robts - an R-package for robust time series and change-point analysis

Alexander Dürre, Roland Fried

with contributions by Tobias Liboschik, Jonathan Rathjens

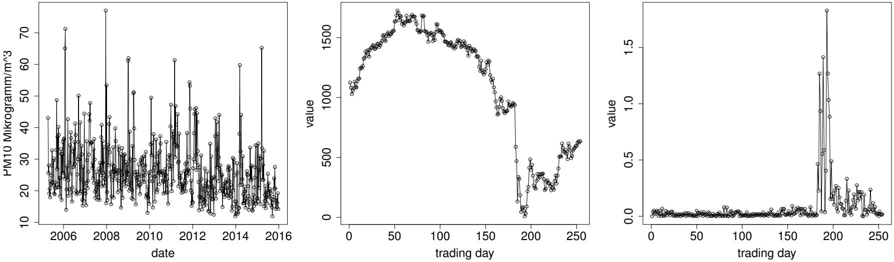
technische universität dortmund

SFB 823

Time series

Time series $\vec{y} = (y_1, \ldots, y_n)'$: real-valued measurements of the same variable observed at equidistant observation times, realization of a random vector $\vec{Y} = (Y_1, \ldots, Y_n)'$.

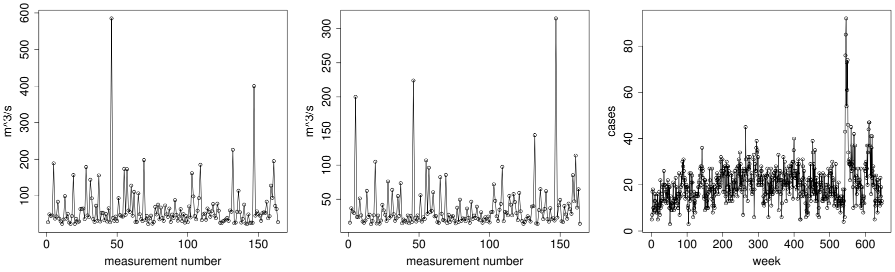
Some examples:
Weekly fine dust averages in Essen (Germany) from 04/2005 till 12/2015 (left).
VW daily opening prices (middle) and their absolute log returns (absolute differences between subsequent log-values, right) in 2015.



Time series

Time series $\vec{y} = (y_1, \ldots, y_n)'$: real-valued measurements of the same variable observed at equidistant observation times, realization of a random vector $(Y_1, \ldots, Y_n)'$.

Further examples:
Large flood events at stations Aue (left) and Niederschlema (middle), river Mulde, Germany.
Weekly number of reported disease cases caused by escherichia coli in the state of NRW (Germany) from 01/2001 to 05/2013 (right).



Outline of the presentation

- $\mathbb{I}$ Estimation of (partial) autocorrelations
  (analysis of linear relationships)
- $\mathbb{II}$ Fitting autoregressive models
  (basic time series modeling)
- $\mathbb{III}$ Estimation of spectral densities
  (analysis of cyclic behavior)
- $\mathbb{IV}$ Change-point detection
  (tests for the basic assumptions)

\+ Simulation results
\+ Applications

$\mathbb{I}$: Robust autocorrelation estimation

Dürre, Fried & Liboschik (2015): Robust estimation of (partial) autocorrelation, *WIREs Comput. Stat.* 7: 205-222.

$\mathbb{I}$: Second order stationarity

$(Y_t : t \in \mathbb{Z})$ stochastic process observed at time points $t = 1, \ldots, n$.

Basic assumption: **second order stationarity** of $(Y_t : t \in \mathbb{Z})$, i.e.,

$$E(Y_t) \equiv \mu \quad \text{and} \quad Var(Y_t) \equiv \gamma(0), \quad t \in \mathbb{Z},$$

and **autocovariances** depending only on time lag h,

$$Cov(Y_t, Y_{t+h}) = \gamma(h), \qquad t, h \in \mathbb{Z}.$$

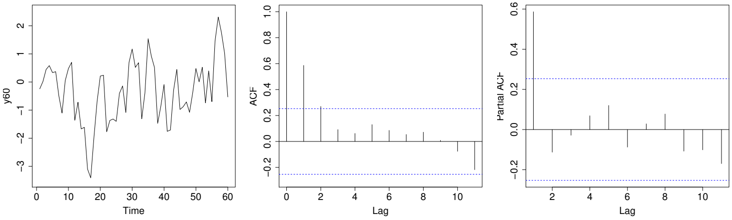
Under these assumptions: **autocorrelations** $\rho(h) = \dfrac{\gamma(h)}{\gamma(0)}$, $h \in \mathbb{N}_0$, measure linear dependence between Y_t and Y_{t+h}, $t, h \in \mathbb{Z}$.
Partial autocorrelations $\pi(h)$, $h \in \mathbb{N}$, measure correlation of Y_t and Y_{t+h} after eliminating linear effects of $Y_{t+1}, \ldots, Y_{t+h-1}$ by regression.

$\mathbb{I}$: Classical estimation

$\vec{y} = (y_1, \ldots, y_n)'$ real-valued time series, realization of $(Y_1, \ldots, Y_n)'$

$$
\begin{aligned}
\hat{\mu} &= \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t, \\
\hat{\gamma}(h) &= \frac{1}{n}\sum_{t=1}^{n-h}(y_t - \bar{y})(y_{t+h} - \bar{y}), \quad h \in \mathbb{N}_0 \\
\hat{\rho}(h) &= \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{1/n\sum_{t=1}^{n-h}(y_t - \bar{y})(y_{t+h} - \bar{y})}{1/n\sum_{t=1}^{n}(y_t - \bar{y})(y_t - \bar{y})}, \quad h \in \mathbb{N}_0.
\end{aligned}
$$

Time series, autocorrelations and partial autocorrelations



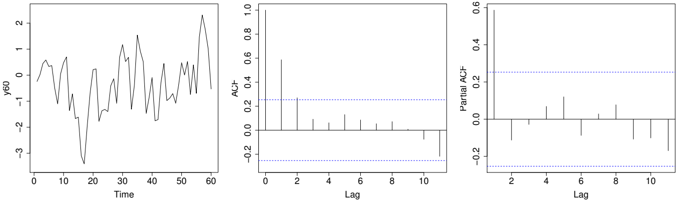
Positive autocorrelations decaying to 0 with increasing time lag. Partial autocorrelations suggest 1st order autoregression: only Y_{t-1} affects Y_t directly.

Interpretation is facilitated by asymptotical confidence bounds calculated assuming all autocorrelations to be zero (white noise process).

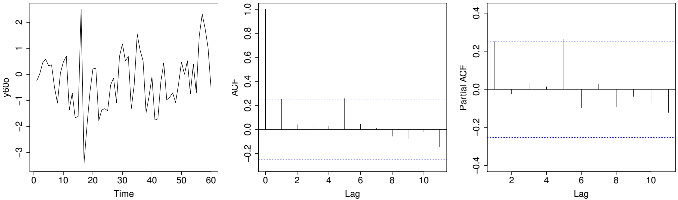
R-functions **acf** and **pacf**, or **acf** with type="partial"

𝕀 : Classical estimation

Time series, autocorrelations and partial autocorrelations



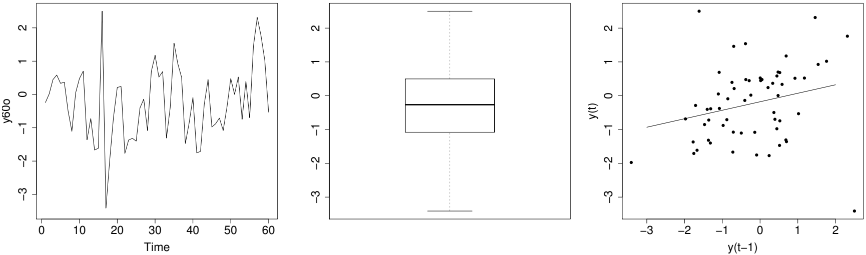
Time series with 1 outlier, autocorrelations and partial autocorrelations



1 moderately large outlier can seriously affect classical estimators

𝕀 : Classical outlier detection

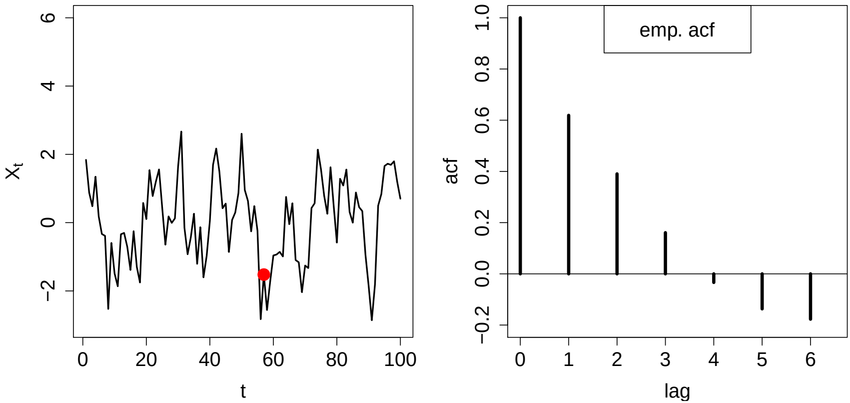
Time series, boxplot of y_t and scatterplot of (y_{t-1}, y_t):



Outlier detection w.r.t. marginal distribution may fail.
Outlier detection in time series needs good estimates
but even moderately large outliers can seriously affect classical estimators.
Note: one outlier affects two pairs (y_t, y_{t+h}) used for estimation of $\rho(h)$.

