Addendum to

Robust Online Scale Estimation in Time Series:

A Model-Free Approach

Sarah Gelper^{1,*}, Karen Schettlinger²,
Christophe Croux¹, and Ursula Gather²

¹ *Faculty of Business and Economics, Katholieke Universiteit Leuven, Naamsestraat
69, 3000 Leuven, Belgium.*

² *Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany.*

This report contains technical details and supplementary material for the article “Robust Online Scale Estimation in Time Series: A Model-Free Approach”. The first section covers the proofs of the results that are presented in the main paper. More specifically, we derive consistency factors, finite sample breakdown points and influence functions for the proposed estimators. The second section contains simulation results for obtaining the finite sample correction factors and the empirical influence functions. References to equations, tables or figures are either to the main article, or to this technical report if starting with I or II.

*Corresponding author. E-mail: sarah.gelper@econ.kuleuven.be, Tel: 0032/16326928, Fax: 0032/1632.67.32

1 Proofs

1.1 Consistency factors

To obtain the asymptotic consistency factor for the Q_{adj}^α estimator, we rely on the local linearity assumption with normally distributed i.i.d. error terms as in equation (3.7). If $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) = F$, then we need to select c_q such that $Q_{adj}^\alpha(F) = \sigma$. The heights for triangles formed by equidistant and consecutive time points are defined by (2.2). Now under the local linearity assumption, substitute (3.7) in (2.2) to obtain

$$h_i = \left| \left(\epsilon_{i+1} - \frac{\epsilon_i + \epsilon_{i+2}}{2} \right) \sigma \right| = \left| v_i \sqrt{\frac{3}{2}} \sigma \right|. \quad (\text{I.1})$$

with $v_i \sim N(0, 1)$.

Since Q_{adj}^α is defined by the correction constant c_q times the α quantile of all heights, it follows that

$$\begin{aligned} P \left(c_q \sqrt{3/2} \sigma |v_i| \leq Q_{adj}^\alpha(F) \right) &= 2 P \left(v_i \leq \frac{Q_{adj}^\alpha(F)}{c_q \sqrt{3/2} \sigma} \right) - 1 = \alpha \\ \Rightarrow \Phi \left(\frac{Q_{adj}^\alpha(F)}{c_q \sqrt{3/2} \sigma} \right) &= \frac{\alpha + 1}{2} \Rightarrow Q_{adj}^\alpha(F) = \Phi^{-1}((\alpha + 1)/2) c_q \sqrt{3/2} \sigma. \end{aligned}$$

Hence, in order to get Fisher consistency, we need to select

$$c_q = \sqrt{2/3} \frac{1}{\Phi^{-1}((\alpha + 1)/2)}.$$

Note that $c_q = (Q_N^\alpha)^{-1}$.

Using definition (2.8) and equation (I.1), we get

$$\begin{aligned} TM_{adj}^\alpha(F) &= c_m E \left(\sigma \sqrt{3/2} |v| \mid \sigma \sqrt{3/2} |v| \leq Q_N^\alpha \sigma \right) \\ &= c_m \frac{\sigma \sqrt{3/2}}{\alpha} \left(2 \int_0^{\sqrt{\frac{2}{3}} Q_N^\alpha} v \varphi(v) dv \right). \\ &= c_m \frac{\sigma \sqrt{6}}{\alpha} [\varphi(0) - \varphi(\sqrt{2/3} Q_N^\alpha)]. \end{aligned}$$

Hence, to get $TM_{adj}^\alpha(F) = \sigma$, we take

$$c_m = \frac{\alpha}{\sqrt{6} [\varphi(0) - \varphi(\sqrt{\frac{2}{3}} Q_N^\alpha)]}.$$

Figure 1: Finite sample breakdown points, explosion and implosion breakdown points for varying values of α for exemplary n .

