression-free and robust estimation of scale for bivariate data. Computational Statistics and Data Analysis 21, 67–85.

Yohai, J., Zamar, H., 1988. High breakdown-point estimates of regression by means of the minimization of an efficient scale. Journal of the American Statistical Association 83-402, 406–413.