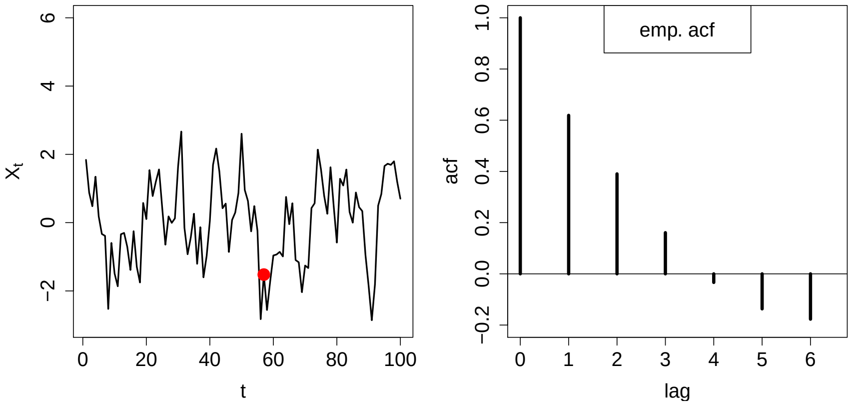
$\mathbb{I}$:Additive isolated outlier

Example: AR(1) process (X_t) with $\phi_1 = 0.6$ and one outlier $\tilde{X}_{t_0} = X_{t_0} + \text{out}$



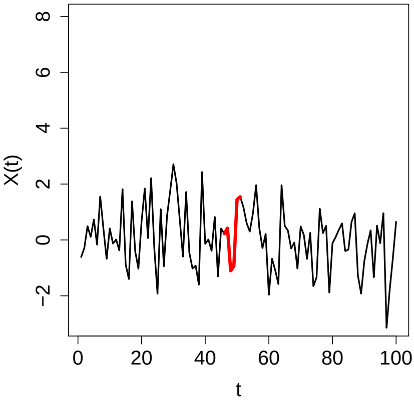
Additive isolated outlier

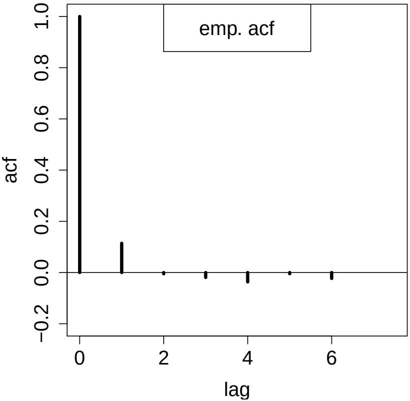
Example: AR(1) process (X_t) with $\phi_1 = 0.6$ and one outlier $\tilde{X}_{t_0} = X_{t_0} + \text{out}$



$\mathbb{I}$: Additive block outliers

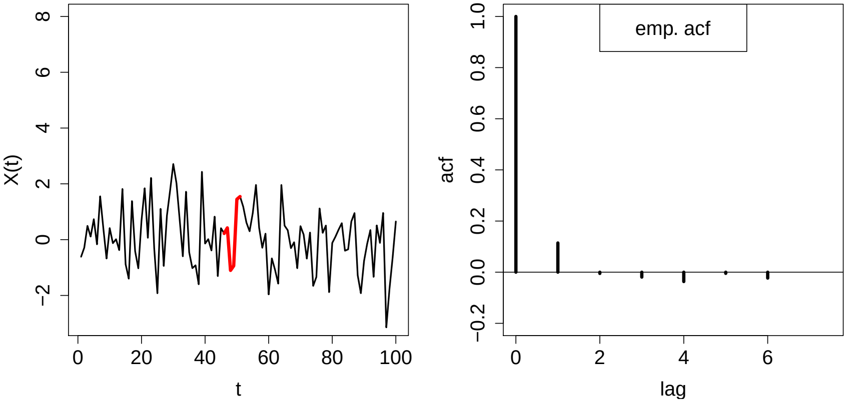
Example: White noise with block outlier of length 4, $\tilde{X}_{[t_0:t_0+k]} = X_{[t_0:t_0+k]} + \text{out}$



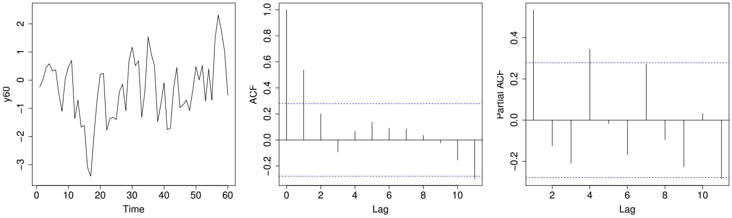


Additive block outliers

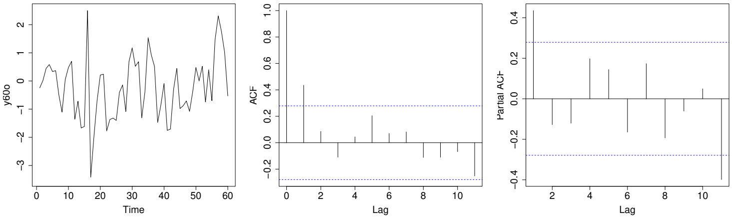
Example: White noise with block outlier of length 4, $\tilde{X}_{[t_0:t_0+k]} = X_{[t_0:t_0+k]} + \text{out}$



Time series, autocorrelations and partial autocorrelations



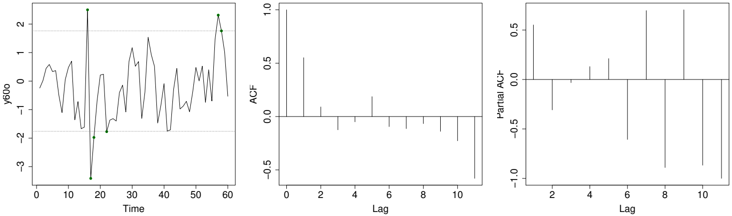
Time series with 1 outlier, autocorrelations and partial autocorrelations



$\mathbb{I}$: Trimming based estimators

acfrob with approach=trim: **trimmed estimation** (Chan & Wei 1992)
replace means in classical estimator by trimmed means
"limiting case": **median estimation** (Chakhchoukh 2010)

Time series with 1 outlier, acf and pacf with $\alpha = 5\%$ trimming



acfrob(data,approach="trim",trim=0.05)
Mildly robust for α close to 0, not very efficient for α close to 0.5.

Separate estimators estimate each $\rho(h)$ individually.
Default choice: approach=GK **Gnanadesikan-Kettenring estimator**
Gnanadesikan & Kettenring (1972), Ma & Genton (2000)

$$\frac{\mathrm{Var}(X+Y)-\mathrm{Var}(X-Y)}{\mathrm{Var}(X+Y)+\mathrm{Var}(X-Y)} = \frac{4\mathrm{Cov}(X,Y)}{2\mathrm{Var}(X)+2\mathrm{Var}(Y)} \overset{\mathrm{Var}(X)=\mathrm{Var}(Y)}{=} \mathrm{Cor}(X,Y)$$

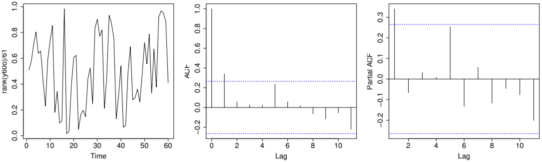
Apply robust variance estimator like Q_n^2 (Rousseeuw & Croux 1993),
$\hat{\sigma}(Z_1,\ldots,Z_n) \approx 2.219 \cdot q_{1/4}(|Z_i - Z_j|, 1 \le i < j \le n)$
Rather efficient in large samples, resists several isolated outliers well.

Separate estimators are quick but possibly not positive-semidefinite.
Positive-semidefiniteness can be enforced by iterative projection algorithm (Al-Homidan 2006) setting psd=TRUE.

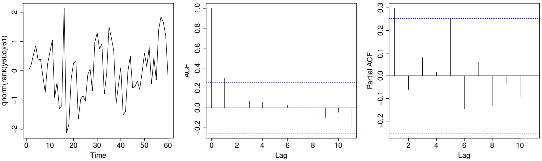
𝕀 : Transformation estimators

Transform data robustly, apply classical estimator to transformed data.

approach=rank: **rank correlation**, transform data using ranks
cor.method=gaussian, spearman, kendall, quadrant
recommendable for integer observations
Ranks of time series with 1 outlier, autocor. and partial autocor.



Gaussian rank transform of data with 1 outlier, autocor. and part. autoc.