For the TMS_{adj}^α defined by (2.9), we get

$$\begin{aligned}
 TMS_{adj}^\alpha(F) &= c_s E \left(\sigma^2 \frac{3}{2} v^2 \middle| |v| \leq \sqrt{\frac{2}{3}} Q_N^\alpha \right)^{1/2} \\
 &= c_s \sqrt{\frac{3}{2}} \sqrt{\frac{2}{\alpha}} \sigma \left(\int_0^{\sqrt{\frac{2}{3}} Q_N^\alpha} v^2 \varphi(v) dv \right)^{1/2} \\
 &= c_s \sqrt{\frac{3}{\alpha}} \sigma \left(-\sqrt{2/3} Q_N^\alpha \varphi(\sqrt{2/3} Q_N^\alpha) + \frac{\alpha+1}{2} - \frac{1}{2} \right)^{1/2}.
 \end{aligned}$$

Fisher consistency is ensured by taking

$$c_s = \frac{\sqrt{\alpha/3}}{\sqrt{\alpha/2 - \sqrt{2/3} Q_N^\alpha \varphi(\sqrt{2/3} Q_N^\alpha)}}.$$

1.2 The finite sample breakdown point

In the following, the maximum value possible for the finite sample breakdown point $\text{fsbp}(S, F_n)$, given by (3.5), of any of the considered scale estimates S ((2.3), (2.6) or (2.7)) at a sample $\mathbf{y}_n$ of size n with empirical distribution function F_n is determined. Further, the corresponding values of α for which the maximum breakdown is achieved are derived. Therefore, consider the quantities

$$A := \left\lceil \frac{n-1 - \lfloor \alpha(n-2) \rfloor}{3} \right\rceil \quad \text{and} \quad B := \lfloor \alpha(n-2) \rfloor.$$

By (3.5), the finite sample breakdown point $\text{fsbp}(S, F_n)$ equals $\min\{A, B\}/n$.

For increasing values of $\alpha \in [0, 1]$, the quantity A is decreasing while B is increasing (see Figure 9). Hence, the maximum breakdown point will be equal to B/n for certain $\alpha \in [\alpha_{\min}, \alpha_1)$, and it will equal A/n for certain $\alpha \in [\alpha_1, \alpha_{\max})$. The maximum value for $\text{fsbp}(S, F_n)$ is reached for any value of α for which $A = B$ whenever this is possible. To determine the bounds $\alpha_{\min}$ and $\alpha_{\max}$, we distinguish between the four cases where either $n+1$, n , $n-1$ or $n+2$ respectively is divisible by four.

Case I: $n \in \{4k-1, k \in \mathbb{N}\}$.

(a) Let $\alpha = \frac{n+1}{4(n-2)} - \varepsilon$ with arbitrary small $\varepsilon > 0$. Then

$$B = \left\lfloor \left(\frac{n+1}{4(n-2)} - \varepsilon \right) (n-2) \right\rfloor = \left\lfloor \frac{n+1}{4} - \varepsilon(n-2) \right\rfloor < \frac{n+1}{4}.$$

Hence $B/n < \frac{n+1}{4n}$ for $\alpha < \alpha_{\min} := \frac{n+1}{4(n-2)}$.

(b) Consider $x \in [n+1, n+5)$ and $\alpha = \frac{x}{4(n-2)}$. Then $B = \lfloor x/4 \rfloor = (n+1)/4$,

$$A = \left\lfloor \frac{n-1-B}{3} \right\rfloor = \left\lfloor \frac{n-1-(n+1)/4}{3} \right\rfloor = \left\lfloor \frac{n+1}{4} - \frac{2}{3} \right\rfloor = \frac{n+1}{4} = B$$

and hence $A/n = B/n = \frac{n+1}{4n}$ for $\alpha \in \left[\frac{n+1}{4(n-2)}, \frac{n+5}{4(n-2)} \right)$.

(c) For $\alpha = \frac{n+5}{4(n-2)}$ it is $B = \lfloor (n+5)/4 \rfloor = (n+1)/4 + 1$ and

$$A = \left\lfloor \frac{n-1-(n+1)/4-1}{3} \right\rfloor = \left\lfloor \frac{n+1}{4} - 1 \right\rfloor = \frac{n+1}{4} - 1.$$

Thus, $A/n < \frac{n+1}{4n}$ for all $\alpha \geq \alpha_{\max} := \frac{n+5}{4(n-2)}$.

