Robust covariance estimation with P_n

Garth Tarr, Samuel Müller and Neville Weber

School of Mathematics and Statistics
THE UNIVERSITY OF SYDNEY

Outline

The robust scale estimator P_n

Covariance estimation using P_n

Autocovariance estimation using P_n

Conclusion and key references

Outline

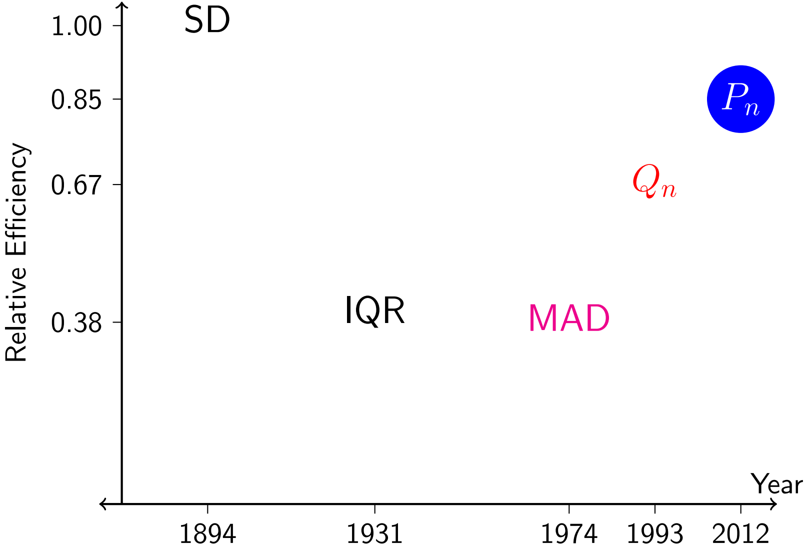
The robust scale estimator P_n

Covariance estimation using P_n

Autocovariance estimation using P_n

Conclusion and key references

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

$$\mathrm{IC}(x;T,F)=\lim_{\epsilon\downarrow 0}\frac{T((1-\epsilon)F+\epsilon\delta_x)-T(F)}{\epsilon}$$

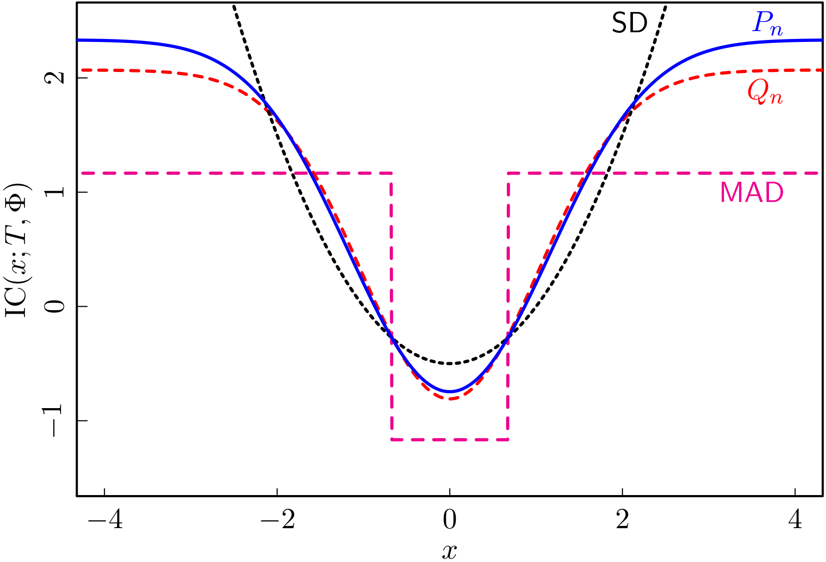
  where δ_x has all its mass at x.
- Serfling (1984) outlines the IC for GL-statistics.

Influence curve 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 $\epsilon>0$,

$$\mathrm{IC}(x;P_n,F)=c\left[\frac{0.75-F(2H_F^{-1}(0.75)-x)}{\int f(2H_F^{-1}(0.75)-x)f(x)\,\mathrm{d}\,x}-\frac{0.25-F(2H_F^{-1}(0.25)-x)}{\int f(2H_F^{-1}(0.25)-x)f(x)\,\mathrm{d}\,x}\right].$$

