Robust scale and autocovariance estimation

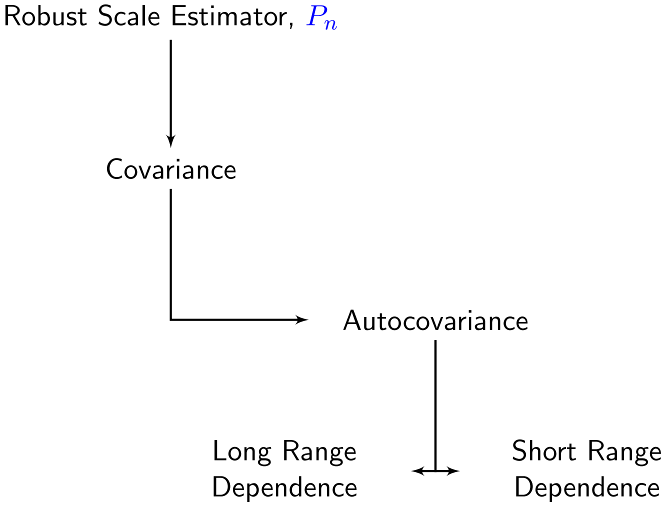
Garth Tarr, Samuel Müller and Neville Weber

EMS 2013

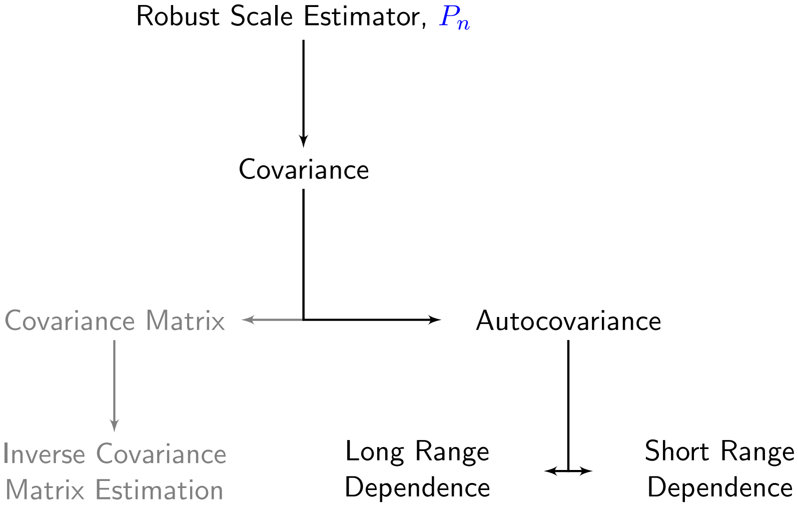
School of Mathematics and Statistics
THE UNIVERSITY OF SYDNEY

THE UNIVERSITY OF SYDNEY

Outline


Robust Scale Estimator, P_n
Covariance
Autocovariance
Long Range Dependence
Short Range Dependence

Outline


Robust Scale Estimator, P_n
Covariance
Covariance Matrix
Autocovariance
Inverse Covariance Matrix Estimation
Long Range Dependence
Short Range Dependence

Outline

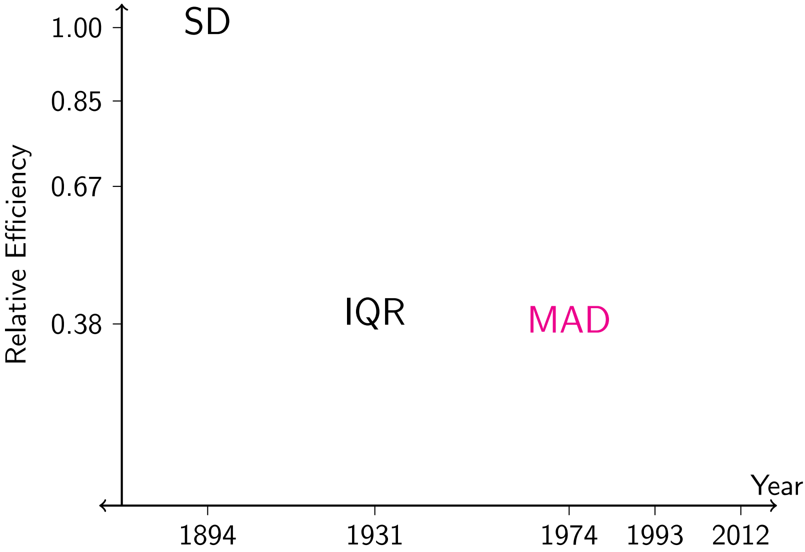
The robust scale estimator P_n

Autocovariance estimation using P_n

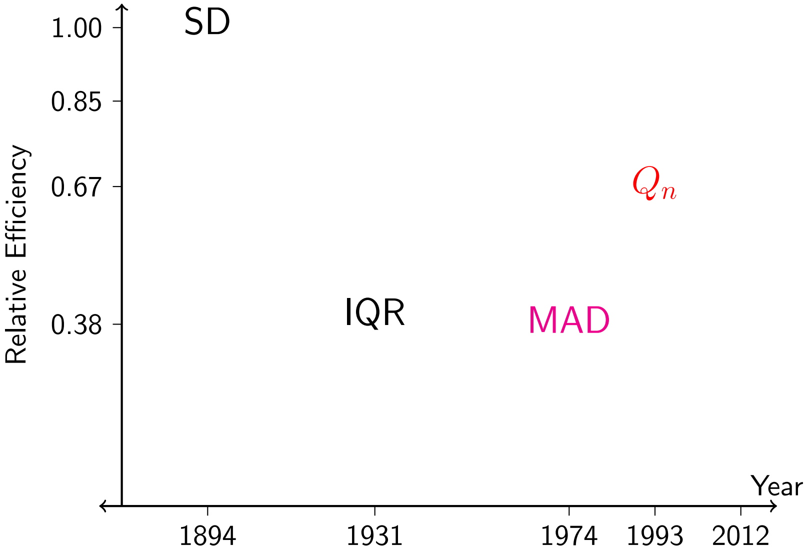
Conclusion and key references

History of scale efficiencies at the Gaussian ($n = 20$)

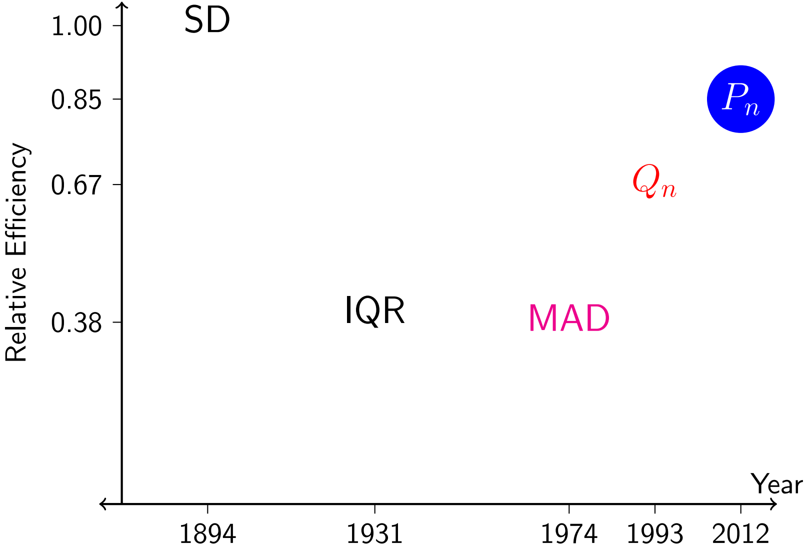
History of scale efficiencies at the Gaussian ($n = 20$)



History of scale efficiencies at the Gaussian ($n = 20$)



History of scale efficiencies at the Gaussian ($n = 20$)



Pairwise mean scale estimator: P_n

- Consider the U-statistic, based on the pairwise mean kernel,

$$U_n(\mathbf{X}) := \binom{n}{2}^{-1} \sum_{i<j} \frac{X_i + X_j}{2}.$$

- Let $H(t) = P((X_i + X_j)/2 \leq t)$ be the cdf of the kernels with corresponding empirical distribution function,

$$H_n(t) := \binom{n}{2}^{-1} \sum_{i<j} \mathbb{I}\left\{\frac{X_i + X_j}{2} \leq t\right\}, \quad \text{for } t \in \mathbb{R}.$$

Definition (Interquartile range of pairwise means)

$$P_n = c\left[H_n^{-1}(0.75) - H_n^{-1}(0.25)\right],$$

where $c \approx 1.048$ is a correction factor to ensure P_n is consistent for the standard deviation when the underlying observations are Gaussian.

Influence curve

- Hampel (1974) defines the influence curve for a functional T at distribution F as
$$\mathrm{IC}(x; T, F) = \lim_{\epsilon \downarrow 0} \frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon}$$
where δ_x has all its mass at x.
- Serfling (1984) outlines the IC for GL-statistics.

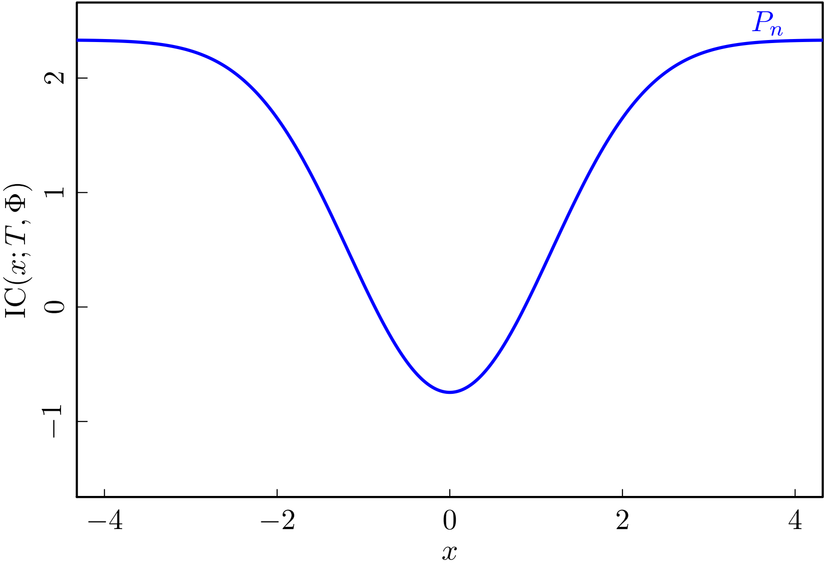
Influence curve

- Hampel (1974) defines the influence curve for a functional T at distribution F as
$$\mathrm{IC}(x;T,F)=\lim_{\epsilon\downarrow 0}\frac{T((1-\epsilon)F+\epsilon\delta_x)-T(F)}{\epsilon}$$
where δ_x has all its mass at x.
- Serfling (1984) outlines the IC for GL-statistics.