From (a), (b), (c) and as A and B are decreasing and increasing in α , respectively, the maximum value for $\text{fsbp}(S, F_n) = \frac{n+1}{4n}$ is reached for $\alpha \in \left[\frac{n+1}{4(n-2)}, \frac{n+5}{4(n-2)} \right)$.

Case II: $n \in \{4k, k \in \mathbb{N}\}$. Analogously to Case I, one can show:

(a) For $\alpha = \frac{n}{4(n-2)} - \varepsilon$, $\varepsilon > 0$, it is $B = \lfloor \frac{n}{4} - \varepsilon(n-2) \rfloor < \frac{n}{4}$ and hence $\frac{1}{n}B < \frac{1}{4}$ for $\alpha < \alpha_{\min} := \frac{n}{4(n-2)}$.

- (b) (i) Consider $x \in [n, n+4)$ and $\alpha = \frac{x}{4(n-2)}$. Then $B = \lfloor x/4 \rfloor = n/4$ and $A = \left\lceil \frac{n-1-n/4}{3} \right\rceil = \left\lceil \frac{n}{4} - \frac{1}{3} \right\rceil = \frac{n}{4}$, and thus $A/n = B/n = 1/4$ for $\alpha \in \left[\frac{n}{4(n-2)}, \frac{n+4}{4(n-2)} \right)$.
- (ii) For $x \in [n+4, n+8)$ and $\alpha = \frac{x}{4(n-2)}$, it is $B = \lfloor x/4 \rfloor = n/4 + 1$ and $A = \left\lceil \frac{n}{4} - \frac{2}{3} \right\rceil = \frac{n}{4}$. Following it is $A/n = 1/4 < B/n$ for $\alpha \in \left[\frac{n+4}{4(n-2)}, \frac{n+8}{4(n-2)} \right)$.
- (c) For $\alpha = \frac{n+8}{4(n-2)}$ it is $B = \lfloor (n+8)/4 \rfloor = n/4 + 2$ and $A = \left\lceil \frac{n}{4} - 1 \right\rceil = \frac{n}{4} - 1$. Hence, it is $A/n < \frac{1}{4} \forall \alpha \geq \alpha_{max} := \frac{n+8}{4(n-2)}$.

Thus, the maximum value $1/4$ for the finite sample breakdown point $\text{fsbp}(S, F_n)$ is reached for $\alpha \in \left[\frac{n}{4(n-2)}, \frac{n+8}{4(n-2)} \right)$.

Case III: $n \in \{4k+1, k \in \mathbb{N}\}$

Analogously one can show that

- (a) $\frac{1}{n}B < \frac{n-1}{4n}$ for $\alpha < \alpha_{min} := \frac{n-1}{4(n-2)}$.
- (b) (i) $\frac{1}{n}A = \frac{1}{n}B = \frac{n-1}{4n}$ for $\alpha \in \left[\frac{n-1}{4(n-2)}, \frac{n+3}{4(n-2)} \right)$.
- (ii) $\frac{1}{n}A = \frac{n-1}{4n} < \frac{1}{n}B$ for $\alpha \in \left[\frac{n+7}{4(n-2)}, \frac{n+11}{4(n-2)} \right)$
- (c) $\frac{1}{n}A < \frac{n-1}{4n} \forall \alpha \geq \alpha_{max} := \frac{n+11}{4(n-2)}$.

Thus, the maximum value $\frac{n-1}{4n}$ for $\text{fsbp}(S, F_n)$ is reached for $\alpha \in \left[\frac{n-1}{4(n-2)}, \frac{n+11}{4(n-2)} \right)$.

Case IV: $n \in \{4k+2, k \in \mathbb{N}\}$

Again, analogously one can show that

- (a) $\frac{1}{n}B < (n-2)/4n$ for $\alpha < 1/4$.
- (b) $\frac{1}{n}B = \frac{n-2}{4n}$ for $\alpha \in \left[\frac{1}{4}, \frac{n+2}{4(n-2)} \right)$.
- (c) $\frac{1}{n}A = \frac{n-2}{4n}$ for $\alpha \in \left[\frac{n+2}{4(n-2)}, \frac{n+14}{4(n-2)} \right)$.
- (d) $\frac{1}{n}A \leq \frac{n-2}{4n} - \frac{1}{n}$ for any $\alpha \geq \frac{n+14}{4(n-2)}$.