Influence curves when $F = \Phi$



Asymptotic variance and relative efficiency

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

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

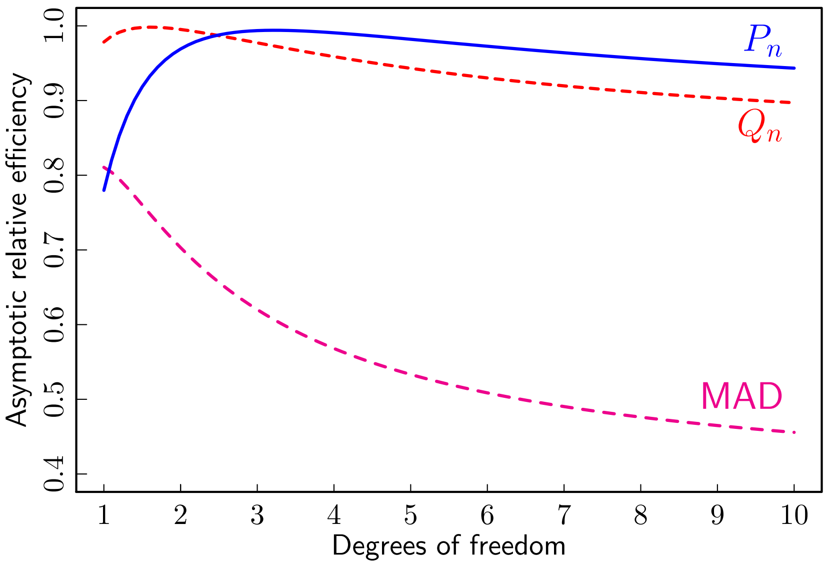
- This equates to an asymptotic efficiency of **0.86** as compared with **0.82** for Q_n and **0.37** for the MAD.

Result

We have a robust scale estimator that is asymptotically more efficient than Q_n at the Gaussian.

But how does it compare at heavier tailed distributions?

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



Outline

From scale to covariance: the GK device

- Gnanadesikan and Kettenring (1972) turned scale into covariance using the **location free** identity

$$\mathrm{cov}(X,Y) = \frac{1}{4\alpha\beta}\left[\mathrm{var}(\alpha X + \beta Y) - \mathrm{var}(\alpha X - \beta Y)\right], \quad (1)$$

where X, Y is a pair of random variables.

- In general, X and Y can have different units, so we set $\alpha = 1/\sqrt{\mathrm{var}(X)}$ and $\beta = 1/\sqrt{\mathrm{var}(Y)}$.
- Replacing **variance** with $\boldsymbol{P_n^2}$ we can similarly construct,

$$\gamma_P(X,Y) = \frac{1}{4\alpha\beta}\left[P_n^2(\alpha X + \beta Y) - P_n^2(\alpha X - \beta Y)\right], \quad (2)$$

where $\alpha = 1/P_n(X)$ and $\beta = 1/P_n(Y)$.

From scale to correlation: the GK device

- Correlation can also be estimated using the identity

$$\operatorname{corr}(X,Y)=\frac{\operatorname{var}(\alpha X+\beta Y)-\operatorname{var}(\alpha X-\beta Y)}{4\alpha\beta\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}}.$$

- Following what Ma and Genton (2001) did with Q_n, to ensure that $|\rho_P|\leq 1$, we use

$$\rho_P(X,Y)=\frac{P_n^2(\alpha X+\beta Y)-P_n^2(\alpha X-\beta Y)}{P_n^2(\alpha X+\beta Y)+P_n^2(\alpha X-\beta Y)}$$

- Covariance (and correlation) matrices can also be constructed on a componentwise basis.
- The same technique can be used to turn other scale estimators into covariance and correlation estimators, e.g. Ma and Genton (2001) study γ_Q and ρ_Q based on Q_n.

Robustness properties

Breakdown value

- P_n can breakdown if 25% of the pairwise means are contaminated.
- This can occur when 13.4% of observations are contaminated.
- γ_P and ρ_P inherit the **13.4%** breakdown value from P_n.
- Optionally, adaptive trimming $\implies$ 50% breakdown value.

Influence curve

- Following Genton and Ma (1999), we have shown that **influence curve** and therefore the **gross error sensitivity** of γ_P and ρ_P can be derived from the IC of P_n.
- Key point: the IC is **bounded** and hence the gross error sensitivity is **finite**.