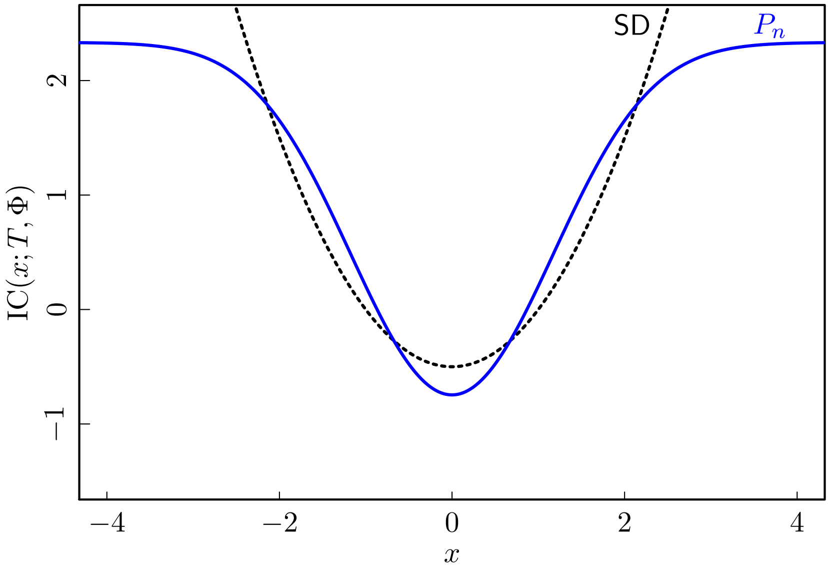
Influence curve for P_n (Tarr, Müller and Weber, 2012)

Assuming that F has derivative $f>0$ on $[F^{-1}(\epsilon),F^{-1}(1-\epsilon)]$ for all $\epsilon>0$,
$$\mathrm{IC}(x;P,F)=c\left[\frac{0.75-F(2H_F^{-1}(0.75)-x)}{\int f(2H_F^{-1}(0.75)-x)f(x)\,\mathrm{d}\,x}-\frac{0.25-F(2H_F^{-1}(0.25)-x)}{\int f(2H_F^{-1}(0.25)-x)f(x)\,\mathrm{d}\,x}\right].$$

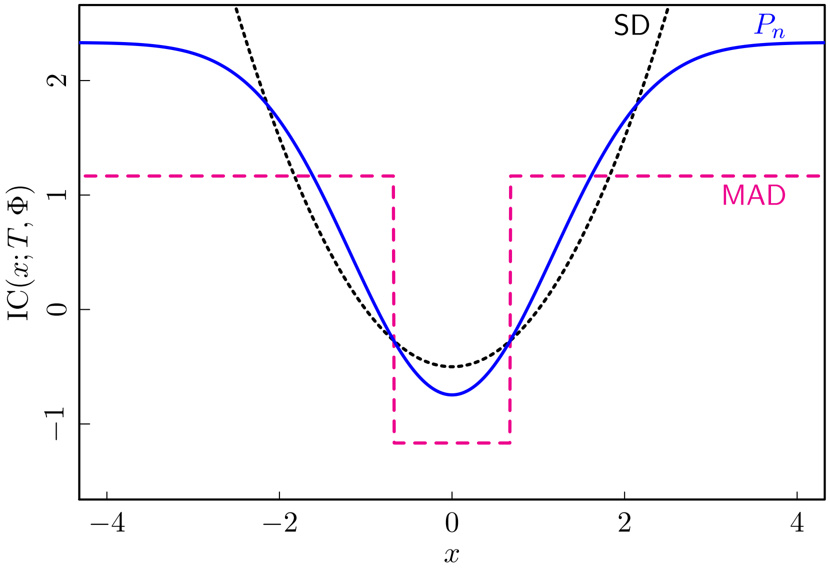
Influence curves when $F = \Phi$



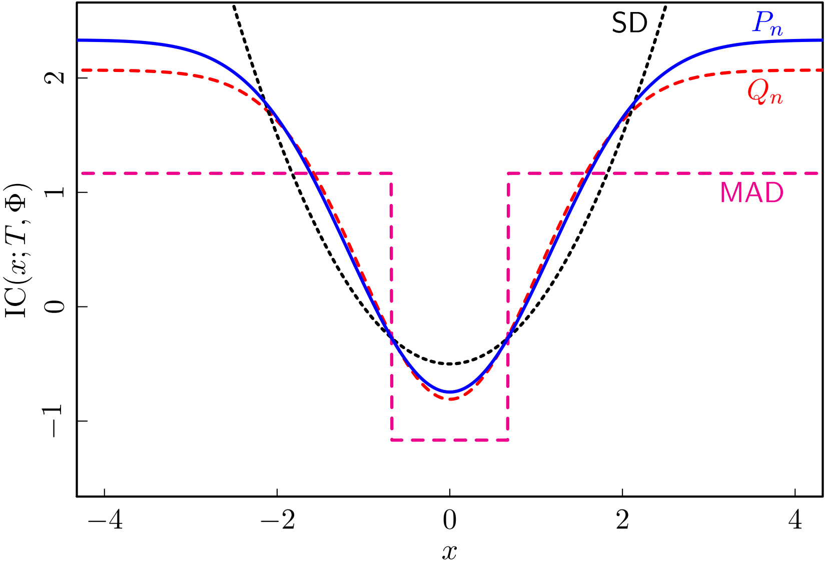
Influence curves when $F = \Phi$



Influence curves when $F = \Phi$



Influence curves when $F = \Phi$



Asymptotic variance and relative efficiency

- Tarr, Müller and Weber (2012) show that when the underlying observations are independent, P_n is asymptotically normal with variance, V, given by the expected square of the influence function.
- When the underlying data are independent Gaussian,

$$V(P_n, \Phi) = \int \mathrm{IC}(x; P, \Phi)^2 \, \mathrm{d}\, \Phi(x) = 0.579.$$

Asymptotic variance and relative efficiency

- Tarr, Müller and Weber (2012) show that when the underlying observations are independent, P_n is asymptotically normal with variance, V, given by the expected square of the influence function.
- When the underlying data are independent Gaussian,

$$V(P_n, \Phi) = \int \mathrm{IC}(x; P, \Phi)^2 \, \mathrm{d}\, \Phi(x) = 0.579.$$

- This equates to an asymptotic efficiency of **0.86** as compared with **0.82** for Q_n and **0.37** for the MAD.

Asymptotic variance and relative efficiency

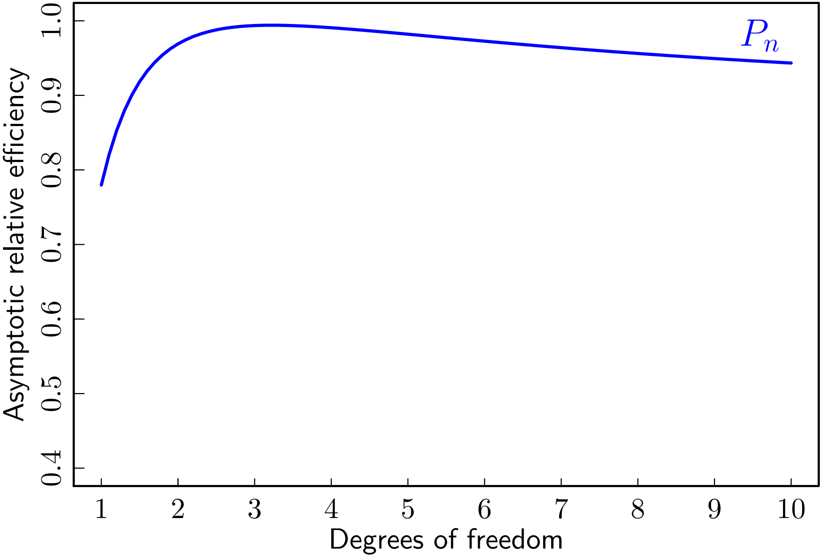
- Tarr, Müller and Weber (2012) show that when the underlying observations are independent, P_n is asymptotically normal with variance, V, given by the expected square of the influence function.
- When the underlying data are independent Gaussian,

$$V(P_n, \Phi) = \int \mathrm{IC}(x; P, \Phi)^2 \,\mathrm{d}\,\Phi(x) = 0.579.$$

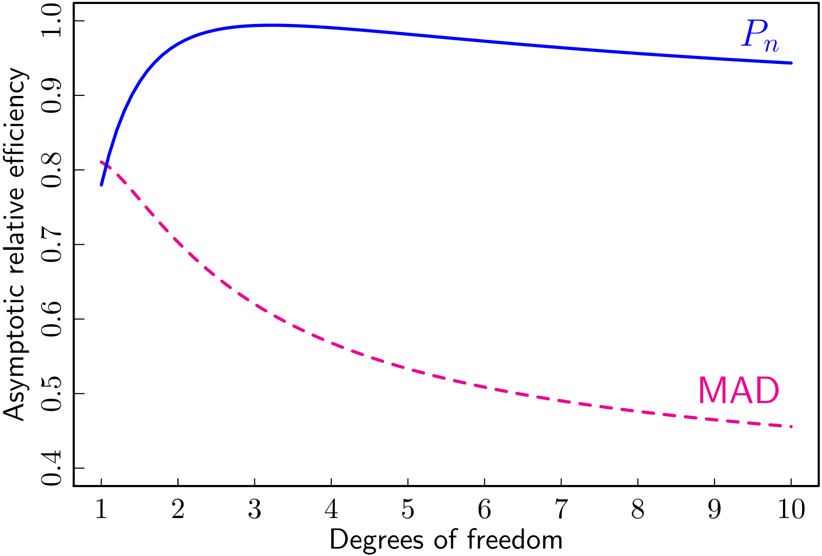
- This equates to an asymptotic efficiency of **0.86** as compared with **0.82** for Q_n and **0.37** for the MAD.

! But how does it compare at heavier tailed distributions?