From (a), (b), (c) and (d) we can conclude that the maximum breakdown point $\text{fsbp}(S, F_n) = \frac{n-2}{4n}$ is reached for $\alpha \in \left[\frac{1}{4}, \frac{n+14}{4(n-2)}\right)$. Note that in this case, the equation $A = B$ does not hold for any value of $\alpha \in [0, 1]$ because:

$$\begin{aligned} A &= B \\ \Leftrightarrow \left\lceil \frac{n-1 - \lfloor \alpha(n-2) \rfloor}{3} \right\rceil &= \lfloor \alpha(n-2) \rfloor \\ \Leftrightarrow \lfloor \alpha(n-2) \rfloor - 1 &< \frac{n-1 - \lfloor \alpha(n-2) \rfloor}{3} \leq \lfloor \alpha(n-2) \rfloor \\ \Leftrightarrow \frac{n-1}{4} &\leq \lfloor \alpha(n-2) \rfloor < \frac{n+2}{4}. \end{aligned}$$

By definition $\lfloor \alpha(n-2) \rfloor \in \mathbb{N}$. However, for $n \in \{4k+2, k \in \mathbb{N}\}$ there is no integer in the interval $\left[\frac{n-1}{4}, \frac{n+2}{4}\right)$.

1.3 Influence functions

For the derivation of equation (3.9), recall that the functional of the Q_{adj}^α estimator is defined by (2.4)

$$Q_{adj}^\alpha(F) = c_q H_F^{-1}(\alpha),$$

for any distribution F of the error terms ϵ_i . Here,

$$H_F(u) = P\left(\left|\epsilon_{i+1} - \frac{\epsilon_i + \epsilon_{i+2}}{2}\right| \leq u\right),$$

for all $u > 0$. The IF as defined in (3.8) then equals

$$\text{IF}(w, Q_{adj}^\alpha, F) = c_q \frac{\partial}{\partial \varepsilon} H_{F_\varepsilon}^{-1}(\alpha) \Big|_{\varepsilon=0} = c_q \frac{-\frac{\partial H_{F_\varepsilon}}{\partial \varepsilon} (H_F^{-1}(\alpha)) \Big|_{\varepsilon=0}}{H'_F(H_F^{-1}(\alpha))}. \quad (\text{I.2})$$

Here, assuming $F = N(0, 1)$, we have

$$H_F^{-1}(\alpha) = H_N^{-1}(\alpha) = Q_N^\alpha = \sqrt{\frac{3}{2}} \Phi^{-1}\left(\frac{\alpha+1}{2}\right),$$

referring to the assumption of normality by the index N .

We now compute numerator and denominator of (I.2) separately. For the numerator, we need the distribution function of the heights from the contaminated

series y_t^ε , which can be written as

$$\begin{aligned} H_{F_\varepsilon}(u) &= (1 - \varepsilon)^3 H_F(u) + \varepsilon(1 - \varepsilon)^2 P\left(\left|w - \frac{y_{t-1} + y_{t+1}}{2}\right| \leq u\right) \\ &\quad + 2\varepsilon(1 - \varepsilon)^2 P\left(\left|y_t - \frac{w + y_{t+1}}{2}\right| \leq u\right) + O(\varepsilon^2), \end{aligned}$$

for any $u > 0$. Computing the derivative of this expression with respect to ε and evaluating in $\varepsilon = 0$ yields

$$\left. \frac{\partial H_{F_\varepsilon}(u)}{\partial \varepsilon} \right|_{\varepsilon=0} = -3 H_F(u) + P\left(\left|w - \frac{y_{t-1} + y_{t+1}}{2}\right| \leq u\right) + P\left(\left|y_t - \frac{w + y_{t+1}}{2}\right| \leq u\right).$$

Since $F = N(0, 1)$, standard calculations give

$$\begin{aligned} \left. \frac{\partial H_{N_\varepsilon}(h)}{\partial \varepsilon} \right|_{\varepsilon=0} &= -3(2\Phi(\sqrt{2/3}h) - 1) + \Phi(\sqrt{2}(h - w)) - \Phi(\sqrt{2}(-h - w)) \\ &\quad + 2(\Phi(\sqrt{(4/5)}((w/2) + h)) - \Phi(\sqrt{(4/5)}((w/2) - h))) \\ &:= G(h, w), \end{aligned} \tag{I.3}$$

and we note that $G(Q_N^\alpha)$ equals expression (3.10).