$\mathbb{I}$: Further autocorrelation estimators

approach=RA: **residual autocovariances** (Bustos & Yohai 1986)
transform data by Huber's (default) or Tukey's ψ-function

approach=filter: transform by **robust filter** (Maronna, Martin & Yohai 2006)

approach=bireg: **separate regressions** of y_{t+h} on y_t (Chang & Politis 2016)

approach=multi: joint estimation from **multivariate covariance matrix**;
estimate covariance matrix by raw or reweighted MCD (default), Stahel-Donoho, S, M or Tyler covariance matrix.
Works well for patchy outliers occurring in blocks but time-consuming.

approach=partrank: **partial autocorrelation estimator** (Masarotto 1987)
estimate partial autocorrelations from forward and backward residuals using Spearman, Kendall, quadrant or Gaussian rank correlation .
Works well if partial autocorrelations main interest.

$\mathbb{I}$: Estimation from multivariate correlations

$$
\mathbb{X} = \underbrace{\begin{pmatrix}
X_1 & 0 & \dots & 0 \\
\vdots & X_1 & \ddots & \vdots \\
\vdots & \vdots & \ddots & 0 \\
\vdots & \vdots & & X_1 \\
X_n & \vdots & & \vdots \\
0 & X_n & & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
0 & \dots & 0 & X_n
\end{pmatrix}}_{p+1}, \quad
\hat{\Xi}(\mathbb{X}) = \begin{pmatrix}
1 & \xi_{1,2} & \dots & \xi_{1,p+1} \\
\xi_{2,1} & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \xi_{p,p+1} \\
\xi_{p+1,1} & \dots & \xi_{p+1,p} & 1
\end{pmatrix}
$$

where $\hat{\Xi}$ robust multivariate correlation estimator

autocorrelation estimation at lag h

$$
\hat{\rho}(h) = \frac{1}{p-h+1}(\xi_{1,h+1} + \ldots + \xi_{p+1-h,p+1})
$$

$\mathbb{I}$: Efficient MCD (Gervini 2003)

calculate robust initial estimators $\hat{\mu}$ and $\hat{\Sigma}$ by MCD
calculate robust Mahalanobis distances r_i

$$r_i = (X_i - \hat{\mu})\hat{\Sigma}^{-1}(X_i - \hat{\mu})$$

identify obs. with b largest Mahalanobis dist. as outliers

$$b = \lceil n \cdot \sup_{x \geq q}(\hat{F}_r(x) - F_r(x)) \rceil$$

calculate emp. correlation from the remaining values

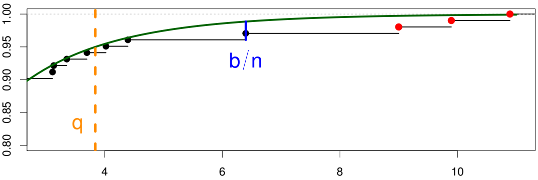


Figure: Example of an empirical distribution function of Mahalanobis distances

$\mathbb{I}$: Available estimators

- emp. acf
- use of transformed values
  - Huber (Duerre et al. 2017)
  - Tukey
  - filter acf
- bivariate estimators
  - Spearman-ρ
  - Kendalls-τ
  - Gaussian rank cor.
  - quandrant cor.
- use of partial correlation
  - partial Masarotto
  - partial Gaussian rank cor.
  - partial Spearmans-ρ
  - partial Kendalls-τ
  - partial quadrant cor.
- GK approach
  - GK-Qn
  - GK-Tau
  - GK-MAD
  - efficient GK
- multivariate methods
  - multi. raw MCD
  - multi. weighted MCD
  - multi. efficient MCD
  - multi. M-estimator
  - multi. S-estimator
  - multi. Tyler-shape
- regression methods
  - MM regression
  - lts regression
  - median regression

$\mathbb{I}$: Available estimators

emp. acf

use of transformed values
- Huber
- Tukey
- filter acf

bivariate estimators
- Spearman-ρ
- Kendalls-τ
- Gaussian rank cor.
- quandrant cor.

use of partial correlation
- partial Masarotto
- partial Gaussian rank cor.
- partial Spearmans-ρ
- partial Kendalls-τ
- partial quadrant cor.

GK approach
- GK-Qn
- GK-Tau
- GK-MAD
- efficient GK

multivariate methods
- multi. raw MCD
- multi. weighted MCD
- multi. efficient MCD
- multi. M-estimator
- multi. S-estimator
- multi. Tyler-shape

regression methods
- MM regression
- lts regression
- median regression

𝕀 : Simulation: one isolated outlier

Gaussian AR(1) with $\phi_1 = 0.8$, $n = 100$
one additive outlier of increasing size, 100 runs

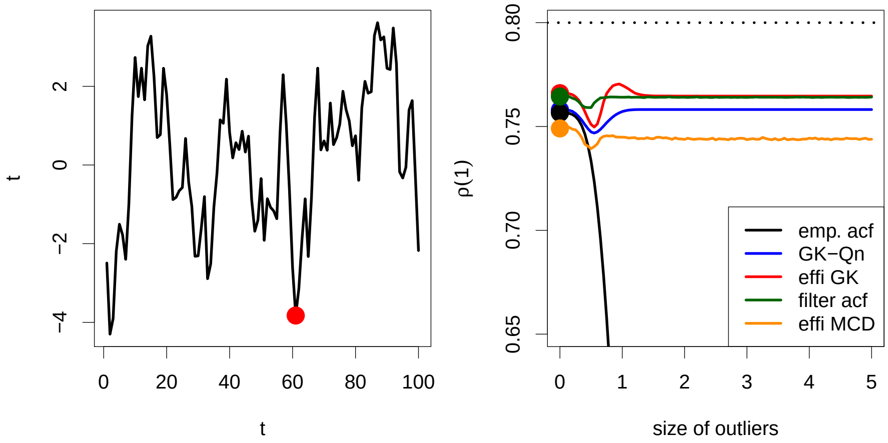


Figure: Example series (left) and average $\hat{\rho}(1)$ depending on outlier size (right)

Gaussian AR(1) with $\phi_1 = 0.8$, $n = 100$
one additive outlier of increasing size, 100 runs

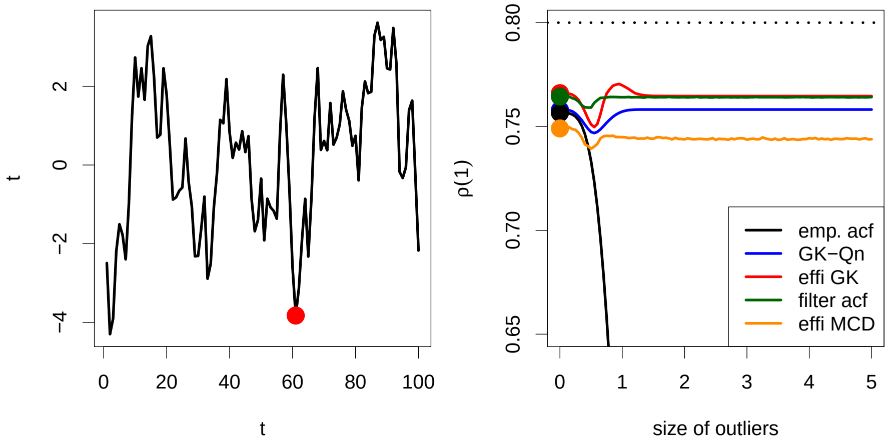


Figure: Example series (left) and average $\hat{\rho}(1)$ depending on outlier size (right)

$\mathbb{I}$: Simulation: many isolated outliers

Gaussian AR(1) with $\phi_1 = 0.8$, $n = 100$
increasing number of outliers, 100 runs

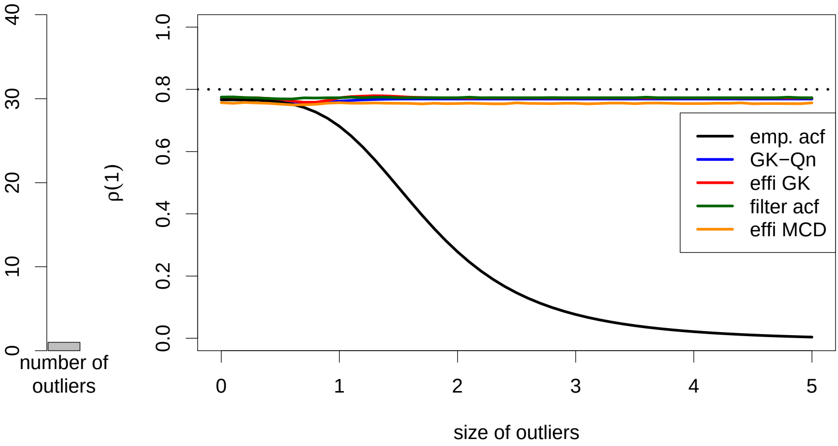


Figure: Average $\rho(1)$ depending on outlier size

$\mathbb{I}$: Simulation: many isolated outliers

Gaussian AR(1) with $\phi_1 = 0.8$, $n = 100$
increasing number of outliers, 100 runs