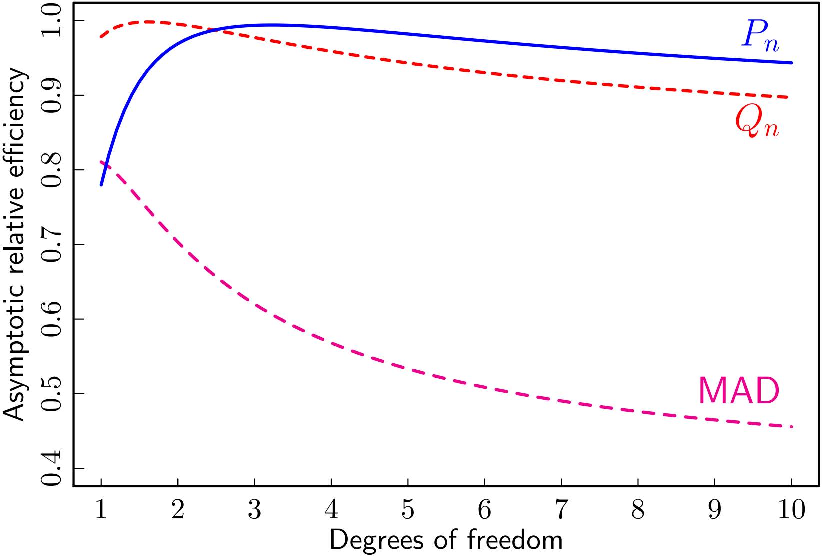
Asymptotic relative efficiency when $f = t_\nu$ for $\nu \in [1, 10]$



Asymptotic relative efficiency when $f = t_\nu$ for $\nu \in [1, 10]$



Asymptotic relative efficiency when $f = t_\nu$ for $\nu \in [1, 10]$



Asymptotic distribution under short range dependence

Assumption 1: Gaussian SRD

Let $\{X_i\}_{i\geq 1}$ be a stationary mean-zero Gaussian process with autocovariances $\gamma(h) = \mathbb{E}(X_1 X_{h+1})$ satisfying $\sum_{h\geq 1} |\gamma(h)| < \infty$.

Result

Under Assumption 1, it can be shown that,

$$\sqrt{n}(P_n - \sigma) = \frac{c_\Phi}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{IC}(X_i, P, \Phi) + o_p(1),$$

hence applying a limit theorem from Arcones (1994) we have

$$\sqrt{n}(P_n - \sigma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\sigma}^2),$$

where $\sigma = \sqrt{\gamma(0)}$ and

$$\tilde{\sigma}^2 = \mathbb{E}\left(\mathrm{IC}^2(X_1, P, \Phi)\right) + 2\sum_{k\geq 1} \mathbb{E}\left(\mathrm{IC}(X_1, P, \Phi)\mathrm{IC}(X_{k+1}, P, \Phi)\right).$$

Asymptotic distribution under long range dependence

Assumption 2: Gaussian LRD

Let $\{X_i\}_{i\geq 1}$ be a stationary mean-zero Gaussian process with autocovariance sequence $\gamma(h) = \mathbb{E}(X_1 X_{h+1})$ satisfying $\gamma(h) = h^{-D}L(h)$, $0 < D < 1$, where L is slowly varying at infinity.

Result

Under Assumption 2,

1. If $D > 1/2$,

$$\sqrt{n}(P_n - \sigma) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\sigma}^2),$$

$$\tilde{\sigma}^2 = \mathbb{E}\,\mathrm{IC}^2(X_1, P, \Phi) + 2\sum_{k\geq 1} \mathbb{E}\,\mathrm{IC}(X_1, P, \Phi)\mathrm{IC}(X_{k+1}, P, \Phi).$$

! Note that this is the same as the SRD result, however the proof is somewhat more involved.

Asymptotic distribution under long range dependence

Assumption 2: Gaussian LRD

Let $\{X_i\}_{i\geq 1}$ be a stationary mean-zero Gaussian process with autocovariance sequence $\gamma(h) = \mathbb{E}(X_1 X_{h+1})$ satisfying $\gamma(h) = h^{-D}L(h)$, $0 < D < 1$, where L is slowly varying at infinity.

Conjecture

Under Assumption 2,

2. If $D < 1/2$,

$$k(D)n^D L(n)^{-1}(P_n - \sigma) \xrightarrow{\mathcal{D}} \frac{\sigma}{2}\left(Z_{2,D}(1) - Z_{1,D}^2(1)\right),$$

where $k(D) = B((1-D)/2, D)$ and B denotes the beta function $Z_{1,D}$ is the standard fBm process and $Z_{2,D}$ is the Rosenblatt process.

! This is consistent with results for other scale estimators, e.g. SD and Q_n (Lévy-Leduc et. al., 2011) and the IQR...

Interquartile range result

Consider equivalent result for the interquartile range,

$$T_n(\mathbf{x}) = c_\Phi \left(F_n^{-1}(3/4) - F_n^{-1}(1/4) \right).$$

Interquartile range result

Consider equivalent result for the interquartile range,

$$T_n(\mathbf{x}) = c_\Phi \left(F_n^{-1}(3/4) - F_n^{-1}(1/4)\right).$$

Result

Under Assumption 2 with $D < 1/2$, T_n satisfies the following limit theorem as $n \to \infty$:

$$k(D)n^D L(n)^{-1}(T_n - \sigma) \xrightarrow{\mathcal{D}} \frac{\sigma}{2}\left(Z_{2,D}(1) - Z_{1,D}^2(1)\right).$$

- In establishing this result we used Wu's (2005) Bahadur representation for sample quantiles under LRD:

$$F_n^{-1}(p) - r = \frac{p - F_n(r)}{\phi(r)} + \frac{\bar{X}_n^2}{2}\frac{\phi'(r)}{\phi(r)} + O(n^{h(D)}L_1(n)).$$

Interquartile range result

Consider equivalent result for the interquartile range,

$$T_n(\mathbf{x}) = c_\Phi \left(F_n^{-1}(3/4) - F_n^{-1}(1/4)\right).$$

Result

Under Assumption 2 with $D < 1/2$, T_n satisfies the following limit theorem as $n \to \infty$:

$$k(D)n^D L(n)^{-1}(T_n - \sigma) \xrightarrow{\mathcal{D}} \frac{\sigma}{2}\left(Z_{2,D}(1) - Z_{1,D}^2(1)\right).$$

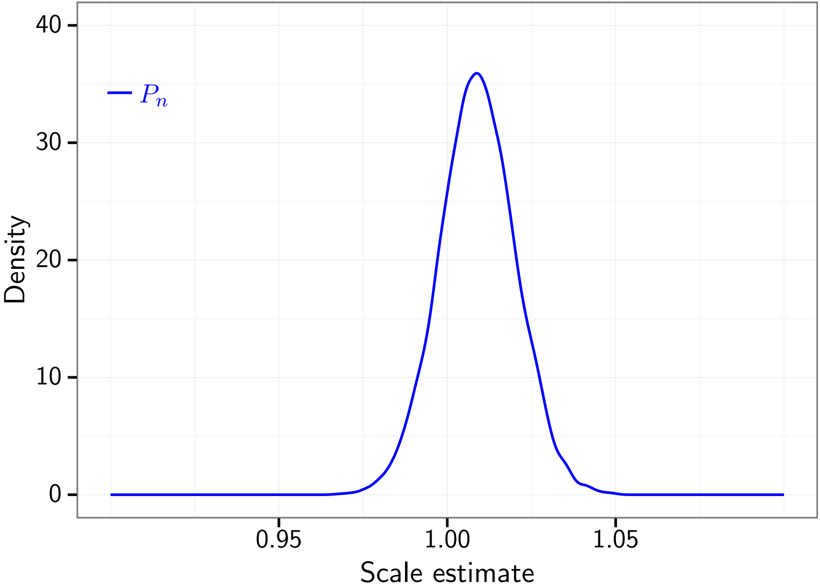
- In establishing this result we used Wu's (2005) Bahadur representation for sample quantiles under LRD:

$$F_n^{-1}(p) - r = \frac{p - F_n(r)}{\phi(r)} + \frac{\bar{X}_n^2}{2}\frac{\phi'(r)}{\phi(r)} + O(n^{h(D)}L_1(n)).$$

! It appears to be non-trivial to extend this result to the pairwise mean empirical distribution function.