Asymptotic efficiency

- The asymptotic variance of $\hat{\gamma}_P$ and $\hat{\rho}_P$ can be found as the expected square of the influence function.
- If X, Y are bivariate Gaussian:

$$V(\gamma_P, \Phi) = 2V(P_n, \Phi)(\sigma_X^2 \sigma_Y^2 + \gamma^2)$$

and

$$V(\rho_P, \Phi) = 2V(P_n, \Phi)(1 - \rho^2)^2.$$

Result

- At the bivariate Gaussian, $\hat{\gamma}_P$ and $\hat{\rho}_p$ maintain their asymptotic efficiency of **86%** relative to the classical covariance and correlation estimators.
- Compared with 82% for $\hat{\gamma}_Q$ and $\hat{\rho}_Q$ or 37% for $\hat{\gamma}_{\text{MAD}}$ and $\hat{\rho}_{\text{MAD}}$

Outline

The robust scale estimator P_n

Covariance estimation using P_n

Autocovariance estimation using P_n

Conclusion and key references

From covariance to autocovariance

Consider the identity,

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

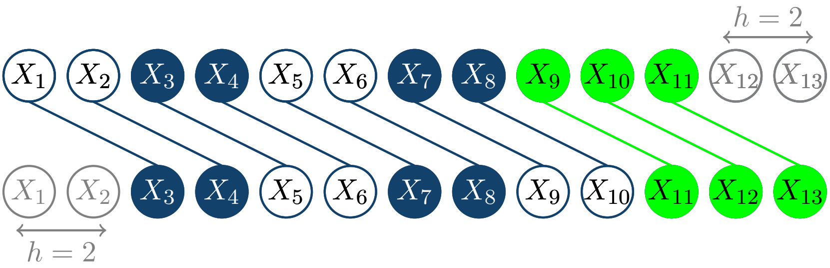
Define, for a weakly stationary series of n observations, $\mathbf{X}_n = \{X_t\}_{1 \leq t \leq n}$,

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

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.

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 ●



Asymptotic distribution under short range dependence

Let $\{X_t\}_{t\geq 1}$ be a stationary mean-zero Gaussian process with autocovariance sequence $\gamma(h) = \mathbb{E}(X_1 X_{h+1})$ satisfying $\sum_{h\geq 1} |\gamma(h)| < \infty$. Then,

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

where

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

Asymptotic distribution under long range dependence

Let $\{X_t\}_{t\geq 1}$ be a stationary mean-zero Gaussian process with autocovariance sequence $\gamma(h) = \mathbb{E}(X_1 X_{h+1})$ satisfying

$$\gamma(h) = h^{-D} L(h), \quad 0 < D < 1,$$

where L is slowly varying at infinity and is positive for large h.

- $D > 1/2$:

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

- $D < 1/2$ (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 fractional Brownian motion process, $Z_{2,D}$ is the Rosenblatt process and

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

Outline

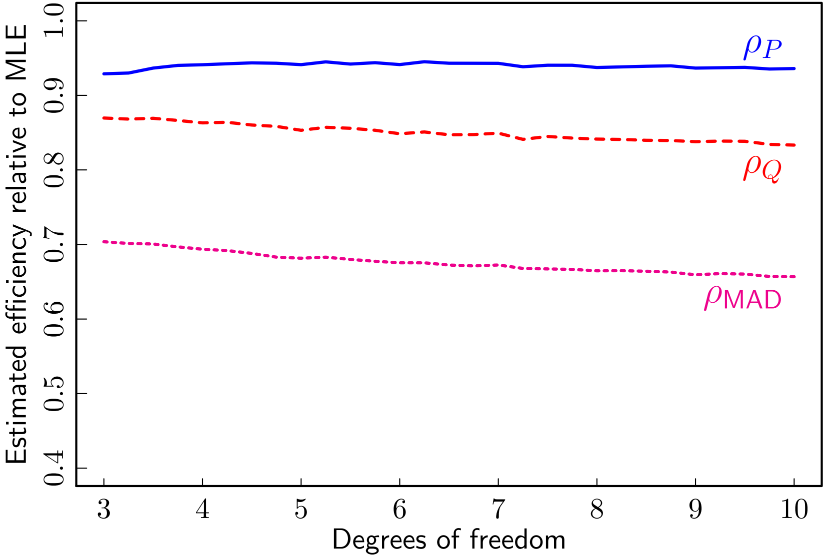
The robust scale estimator P_n

Covariance estimation using P_n

Autocovariance estimation using P_n

Conclusion and key references

Bivariate t_ν for $\nu \in [3, 10]$, $n = 20$ and $\rho = 0.5$.



Summary

1. Aim

- Efficient, robust and widely applicable scale, covariance and correlation estimators.

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.
- High relative efficiency in finite samples.
- Robustness properties transfer from P_n to γ_P and ρ_P.

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.

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

References

Ma, Y. and Genton, M. (2001).
Highly robust estimation of dispersion matrices
Journal of Multivariate Analysis, 78(1):11–36.

Rousseeuw, P. and Croux, C. (1993).
Alternatives to the median absolute deviation.
Journal of the American Statistical Association, 88(424):1273–1283.

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.

Tarr, G., Müller, S. and Weber, N.C., (Preprint).
Asymptotics of the P_n scale estimator under Gaussian long range dependence.