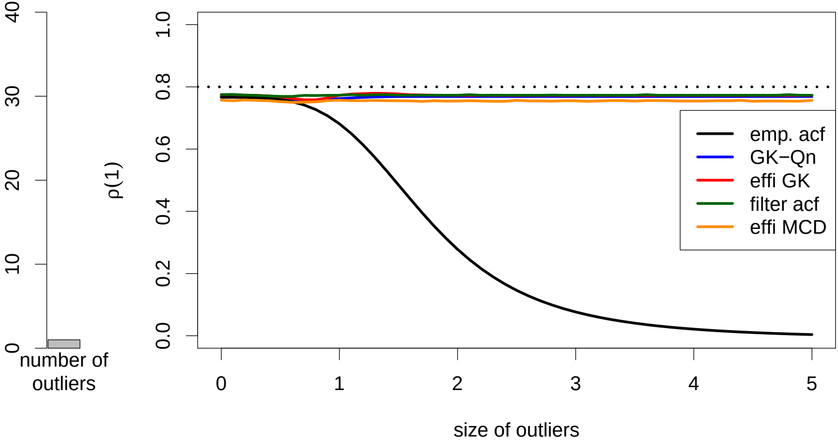


Figure: Average $\rho(1)$ depending on outlier size

𝕀 : Simulation: many isolated outliers

Gaussian AR(1) with $\phi_1 = 0.8$, $n = 100$
increasing number of outliers, 100 runs

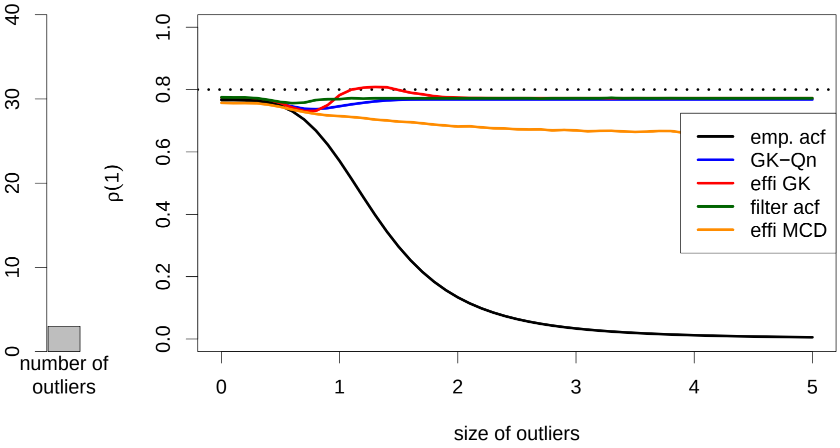


Figure: Average $\rho(1)$ depending on outlier size

𝕀 : Simulation: block of 2 outliers

Gaussian white noise, $n = 100$
increasing outlier size, 100 runs

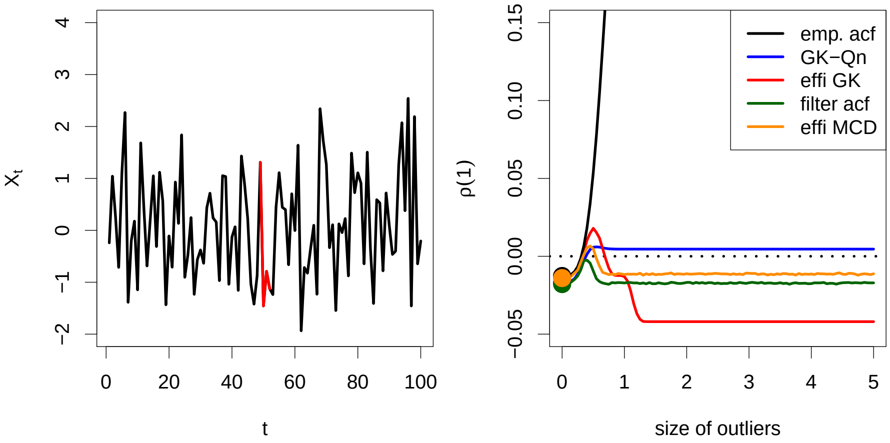


Figure: Example series (left) and average $\hat{\rho}(1)$ depending on outlier size (right)

$\mathbb{I}$: Simulation: block of 2 outliers

Gaussian white noise, $n = 100$
increasing outlier size, 100 runs

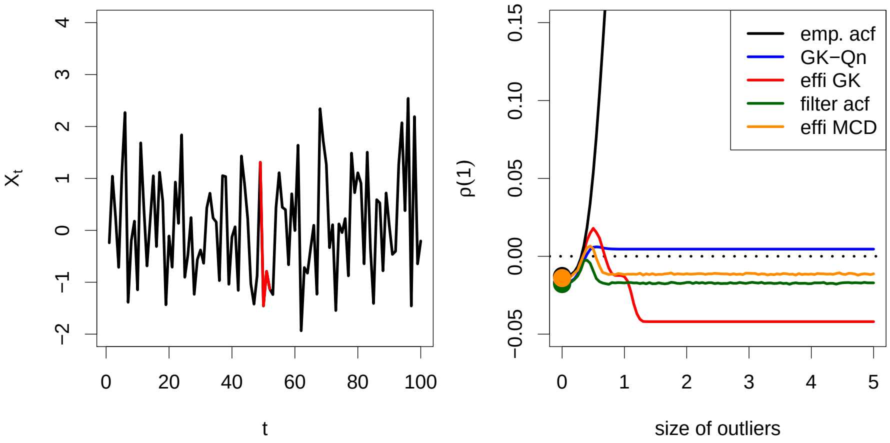


Figure: Example series (left) and average $\hat{\rho}(1)$ depending on outlier size (right)

$\mathbb{I}$: Simulation: block outlier of increasing length

Gaussian white noise, $n = 100$
increasing number of outliers, 100 runs

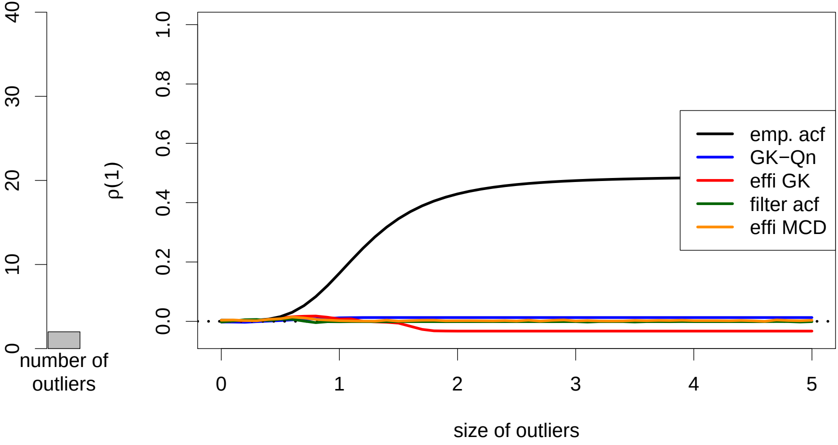


Figure: Average $\hat{\rho}(1)$ depending on outlier size (right)

𝕀 : Simulation: block outlier of increasing length

Gaussian white noise, $n = 100$
increasing number of outliers, 100 runs

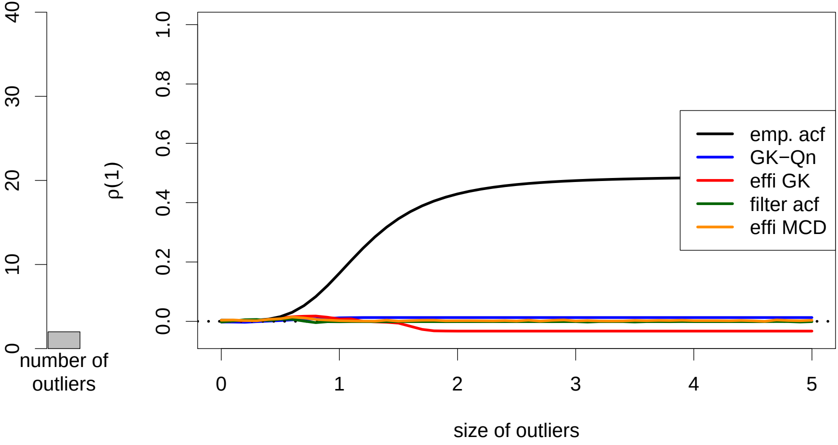


Figure: Average $\hat{\rho}(1)$ depending on outlier size (right)

$\mathbb{I}$: Simulation: block outlier of increasing length

Gaussian white noise, $n = 100$
increasing number of outliers, 100 runs

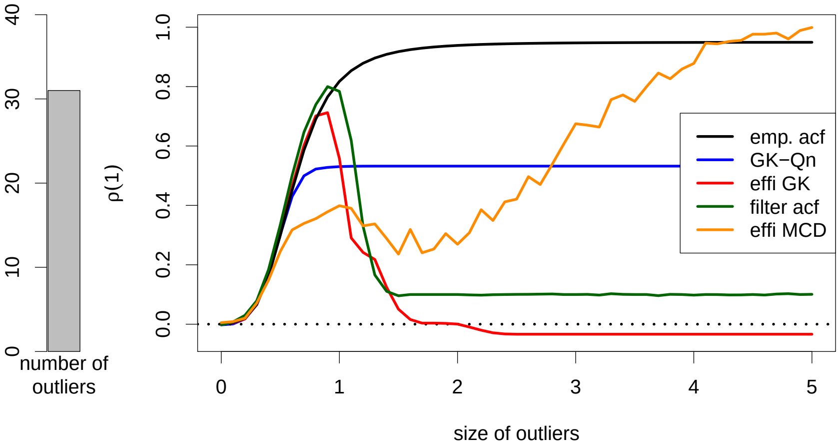
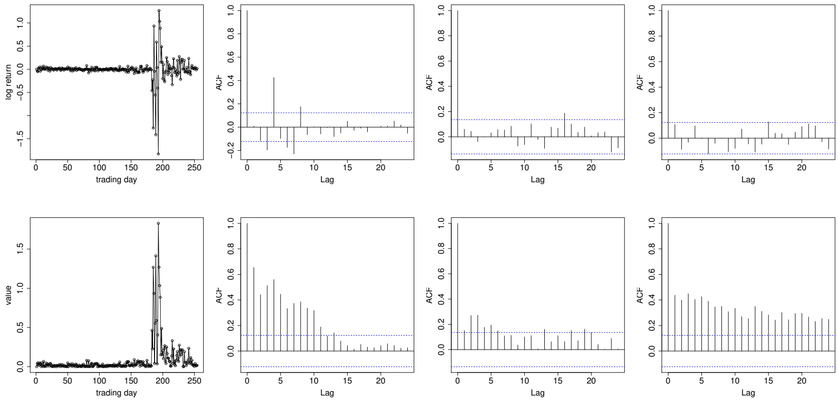