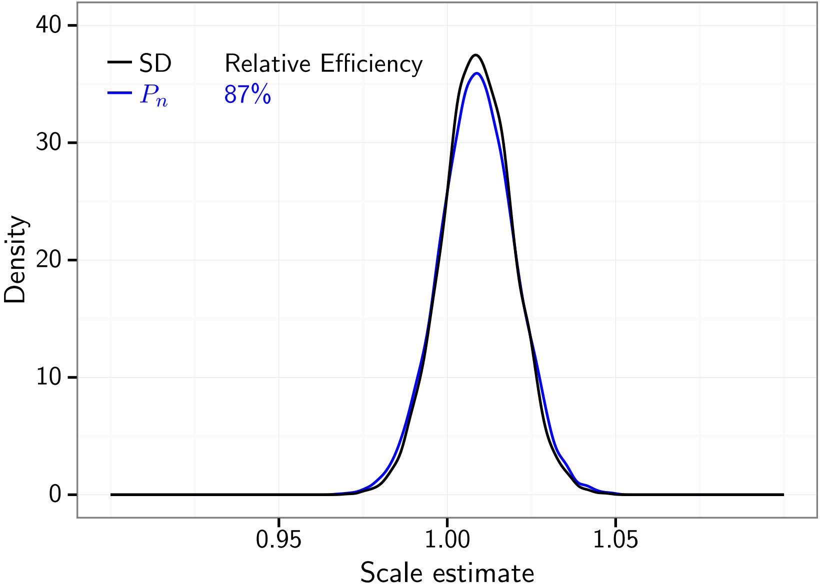
Empirical densities

ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



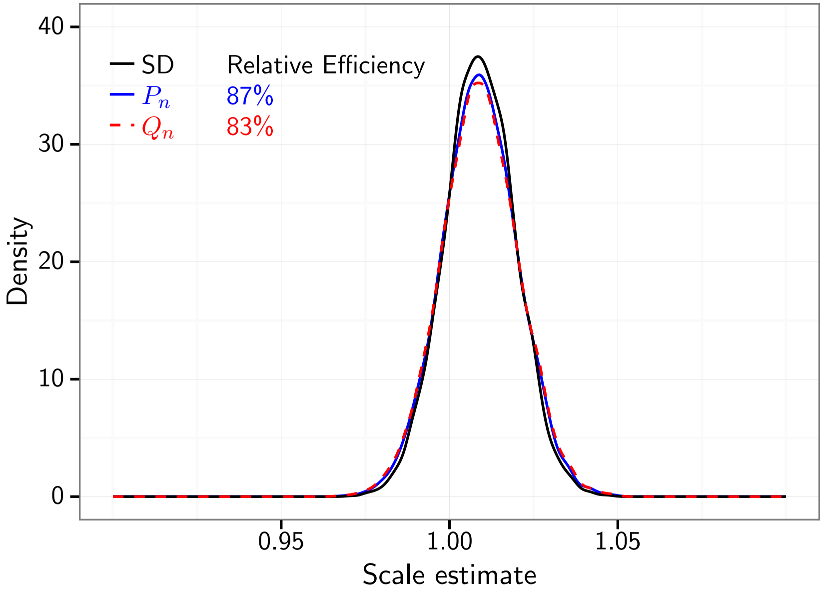
Empirical densities

ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



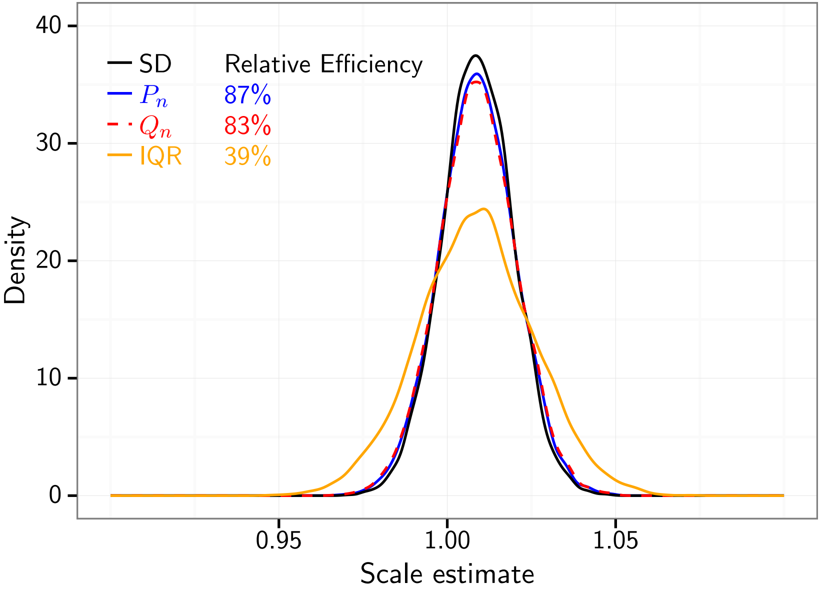
Empirical densities

ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



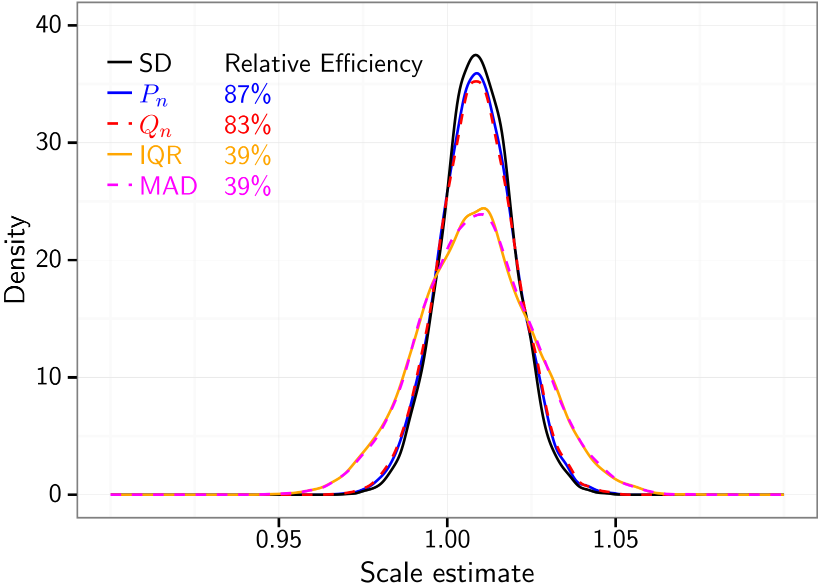
Empirical densities

ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



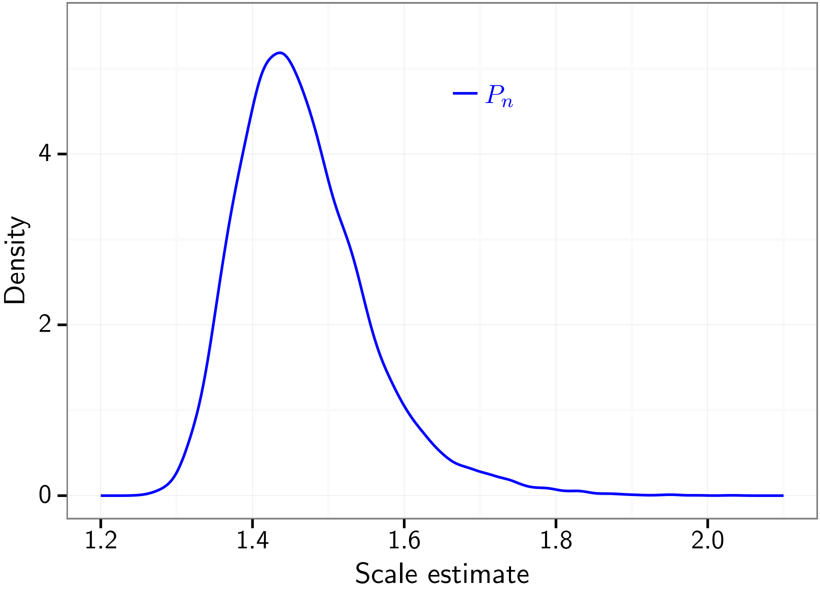
Empirical densities

ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



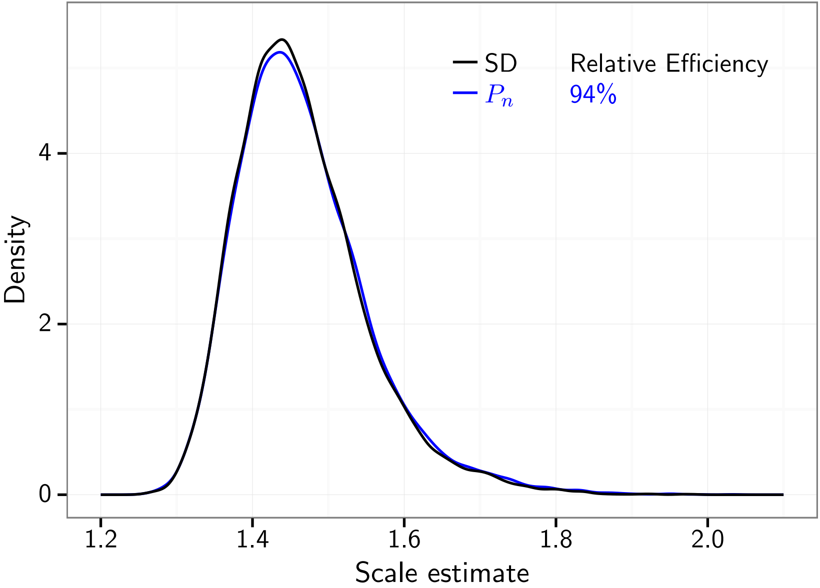
Empirical densities

ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



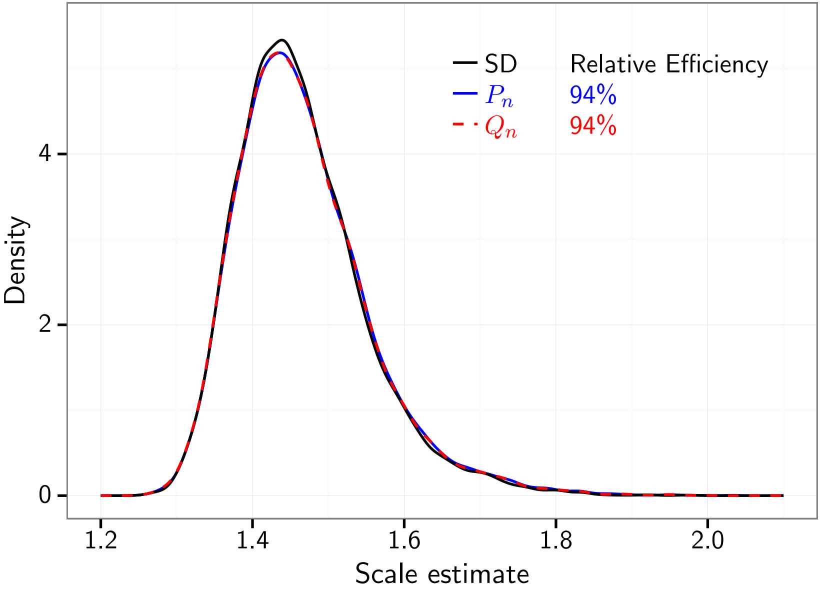
Empirical densities

ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



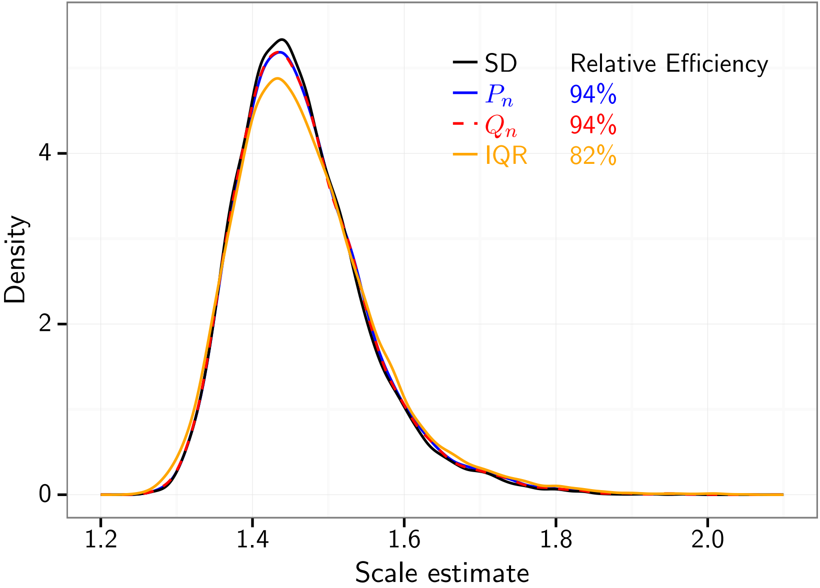
Empirical densities

ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



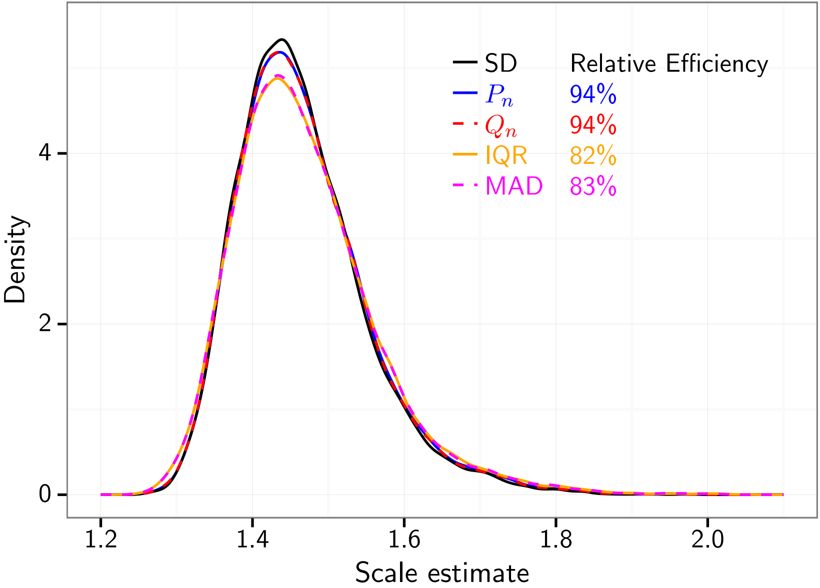
Empirical densities

ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$

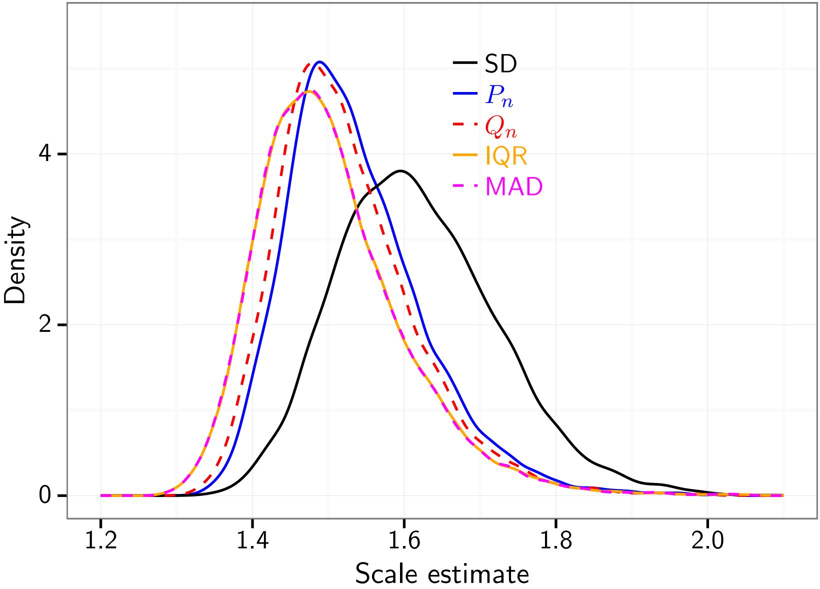


Empirical densities

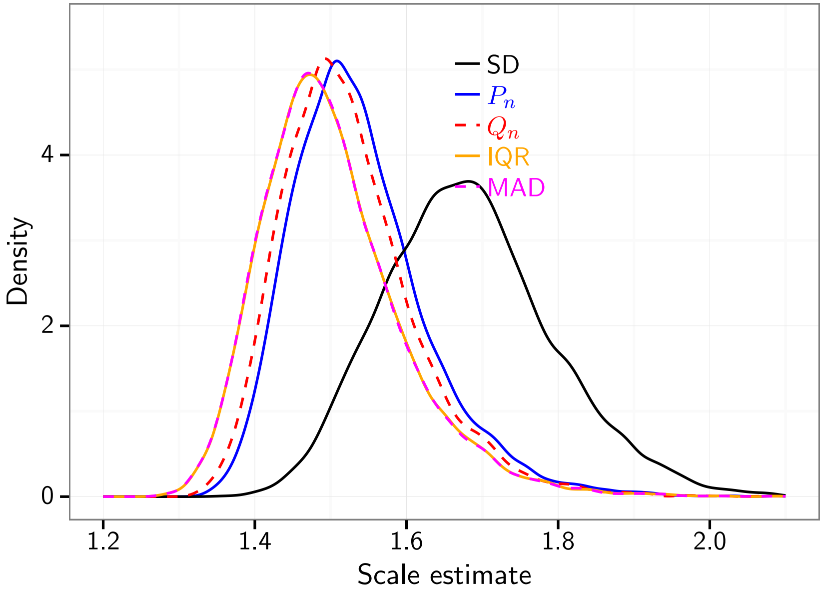
ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



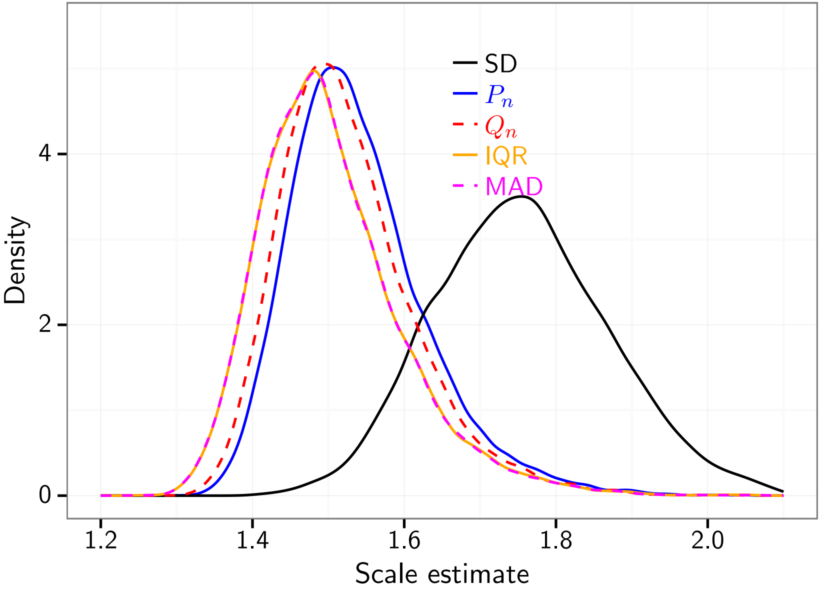
Empirical densities with 2% contamination at 5



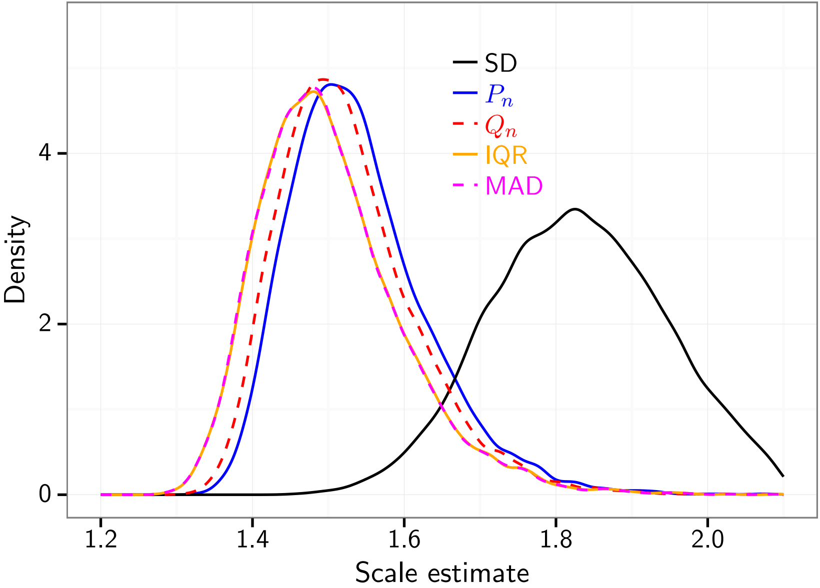
Empirical densities with 2% contamination at 6



Empirical densities with 2% contamination at 7



Empirical densities with 2% contamination at 8