For the denominator of expression (I.2), we note that

$$H_F(u) = P(h \leq u) = P\left(\left|y_{t+1} - \frac{y_t + y_{t+2}}{2}\right| \leq u\right) = \Phi(\sqrt{2/3}u) - \Phi(-\sqrt{2/3}u)$$

from which it follows that

$$\left. \frac{\partial H_F(u)}{\partial u} \right|_{u=Q_N^\alpha} = 2\sqrt{2/3} \varphi(\sqrt{2/3} Q_N^\alpha). \tag{I.4}$$

The expression of the $IF(Q_{adj}^\alpha, N(0, 1), w)$ for normal distributed error terms follows from equations (I.2), (I.3) and (I.4).

The functionals TM_{adj}^α and TMS_{adj}^α both take the same form

$$M : F \mapsto M(F) = c E(h^p | h \leq Q_F^\alpha)^{1/p} = \frac{c}{\alpha} \left(\int_0^{Q_F^\alpha} h^p dH_F(h) \right)^{1/p}, \tag{I.5}$$

where $p = 1$ for the TM_{adj}^α and $p = 2$ for the TMS_{adj}^α , c is either c_m (equation (2.11)) or c_s (equation (2.12)) and $Q_F^\alpha = H_F^{-1}(\alpha)$. The influence function, as defined in equation (3.8), is given by

$$IF(w, M, F) = \left. \frac{\partial}{\partial \varepsilon} \frac{c}{\alpha} \left(\int_0^{Q_{F_\varepsilon}^\alpha} h^p dH_{F_\varepsilon}(h) \right)^{1/p} \right|_{\varepsilon=0},$$

and the chain rule delivers

$$\text{IF}(w, M, F) = \frac{c}{\alpha} \frac{1}{p} E(h^q | h \leq Q_F^\alpha)^{(1/p)-1} \frac{\partial}{\partial \varepsilon} \int_0^{Q_{F_\varepsilon}^\alpha} h^p dH_{F_\varepsilon}(h) \Big|_{\varepsilon=0}.$$

We know that $E(h^p | h \leq Q_F^\alpha) = 1/c^p$, so

$$\begin{aligned} \text{IF}(w, M, F) &= \frac{c^p}{p\alpha} \left[\frac{\partial}{\partial p} \int_0^p h^p dH_F(h) \Big|_{p=Q_F^\alpha} \frac{\partial}{\partial \varepsilon} Q_{F_\varepsilon}^\alpha \Big|_{\varepsilon=0} \right. \\ &\quad \left. + \frac{\partial}{\partial \varepsilon} \int_0^{Q_F^\alpha} h^p dH_{F_\varepsilon}(h) \Big|_{\varepsilon=0} \right] \\ &:= \frac{c^p}{p\alpha} (T_1 + T_2). \end{aligned}$$

We will now separately examine the first and second term T_1 and T_2 .

Applying the Leibnitz rule to the first term yields

$$T_1 = (Q_F^\alpha)^p H'_F(Q_F^\alpha) \frac{\partial}{\partial \varepsilon} Q_{F_\varepsilon}^\alpha \Big|_{\varepsilon=0}.$$

Recall now that the influence function of Q_{adj}^α can be rewritten as

$$\text{IF}(w, Q_{adj}, N(0, 1)) = c_q \frac{\partial}{\partial \varepsilon} Q_{N_\varepsilon}^\alpha \Big|_{\varepsilon=0} = c_q \frac{-G(Q_N^\alpha, w)}{H'_N(Q_N^\alpha)},$$