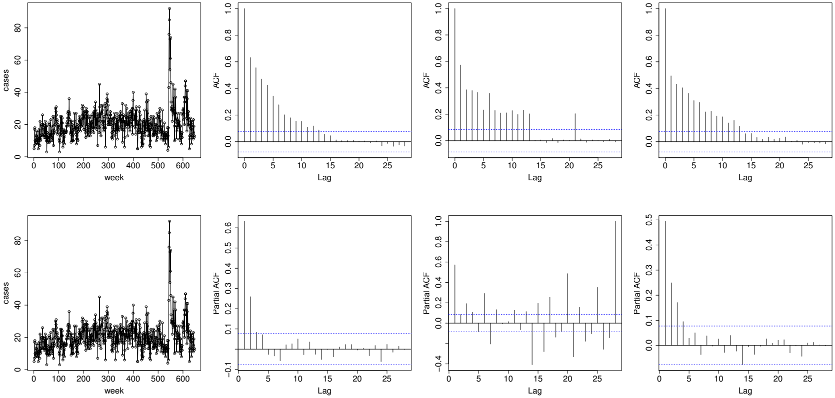
Figure: Average $\hat{\rho}(1)$ depending on outlier size (right)

VW log returns (top) and absolute log returns (bottom) in 2015: autocorrelation (left) and robust autocorrelation estimates by GK (center) and ranks (right).



I : Analysis of escherichia coli cases

Escherichia coli cases: autocorrelation (top) and partial autocorrelation (bottom) estimates and robust estimates by GK (center) and ranks (right).



$\mathbb{III}$: Estimation of autoregressive models

Autocorrelation estimates reliable only for time lags $h << n$.
Autoregressive models constitute simple approach to estimation and prediction.

Dürre, Fried & Liboschik (2018+): Robust time series analysis: the R-package robts, *Work in progress*.

$\mathbb{III}$: Autoregressive model of order p

$(Y_t : t \in \mathbb{Z})$ second order stationary, observed at time points $t = 1, \ldots, n$.

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + e_t, \; t \in \mathbb{Z}$$

Condition **CS** for existence of causal stationary solution:
All roots of $1 - \phi_1 y - \ldots - \phi_p y^p = 0$ larger than 1 in absolute value.
For $p = 1$: $|\phi_1| < 1$.

Under these assumptions:

$$E(Y_t) = \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p}$$

Autocorrelations $\rho(h)$ decay to 0 exponentially fast for lags $h > p$.
Partial autocorrelations $\pi(h)$, $h \in \mathbb{N}$, are 0 for $h > p$.

$\mathbb{III}$: Classical estimation

$\vec{y} = (y_1, \ldots, y_n)'$ real-valued time series, realization of $(Y_1, \ldots, Y_n)'$

(Exact) Maximum likelihood (ML) estimation maximizes
$f(\vec{y}) = f(y_{p+1}, \ldots, y_n | y_1, \ldots, y_p) \cdot f(y_1, \ldots, y_p)$ numerically.

Conditional maximum likelihood (CML) maximizes $f(y_{p+1}, \ldots, y_n | y_1, \ldots, y_p)$.
Gaussian case: ordinary (conditional) least squares estimator in regression model.
Simpler than ML, same asymptotic efficiency for fixed p, but **CS** possibly not fulfilled.

Yule-Walker (YW) translates estimated autocorrelations $\rho(1), \ldots, \rho(p)$ into $\phi_1, \ldots, \phi_p$.
Simpler than ML, less efficient in small Gaussian samples, **CS** fulfilled.

In case of **ML** and **CML**, ϕ_0 can be estimated jointly with $\phi_1, \ldots, \phi_p$ or via $\hat{\mu} = \bar{y}$.

R-function **ar** or **arima**, method=c("mle","ols","yw")
For clean data: 0.578, 0.587, 0.587 instead of 0.7.
With 1 outlier: 0.247, 0.251, 0.251 instead of 0.7.

$\mathbb{III}$: Robust AR fitting

Regression estimators:
gm: generalized M-estimator from robust normal equations for successive estimation of $\phi_1, \ldots, \phi_p$ using Huber's ψ-function for residuals and Tukey's ψ-function for regressors.
regression: MM-estimator lmrob from robustbase based on Tukey's ψ-function
filter: minimize τ-scale $\frac{\sigma_0^2}{n-p} \sum_{t=p+1}^{n} min \left\{ \left(\frac{e_{t,h-1} - \phi_h \tilde{y}_{t-h}}{\sigma_0} \right)^2, u^2 \right\}, \; h = 1, \ldots, p$
with filtered regressors $\tilde{y}_{t-1}, \ldots, \tilde{y}_{t-p}$.

Yule-Walker (YW) estimation from estimated autocorrelations $\rho(1), \ldots, \rho(p)$.

arrob(method=c("gm","regression","filter","yw"), options locfn=, scalefn=)
For clean data: 0.332, 0.586, 0.452, 0.536 instead of 0.7.
With 1 outlier: 0.228, 0.499, 0.509, 0.435 instead of 0.7.

$\mathbb{III}$: Robust Yule-Walker estimation

Calculate AR coefficients $\phi_1, \ldots, \phi_p$ by solving Yule-Walker equations

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \ldots & \rho(p-1) \\ \rho(1) & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho(1) \\ \rho(p-1) & \ldots & \rho(1) & 1 \end{pmatrix} \cdot \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}$$

Calculate residuals $\epsilon_p^{(p)}, \ldots, \epsilon_n^{(p)}$

Choose model order p with smallest robust BIC

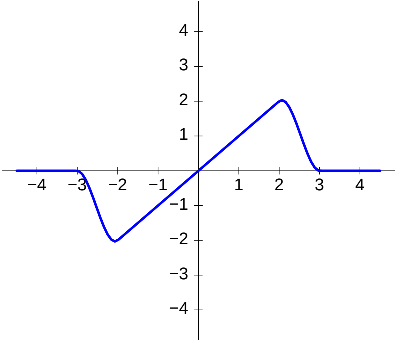
$$\log\left(\hat{\sigma}^2\left(\epsilon_p^{(p)}, \ldots, \epsilon_n^{(p)}\right)\right) + \frac{\log(n-p)2p}{n-p}$$

where $\hat{\sigma}$ robust scale estimator, e.g. Qn.

𝕀𝕀 : Robust filter, Masreliez (1975), Maronna et al. (2006)

For given AR model with $\phi_1, \ldots, \phi_p$ calculate iteratively:

prediction $\hat{X}_t$ based on filtered values $Y_{t-1}, \ldots Y_{t-p}$
residual $\epsilon_t = X_t - \hat{X}_t$
filtered value $Y_t = \hat{X}_t + \psi(\epsilon_t)$



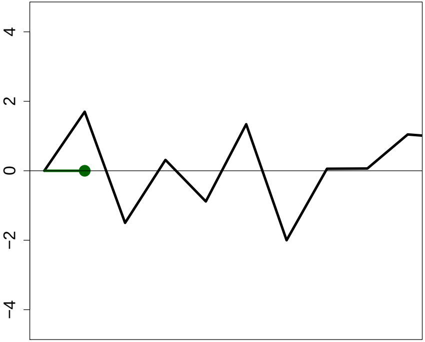


Figure: Possible Ψ−function (left) and filtered time series (right)

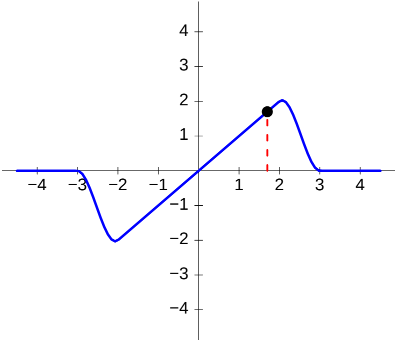
$\mathbb{III}$: Robust filter, Masreliez (1975), Maronna et al. (2006)