Outline

The robust scale estimator P_n

Autocovariance estimation using P_n

Conclusion and key references

From scale to autocovariance

- Gnanadesikan and Kettenring (1972) highlighted an identity that relates scale and covariance.
- In the autocovariance setting,

$$\gamma(h) = \operatorname{cov}(X_1, X_{h+1}) = \frac{1}{4}\left[\operatorname{var}(X_1 + X_{h+1}) - \operatorname{var}(X_1 - X_{h+1})\right].$$

- For a series of n observations, $\mathbf{X}_n = \{X_t\}_{1 \leq t \leq n}$,

$$\hat{\gamma}_P(h) = \frac{1}{4}\left[P^2_{n-h}(\mathbf{X}_{1:n-h} + \mathbf{X}_{h+1:n}) - P^2_{n-h}(\mathbf{X}_{1:n-h} - \mathbf{X}_{h+1:n})\right].$$

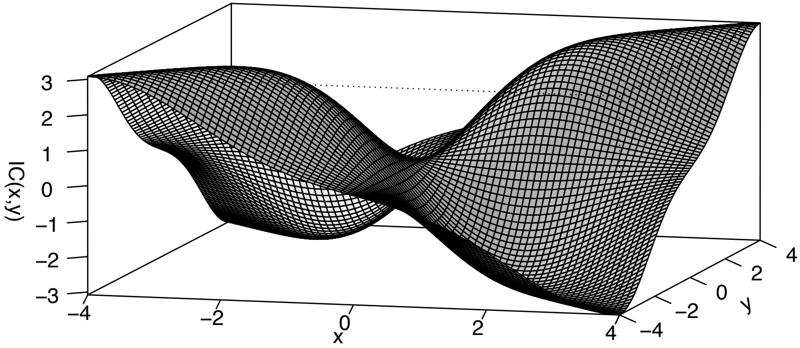
  where $\mathbf{X}_{1:n-h}$ are the first $n-h$ observations in $\mathbf{X}_n$ and $\mathbf{X}_{h+1:n}$ are the last $n-h$ observations.
- The same technique can be used to turn other scale estimators into covariance and correlation estimators, e.g. Ma and Genton (2000) study γ_Q based on Q_n.

Robustness properties

- Following Genton and Ma (1999), we have shown that **influence curve**, $\mathrm{IC}((x,y),\gamma_P,\Phi)$, and therefore the **gross error sensitivity** of γ_P can be derived from the IC of P_n.

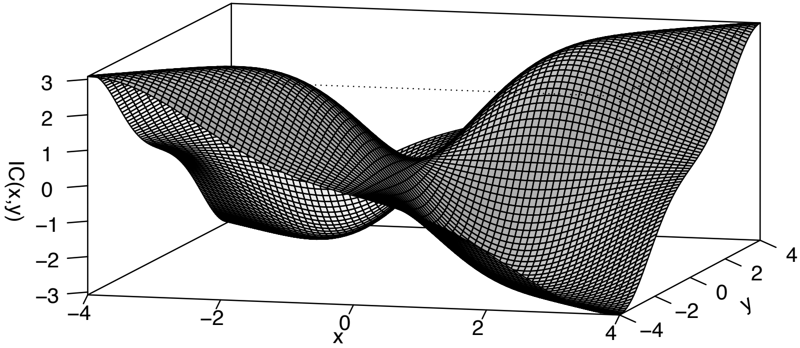
Robustness properties

- Following Genton and Ma (1999), we have shown that **influence curve**, $\mathrm{IC}((x,y),\gamma_P,\Phi)$, and therefore the **gross error sensitivity** of γ_P can be derived from the IC of P_n.