hence for $F = N(0, 1)$

$$T_1 = -(Q_N^\alpha)^p G(Q_N^\alpha, w).$$

To obtain the second term T_2 , we use the expression for G in (I.3) and obtain

$$\begin{aligned} T_2 &= \int_0^{Q_N^\alpha} h^p dG(h) \\ &= -3 \int_0^{Q_N^\alpha} h^p dH_F(h) + \int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{2}(h-w)) \\ &\quad - \int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{2}(-h-w)) \\ &\quad + 2 \int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{(4/5)}((w/2)+h) - 2c/\alpha) \\ &\quad - 2 \int_0^{Q_N^\alpha} h^p d\Phi(\sqrt{(4/5)}((w/2)-h)) \\ &= -3 \frac{\alpha}{c^p} + \int_0^{Q_N^\alpha} h^p \left[\sqrt{2}(\varphi(\sqrt{2}(w+h)) + \varphi(\sqrt{2}(w-h))) \right. \\ &\quad \left. + 2\sqrt{4/5}(\varphi(\sqrt{4/5}(w/2+h)) + \varphi(\sqrt{4/5}(w/2-h))) \right] dh. \end{aligned}$$

Defining

$$I_{a,b}^p = \int_0^{Q_N^\alpha} h^p \varphi(aw + bh) dh, \quad (\text{I.6})$$

results in expression (3.11) for $\text{IF}(w, M, N(0, 1))$

In expression (3.11) of the influence function, there still appear integrals which have to be computed. In practice, we can do this by numerical integration techniques; or we can derive an analytic expression for $I_{a,b}^p$. Using the substitution method for solving integrals, define $u = aw + bh$ and rewrite equation (I.6) as

$$I_{a,b}^p = \int_{aw}^{aw+bQ} \left(\frac{u - aw}{b} \right)^p \varphi(u) \frac{1}{b} du.$$

We now distinguish between $p = 1$ and $p = 2$. For $p = 1$ one can verify that

$$I_{a,b}^1 = \frac{\varphi(aw) - \varphi(aw + bQ)}{b^2} + \frac{aw}{b^2} (\Phi(aw) - \Phi(aw + bQ)),$$

and for $p = 2$, after some calculations, we get

$$I_{a,b}^2 = \frac{1}{b^3} [(aw - bQ)\varphi(aw + bQ) - aw\varphi(aw) + (1 + a^2w^2)\Phi(aw + bQ) - (1 + a^2w^2)\Phi(aw)]$$

In the special case where $\alpha = 1$, we obtain simplified expressions for the mean over all heights and the mean over all squared heights:

$$\begin{aligned} \text{IF}(w, TM_{adj}^1, N(0, 1)) &= c_m \left[-3\sqrt{6}\varphi(0) + 2w\Phi(\sqrt{2}w) + \sqrt{2}\varphi(\sqrt{2}w) - w \right. \\ &\quad \left. + 2(\sqrt{5}\varphi(\sqrt{1/5}w) + w\Phi(\sqrt{1/5}w) - w/2) \right], \end{aligned}$$

and

$$\text{IF}(w, TMS_{adj}^1, N(0, 1)) = \frac{1}{2}(w^2 - 1).$$

2 Simulations

2.1 Correction factors

In definitions (2.3), (2.6) and (2.7), consistency factors are used to achieve Fisher consistency. However, these estimators may still be biased for finite samples;

n	$\alpha = \frac{n+1}{4(n-2)}$			$\alpha = 0.5$			$\alpha = 1$	
	Q_{adj}^α	TM_{adj}^α	TMS_{adj}^α	Q_{adj}^α	TM_{adj}^α	TMS_{adj}^α	TM_{adj}^α	TMS_{adj}^α
10	2.18	4.40	4.07	1.27	2.61	2.32	1.02	0.85
15	2.43	4.90	4.42	1.34	2.76	2.42	1.02	0.84
20	2.61	5.27	4.70	1.24	2.57	2.24	1.02	0.84
50	2.46	4.98	4.34	1.22	2.53	2.18	1.02	0.82
100	2.57	5.17	4.49	1.22	2.52	2.18	1.02	0.82
200	2.57	5.18	4.49	1.21	2.52	2.17	1.02	0.82
∞	2.56	5.17	4.47	1.21	2.51	2.16	1.02	0.82

Table 1: Correction factors obtained by 10000 simulation runs for several values of n and α , together with the consistency factors ($n \rightarrow \infty$), under the assumption of Gaussian noise.