For given AR model with $\phi_1, \ldots, \phi_p$ calculate iteratively:

prediction $\hat{X}_t$ based on filtered values $Y_{t_1}, \ldots Y_{t-p}$

residual $\epsilon_t = X_t - \hat{X}_t$

filtered value $Y_t = \hat{X}_t + \psi(\epsilon_t)$



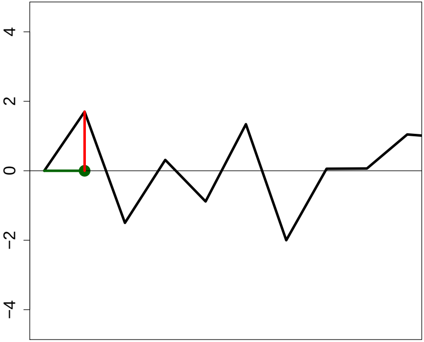


Figure: Possible Ψ−function (left) and filtered time series (right)

𝕀𝕀 : Robust filter, Masreliez (1975), Maronna et al. (2006)

For given AR model with $\phi_1, \ldots, \phi_p$ calculate iteratively:

prediction $\hat{X}_t$ based on filtered values $Y_{t-1}, \ldots Y_{t-p}$

residual $\epsilon_t = X_t - \hat{X}_t$

filtered value $Y_t = \hat{X}_t + \psi(\epsilon_t)$

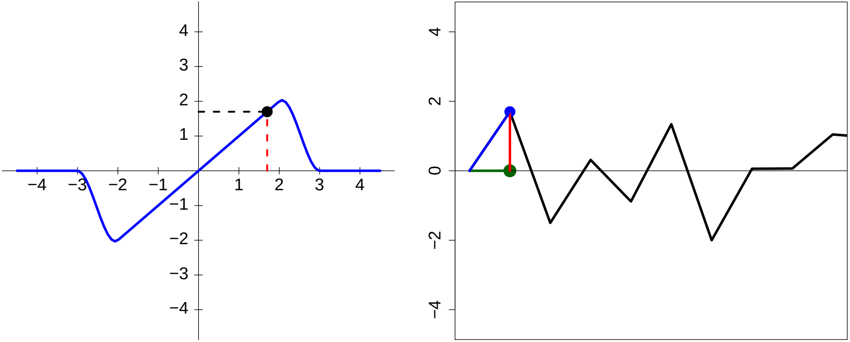


Figure: Possible Ψ−function (left) and filtered time series (right)

$\mathbb{III}$: Robust filter, Masreliez (1975), Maronna et al. (2006)

For given AR model with $\phi_1, \ldots, \phi_p$ calculate iteratively:

prediction $\hat{X}_t$ **based on filtered values** $Y_{t-1}, \ldots Y_{t-p}$

residual $\epsilon_t = X_t - \hat{X}_t$

filtered value $Y_t = \hat{X}_t + \psi(\epsilon_t)$

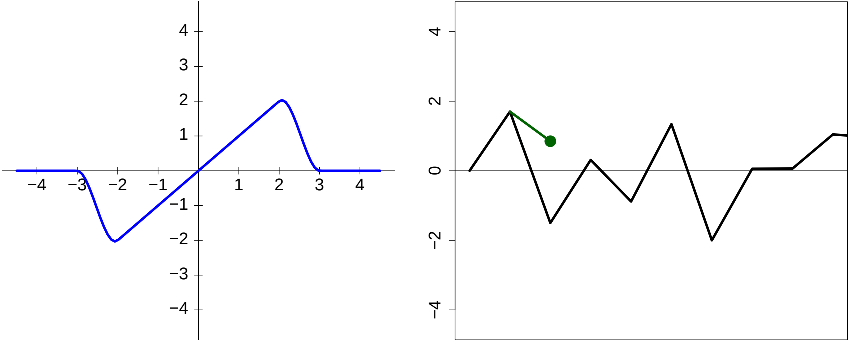


Figure: Possible $\Psi-$function (left) and filtered time series (right)

Robust filter, Masreliez (1975) and Maronna et al. (2006)

For given AR model with $\phi_1, \ldots, \phi_p$ calculate iteratively:

prediction $\hat{X}_t$ based on filtered values $Y_{t-1}, \ldots Y_{t-p}$

residual $\epsilon_t = X_t - \hat{X}_t$

filtered value $Y_t = \hat{X}_t + \psi(\epsilon_t)$

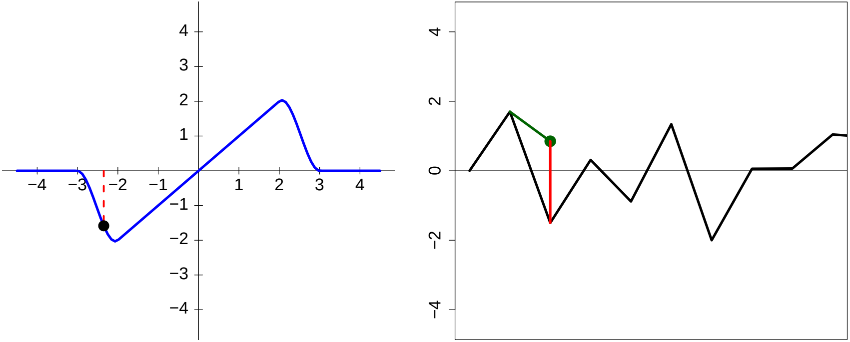


Figure: Possible Ψ−function (left) and filtered time series (right)

Robust filter, Masreliez (1975) and Maronna et al. (2006)

For given AR model with $\phi_1, \ldots, \phi_p$ calculate iteratively:

prediction $\hat{X}_t$ based on filtered values $Y_{t-1}, \ldots Y_{t-p}$

residual $\epsilon_t = X_t - \hat{X}_t$

filtered value $Y_t = \hat{X}_t + \psi(\epsilon_t)$

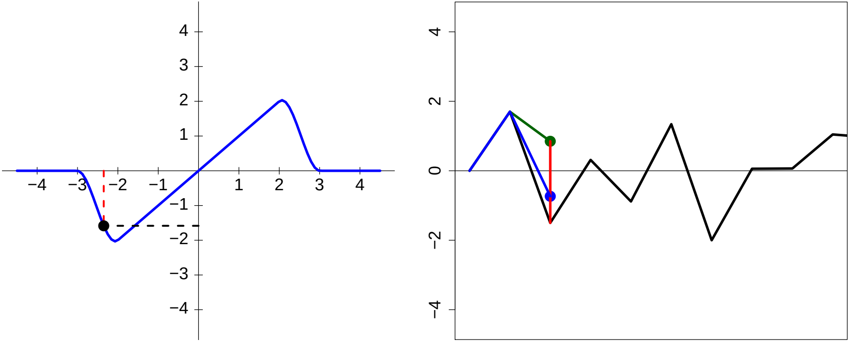


Figure: Possible Ψ−function (left) and filtered time series (right)

Robust filter, Masreliez (1975) and Maronna et al. (2006)

For given AR model with $\phi_1, \ldots, \phi_p$ calculate iteratively:

prediction $\hat{X}_t$ based on filtered values $Y_{t-1}, \ldots Y_{t-p}$

residual $\epsilon_t = X_t - \hat{X}_t$

filtered value $Y_t = \hat{X}_t + \psi(\epsilon_t)$

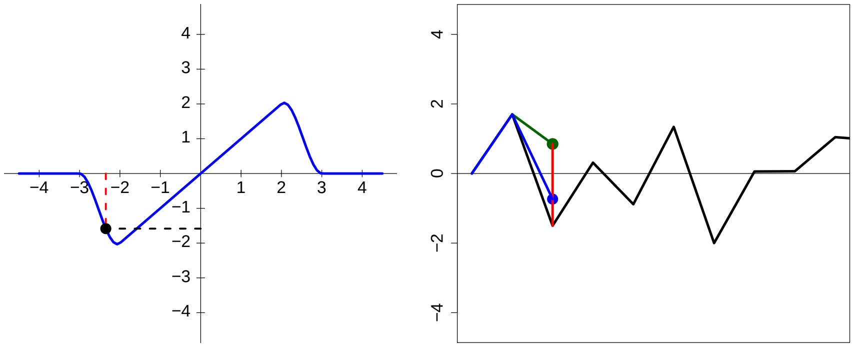


Figure: Possible Ψ−function (left) and filtered time series (right)

Ⅲ : Simulation: AR model estimation without outliers

random AR order: $p \sim Pois(\lambda = 2)$

random AR coefficients: $\phi_i \sim U[0,1] - U[0,1]$ (replicated if CS not fulfilled)

standard normal innovations, $n = 100$, 1000 runs

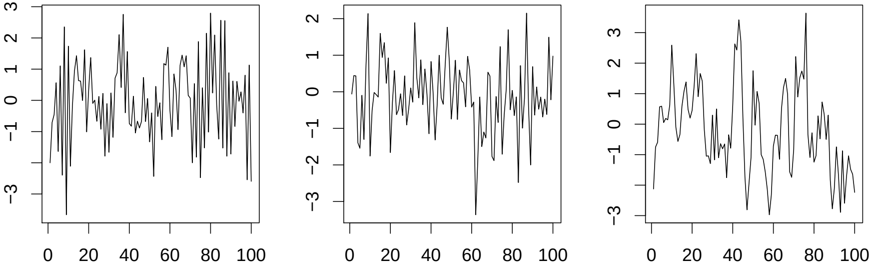


Figure: Three of the simulated AR time series

III : Simulation: AR model estimation without outliers

estimator	mean $\|\|\hat{\phi} - \phi\|\|_2$
emp. acf	0.14
mult. Tyler	0.16
mult. M	0.16
mult. S	0.16
GRK	0.16
part. GRK	0.16
effi MCD	0.16
weight. MCD	0.16
raw MCD	0.16
p. Kendall	0.17
effi GK	0.17
p. Spearman	0.17
filter acf	0.17
acf MM-reg.	0.17
Spearman	0.17
Kendall	0.17

estimator	mean $\|\|\hat{\phi} - \phi\|\|_2$
Tukey	0.17
p. Masarotto	0.17
GK Tau	0.17
Huber	0.17
GK Qn	0.18
Masarotto	0.18
MM-reg*	0.19
acf lts-reg.	0.2
acf median-reg.	0.2
GM*	0.2
filter ar*	0.21
quadrant	0.21
p. quadrant	0.21
GK MAD	0.21
median cor.	0.22
trimmed cor.	0.25