Robustness properties

- Following Genton and Ma (1999), we have shown that **influence curve**, $\mathrm{IC}((x,y),\gamma_P,\Phi)$, and therefore the **gross error sensitivity** of γ_P can be derived from the IC of P_n.



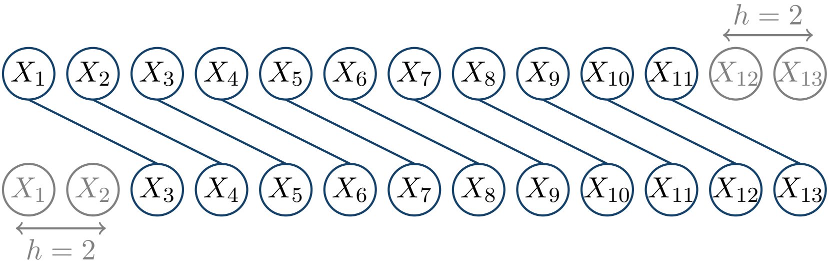
! IC is **bounded** and hence the gross error sensitivity is **finite**.

Breakdown value

- Ma and Genton (2000) show that the breakdown value for autocovariance estimators is (roughly) half that of the corresponding covariance estimator.
- Consider $n = 13$

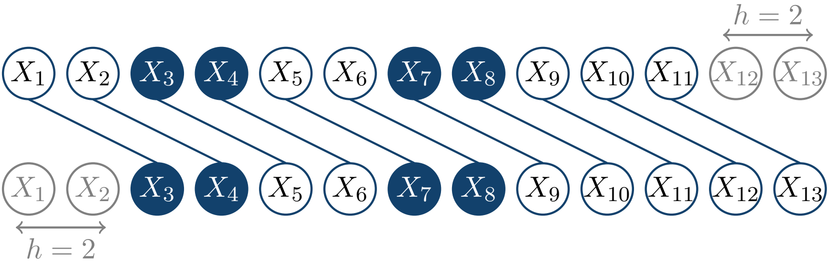
Breakdown value

- Ma and Genton (2000) show that the breakdown value for autocovariance estimators is (roughly) half that of the corresponding covariance estimator.
- Consider $n = 13$, autocovariance at lag $h = 2$
- Working with $\mathbf{X}_{1:11} \pm \mathbf{X}_{3:13}$



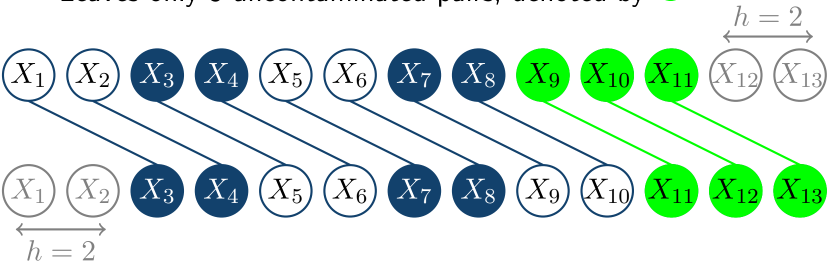
Breakdown value

- Ma and Genton (2000) show that the breakdown value for autocovariance estimators is (roughly) half that of the corresponding covariance estimator.
- Consider $n = 13$, autocovariance at lag $h = 2$
- Working with $\mathbf{X}_{1:11} \pm \mathbf{X}_{3:13}$
- 4 contaminated observations, denoted by ●



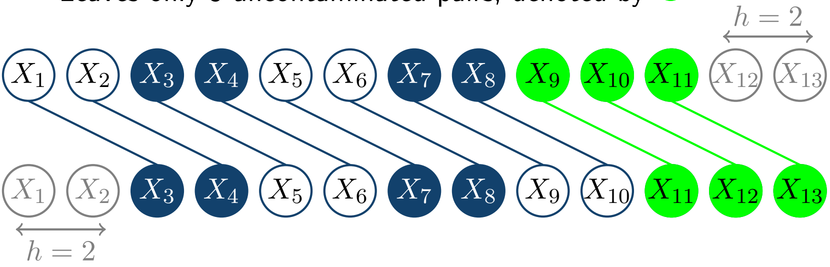
Breakdown value

- Ma and Genton (2000) show that the breakdown value for autocovariance estimators is (roughly) half that of the corresponding covariance estimator.
- Consider $n = 13$, autocovariance at lag $h = 2$
- Working with $\mathbf{X}_{1:11} \pm \mathbf{X}_{3:13}$
- 4 contaminated observations, denoted by ●
- Leaves only 3 uncontaminated pairs, denoted by ●



Breakdown value

- Ma and Genton (2000) show that the breakdown value for autocovariance estimators is (roughly) half that of the corresponding covariance estimator.
- Consider $n = 13$, autocovariance at lag $h = 2$
- Working with $\mathbf{X}_{1:11} \pm \mathbf{X}_{3:13}$
- 4 contaminated observations, denoted by ●
- Leaves only 3 uncontaminated pairs, denoted by ●



! $\hat{\gamma}_P(h)$ has roughly half the **13.4%** breakdown value of P_n.

Asymptotic distribution under short range dependence

Result

If $\{X_i\}_{i\geq 1}$ satisfy Assumption 1 (SRD) then the sequences $\{X_i + X_{i+h}\}_{i\geq 1}$ and $\{X_i - X_{i+h}\}_{i\geq 1}$ also satisfy Assumption 1, hence we can use the previous results and apply the functional delta method to show,

$$\sqrt{n}(\hat{\gamma}_P(h) - \gamma(h)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \breve{\sigma}^2(h)),$$

where

$$\begin{aligned}\breve{\sigma}^2(h) &= \mathbb{E}\left[\mathrm{IC}^2((X_1, X_{1+h}), \gamma_P, \Phi)\right] \\ &\quad + 2\sum_{k\geq 1} \mathbb{E}\left[\mathrm{IC}((X_1, X_{1+h}), \gamma_P, \Phi)\,\mathrm{IC}((X_{k+1}, X_{k+1+h}), \gamma_P, \Phi)\right].\end{aligned}$$

Asymptotic distribution under long range dependence

Result

Under Assumption 2 (LRD),

1. $D > 1/2$ (same limiting distribution as SRD):

$$\sqrt{n}(\hat{\gamma}_P(h) - \gamma(h)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \breve{\sigma}^2(h)).$$

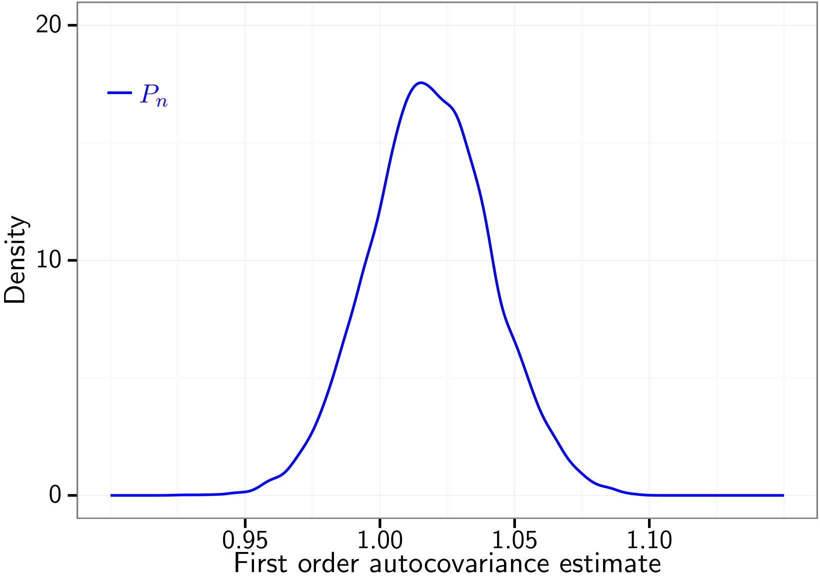
2. $D < 1/2$ (based on conjectured result):

$$\frac{k(D)n^D}{\tilde{L}(n)}(\hat{\gamma}_P(h) - \gamma(h)) \xrightarrow{\mathcal{D}} \frac{\gamma(0) + \gamma(h)}{2}(Z_{2,D}(1) - Z_{1,D}^2(1)),$$

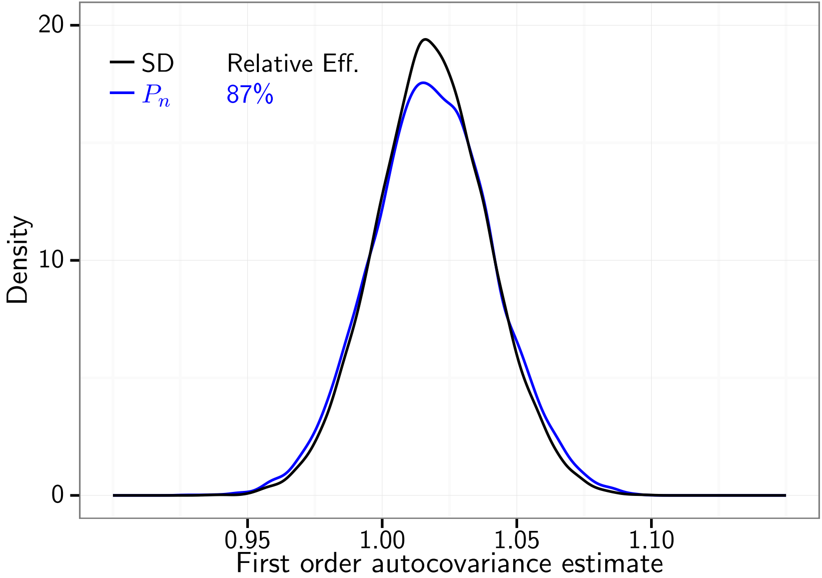
where $k(D) = \mathrm{Beta}((1-D)/2, D)$, $Z_{1,D}$ is the standard fBm process, $Z_{2,D}$ is the Rosenblatt process and

$$\tilde{L}(n) = 2L(n) + L(n+h)(1+h/n)^{-D} + L(n-h)(1-h/n)^{-D}.$$

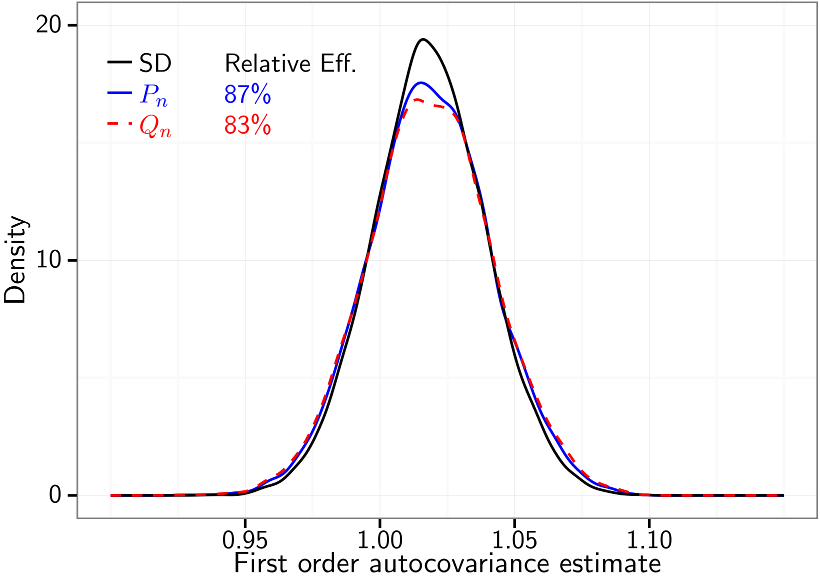
Empirical densities $\hat{\gamma}_\bullet(1)$ ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