replacing the consistency factors by correction factors then ensures unbiasedness of the scale estimators. Those correction factors c_q^n , c_m^n and c_s^n depend on the window width n and differ slightly from the consistency factors. To obtain the correction factors, we assume linearity and a constant scale σ within the time window considered, as in (3.7). Here, F_0 will be taken to be the standard normal $N(0, 1)$. Table 1 gives values of the correction factors under the assumption of i.i.d. Gaussian error terms and based on 10000 simulation runs for different window widths and different values of α . In the first column, we choose α such that the optimal finite sample breakdown point is achieved, as is explained in Section 3.1. From Table 1 it is clear that as the window width increases, the correction factors converge to their asymptotic values. These asymptotic values are the consistency factors in the definition of the functionals in equations (2.3), (2.6) and (2.7).

In practice, we suggest using Q_{adj}^α with $\alpha = 0.5$ because this yields a good compromise between robustness and efficiency. For application to real data, we advise using the finite sample correction factor for this estimator. Though this can be obtained for any sample size by simulation, it might be rather time consuming.

Approximation of the finite sample correction factors

Figure 2: Approximation of the finite sample correction factors for the Q_{adj}^α for α equal to 0.5. The dots represent the simulated value of d_n for 10000 simulations and the solid line is the approximation using formula (II.1).

Therefore, we propose a simple approximative formula for the finite sample correction factor of the $Q_{adj}^{0.5}$ under the assumption of i.i.d. normal error terms. The correction factor c_q^n can be written as a multiple of its asymptotic value:

$$c_q^n = d_n c_q .$$

As c_q^n approaches c_q for n tending to infinity, we know that d_n attains one. We run 10000 simulations for varying window widths, up to $n = 200$ observations. A plot of the simulated values for d_n can be found in Figure 2. The shape of the plot is well approximated by a function of the form

$$d_n = \frac{n}{n - a},$$

where the constant a can be estimated by an OLS regression of the following model:

$$n(d_n - 1) = a d_n + u_t$$

which yields $a = 0.44$. Since c_q equals 1.21 for $\alpha = 0.5$, we get

$$d_n \approx \frac{n}{n - 0.44} \Rightarrow c_q^n \approx 1.21 \frac{n}{n + 0.44}. \quad (\text{II.1})$$

This approximation gives values close to the large scale simulated values of d_n , as can be seen by the solid line in Figure 2, and is easy to use in practice.

For the pure descriptive purpose of monitoring scale, one will often not be willing to make distributional assumptions on the noise component. If this is the case, then the estimators may be computed omitting the correction factors. The estimated sequence S_t will of course be the same, up to scalar multiplication.

2.2 Empirical influence functions

The theoretical influence function in equation (3.8) is an asymptotic concept. In practice however, we work with finite sample sizes or, for the considered application, window widths. A finite sample equivalent of the influence function defined in (3.8) is given by the Empirical Influence Function (EIF). It is a simulation based curve and defined by

$$\text{EIF}(w; S, F, n, N_s) = \frac{n}{N_s} \sum_{s=1}^{N_s} (S(\tilde{\mathbf{y}}^s) - S(\mathbf{y}^s)), \quad (\text{II.2})$$

where $n = 20$ the window width, and $N_s = 10000$ is the number of simulations. For every simulated sample $\mathbf{y}^s$ of size n from a $N(0, 1)$ distribution the scale estimate is computed. This sample is then contaminated by replacing one randomly selected observation by the value w , resulting in a contaminated series $\tilde{\mathbf{y}}^s$ from which again the scale estimator S is computed. Figure 3 shows the EIF of the different adjacent type estimators, similar to Figure 1 in the main article, for $\alpha = 1$ and $\alpha = \alpha_{opt} \approx 0.29$ as defined in equation (3.6). It is striking how close the theoretical and empirical influence functions are already for $n = 20$.

Figure 3: Empirical Influence functions (EIF) for the Q_{adj}^α , TM_{adj}^α , and TMS_{adj}^α estimators, for α corresponding to the optimal finite sample breakdown point and for $\alpha = 1$.

References

Gelper, S., Schettlinger, K., Croux, C. and Gather, U., 2008. Robust online scale estimation in time series: A model-free approach.