1000 random Gaussian time series as before

random outlier-type: $P(\text{block}) = P(\text{isolated}) = 1/2$

random number of outliers: $n_{\text{out}} \sim U\{1, 20\}$ at random positions

random outlier size: $s_i \sim N(0, 100)$

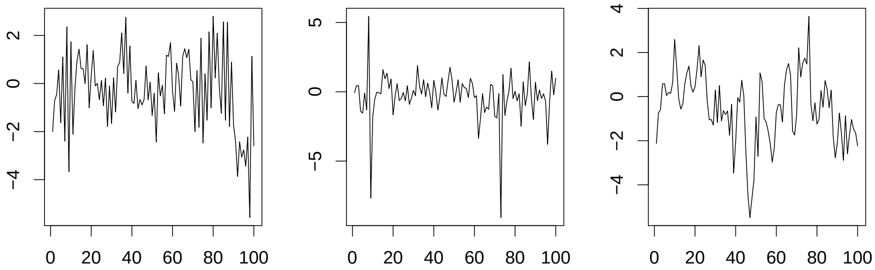


Figure: Three of the simulated AR time series with outliers

Simulation: AR model estimation with outliers

estimator	mean $\|\|\hat{\pi} - \pi\|\|_2$
filter acf	0.25
filter ar*	0.26
effi MCD	0.26
weight. MCD	0.26
raw MCD	0.26
multi S	0.27
Tukey	0.29
effi GK	0.29
GM*	0.29
GK Qn	0.30
GK Tau	0.30
GK MAD	0.30
acf lts-reg.	0.30
Spearman	0.30
p. Kendall	0.30
Kendall	0.31

estimator	mean $\|\|\hat{\phi} - \phi\|\|_2$
p. Quadrant	0.31
p. Spearman	0.31
Quadrant	0.31
median cor.	0.31
GRK	0.31
Huber	0.32
p. GRK	0.32
Masarotto	0.32
p. Masarotto	0.32
mult. Tyler	0.33
acf MM reg.	0.33
MM reg.*	0.33
trimmed cor.	0.33
multi M	0.34
acf median reg.	0.38
emp. acf	0.50

$\mathbb{III}$: Model selection

Classical: use partial autocorrelations or penalized criteria like

$$AIC = \log(\hat{\sigma}^2_{ML}) + 2p/n, \qquad BIC = \log(\hat{\sigma}^2_{ML}) + \log(n)p/n.$$

BIC asymptotically consistent.
AIC overfits in large samples but asymptotically efficient prediction.

Robust: use robust partial autocorrelations or penalized criteria like

$$AIC = \ln(\hat{\sigma}^2_R) + 2p/(n-p).$$

AIC default, option e.g. aicpenalty = function(p) log(length(data)-p) * p

Order $p = 1$ selected for simulated example except $p = 17$ by ar(method=" ols") for contaminated data and $p = 2$ by arrob(method="yw").

$\mathbb{III}$: Estimation of spectral densities

Dürre, Fried & Liboschik (2015): Robust estimation of (partial) autocorrelation, *WIREs Comput. Stat.* 7: 205-222.

$\mathbb{III}$: Spectral density

$(Y_t : t \in \mathbb{Z})$ second order stationary, observed at time points $t = 1, \ldots, n$.

$$S(f) = 1 + 2\sum_{h=1}^{\infty} \rho(h)\cos(2\pi fh), \quad 0 \leq f \leq 0.5$$

Contributions of sinus-cycles with different frequencies to process variance $\gamma(0)$.

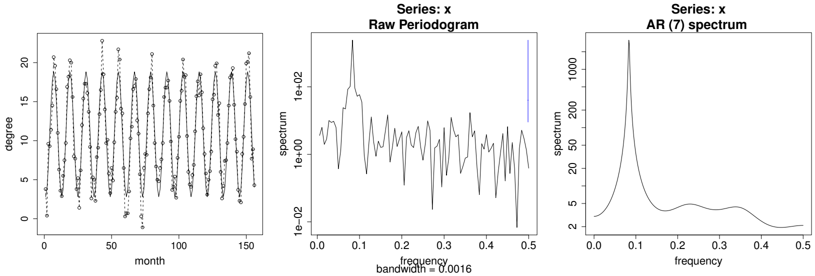
Bijective transformation of autocorrelation function.

For white noise with $\rho(h) = 0,\ h = 1, 2, \ldots$ we get $S(f) \equiv 1$.

Estimation by fitting sine curves with different frequencies ("periodogram") or from fitted autocorrelations.

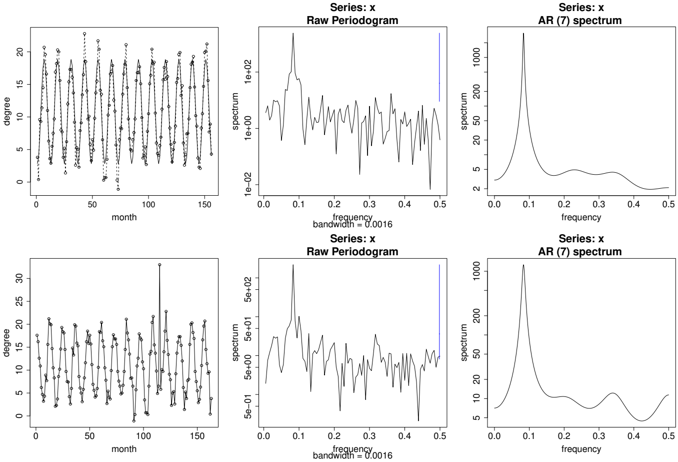
spectrum(method=c("pgram","ar"))

Monthly average temperature in Düsseldorf with fitted sine function (left) and its periodogram (center) or ar spectral estimate (right).



Periodogram indicates existence of a cycle with frequency 1/12, i.e., length 12.

Temperatures (top) and with a value of 33 instead of 3.3 (bottom): periodogram (center) or ar (right) spectral estimate.



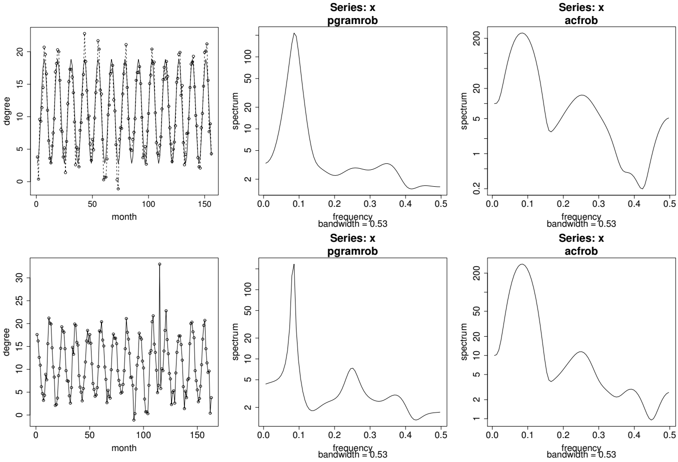
𝕀𝕀𝕀 : Robust estimation

spectrum(method=c("pgram")):
prewhitening estimator described by Maronna, Martin & Yohai (2006, ch. 8.14): fit AR-model robustly, transform prediction residuals by robust ψ-function, combine spectral density estimates from AR-fit and from transformed residuals.

spectrum(method=c("acf")):
estimate autocorrelations robustly and plug them into the above formula for $S(f)$.

III : Temperature data - robust analysis

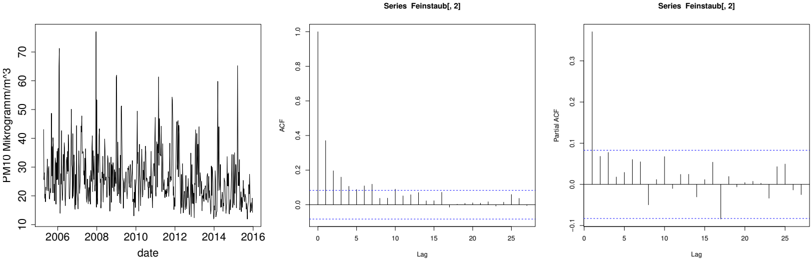
Temperatures (top) and with a value of 33 instead of 3.3 (bottom): periodogram (center) or acf (right) spectral estimate.