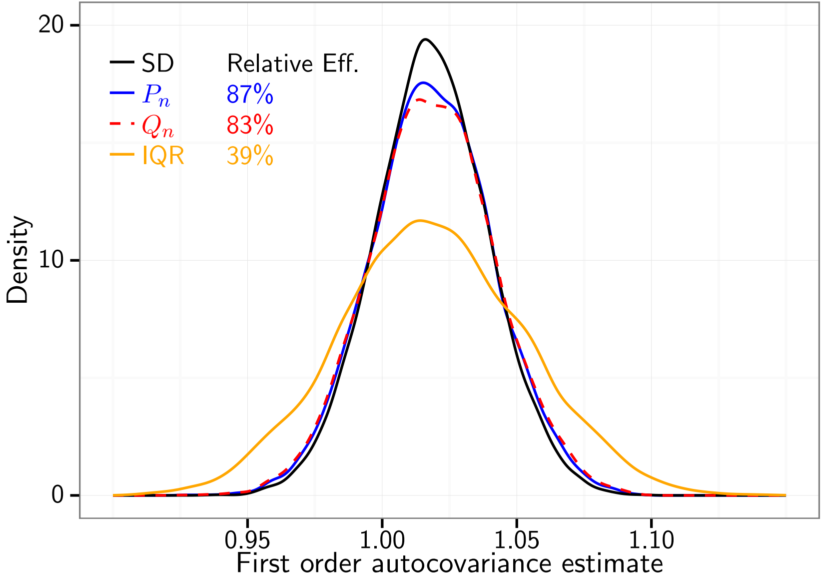
Empirical densities $\hat{\gamma}_{\bullet}(1)$ ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



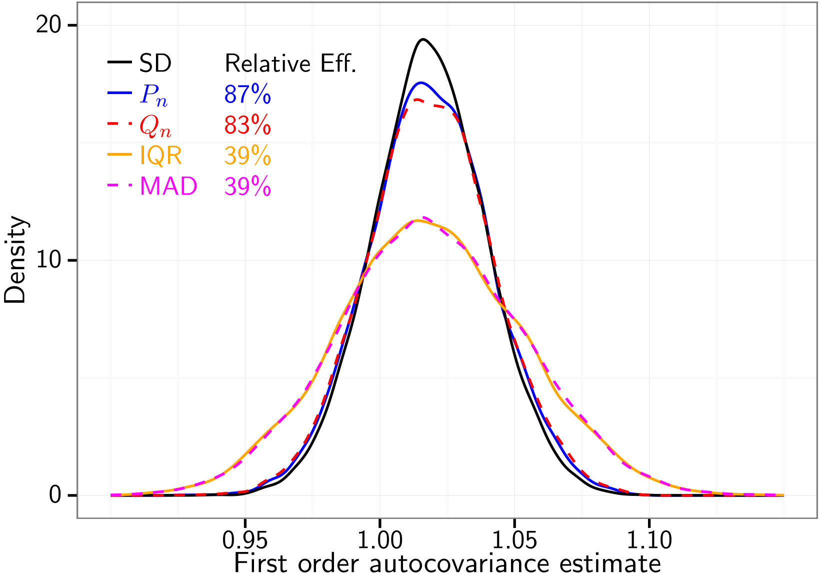
Empirical densities $\hat{\gamma}_\bullet(1)$ ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



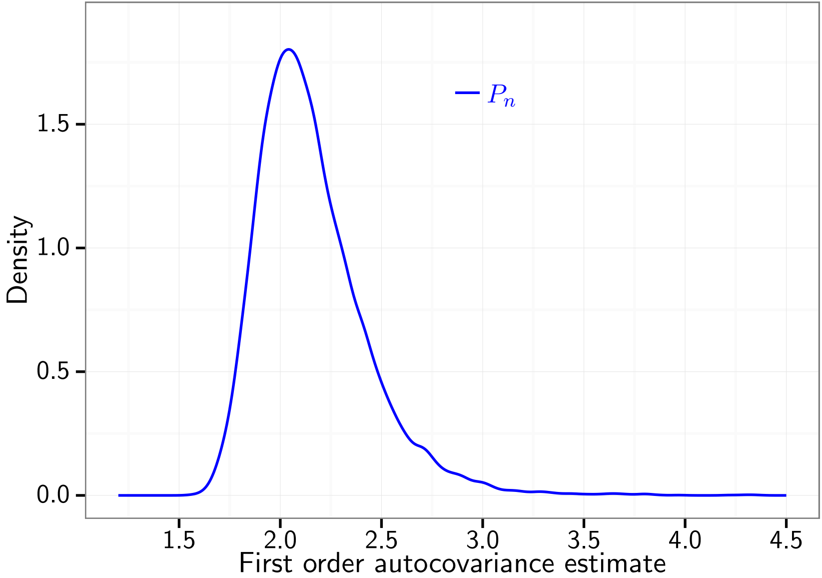
Empirical densities $\hat{\gamma}_{\bullet}(1)$ ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



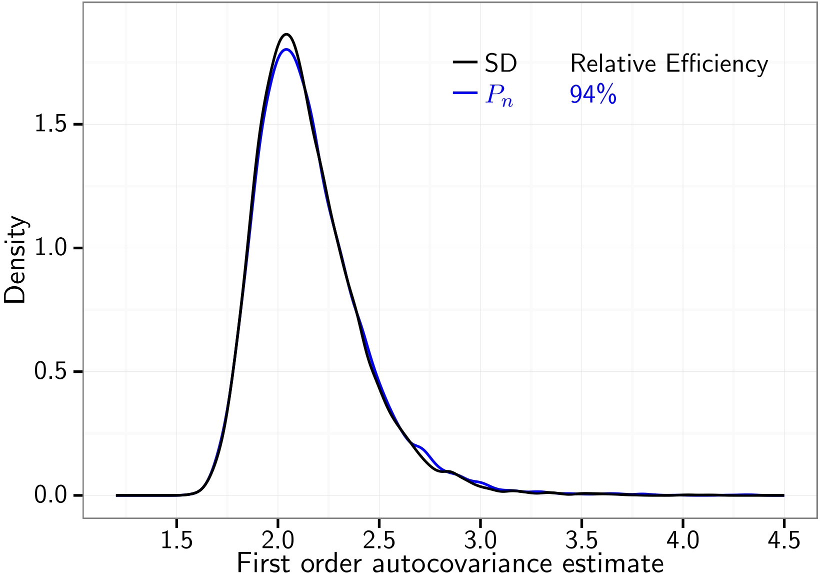
Empirical densities $\hat{\gamma}_{\bullet}(1)$ ARFIMA(0, 0.1, 0); $D = 0.8$; $n = 5000$



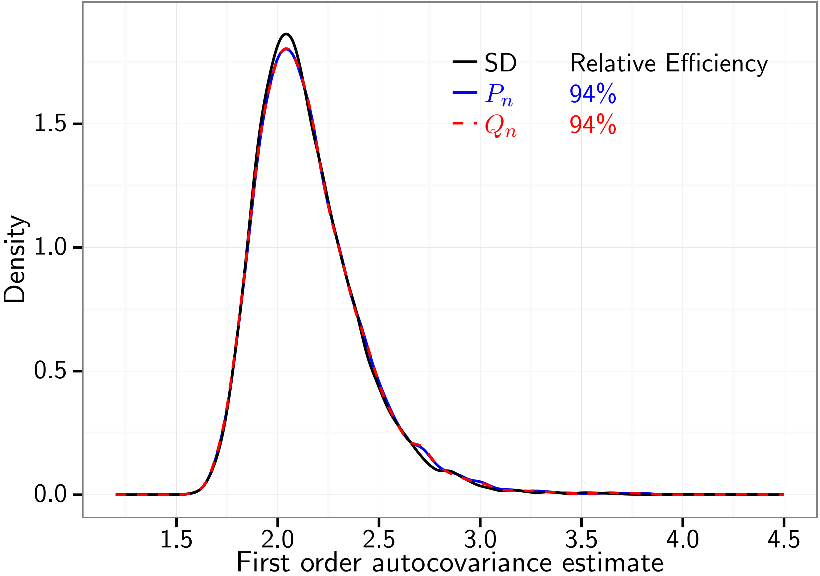
Empirical densities $\hat{\gamma}_\bullet(1)$ ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



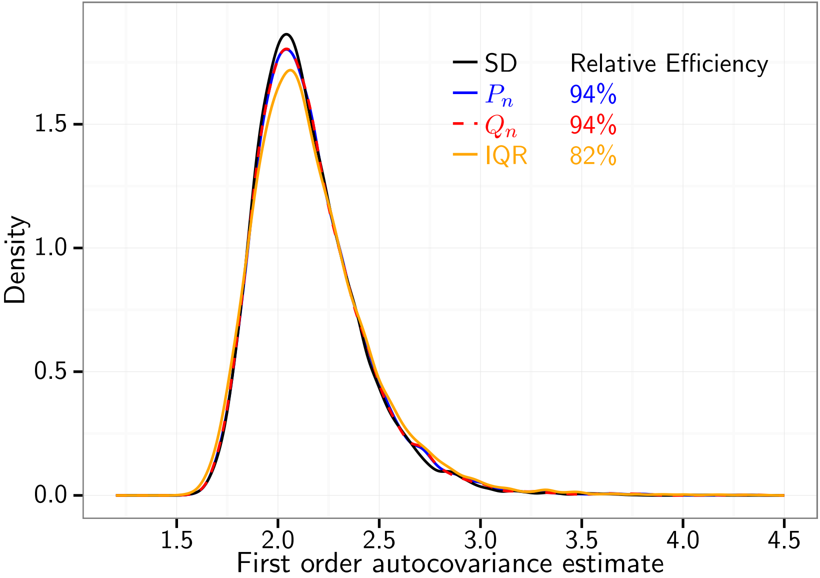
Empirical densities $\hat{\gamma}_{\bullet}(1)$ ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



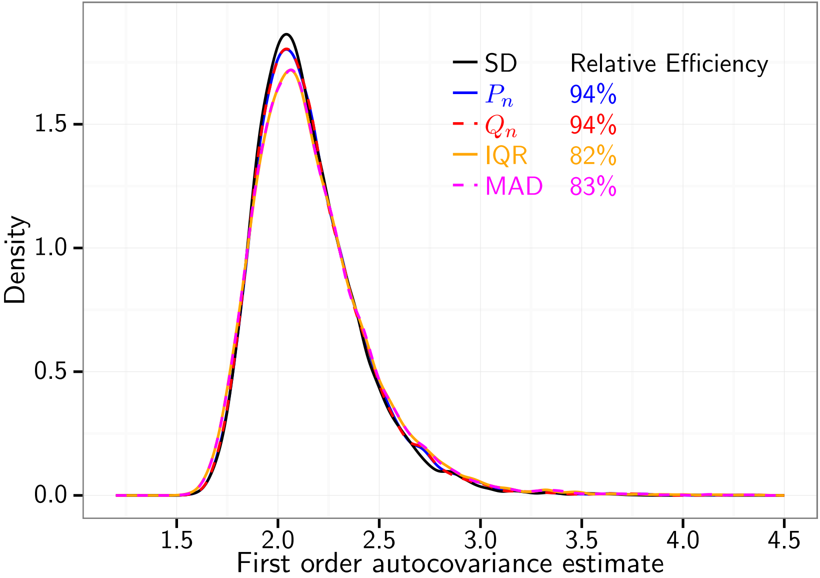
Empirical densities $\hat{\gamma}_\bullet(1)$ ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



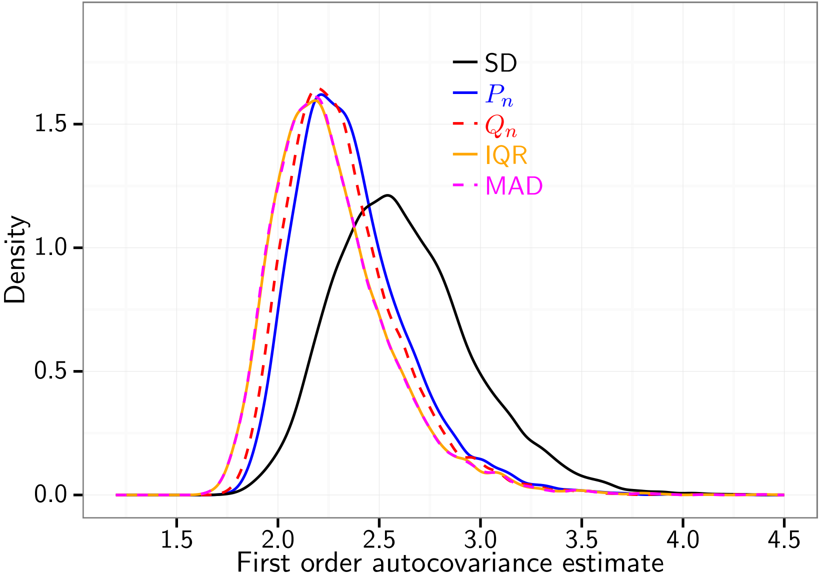
Empirical densities $\hat{\gamma}_{\bullet}(1)$ ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



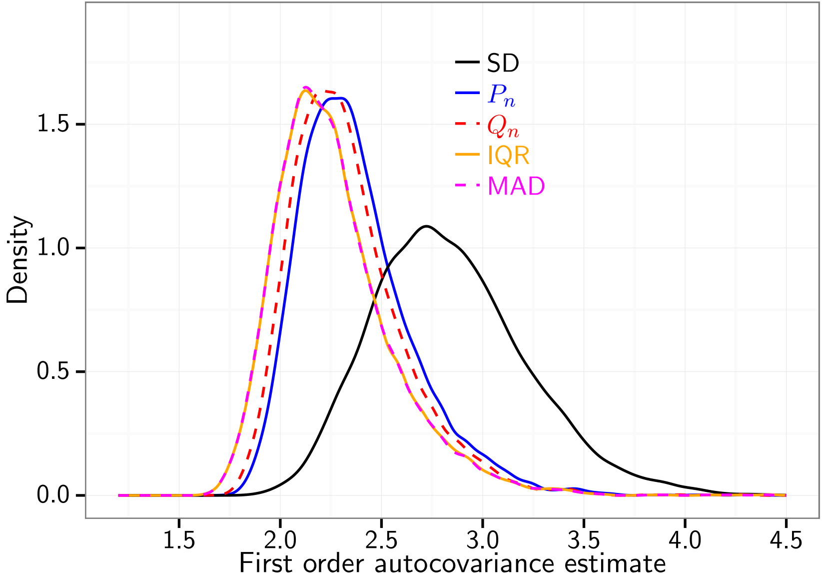
Empirical densities $\hat{\gamma}_\bullet(1)$ ARFIMA(0, 0.45, 0); $D = 0.1$; $n = 5000$



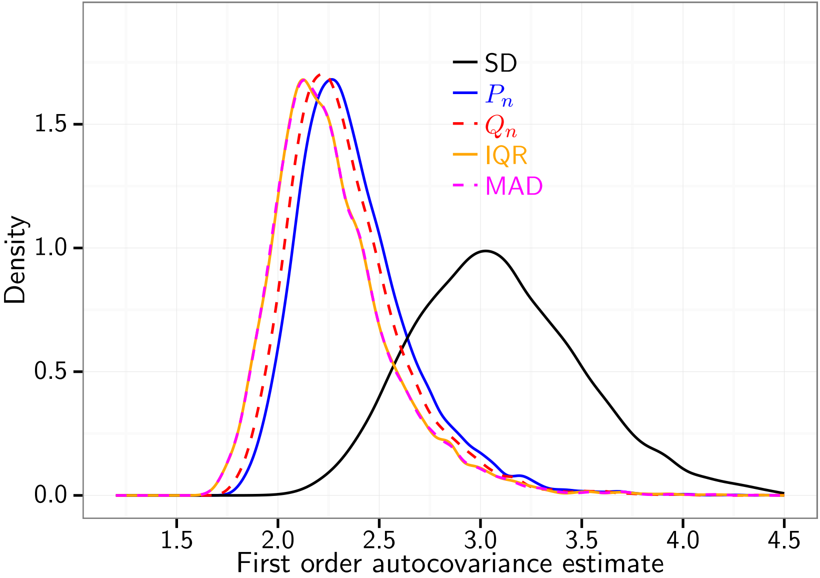
Empirical densities $\hat{\gamma}_{\bullet}(1)$ with 2% contamination at 5



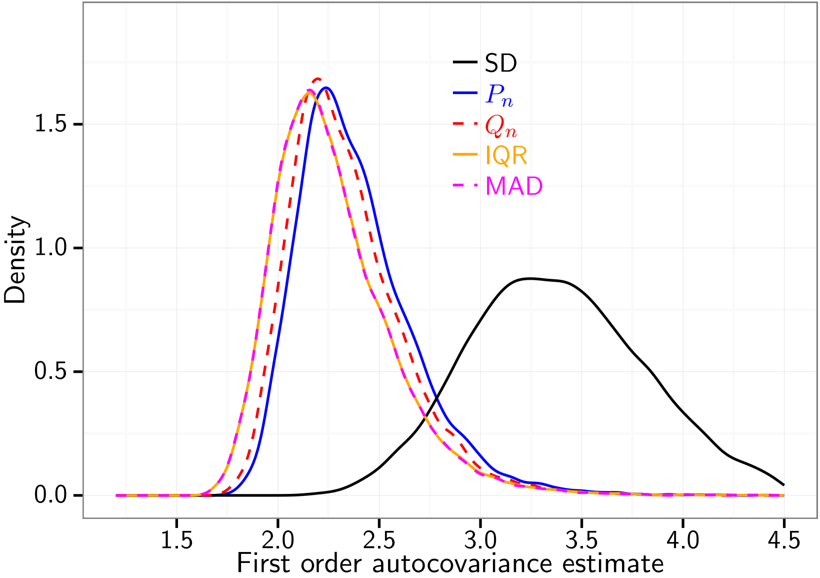
Empirical densities $\hat{\gamma}_\bullet(1)$ with 2% contamination at 6



Empirical densities $\hat{\gamma}_{\bullet}(1)$ with 2% contamination at 7



Empirical densities $\hat{\gamma}_\bullet(1)$ with 2% contamination at 8



Outline

The robust scale estimator P_n

Autocovariance estimation using P_n

Conclusion and key references

Summary

1. Aim

- Efficient and robust scale and autocovariance estimation.

Summary

1. Aim

- Efficient and robust scale and autocovariance estimation.

2. Method

- P_n scale estimator transformed using the GK identity.

Summary

1. Aim

- Efficient and robust scale and autocovariance estimation.

2. Method

- P_n scale estimator transformed using the GK identity.

3. Results

- 86% asymptotic efficiency at the Gaussian and high asymptotic efficiency at heavier tailed distributions.
- Robustness properties transfer from P_n to γ_P.
- In certain LRD settings robust estimators have very high efficiencies relative to the standard deviation.

References

Genton, M. and Ma, Y. (1999).
Robustness properties of dispersion estimators.
Statistics & Probability Letters, 44(4):343–350.

Gnanadesikan, R. and Kettenring J. R. (1972).
Robust estimates, residuals and outlier detection with multiresponse data
Biometrics, 28(1):81–124.

Hampel, F. (1974).
The influence curve and its role in robust estimation.
Journal of the American Statistical Association, 69(346):383–393.

Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M.S. and Reisen, V.A. (2011).
Robust estimation of the scale and of the autocovariance function of Gaussian short- and long-range dependent processes
Journal of Time Series Analysis, 32(2):135–156.

References

Ma, Y. and Genton, M. (2000).
Highly robust estimation of the autocovariance function
Journal of Time Series Analysis, 21(6):663–684.

Serfling, R. J. (1984).
Generalized L-, M-, and R-statistics.
The Annals of Statistics, 12(1):76–86.

Tarr, G., Müller, S. and Weber, N.C., (2012).
A robust scale estimator based on pairwise means.
Journal of Nonparametric Statistics, 24(1):187–199.

Wu, W.B., (2005).
On the Bahadur representation of sample quantiles for dependent sequences.
Annals of Statistics, 33(4):1934–1963.