$\mathbb{IV}$: Robust change-point detection

Dehling, Fried & Wendler (2018+): A robust method for shift detection in time series.

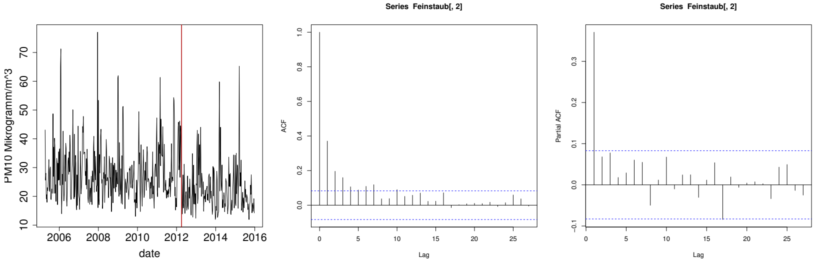
Weekly fine dust averages in Essen (Germany)



Sample (partial) autocorrelations point at stationary AR(1).

𝕀𝕍 : Example: fine dust

Weekly fine dust averages in Essen (Germany)

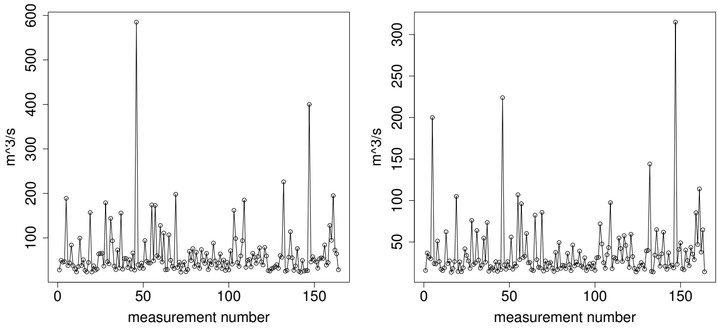


CUSUM-statistic for location shift: 3.57»1.36 (critical value at 5% significance), estimated changepoint early april 2012.

changerob(method="CUSUM")

𝕀𝕍 : Example: floods at the river Mulde

Floods in Aue (left) and Niederschlema (right).



Are there changes because of water engineering, etc.?
CUSUM-statistic for location shift: 0.58 for Aue, 0.82 for Niederschlema.

$(Y_t : t \in \mathbb{Z})$ 2nd order stationary, short-range dependent process.

$$
\begin{aligned}
T_C &= \max_{t=1,\ldots,n} \frac{t}{\sqrt{n}} \left| \frac{1}{t}\sum_{i=1}^{t} Y_i - \frac{1}{n}\sum_{j=1}^{n} Y_j \right| \\
&= \max_{t=1,\ldots,n} \frac{1}{n\sqrt{n}} \left| \sum_{i=1}^{t}\sum_{j=t+1}^{n} (Y_i - Y_j) \right| \\
&= \max_{t=1,\ldots,n} \frac{t(n-t)}{n\sqrt{n}} \left| \frac{1}{t}\sum_{i=1}^{t} Y_i - \frac{1}{n-t}\sum_{j=t+1}^{n} Y_j \right| \\
&\xrightarrow{H_0} \sigma_2 \sup_{0\le\lambda\le 1} |W(\lambda) - \lambda W(1)|, \\
\sigma_2^2 &= \sum_{h=-\infty}^{\infty} Cov(Y_1, Y_h)
\end{aligned}
$$

$\mathbb{IV}$: CUSUM-type tests

$(Y_t : t \in \mathbb{Z})$ 2nd order stationary, short-range dependent process.

$$
\begin{aligned}
T_1 &= \max_t \frac{t}{\sqrt{n}} \left| \hat{\theta}_{1:t} - \frac{1}{n} \hat{\theta}_{1:n} \right| \\
T_2 &= \max_t \frac{1}{n\sqrt{n}} \left| \sum_{i=1}^{t} \sum_{j=t+1}^{n} h(Y_i - Y_j) \right| \\
T_3 &= \max_t \frac{t(n-t)}{n\sqrt{n}} \left| \hat{\Delta}_{t:n-t} \right| \\
&\overset{?}{\longrightarrow} \sigma \sup_{0 \le \lambda \le 1} |W(\lambda) - \lambda W(1)|
\end{aligned}
$$

$\mathbb{IV}$: Implementation in changerob

$(Y_t : t \in \mathbb{Z})$ 2nd order stationary, short-range dependent process.

method="CUSUM": CUSUM-test

$$T_C = \max_t \frac{t}{\sqrt{n}} \left| \hat{\theta}_{1:t} - \frac{1}{n} \hat{\theta}_{1:n} \right|, \ \sigma_2^2 = \sum_{h=-\infty}^{\infty} Cov(Y_1, Y_h)$$

method="Wilcoxon": Change-test using Wilcoxon-Mann-Whitney

$$T_W = \max_t \frac{1}{n\sqrt{n}} \left| \sum_{i=1}^{t} \sum_{j=t+1}^{n} (1(Y_i - Y_j < 0) - 0.5) \right|, \ \sigma_1^2 = \sum_{h=-\infty}^{\infty} Cov(F(Y_1), F(Y$$

method="HL": Change-test using Hodges-Lehmann 2-sample estimator

$$T_{HL} = g(0) \max_t \frac{t(n-t)}{n\sqrt{n}} \left| med_{i \leq t < j} Y_i - Y_j \right|, \ \sigma_1^2 = \sum_{h=-\infty}^{\infty} Cov(F(Y_1), F(Y_h))$$

$\mathbb{IV}$: Estimation of long run variance

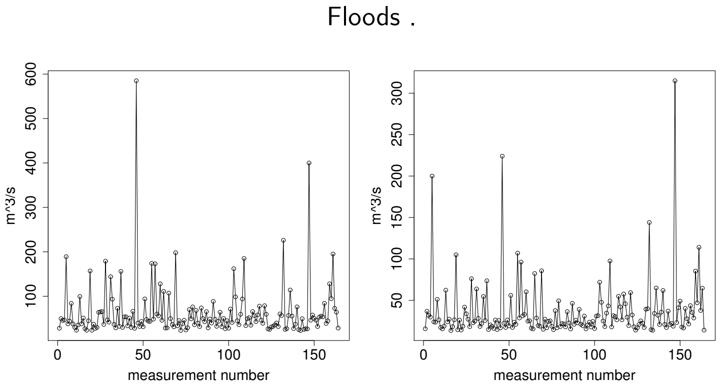
$(Y_t : t \in \mathbb{Z})$ 2nd order stationary, short-range dependent process.

changerob(var.method="window"): subsampling estimator
σ_2^2 and σ_1^2 are the asymptotical variances of $\overline{Y}_n$ and $\overline{F(Y)}_n$.
Idea: estimate σ_1^2 & σ_2^2 from variability of $\overline{Y}_m$ & $\overline{F(Y)}_m$ in subsamples of length $1 << m << n$.
Carlstein's (1986) rule for choice of m works well if AR(1) fits well.

changerob(var.method="acf"): kernel smoothing of autocorrelations
Idea: approximate sum of all autocovariances by weighted sum using Bartlett or flat-top kernel; bandwidth chosen such that estimates at larger lags are small.
Gives often somewhat conservative tests.

changerob(var.method="acfextra"): extrapolate first p autocorrelations
Idea: fit AR(p)-model to first p autocorrelations and extrapolate.
Works nicely if AR(p)-model fits well.

𝕀𝕍 : Example: floods in Aue (left) and Niederschlema (right)

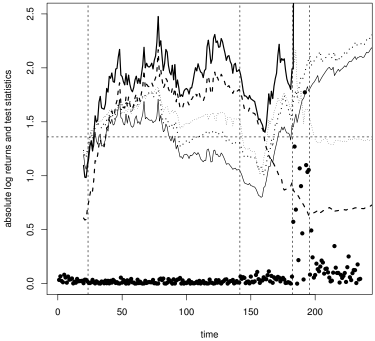


Are there changes because of water engineering, etc.?

CUSUM:	Aue 0.58	Niederschlema 0.82	0.80 under independence
Wilcoxon:	Aue 0.72	Niederschlema 1.09	1.23⋆ under independence
HL:	Aue 0.70	Niederschlema 1.27⋆	1.38⋆⋆ under independence

$\mathbb{IV}$: Example: VW absolute log returns in 2015

Change-test statistics calculated from first $m = 20, \ldots, 254$ observations:



CUSUM using ordinary (thin dots) or Huberized observations (bold dots), Wilcoxon (thin solid), 2- (bold solid) or 1-sample (bold dashes) HL statistic. Horizontal dashed line: critical value at 5% significance. Vertical lines mark changes detected by 2-sample HL-change-test with binary segmentation.

Constancy of the level of the absolute values means stability of the variance.
CUSUM-type statistics get confused by multiple shifts into opposite directions.
HL change-test still works well.

Summary

- I Estimation of (partial) autocorrelations
  (analysis of linear relationships)
- II Fitting autoregressive models
  (basic time series modeling)
- III Estimation of spectral densities
  (analysis of cyclic behavior)
- IV Change-point detection
  (tests for the basic assumptions)

Choice of suitable robust estimators and tests should be guided by knowledge of the data (e.g. discrete or continuous) and the expected pattern and sizes of outliers or change-points.